Individual Participant Data Meta-Analysis for a Binary Outcome: One-Stage or Two-Stage?

Thomas P. A. Debray, Karel G. M. Moons, Ghada Mohammed Abdallah Abo-Zaid, Hendrik Koffijberg, Richard David Riley

1 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands, 2 European Centre for Environment and Human Health, Peninsula College of Medicine and Dentistry, University of Exeter, at the Knowledge Spa, Royal Cornwall Hospital, Truro, Cornwall, United Kingdom, 3 School of Health and Population Sciences, and School of Mathematics, Public Health, Epidemiology and Biostatistics, College of Medical and Dental Sciences, University of Birmingham, Birmingham, United Kingdom

Abstract

Background: A fundamental aspect of epidemiological studies concerns the estimation of factor-outcome associations to identify risk factors, prognostic factors and potential causal factors. Because reliable estimates for these associations are important, there is a growing interest in methods for combining the results from multiple studies in individual participant data meta-analyses (IPD-MA). When there is substantial heterogeneity across studies, various random-effects meta-analysis models are possible that employ a one-stage or two-stage method. These are generally thought to produce similar results, but empirical comparisons are few.

Objective: We describe and compare several one- and two-stage random-effects IPD-MA methods for estimating factor-outcome associations from multiple risk-factor or predictor finding studies with a binary outcome. One-stage methods use the IPD of each study and meta-analyse using the exact binomial distribution, whereas two-stage methods reduce evidence to the aggregated level (e.g. odds ratios) and then meta-analyse assuming approximate normality. We compare the methods in an empirical dataset for unadjusted and adjusted risk-factor estimates.

Results: Though often similar, on occasion the one-stage and two-stage methods provide different parameter estimates and different conclusions. For example, the effect of erythema and its statistical significance was different for a one-stage (OR = 1.35, p = 0.03) and univariate two-stage (OR = 1.55, p = 0.12). Estimation issues can also arise: two-stage models suffer unstable estimates when zero cell counts occur and one-stage models do not always converge.

Conclusion: When planning an IPD-MA, the choice and implementation (e.g. univariate or multivariate) of a one-stage or two-stage method should be prespecified in the protocol as occasionally they lead to different conclusions about which factors are associated with outcome. Though both approaches can suffer from estimation challenges, we recommend employing the one-stage method, as it uses a more exact statistical approach and accounts for parameter correlation.
methods for meta-analysis and evidence synthesis in this area, in order to summarise the factor-outcome associations in the current evidence-base [2,14–20], as commonly applied in intervention research [21–25]. However, due to numerous problems of published primary studies investigating factor-outcome associations, especially publication bias and selective reporting, meta-analyses based on published results are notoriously prone to bias [0,10]. Problems with such aggregate data also arise in clinical research when differential treatment effects by patient characteristics are of concern [26]. Problems with such aggregate data also arise in clinical research when differential treatment effects by patient characteristics are of concern [26]. Problems with such aggregate data also arise in clinical research when differential treatment effects by patient characteristics are of concern [26].

Thus there is increasing interest in obtaining individual participant data (IPD) from these studies to facilitate a more reliable meta-analysis.

When IPD are available, meta-analysis is usually performed using a two-stage approach [24]. Each study is summarized by its factor-outcome association estimate and variance in the first stage, and these aggregate data (AD) are then appropriately combined across studies in the second stage. In this manner, a summary effect size, such as the odds or hazard ratio, is produced for each factor-outcome association of interest [27] whilst potentially accounting for between-study heterogeneity (e.g. due to different participant characteristics, methods of measurements, and undergone treatments) [12,19,28–35]. An alternative method for IPD meta-analysis (IPD-MA) is a one-stage approach which synthesises the IPD from all studies in a single step, whilst accounting for clustering of patients within studies [36–38]. Assuming the sufficient AD are obtained from each study for the two-stage method, it is widely believed that one-stage and two-stage methods lead to similar conclusions [39–41]; however, empirical comparisons are relatively few. Indeed, because the design and implementation of one-stage and two-stage random-effects models may substantially differ, it is important to ascertain whether the choice of method can influence the final conclusions about whether a factor has a (statistically) significant association with the outcome.

In a recent empirical evaluation using a meta-analysis of 24 randomised trials of antiplatelets to prevent preeclampsia, Stewart et al. [37] conclude that ‘two-stage and one-stage approaches to analysis produce similar results’ and ‘where an IPD review evaluates effectiveness based on sufficient data from randomised controlled trials, one-stage statistical analyses may not add much value to simpler two-stage approaches’. It is important to consider if this recommendation is valid in other empirical examples, and if it translates to epidemiological studies. In particular, epidemiological studies of factor-outcome associations may be affected by several covariates, namely confounders (in causal factor studies) or other predictors (in predictor finding studies) [33,42,43]. This situation may also arise in clinical trials when interactions occur between treatment effects and covariates, or when adjustment is needed for prognostic factors that are unbalanced between groups. Thus the random-effects framework needs to accommodate these covariates during modeling in order to estimate factor-outcome associations after adjusting for other factors. Factors that are strongly associated with the outcome might retain their association even when adjusting for other variables. However, there has again been little comparison of one-stage and two-stage IPD-MA methods when adjustment is required [44,45].

