

# metaplus: An R Package for the Analysis of Robust Meta-Analysis and Meta-Regression

by Ken J. Beath

**Abstract** The `metaplus` package is described with examples of its use for fitting meta-analysis and meta-regression. For either meta-analysis or meta-regression it is possible to fit one of three models: standard normal random effect,  $t$ -distribution random effect or mixture of normal random effects. The latter two models allow for robustness by allowing for a random effect distribution with heavier tails than the normal distribution, and for both robust models the presence of outliers may be tested using the parametric bootstrap. For the mixture of normal random effects model the outlier studies may be identified through their posterior probability of membership in the outlier component of the mixture. Plots allow the results of the different models to be compared. The package is demonstrated on three examples: a meta-analysis with no outliers, a meta-analysis with an outlier and a meta-regression with an outlier.

## Introduction

Meta-analysis is a method of combining the results of different studies to produce one overall result (Sutton et al., 2000). Meta-regression is an extension to meta-analysis which allows study-specific effect sizes to change depending on study-specific covariates. For example there may be studies comparing a drug to placebo, with varying doses of the drug used in the different studies. It is possible that the effectiveness of the drug will vary with dose, in a linear or nonlinear relationship, and by including this in the model the unexplained variation is reduced.

One of the difficulties in combining studies is that the differences between studies may be greater than would be indicated by the variation within each study. This is allowed for by the random effect model where the effect for each study has two components: an overall effect and a random component specific to each study, with the random component traditionally assumed to have a normal distribution. The model without a random effect is known as the fixed effect model, which is equivalent to a random effect model with zero variance for the random effect.

One difficulty is that the assumption of a normally distributed random effect may be unrealistic, with a particular violation that the tails are heavier than would be expected. While it has been shown that results are robust to moderate violations of the normality assumption (Kontopantelis and Reeves, 2012), this does not apply to more extreme cases. One solution to this is to use an alternative to the normal distribution for the random effect, for example the  $t$ -distribution, as described in Lee and Thompson (2008) and Baker and Jackson (2008), the Laplace distribution (Demidenko, 2013, Section 5.1.5), a non-parametric (Branscum and Hanson, 2008) or a semi-parametric (Burr and Doss, 2005) random effect distribution. This, however, does not identify which studies are unusual. A traditional method of identifying outliers is through residual diagnostics and this has been applied to meta-analysis by Viechtbauer and Cheung (2010). However, the effect of the outliers on the fitted model may cause them to be masked (Atkinson, 1986). This occurs when the outliers affect the fitted model to the extent that the unusual observations no longer appear unusual. A method to avoid this is deletion of residuals, used in Viechtbauer and Cheung (2010), but this is only effective for single outliers. It can be extended to allow multiple outliers but with the need to fit a large number of models. A method to avoid the problem of multiple outliers is described by Gumedze and Jackson (2011). They assume that studies are either normal or are outlier studies from a random effect distribution with a higher variance. Only one study is assumed to be an outlier, with each study tested in turn, but multiple outliers then allowed for using order statistics. Beath (2014) noted the similarity of this model to a mixture model, which also allows for a more general fitting algorithm and a statistical test for the presence of outliers and indication of which studies are outliers.

The purpose of the `metaplus` package (Beath et al., 2016) is to fit the two robust models with random effects based on the  $t$ -distribution and the mixture of normals, as well as the standard normal random effects model. It is not designed to replace a more general meta-analysis package, such as `metafor` (Viechtbauer, 2010) but to provide additional specialised analyses. In producing forest plots, it builds upon the functionality of the `metafor` package, allowing the various models to be compared.

## Models

The random effect meta-analysis model assumes that the observed treatment effect  $Y_i$  for study  $i$  is

$$Y_i = \mu + E_i + \epsilon_i,$$

where  $\mu$  is the overall mean for the studies,  $E_i$  is a random effect with mean zero, and  $\epsilon_i$  is a normally distributed error with variance  $\sigma_i^2$  for study  $i$ , where the within study variance  $\sigma_i^2$  is assumed to be known.

