THE STATA JOURNAL

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

For more information on the *Stata Journal*, including information for authors, see the webpage

<div align="center">http://www.stata-journal.com</div>

The *Stata Journal* is indexed and abstracted in the following:

- CompuMath Citation Index®
- Current Contents/Social and Behavioral Sciences®
- RePEc: Research Papers in Economics
- Science Citation Index Expanded (also known as SciSearch®)
- Scopus™
- Social Sciences Citation Index®

metaan: Random-effects meta-analysis

Evangelos Kontopantelis
National Primary Care
Research & Development Centre
University of Manchester
Manchester, UK
e.kontopantelis@manchester.ac.uk

David Reeves
Health Sciences Primary Care
Research Group
University of Manchester
Manchester, UK
david.reeves@manchester.ac.uk

Abstract. This article describes the new meta-analysis command `metaan`, which can be used to perform fixed- or random-effects meta-analysis. Besides the standard DerSimonian and Laird approach, `metaan` offers a wide choice of available models: maximum likelihood, profile likelihood, restricted maximum likelihood, and a permutation model. The command reports a variety of heterogeneity measures, including Cochran's Q, I^2, H_M^2, and the between-studies variance estimate $\hat{\tau}^2$. A forest plot and a graph of the maximum likelihood function can also be generated.

Keywords: st0201, metaan, meta-analysis, random effect, effect size, maximum likelihood, profile likelihood, restricted maximum likelihood, REML, permutation model, forest plot

1 Introduction

Meta-analysis is a statistical methodology that integrates the results of several independent clinical trials in general that are considered by the analyst to be "combinable" (Huque 1988). Usually, this is a two-stage process: in the first stage, the appropriate summary statistic for each study is estimated; then in the second stage, these statistics are combined into a weighted average. Individual patient data (IPD) methods exist for combining and meta-analyzing data across studies at the individual patient level. An IPD analysis provides advantages such as standardization (of marker values, outcome definitions, etc.), follow-up information updating, detailed data-checking, subgroup analyses, and the ability to include participant-level covariates (Stewart 1995; Lambert et al. 2002). However, individual observations are rarely available; additionally, if the main interest is in mean effects, then the two-stage and the IPD approaches can provide equivalent results (Olkin and Sampson 1998).

This article concerns itself with the second stage of the two-stage approach to meta-analysis. At this stage, researchers can select between two main approaches—the fixed-effects (FE) or the random-effects model—in their efforts to combine the study-level summary estimates and calculate an overall average effect. The FE model is simpler and assumes the true effect to be the same (homogeneous) across studies. However, homogeneity has been found to be the exception rather than the rule, and some degree of true effect variability between studies is to be expected (Thompson and Pocock 1991). Two sorts of between-studies heterogeneity exist: clinical heterogeneity stems from dif-

ferences in populations, interventions, outcomes, or follow-up times, and methodological heterogeneity stems from differences in trial design and quality (Higgins and Green 2009; Thompson 1994). The most common approach to modeling the between-studies variance is the model proposed by DerSimonian and Laird (1986), which is widely used in generic and specialist meta-analysis statistical packages alike. In Stata, the DerSimonian–Laird (DL) model is used in the most popular meta-analysis commands—the recently updated `metan` and the older but still useful `meta` (Harris et al. 2008). However, the between-studies variance component can be estimated using more-advanced (and computationally expensive) iterative techniques: maximum likelihood (ML), profile likelihood (PL), and restricted maximum likelihood (REML) (Hardy and Thompson 1996; Thompson and Sharp 1999). Alternatively, the estimate can be obtained using nonparametric approaches, such as the permutations (PE) model proposed by Follmann and Proschan (1999).

We have implemented these models in `metaan`, which performs the second stage of a two-stage meta-analysis and offers alternatives to the DL random-effects model. The command requires the studies' effect estimates and standard errors as input. We have also created `metaeff`, a command that provides support in the first stage of the two-stage process and complements `metaan`. The `metaeff` command calculates for each study the effect size (standardized mean difference) and its standard error from the input parameters supplied by the user, using one of the models described in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins and Green 2006). For more details, type `ssc describe metaeff` in Stata or see Kontopantelis and Reeves (2009).

