

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



Morris, TP; Fisher, DJ; Kenward, MG; Carpenter, JR (2018) Meta-analysis of Gaussian individual patient data: two-stage or not two-stage? *Statistics in medicine*. ISSN 0277-6715 DOI: <https://doi.org/10.1002/sim.7589>

Downloaded from: <http://researchonline.lshtm.ac.uk/4646152/>

DOI: [10.1002/sim.7589](https://doi.org/10.1002/sim.7589)



Usage Guidelines

Please refer to usage guidelines at <http://researchonline.lshtm.ac.uk/policies.html> or alternatively contact researchonline@lshtm.ac.uk.

Available under license: <http://creativecommons.org/licenses/by/2.5/>

RESEARCH ARTICLE

Meta-analysis of Gaussian individual patient data: two-stage or not two-stage?

Tim P. Morris¹  | David J. Fisher¹ | Michael G. Kenward²  | James R. Carpenter^{1,3}¹London Hub for Trials Methodology Research, MRC Clinical Trials Unit at UCL, London, UK²Ashkirk, UK³Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK**Correspondence**Tim P. Morris, London Hub for Trials Methodology Research, MRC Clinical Trials Unit at UCL, London, UK.
Email: tim.morris@ucl.ac.uk**Present Address**

90 High Holborn, London WC1V 6LJ, UK

Funding information

Medical Research Council, Grant/Award Number: MC_UU_12023/21 and MC_UU_12023/29

Quantitative evidence synthesis through meta-analysis is central to evidence-based medicine. For well-documented reasons, the meta-analysis of individual patient data is held in higher regard than aggregate data. With access to individual patient data, the analysis is not restricted to a “two-stage” approach (combining estimates and standard errors) but can estimate parameters of interest by fitting a single model to all of the data, a so-called “one-stage” analysis. There has been debate about the merits of one- and two-stage analysis. Arguments for one-stage analysis have typically noted that a wider range of models can be fitted and overall estimates may be more precise. The two-stage side has emphasised that the models that can be fitted in two stages are sufficient to answer the relevant questions, with less scope for mistakes because there are fewer modelling choices to be made in the two-stage approach. For Gaussian data, we consider the statistical arguments for flexibility and precision in small-sample settings. Regarding flexibility, several of the models that can be fitted only in one stage may not be of serious interest to most meta-analysis practitioners. Regarding precision, we consider fixed- and random-effects meta-analysis and see that, for a model making certain assumptions, the number of stages used to fit this model is irrelevant; the precision will be approximately equal. Meta-analysts should choose modelling assumptions carefully. Sometimes relevant models can only be fitted in one stage. Otherwise, meta-analysts are free to use whichever procedure is most convenient to fit the identified model.

KEYWORDS

individual-patient data, meta-analysis, one-stage, two-stage

1 | INTRODUCTION

Quantitative synthesis of evidence from multiple studies through meta-analysis is a cornerstone of medical, psychological, and social research. Historically, parameter estimation in meta-analysis has involved two stages, taking (published) estimates of effect and standard errors from each study (first stage), then combining them (second stage), assigning weights to each study, with weights most often based on the inverse variance.

Over the last 2 decades, the bar for high-quality systematic reviews has been raised, particularly by reviewers who have collected participant-level data for included studies and analysed these datasets to obtain their own trial-level estimates

.....
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

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

in place of published estimates. The merits of using individual participant data (IPD) are well documented¹⁻⁴: Data can be checked by the reviewer; consistent inclusion and exclusion criteria can be applied; participants can be included who were not in the original reports; longer follow-up may be included than at the time of the primary study report; analyses can be executed identically for each study, meaning that it is more reasonable to combine estimates; and there is greater scope to investigate interactions. Collecting and checking IPD is a time-consuming, diplomatic, and expensive exercise, so when IPD has been obtained, the advantages should be fully leveraged.

When IPD have been obtained, the analysis is not restricted to combining estimates and standard errors from published data. Instead, the whole analysis can be done in one stage. The merits of “one-stage” compared to “two-stage” approaches have been debated.⁵⁻⁸ Those who argue for one-stage methods tend to be statistically trained, while those who argue for two-stage methods are typically meta-analysts whose primary training is not statistical. (Executing a thorough systematic review and IPD meta-analysis involves many non-statistical skills.) There seems to be support for both perspectives.

The views of a one-stage exponent are exemplified by an anonymous reviewer of one of the authors' (DJF) work on a Stata package to perform two-stage meta-analysis of IPD.³ The reviewer was ambivalent. Despite approving the package, he or she was convinced that two-stage methods have little to offer, writing

The debate is ongoing but the statistical merits of one-stage IPD far outweigh the benefits of a two-stage approach ...

... the merit of the one-stage is doing an analysis that is definitely more accurate ...

... a simulation study is needed to quantify how much better one-stage is in certain scenarios ...

I do not think it [two-stage meta-analysis] should be recommended for use when a one-stage approach is possible

These opinions were apparently informed by Mathew and Nordström's paper,⁶ which shows that a two-stage procedure can be at best as asymptotically efficient as a one-stage procedure. This is in contrast to the view of Burke et al who argue that “differences arise because of different modelling assumptions, rather than the choice of one-stage or two-stage itself.”⁹

The two-stage perspective tends to be that the models and their assumptions are transparent and of scientific value and are sufficient to address relevant substantive questions. There is arguably less transparency in the reporting of models used with one-stage analysis, as evidenced by two recent reviews.^{10,11} Simmonds et al found that “One-stage methods were in general more poorly described, perhaps because of the greater complexity involved in describing properties of regression models.”¹⁰ The main choices for two-stage analysis are around whether to model treatment effects as fixed or random and (a closely related choice) how much weight studies should contribute to the overall estimate. The apparent simplicity of two-stage analysis, combined with the longer history of two-stage estimation in the context of meta-analysis, may be why descriptions tend to be better.

The aim of this paper is to consider the statistical arguments for one- and two-stage analysis in the context of Gaussian outcome data, and particularly to identify differences in precision for models that can be fitted in either way. This is done in the finite-sample setting; the reality we all work in. That is, our results allow explicitly for small numbers of patients-per-study and the implied uncertainty about study-specific estimates.

To aid clarity, we focus here on the simple setting where the IPD include a quantitative outcome, an indicator of treatment assignment, a participant identifier, and a study identifier. It is assumed that the focus of the meta-analysis is on the treatment effect and that no covariates are present. This is an inadequate summary of a meta-analysis,¹² but this is often the primary focus and permits a clear discussion of the arguments for and against one- and two-stage approaches.

As evidenced by the above quotes, two-stage procedures can receive a bad press from statisticians, and we seek to establish the extent to which such comments are justified, not asymptotically, but in practice.

2 | MODELS FOR META-ANALYSIS

Even for the relatively simple structure of datasets considered here, there are several possible models for performing a meta-analysis. Borrowing heavily from Senn,¹³ Table 1 lists seven that might be of interest. We define our notation below before discussing the models in Table 1.

2.1 | Notation

Let y denote outcome and x denote treatment assignment¹⁴ (coded $-0.5/+0.5$). Let $i = 1, \dots, I$ index studies and $j = 1, \dots, n_i$ index patients within a study, so that y_{ij} is the outcome for the j th patient in the i th study. Let α be the intercept term and β be the treatment effect, the parameter of central interest. The interpretation of α and β will depend on which

TABLE 1 Models for individual patient data meta-analysis. Model II is the “fixed-effects meta-analysis” and model V is the “random-effects meta-analysis”

Model	σ_i^2	one-stage? two-stage?	Example one-stage Stata code	Example two-stage Stata code
I	$ry_{ij} = \alpha + \beta x_{ij} + \epsilon_{ij}$	Yes No	Common ...	regress y x ...
II	$ry_{ij} = (\alpha + a_i) + \beta x_{ij} + \epsilon_{ij}$	Yes Yes	Study-specific	mixed y x i.study, residuals(,by(study)) reml ipdmetan, study(study): regress y x
III	$y_{ij} = \alpha_i + \beta x_{ij} + \epsilon_{ij}$ $\alpha_i = \alpha + a_i$ where $a_i \sim N(0, \lambda^2)$	Yes No	Common	mixed y x study: , reml ...
IV	$y_{ij} = (\alpha + a_i) + (\beta + b_i)x_{ij} + \epsilon_{ij}$	Yes No	Common	regress y x##study ...
V	$y_{ij} = (\alpha + a_i) + \beta_i x_{ij} + \epsilon_{ij}$ $\beta_i = \beta + b_i$ where $b_i \sim N(0, \tau^2)$	Yes Yes	Study-specific	mixed y x i.study study: x, nocons residuals(,by(study)) reml df(kr) $b_{ipdmetan}$, study(study) re(kr): regress y x
VI	$y_{ij} = \alpha_i + (\beta + b_i)x_{ij} + \epsilon_{ij}$ $\alpha_i = \alpha + a_i$ where $a_i \sim N(0, \lambda^2)$.	Yes Yes ^a	Study-specific	mixed y x##study study: , residuals(,by(study)) reml df(kr) $b_{ipdmetan}$, study(study) re poolvar(_cons): regress y x
VII	$y_{ij} = \alpha_i + \beta_i x_{ij} + \epsilon_{ij}$ $\alpha_i = \alpha + a_i$ and $\beta_i = \beta + b_i$ where $(a_i, b_i) \sim \text{BVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \lambda & \rho \\ \rho & \tau \end{bmatrix}\right)$	Yes Yes	Study-specific	mixed y x study: x, residuals(,by(study)) reml b_{mvmeta_make} regress y x, by(study) saving(data) replace names(b v) then b_{mvmeta} b v, reml

^aWhile the model can technically be fitted in a two-stage way, it only obtains an estimate of α , not β so is practically useless for meta-analysis of treatment effects.