An extension, known as meta-regression, to the random effect meta-analysis model is to include covariates to explain the heterogeneity (Sutton et al., 2000, p. 51). Incorporating this into the meta-analysis model we obtain

$$Y_i = \mu + X_i^T \beta + E_i + \epsilon_i,$$

where  $X_i$  is a vector of covariate values for study  $i$ , and  $\beta$  is a vector of the corresponding parameters.

In **metaplus** there are three available random effect distributions:

**Normal:** The probability density function for study  $i$  is

$$f(Y_i|X_i; \mu, \tau) = \frac{1}{\sqrt{2\pi(\sigma_i^2 + \tau^2)}} \exp\left(-\frac{(Y_i - \mu - X_i^T \beta)^2}{2(\sigma_i^2 + \tau^2)}\right).$$

**Robust  $t$ -distribution:** This distribution was introduced as one of a number of distributions for robust meta-analysis by Lee and Thompson (2008) and Baker and Jackson (2008). This approach replaces the normal random effect distribution with a  $t$ -distribution. The degrees of freedom ( $\nu$ ) of the  $t$ -distribution control the heaviness of the tails, and are estimated from the data, using  $\nu^{-1}$  as the parameter for numerical advantages. The probability density function no longer has a closed-form expression, requiring integration over the  $t$ -distribution random effect as

$$f(Y_i|X_i; \mu, \tau, \nu) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{(Y_i - \mu - X_i^T \beta - \eta)^2}{2\sigma_i^2}\right) g(\eta|\tau, \nu) d\eta,$$

where  $g(\eta|\tau, \nu)$  is the density function of a scaled  $t$ -distribution with  $\nu$  degrees of freedom

$$g(\eta|\tau, \nu) = \frac{\Gamma((\nu+1)/2)}{\tau\sqrt{\pi\nu}\Gamma(\nu/2)} \left(1 + \frac{\eta^2}{\nu\tau^2}\right)^{-((\nu+1)/2)}.$$

**Robust mixture:** This assumes that a study can belong to one of two classes, where each class is a standard random effect model with the same mean but different random effect variance, which is higher for the outlier class (Beath, 2014). The robust meta-regression model takes the form

$$Y_{i|k} = \mu + X_i^T \beta + E_{i|k} + \epsilon_i,$$

where  $\epsilon_i$  is as for the standard model, but  $E_{i|k}$  is now a random effect dependent on the class, where  $k = 1, 2$  indexes the classes, with  $k = 1$  corresponding to standard studies and  $k = 2$  to outlier studies, with random effect variances  $\tau_1^2, \tau_2^2$  respectively, with the restriction that  $\tau_2^2 > \tau_1^2$ , and again zero mean. The probability density function becomes the weighted sum of the probability density function for each class, with weights equal to the proportion of studies in each class  $\pi_1, \pi_2$  for the standard and outlier studies, respectively:

$$f(Y_i|X_i; \mu, \tau_1, \tau_2, \pi_1, \pi_2) = \sum_{k=1}^2 \pi_k \frac{1}{\sqrt{2\pi}} \left(\frac{1}{\sigma_i^2 + \tau_k^2}\right)^{1/2} \exp\left(-\frac{1}{2} \frac{(Y_i - \mu - X_i^T \beta)^2}{\sigma_i^2 + \tau_k^2}\right)$$

with the constraints that  $\pi_1 + \pi_2 = 1$  and  $0 \leq \pi_i \leq 1$ .