The `metaan` command does not offer the plethora of options `metan` does for inputting various types of binary or continuous data. Other useful features in `metan` (unavailable in `metaan`) include stratified meta-analysis, user-input study weights, vaccine efficacy calculations, the Mantel–Haenszel FE method, L'Abbe plots, and funnel plots. The REML model, assumed to be the best model for fitting a random-effects meta-analysis model even though this assumption has not been thoroughly investigated (Thompson and Sharp 1999), has recently been coded in the updated meta-regression command `metareg` (Harbord and Higgins 2008) and the new multivariate random-effects meta-analysis command `mvmeta` (White 2009). However, the output and options provided by `metaan` can be more useful in the univariate meta-analysis context.

2   The metaan command

2.1   Syntax

`metaan` *varname1 varname2* $\big[\mathit{if}\,\big]$ $\big[\mathit{in}\,\big]$, {`fe`|`dl`|`ml`|`reml`|`pl`|`pe`} $\big[$`varc`
   `label(`*varname*`)` `forest` `forestw(`#`)` `plplot(`*string*`)`$\big]$

where
    varname1 is the study effect size.
    varname2 is the study effect variation, with standard error used as the default.

2.2   Options

`fe` fits an FE model that assumes there is no heterogeneity between the studies. The model assumes that within-study variances may differ, but that there is homogeneity of effect size across studies. Often the homogeneity assumption is unlikely, and variation in the true effect across studies is to be expected. Therefore, caution is required when using this model. Reported heterogeneity measures are estimated using the `dl` option. You must specify one of `fe`, `dl`, `ml`, `reml`, `pl`, or `pe`.

`dl` fits a DL random-effects model, which is the most commonly used model. The model assumes heterogeneity between the studies; that is, it assumes that the true effect can be different for each study. The model assumes that the individual-study true effects are distributed with a variance τ^2 around an overall true effect, but the model makes no assumptions about the form of the distribution of either the within-study or the between-studies effects. Reported heterogeneity measures are estimated using the `dl` option. You must specify one of `fe`, `dl`, `ml`, `reml`, `pl`, or `pe`.

`ml` fits an ML random-effects model. This model makes the additional assumption (necessary to derive the log-likelihood function, and also true for `reml` and `pl`, below) that both the within-study and the between-studies effects have normal distributions. It solves the log-likelihood function iteratively to produce an estimate of the between-studies variance. However, the model does not always converge; in some cases, the between-studies variance estimate is negative and set to zero, in which case the model is reduced to an `fe` specification. Estimates are reported as missing in the event of nonconvergence. Reported heterogeneity measures are estimated using the `ml` option. You must specify one of `fe`, `dl`, `ml`, `reml`, `pl`, or `pe`.

`reml` fits an REML random-effects model. This model is similar to `ml` and uses the same assumptions. The log-likelihood function is maximized iteratively to provide estimates, as in `ml`. However, under `reml`, only the part of the likelihood function that is location invariant is maximized (that is, maximizing the portion of the likelihood that does not involve μ if estimating τ^2, and vice versa). The model does not always converge; in some cases, the between-studies variance estimate is negative and set to zero, in which case the model is reduced to an `fe` specification. Estimates are reported as missing in the event of nonconvergence. Reported heterogeneity measures are estimated using the `reml` option. You must specify one of `fe`, `dl`, `ml`, `reml`, `pl`, or `pe`.

`pl` fits a PL random-effects model. This model uses the same likelihood function as `ml` but takes into account the uncertainty associated with the between-studies variance estimate when calculating an overall effect, which is done by using nested iterations to converge to a maximum. The confidence intervals (CIs) provided by the model are asymmetric, and hence so is the diamond in the forest plot. However, the model

does not always converge. Values that were not computed are reported as missing. Reported heterogeneity measures are estimated using the `ml` option because $\widehat{\mu}$ and $\widehat{\tau}^2$, the effect and between-studies variance estimates, are the same. Only their CIs are reestimated. The model also provides a CI for the between-studies variance estimate. You must specify one of `fe`, `dl`, `ml`, `reml`, `pl`, or `pe`.

`pe` fits a PE random-effects model. This model can be described in three steps. First, in line with a null hypothesis that all true study effects are zero and observed effects are due to random variation, a dataset of all possible combinations of observed study outcomes is created by permuting the sign of each observed effect. Then, the `dl` model is used to compute an overall effect for each combination. Finally, the resulting distribution of overall effect sizes is used to derive a CI for the observed overall effect. The CI provided by the model is asymmetric, and hence so is the diamond in the forest plot. Reported heterogeneity measures are estimated using the `dl` option. You must specify one of `fe`, `dl`, `ml`, `reml`, `pl`, or `pe`.