^bNote that ipdmetan and mvmeta are user-written Stata packages, which can be installed in Stata using “ . ssc install command_name.”

model is used; for example, in model I, α represents mean outcome (because treatment is coded $-0.5/+0.5$ and it is assumed that randomisation is 1:1).

There are 3 possibilities for modelling the main effect (intercept) of study:

$\alpha_i = \alpha, \forall i$	a single constant; studies share an intercept
$\alpha_i = \alpha + a_i$	fixed-study intercept
$\alpha_i = \alpha + a_i$ where $a_i \sim N(0, \lambda^2)$	random-study intercept

Similarly, for the treatment effect,

$\beta_i = \beta, \forall i$	common treatment effect (no interaction)
$\beta_i = \beta + b_i$	fixed treatment effects which are different for each study (fixed treatment-by-study interaction terms)
$\beta_i = \beta + b_i$ where $b_i \sim N(0, \tau^2)$	random treatment effect (random treatment-by-study interaction)

Let $\varepsilon_{ij} \sim N(0, \sigma_i^2)$. With a one-stage analysis, σ_i^2 can be allowed to vary by study or restricted such that $\sigma_i^2 = \sigma^2 \forall i$ or some combination of these. For example, a fixed-effects model based on ordinary least squares (OLS) implies $\alpha_i = \alpha + a_i$ (with a_i representing fixed differences in intercepts for $i = 1, \dots$), $\beta_i = \beta$, and $\sigma_i^2 = \sigma^2$. Two-stage analysis by default allows σ_i^2 to vary by i (we have referred to this as *default* because, at the time of writing, we believe all software packages which perform stage two do so under this assumption). However, the assumption that $\sigma_i^2 = \sigma^2$ can also be invoked with a two-stage analysis (see Olkin and Sampson⁵). To do this, relative study weights are proportional to a function of n_i rather than of $\hat{\sigma}_i^2$. Weighting in this way exactly recovers the one-stage OLS estimate.¹⁵ We return to this point later.

We distinguish here between two types of variance we will refer to, particularly for understanding the workings provided in the appendices. We denote the true variance of the estimator of β by $\text{Var}(\hat{\beta})$. In the frequentist sense, this is the long-run variance of $\hat{\beta}$ under repeated sampling. Second, the estimated variance is denoted by $\widehat{\text{Var}}(\hat{\beta})$. This is what is estimated in a specific realisation of a meta-analysis. A good variance estimator should have expectation $\text{Var}(\hat{\beta})$.

2.2 | Flexibility and estimands

Note the column of Table 1 stating whether the model can be fitted in one stage (top) or two (bottom). Of the models considered, (I), (III), and (IV) cannot be fitted in two stages. Further, although model (VI) can be fitted in two stages, it does not provide an estimate of β . This restriction is what is meant when one-stage approaches are promoted for their flexibility.

Note that, provided that the available evidence is representative and the model specific assumptions hold, any of the models listed in Table 1 could provide an unbiased estimate of the treatment effect and its variance. The practical settings for which this is true may be rather limited. For example, if each study used simple randomisation and the same allocation ratio (not necessarily 1:1), then $\hat{\beta}$ would be unbiased for all the models in Table 1. The interesting point is that this depends only on the design of the included studies and not on the true data-generating model. We refer interested readers to Parzen et al¹⁶ or Kahan and Morris.¹⁷

3 | A COMPARISON OF ONE- AND TWO-STAGE APPROACHES FOR LINEAR MODELS

3.1 | Fixed-effects model (model II)

We wish to compare the variance of one- and two-stage fixed-effect meta-analysis. There are 2 approaches to this.

1. Compare the variance of “default” or most natural one- and two-stage fixed-effect estimators.
2. Compare the variance of one- and two-stage estimators when the models and associated assumptions are the same.

The former is an apples-vs-oranges comparison. It is interesting because one-stage fixed-effect estimators are most naturally estimated by OLS (this is the default in all major software), which assumes a common variance across studies (previously described as “scarcely credible” in the context of meta-analysis).¹⁵ Meanwhile the two-stage inverse-variance estimator most naturally assumes heterogeneity of variances across studies (again this is the default in software). Assumptions about within-study variances impact on the relative weights of studies and thus on the overall estimate of β and its variance. Results may differ between one- and two-stage analysis due to subtle differences in modelling assumptions, such as this one. It is important not to attribute this to the number of stages in computation. Therefore, the comparison

of defaults is very much a comparison of the modelling assumptions, rather than of the procedure by which the models are estimated (ie, in one or two stages).

The most natural one-stage analysis is to fit a model based on OLS, where the theory is well understood. This assumes σ_i^2 is common across studies. The theory is less well understood when σ_i^2 are allowed to be study specific, and the model is estimated based on variance-weighted least squares. The fact that the `mixed` command is needed to fit this model demonstrates that it is not a natural one-stage model to fit. We derive the one-stage estimator of β and variance estimator allowing study-specific σ_i^2 in Appendix A.1. These turn out to be the same as the two-stage inverse-variance estimator and the usual asymptotic inverse-variance formula:

$$\widehat{\text{Var}}(\hat{\beta}) = \sum_{i=1}^I \frac{4\hat{\sigma}_i^2}{n_i}. \quad (1)$$

For details, see Appendix A.1.

Thus, when we allow study-specific variances, computing our estimates in one or two stages would lead to identical point estimates and variance estimates. (In fact, the variance estimators are biased downwards in finite samples because they assume that the $\hat{\sigma}_i^2$ are known rather than estimated, as is assumed in Mathew and Nordström.⁶) This result is related to the result of Olkin and Sampson,⁵ when both models impose a shared σ_i^2 , and to Lunn et al,¹⁸ who present a two-stage strategy for fitting a full Bayesian model.

We focus on comparisons of variance here and, to do so, needed to derive some new theoretical results. Appendix A works through the mathematics, while the rationale and results are given here. The overall aim is to calculate the expected value of the variance for the OLS and inverse-variance estimators. This involved:

1. Calculation of the expected value of the asymptotic variance formula (1) for the inverse-variance estimator (noting again that this variance estimator is biased downwards in small samples and so produces misleading conclusions) (Appendix A.2).
2. Derivation of a new variance formula which explicitly acknowledges that the $\hat{\sigma}_i^2$ are estimates and so is unbiased with practical sample sizes (Appendix A.3). This tends to the standard formula (1) as all $n_i \rightarrow \infty$ but is less biased when any n_i is small. This is

$$\text{Var}(\hat{\beta}) \simeq \frac{1}{\widehat{W}} + \sum_{i=1}^I \frac{2\sigma_i^4}{n_i - 1} \left(\frac{\hat{\beta}_i - \hat{\beta}}{\widehat{W}} \right)^2, \quad (2)$$

where $1/\widehat{W}$ is the standard inverse-variance formula (1).

3. Calculation of the expectation of the new small-sample variance formula (2) (Appendix A.4).

The expected value of the asymptotic and small-sample variance formulas are then compared with the version that assumes common $\hat{\sigma}_i^2$.

Results are shown in Figure 1, showing very little effect of the number of studies I (the 40 translucent curves are almost on top of each other) but some effect of the number of patients per study n_i . However, this effect becomes small with more than about 25 patients per study. This is shown for scenarios where all studies contain the same number of participants but could equally be done with specific incidences of unbalanced data.

This demonstrates that assuming σ_i^2 is the same across studies reduces variance of $\hat{\beta}$ when n_i is small. This being the assumption underlying OLS, it is typically associated with one-stage analysis. Two-stage fixed-effects meta-analysis typically allows σ_i^2 to be study specific. These are merely defaults, and it is not necessary for either procedure to use the default assumption.

3.2 | Random-effects model (model V)

Here, we consider model V of Table 1, often designated as “random-effects” meta-analysis, because the I treatment effects are assumed to follow a probability distribution. There is a bewildering range of competing two-stage estimators for this model.³

We will consider the restricted maximum likelihood estimator. It is attractive because it is more efficient than moment-based estimators such as DerSimonian and Laird's method but can provide consistent estimates of variance parameters. For finite-sample inference with restricted maximum likelihood (REML) estimation, Kenward and Roger's approximate correction to the covariance matrix and degrees of freedom is the best available option.¹⁹

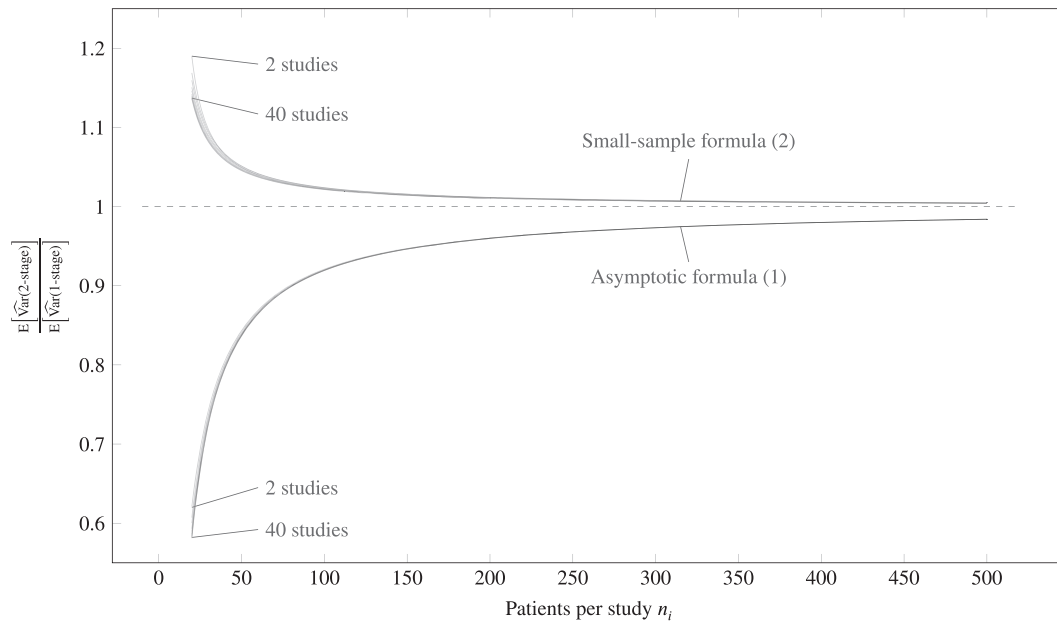


FIGURE 1 Ratio of two- to one-stage expected value of variances, using the asymptotic variance formula (1) (lines below 1) and the small sample formula (2) (lines above 1). The plot shows 39 translucent lines as the number of studies increases from $i = 2, \dots, 40$. The lines for smaller i are higher for both panels, but differences are negligible

It is possible to fit the random-effects meta-analysis REML realisation of the Kenward–Roger correction in one stage via `mixed` in Stata or `proc mixed` in SAS. The covariance and degrees of freedom for two-stage meta-analysis of a single parameter has not previously been derived; we provide the derivation in Appendix B.