## Profile likelihood based confidence intervals

A difficulty with the use of standard maximum likelihood techniques for random effect models is that they produce biased estimates for the variance of the random effect, which results in biased estimates of the standard errors for the parameters of interest, and therefore poor coverage using Wald-type confidence intervals. The solution for meta-analysis has been the use of Wald-type confidence intervals

obtained from models fitted using residual maximum likelihood (REML), but this is difficult for the robust models. However, profile likelihood based confidence intervals (Pawitan, 2001, p. 61) have been found to be superior (Hardy and Thompson, 1996), and these are used for all fitted models. The profile likelihood based confidence intervals are obtained from routines based on the `mle2` function in the package `bbmle` (Bolker and R Core Team, 2014) which provides an extended version of `mle`. The  $p$ -values are calculated using the likelihood ratio test statistic so that they are consistent with the confidence intervals.

### Parametric bootstrap

Testing for the need for the robust distributions requires a test of  $\nu = \infty$ , or equivalently  $\nu^{-1} = 0$  for the  $t$ -distribution and  $\pi_2 = 0$  for the robust mixture. Both tests involve a test of a parameter on the boundary of the parameter space, so the usual asymptotic theory cannot be used. One solution is the parametric bootstrap (McLachlan, 1987), which involves simulating data sets under the null hypothesis and calculating the likelihood ratio test statistic for each simulated data set. The observed test statistic is then compared to the simulated test statistics to determine the  $p$ -value.

### Other computational details

For both robust models the starting values are important, as the optimisation used to obtain the maximum likelihood may converge to a local minimum. For the  $t$ -distribution a standard normal random effect model is first fitted. The parameter estimates from this model together with a range of values of the  $t$ -distribution degrees of freedom are used as starting parameter values for the  $t$ -distribution random effect model. From these fitted models the model with the maximum likelihood is chosen as the final fitted model.

For the  $t$ -distribution random effect model numerical integration is used to obtain the marginal likelihood, with a choice of either adaptive quadrature or adaptive Gauss-Hermite quadrature. In general adaptive quadrature was found to be superior; however it was required to use adaptive Gauss-Hermite quadrature when the standard errors of studies are unusually small. Another difficulty is that the model is not identifiable when  $\tau^2 = 0$  as the likelihood is no longer dependent on the  $t$ -distribution degrees of freedom, and this causes difficulties with the optimisation. To avoid this a model was fitted with  $\nu^{-1} = 0$ , to allow  $\tau^2 = 0$ , and the likelihood from this model used if it was equal or larger than given by the optimisation with  $\nu$  unconstrained.

For the robust mixture model a generalized EM (GEM) algorithm is used. The usual method for generating starting values for a mixture model using the EM algorithm, as described in McLachlan and Peel (2000, p. 55), is to randomly allocate subjects to each group in the initial E step. This is repeated for a number of random allocations and the resulting model fit with the highest maximum likelihood used as the fitted model. For the outlier models this usually requires a large number of random allocations, and therefore model fits, due to the small number in the outlier class.

The method used in **metaplus** is to systematically generate the initial outliers in the E step with an increasing number of initial outliers, starting with no outliers. For a given number of outliers in the selected initial set all possible initial sets are fitted with the restriction that each set of initial outliers builds on the best set of initial outliers found for the previous number of outliers. For example if, when considering single initial outliers, study 10 as the initial outlier produces the highest maximum likelihood then study 10 would be included in all pairs of studies when considering models with two initial outliers. When the maximum likelihood does not increase the process is stopped.

### Using package **metaplus**

The main function available in **metaplus** is `metaplus`, with associated methods `outlierProbs` and `testOutliers` specific to **metaplus**, with the arguments for each shown in Table 1. The function `metaplus` fits a meta-analysis model to the studies, with results extracted using `summary`, and plotted using `plot`. The `plot` method makes use of the forest method in **metafor** allowing the same customisations of the plots. An additional argument specific to `plot` in **metaplus** is `extrameta`, which allows for extra meta-analysis results to be plotted. This allows for different models (i.e. standard and robust) to be compared, or for meta-regression to show the overall effect at different values of the covariates. An alternative method of plotting is to use `forestplot` (Gordon and Lumley, 2015) which allows some other customisations, but will require combining the data from the studies and summaries. The method `testOutliers` tests for the presence of outliers for the robust models using the parametric bootstrap. The method `outlierProbs` determines the posterior probability of each

