



**RESEARCH ARTICLE**

# Multistep estimators of the between-study variance: The relationship with the Paule-Mandel estimator

Robbie C. M. van Aert<sup>1</sup>  | Dan Jackson<sup>2</sup> <sup>1</sup>Methodology and Statistics, Tilburg University, Tilburg, The Netherlands<sup>2</sup>Statistical Innovation Group, Advanced Analytics Centre, AstraZeneca, Cambridge, UK**Correspondence**Robbie C. M. van Aert, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands.  
Email: R.C.M.vanAert@tilburguniversity.edu**Funding information**

Netherlands Organization for Scientific Research (NWO), Grant/Award Number: 406-13-050

A wide variety of estimators of the between-study variance are available in random-effects meta-analysis. Many, but not all, of these estimators are based on the method of moments. The DerSimonian-Laird estimator is widely used in applications, but the Paule-Mandel estimator is an alternative that is now recommended. Recently, DerSimonian and Kacker have developed two-step moment-based estimators of the between-study variance. We extend these two-step estimators so that multiple (more than two) steps are used. We establish the surprising result that the multistep estimator tends towards the Paule-Mandel estimator as the number of steps becomes large. Hence, the iterative scheme underlying our new multistep estimator provides a hitherto unknown relationship between two-step estimators and Paule-Mandel estimator. Our analysis suggests that two-step estimators are not necessarily distinct estimators in their own right; instead, they are quantities that are closely related to the usual iterative scheme that is used to calculate the Paule-Mandel estimate. The relationship that we establish between the multistep and Paule-Mandel estimator is another justification for the use of the latter estimator. Two-step and multistep estimators are perhaps best conceptualized as approximate Paule-Mandel estimators.

**KEYWORDS**

estimation, iterative scheme, method of moments, random-effects meta-analysis

## 1 | INTRODUCTION

Meta-analysis statistically combines effect size estimates from different studies in order to calculate a quantitative summary of the evidence base. Two important outcomes from a meta-analysis are the estimates of the overall effect size and the between-study variance (the variance of the studies' true effect sizes). Between-study heterogeneity refers to the possibility that there is more variation in the studies' observed effect sizes than what would be expected by sampling variability alone<sup>1,2</sup> and is often present in meta-analyses.<sup>3-5</sup> Characteristics of the included studies (eg, differences between populations from which participants were sampled or treatments across studies) can be incorporated as moderators in meta-regressions to explore and explain the between-study heterogeneity.<sup>6-8</sup> However, random-effects meta-analyses are often used to account for, but not explain, between-study heterogeneity.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2018 The Authors. *Statistics in Medicine* published by John Wiley & Sons Ltd.

A wide variety of estimators are available for the between-study variance. Two recent papers<sup>9,10</sup> review existing research on these estimators and recommended either the Paule-Mandel (PM) estimator<sup>11</sup> or the restricted maximum likelihood (REML) estimator.<sup>12</sup> However, the DerSimonian-Laird (DL) estimator is most often used in practice.<sup>5,13,14</sup> The popularity of the DL estimator is due to its simplicity, because it is calculated from an easily computed noniterative method and also because it is already familiar to applied meta-analysts. In this paper, we focus on estimators that are motivated by the method of moments, which includes the DL and PM estimators, but not REML.

In particular, we use the general method of moments estimator (ie, with an arbitrary set of weights for the effect sizes) proposed by DerSimonian and Kacker<sup>15</sup> to develop a new multistep DL estimator. This idea extends the two-step DL (DL<sub>2</sub>) estimator, which was also proposed by DerSimonian and Kacker.<sup>15</sup> The usual (one-step) DL estimator uses the inverse of the studies' within-study sampling variances as weights to estimate the between-study variance. In the two-step-estimation procedure, the estimate of the usual DL estimator is calculated in the first step and this estimate is then included in the weights of the second step. Full details of the DL<sub>2</sub> estimator are provided in Section 3. The statistical properties of the DL<sub>2</sub> estimator are largely unknown, because the method has rarely been topic of further study. Bhaumik et al<sup>16</sup> studied the statistical properties of the DL<sub>2</sub> estimator and concluded that for rare events, both the DL<sub>2</sub> and PM estimators are negatively biased. It was our initial intuition that allowing the number of steps to tend to infinity in our new multistep estimator would define a new type of estimator. However, working empirically to begin with and then mathematically, we will demonstrate that the PM estimator is obtained if the number of steps tends towards infinity. Hence, we will instead establish the relationship between the two-step estimators and PM estimator, which is another justification for the use of the PM estimator.

The rest of the paper is set out as follows. We continue with describing the random-effects model for meta-analysis in Section 2. In Section 3, we describe three existing moments-based estimators, DL, DL<sub>2</sub>, and PM. Our new multistep estimator is introduced in Section 4. Subsequently, we apply these estimators to three contrasting examples in Section 5 where we empirically show that the multistep estimator tends towards the PM estimator as the number of steps becomes large, where this convergence occurs quickly in practice. Section 6 contains mathematics that formally establishes the relationship between the multistep estimators and PM estimator. We explore the use of meta-regression models in Section 7, and we conclude with a short discussion in Section 8.

## 2 | THE RANDOM-EFFECTS MODEL

The random-effects model assumes that the effect size estimates  $y_i$ ,  $i=1, \dots, n$ , are extracted from separate studies. This model can be written as

$$y_i = \mu + \mu_i + \epsilon_i, \quad (1)$$

where  $\mu$  is the average true effect size,  $\mu_i$  is a random effect indicating the difference between the  $i$ th study's true effect size and  $\mu$ , and  $\epsilon_i$  is the  $i$ th study's sampling error. It is commonly assumed that  $\mu_i \sim N(0, \tau^2)$  where  $\tau^2$  is the between-study variance and  $\epsilon_i \sim N(0, \sigma_i^2)$ , where  $\sigma_i^2$  is the within-study sampling variance of the  $i$ th study. Furthermore, all  $\mu_i$  and  $\epsilon_i$  are assumed to be mutually independent. The within-study sampling variances  $\sigma_i^2$  are usually estimated in practice and then assumed to be known in the analysis. We will emphasize that the  $\sigma_i^2$  are estimated by writing  $\hat{\sigma}_i^2$  as their estimates.

The parameter  $\mu$  is usually of primary interest. The usual method for making inferences about  $\mu$  initially estimates  $\tau^2$  and then treats the resulting estimate as fixed and known.<sup>9,17</sup> Hence, the conventional weights in the random-effects model,  $1/(\hat{\sigma}_i^2 + \hat{\tau}^2)$ , are treated as fixed and known and the usual inferential procedure for  $\mu$  is straightforward.<sup>8</sup> However, the estimate of the between-study variance,  $\hat{\tau}^2$ , is our primary interest here with moment-based estimators as our focus.

## 3 | MOMENT-BASED METHODS FOR ESTIMATING THE BETWEEN-STUDY VARIANCE

Most of the moment-based estimators for  $\tau^2$  are a special case of a general method of moments estimator.<sup>15</sup> To derive this general estimation method, DerSimonian and Kacker<sup>15</sup> propose methodology for estimating  $\tau^2$  using an arbitrary set of weights  $a_i$ ,  $i=1, \dots, n$ , where all  $a_i$  are fixed positive constants. To estimate  $\tau^2$ , DerSimonian and Kacker<sup>15</sup> propose equating  $\sum_{i=1}^n a_i (y_i - \hat{\mu})^2$ , where  $\hat{\mu} = \sum_{i=1}^n a_i y_i / \sum_{i=1}^n a_i$ , to its expected value. As explained by DerSimonian and Kacker,<sup>15</sup> this results in the estimating equation

$$\hat{\tau}_{MM}^2 = \frac{[\sum_{i=1}^n a_i (y_i - \hat{\mu})^2] - [\sum_{i=1}^n a_i \hat{\sigma}_i^2 - \sum_{i=1}^n a_i^2 \hat{\sigma}_i^2 / \sum_{i=1}^n a_i]}{\sum_{i=1}^n a_i - \sum_{i=1}^n a_i^2 / \sum_{i=1}^n a_i}, \quad (2)$$