The one- and two-stage versions of these models were compared in a simulation study. The simulation design follows. This is structured as aims, data-generating mechanism, methods, estimand, and performance measures.

- **Aims**

The aim of the simulation study is to compare one- and two-stage procedures in terms of (1) precision of meta-analysis with random treatment effects and fixed-study intercepts and (2) coverage of 95% confidence intervals. Both are with respect to β , the overall average treatment effect.

- **Data-generating mechanism**

Individual participant data were simulated for independently 2 up to 40 studies from model V of Table 1, that is, studies had fixed intercepts and random treatment effects (drawn from a Gaussian distribution). One thousand simulated datasets were produced using Stata 14's default random-number generator (64-bit Mersenne twister). For the data-generating mechanism we used, parameter values were set to $\alpha = a_i = 0 \forall i$, $\beta = 0$ with $E[\hat{\tau}^2] = 1$, and $E[\hat{\sigma}^2] = 1$ with $\text{Var}(\hat{\sigma}^2) = 50/\chi_{50}^2$ (where χ_{50}^2 denotes a random draw from a χ^2 distribution on 50 df). Study sizes were unbalanced. Following Rucker et al,²⁰ based on Galandi et al,²¹ study sizes n_i were drawn from a log-normal distribution with $E(\log n_i) = 3.798$ and $\text{SD}(\log n_i) = 1.104$, with n_i rounded to the nearest integer or to 20 for any drawn values < 20 .

- **Methods**

One-stage analysis was done in SAS 9.3 using `proc mixed`, while two-stage analysis³ used `ipdmetan` in Stata 14, based on the same simulated datasets. The choice of SAS for one-stage models was for computational speed: Both packages return the same results, but SAS currently gets there faster. The only difference in the methods is that the two-stage variance estimates were based on expected information and the one-stage on observed.

- **Estimand**

The estimand of interest is the meta-analytic estimate of average overall treatment effect β . As earlier mentioned, we acknowledge that this is not an adequate summary of a meta-analysis but is usually the parameter of central interest.

- **Performance measures**

The key performance measure is precision: the inverse of the empirical variance of $\hat{\beta}$.²² This is estimated for two-stage relative to one-stage (see White²²) and presented as “% gain”. The empirical variance is based only on the REML point estimate and so does not depend on the Kenward–Roger adjustment. To assess the new variance and df adjustments, we also compare the coverage of nominal 95% confidence intervals.

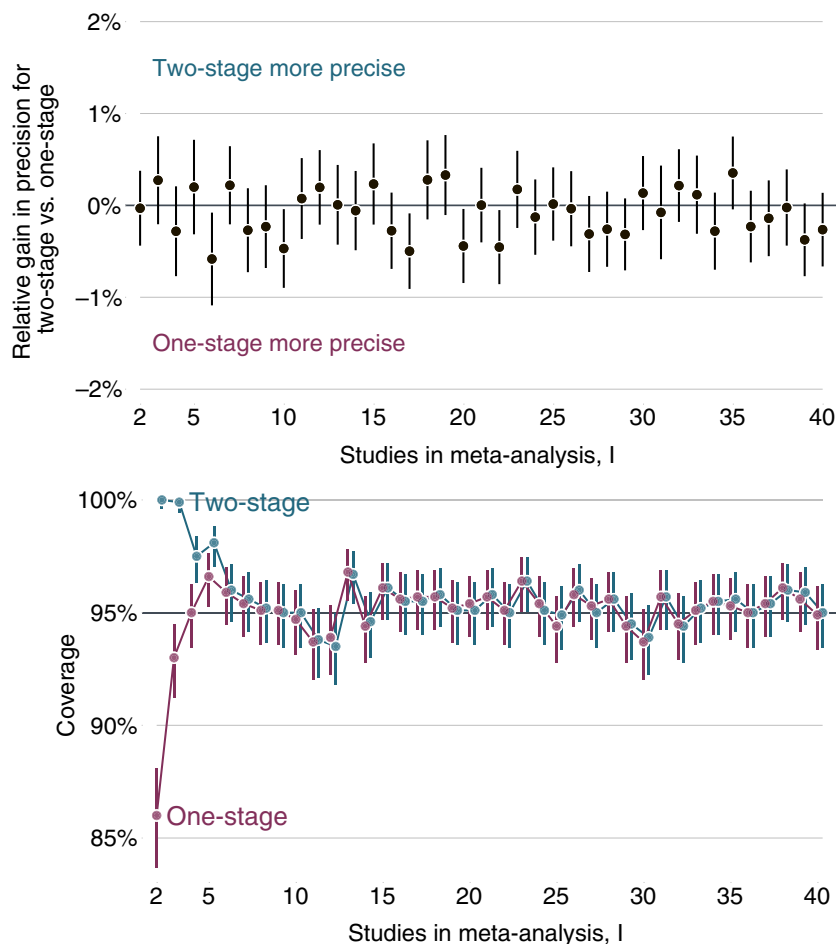


FIGURE 2 Simulation study results from 1000 repetitions plotted for 2 up to 40 studies in a meta-analysis. Upper panel: relative % increase in precision for two- vs one-stage restricted maximum likelihood estimation. Lower panel: coverage of Kenward–Roger confidence intervals

The Stata files required to generate data and run one- and two-stage analyses are included as a supplement to this article. Also included is the simulation analysis of the results (though the code to produce figures is not).²²

Figure 2 plots results for relative precision and coverage. All results are accompanied by 95% Monte Carlo confidence intervals to describe simulation uncertainty.²²

The upper panel displays relative precision, showing that the two-stage REML estimator has precision extremely similar to its one-stage counterpart under this data-generating mechanism. Although we are not able to obtain an analytic result to support whether this is a general or specific result, it appears to be in line with the equivalence of one- and two-stage for fixed-effects meta-analysis when the same models are fitted.

Coverage results (the lower panel of Figure 2) indicate that both methods tend to have good coverage, except when the number of studies is below 5. Inspection of the simulation results shows that variance estimates are accurate, on average very close to the true (empirical) variance. The conservative intervals appear to be due to the df estimated by two-stage Kenward–Roger approach being conservative (two-stage) and anti-conservative (one-stage). For both procedures, the variance formulas closely matched the empirical variance of the REML estimate of β . The issue with confidence interval coverage is then due to the df used. A comparison of the one- and two-stage df across simulation runs showed that the one-stage df were slightly higher than the two-stage counterpart. This is apparently due to the approximations involved in the Taylor expansions, which begin to fail in different directions at very small study sample sizes.

4 | ILLUSTRATION OF METHODS FOR INDANA META-ANALYSIS

The INDANA data includes IPD from 10 trials of antihypertensive drugs in patients at high risk of cardiovascular disease.^{23,24} Several outcome variables were collected, such as death, stroke, systolic blood pressure, diastolic blood pressure, and cholesterol. We focus here on systolic blood pressure (SBP), which was recorded annually at years 1 to 5 in

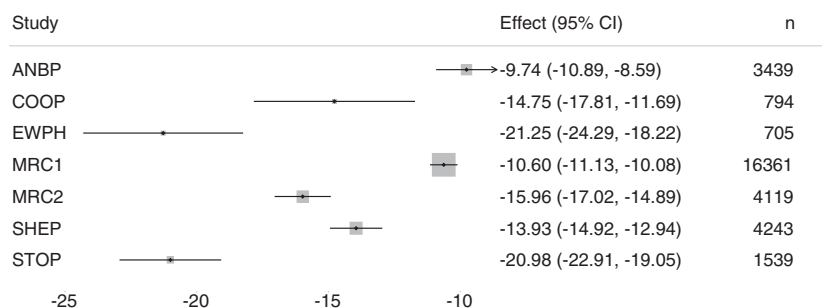


FIGURE 3 Forest plot of mean difference in systolic blood pressure at 1 year by randomised arm, for 7 trials in the INDANA data. The size of squares is based on fixed-effect weights with study-specific σ_i^2 . CI, confidence interval

TABLE 2 Results from meta-analysis of INDANA data: overall mean difference in systolic blood pressure under different meta-analysis models

M-A model	Estimation procedure	$\hat{\beta}$ (95% CI), mm Hg
Fixed effect, shared σ_i^2	One-stage	-12.5254 (-12.9128, -12.1382)
	Two-stage	-12.5255 (-12.9112, -12.1398)
Fixed effect, study-specific σ_i^2	One-stage ^a	-12.3404 (-12.7236, -11.9571)
	Two-stage ^a	-12.3404 (-12.7236, -11.9571)
Random effects, study-specific σ_i^2 (REML with Kenward-Roger adjustment)	One-stage	-15.2069 (-19.3984, -11.0153)
	Two-stage	-15.2068 (-18.5534, -11.8602)

^aConfidence intervals (CIs) computed using asymptotic variance formula (1).

seven of the 10 studies and never in the remaining three. As previously, we consider a simple analysis involving a linear regression of SBP on randomised group. Analysis was restricted to participants with observed values of SBP at 1 year.