metaplus() arguments	
yi	Vector of observed effect sizes corresponding to each study.
sei	Vector of observed standard errors corresponding to each study.
mods	Data frame of covariates corresponding to each study (only required for a meta-regression model).
random	The type of random effect distribution. One of "normal", "t-dist", "mixture", for standard normal, <i>t</i> -distribution or mixture of normals, respectively.
label	The label to be used for this model when producing the summary line on the forest plot. This allows for identification of the model when comparing multiple models.
plotci	Should a diagnostic plot for the profile likelihood be made? See the package <b>bbmle</b> documentation for further details.
justfit	Should the model only be fitted? If only the model is fitted then profiling and likelihood ratio test statistics are not calculated. This is useful for bootstrapping to reduce computation time.
slab	Vector of character strings corresponding to each study. This is used only to label the plots.
useAGQ	Should adaptive Gauss-Hermite quadrature be used with the <i>t</i> -distribution random effect model. This may be used when there are numerical problems due to small standard errors.
quadpoints	Number of quadrature points for the adaptive Gauss-Hermite quadrature.
data	Optional data frame in which to search for other variables.
outlierProbs() arguments	
object	"metaplus" object.
testOutliers() arguments	
object	"metaplus" object.
R	Number of simulations used in the parametric bootstrap.

**Table 1:** Arguments for functions and methods of the **metaplus** package.

study being an outlier for the normal mixture model. The returned object has an associated plot method to plot the outlier probabilities. The returned results are shown in Table 2.

## Examples

In the following examples, both robust options are used to demonstrate the capabilities of the package. In practice it will be required to choose which model to use when determining the final result. This should be the better fitting model, which can be determined using either AIC or BIC. Where the outliers are extreme the *t*-distribution will fit poorly requiring the use of the mixture distribution. In other cases the *t*-distribution will be preferred as it uses one less parameter, also making it less likely to produce unstable results which will be shown in the confidence interval profile plot. Where there is little difference between the fits the mixture distribution may be preferred as it allows identification of the outlier studies.

### Intravenous magnesium in acute myocardial infarction

A number of studies have been performed to determine the effectiveness of intravenous magnesium in acute myocardial infarction, and a meta-analysis is performed in Sterne et al. (2001). The studies have caused considerable controversy, as the results of a single large study ISIS-4 (ISIS-4: Collaborative Group, 1995) contradicts the results of a meta-analysis. Higgins and Spiegelhalter (2002) discuss some of the history and some suggested methods from a Bayesian perspective, Woods (2002) comments on the variability between studies due to timing of infusion, and Downing (1999) on the higher level of dose used in ISIS-4, with a more recent meta-analysis by Li et al. (2009). Of interest is whether,

metapplus()	
results	Matrix containing columns for estimate, lower and upper 95% confidence interval and $p$ -value. If <code>justfit = TRUE</code> then only the parameter estimates are returned.
yi	Vector of observed effect sizes.
sei	Vector of observed standard errors corresponding to each effect size.
mods	Data frame of covariates corresponding to each study (only returned from a meta-regression model).
fittedmodel	Final model returned from <b>bbmle</b> .
justfit	Value of <code>justfit</code> passed to <code>metapplus</code> .
random	Type of random effect.
slab	Vector of character strings corresponding to each study. This is used to label the forest plot.
outlierProbs()	
outlier.prob	Vector of posterior probabilities that the study is an outlier corresponding to each study.
slab	Vector of labels for the studies.
testOutliers()	
pvalue	$p$ -value obtained from the parametric bootstrap.
observed	Observed value of the likelihood ratio test statistic.
sims	Vector of simulated values of the test statistic under the null hypothesis.