where negative estimates  $\hat{\tau}_{MM}^2$  from (2) are truncated to zero (because  $\tau^2 \geq 0$ ). An often overlooked point is that the calculation of the expectation of  $\sum_{i=1}^n a_i (y_i - \hat{\mu})^2$ , which gives rise to the estimating Equation 2, ignores the uncertainty in the  $\hat{\sigma}_i^2$  and has taken  $\sigma_i^2 = \hat{\sigma}_i^2$ . Although when presenting Equation 2, we have emphasized that the estimates  $\hat{\sigma}_i^2$  are used in the calculation; this does not clearly convey the fact that the estimation does not take their uncertainty into account. Kulinskaya and Dollinger<sup>18</sup> and Hoaglin<sup>19</sup> criticize moment-based methods for this type of reason, because ignoring uncertainty in  $\hat{\sigma}_i^2$  may cause bias in the estimate of  $\tau^2$  especially if the sample size of the studies is small. By ignoring the uncertainty in the within-study variances, we have that  $E[\tau_{MM}^2] = \tau^2$  before truncation to zero, but the truncated estimator is positively biased.<sup>20,21</sup>

### 3.1 | The DerSimonian-Laird (DL) estimator

The DL estimator,  $\hat{\tau}_{DL}^2$ , is obtained by taking  $a_i = 1/\hat{\sigma}_i^2$  in Equation 2. We then have  $\sum_{i=1}^n a_i \hat{\sigma}_i^2 - \sum_{i=1}^n a_i^2 \hat{\sigma}_i^2 / \sum_{i=1}^n a_i = n-1$ , so that Equation 2 simplifies when using this standard set of weights. Negative estimates are again truncated to zero. Uncertainty in  $\hat{\sigma}_i^2$  is, as in Equation 2, neglected by treating the weights  $a_i = 1/\hat{\sigma}_i^2$  as fixed constants. This may result in bias when estimating  $\tau^2$  using the DL estimator especially if sample sizes of the studies is small.<sup>18,19</sup>

### 3.2 | The two-step DerSimonian-Laird estimator

DerSimonian and Kacker<sup>15</sup> propose an alternative estimator that is an extension of the DL estimator. The usual DL estimate  $\hat{\tau}_{DL}^2$ , described in the previous section, is calculated in the first step. The two-step DL (DL<sub>2</sub>) estimator adds a second step by incorporating  $\hat{\tau}_{DL}^2$  into the weights and computes  $\hat{\tau}_{DL_2}^2$  using estimating Equation 2 with  $a_i = 1/(\hat{\sigma}_i^2 + \hat{\tau}_{DL}^2)$ .

To describe the two-step DL estimator more explicitly, and also to define the PM and multistep DL estimators below, it is convenient to define the quantity

$$Q_{GEN}(\tau^2) = \sum_{i=1}^n \frac{(y_i - \hat{\mu}(\tau^2))^2}{\hat{\sigma}_i^2 + \tau^2}, \quad (3)$$

where  $\hat{\mu}(\tau^2) = \sum_{i=1}^n y_i / (\hat{\sigma}_i^2 + \tau^2) / \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \tau^2)$ . Then  $Q_{GEN}(0)$  is the usual  $Q$  statistic used in meta-analysis.<sup>22,23</sup> From Equation 2 with  $a_i = 1/(\hat{\sigma}_i^2 + \hat{\tau}_{DL}^2)$ , we have

$$\hat{\tau}_{DL_2}^2 = \frac{Q_{GEN}(\hat{\tau}_{DL}^2) - \left[ \sum_{i=1}^n \hat{\sigma}_i^2 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL}^2) - \sum_{i=1}^n \hat{\sigma}_i^4 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL}^2)^2 / \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL}^2) \right]}{\sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL}^2) - \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL}^2)^2 / \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL}^2)}, \quad (4)$$

where we again truncate negative estimates to zero. The weights  $a_i = 1/(\hat{\sigma}_i^2 + \hat{\tau}_{DL}^2)$  are intuitively appealing, because we then weight by estimates of the studies' total precisions which are also the standard weights when making inferences about  $\mu$  in the random-effects model.<sup>8,24</sup> Using these weights raises further statistical issues, because they are now functions of both the  $\hat{\sigma}_i^2$  and the estimated between-study variance  $\hat{\tau}_{DL}^2$ . There is statistical error in both of these estimated variance components, and so treating the weights  $a_i = 1/(\hat{\sigma}_i^2 + \hat{\tau}_{DL}^2)$  as fixed constants continues to have the potential to have unfortunate implications for the estimation.

It is possible to use other estimators in the first step, and DerSimonian and Kacker<sup>15</sup> also propose using the Cochran analysis of variance (ANOVA) estimator<sup>22,25</sup> that is based on an unweighted sum of squares for this purpose. However, the DL estimator is so common in application that we only explore the use of two-step and multistep estimators that use this particular estimator. Nonetheless, our main results will apply regardless of the type of estimator used in the first step as we will explain below. Hence, generalizability of our results is not restricted by using the DL estimator in the first step, but the results also apply if, for instance, the Cochran ANOVA estimator is used in the first step.

### 3.3 | The Paule-Mandel (PM) estimator

Another moment-based estimator for  $\tau^2$  is the PM estimator.<sup>11</sup> This estimation method exploits the fact that  $Q_{GEN}(\tau^2) \sim \chi_{n-1}^2$  under the assumptions made in the random-effects model (normal sampling distribution for all  $y_i$  and known within-study variances  $\sigma_i^2$ ). Hence,  $\hat{\tau}_{PM}^2$  is obtained by matching  $Q_{GEN}(\tau^2)$  to its expected value and  $\hat{\tau}_{PM}^2$  is the solution to

$$Q_{GEN}(\hat{\tau}_{PM}^2) = n-1. \tag{5}$$

For any given dataset,  $Q_{GEN}(\tau^2)$  is a monotonically decreasing continuous function of  $\tau^2$ . As a consequence, Equation 5 always provides a unique estimate<sup>15,26-28</sup> if  $Q_{GEN}(0) \geq (n-1)$ . If  $Q_{GEN}(0) < (n-1)$ , then no positive solution to the estimating Equation 5 exists, and we take  $\hat{\tau}_{PM}^2 = 0$ . The estimating Equation 5 is nonlinear and so must be solved numerically, but this is straightforward in practice. An empirical Bayes estimator for estimating  $\tau^{229,30}$  was developed independently, but this has subsequently been shown to be equivalent to the PM estimator.<sup>9,28</sup>

Unlike the DL and DL<sub>2</sub> estimator and other moment-based estimators, the PM estimator does not directly use estimating Equation 2. This is because the general method of moments treats the weights  $a_i$  as fixed (and therefore known) constants, but the PM estimator uses weights  $1/(\hat{\sigma}_i^2 + \tau^2)$  that are explicitly unknown (because  $\tau^2$  is unknown). The PM estimator is motivated using the method of moments, but otherwise, there is no direct connection between the PM estimator and other moment-based estimators. We introduce our new multistep estimator in the next section, and we will illustrate the relationship between the PM and the two-step estimator.