The aim of this article is to describe and empirically evaluate possible one-stage and two-stage IPD-MA models for synthesizing (causal or predictive) factor-outcome association estimates across multiple studies where a continuous or binary factor is of interest in relation to a binary outcome. It is therefore similar in spirit to a recent description of methods for meta-analysis of time-to-event outcomes [46]. The methods are compared using an empirical example, to illustrate their advantages, differences and accessibility. Our methods all assume that between-study heterogeneity in baseline risk and factor-outcome associations exists, as it likely in practice, and so we only consider random-effects IPD-MA models.

Motivating Example

Deep Vein Thrombosis (DVT) is a blood clot that forms in a vein in the body (usually in the lower leg or thigh). A (part of such) clot can break off and be carried through the bloodstream to the lungs and there cause a blockage (pulmonary embolism), preventing oxygenation of the blood and potentially causing death. The diagnosis DVT presence or absence can (ultimately) be made using repeated leg ultrasound, which requires patient referral and is to some extent burdening and time and money consuming. Hence, it is desirable to predict the presence or absence of DVT without having to refer patients for more cumbersome testing, by rather using easy to obtain predictors from their patient history, physical examination and simple blood assays. For this reason, in patients with a suspected DVT various studies aimed at estimating which factors – out of a range of candidate factors – are indeed associated with the presence or absence of DVT; in other words, which factors are useful diagnostic predictors of the probability that a patient truly has DVT.

A previous systematic review collected the IPD of patients with a suspected DVT from 13 studies (n=10002), and this IPD contains information about the patients’ history, physical examination and results from a biomarker test (Table 1 and Table S1) [47]. In this article, we use these data to illustrate the described meta-analysis methods for identifying important risk factors. We assume random effects for factor-outcome associations as the presence of heterogeneity between studies is expected due to differences in locale, setting and time. Detailed information about the included studies and predictors is available in Table S1 and Table S2.

Methods

This section describes the framework for random-effects IPD-MA modeling of risk factor (predictor finding) studies with a binary outcome. Hereto, it identifies two sources of data: IPD and AD. IPD is represented by patient-level factor values (covariates) and outcomes, whereas AD consists of study-level summaries such as the estimated log odds ratios and corresponding standard errors for the factor-outcome associations reported [48]. We describe two-stage and then one-stage IPD-MA approaches [49] and describe how to account for differences in baseline risk across studies (clustering). Further, we show how to extend these methods to adjust for known risk factors, and evaluate some important estimation difficulties that arise when relatively few data are available. The DVT data is used to illustrate the methods and to identify some important differences.
Two-stage IPD Methods

First stage. In a two-stage method, the IPD are first analyzed separately in each study using an appropriate statistical method for binary outcome data. For example, consider where a single risk factor is of interest, then the logistic regression model is:

$$y_i \sim \text{Bernoulli}(p_i) \quad (\text{Model } 1)$$

$$\logit(p_i) = \alpha + \beta x_i$$

with unknown parameters $\alpha$ (intercept) and $\beta$ (slope representing the association between factor $x$ and binary outcome $y$). The logit outcome probability for subject $i$, $p_i$, is then a linear function of the factor $x_i$. The resulting estimates from study $j$ are denoted as $\hat{\alpha}_j$ (intercept) and $\hat{\beta}_j$ (log odds ratio). Consequently, the first step yields the intercept and the factor-outcome association estimates, and their associated within-study covariance matrix (containing the variance of the intercept $\text{var}(\hat{\alpha}_j)$ and each association $\text{var}(\hat{\beta}_j)$, as well as their respective covariances $\text{cov}(\hat{\alpha}_j, \hat{\beta}_j)$) for each individual study. By utilising all the model parameter estimates, their variances and their correlation (covariance), the original IPD is reduced to AD for each study [50,51]. If IPD are not available, such AD may alternatively be sought from study publications or study authors.

Second Stage. Option 1. Full (bivariate) meta-analysis AD model: The AD are combined by a bivariate random-effects model that simultaneously synthesises the factor-outcome association (beta) estimates and the baseline risk (intercept) estimates whilst accounting for their correlation. The model assumes that the true underlying effect of the $j^{th}$ study (asymptotically) arises from a multivariate normal (MVN) distribution [53], and incorporates within- and between-study covariance. Specifically, the model fits the following marginal distributions:

$$\begin{bmatrix} \hat{\alpha}_j \\ \hat{\beta}_j \end{bmatrix} \sim \text{MVN}$$

$$(\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \begin{bmatrix} \tau_{\alpha}^2 & \tau_{\alpha\beta} \\ \tau_{\alpha\beta} & \tau_{\beta}^2 \end{bmatrix} + \begin{bmatrix} \text{var}(\hat{\alpha}_j) & \text{cov}(\hat{\alpha}_j, \hat{\beta}_j) \\ \text{cov}(\hat{\alpha}_j, \hat{\beta}_j) & \text{var}(\hat{\beta}_j) \end{bmatrix})$$

Table 1. Overview of the DVT data.