`varc` specifies that the study-effect variation variable, *varname2*, holds variance values. If this option is omitted, `metaan` assumes that the variable contains standard-error values (the default).

`label(`*varname*`)` selects labels for the studies. One or two variables can be selected and converted to strings. If two variables are selected, they will be separated by a comma. Usually, the author names and the year of study are selected as labels. The final string is truncated to 20 characters.

`forest` requests a forest plot. The weights from the specified analysis are used for plotting symbol sizes (`pe` uses `dl` weights). Only one graph output is allowed in each execution.

`forestw(`#`)` requests a forest plot with adjusted weight ratios for better display. The value can be in the $[1, 50]$ range. For example, if the largest to smallest weight ratio is 60 and the graph looks awkward, the user can use this command to improve the appearance by requesting that the weight be rescaled to a largest/smallest weight ratio of 30. Only the weight squares in the plot are affected, not the model. The CIs in the plot are unaffected. Only one graph output is allowed in each execution.

`plplot(`*string*`)` requests a plot of the likelihood function for the average effect or between-studies variance estimate of the `ml`, `pl`, or `reml` model. The `plplot(mu)` option fixes the average effect parameter to its model estimate in the likelihood function and creates a two-way plot of τ^2 versus the likelihood function. The `plplot(tsq)` option fixes the between-studies variance to its model estimate in the likelihood function and creates a two-way plot of μ versus the likelihood function. Only one graph output is allowed in each execution.

2.3   Saved results

`metaan` saves the following in `r()` (some varying by selected model):

Scalars
`r(Hsq)`	heterogeneity measure H_M^2
`r(Q)`	Cochran's Q value
`r(df)`	degrees of freedom
`r(effvar)`	effect variance
`r(efflo)`	effect size, lower 95% CI
`r(Isq)`	heterogeneity measure I^2
`r(Qpval)`	p-value for Cochran's Q
`r(eff)`	effect size
`r(effup)`	effect size, upper 95% CI

In addition to the standard results, `metaan, fe` and `metaan, dl` save the following in `r()`:

Scalars
`r(tausq_dl)`	$\hat{\tau}^2$, from the DL model

In addition to the standard results, `metaan, ml` saves the following in `r()`:

Scalars
`r(tausq_dl)`	$\hat{\tau}^2$, from the DL model
`r(conv_ml)`	ML convergence information
`r(tausq_ml)`	$\hat{\tau}^2$, from the ML model

In addition to the standard results, `metaan, reml` saves the following in `r()`:

Scalars
`r(tausq_dl)`	$\hat{\tau}^2$, from the DL model
`r(conv_reml)`	REML convergence information
`r(tausq_reml)`	$\hat{\tau}^2$, from the REML model

In addition to the standard results, `metaan, pl` saves the following in `r()`:

Scalars
`r(tausq_dl)`	$\hat{\tau}^2$, from the DL model
`r(tausqlo_pl)`	$\hat{\tau}^2$ (PL), lower 95% CI
`r(cloeff_pl)`	convergence information, PL effect size (lower CI)
`r(ctausqlo_pl)`	convergence information, PL $\hat{\tau}^2$ (lower CI)
`r(conv_ml)`	ML convergence information
`r(tausq_pl)`	$\hat{\tau}^2$, from the PL model
`r(tausqup_pl)`	$\hat{\tau}^2$ (PL), upper 95% CI
`r(cupeff_pl)`	convergence information, PL effect size (upper CI)
`r(ctausqup_pl)`	convergence information, PL $\hat{\tau}^2$ (upper CI)

In addition to the standard results, `metaan, pe` saves the following in `r()`:

Scalars
`r(tausq_dl)`	$\hat{\tau}^2$, from the DL model
`r(exec_pe)`	information on PE execution

In each case, heterogeneity measures H_M^2 and I^2 are computed using the returned between-variances estimate $\widehat{\tau}^2$. Convergence and PE execution information is returned as 1 if successful and as 0 otherwise. `r(effvar)` cannot be computed for PE. `r(effvar)` is the same for ML and PL, but for PL the CIs are "amended" to take into account the $\widehat{\tau}^2$ uncertainty.