#### 4 | THE MULTISTEP DERSIMONIAN AND LAIRD ESTIMATOR

In this section, we develop the multistep DL estimator as a natural extension of the DL<sub>2</sub> estimator. From Equation 4, we have that the DL<sub>2</sub> estimator is simply the estimate from the more general estimating Equation 2 where the weights are  $a_i = 1/(\hat{\sigma}_i^2 + \hat{\tau}_{DL}^2)$ . The key observation is that the two-step estimator uses weights that are the reciprocal of the estimated total study variances, where the between-study variance is estimated using the usual DL estimator. A natural way to extend this estimator to define a three-step estimator is to use weights that are reciprocal of the estimated total study variances, where the between-study variance is estimated using the DL<sub>2</sub> estimator. Hence, we define  $\hat{\tau}_{DL_3}^2$  to be

$$\hat{\tau}_{DL_3}^2 = \frac{Q_{GEN}(\hat{\tau}_{DL_2}^2) - \left[ \sum_{i=1}^n \hat{\sigma}_i^2 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_2}^2) - \sum_{i=1}^n \hat{\sigma}_i^2 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_2}^2)^2 / \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_2}^2) \right]}{\sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_2}^2) - \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_2}^2)^2 / \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_2}^2)},$$

where as before, we truncate negative estimates to zero. We can then define a four-step estimator in a similar way, using Equation 2 with weights  $a_i = 1/(\hat{\sigma}_i^2 + \hat{\tau}_{DL_3}^2)$ , and then a five-step estimator using weights  $a_i = 1/(\hat{\sigma}_i^2 + \hat{\tau}_{DL_4}^2)$ , and so on. In general, we define the  $(k+1)$ th step DL estimator as

$$\hat{\tau}_{DL_{k+1}}^2 = \frac{Q_{GEN}(\hat{\tau}_{DL_k}^2) - \left[ \sum_{i=1}^n \hat{\sigma}_i^2 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2) - \sum_{i=1}^n \hat{\sigma}_i^2 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2)^2 / \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2) \right]}{\sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2) - \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2)^2 / \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2)}, \tag{6}$$

for  $k \geq 1$ , where  $\hat{\tau}_{DL_1}^2$  is defined to be the usual DL estimator  $\hat{\tau}_{DL}^2$ . As usual, we truncate the resulting estimate from Equation 6 to zero if the solution is negative. Written explicitly in terms of this truncation, the  $(k+1)$ th step DL estimator is

$$\hat{\tau}_{DL_{k+1}}^2 = \max \left( 0, \frac{Q_{GEN}(\hat{\tau}_{DL_k}^2) - \left[ \sum_{i=1}^n \hat{\sigma}_i^2 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2) - \sum_{i=1}^n \hat{\sigma}_i^2 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2)^2 / \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2) \right]}{\sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2) - \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2)^2 / \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2)} \right). \tag{7}$$

In practice, we compute  $\hat{\tau}_{DL_k}^2$  recursively by first computing  $\hat{\tau}_{DL}^2$ , then  $\hat{\tau}_{DL_2}^2$ , then  $\hat{\tau}_{DL_3}^2$ , and so on until we reach the required value of  $k$ . However, all of these estimators are available in closed form and so it is in principle also possible to write  $\hat{\tau}_{DL_k}^2$  in this way. Assuming that the limit exists, we define  $\lim_{k \rightarrow \infty} \hat{\tau}_{DL_k}^2 = \hat{\tau}_{DL_\infty}^2$ . We will see below that, whenever convergence occurs,  $\hat{\tau}_{DL_\infty}^2 = \hat{\tau}_{PM}^2$ , so that instead of defining a new estimator, we establish the relationship between existing estimates by taking this limit.

## 5 | EXAMPLES

In this section, we apply the DL, PM,  $DL_2$ , and multistep DL estimators to three contrasting examples. Having illustrated our main findings empirically using these examples, we will demonstrate them mathematically in Section 6.

### 5.1 | Characteristics of the 3 examples

Our first example is a meta-analysis by Bangert-Drowns et al<sup>31</sup> studying the effect of school-based writing-to-learn interventions on academic achievement. This meta-analysis consists of 48 estimated standardized mean differences (ie, Hedges'  $g$ ). The second example is obtained from Sterne et al<sup>32</sup> and is a meta-analysis on the effectiveness of intravenous magnesium in acute myocardial infarction. This meta-analysis consists of 16 estimated log odds ratios. The third example is a meta-analysis on the efficacy of two treatments for post-traumatic stress disorder.<sup>33</sup> This meta-analysis consists of 10 standardized mean differences. The metafor package<sup>34</sup> was used to calculate the DL and PM estimators, and we used our own bespoke code to recursively calculate the multistep  $DL_k$  estimator. R code for applying these estimators to the examples is available via <https://osf.io/paqzm/>.

### 5.2 | Results

Table 1 shows the DL,  $DL_2$ ,  $DL_k$ , and PM estimates of  $\tau^2$  for all three examples. For each example, we calculated the multistep DL estimator until the  $(k+1)$ th step DL estimator was the same as the  $k$ th step estimator up to 4 decimal places. Convergence was taken to have been reached at this point, so that any further steps would result in the same estimate to this level of numerical accuracy. From Table 1, we can see that this convergence was reached in 6, 10, and 4 steps, for example, one, two, and three, respectively. Furthermore, we can see that in each case, the  $DL_2$  estimate is closer to the PM estimate than the DL estimate and that the  $DL_k$  estimate converges to the PM estimate. The way in which this convergence occurred was different for each example. For the first example obtained from Bangert-Drowns et al,<sup>31</sup> the DL estimate was notably less than the PM estimate. Then the  $DL_2$  estimate took a large step towards the PM estimator and after this convergence was quickly reached. For the second example obtained from Sterne et al,<sup>32</sup> the DL estimate was notably greater than the PM estimate and once again, the  $DL_2$  estimate took a large step towards the PM estimator (and in fact “overshot” this). Convergence of the multistep DL estimator was reasonably fast although the sequence produced by the  $DL_k$  estimates was not monotone until  $k \geq 7$ . For the third example obtained from Ho et al,<sup>33</sup> the DL and PM estimators are similar and convergence was very quickly reached.

**TABLE 1** The DerSimonian-Laird (DL), two-step DerSimonian-Laird ( $DL_2$ ), multistep DerSimonian-Laird ( $DL_k$ ) (where  $k$  refers to the  $k$ th step), and Paule-Mandel (PM) estimates for the three example datasets

| Estimate  | Bangert-Drowns et al <sup>31</sup> | Sterne et al <sup>32</sup> | Ho et al <sup>33</sup> |
|-----------|------------------------------------|----------------------------|------------------------|
| DL        | 0.0455                             | 0.2239                     | 0.0076                 |
| $DL_2$    | 0.0652                             | 0.1587                     | 0.0078                 |
| $DL_3$    | 0.0684                             | 0.1841                     | 0.0079                 |
| $DL_4$    | 0.0688                             | 0.1736                     | 0.0079                 |
| $DL_5$    | 0.0689                             | 0.1778                     |                        |
| $DL_6$    | 0.0689                             | 0.1761                     |                        |
| $DL_7$    |                                    | 0.1768                     |                        |
| $DL_8$    |                                    | 0.1765                     |                        |
| $DL_9$    |                                    | 0.1766                     |                        |
| $DL_{10}$ |                                    | 0.1766                     |                        |
| PM        | 0.0689                             | 0.1766                     | 0.0079                 |

### 5.3 | Conclusions

Although the way in which the multistep DL estimator converged to the PM estimator was different in each case, all three examples illustrated the surprising finding that  $\lim_{k \rightarrow \infty} \hat{\tau}_{DLk}^2 = \hat{\tau}_{DL\infty}^2 = \hat{\tau}_{PM}^2$ . A large number of simulations (see <https://osf.io/dpuzs/> for R code) using  $a_i = 1/\hat{\sigma}_i^2$  and  $a_i = 1/(\hat{\sigma}_i^2 + \hat{\tau}^2)$ , where  $\hat{\tau}^2$  is either the DL estimate or the Cochran ANOVA estimate, as study weights in the first step confirmed that multistep estimators converge to the PM estimator. Hence, this indicates that convergence was not only a property of the selected data sets and that convergence also occurred if the DL estimator was not used in the first step. Our findings are in agreement with the observation by DerSimonian and Kacker<sup>15</sup> that two-step estimators better approximate the method of Paule and Mandel, and the conclusion by Bhaumik et al<sup>16</sup> that performance of the DL<sub>2</sub> and PM estimator is similar. This is because we have observed that DL<sub>2</sub> is the second step in an iterative scheme that takes us from  $\hat{\tau}_{DL}^2$  to  $\hat{\tau}_{PM}^2$ .