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>ddidm dich = 1</th>
<th>notraum = 1</th>
<th>coag = 1</th>
<th>eryt = 1</th>
<th>sex = 1</th>
<th>malign = 1</th>
<th>par = 1</th>
<th>tend = 1</th>
<th>leg = 1</th>
<th>caldiff3 = 1</th>
<th>pit = 1</th>
<th>vein = 1</th>
<th>alt-diagn = 1</th>
<th>surg = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1028(131)</td>
<td>472 (117)</td>
<td>743 (104)</td>
<td>19 (3)</td>
<td>382 (52)</td>
<td>376 (66)</td>
<td>54 (15)</td>
<td>13 (2)</td>
<td>562 (69)</td>
<td>232 (46)</td>
<td>311 (72)</td>
<td>704(108)</td>
<td>155 (28)</td>
<td>669 (22)</td>
<td>81 (11)</td>
</tr>
<tr>
<td>2</td>
<td>814(318)</td>
<td>598 (313)</td>
<td>169 (89)</td>
<td>353 (186)</td>
<td>419(196)</td>
<td>127 (57)</td>
<td>217 (58)</td>
<td>75 (34)</td>
<td>160 (97)</td>
<td>26 (10)</td>
<td>303 (106)</td>
<td>357(100)</td>
<td>43 (16)</td>
<td>448 (26)</td>
<td>168(45)</td>
</tr>
<tr>
<td>3</td>
<td>153 (26)</td>
<td>82 (19)</td>
<td>51 (20)</td>
<td>59 (19)</td>
<td>73 (22)</td>
<td>25 (8)</td>
<td>74 (3)</td>
<td>24 (10)</td>
<td>67 (64)</td>
<td>22 (7)</td>
<td>169 (78)</td>
<td>193 (70)</td>
<td>31 (9)</td>
<td>44 (2)</td>
<td>14 (4)</td>
</tr>
<tr>
<td>4</td>
<td>1756(411)</td>
<td>664 (238)</td>
<td>607 (251)</td>
<td>426 (210)</td>
<td>950(272)</td>
<td>283 (92)</td>
<td>906 (92)</td>
<td>197(87)</td>
<td>132 (69)</td>
<td>44 (15)</td>
<td>125 (55)</td>
<td>238 (58)</td>
<td>34 (13)</td>
<td>30 (3)</td>
<td>23 (5)</td>
</tr>
<tr>
<td>5</td>
<td>791 (126)</td>
<td>572 (90)</td>
<td>322 (79)</td>
<td>490 (74)</td>
<td>155 (32)</td>
<td>300 (43)</td>
<td>105(25)</td>
<td>23 (9)</td>
<td>161 (31)</td>
<td>47 (21)</td>
<td>93 (33)</td>
<td>97 (29)</td>
<td>39 (9)</td>
<td>114 (8)</td>
<td>16 (6)</td>
</tr>
<tr>
<td>6</td>
<td>1075(190)</td>
<td>494 (118)</td>
<td>217 (67)</td>
<td>303 (106)</td>
<td>357(100)</td>
<td>43 (16)</td>
<td>448 (26)</td>
<td>168(45)</td>
<td>107 (61)</td>
<td>75 (18)</td>
<td>176 (32)</td>
<td>43 (16)</td>
<td>154 (25)</td>
<td>42 (9)</td>
<td>16 (6)</td>
</tr>
<tr>
<td>7</td>
<td>429 (61)</td>
<td>203 (41)</td>
<td>30 (13)</td>
<td>96 (33)</td>
<td>87 (24)</td>
<td>33 (5)</td>
<td>176 (17)</td>
<td>25 (6)</td>
<td>129 (208)</td>
<td>583 (164)</td>
<td>556 (194)</td>
<td>799(193)</td>
<td>257 (82)</td>
<td>782 (98)</td>
<td>181(54)</td>
</tr>
<tr>
<td>8</td>
<td>325 (52)</td>
<td>161 (31)</td>
<td>47 (21)</td>
<td>93 (33)</td>
<td>97 (29)</td>
<td>39 (9)</td>
<td>114 (8)</td>
<td>16 (6)</td>
<td>814 (318)</td>
<td>562 (69)</td>
<td>232 (46)</td>
<td>311 (72)</td>
<td>704 (108)</td>
<td>155 (28)</td>
<td>669 (22)</td>
</tr>
<tr>
<td>9</td>
<td>1295(289)</td>
<td>924 (208)</td>
<td>583 (164)</td>
<td>556 (194)</td>
<td>799(193)</td>
<td>257 (82)</td>
<td>782 (98)</td>
<td>181(54)</td>
<td>1028 (131)</td>
<td>562 (69)</td>
<td>232 (46)</td>
<td>311 (72)</td>
<td>704 (108)</td>
<td>155 (28)</td>
<td>669 (22)</td>
</tr>
<tr>
<td>10</td>
<td>436 (42)</td>
<td>222 (28)</td>
<td>168 (28)</td>
<td>66 (13)</td>
<td>91 (18)</td>
<td>1 (0)</td>
<td>119 (5)</td>
<td>58 (9)</td>
<td>814 (318)</td>
<td>562 (69)</td>
<td>232 (46)</td>
<td>311 (72)</td>
<td>704 (108)</td>
<td>155 (28)</td>
<td>669 (22)</td>
</tr>
<tr>
<td>11</td>
<td>541 (121)</td>
<td>239 (74)</td>
<td>152 (67)</td>
<td>162 (63)</td>
<td>270 (86)</td>
<td>38 (16)</td>
<td>313 (22)</td>
<td>96 (34)</td>
<td>1295 (289)</td>
<td>924 (208)</td>
<td>583 (164)</td>
<td>556 (194)</td>
<td>799(193)</td>
<td>257 (82)</td>
<td>782 (98)</td>
</tr>
<tr>
<td>12</td>
<td>550 (55)</td>
<td>176 (21)</td>
<td>83 (16)</td>
<td>114 (30)</td>
<td>251 (40)</td>
<td>28 (8)</td>
<td>245 (16)</td>
<td>39 (7)</td>
<td>1028 (131)</td>
<td>562 (69)</td>
<td>232 (46)</td>
<td>311 (72)</td>
<td>704 (108)</td>
<td>155 (28)</td>
<td>669 (22)</td>
</tr>
<tr>
<td>13</td>
<td>809 (42)</td>
<td>258 (22)</td>
<td>75 (8)</td>
<td>153 (18)</td>
<td>196 (17)</td>
<td>32 (3)</td>
<td>399 (9)</td>
<td>45 (4)</td>
<td>1295 (289)</td>
<td>924 (208)</td>
<td>583 (164)</td>
<td>556 (194)</td>
<td>799(193)</td>
<td>257 (82)</td>
<td>782 (98)</td>
</tr>
</tbody>
</table>