3   Methods

The `metaan` command offers six meta-analysis models for calculating a mean effect estimate and its CIs: FE model, random-effects DL method, ML random-effects model, REML random-effects model, PL random-effects model, and PE method using a DL random-effects model. Models of the random-effects family take into account the identified between-studies variation, estimate it, and usually produce wider CIs for the overall effect than would an FE analysis. Brief descriptions of the models have been provided in section 2.2. In this section, we will provide a few more details and practical advice in selecting among the models. Their complexity prohibits complete descriptions in this article, and users wishing to look into model details are encouraged to refer to the original articles that described them (DerSimonian and Laird 1986; Hardy and Thompson 1996; Follmann and Proschan 1999; Brockwell and Gordon 2001).

The three ML models are iterative and usually computationally expensive. ML and PL derive the μ (overall effect) and τ^2 estimates by maximizing the log-likelihood function in (1) under different conditions. REML estimates τ^2 and μ by maximizing the restricted log-likelihood function in (2).

$$\log L(\mu, \tau^2) = -\frac{1}{2} \left[\sum_{i=1}^{k} \log \left\{ 2\pi \left(\widehat{\sigma}_i^2 + \tau^2 \right) \right\} + \sum_{i=1}^{k} \frac{(\widehat{y}_i - \mu)^2}{\widehat{\sigma}_i^2 + \tau^2} \right], \quad \mu \in \Re \ \& \ \tau^2 \geq 0 \quad (1)$$

$$\log L'(\mu, \tau^2) = -\frac{1}{2} \left[\sum_{i=1}^{k} \log \left\{ 2\pi \left(\widehat{\sigma}_i^2 + \tau^2 \right) \right\} + \sum_{i=1}^{k} \frac{(\widehat{y}_i - \widehat{\mu})^2}{\widehat{\sigma}_i^2 + \tau^2} \right]$$
$$- \frac{1}{2} \log \sum_{i=1}^{k} \frac{1}{\widehat{\sigma}_i^2 + \tau^2}, \quad \widehat{\mu} \in \Re \ \& \ \tau^2 \geq 0 \quad (2)$$

where k is the number of studies to be meta-analyzed, $\widehat{y}_i$ and $\widehat{\sigma}_i^2$ are the effect and variance estimates for study i, and $\widehat{\mu}$ is the overall effect estimate.

ML follows the simplest approach, maximizing (1) in a single iteration loop. A criticism of ML is that it takes no account of the loss in degrees of freedom that results from estimating the overall effect. REML derives the likelihood function in a way that adjusts for this and removes downward bias in the between-studies variance estimator. A useful description for REML, in the meta-analysis context, has been provided by Normand (1999). PL uses the same likelihood function as ML, but uses nested iterations to take into account the uncertainty associated with the between-studies variance estimate when calculating an overall effect. By incorporating this extra factor of uncertainty, PL yields

CIs that are usually wider than for DL and also are asymmetric. PL has been shown to outperform DL in various scenarios (Brockwell and Gordon 2001).

The PE model (Follmann and Proschan 1999) can be described as follows: First, in line with a null hypothesis that all true study effects are zero and observed effects are due to random variation, a dataset of all possible combinations of observed study outcomes is created by permuting the sign of each observed effect. Next the `dl` model is used to compute an overall effect for each combination. Finally, the resulting distribution of overall effect sizes is used to derive a CI for the observed overall effect.

Method performance is known to be affected by three factors: the number of studies in the meta-analysis, the degree of heterogeneity in true effects, and—provided there is heterogeneity present—the distribution of the true effects (Brockwell and Gordon 2001). Heterogeneity, which is attributed to clinical or methodological diversity (Higgins and Green 2006), is a major problem researchers have to face when combining study results in a meta-analysis. The variability that arises from different interventions, populations, outcomes, or follow-up times is described by clinical heterogeneity, while differences in trial design and quality are accounted for by methodological heterogeneity (Thompson 1994). Traditionally, heterogeneity is tested with Cochran's Q, which provides a p-value for the test of homogeneity, when compared with a χ^2_{k-1} distribution where k is the number of studies (Brockwell and Gordon 2001). However, the test is known to be poor at detecting heterogeneity because its power is low when the number of studies is small (Hardy and Thompson 1998). An alternative measure is I^2, which is thought to be more informative in assessing inconsistency between studies. I^2 values of 25%, 50%, and 75% correspond to low, moderate, and high heterogeneity, respectively (Higgins et al. 2003). Another measure is H^2_M, the measure least affected by the value of k. It takes values in the $[0, +\infty)$ range, with 0 indicating perfect homogeneity (Mittlböck and Heinzl 2006). Obviously, the between-studies variance estimate $\widehat{\tau}^2$ can also be informative about the presence or absence of heterogeneity.