## 6 | PROVING (WHEN CONVERGENCE OCCURS) THAT THE MULTISTEP ESTIMATOR CONVERGES TO THE PAULE-MANDEL ESTIMATOR

As explained above, in addition to our three examples, many simulated datasets have shown that multistep estimators converge to the PM estimator. In this section, we provide mathematical proofs to formally establish this limit. We will explain why it is not necessary that the DL estimator is used in the first step, so that our findings apply to multistep estimators regardless of the nature of the estimation used in the first step.

### 6.1 | Lemma: agreement with respect to truncation to zero of the DerSimonian and Laird and Paule-Mandel estimators

We start by proving the lemma that the DL and the PM estimators always agree in the sense that, for any given dataset, they are either both zero (if  $Q_{GEN}(0) \leq (n-1)$ ) or both positive (if  $Q_{GEN}(0) > (n-1)$ ). It is conceptually appealing that these two estimators agree in this way, and this is easily proved, but we do not think that this result has been stated previously.

Proof: If  $Q_{GEN}(0) < (n-1)$ , where  $Q_{GEN}(\tau^2)$  is defined in Equation 3, then the PM estimator is truncated to zero as explained in Section 3.3. Furthermore, the first term in the numerator of Equation 2 is also  $Q_{GEN}(0)$  when the DL weights of  $a_i = 1/\hat{\sigma}_i^2$  are used. As noted in Section 3.1, we then also have  $\sum_{i=1}^n a_i \hat{\sigma}_i^2 - \sum_{i=1}^n a_i^2 \hat{\sigma}_i^2 / \sum_{i=1}^n a_i = (n-1)$  in the numerator of Equation 2. Hence, the DL estimator is also truncated to zero if  $Q_{GEN}(0) < (n-1)$ . If  $Q_{GEN}(0) = (n-1)$ , then immediately from their estimating equations, both the DL and PM estimators are zero. Finally, if  $Q_{GEN}(0) > (n-1)$ , then no truncation for either estimator is required, so that the DL and PM estimators are both positive.

### 6.2 | Proving that if convergence of the multistep estimator occurs, then it is to the Paule-Mandel estimate

Having established our lemma, we will prove that the estimate of the multistep estimator equals the PM estimate if convergence occurs. We will prove this first for cases where the convergence is to a positive estimate and then to an estimate of zero.

#### 6.2.1 | The case where the estimate converged to is positive

Assume that convergence occurs and the resulting estimate is positive, so that  $\hat{\tau}_{DLk+1}^2 = \hat{\tau}_{DLk}^2 = \hat{\tau}^2 > 0$ . We substitute  $\hat{\tau}_{DLk+1}^2 = \hat{\tau}_{DLk}^2 = \hat{\tau}^2$  into Equation 6, where this equation correctly describes the iteration from DL<sub>k</sub> to DL<sub>k+1</sub> (because the estimate is positive and no truncation is necessary). Then solving the resulting equation for  $Q_{GEN}(\hat{\tau}^2)$  results in

$$Q_{GEN}(\hat{\tau}^2) = \sum_{i=1}^n \frac{\hat{\sigma}_i^2 + \hat{\tau}^2}{\hat{\sigma}_i^2 + \hat{\tau}^2} - \frac{\sum_{i=1}^n (\hat{\sigma}_i^2 + \hat{\tau}^2) / (\hat{\sigma}_i^2 + \hat{\tau}^2)^2}{\sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}^2)} = (n-1),$$

which from Equation 5 means that  $\hat{\tau}^2 = \hat{\tau}_{PM}^2$ .

## 6.2.2 | The case where the estimate converged to is zero

Assume that convergence occurs and the resulting estimate is either zero or truncated to zero, so that  $\hat{\tau}_{DL_{k+1}}^2 = \hat{\tau}_{DL_k}^2 = \hat{\tau}^2 = 0$ . If we substitute  $\hat{\tau}_{DL_{k+1}}^2 = \hat{\tau}_{DL_k}^2 = \hat{\tau}^2 = 0$  into Equation 7, the term in square brackets of 7 simplifies to  $(n-1)$  and this equation becomes

$$0 = \max\left(0, \frac{Q_{GEN}(0) - (n-1)}{c}\right), \quad (8)$$

where  $c = \sum_{i=1}^n 1/\hat{\sigma}_i^2 - \sum_{i=1}^n 1/\hat{\sigma}_i^4 / \sum_{i=1}^n 1/\hat{\sigma}_i^2 > 0$ . Equation 8 is satisfied only if  $Q_{GEN}(0) - (n-1) \leq 0$ , from which the lemma in Section 6.1 implies that both the DL and PM estimators are zero (which is also the assumed value of  $\hat{\tau}_{DL_{k+1}}^2 = \hat{\tau}_{DL_k}^2 = \hat{\tau}^2$ ). Hence, if the convergence of the multistep estimator is to zero, then the PM estimate is also zero, so that  $\hat{\tau}^2 = \hat{\tau}_{PM}^2$ .

## 6.2.3 | Failure of convergence of the multistep estimator

Although we have observed convergence of the multistep estimators in thousands of simulated datasets (see <https://osf.io/dpuzs/>), it is possible to create examples where the multistep estimator does not converge. As a concrete example of nonconvergence, imagine a meta-analysis with 4 effect sizes  $y_1 = -0.2$ ,  $y_2 = 0.1$ ,  $y_3 = -0.05$ , and  $y_4 = -0.3$ , with corresponding  $\sigma_1^2 = \sigma_2^2 = 0.01$  and  $\sigma_3^2 = \sigma_4^2 = 0.2$ . The DL estimate is  $\hat{\tau}_{DL}^2 = 0.016$ . Using this  $\hat{\tau}_{DL}^2$  in estimating Equation 4 gives  $\hat{\tau}_{DL_2}^2 = 0$ . Hence,  $\hat{\tau}_{DL_3}^2$  is then the usual DL estimator and, instead of achieving convergence, the multistep estimator oscillates between 0.016 and 0, and does not converge to  $\hat{\tau}_{PM}^2 = 0.0066$ . The difficulties for achieving convergence in this example would appear to be because the DL and PM estimates differ so substantially and also because the within-study variances are of different magnitudes (so that  $Q_{GEN}(\tau^2)$  is sensitive to the value of  $\tau^2$  when this is small). This example is a counterexample to the conjecture that the multistep estimator always converges to the PM estimator.

## 6.2.4 | Conclusions

Regardless of whether or not the convergence of the multistep estimator is to a positive estimate, we have proved that if convergence occurs, then this is to the PM estimate. Simulating thousands of meta-analyses (see <https://osf.io/dpuzs/>) did not reveal the convergence problems suggesting that these problems only occur in rare cases such as the artificial one described above. We conclude that that, in practice, multistep estimators converge to the PM estimate and also that they cannot converge to anything other than the PM estimate.

Although the finding that multistep estimators may not converge reduces the utility of our analysis, our analytical results are more general than might be supposed, because it is not limited to using the DL estimator in the first step. All that is necessary for our results is that subsequent steps weight by the reciprocal of the estimated total study variances where the estimated between-study variance is the estimate at the previous step. Hence, our work establishes a link between multistep estimators per se and the PM estimator rather than between just the  $DL_k$  and PM estimators.

## 6.3 | The relationship with an established Newton-Raphson method for calculating the Paule-Mandel estimate

DerSimonian and Kacker<sup>15</sup> propose a Newton-Raphson algorithm for calculating the PM estimate (see their Appendix A). This algorithm sets  $\hat{\tau}_{PM}^2$  to zero if  $Q_{GEN}(0) \leq (n-1)$ . If  $Q_{GEN}(0) > (n-1)$ , then  $\hat{\tau}_{PM}^2 > 0$  and an initial value for the algorithm must be chosen. Then the Newton-Raphson algorithm takes  $\hat{\tau}_{k+1}^2 = \hat{\tau}_k^2 + \Delta\hat{\tau}_{NR}^2$ , where

$$\Delta\hat{\tau}_{NR}^2 = \frac{Q_{GEN}(\hat{\tau}_k^2) - (n-1)}{\sum_{i=1}^n \frac{1}{(\hat{\sigma}_i^2 + \hat{\tau}_k^2)^2} (y_i - \hat{\mu}(\hat{\tau}_k^2))^2}, \quad (9)$$

where  $\hat{\mu}(\hat{\tau}_k^2) = \sum_{i=1}^n y_i / (\hat{\sigma}_i^2 + \hat{\tau}_k^2) / \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_k^2)$ . Negative estimates are truncated to zero, and the algorithm keeps iterating until convergence is reached. Jackson et al<sup>35</sup> explain how to generalize this Newton-Raphson procedure so that it can be applied to meta-regression models.