Observed factor level counts (for which dvt = 1) for binary risk factors in each study of the DVT case study. Entries are left blank for studies that did not measure the corresponding factor.

doi:10.1371/journal.pone.0060650.t001
with unknown parameters $\alpha$, $\beta$, $\tau_\alpha$, $\tau_\beta$ and $\tau_{ab}$. Here, $\alpha$ and $\beta$ represent the average baseline risk and factor-outcome association across studies, respectively, $\tau_\alpha$ and $\tau_\beta$ describe their respective degree of heterogeneity between studies, and $\tau_{ab}$ their between-study covariance.

**Option 2. Traditional (univariate) meta-analysis AD model:** Most researchers ignore within-study and between-study covariances in parameter estimates and thus assume that $\text{cov}(\gamma_j,\beta_j)$ and $\tau_{ab}$ equal 0 [51]. Essentially, this reduces Model 1 to a univariate meta-analysis of the factor-outcome association, and is similar to the commonly applied DerSimonian and Laird's classical random-effects meta-analysis model [21,54], where:

\[
\beta_j \sim N\left(\beta, \tau_\beta^2 + \text{var}(\beta_j)\right)
\]

(Model 2)

with unknown parameters $\beta$ and $\tau_\beta$. This model no longer synthesises the baseline risk across studies, and just pools the factor-outcome associations.

**One-stage Methods**

In a one-stage method, the IPD from all studies are modeled simultaneously whilst accounting for the clustering of subjects within studies. The one-stage IPD-MA framework is a (multilevel) logistic regression model with random effects. Different specifications are possible, as now described.

**Option 1. Fully (bivariate) random-effects one-stage model:** Here, as in Model 1, random effects are specified for both the intercept and the slope, and their between-study covariance is modelled

\[
y_{ij} \sim \text{Bernoulli}(p_{ij}) \quad \logit(p_{ij}) = x_i + \beta_j x_{ij} \quad \text{(Model 3)}
\]

\[
\begin{bmatrix}
\gamma_i \\
\beta_j
\end{bmatrix} \sim \text{MVN}\left(
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix},
\begin{bmatrix}
\tau_\alpha^2 & \tau_\alpha \tau_{ab} \\
\tau_\alpha \tau_{ab} & \tau_{ab}^2, \tau_\beta^2
\end{bmatrix}
\right)
\]

where $i$ indicates observations at the individual level and $j$ again represents the study level. Note that $\gamma_i$ and $\beta_j$ are not explicitly estimated (in contrast to Model 1, where it represents the AD from the individual studies) but follow from the unknown parameters $\alpha$, $\beta$, $\tau_\alpha$, $\tau_\beta$ and $\tau_{ab}$. These parameters have the same interpretation as those from Model 1.

**Option 2. Reduced random-effects one-stage model:** In a reduced one-stage model, independent random effects are assumed for the intercept and slope in order to avoid estimating the between-study covariance, which can often be problematic:

\[
y_{ij} \sim \text{Bernoulli}(p_{ij}) \quad \logit(p_{ij}) = x_i + \beta_j x_{ij} \quad \text{(Model 4)}
\]

\[
\begin{bmatrix}
\gamma_i \\
\beta_j
\end{bmatrix} \sim \text{MVN}\left(
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix},
\begin{bmatrix}
\tau_\alpha^2 & 0 \\
0 & \tau_\beta^2
\end{bmatrix}
\right)
\]

**Option 3. Stratified one-stage model:** Finally, it is possible to reduce the number of assumptions by estimating a stratified one-stage model. This model no longer estimates an underlying average for the intercepts but rather estimates a separate intercept for each study. Thus the between-study normality assumption for the intercept term is no longer required for $\gamma_i$, and there is no need to estimate a between-study covariance term. However, heterogeneity in the the factor-outcome association is still modelled using a random effect:

\[
y_{ij} \sim \text{Bernoulli}(p_{ij}) \\
\logit(p_{ij}) = \sum_{m=1}^{M} (x_{im} I_{m-j}) + \beta_j x_{ij} \quad \text{(Model 5)}
\]

\[
\beta_j \sim N\left(\beta, \tau_\beta^2\right)
\]

where the indicator term $I_{m-j}$ indicates that a separate intercept should be estimated for each study $j = 1, \ldots, M$. Similar to Model 3 and Model 4, $\beta_j$ is not explicitly estimated but follows from the unknown parameters $\alpha_1, \ldots, \alpha_M$, $\beta$ and $\tau_\beta$.

**Extending the One-stage and Two-stage Models to Examine Multiple Risk Factors**

Previously, we described models for summarizing unadjusted factor-outcome associations. Although these models are fairly straightforward to implement, it is well known that factor-outcome associations are often influenced by extraneous variables rendering exposure groups incomparable. This situation may, for instance, arise when associations are estimated from cohort and cross-sectional studies (prognostic research) or treatment-by-patient-characteristic interactions occur (intervention research). In addition, several authors have recommended that each factor should be studied for their incremental (causal or predictive) value beyond established risk factors [55,56]. This raises the need for multivariable analyses, where the factor-outcome association under investigation is adjusted for potential confounders or other known predictors. Consequently, the methods from previous section performing a univariate (or bivariate) meta-analysis need to be extended to perform a (multivariate) meta-analysis where the factor-outcome associations (and intercept) are adjusted for $K$ additional factors.