The situation is shown in Figure 3, a forest plot of the within-study estimates of the mean difference in SBP at 1 year. The trials are highly heterogeneous in many ways, and this is apparent in the estimated treatment effects. A fixed-effect meta-analysis will thus produce a confidence interval that is too narrow and relates to what is arguably an ill-defined target parameter, as the assumption of a constant treatment effect is clearly not supported. Meanwhile, the smaller trials seem to have larger effects, so that the random-effects point estimate may be too large. Our purpose here is not to produce the most appropriate summary of the overall effect in terms of the point estimate and confidence interval; instead, we aim to demonstrate the impact of different modelling assumptions and of the number of stages used to fit a given model, on these quantities.

We consider three different modelling assumptions, matching the models considered in Section 2. First, a fixed-effect meta-analysis where σ_i^2 is assumed to be equal in all studies; second, a fixed-effect meta-analysis under the assumption that σ_i^2 are different across studies; and third, a random-effects meta-analysis also allowing σ_i^2 to vary across studies. Table 2 gives the results of applying one- and two-stage procedures for parameter estimation. Results are intentionally given to 4 decimal places. We see clear differences between different modelling assumptions, the largest difference being between the fixed- and random-effects models. However, the different procedures produce almost identical estimates of β and very similar confidence intervals.

5 | DISCUSSION

Focusing on practical sample sizes (rather than asymptotic results), this paper attempts to resolve the debate about using one- or two-stage estimation methods in meta-analysis. We have considered Gaussian data and demonstrated theoretically (for fixed effect) and via simulation (for random-effects models), that provided the same underlying model is used, inference from one- and two-stage procedures is practically equivalent for 2 models which are practically relevant.

To make progress, we focused on a very simple data structure and considered combining evidence from randomised studies. We considered the main arguments regarding “flexibility” and precision. Table 1 shows that a wider range of statistical models can be fitted in a one-stage framework; however, for many meta-analyses, this flexibility is not needed.

To study precision, we derived some new results. Specifically, the one-stage fixed-effects estimator of β which allows for study-specific variances; a new variance estimator which admits that $\hat{\sigma}_i^2$ are estimated and not known (again for the fixed-effect model allowing study-specific variances); and estimated the expectation of this formula, comparing it to the

expectation of the OLS variance, where $\hat{\sigma}_i^2$ is assumed to be equal across all studies. For random-effect meta-analysis, we also calculated and implemented the Kenward–Roger small sample correction.

In practical sample sizes, our work shows that the arguments regarding precision are largely redundant, as previously shown asymptotically for more general models.⁷ Mathew and Nördstrom claim to show that a one-stage analysis will always be at least as precise as two-stage, but the only example of an inequality is for a meta-analysis with no study effects at all, which is of no practical interest.⁶ Appendix A.1 shows that the fixed-effect model can be fitted using weighted one-stage estimation, which returns the same estimates as would be obtained through inverse-variance weighted two-stage estimation. Meanwhile, Senn¹⁵ shows how OLS estimates can be recovered from summary data in two stages. The results displayed in Figure 1 show that the latter is more precise than the former. Thus, analysis in two stages may be less precise, more precise, or identical to two stages, depending on the modelling assumptions made by the computational procedure. In summary, these papers do not contradict our key finding, which is that for equivalent models, fitting by a one- or two-stage procedure gives practically equivalent results.

Improvements in precision must be attributed to different modelling assumptions, that is, the precision argument for one-stage procedures is actually about *flexibility* again. A different model may be more or less precise: In general, we buy information with assumptions. We argue that model choice should never be based on precision. Opting to fit a model in one stage or two based on the perceived efficiency is in fact doing exactly this. In practice, it is important to make modelling assumptions that are tenable and fit the model as using whatever procedure allows for these assumptions. For the assumptions which may change between one- and two-stage procedures, see Burke et al.⁹

Due to the simple data structure considered, we have not touched on covariate adjustment here. Adjusting for the main effects of covariates may be desirable, either to obtain a valid estimate of $\text{Var}(\hat{\beta}_i)$ (for example, where covariates have been balanced by design²⁵) or to increase power.²⁶ Here, a one-stage model can estimate covariate effects to have the same values across all studies. Two-stage analysis relaxes this assumption by fitting a model which implies separate covariate effects for each study, which we conjecture is the correct way to ensure a valid estimate of $\text{Var}(\hat{\beta}_i)$. To fit a one-stage version of this model, covariate–study interaction terms need to be included and, if the modelling assumptions are identical, our results would hold. If the covariate effects truly differ by study, then the model that incorporates this (whether fitted by a one- or two-stage procedure) will tend to give more precise estimates; the converse is also true.

For meta-analysis of patient-level treatment–covariate interactions, using the two-stage procedure of fitting a model including interactions and combining these within-study interactions in a second-stage meta-analysis, guards against ecological bias.^{27–29} As always, a practically equivalent model can be fitted using a one-stage procedure; however, in this setting, care needs to be taken over the parameterisation and study-specific covariate means must be adjusted for.^{28,29} This may not be intuitive, and in this context, a one-stage analysis should aim to estimate the same effect as a two-stage analysis. Here, a model that can be fitted in one stage but not two may actually be misleading (“deluded” in the parlance of Fisher et al).²⁸

For practical meta-analysis, a clear description of the model and assumptions is important. See Burke et al⁹ for a description of practical assumptions which may differ between one- and two-stage meta-analysis: The relevance of each of these should be considered and described.

When fitting a model in two stages, it is important to describe whether a fixed- or random-effects model is being used. For fixed-effect models, the assumption about common or study-specific $\hat{\sigma}_i^2$ then needs to be described. For random effects, the random-effects model should be described. The clearest way of doing this is in terms of the estimator of tau squared (eg, DerSimonian–Laird or REML³) followed by any additional details such as use of the Hartung–Knapp/Sidik–Jonkman or Kenward–Roger variance correction.

When fitting a model in one stage, some modelling aspects to describe may be whether random effects were estimated and the parameters they were attached to; whether $\hat{\sigma}_i^2$ was shared or study specific; any correction to the covariance matrix; whether parameters for covariate effects were constrained to be equal or allowed to vary across studies; and how interactions were estimated.

We encourage those involved in meta-analysis methodology to focus more on consideration of models and assumptions and less on the procedures used for parameter estimation. The procedure used (one or two stages) is a computational tool; the model being fitted is what the tool aims to construct, and a tool that does the required job conveniently should be favoured.

ACKNOWLEDGEMENTS

We are grateful to the investigators who provided the INDANA data and, in turn, to all the trial teams and patients involved in the various studies.

This work was supported by the Medical Research Council (grant numbers MC_UU_12023/21 and MC_UU_12023/29).

CONFLICTS OF INTEREST

The meta-analysis group of the MRC Clinical Trials Unit at UCL have historically used two-stage approaches. Defending the approach may buttress our colleagues' work. DJF wrote the `ipdmeta` Stata package, which performs two-stage meta-analysis; our colleague Ian White wrote the `mvmeta` Stata package; defending two-stage approaches may increase the use of these package and citations of the articles. TPM came into this research expecting one-stage analysis to be better than two-stage and now favours neither approach other than the most convenient. The authors declare that they have no other conflicts of interest.