**Table 2:** Results reported by functions and methods of the **metapplus** package.

given the heterogeneity between studies, the ISIS-4 study is unusual. The data have been obtained in the form of log odds ratios for mortality where negative values correspond to treatment benefit, but if raw data in the form of number of events per number of patients is available, then these can be converted using, for example, the `escalc` function in the **metafor** package. The standard random effect meta-analysis can be performed, and the parameter estimates obtained as follows:

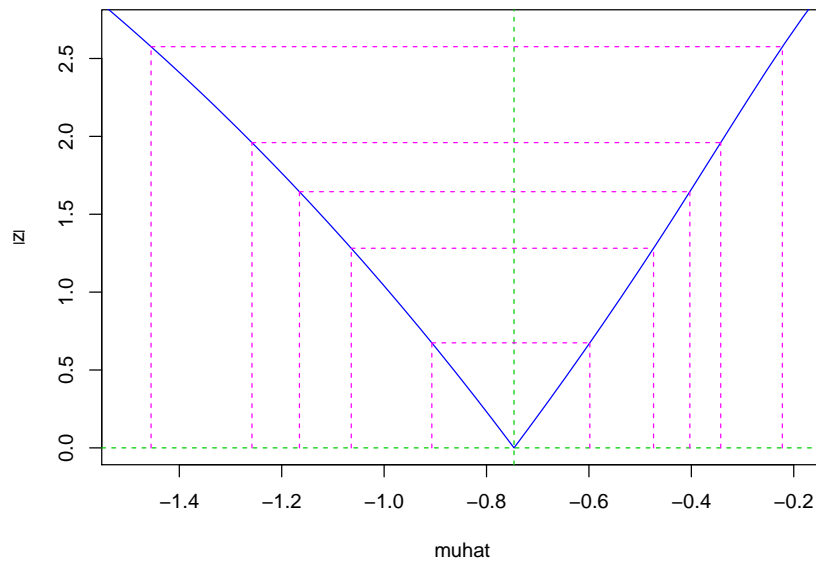
```
> mag.meta <- metapplus(yi, sei, slab = study, data = mag)
> summary(mag.meta)
```

```
      Est. 95% ci.lb 95% ci.ub  pvalue
muhat -0.7463   -1.2583   -0.3428 0.000501
tau2    0.2540
```

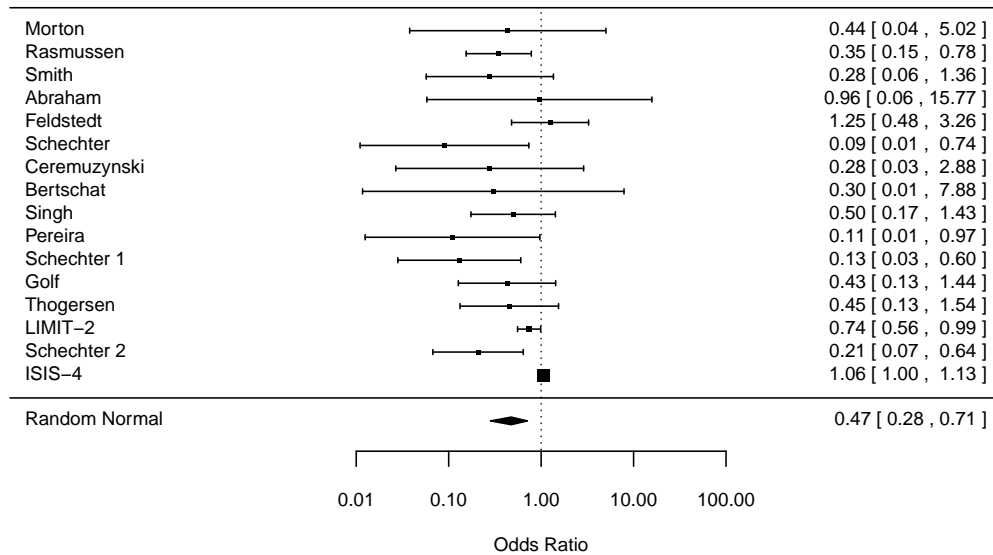
```
      logLik      AIC      BIC
-19.68459 43.36918 44.91436
```