**Extended two-stage models.** For the two-stage method, multivariable logistic regression models are estimated in each study:

\[
y_i \sim \text{Bernoulli}(p_i) \\
\logit(p_i) = x_i + \sum_{k=1}^{K} \theta_k z_{ik} \quad \text{(Model 2.2)}
\]

which yields an intercept $\hat{\gamma}_i$, a risk factor-outcome association $\hat{\beta}_i$, confounder-outcome associations $\hat{\theta}_1, \ldots, \hat{\theta}_K$ and a within-study covariance matrix $\Sigma_j$ for each study. A summary estimate for the regression coefficients and model intercept can be obtained by extending the bivariate random-effects model from Model 1 into a multivariate generalization [43,52,57,58].
Extended one-stage models. The fully random-effects one-stage model with multiple risk factors is specified as follows:

\[
y_j \sim \text{Bernoulli}(p_{y_j}) \\
\logit(p_{y_j}) = x_j + \beta_j x_j + \sum_{k=1}^{K} (\theta_j z_{y_j})_k
\]

\[
\begin{bmatrix}
\hat{\alpha}_j \\
\hat{\beta}_j \\
\hat{\theta}_j \\
\vdots \\
\hat{\theta}_K
\end{bmatrix} \sim \text{MVN}
\]

\[
\begin{bmatrix}
x \\
\beta \\
\theta_1 \\
\vdots \\
\theta_K
\end{bmatrix} \\
\begin{bmatrix}
\tau_x^2 & \tau_{x\beta} & \tau_{x\theta_1} & \cdots & \tau_{x\theta_K} \\
\tau_{x\beta} & \tau_{\beta}^2 & \tau_{\beta\theta_1} & \cdots & \tau_{\beta\theta_K} \\
\tau_{x\theta_1} & \tau_{\theta_1\theta_1} & \tau_{\theta_1\theta_2} & \cdots & \tau_{\theta_1\theta_K} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\tau_{x\theta_K} & \tau_{\theta_K\theta_1} & \tau_{\theta_K\theta_2} & \cdots & \tau_{\theta_K\theta_K}
\end{bmatrix} + \Sigma_j
\]  

(Model A)

Usually researchers assume zero within-study and between-study correlation, and so perform a separate univariate meta-analysis to each factor-outcome and confounder-outcome association separately; that is Model 2 is fitted for each of the log odds ratio terms separately (Model B).

Alternatively, a reduced one-stage model can be estimated by assuming independent random effects for \(x, \beta, \theta_1, \ldots, \theta_K\), i.e. the off-diagonal terms in Model - C are set to 0 (Model D).

Finally, it is possible to reduce the number of random effects by stratifying the intercepts and/or predictors for which a summary estimate is not of interest. For example, one-stage stratified model that estimates a separate intercept for each study can be achieved as follows:

\[
y_j \sim \text{Bernoulli}(p_{y_j}) \\
\logit(p_{y_j}) = x_j + \beta_j x_j + \sum_{k=1}^{K} (\theta_j z_{y_j})_k
\]

\[
\begin{bmatrix}
\hat{\beta}_j \\
\hat{\theta}_j \\
\end{bmatrix} \sim \text{MVN}
\]

\[
\begin{bmatrix}
x \\
\beta \\
\theta_1 \\
\vdots \\
\theta_K
\end{bmatrix} \\
\begin{bmatrix}
\tau_{\beta}^2 & \tau_{\beta\theta_1} & \tau_{\beta\theta_2} & \cdots & \tau_{\beta\theta_K} \\
\tau_{\beta\theta_1} & \tau_{\theta_1\theta_1} & \tau_{\theta_1\theta_2} & \cdots & \tau_{\theta_1\theta_K} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\tau_{\beta\theta_K} & \tau_{\theta_K\theta_1} & \tau_{\theta_K\theta_2} & \cdots & \tau_{\theta_K\theta_K}
\end{bmatrix}
\]  

(Model E)

Stratification on all confounders may, however, not always be feasible due to sample size constraints. For this reason, we generally recommend to model separate intercept terms and to assume random effects for all predictor effects (and hence reduce model complexity by introducing additional assumptions). The underlying rationale is that accurate estimates for confounding parameters are usually not required. Although this simplification may introduce bias in all parameter estimates, baseline risks are likely most affected because they capture all unexplained variation. A non-parametric modeling approach for the intercept terms may thus better accommodate model misspecification.

Estimation Procedures and Approximations

In the two-stage methods, the first stage model (logistic regression in each study) is estimated using maximum likelihood (ML). In the second stage, the AD meta-analysis models are estimated using, for example, methods of moment (MOM) or restricted maximum likelihood (REML) [21,52,54,59,60]. This can be implemented in numerous software, with packages such as lme4 in R, Proc Mixed in SAS and mmeta in STATA. However, difficulties may arise in the first or second stage estimation. For risk factors that are binary, if zero cell counts occur in some of the included studies (e.g. when all patients with the risk factor presence also have the outcome), the likelihood function may not converge or converges in an unstable factor-outcome association. This problem is also known as [partial] separation [61,62], and can be overcome by penalization [63–67] or adding a continuity correction [68,69]. A second problem may arise when the number of included studies is small as estimation of between-study covariance may become problematic [43,52,70].