The test for heterogeneity is often used as the basis for applying an FE or a random-effects model. However, the often low power of the Q test makes it unwise to base a decision on the result of the test alone. Research studies, even on the same topic, can vary on a large number of factors; hence, homogeneity is often an unlikely assumption and some degree of variability between studies is to be expected (Thompson and Pocock 1991). Some authors recommend the adoption of a random-effects model unless there are compelling reasons for doing otherwise, irrespective of the outcome of the test for heterogeneity (Brockwell and Gordon 2001).

However, even though random-effects methods model heterogeneity, the performance of the ML models (ML, REML, and PL) in situations where the true effects violate the assumptions of a normal distribution may not be optimal (Brockwell and Gordon 2001; Hardy and Thompson 1998; Böhning et al. 2002; Sidik and Jonkman 2007). The number of studies in the analysis is also an issue, because most meta-analysis models (including DL, ML, REML, and PL—but not PE) are only asymptotically correct; that is, they provide the theoretical 95% coverage only as the number of studies increases (approaches infinity). Method performance is therefore affected when the number of studies is small,

but the extent depends on the model (some are more susceptible), along with the degree of heterogeneity and the distribution of the true effects (Brockwell and Gordon 2001).

4 Example

As an example, we apply the `metaan` command to health-risk outcome data from seven studies. The information was collected for an unpublished meta-analysis, and the data are available from the authors. Using the `describe` and `list` commands, we provide details of the dataset and proceed to perform a univariate meta-analysis with `metaan`.

```
. use metaan_example

. describe
Contains data from metaan_example.dta
  obs:             7
 vars:             4                          19 Apr 2010 12:19
 size:           560 (99.9% of memory free)
─────────────────────────────────────────────────────────────────────────────
              storage   display    value
variable name   type    format     label     variable label
─────────────────────────────────────────────────────────────────────────────
study           str16   %16s                 First author and year
outcome         str48   %35s                 Outcome description
effsize         float   %9.0g                effect sizes
se              float   %9.0g                SE of the effect sizes
─────────────────────────────────────────────────────────────────────────────
Sorted by:   study  outcome

. list study outcome effsize se, noobs clean

              study                        outcome     effsize          se
       Bakx A, 1985   Serum cholesterol (mmol/L)    -.3041526    .0958199
   Campbell A, 1998                         Diet     .2124063    .0812414
      Cupples, 1994                          BMI     .0444239     .090661
         Eckerlund                           SBP    -.3991309      .12079
        Moher, 2001        Cholesterol (mmol/l)    -.9374746    .0691572
    Woolard A, 1995      Alcohol intake (g/week)    -.3098185     .206331
    Woolard B, 1995      Alcohol intake (g/week)    -.4898825    .2001602
```

```
. metaan effsize se, pl label(study) forest

Profile Likelihood method selected
```

Study	Effect	[95% Conf. Interval]		% Weight
Bakx A, 1985	−0.304	−0.492	−0.116	15.09
Campbell A, 1998	0.212	0.053	0.372	15.40
Cupples, 1994	0.044	−0.133	0.222	15.20
Eckerlund	−0.399	−0.636	−0.162	14.49
Moher, 2001	−0.937	−1.073	−0.802	15.62
Woolard A, 1995	−0.310	−0.714	0.095	12.01
Woolard B, 1995	−0.490	−0.882	−0.098	12.19
Overall effect (pl)	−0.308	−0.622	0.004	100.00

```
ML method succesfully converged
PL method succesfully converged for both upper and lower CI limits

Heterogeneity Measures
```

	value	df	p-value
Cochrane Q	139.81	6	0.000
I^2 (%)	91.96		
H^2	11.44		

	value	[95% Conf. Interval]	
tau^2 est	0.121	0.000	0.449

```
Estimate obtained with Maximum likelihood - Profile likelihood provides the CI
PL method succesfully converged for both upper and lower CI limits of the tau^2
> estimate
```