Adding the argument `plotci = TRUE` will produce a plot giving details of the profile confidence intervals, as shown in Figure 1. The basis of the plot is that the profile log likelihood in the region of the maximum likelihood estimate should be asymptotically quadratic. As differences from a quadratic are difficult to determine by eye, a transformation is performed to the  $z$  scale, so that the curve should follow a straight line. Rather than plotting  $z$ ,  $|z|$  is plotted so that the curve should then be in the form of a symmetric “V” (Bolker and R Core Team, 2014). In this case, the shape is not symmetric, so this does not hold, although the difference is not large enough to be important. This is confirmed by the lack of symmetry of the confidence interval for `muhat`. An important variation from the “V” occurs when either half of the curve may not be monotonic, indicating that the profile likelihood is multi-modal and if this occurs in a region affecting the confidence interval then the calculated confidence interval may be incorrect. It may also be an indication that the model used is incorrect or that there is insufficient data for the fitted model.

The forest plot showing the studies and overall effect can be obtained using `plot(mag.meta)`. The **metapplus** package uses the forest plot capabilities of the **metafor** package which allows the arguments for the forest plot in **metafor** to be used when plotting. As the results for the magnesium studies are log odds ratios it is more useful to produce plots with units of odds ratios. This can be obtained by annotating the horizontal axis with odds ratios corresponding to the log odds, and requesting



**Figure 1:** Profile plot for intravenous magnesium in acute myocardial infarction using the normal random effect model.



**Figure 2:** Forest plot for magnesium studies for mortality using the normal random effect model.

an exponential transformation for the coefficients, as shown in the following code, and the plot is shown in Figure 2. The documentation for the `metafor` package should be investigated for further modifications. Under some systems the characters will not be properly spaced. This can be solved by using the `extrafont` (Chang, 2014) package and a fixed width font, for example 'Courier New'.

```
> plot(mag.meta, atransf = exp, at = log(c(.01, .1, 1, 10, 100)),
+ xlab = "Odds Ratio", cex = 0.75)
```

The meta-analysis is repeated using a *t*-distribution for the random effect by adding the `random = "t-dist"` argument. From the summary the estimate of `vinv`, the inverse degrees of freedom, is zero corresponding to infinite degrees of freedom, or a normal distribution. The BIC is also a guide, with an increase for the *t*-distribution model indicating that a standard normal is the correct model.

```
> mag.tdist <- metaplust(yi, sei, slab = study, random = "t-dist", data = mag)
> summary(mag.tdist)
```

```
Est. 95% ci.lb 95% ci.ub pvalue
muhat -0.7463 -1.2583 -0.3430 0.000501
tau2 0.2540
```

```
vinv 0.0000
```

```
      logLik      AIC      BIC
-19.68459 45.36918 47.68695
```

This can be confirmed with the `testOutliers` command, which performs a parametric bootstrap to obtain the null distribution of the likelihood ratio test statistic for the test that  $\nu^{-1} = 0$ , required as the test of the parameter is on the boundary of the parameter space. Note that this may take some time for the default of 999 simulations, of the order of one hour or longer depending on the number of studies, so initial investigation may be performed with a smaller number of simulations, with consequently lower accuracy.

```
> summary(testOutliers(mag.tdist))
```

```
Observed LRT statistic 0.0 p value 1
```