We can also calculate the corresponding  $\Delta\hat{\tau}^2$  when using Equation 6 in the iterative scheme that produces our multistep estimators as  $\Delta\hat{\tau}^2 = \hat{\tau}_{DL_{k+1}}^2 - \hat{\tau}_{DL_k}^2$ . From Equation 6, this is

$$\Delta\hat{\tau}^2 = \frac{Q_{GEN}(\hat{\tau}_{DL_k}^2) - \left[ \sum_{i=1}^n \hat{\sigma}_i^2 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2) - \sum_{i=1}^n \hat{\sigma}_i^2 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2)^2 / \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2) \right]}{\sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2) - \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2)^2 / \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2)} - \hat{\tau}_{DL_k}^2.$$

Putting the right-hand side of the numerator over a common denominator results in

$$\Delta\hat{\tau}^2 = \frac{Q_{GEN}(\hat{\tau}_{DL_k}^2) - (n-1)}{\sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2) - \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2)^2 / \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2)}. \tag{10}$$

Equation 10 also illustrates why the multistep estimator converges to the PM estimator in practice. This is because the multistep estimator converges if and only if  $\Delta\hat{\tau}^2 = 0$ , so that  $Q_{GEN}(\hat{\tau}_{DL_k}^2) - (n-1) = 0$  and  $\hat{\tau}_{DL_k}^2 = \hat{\tau}_{PM}^2$ . If instead the PM estimate has not been converged to, Equation 10 shows that the estimator takes a step in the direction of the PM estimate in the  $k$ th step, because if  $Q_{GEN}(\hat{\tau}_{DL_k}^2) < (n-1)$ , then  $\Delta\hat{\tau}^2 < 0$  and if  $Q_{GEN}(\hat{\tau}_{DL_k}^2) > (n-1)$ , then  $\Delta\hat{\tau}^2 > 0$ .

Comparing Equations 9 and 10, we can also see that the iterative scheme for the multistep estimator is closely related to the established Newton-Raphson method for calculating  $\hat{\tau}_{PM}^2$ . In Appendix A, we show that the expectation of the denominator of Equation 9, under the model  $y_i \sim N(\mu, \hat{\sigma}_i^2 + \hat{\tau}_k^2)$  and where the  $y_i$  are independent (where we suppress the distinction between  $\hat{\tau}_k^2$  and  $\hat{\tau}_{DL_k}^2$ ), is equal to the denominator of Equation 10. This is reminiscent of the relationship between Fisher's scoring and Newton-Raphson methods in maximum likelihood estimation. This is because Fisher's scoring algorithm solves the likelihood-based estimating equation by replacing the observed information in the denominator in a Newton-Raphson procedure by its expectation (the expected information). This observation provides us with intuition into why multistep estimators tend towards the PM estimator as the number of steps becomes large.

## 7 | THE RANDOM-EFFECTS META-REGRESSION MODEL

For ease of exposition, we have presented our main results for random-effects meta-analyses, but these are readily extended to meta-regression models where study level covariate effects are included in the model. To establish that our results generalize in this way, we consider meta-regression models with an arbitrary number of covariates in this section. All of the results in this section simplify to those shown previously.

The random-effects meta-regression model is an extension of model 1, where we assume that

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + \mu_i + \epsilon_i,$$

where  $\mathbf{x}_i$  is the  $1 \times p$  row vector of covariates associated with this study and  $\boldsymbol{\beta}$  is the  $p \times 1$  column vector of regression parameters of interest. Unless an intercept-free regression is required, the first "covariate" in each study is taken to be one to include the intercept. A matrix formulation of this standard model is

$$\mathbf{Y} | \mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \Delta + \tau^2 \mathbf{I}), \tag{11}$$

where  $\mathbf{Y}$  is a column vector containing the  $y_i$ ,  $\mathbf{X}$  is the  $n \times p$  design matrix (sometimes referred to as the model matrix) whose  $i$ th row is  $\mathbf{x}_i$ ,  $\Delta = \text{diag}(\hat{\sigma}_i^2)$  and  $\mathbf{I}$  is the  $n \times n$  identity matrix. The parameter  $\tau^2$  in model 11 is called the residual between-study variance and describes the heterogeneity in the effect size estimates that is not explained by the covariates.

### 7.1 | The general method of moments for meta-regression

Jackson et al<sup>35</sup> generalize the general method of moments (Equation 2) to the meta-regression setting. They define  $\mathbf{A} = \text{diag}(a_i)$ , a diagonal matrix containing the weights, and  $\mathbf{B} = \mathbf{A} - \mathbf{A}\mathbf{X}(\mathbf{X}^t\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{A}$ . They also define the  $Q_a$  statistic



$$Q_a = \mathbf{Y}^t \mathbf{B} \mathbf{Y}.$$

Jackson et al<sup>35</sup> use the subscript  $a$  to emphasize that the weights  $a_i$  are used, and so use the notation  $Q_a$  for this quadratic form. This  $Q_a$  statistic reduces to the quadratic form in the numerator of Equation 2 in the meta-analysis setting. Jackson et al<sup>35</sup> show that the meta-regression version of the generalized method of moments in Equation 2 is

$$\hat{\tau}_{MM}^2 = \frac{Q_a - \text{tr}(\mathbf{B}\Delta)}{\text{tr}(\mathbf{B})}, \quad (12)$$

where  $\text{tr}(\cdot)$  denotes the trace of a matrix and  $\text{tr}(\mathbf{B}) > 0$ . As in the meta-analysis setting, we truncate  $\hat{\tau}_{MM}^2$  when the solution to Equation 12 is negative.

## 7.2 | Paule-Mandel and DerSimonian and Laird estimators for meta-regression

### 7.2.1 | The Paule-Mandel estimator

The PM-type estimator in the meta-regression setting proposed by Jackson et al<sup>35</sup> uses weights  $a_i = 1/(\hat{\sigma}_i^2 + \tau^2)$  when computing the  $Q_a$  statistic. We denote the resulting  $Q_a$  statistic using the notation  $Q_{GEN}(\tau^2)$  in order to emphasize the dependence of the weights on the unknown parameter  $\tau^2$ . This is a direct generalization of  $Q_{GEN}(\tau^2)$  in Equation 3. Since  $Q_{GEN}(\tau^2)$  follows a  $\chi^2$ -distribution with  $n-p$  degrees of freedom, the PM estimator is obtained by solving

$$Q_{GEN}(\hat{\tau}_{PM}^2) = n-p. \quad (13)$$

If  $Q_{GEN}(0) < n-p$ , then because for any given dataset  $Q_{GEN}(\tau^2)$  is a monotonically decreasing continuous function in  $\tau^2$ , there is no solution to this equation and we take  $\hat{\tau}_{PM}^2 = 0$ .<sup>35</sup> Following similar arguments as in the meta-analysis case, if  $Q_{GEN}(0) \leq n-p$ , then  $\hat{\tau}_{PM}^2 = 0$  and if  $Q_{GEN}(0) > n-p$ , then  $\hat{\tau}_{PM}^2 > 0$ .