The PL model used in the example converged successfully, as did ML, whose convergence is a prerequisite. The overall effect is not found to be significant at the 95% level, and there is considerable heterogeneity across studies, according to the measures. The model also displays a 95% CI for the between-studies variance estimate $\hat{\tau}^2$ (provided that convergence is achieved, as is the case in this example). The forest plot created by the command is displayed in figure 1.
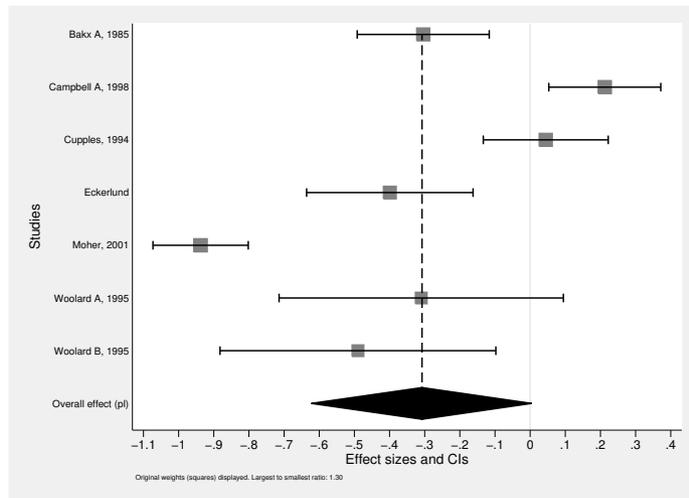
(Continued on next page)

Figure 1. Forest plot displaying PL meta-analysis

When we reexecute the analysis with the `plplot(mu)` and `plplot(tsq)` options, we obtain the log-likelihood function plots shown in figures 2 and 3.
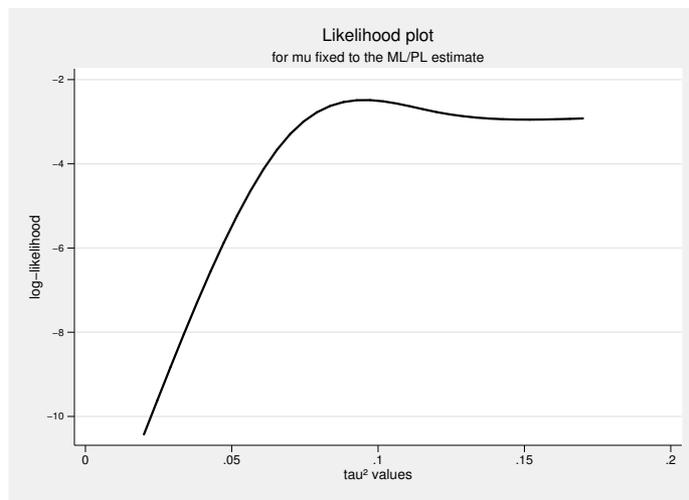

Figure 2. Log-likelihood function plot for μ fixed to the model estimate
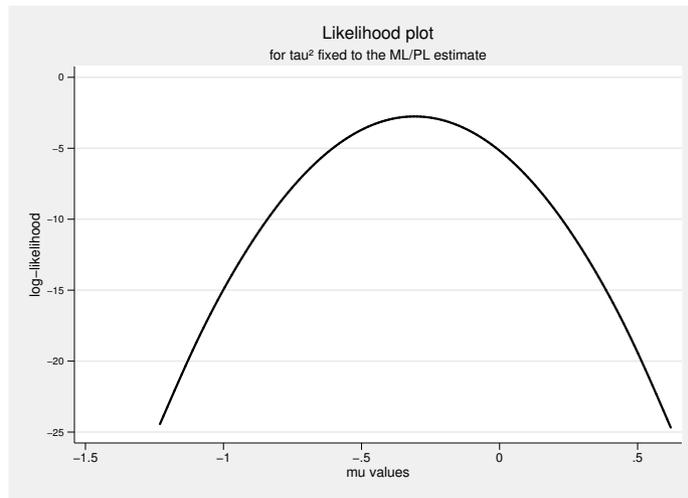
Figure 3. Log-likelihood function plot for τ^2 fixed to the model estimate

5 Discussion

The `metaan` command can be a useful meta-analysis tool that includes newer and, in certain circumstances, better-performing models than the standard DL random-effects model. Unpublished results exploring model performance in various scenarios are available from the authors. Future work will involve implementing more models in the `metaan` command and embellishing the forest plot.