The analysis can be repeated using the robust mixture distribution for the random effect. The variance of both the random effect for standard studies (`tau2`) and for outlier studies (`tau2out`) are very close indicating that there are no outlier studies and this is confirmed by the outlier test.

```
> mag.mix <- metaplust(yi, sei, slab = study, random = "mixture", data = mag)
> summary(mag.mix)
```

```
      Est. 95% ci.lb 95% ci.ub  pvalue
muhat  -0.7463147 -1.2593989 -0.3427085 0.000777
tau2    0.2539981
tau2out 0.2540892
Outlier prob. 0.0001904
```

```
      logLik      AIC      BIC
-19.68459 47.36918 50.45954
```

```
> summary(testOutliers(mag.mix))
```

```
Observed LRT statistic 0.0 p value 1
```

### CDP choline for cognitive and behavioural disturbances

This meta-analysis evaluates the effect of CDP choline for cognitive and behavioural disturbances associated with chronic cerebral disorders in the elderly (Fioravanti and Yanagi, 2005) using standardised mean differences of memory measures as the outcome. A study (Bonavita 1983) was previously determined to be an outlier by Gumedze and Jackson (2011). A standard random effect meta-analysis will be fitted first, as previously.

```
> cdp.meta <- metaplust(yi, sei, slab = study, data = cdp)
> summary(cdp.meta)
```

```
      Est. 95% ci.lb 95% ci.ub  pvalue
muhat 0.38944 0.07269 0.76634 0.0218
tau2 0.14666
```

```
      logLik      AIC      BIC
-8.198544 20.39709 21.00226
```

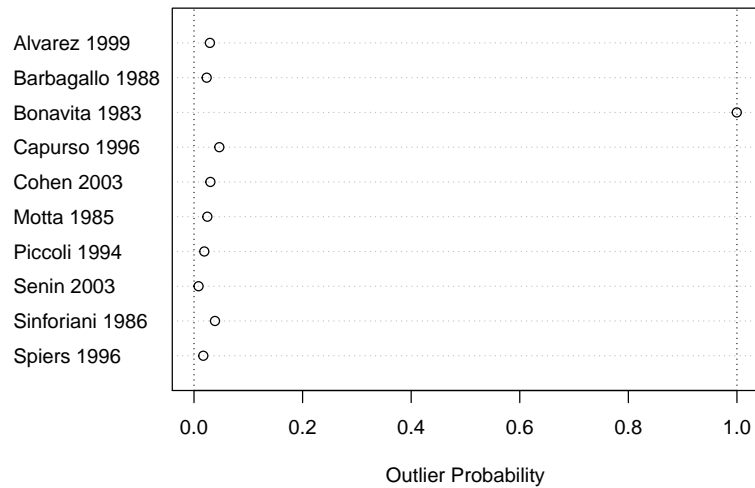
A robust model using the *t*-distribution is fitted with the following code.

```
> cdp.tdist <- metaplust(yi, sei, slab = study, random = "t-dist", data = cdp)
> summary(cdp.tdist)
```

```
      Est. 95% ci.lb 95% ci.ub  pvalue
muhat 1.946e-01 5.296e-02 3.610e-01 0.00899
tau2 4.478e-05
vinv 2.024e+00
```

```
      logLik      AIC      BIC
-4.057683 14.11537 15.02312
```

```
> summary(testOutliers(cdp.tdist))
```



**Figure 3:** Outlier probabilities for CDP studies from the robust mixture random effect model.

Observed LRT statistic 8.3 p value 0.001

As a rough guide, the decrease in AIC and BIC demonstrates that the model is an improvement, and this is confirmed with the outlier test. The fit is repeated using the robust mixture.

```
> cdp.mix <- metaplust(yi, sei, slab = study, random = "mixture", data = cdp)
> summary(cdp.mix)
```

	Est.	95% ci.lb	95% ci.ub	pvalue
muhat	0.1910	0.0563	0.3479	0.00711
tau2	0.0000			
tau2out	3.1558			
Outlier prob.	0.1237			

logLik	AIC	BIC
-3.007145	14.01429	15.22463

```
> summary(testOutliers(cdp.mix))
```