ORCID

Tim P. Morris  <http://orcid.org/0000-0001-5850-3610>

Michael G. Kenward  <http://orcid.org/0000-0003-0808-4192>

REFERENCES

1. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval Health Prof.* 2002;25(1):76-97.
2. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ.* 2010;340:c221.
3. Fisher DJ. Two-stage individual participant data meta-analysis and generalized forest plots. *Stata J.* 2015;15(2):369-396.
4. Tierney JF, Vale C, Riley R, et al. Individual participant data (IPD) meta-analyses of randomised controlled trials: guidance on their use. *PLoS Med.* 2015;12(7):e1001855+.
5. Olkin I, Sampson A. Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics.* 1998;54(1):317-322.
6. Mathew T, Nordstrom K. Comparison of one-step and two-step meta-analysis models using individual patient data. *Biom J.* 2010;52(2):271-287.
7. Lin DY, Zeng D. On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika.* 2010;97(2):321-332.
8. Lin DY, Zeng D. Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. *Genet Epidemiol.* 2010;34(1):60-66.
9. Burke DL, Ensor J, Riley RD. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Statist Med.* 2017;36(5):855-875.
10. Simmonds M, Stewart G, Stewart L. Decade of individual participant data meta-analyses: a review of current practice. *Contemp Clin Trials.* 2015;45(Pt A):76-83.
11. Debray TPA, Moons KGM, Valkenhoef G, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Res Synth Methods.* 2015;6(4):293-309.
12. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Series A (Statistics in Society).* 2009;172(1):137-159.
13. Senn S. The many modes of meta. *Drug Inf J.* 2000;34:535-54.
14. Tudur Smith C, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. *Stat Med.* 2005;24(9):1307-1319.
15. Senn SJ. Hans van Houwelingen and the art of summing up. *Biom J.* 2010;52(1):85-94.
16. Parzen M, Lipsitz SR, Dear KBG. Does clustering affect the usual test statistics of no treatment effect in a randomized clinical trial? *Biom J.* 1998;40(4):385-402.
17. Kahan BC, Morris TP. Assessing potential sources of clustering in individually randomised trials. *BMC Med Res Methodol.* 2013;13(1):58+.
18. Lunn D, Barrett J, Sweeting M, Thompson S. Fully Bayesian hierarchical modelling in two stages, with application to meta-analysis. *J R Stat Soc Ser A Stat Soc Ser C (Appl Stat).* 2013;62(4):551-572.
19. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics.* 1997;53:983-997.
20. Rücker G, Schwarzer G, Carpenter JR. Arcsine test for publication bias in meta-analyses with binary outcomes. *Stat Med.* 2008;27(5):746-763.
21. Galandi D, Schwarzer G, Antes G. The demise of the randomised controlled trial: bibliometric study of the German-language health care literature, 1948 to 2004. *BMC Med Res Methodol.* 2006;6:30.
22. White IR. `simsum`: analyses of simulation studies including Monte Carlo error. *Stata J.* 2010;10(3):369-385.
23. Gueyffier F, Boutitie F, Boissel JP, et al. INDANA: a meta-analysis on individual patient data in hypertension. Protocol and preliminary results. *Thérapie.* 1995;50(4):353-362.
24. Gueyffier F, Boutitie F, Boissel JP, et al. Effect of antihypertensive drug treatment on cardiovascular outcomes in women and men. A meta-analysis of individual patient data from randomized, controlled trials. The INDANA Investigators. *Ann Intern Med.* 1997;126(10):761-767.
25. Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. *Stat Med.* 2012;31(4):328-340.

26. Kahan BC, Jairath V, Doré CJ, Morris TP. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials*. 2014;15(1):139+.
27. Fisher DJ, Copas AJ, Tierney JF, Parmar MKB. A critical review of methods for the assessment of patient-level interactions in individual participant data meta-analysis of randomized trials, and guidance for practitioners. *J Clin Epidemiol*. 2011;64(9):949-967.
28. Fisher DJ, Carpenter JR, Morris TP, Freeman SC, Tierney JF. Meta-analytical methods to identify who benefits most from treatments: daft, deluded, or deft approach? *BMJ*. 2017;356:j573+.
29. Hua H, Burke DL, Crowther MJ, Ensor J, Tudur Smith C, Riley RD. One-stage individual participant data meta-analysis models: estimation of treatment-covariate interactions must avoid ecological bias by separating out within-trial and across-trial information. *Statist Med*. 2017;36(5):772-789.
30. Kendall MG, Stuart A. *The Advanced Theory of Statistics*, Vol. 1: Distribution Theory. 10. :232+. Bucks, UK: Charles Griffin & Company Limited third ed; 1969.
31. Kenward MG, Roger JH. An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Comput Stat Data Anal*. 2009;53(7):2583-2595.

How to cite this article: Morris TP, Fisher DJ, Kenward MG, Carpenter JR. Meta-analysis of Gaussian individual patient data: two-stage or not two-stage? *Statistics in Medicine*. 2018;1–20. <https://doi.org/10.1002/sim.7589>

APPENDIX A: PROPERTIES OF WEIGHTED ESTIMATORS

The below sections derive the variance of the fixed-effects inverse-variance estimator. Unless otherwise stated, summations are over all i , the study index.

A.1 | Variance estimator for a one-stage analysis assuming varying variances

The fixed effects model is $Y_{ij} = \alpha_i + \beta X_{ij} + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim N(0, \sigma^2)$, i indexes studies and j indexes patients. This model is estimated by $(X'X)^{-1}X'Y$ and covariance matrix by $(X'X)^{-1}\hat{\sigma}^2$. We require the expectation of $\widehat{\text{Var}}(\hat{\beta})$.

Assume that we have I studies, each with n_i patients. Define the $\sum n_i \times (I + 1)$ design matrix

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & \dots & 0 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

The first I columns indicate (exclusive) membership in study $i = 1, \dots, I$ and the $(I + 1)$ th indicates treatment assignments. For a one-stage analysis, $\widehat{\text{Var}}(\hat{\beta})$ is the $(I + 1)$ th diagonal element of $\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$. First, $(\mathbf{X}'\mathbf{X})$ is $(I + 1) \times (I + 1)$:

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n_1 & 0 & \dots & n_1/2 \\ 0 & n_2 & & n_2/2 \\ \vdots & \vdots & \ddots & \vdots \\ n_1/2 & n_2/2 & \dots & \sum n_i/2 \end{pmatrix}.$$

The inverse is

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} & & & \dots & -2/\sum n_i \\ & & & & -2/\sum n_i \\ & & & & \vdots \\ -2/\sum n_i & -2/\sum n_i & \dots & & 4/\sum n_i \end{pmatrix}.$$

We are interested in the last element (variance of the treatment effect), which is $\widehat{\text{Var}}(\hat{\beta}) = \frac{4\sigma^2}{\sum n_i}$.

The shared- σ^2 assumption can be relaxed in the following way: Construct a $\sum n_i \times \sum n_i$ diagonal matrix of weights W , where the nonzero elements are $1/\hat{\sigma}_i^2$; the variance-covariance matrix is then estimated as $(X'WX)^{-1}\hat{\sigma}^2$. The matrix $X'WX$ is

$$X'WX = \left(\begin{array}{cc|c} \frac{n_1}{\sigma_1^2} & 0 & \frac{n_1}{2\sigma_1^2} \\ & \ddots & \\ 0 & & \frac{n_I}{2\sigma_I^2} \\ \hline \frac{n_1}{2\sigma_1^2} & & \sum \frac{n_i}{2\sigma_i^2} \end{array} \right).$$

The solid lines show how the matrix can be partitioned to invert the $(I+1)$ th diagonal element (variance of the treatment effect), which is

$$\left(\sum \frac{n_i}{4\sigma_i^2} \right)^{-1}.$$

This is identical to the standard inverse-variance formula used by a two-stage procedure. The algebra can also be worked through for the estimator of β to include weights W , that is $(X'WX)^{-1}X'Wy$. The one- and two-stage procedures then give identical results for $\hat{\beta}$. This complements the result described and worked through in Senn,¹⁵ taken from Olkin and Sampson,⁵ which shows that a two-stage procedure can obtain one-stage estimates when modified to make the same assumptions of equal $\sigma_i^2 \forall i$.

A.2 | Expected value of standard two-stage variance estimator

Let $i = 1, \dots, I$ index studies and $j = 1, \dots, n_i$ index individuals. The unadjusted estimator of β for active vs control is

$$\hat{\beta}_i = \frac{2}{n_i} \sum_{\text{actv.}} y_j - \frac{2}{n_i} \sum_{\text{cont.}} y_j$$

with variance for study i estimated by

$$\widehat{\text{Var}}(\hat{\beta}_i) = \frac{4\hat{\sigma}_i^2}{n_i}.$$

This first stage is repeated within each study. Define weights $\hat{w}_i = n/4\hat{\sigma}_i^2$ and let

$$\widehat{W} = \sum_{i=1}^I \hat{w}_i.$$

For the second stage, the inverse-variance estimator of the overall treatment effect is

$$\hat{\beta} = \frac{\sum \hat{w}_i \hat{\beta}_i}{\widehat{W}}. \tag{A1}$$

The variance of β is almost always estimated by

$$\widehat{\text{Var}}(\hat{\beta}) = \frac{1}{\widehat{W}} = \sum_{i=1}^I \frac{4\hat{\sigma}_i^2}{n_i}. \tag{A2}$$

We now calculate the expectation of $\widehat{\text{Var}}(\hat{\beta})$ for 2 studies. Note that

$$\begin{aligned}\hat{\sigma}_i^2 &\sim \frac{\chi_{n_i-2}^2 \sigma_i^2}{n_i - 2}, \\ \text{so } E[\hat{\sigma}_i^2] &= \sigma_i^2, \\ \text{and } \text{Var}(\hat{\sigma}_i^2) &= \frac{2\sigma_i^4}{n_i - 1}. \\ \text{Write } \hat{\sigma}_i^2 &\sim \sigma_i^2 + \varepsilon_i, \\ \text{where } \varepsilon_i &\sim \left(0, \frac{2\sigma_i^4}{n_i - 1}\right).\end{aligned}\tag{A3}$$

For a meta-analysis of two studies, our variance estimator is

$$\widehat{\text{Var}}(\hat{\beta}) = 4(n_1 \hat{\sigma}_1^{-2} + n_2 \hat{\sigma}_2^{-2})^{-1}.$$

Substituting $x = \hat{\sigma}_1^2$, $y = \hat{\sigma}_2^2$, write $\widehat{\text{Var}}(\hat{\beta})$ as

$$f(x, y) = 4(n_1 x^{-1} + n_2 y^{-1})^{-1}.$$

The first and second partial derivatives for x and y are

$$\begin{aligned}\frac{\partial f}{\partial x} &= n_1 x^{-2} f^2, \\ \frac{\partial f}{\partial y} &= n_2 y^{-2} f^2, \\ \frac{\partial^2 f}{\partial x^2} &= -2n_1 x^{-3} f^2 + 2n_1^2 x^{-4} f^3, \\ \frac{\partial^2 f}{\partial y^2} &= -2n_2 y^{-3} f^2 + 2n_2^2 y^{-4} f^3.\end{aligned}$$