6 Acknowledgments

We would like to thank the authors of `meta` and `metan` for all their work and the anonymous reviewer whose useful comments improved the article considerably.

7 References

Böhning, D., U. Malzahn, E. Dietz, P. Schlattmann, C. Viwatwongkasem, and A. Biggeri. 2002. Some general points in estimating heterogeneity variance with the DerSimonian–Laird estimator. *Biostatistics* 3: 445–457.

Brockwell, S. E., and I. R. Gordon. 2001. A comparison of statistical methods for meta-analysis. *Statistics in Medicine* 20: 825–840.

DerSimonian, R., and N. Laird. 1986. Meta-analysis in clinical trials. *Controlled Clinical Trials* 7: 177–188.

Follmann, D. A., and M. A. Proschan. 1999. Valid inference in random effects meta-analysis. *Biometrics* 55: 732–737.

Harbord, R. M., and J. P. T. Higgins. 2008. Meta-regression in Stata. *Stata Journal* 8: 493–519.

Hardy, R. J., and S. G. Thompson. 1996. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* 15: 619–629.

———. 1998. Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine* 17: 841–856.

Harris, R. J., M. J. Bradburn, J. J. Deeks, R. M. Harbord, D. G. Altman, and J. A. C. Sterne. 2008. metan: Fixed- and random-effects meta-analysis. *Stata Journal* 8: 3–28.

Higgins, J. P. T., and S. Green. 2006. *Cochrane Handbook for Systematic Reviews of Interventions Version 4.2.6.*
http://www2.cochrane.org/resources/handbook/Handbook4.2.6Sep2006.pdf.

———. 2009. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2.* http://www.cochrane-handbook.org/.

Higgins, J. P. T., S. G. Thompson, J. J. Deeks, and D. G. Altman. 2003. Measuring inconsistency in meta-analyses. *British Medical Journal* 327: 557–560.

Huque, M. F. 1988. Experiences with meta-analysis in NDA submissions. *Proceedings of the Biopharmaceutical Section of the American Statistical Association* 2: 28–33.

Kontopantelis, E., and D. Reeves. 2009. MetaEasy: A meta-analysis add-in for Microsoft Excel. *Journal of Statistical Software* 30: 1–25.

Lambert, P. C., A. J. Sutton, K. R. Abrams, and D. R. Jones. 2002. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology* 55: 86–94.

Mittlböck, M., and H. Heinzl. 2006. A simulation study comparing properties of heterogeneity measures in meta-analyses. *Statistics in Medicine* 25: 4321–4333.

Normand, S.-L. T. 1999. Meta-analysis: Formulating, evaluating, combining, and reporting. *Statistics in Medicine* 18: 321–359.

Olkin, I., and A. Sampson. 1998. Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics* 54: 317–322.

Sidik, K., and J. N. Jonkman. 2007. A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine* 26: 1964–1981.

Stewart, L. A. 1995. Practical methodology of meta-analyses (overviews) using updated individual patient data. *Statistics in Medicine* 14: 2057–2079.

Thompson, S. G. 1994. Systematic review: Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal* 309: 1351–1355.

Thompson, S. G., and S. J. Pocock. 1991. Can meta-analyses be trusted? *Lancet* 338: 1127–1130.

Thompson, S. G., and S. J. Sharp. 1999. Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine* 18: 2693–2708.

White, I. R. 2009. Multivariate random-effects meta-analysis. *Stata Journal* 9: 40–56.

About the authors

Evangelos (Evan) Kontopantelis is a research fellow in statistics at the National Primary Care Research and Development Centre, University of Manchester, England. His research interests include statistical methods in health sciences with a focus on meta-analysis, longitudinal data modeling, and large clinical database management.

David Reeves is a senior research fellow in statistics at the Health Sciences Primary Care Research Group, University of Manchester, England. David has worked as a statistician in health services research for nearly three decades, mainly in the fields of learning disability and primary care. His methodological research interests include the robustness of statistical methods, the analysis of observational studies, and applications of social network analysis methods to health systems.