One-stage methods involve the estimation of a mixed effects (multilevel) model which is often high dimensional [67]. For this reason, numerical integration is often achieved through approximate methods such as adaptive Gauss-Hermite Quadrature [29,71–73]. Although estimation becomes more precise as the number of quadrature points increases, it often gives rise to computational difficulties and convergence problems [74]. Furthermore, it has been demonstrated that the one-stage method may yield (downwardly) biased variance parameters when studies are small or limited in number [29,75–77]. The one-stage method may also produce downwardly biased coefficient estimates when an incorrect model is specified, for instance when random effects are wrongly assumed [78]. This may increase type-II errors. Although these issues could be reduced by penalization, there is a lack of REML procedures due to the computational difficulty of the second-order Laplace approximation [75].
**Case Studies**

In this section, we illustrate the benefits, limitations, and differences of one-stage and two-stage methods in the DVT data. For all case studies, in the two-stage models we used MLE in the first stage and MLE, REML or MOM in the second stage. For the one-stage models we used adaptive Gauss-Hermite Quadrature with 1 (Laplacian approximation) and 5 quadrature points.

In the first case study, we performed meta-analyses to estimate the unadjusted factor-outcome association for 16 risk factors using each of the models described above, and we examined the obtained log odds ratio ($\hat{\beta}$), standard error (S.E.), between-study variability ($\tau^2$), and between-study correlation ($\rho_{ab}$). The models considered are: full bivariate two-stage meta-analysis (Model 1), traditional univariate two-stage meta-analysis (Model 2), fully random-effects one-stage meta-analysis (Model 3), reduced random-effects one-stage meta-analysis (Model 4) and stratified one-stage meta-analysis (Model 5). For two-stage methods, we penalized the likelihood using Jeffreys invariant prior in datasets with (partial) separation in order to stabilize study-specific estimates [63,64].

In the second case study, we performed meta-analyses to investigate the risk factor *dmdich* adjusted for 5 covariates (*malign*, *surg* and *calffb*). Here, we estimated the following models: extended full two-stage model (Model 1+Model 2), extended reduced two-stage model (Model 2+Model 3), extended full one-stage model (Model 3), extended reduced one-stage model (Model 4) and extended stratified one-stage model (Model 5).

For all models, we calculated $p$-values (with $\alpha=0.05$) and corresponding 95% confidence intervals for the estimated odds ratios, according to:

$$\hat{\beta} \pm z^* \sqrt{SE(\hat{\beta})^2}$$

where $z^*$ is the 0.975 percentile of the standardized normal distribution. Finally, we calculated 95% prediction intervals to indicate a range for the predicted odds ratio in a new study [32,34]. Assuming the random effects are normally distributed with between-study standard deviation, then an approximate 95% prediction interval for the factor-outcome association in an unspecified study can be obtained as:

$$\hat{\beta} \pm t_{M-2}^* \sqrt{\tau^2 + \frac{1}{M} + SE(\hat{\beta})^2}$$

where $\hat{\beta}$ is the estimate of the average factor-outcome association across studies, and $t_{M-2}^*$ is the 0.975 percentile of the Student’s $t$ distribution with $M-2$ degrees of freedom, where $M$ is the number of studies in the meta-analysis.

All models were implemented in R 2.15.1 using Linux Mint 14 Nadia (MATE 64-bit) and incorporated the packages *lme4* (v0.999999-0), *mmeta* (v0.3.4), *logistf* (v1.10) and *metanmisc* (v0.0.4). Additional source code is available in Supporting Information S1.

**Results**

**One-stage Versus Two-stage Methods**

Results in Table 2 and Table S3 indicate that one- and two-stage methods often yield similar estimates for pooled factor-outcome associations, but importantly not always. For example, for the factor *par* we found an odds ratio of 1.45 (Model~1 using MLE) versus 1.32 (Model~5 using MLE). Occasionally, differences led to the one-stage and two-stage models disagreeing upon statistical significance (e.g. *eet*). These differences mainly occurred when relatively few data were available per study (*coag* and *par*), or relatively few studies were at hand (*eet* and *ddimdich*). For instance, the OR of *eet* was 1.52 (95% CI: 0.93 to 2.47) for the univariate two-stage approach (using DerSimonian and Laird’s MOM estimator), versus 1.35 (95% CI: 1.03 to 1.77) for the stratified one-stage approach. Furthermore, one-stage and two-stage methods tend to provide different estimates for standard errors and between-study heterogeneity parameters, leading to different prediction intervals. For instance, the prediction interval for the odds ratio of *ddimdich* ranged from 8.65 to 36.20 (Model 2 using MLE), versus 14.24 to 24.17 (Model 5 using MLE). Although usually they give similar results, the univariate two-stage method (Model 2) sometimes obtains different conclusions to the bivariate two-stage method (Model 1). For instance, for *eet* we respectively found an odds ratio of 1.55 ($p=0.115$) versus 1.38 ($p=0.043$) when REML was used as estimation procedure. Finally, the bivariate two-stage method (Model 1) often gives more similar results to the one-stage method. For the factor *eet*, we found OR = 1.37 with $p=0.036$ using Model 1 (bivariate two-stage model), versus OR = 1.37 with $p=0.037$ for Model 3 (bivariate one-stage model), OR = 1.39 with $p=0.046$ for Model 4 (reduced one-stage model) and OR = 1.35 with $p=0.029$ for Model 5 (stratified one-stage model). These estimates were all somewhat different to the results for Model 2 (univariate two-stage MoM) where OR = 1.52 with $p=0.094$.