Denote f_x after evaluating $\frac{\partial f}{\partial x}$ at $\mu_x = \sigma_1^2$, $\mu_y = \sigma_2^2$, and similarly for f_y, f_{xx}, f_{yy} . By a Taylor expansion,

$$\begin{aligned}f(x, y) &\simeq f(\mu_x, \mu_y) + (\varepsilon_x, \varepsilon_y) \begin{pmatrix} f_x \\ f_y \end{pmatrix} + \frac{1}{2} (\varepsilon_x, \varepsilon_y) \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} \begin{pmatrix} \varepsilon_x \\ \varepsilon_y \end{pmatrix} \\ &= f(\mu_x, \mu_y) + \varepsilon_x f_x + \varepsilon_y f_y + \frac{1}{2} (\varepsilon_x^2 f_{xx} + 2\varepsilon_x \varepsilon_y f_{xy} + \varepsilon_y^2 f_{yy}).\end{aligned}$$

Note that $E[\varepsilon_x] = E[\varepsilon_y] = 0$, and studies are uncorrelated, so $E[\varepsilon_x \varepsilon_y] = 0$, then

$$E[\text{Var}(\hat{\beta})] \simeq f(\mu_x, \mu_y) + \frac{1}{2} (E[\varepsilon_x^2] f_{xx}) + \frac{1}{2} (E[\varepsilon_y^2] f_{yy}).\tag{A4}$$

Within each study,

$$E[\varepsilon_i^2] = \frac{2\sigma_i^4}{n_i - 1}.$$

Generalising to I studies,

$$\begin{aligned}f(x_1, \dots, x_I) &= 4(n_1 x_1^{-1} + \dots + n_I x_I^{-1})^{-1} \\ f_{x_i x_i} &= 2(n_i x_i^{-4} f^3 - n_i x_i^{-3} f^2).\end{aligned}$$

This gives us that

$$\begin{aligned}
 E[\widehat{\text{Var}}(\hat{\beta})] &\simeq f(\sigma_1^2, \dots, \sigma_I^2) + \frac{1}{2} \sum_{i=1}^I E[\varepsilon_i^2] f_{\sigma_i^2, \sigma_i^2} \\
 &= 4 \left(\sum_{i=1}^I n_i \sigma_i^{-2} \right)^{-1} + \frac{1}{2} \sum_{i=1}^I \left(\frac{2\sigma_i^2}{n_i - 1} f_{\sigma_i^2, \sigma_i^2} \right).
 \end{aligned}
 \tag{A5}$$

This result is used to compare the expectation of (A2) to the OLS fixed-effect estimator (results given in Figure 1, upper panel).

A.3 | An improved variance estimator to partner the inverse-variance estimator of $\hat{\beta}$

Because the standard variance formula given in (A2) is biased downwards in finite samples, we derive an improved estimator for the variance of (A1). Let

$$\hat{\beta} = f(\hat{\beta}_1, \hat{\sigma}_1, \dots, \hat{\beta}_I, \hat{\sigma}_I) = \frac{\sum \hat{w}_i \hat{\beta}_i}{\widehat{W}}.$$

First derivatives wrt $\hat{\beta}_i$ and \hat{w}_i are

$$\begin{aligned}
 \frac{\partial f}{\partial \hat{w}_i} &= \frac{\beta_i}{\widehat{W}} - \frac{\hat{\beta}}{\widehat{W}}, \\
 \frac{\partial f}{\partial \hat{\beta}_i} &= \frac{\hat{w}_i}{\widehat{W}}.
 \end{aligned}$$

By a Taylor expansion of $f()$, we get

$$\text{Var}(\hat{\beta}) \simeq \left(\frac{\partial f}{\partial \hat{\beta}_1}, \frac{\partial f}{\partial \hat{w}_1}, \dots, \frac{\partial f}{\partial \hat{\beta}_I}, \frac{\partial f}{\partial \hat{w}_I} \right) \times \begin{pmatrix} \frac{4\hat{\sigma}_1^2}{n_1} & 0 & \dots & \mathbf{0} & \mathbf{0} \\ 0 & \frac{2\hat{\sigma}_1^4}{n_1-1} & & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & & \frac{4\hat{\sigma}_I^2}{n_I} & 0 \\ \mathbf{0} & \mathbf{0} & \dots & 0 & \frac{2\hat{\sigma}_I^4}{n_I-1} \end{pmatrix} \times \begin{pmatrix} \frac{\partial f}{\partial \hat{\beta}_1} \\ \frac{\partial f}{\partial \hat{w}_1} \\ \vdots \\ \frac{\partial f}{\partial \hat{\beta}_I} \\ \frac{\partial f}{\partial \hat{w}_I} \end{pmatrix}
 \tag{A6}$$

$$\begin{aligned}
 &= \sum \frac{1}{\hat{w}_i} \left(\frac{\hat{w}_i}{\widehat{W}} \right)^2 + \sum \frac{2\hat{\sigma}_i^4}{n_i - 1} \left(\frac{\hat{\beta}_i - \hat{\beta}}{\widehat{W}} \right)^2 \\
 &= \frac{1}{\widehat{W}} + \sum_{i=1}^2 \frac{2\sigma_i^4}{n_i - 1} \left(\frac{\hat{\beta}_i - \hat{\beta}}{\widehat{W}} \right)^2.
 \end{aligned}
 \tag{A7}$$

The bold 0s in (A6) are always equal to 0 in meta-analysis provided studies can be assumed independent. The roman 0s are set to 0 here because we are considering linear models, where an estimate and its variance are independent. Note that when β_i and $\text{Var}(\beta_i)$ cannot be assumed independent, as with binary outcomes, the roman 0s in (A6) need to be replaced with the correct terms.

(A7) provides a variance estimator for Gaussian data which improves on (A2) in finite samples.

A.4 | Expected value of the variance estimator (A7)

The estimator is given by A1, where

$$\begin{aligned}
 \hat{w}_i &= \frac{n_i}{4\hat{\sigma}_i^2} \sim \frac{n_i v_i X_i}{4\sigma_i^2} \\
 \text{for } X_i^{-1} &\sim \chi_{v_i}^2, \\
 \text{and } \hat{\beta}_i &\sim N\left(\beta; \frac{4\sigma_i^2}{n_i}\right).
 \end{aligned}$$

$\hat{\beta}_i$ and \hat{w}_i are assumed to be independent. Note that \hat{w}_i can be expressed as a function of the random variable X_i . Let X represent the set of all X_i .

We want the variance of $\hat{\beta}$.

$$\begin{aligned}\text{Var}(\hat{\beta}) &= E_X \{ \text{Var}(\hat{\beta}|X) \} + \text{Var}_X \{ E(\hat{\beta}_i|X) \} \\ &= E_X \left\{ \frac{\hat{w}_i^2 \frac{4\sigma_i^2}{n_i}}{(\sum \hat{w}_i)^2} \right\} + \text{Var}_X(\beta) \\ &= E_X \left\{ \frac{A}{B} \right\}, \\ \text{for } A &= \sum e_i X_i^2 \\ \text{and } B &= \left(\sum f_i X_i \right)^2, \\ \text{where } e_i &= \frac{n_i v_i^2}{4\sigma_i^2} \\ \text{and } f_i &= \frac{n_i v_i}{4\sigma_i^2} = \frac{e_i}{v_i}.\end{aligned}$$

The generic problem is that we want the expectation of the ratio

$$R = \frac{A}{B}$$

for A and B as defined above with arbitrary positive constants e_i and f_i .

Use the second-order expression³⁰

$$E(R) \simeq \frac{E(A)}{E(B)} - \frac{\text{Cov}(A, B)}{E(B)^2} + \frac{\text{Var}(B)E(A)}{E(B)^3}. \quad (\text{A8})$$

The calculation of the required moments follows.

$$\begin{aligned}E(A) &= \sum_i \frac{e_i}{(v_i - 2)(v_i - 4)}, \\ E(B) &= E_X \left\{ \left(\sum_i f_i X_i \right)^2 \right\} \\ &= E \left(\sum_i \sum_j f_i f_j X_i X_j \right) \\ &= \sum_i f_i^2 E(X_i^2) + \sum_{i \neq j} \sum_j f_i f_j E(X_i) E(X_j) \\ &= \sum_i \frac{f_i^2}{(v_i - 2)(v_i - 4)} + \sum_{i \neq j} \sum_j f_i f_j \frac{1}{(v_i - 2)(v_j - 2)}, \\ \text{Cov}(A, B) &= \text{Cov} \left\{ \sum e_i X_i^2, \left(\sum f_i X_i \right)^2 \right\} \\ &= \sum_i \sum_j \sum_k e_i f_j f_k \text{Cov}(X_i^2, X_j X_k) \\ &= \sum_i e_i f_i^2 \text{Var}(X_i^2) + \sum_{i \neq j} \sum_j 2e_i f_i f_j \text{Cov}(X_i^2, X_i X_j),\end{aligned}$$

$$\text{and Var}(X_i^2) = \frac{8(v_i - 5)}{(v_i - 2)^2(v_i - 4)^2(v_i - 6)(v_j - 2)},$$

$$\text{Cov}(X_i^2, X_i X_j) = \frac{4}{(v_i - 2)^2(v_i - 4)(v_i - 6)(v_j - 2)},$$