### 7.2.2 | The DerSimonian and Laird estimator

The standard weights of  $a_i = 1/\hat{\sigma}_i^2$  produce a DL-type estimator of  $\tau^2$  when using Equation 12, so that this estimator is just a special case of the general method of moments. We then have  $\mathbf{A} = \Delta^{-1}$  so that  $\mathbf{B} = \Delta^{-1} - \Delta^{-1} \mathbf{X}(\mathbf{X}^t \Delta^{-1} \mathbf{X})^{-1} \mathbf{X}^t \Delta^{-1}$ . Hence, with these weights, the numerator of Equation 12 becomes

$$Q_{GEN}(0) - \text{tr}(\mathbf{B}\Delta) = Q_{GEN}(0) - \text{tr}(\Delta^{1/2} \mathbf{B} \Delta^{1/2}),$$

where this final equality is because  $\text{tr}(\mathbf{C}\mathbf{D}) = \text{tr}(\mathbf{D}\mathbf{C})$ , where  $\mathbf{C}$  and  $\mathbf{D}$  are square matrices of the same size, and because  $\Delta^{1/2} \Delta^{1/2} = \Delta$ . We can then further simplify this expression by taking

$$\text{tr}(\Delta^{1/2} \mathbf{B} \Delta^{1/2}) = n-p.$$

This identity is because  $\text{tr}(\Delta^{1/2} \mathbf{B} \Delta^{1/2}) = \text{tr}(\mathbf{I}) - \text{tr}(\Delta^{-1/2} \mathbf{X}(\mathbf{X}^t \Delta^{-1} \mathbf{X})^{-1} \mathbf{X}^t \Delta^{-1/2})$ , where  $\text{tr}(\mathbf{I}) = n$  and  $\text{tr}(\Delta^{-1/2} \mathbf{X}(\mathbf{X}^t \Delta^{-1} \mathbf{X})^{-1} \mathbf{X}^t \Delta^{-1/2}) = p$ . This final equality follows from the observation that the hat matrix corresponding to a design matrix  $\mathbf{X}$  is given by  $\mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t$ , where  $\text{tr}(\mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t) = \text{tr}(\mathbf{X}^t \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1})$ . For an identifiable regression  $\mathbf{X}^t \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1}$  is a  $p \times p$  identity matrix, which results in the well-known result that the trace of the hat matrix is  $p$ . Then we simply observe that  $\Delta^{-1/2} \mathbf{X}(\mathbf{X}^t \Delta^{-1} \mathbf{X})^{-1} \mathbf{X}^t \Delta^{-1/2}$  is the hat matrix corresponding to the design matrix  $\Delta^{-1/2} \mathbf{X}$ , so that its trace is also  $p$ . The numerator of Equation 12 therefore simplifies to  $Q_{GEN}(0) - (n-p)$  for the DL estimator.

## 7.3 | Multistep estimators for meta-regression

We can motivate multistep estimators of  $\tau^2$  for meta-regression in exactly the same way as in meta-analysis. For example, using the DL estimator, we first calculate  $\hat{\tau}_{DL}^2$  using Equation 12 and weights of  $a_i = 1/\hat{\sigma}_i^2$ , truncating the estimate to zero if the solution is negative. We can then calculate  $\hat{\tau}_{DL_2}^2$  using Equation 12 and weights of  $a_i = 1/(\hat{\sigma}_i^2 + \hat{\tau}_{DL}^2)$ , from which we can then calculate  $\hat{\tau}_{DL_3}^2$  and so on. In general, we calculate  $\hat{\tau}_{DL_{k+1}}^2$  using Equation 12 with weights of

$a_i = 1/(\hat{\sigma}_i^2 + \hat{\tau}_{DL_k}^2)$ . Any negative solutions are truncated to zero. This process generalizes the multistep estimators for meta-analysis described in Section 4.

Let  $\mathbf{A}_{\hat{\tau}_{DL_k}^2} = (\Delta + \hat{\tau}_{DL_k}^2 \mathbf{I})^{-1}$  denote the diagonal matrix containing the weights when computing the  $(k+1)$ th step DL estimator, for  $k \geq 1$ . Let  $\mathbf{B}_{\hat{\tau}_{DL_k}^2}$  denote the corresponding matrix  $\mathbf{B}$  computed using  $\mathbf{A}_{\hat{\tau}_{DL_k}^2}$ . From Equation 12, we can then write

$$\hat{\tau}_{DL_{k+1}}^2 = \frac{Q_{GEN}(\hat{\tau}_{DL_k}^2) - \text{tr}(\mathbf{B}_{\hat{\tau}_{DL_k}^2} \Delta)}{\text{tr}(\mathbf{B}_{\hat{\tau}_{DL_k}^2})}, \quad (14)$$

for  $k \geq 1$ , where we truncate the resulting estimate to zero if the solution is negative. Equation 14 is a direct generalization of Equation 6 for meta-regression. Written explicitly in terms of the truncation, the  $(k+1)$ th step estimator is

$$\hat{\tau}_{DL_{k+1}}^2 = \max\left(0, \frac{Q_{GEN}(\hat{\tau}_{DL_k}^2) - \text{tr}(\mathbf{B}_{\hat{\tau}_{DL_k}^2} \Delta)}{\text{tr}(\mathbf{B}_{\hat{\tau}_{DL_k}^2})}\right), \quad (15)$$

and Equation 15 is a direct generalization of Equation 7.

## 7.4 | Lemma: Agreement with respect to truncation to zero of the DL and PM estimators

In this section, we generalize the lemma in Section 6.1 for meta-regression. As explained above, the PM estimator is positive if and only if  $Q_{GEN}(0) > n-p$ . As also explained above, the numerator of Equation 12 simplifies to  $Q_{GEN}(0) - (n-p)$  when using the DL estimator ( $a_i = 1/\hat{\sigma}_i^2$ ). Hence, the DL estimator is also positive if and only if  $Q_{GEN}(0) > n-p$ . If instead  $Q_{GEN}(0) \leq n-p$ , then both the DL and PM estimators are zero. We therefore have established that the type of agreement described in Section 6.1 between the PM and DL estimators also applies in the meta-regression setting.

## 7.5 | Proving that if convergence occurs, then it is to the Paule-Mandel estimate

Although we do not prove that the multistep estimator always converges for the meta-regression model, we have also simulated thousands of datasets and conducted meta-regression analyses with one continuous covariate (see <https://osf.io/5wqvdl/> for R code) and did not observe any convergence problems. We have established that artificial examples can be created where the multistep estimator does not converge but that this nonconvergence is unlikely to occur in practice in both meta-analysis and meta-regression. In the remainder of this section, we generalize the results in Section 6.2 for meta-regression.

### 7.5.1 | The case where the estimate converged to is positive

Assume that convergence occurs and the resulting estimate is positive, so that  $\hat{\tau}_{DL_{k+1}}^2 = \hat{\tau}_{DL_k}^2 = \hat{\tau}^2 > 0$ . We substitute  $\hat{\tau}_{DL_{k+1}}^2 = \hat{\tau}_{DL_k}^2 = \hat{\tau}^2$  into Equation 14, where this equation correctly describes the iteration from  $DL_k$  to  $DL_{k+1}$  (because the estimate is positive and no truncation is necessary). Then solving the resulting equation for  $Q_{GEN}(\hat{\tau}^2)$  results in

$$Q_{GEN}(\hat{\tau}^2) = \text{tr}(\mathbf{B}_{\hat{\tau}^2}(\Delta + \hat{\tau}^2 \mathbf{I})) = (n-p),$$

where the final equality follows from an argument involving a hat matrix that is very similar to the one made in Section 7.2.2. Equation 13 implies that  $\hat{\tau}^2 = \hat{\tau}_{PM}^2$ .