**Estimation of Correlation between Random Effects**

As previously described, only the full one- and two-stage models (Model 1 & Model 3) estimate a parameter for the correlation between random effects. Results in Table 2 demonstrate that these models often yield correlation estimates that are close to +1 or −1, particularly when insufficient data are available and MLE is used. If correlations between random effects are assumed zero (Model 2 & Model 4), we noticed that parameter estimates may considerably change and thereby affect the calculation of $p$-values and prediction intervals. A good example is the unadjusted factor *coag*, where the prediction interval for the OR ranged from 0.62 to 2.47 (Model 1 with MLE) versus 0.75 to 2.23 (Model 2 with MLE), and the corresponding $p$-value decreased from 0.172 to 0.078. Similar findings were obtained for the adjusted analyses (Table 2). Finally, results indicate that the estimated correlation between random effects tends to be less extreme when REML is used (Table 3). The factor *surg* is a good example, as $\rho_{ab}$ decreased from −0.90 (MLE) to −0.65 (REML).

**Estimation of Stratified Models**

It is possible to avoid estimating correlation between random effects without assuming independence by using a stratified one-stage model, for example where a separate intercept is estimated for each study (Model 5) and, in the adjusted analyses, where predictors not of key interest are also stratified. Results indicate that the estimation of a separate intercept for each study (Model 5) tends to decrease the standard errors and between-study heterogeneity of factor-outcome associations (unless between-study correlations are +1 or −1). This, in turn, resulted in smaller prediction intervals for estimated odds ratios. For instance, the prediction interval for the unadjusted OR of *dmdich* ranged from 8.08 to 38.78 (Model 3), versus 14.24 to 24.17 (Model 5).

**Estimation of One-stage Models**

One-stage models were estimated with 1 and 5 quadrature points, and sometimes suffered from convergence problems (e.g. *par* and *coag* in Table 2 where positive indefiniteness occurred.
**Table 2.** Estimated unadjusted factor-outcome associations for the DVT case study.

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Model</th>
<th>Estimation</th>
<th>$\beta$</th>
<th>S.E.($\beta$)</th>
<th>OR</th>
<th>95% CI</th>
<th>95% PI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0achst (4^4)</td>
<td>3</td>
<td>MLE</td>
<td>0.21</td>
<td>0.16</td>
<td>0.22</td>
<td>0.98</td>
<td>1.24</td>
<td>0.91 to 1.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>REML</td>
<td>0.22</td>
<td>0.17</td>
<td>0.29</td>
<td>0.82</td>
<td>1.24</td>
<td>0.88 to 1.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MLE</td>
<td>0.26</td>
<td>0.15</td>
<td>0.15</td>
<td>1.29</td>
<td>1.29</td>
<td>0.97 to 1.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>REML</td>
<td>0.26</td>
<td>0.16</td>
<td>0.23</td>
<td>1.29</td>
<td>1.29</td>
<td>0.94 to 1.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MOM</td>
<td>0.26</td>
<td>0.16</td>
<td>0.21</td>
<td>1.29</td>
<td>1.29</td>
<td>0.95 to 1.76</td>
</tr>
<tr>
<td>%ddimisch (6)</td>
<td>3</td>
<td>MOM</td>
<td>0.28</td>
<td>0.13</td>
<td>0.21</td>
<td>1.33</td>
<td>1.33</td>
<td>1.03 to 1.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MLE</td>
<td>0.28</td>
<td>0.13</td>
<td>0.19</td>
<td>1.32</td>
<td>1.32</td>
<td>1.03 to 1.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MLE</td>
<td>0.32</td>
<td>0.15</td>
<td>0.10</td>
<td>1.37</td>
<td>1.37</td>
<td>1.02 to 1.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MLE</td>
<td>0.32</td>
<td>0.16</td>
<td>0.13</td>
<td>1.38</td>
<td>1.38</td>
<td>1.01 to 1.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MLE</td>
<td>0.30</td>
<td>0.14</td>
<td>0.00</td>
<td>1.35</td>
<td>1.35</td>
<td>1.03 to 1.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MLE</td>
<td>0.44</td>
<td>0.28</td>
<td>0.39</td>
<td>1.55</td>
<td>1.55</td>
<td>0.90 to 2.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MOM</td>
<td>0.42</td>
<td>0.25</td>
<td>0.33</td>
<td>1.52</td>
<td>1.52</td>
<td>0.93 to 2.47</td>
</tr>
<tr>
<td>0eryt (3)</td>
<td>3</td>
<td>MLE</td>
<td>0.31</td>
<td>0.15</td>
<td>0.10</td>
<td>1.37</td>
<td>1.37</td>
<td>1.02 to 1.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MLE</td>
<td>0.31</td>
<td>0.15</td>
<td>0.10</td>
<td>1.37</td>
<td>1.37</td>
<td>1.02 to 1.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MLE</td>
<td>0.33</td>
<td>0.17</td>
<td>0.14</td>
<td>1.39</td>
<td>1.39</td>
<td>1.01 to 1.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MLE</td>
<td>0.33</td>
<td>0.17</td>
<td>0.14</td>
<td>1.39</td>
<td>1.39</td>
<td>1.01 to 1.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MLE</td>
<td>0.30</td>
<td>0.14</td>
<td>0.00</td>
<td>1.35</td>
<td>1.35</td>
<td>1.03 to 1.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MLE</td>
<td>0.30</td>
<td>0.14</td>
<td>0.00</td>
<td>1.35</td>
<td>1.35</td>
<td>1.03 to 1.77</td>
</tr>
<tr>
<td>0hacht (4^4)</td>
<td>3</td>
<td>MLE</td>
<td>0.08</td>
<td>0.17</td>
<td>0.18</td>
<td>−1.00</td>
<td>1.09</td>
<td>0.77 to 1.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MLE</td>
<td>−0.03</td>
<td>0.15</td>
<td>0.00</td>
<td>0.98</td>
<td>0.98</td>
<td>0.73 to 1.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MLE</td>
<td>−0.03</td>
<td>0.15</td>
<td>0.00</td>
<td>0.97</td>
<td>0.97</td>
<td>0.72 to 1.30</td>
</tr>
</tbody>
</table>