$$\begin{aligned} E(B^2) &= E \left\{ \left(\sum_i f_i X_i \right)^4 \right\} \\ &= E \left(\sum_i \sum_j \sum_k \sum_l f_i f_j f_k f_l X_i X_j X_k X_l \right) \\ &= \sum_i f_i^4 E(X_i^4) + 4 \sum_{i \neq j} \sum_j f_i^3 f_j E(X_i^3 X_j) \\ &\quad + 3 \sum_{i \neq j} \sum_j f_i^2 f_j^2 E(X_i^2) E(X_j^2) \\ &\quad + 6 \sum_{i \neq j} \sum_{j \neq k} \sum_k f_i^2 f_j f_k E(X_i^2) E(X_j) E(X_k) \\ &\quad + \sum_{i \neq j} \sum_{j \neq k} \sum_{k \neq l} \sum_l f_i f_j f_k f_l E(X_i) E(X_j) E(X_k) E(X_l) \\ &= \sum_i f_i^4 \frac{1}{(v_i - 2)(v_i - 4)(v_i - 6)(v_i - 8)} \\ &\quad + 4 \sum_{i \neq j} \sum_j f_i^3 f_j \frac{1}{(v_i - 2)(v_i - 4)(v_i - 6)(v_j - 2)} \\ &\quad + 3 \sum_{i \neq j} \sum_j f_i^2 f_j^2 \frac{1}{(v_i - 2)(v_i - 4)(v_j - 2)(v_j - 4)} \\ &\quad + 6 \sum_{i \neq j} \sum_{j \neq k} \sum_k f_i^2 f_j f_k \frac{1}{(v_i - 2)(v_i - 4)(v_j - 2)(v_k - 2)} \\ &\quad + \sum_{i \neq j} \sum_{j \neq k} \sum_{k \neq l} \sum_l f_i f_j f_k f_l \frac{1}{(v_i - 2)(v_j - 2)(v_k - 2)(v_l - 2)}. \end{aligned}$$

$$\text{Then, Var}(B) = E(B^2) - E(B)^2.$$

Substituting the expectations $E(A)$, $E(B)$, $\text{Cov}(A, B)$, and $\text{Var}(B)$ into (A8) gives an expression for the actual variance of $\hat{\beta}$.

This has been verified in simulation; the Mata code to calculate the formula and verify (A8) and its components is provided as a supplement. Also included is the log file. This is run for 7 studies where, arbitrarily, $f = [1, 2, 3, 4, 5, 6, 7]$ and $e = [10, 5, 3\frac{1}{3}, 2.5, 2, 1\frac{2}{3}, 1\frac{3}{7}]$. We simulated the X_i s from $1/\chi_{v_i}^2$ and calculated

$$\begin{aligned} A &= \sum_{i=1}^7 e_i X_i^2, \\ B &= \left(\sum_{i=1}^7 f_i X_i \right)^2, \\ \text{and } R &= \frac{A}{B}. \end{aligned}$$

We provide 2 simulation results. First, $\text{Var}(R)$ is estimated empirically from the simulated values; second, it is calculated using (A8) where empirical values of $E(A)$, $E(B)$, $\text{Var}(B)$, and $\text{Cov}(A, B)$ are substituted for the theoretical expectations. These results, and those based on the expectations, are given in Table A1 for 1×10^{10} repetitions. A tiny error in the

TABLE A1 Simulation-based check of variance formula (1×10^{10} repetitions)

	Empirical	Approximate empirical, based on (A8)	Theoretical based on (A8)
E(A)	...	0.02695	0.02695
E(B)	...	1.17552	1.17554
Var(B)	...	0.11525	0.11523
Cov(AB)	...	0.00157	0.00157
Var(A/B)	0.023618	0.023706	0.023704
Ratio vs empirical	1	1.00374	1.00364

approximation (A8) is seen, but the empirical and theoretical components of (A8) are near identical. This justifies the use of the theoretical formula to compare expected variances of the inverse-variance and OLS estimators (as in Figure 1).

APPENDIX B: KENWARD-ROGER CORRECTION FOR TWO-STAGE RANDOM-EFFECTS META-ANALYSIS ESTIMATED BY REML

This section derives the Kenward–Roger variance estimator and degrees of freedom for two-stage meta-analysis using the notation of Kenward and Roger.^{19,31} We first recap the general approach and notation before applying it to the meta-analysis model.

The general linear mixed-effects model can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e},$$

where \mathbf{y} is a $(k \times 1)$ vector of random variables, \mathbf{X} is a $(k \times p)$ matrix of known constants for the $(p \times 1)$ fixed-effect parameter vector $\boldsymbol{\beta}$, \mathbf{Z} is the $(k \times q)$ design matrix for the $(q \times 1)$ random-effects parameter vector $\boldsymbol{\gamma}$, and \mathbf{e} is a $(k \times 1)$ vector of random error terms.

We assume $E[\boldsymbol{\gamma}] = 0$, $E[\mathbf{e}] = 0$, and $\text{Cov}[\boldsymbol{\gamma}, \mathbf{e}] = 0$. Define \mathbf{D} as the $(q \times q)$ covariance matrix of the random-effects parameters in $\boldsymbol{\gamma}$, and \mathbf{R} as the $(k \times k)$ covariance matrix of \mathbf{e} . Then, $\boldsymbol{\Sigma}$, the $(k \times k)$ covariance matrix of \mathbf{y} , is equal to $\mathbf{Z}\mathbf{D}\mathbf{Z}^\top + \mathbf{R}$. After assuming normality of the random terms in the model, we obtain $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$.

Kenward and Roger describe $\boldsymbol{\Sigma}$ under the general linear mixed-effects model as being an unknown, block diagonal, $(k \times k)$ variance-covariance matrix whose elements are well-behaved functions of $\boldsymbol{\sigma}(r \times 1)$.¹⁹ The standard approach would be to use REML to find an estimate $\hat{\boldsymbol{\sigma}}$ for $\boldsymbol{\sigma}$ and proceed to find a REML estimator $\hat{\boldsymbol{\beta}} = \boldsymbol{\Phi}(\hat{\boldsymbol{\sigma}})\mathbf{X}^\top\boldsymbol{\Sigma}(\hat{\boldsymbol{\sigma}})^{-1}\mathbf{Y}$ for $\boldsymbol{\beta}$, with variance $\hat{\boldsymbol{\Phi}}(\hat{\boldsymbol{\sigma}}) = [\mathbf{X}^\top\boldsymbol{\Sigma}(\hat{\boldsymbol{\sigma}})^{-1}\mathbf{X}]^{-1}$, where $\boldsymbol{\Sigma}$ and hence $\boldsymbol{\Phi}$ are dependent on $\hat{\boldsymbol{\sigma}}$.

The standard REML estimator of $\text{Var}(\hat{\boldsymbol{\beta}})$, $\hat{\boldsymbol{\Phi}} = [\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{X}]^{-1}$, is the variance-covariance matrix of the asymptotic limiting distribution of $\hat{\boldsymbol{\beta}}$ based on the REML estimation of $\boldsymbol{\sigma}$. Kenward and Roger obtain a better approximation than $\hat{\boldsymbol{\Phi}}$ to the finite-sample variance-covariance matrix of $\hat{\boldsymbol{\beta}}$.

The corrected estimator can be written as

$$\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}} + 2\hat{\boldsymbol{\Phi}} \left\{ \sum_{i=1}^r \sum_{j=1}^r W_{ij} \left(\mathbf{Q}_{ij} - \mathbf{P}_i \hat{\boldsymbol{\Phi}} \mathbf{P}_j - \frac{1}{4} \mathbf{R}_{ij} \right) \right\} \hat{\boldsymbol{\Phi}} - \frac{1}{4} \sum_{s,t=1}^r W_{st} V_t \hat{\boldsymbol{\Phi}} \mathbf{P}_s \hat{\boldsymbol{\Phi}},$$

where

$$\begin{aligned} \mathbf{P}_i &= \mathbf{X}^\top \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i} \mathbf{X}, \\ \mathbf{Q}_{ij} &= \mathbf{X}^\top \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i} \boldsymbol{\Sigma} \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_j} \mathbf{X}, \\ \mathbf{R}_{ij} &= \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \sigma_i \partial \sigma_j} \boldsymbol{\Sigma}^{-1} \mathbf{X}, \end{aligned}$$

and W_{ij} is the (i, j) th element of the variance-covariance matrix of σ , that is, of $\mathbf{W} = \text{Var}[\sigma]$, which may be approximated by the inverse of the observed or expected information matrix. Kenward and Roger give the following expression for the expected information:

$$I(\hat{\sigma})_{ij} = \frac{1}{2} \text{tr} \left(\frac{\partial \Sigma^{-1}}{\partial \sigma_i} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_j} \Sigma \right) - \frac{1}{2} \text{tr} \left(2 \hat{\Phi} \mathbf{Q}_{ij} - \hat{\Phi} \mathbf{P}_i \hat{\Phi} \mathbf{P}_j \right).$$

The third term of $\hat{\Phi}_A$, and therefore V_i and \mathbf{S}_{ij} , may be ignored if the covariance matrix Σ is linear in σ . Those quantities are defined as follows:

$$V_i = \text{tr} \left\{ \mathbf{s} \frac{\partial \Sigma}{\partial \sigma_i} \right\} - 2 \text{tr} \left\{ \left(\mathbf{X}^\top \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_i} \mathbf{S} \mathbf{X} \right) \hat{\Phi} \right\} + \text{tr} \left\{ \left(\mathbf{X}^\top \mathbf{S} \mathbf{X} \right) \hat{\Phi} \left(\mathbf{X}^\top \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_i} \Sigma^{-1} \mathbf{X} \right) \hat{\Phi} \right\},$$

$$\mathbf{S}_{ij} = \Sigma^{-1} \left(\sum_{i,j=1}^r W_{ij} \frac{\partial^2 \Sigma}{\partial \sigma_i \partial \sigma_j} \right) \Sigma^{-1}.$$

B.1 | The random-effects meta-analysis model

For the random-effects meta-analysis model, \mathbf{y} consists of the I treatment effect estimates β_i , $\mathbf{X} = \mathbf{1}_i$, β includes the “pooled” effect size β , and Σ is diagonal with elements $\Sigma_{ii} = \hat{\sigma}_i^2 + \hat{\tau}^2$, where $\hat{\sigma}_i^2$ is the estimated variance of β_i and $\hat{\tau}^2$ is the estimated between-study heterogeneity.