### 7.5.2 | The case where the estimate converged to is zero

Assume that convergence occurs and the resulting estimate is either zero or truncated to zero, so that  $\hat{\tau}_{DL_{k+1}}^2 = \hat{\tau}_{DL_k}^2 = \hat{\tau}^2 = 0$ . If we substitute  $\hat{\tau}_{DL_{k+1}}^2 = \hat{\tau}_{DL_k}^2 = \hat{\tau}^2 = 0$  into Equation 15, then this equation becomes

$$0 = \max\left(0, \frac{Q_{GEN}(0) - (n-p)}{c}\right),$$

where  $c = \text{tr}(\mathbf{B}_0) > 0$ . Equation 8 is satisfied only if  $Q_{GEN}(0) - (n-p) \leq 0$ , from which the lemma in Section 7.4 implies that both the DL and PM estimators are zero (which is also the assumed value of  $\hat{\tau}_{DL_{k+1}}^2 = \hat{\tau}_{DL_k}^2 = \hat{\tau}^2$ ). Hence, if the convergence of the multistep estimator is to zero, then the PM estimate is also zero, so that  $\hat{\tau}^2 = \hat{\tau}_{PM}^2$ . We have therefore established that multistep estimates also converge to the PM estimator in meta-regression models whenever convergence occurs.

## 8 | DISCUSSION

Two-step estimators have recently been presented as estimators of the between-study variance. We have extended these two-step estimators to a multistep estimator and show by means of empirical examples, simulations, and also analytically that the multistep estimator converges to the PM estimator if the number of steps is sufficiently large. This convergence occurs quickly in practice. Although examples can be produced where the multistep estimator does not converge, we have shown that the PM estimator is obtained in the limit when convergence is obtained and that convergence problems seldom occur in practice. Hence, our analysis suggests that the two-step estimators are better conceptualized as part of the usual iterative scheme that is used to calculate estimates using the PM estimator. Our findings also clarify why previous work<sup>15,16</sup> observed that the  $DL_2$  estimator was closer to the PM estimator than the DL estimator. We therefore suggest that the two-step estimators, as well as the proposed multistep estimator, are not seen as truly distinct estimators but as steps in an iterative procedure that results in the PM estimator.

Now that REML and the PM estimator are computationally feasible and established in standard software, we align ourselves with those who argue that these estimators should be preferred over the DL estimator.<sup>9,10</sup> The case for REML becoming the default estimation method is now strong. However, the PM estimator is a viable alternative that is currently the best estimator that uses the method of moments. An advantage of the PM estimator compared to REML is that, in a small proportion of meta-analyses, REML suffers from convergence problems.<sup>5</sup> A byproduct of our work is the development of a new iterative scheme that can be used to calculate the PM estimator.

Our work is a good example of scientists exploring an issue of interest with the expectation of discovering something new and then making new, but unanticipated, discoveries. However, discovering the link between the multistep and PM estimator is in some respects even more satisfying than inventing a new class of estimators of the between-study variance. We have already explained that the PM estimator has been found to be equivalent to the empirical Bayes estimator, and our results provide another justification for the use of the PM estimator. This estimator would therefore seem to have a very wide variety of justifications and connections with other approaches, which suggests that it has a useful role in both methodological and applied work.

The estimation equation for the multistep estimator in Equation 6 closely resembles a fixed-point iteration problem,<sup>36</sup> because the estimate of the between-study variance in the previous step is included in the weights of the estimation equation in the next step. Studying the multistep estimator using methods for fixed-point iteration may yield further insights into the characteristics of meta-analysis datasets where convergence problems occur. We leave this as an opportunity for future research, which would probably be best undertaken by experts in numerical analysis.

We have considered the random-effects models for meta-analysis and meta-regression. Both of these models assume that the outcome data are independent. More sophisticated models that allow for correlated data include multivariate meta-analysis<sup>37</sup> and network meta-analysis.<sup>38</sup> Jackson et al<sup>39</sup> have already developed PM estimators for network meta-analysis, but our connection between multistep and PM estimators provides an alternative possibility for motivating them. There is currently no PM estimator for the between-study covariance matrix in multivariate meta-analysis, but two extensions of the DL estimator have been proposed.<sup>40,41</sup> Generalizing one or both of these estimators to allow an arbitrary set of weights, and so develop a general method of moments estimator, could then motivate the development of multistep estimators in the context of multivariate meta-analysis. When convergence is reached as the number of steps becomes large, PM estimators of the between-study covariance matrix could then be defined in this limit. However, considerable methodological development is needed to extend our work to the network and multivariate meta-analysis settings, because this would first require the development of a generalized method of moments for correlated outcome data. We therefore leave this as a tantalizing possibility for further work. However, enthusiasm for this idea is likely to be mitigated by the finding that the multistep estimator does not always converge. Matters will become more complicated in the multivariate setting and some convention for defining a PM estimator in this way when convergence is not obtained would be needed.

To summarize, we have extended the two-step estimator so that multiple steps can be used and reproduced the PM estimator in the limit when the number of steps are sufficiently large. The PM estimator therefore has another justification as a result of its relationship with the proposed multistep estimator. We suggest that the meta-analysis

community should no longer consider the two-step and multistep estimators to be truly distinct estimators but should instead regard these type of estimators as approximate PM estimators.

## ACKNOWLEDGEMENT

The authors thank Marcel A. L. M. van Assen for his comments on a previous version of this paper. Robbie C. M. van Aert was supported by grant 406-13-050 from the Netherlands Organization for Scientific Research (NWO).

## ORCID

Robbie C. M. van Aert  <http://orcid.org/0000-0001-6187-0665>

Dan Jackson  <http://orcid.org/0000-0002-4963-8123>

## REFERENCES

1. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188.
2. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539-1558.
3. Higgins JPT. Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *Int J Epidemiol*. 2008;37(5):1158-1160.
4. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc*. 2009;172(1):137-159.
5. Kontopantelis E, Springate DA, Reeves D. A re-analysis of the Cochrane Library data: the dangers of unobserved heterogeneity in meta-analyses. *PLoS One*. 2013;8(7):e69930.
6. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med*. 1999;18(20):2693-2708.
7. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med*. 2002;21(4):589-624.
8. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. Chichester, UK: John Wiley & Sons, Ltd.; 2009.
9. Veroniki AA, Jackson D, Viechtbauer W, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res Synth Methods*. 2016;7(1):55-79.
10. Langan D, Higgins JPT, Simmonds M. Comparative performance of heterogeneity variance estimators in meta-analysis: a review of simulation studies. *Res Synth Methods*. 2016;8(2):181-198.
11. Paule RC, Mandel J. Consensus values and weighting factors. *J Res Nat Bur Stand*. 1982;87(5):377-385.
12. Raudenbush SW. Analyzing effect sizes: random-effects models. In: Cooper H, Hedges LV, Valentine JC, eds. *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation; 2009:295-315.
13. Jackson D, Bowden J, Baker R. How does the DerSimonian and Laird procedure for random effects meta-analysis compare with its more efficient but harder to compute counterparts?. *J Stat Plan Inference*. 2010;140(4):961-970.
14. Wiksten A, Rücker G, Schwarzer G. Hartung-Knapp method is not always conservative compared with fixed-effect meta-analysis. *Stat Med*. 2016;35(15):2503-2515.
15. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials*. 2007;28(2):105-114.
16. Bhaumik DK, Amatya A, Normand SLT, et al. Meta-analysis of rare binary adverse event data. *J Am Stat Assoc*. 2012;107(498):555-567.
17. Biggerstaff BJ, Tweedie RL. Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. *Stat Med*. 1997;16(7):753-768.
18. Kulinskaya E, Dollinger MB. An accurate test for homogeneity of odds ratios based on Cochran's Q-statistic. *BMC Med Res Methodol*. 2015;15(49):1-19.
19. Hoaglin DC. Misunderstandings about Q and 'Cochran's Q test' in meta-analysis. *Stat Med*. 2016;35(4):485-495.
20. Rukhin AL. Estimating heterogeneity variance in meta-analysis. *J R Stat Soc Series B Stat Methodol*. 2013;75(3):451-469.
21. Viechtbauer W. Bias and efficiency of meta-analytic variance estimators in the random-effects model. *J Educ Behav Stat*. 2005;30(3):261-293.
22. Cochran WG. The combination of estimates from different experiments. *Biometrics*. 1954;10(1):101-129.
23. Hoaglin DC. Shortcomings of an approximate confidence interval for moment-based estimators of the between-study variance in random-effects meta-analysis. *Res Synth Methods*. 2016;7(4):459-461.