Note: OR = Odds Ratio; 95% CI = 95% Confidence Interval; 95% PI = 95% Prediction Interval; p-value = p-value for the test of the null hypothesis that the true parameter is zero.
when 5 quadrature points were used). Possibly, these problems are related to poor model specification. Parameter estimates were similar for 1 and 5 quadrature points in the unadjusted analyses, however, some small differences occurred in the adjusted analyses (e.g. ddimdich in Table 3).

Discussion

We have described several random-effects IPD-MA models that implement a one-stage or two-stage method, where one desires to evaluate a potential causal (risk) factor or predictor of outcome. We detailed how they can be estimated and also extended to adjust for other factors. Despite the conventional belief that one-stage and two-stage methods yield similar conclusions [35–37], our empirical investigation shows that this is not always the case. Specifically, we found that different estimates for pooled effects, standard errors, between-study heterogeneity and correlation between random effects can result from choosing a different method (one-stage or two-stage), choosing a different estimation procedure (MLE, REML, MOM, number of quadrature points) and choosing a different model specification (independent random effects, joint random effects, stratified estimation). Although these differences were usually not substantial, in the DVT example they lead to discrepancies concerning the statistical significance of age, duration of symptoms, family history of thrombofilia, presence of erythema, presence of paresis and (dichotomized) D-dimer value.

Thus, importantly the choice of IPD-MA method may actually influence the conclusions about which factors are thought to be risk factors. This makes it desirable to pre-specify in a study protocol what meta-analysis method will be used, to avoid unjustified post-hoc analyses being performed to achieve statistical significance. We generally recommend that the one-stage method should be used. This method models the exact binomial distribution of the data in each study, and does not require a continuity correction when (partial) separation occurs [61–64,67]. The one-stage method may therefore produce more reliable results than the two-stage method when few studies or few subjects per study are available, as the two-stage method incorrectly assumes asymptotic normality (for the log odds ratio estimates from each study) in such scenarios [67]. The one-stage method further facilitates the adjustment for other factors, which is particularly important in non-randomised settings. In addition, one-stage models are more flexible, for example making the implementation of non-linear associations and interactions straightforward [24,37,79–82]. Finally, stratification in one-stage models avoids the need for estimating correlations between random effects. One can simply estimate study-specific intercepts and slopes and place the random effect only on the factor of interest.

Although we focused on IPD-MA of prognostic factors in this article, the two-stage methods can also be applied when only AD data is available for the included studies. These methods are usually preferred because sharing of IPD is often unfeasible due to,
for instance, confidentiality agreements. Results from our empirical example demonstrate that the full two-stage model, which pooling the AD accounts for heterogeneity of baseline risk and risk factors, their within-study and between-study correlation, tends to yield most consistent results with the one-stage model. The full two-stage method is a bivariate meta-analysis, which by additionally using the correlation between parameter estimates, is known to have benefits over a univariate me-analysis [43]. The methods presented here could further be extended using methods allowing for the combination of IPD with AD [49,83,84]. Potential limitations such as missing data in a subset of studies could be overcome using imputation methods that account for clustering. A Bayesian approach would be the most promising, as it would permit specification of the imputation model alongside the one-stage model, resolving some estimation limitations of the current approaches [32,85,86]. Furthermore, Bayesian approaches facilitate sensitivity analyses through adjusting prior specification, and permit the the robustness of fitted models to be evaluated. This is particularly useful when few studies are available and estimated parameters of one- and two-stage models may be severely biased due to estimation difficulties. Future research is needed to evaluate the performance of the described methods, and to compare their accuracy and coverage with Bayesian alternatives.

In summary, the choice of one-stage or two-stage method for performing a random-effects IPD-MA may influence the statistical identification of risk factors (predictors) for a binary outcome. When the number of studies in the meta-analysis are large and the number of events in each study are not few, we agree with Stewart et al [37] that a two-stage method will usually suffice. However, we generally recommend that a one-stage IPD-MA method is used as this models the exact binomial distribution, accounts for within-study parameter correlation, offers more flexibility in the model specification and avoids continuity corrections. It is therefore particularly preferable when few studies or few events in some studies are available.

Supporting Information

Supporting Information S1 Full model specifications and R code for implementation. (GZ)

Table S1 Overview of the DVT datasets. (PDF)

Table S2 Overview of the variables in the DVT datasets. (PDF)

Table S3 Additional results Case Study. (PDF)

Author Contributions

Conceived and designed the experiments: TPD KGM GAZ HK RDR. Performed the experiments: TPD RDR. Analyzed the data: TPD KGM HK RDR. Contributed reagents/materials/analysis tools: TPD KGM GAZ HK RDR. Wrote the paper: TPD KGM GAZ HK RDR.

References