We proceed by noting that Kenward and Roger's $\sigma_1 = \tau^2$ (note that this is distinct from σ_i , the within-study standard deviation) and also define $w_i = (\sigma_i^2 + \sigma_1)^{-1}$, so that $\Sigma_{ii} = w_i^{-1}$. Then,

$$\widehat{\text{Var}}(\hat{\beta}) = \hat{\Phi} = \frac{1}{\sum_i w_i},$$

$$\left(\frac{\partial \Sigma^{-1}}{\partial \sigma_1} \right)_{ii} = -w_i^2,$$

$$\mathbf{P}_1 = -\sum_i w_i^2,$$

$$\mathbf{Q}_{11} = \sum_i w_i^3.$$

Note that Σ is linear in σ_1 , so that $\frac{\partial^2 \Sigma}{\partial \sigma_1^2} = \mathbf{0}$. This immediately implies that $\mathbf{R}_{11} = \mathbf{0}$ and also that the entire third term of the above expression for $\hat{\Phi}_A$ is 0.

Then,

$$\mathbf{Q}_{11} - \mathbf{P}_1 \hat{\Phi} \mathbf{P}_1 = \sum_i w_i^3 - \frac{(\sum_i w_i^2)^2}{\sum_i w_i}$$

$$\Rightarrow \hat{\Phi}_A = \frac{1}{\sum_i w_i} + \frac{2W_{11}}{(\sum_i w_i)^2} \left[\sum_i w_i^3 - \frac{(\sum_i w_i^2)^2}{\sum_i w_i} \right]$$

$$= \frac{1}{\sum_i w_i} \left\{ 1 + 2W_{11} \left(\frac{\sum_i w_i^3}{\sum_i w_i} - \left[\frac{\sum_i w_i^2}{\sum_i w_i} \right]^2 \right) \right\},$$

where $W_{11} = \text{Var}[\sigma_1]$ as explained earlier.

The expected information works out to be

$$I_E(\hat{\sigma}_1) = \frac{1}{2} \sum_i w_i^2 - \frac{\sum_i w_i^3}{\sum_i w_i} + \frac{1}{2} \left[\frac{\sum_i w_i^2}{\sum_i w_i} \right]^2.$$

If we wish to use the observed information, this requires the second derivative of the log-likelihood with respect to σ_1 .

$$\log \mathcal{L}(\sigma_1) = \text{constant} + \frac{1}{2} \sum_i \log(w_i) - \frac{1}{2} \log \sum_i w_i - \frac{1}{2} \sum_i w_i (\beta_i - \hat{\beta})^2$$

Note that $\hat{\beta}$ is itself a function of σ_1 and must also be differentiated. Concentrating first on this term, let $v_i = \beta_i - \hat{\beta}$ so that the final term becomes $-\frac{1}{2} \sum_i w_i v_i^2$. Then,

$$-\frac{1}{2} \frac{\partial}{\partial \sigma_1} \sum_i w_i v_i^2 = - \sum_i w_i v_i v_i' - \frac{1}{2} \sum_i w_i' v_i^2.$$

Now, using the quotient rule and rearranging,

$$\begin{aligned} v_i' &= \frac{\partial}{\partial \sigma_1} \frac{\sum_i w_i \beta_i}{\sum_i w_i} = - \frac{\sum_i w_i^2 v_i}{\sum_i w_i} \quad \forall i \\ &\Rightarrow - \sum_i w_i v_i v_i' = \frac{\sum_i w_i^2 v_i}{\sum_i w_i} \sum_i w_i v_i. \\ \text{But } \sum_i w_i v_i &= \sum_i w_i \beta_i - \sum_i w_i \frac{\sum_i w_i \beta_i}{\sum_i w_i} = 0 \\ &\Rightarrow -\frac{1}{2} \frac{\partial}{\partial \sigma_1} \sum_i w_i v_i^2 = -\frac{1}{2} \sum_i w_i' v_i^2 = \frac{1}{2} \sum_i w_i^2 v_i^2 \\ &\Rightarrow -\frac{1}{2} \frac{\partial^2}{\partial \sigma_1^2} \sum_i w_i v_i^2 = -\frac{(\sum_i w_i^2 v_i)^2}{\sum_i w_i} - \sum_i w_i^3 v_i^2 \text{ by the product rule.} \end{aligned}$$

Now, we go back and calculate the second differentials of the first two (non-constant) terms:

$$\begin{aligned} \frac{1}{2} \frac{\partial^2}{\partial \sigma_1^2} \sum_i \log(w_i) &= \frac{1}{2} \sum_i w_i^2, \\ -\frac{1}{2} \frac{\partial^2}{\partial \sigma_1^2} \log \sum_i w_i &= \frac{1}{2} \left[\frac{\sum_i w_i^2}{\sum_i w_i} \right]^2 - \frac{\sum_i w_i^3}{\sum_i w_i}. \end{aligned}$$

Hence, the expression for observed information is

$$I_O(\sigma_1) = -\frac{\partial^2}{\partial \sigma_1^2} \log \mathcal{L}(\sigma_1) = \frac{[\sum_i w_i^2 (\beta_i - \hat{\beta})]^2}{\sum_i w_i} + \sum_i w_i^3 (\beta_i - \hat{\beta})^2 - I_E(\tau^2).$$

Finally,

$$\widehat{\text{Var}}(\hat{\beta}) = \hat{\Phi}_A = \frac{1}{\sum_i w_i} \left\{ 1 + \frac{2 \left(\frac{\sum_i w_i^3}{\sum_i w_i} - \left[\frac{\sum_i w_i^2}{\sum_i w_i} \right]^2 \right)}{I(\sigma_1)} \right\},$$

where $I(\sigma_1)$ may be either $I_E(\sigma_1)$ or $I_O(\sigma_1)$.

Kenward and Roger consider inferences of the form $\mathbf{L}\boldsymbol{\beta}$ for \mathbf{L} a fixed ($l \times p$) matrix, where l is the number of simultaneous inferences and p is the number of elements of $\boldsymbol{\beta}$. Here, $l = p = 1$ (since we simply wish to estimate β), hence $\mathbf{L} = \mathbf{1}$ also. Furthermore, since $l = 1$, the test distribution is t rather than F , and the scaling factor λ should also be equal to 1.

To calculate approximate degrees of freedom m , working through the formulae of Kenward and Roger^{19(p987)}:

$$\begin{aligned}
 A_1 &= \sum_i \sum_j W_{ij} \text{tr}(\Theta \Phi P_i \Phi) \text{tr}(\Theta \Phi P_j \Phi), \\
 A_2 &= \sum_i \sum_j W_{ij} \text{tr}(\Theta \Phi P_i \Phi \Theta \Phi P_j \Phi), \\
 \Theta &= L(L^\top \Phi L)^{-1} L^\top = \Phi^{-1}, \\
 \Rightarrow A_1 = A_2 &= W_{11} P_1^2 \hat{\Phi}_A^2 \\
 g &= \frac{(l+1)A_1 - (l+4)A_2}{(l+2)A_2} = \frac{2A_1 - 5A_1}{3A_1} = -1, \\
 \Rightarrow 3l + 2(1-g) = 7 &\Rightarrow c_1 = \frac{g}{3l+2(1-g)} = -\frac{1}{7}, \\
 c_2 &= \frac{l-g}{3l+2(1-g)} = \frac{2}{7}, \\
 \text{and } c_3 &= \frac{l+2-g}{3l+2(1-g)} = \frac{4}{7}.
 \end{aligned}$$

Next,

$$\begin{aligned}
 E[F] &= \left(1 - \frac{A_2}{l}\right)^{-1} = (1 - A_1)^{-1}, \\
 B &= \frac{1}{2l} (A_1 + 6A_2) = \frac{7}{2} A_1,
 \end{aligned}$$

and

$$\begin{aligned}
 V[F] &= \frac{2}{l} \left\{ \frac{1 + c_1 B}{(1 - c_2 B)^2 (1 - c_3 B)} \right\} = \frac{2 - A_1}{(1 - A_1)^2 (1 - 2A_1)} \\
 \Rightarrow \rho &= \frac{V[F]}{2E[F]^2} = \frac{1 - \frac{1}{2}A_1}{1 - 2A_1}, \\
 \Rightarrow m &= 4 + \frac{l+2}{l\rho - 1} = \frac{2}{A_1} = \frac{2}{W_{11} P_1^2 \hat{\Phi}_A^2}, \\
 \text{and } \lambda &= \frac{m}{E[F](m-2)} = 1.
 \end{aligned}$$

If W_{11} is based on the observed information matrix, the degrees of freedom are

$$m_O = \frac{2 \frac{[\sum_i w_i^2 (\beta_i - \hat{\beta})]^2}{\sum_i w_i} + 2 \sum_i w_i^3 (\beta_i - \hat{\beta})^2 - \sum_i w_i^2 + 2 \frac{\sum_i w_i^3}{\sum_i w_i} - \left[\frac{\sum_i w_i^2}{\sum_i w_i} \right]^2}{\hat{\Phi}_A^2 (\sum_i w_i^2)^2}$$

or if based on the expected information matrix:

$$m_E = \frac{\sum_i w_i^2 - 2 \frac{\sum_i w_i^3}{\sum_i w_i} + \left[\frac{\sum_i w_i^2}{\sum_i w_i} \right]^2}{\hat{\Phi}_A^2 (\sum_i w_i^2)^2}.$$