24. Shadish WR, Haddock CK. Combining estimates of effect size. In: Cooper H, Hedges LV, Valentine JC, eds. *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation; 2009:257-277.
25. Hedges LV. A random effects model for effect sizes. *Psychol Bull*. 1983;93(2):388-395.
26. Knapp G, Biggerstaff BJ, Hartung J. Assessing the amount of heterogeneity in random-effects meta-analysis. *Biom J*. 2006;48(2):271-285.
27. Viechtbauer W. Confidence intervals for the amount of heterogeneity in meta-analysis. *Stat Med*. 2007;26(1):37-52.
28. Viechtbauer W, López-López JA, Sánchez-Meca J, Marín-Martínez F. A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychol Methods*. 2015;20(3):360-374.
29. Morris CN. Parametric empirical Bayes inference: theory and applications. *J Am Stat Assoc*. 1983;78(381):47-55.
30. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. *Stat Med*. 1995;14(4):395-411.
31. Bangert-Drowns RL, Hurley MM, Wilkinson B. The effects of school-based writing-to-learn interventions on academic achievement: a meta-analysis. *Rev Educ Res*. 2004;74(1):29-58.
32. Sterne JAC, Bradburn MJ, Egger M. Meta-analysis in Stata. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic Reviews in Health Care: Meta-Analysis in Context*. 2nd ed. London: BMJ Books; 2001:347-369.
33. Ho MSK, Lee CW. Cognitive behaviour therapy versus eye movement desensitization and reprocessing for post-traumatic disorder—is it all in the homework then?. *Eur Rev Appl Psychol*. 2012;62(4):253-260.
34. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36(3):1-48.
35. Jackson D, Turner RM, Rhodes K, Viechtbauer W. Methods for calculating confidence and credible intervals for the residual between-study variance in random effects meta-regression models. *BMC Med Res Methodol*. 2014;14(1):103.
36. Adams RA, Essex C. *Calculus: A Complete Course*. 8th ed. Toronto: Pearson; 2013.
37. Jackson D, Riley R, White IR. Multivariate meta-analysis: potential and promise. *Stat Med*. 2011;30(20):2481-2498.
38. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods*. 2012;3(2):80-97.
39. Jackson D, Veroniki AA, Law M, Tricco AC, Baker R. Paule-Mandel estimators for network meta-analysis with random inconsistency effects. *Res Synth Methods*. 2017;8(4):416-434.
40. Jackson D, White IR, Thompson SG. Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. *Stat Med*. 2010;29(12):1282-1297.
41. Jackson D, White IR, Riley RD. A matrix-based method of moments for fitting the multivariate random effects model for meta-analysis and meta-regression. *Biom J*. 2013;55(2):231-245.

**How to cite this article:** van Aert RCM, Jackson D. Multistep estimators of the between-study variance: The relationship with the Paule-Mandel estimator. *Statistics in Medicine*. 2018;37:2616–2629. <https://doi.org/10.1002/sim.7665>

## APPENDIX A

In this appendix, we prove that

$$\mathbb{E} \left[ \sum_{i=1}^n \frac{1}{(\hat{\sigma}_i^2 + \hat{\tau}_k^2)^2} (y_i - \hat{\mu}(\hat{\tau}_k^2))^2 \right] = \sum_{i=1}^n \frac{1}{\hat{\sigma}_i^2 + \hat{\tau}_k^2} - \frac{\sum_{i=1}^n \frac{1}{(\hat{\sigma}_i^2 + \hat{\tau}_k^2)^2}}{\sum_{i=1}^n \frac{1}{\hat{\sigma}_i^2 + \hat{\tau}_k^2}},$$

under the model  $y_i \sim N(\mu, \hat{\sigma}_i^2 + \hat{\tau}_k^2)$ , where all  $y_i$  are independent and  $\hat{\mu}(\hat{\tau}_k^2) = \sum_{i=1}^n y_i / (\hat{\sigma}_i^2 + \hat{\tau}_k^2) / \sum_{i=1}^n 1 / (\hat{\sigma}_i^2 + \hat{\tau}_k^2)$ .

To simplify the notation, let  $w_i = 1 / (\hat{\sigma}_i^2 + \hat{\tau}_k^2)$ . Then the required expectation is

$$\begin{aligned} \mathbb{E} \left[ \sum_{i=1}^n w_i^2 (y_i - \hat{\mu}(\hat{\tau}_k^2))^2 \right] &= \mathbb{E} \left[ \sum_{i=1}^n w_i^2 (y_i - \mu + \mu - \hat{\mu}(\hat{\tau}_k^2))^2 \right] \\ &= \mathbb{E} \left[ \sum_{i=1}^n w_i^2 (y_i - \mu)^2 \right] + \mathbb{E} \left[ \sum_{i=1}^n w_i^2 (\hat{\mu}(\hat{\tau}_k^2) - \mu)^2 \right] - 2 \mathbb{E} \left[ \sum_{i=1}^n w_i^2 (y_i - \mu) (\hat{\mu}(\hat{\tau}_k^2) - \mu) \right]. \end{aligned} \quad (\text{A1})$$

The first term in Equation A1 is

$$\mathbb{E} \left[ \sum_{i=1}^n w_i^2 (y_i - \mu)^2 \right] = \sum_{i=1}^n w_i^2 \mathbb{E} [(y_i - \mu)^2] = \sum_{i=1}^n w_i^2 \text{Var}[y_i] = \sum_{i=1}^n w_i. \quad (\text{A2})$$

The second term in Equation A1 is

$$\mathbb{E} \left[ \sum_{i=1}^n w_i^2 (\hat{\mu}(\hat{\tau}_k^2) - \mu)^2 \right] = \sum_{i=1}^n w_i^2 \mathbb{E} [(\hat{\mu}(\hat{\tau}_k^2) - \mu)^2] = \text{Var}[\hat{\mu}(\hat{\tau}_k^2)] \sum_{i=1}^n w_i^2,$$

where  $\text{Var}[\hat{\mu}(\hat{\tau}_k^2)] = \text{Var}[\sum_{i=1}^n w_i y_i / \sum_{i=1}^n w_i] = 1 / \sum_{i=1}^n w_i$ . Hence, the second term in A1 is equal to

$$\frac{\sum_{i=1}^n w_i^2}{\sum_{i=1}^n w_i}. \quad (\text{A3})$$

The third term in Equation A1 is

$$-2 \mathbb{E} \left[ \sum_{i=1}^n w_i^2 (y_i - \mu) (\hat{\mu}(\hat{\tau}_k^2) - \mu) \right] = -2 \sum_{i=1}^n w_i^2 \text{Cov} \left[ y_i, \frac{\sum_{j=1}^n w_j y_j}{w_+} \right],$$

where  $w_+ = \sum_{i=1}^n w_i$  and we have used the definition of  $\hat{\mu}(\hat{\tau}_k^2)$  and use the summation over  $j$  to compute it, because  $\hat{\mu}(\hat{\tau}_k^2)$  takes the same value for all  $i$  in the summation. The  $y_i$  are independent so that all covariances in the above summation are zero unless  $i=j$ . Hence, the third term is

$$-2 \sum_{i=1}^n w_i^2 \text{Cov} \left[ y_i, \frac{\sum_{j=1}^n w_j y_j}{w_+} \right] = -2 \sum_{i=1}^n w_i^2 \text{Cov} \left[ y_i, \frac{w_i y_i}{w_+} \right] = -2 \frac{\sum_{i=1}^n w_i^2 \text{Cov}[y_i, w_i y_i]}{w_+},$$

where  $\text{Cov}[y_i, w_i y_i] = w_i \text{Cov}(y_i, y_i) = w_i \text{Var}(y_i) = 1$  so that the third term is

$$-2 \frac{\sum_{i=1}^n w_i^2}{w_+}. \quad (\text{A4})$$

The summation of Equations A2, A3, and A4, recalling that  $w_i = 1 / (\hat{\sigma}_i^2 + \hat{\tau}_k^2)$ , gives the required expectation.