Observed LRT statistic 10.4 p value 0.001

The output from the robust mixture model has an interesting feature. For standard studies the estimated random effect variance is zero, indicating that only the outlier studies are contributing to the heterogeneity. The posterior probability of each study being an outlier can be obtained as:

```
> cdp.mix.outlierProbs <- outlierProbs(cdp.mix)
```

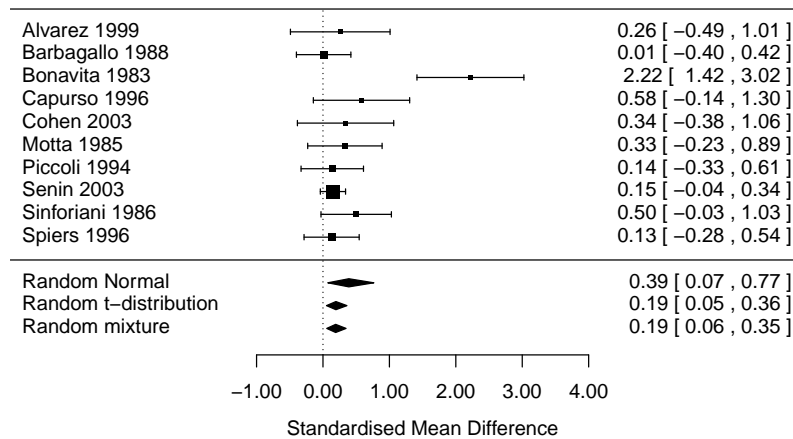
and plotted using `plot(cdp.mix.outlierProbs)` in Figure 3. This shows clearly that Bonavita 1983 has a posterior probability of nearly 1.0 of being an outlier. The other studies have a non-zero posterior probability of being outliers, as there is an overlap between the distribution of the standard and outlier studies, but are relatively close to zero.

Lastly, a forest plot with the results of all three models is generated, using the `extrameta` parameter to add the robust models, i.e. `plot(cdp.meta, extrameta = list(cdp.tdist, cdp.mix))`, and these are shown in Figure 4, where it can be noted that Bonavita 1983 has an unusually high value. The effect of the robust models is to down-weight the Bonavita 1983 study, which has the consequence of both reducing the overall effect estimate and its standard error.

### Exercise for depression

This example is a meta-analysis of trials of exercise in the management of depression (Lawlor and Hopker, 2001). Higgins and Thompson (2004) used the data as an example of meta-regression using a number of covariates, which will be limited here to a single covariate, the duration of trial. The outcome is effect size calculated using Cohen's method. First the meta-analysis using standard normal random effect and the robust mixture model are performed. The data will be ordered by duration to assist in identifying a variation from the linear relationship.





**Figure 4:** Forest plot for CDP studies (standardised mean difference for memory measures) with summaries.

```
> exercise <- exercise[order(exercise$duration), ]
> exercise.meta <- metaplust(smd, sqrt(varsm), mods = duration, slab = study,
+ data = exercise)
> summary(exercise.meta)

      Est. 95% ci.lb 95% ci.ub  pvalue
muhat  -2.8994   -4.3006  -1.5222 0.000884
tau2    0.1171
duration 0.2078   0.0584   0.3632 0.011570

      logLik      AIC      BIC
-8.133435 22.26687 23.17462

> exercise.mix <- metaplust(smd, sqrt(varsm), mods = duration, slab = study,
+ random = "mixture", data = exercise)
> summary(exercise.mix)

      Est. 95% ci.lb 95% ci.ub  pvalue
muhat  -2.88472  -4.11082  -1.48262 0.000649
tau2    0.00000
tau2out 0.59398
Outlier prob. 0.25169
duration 0.21086  0.07808  0.34586 0.007052

      logLik      AIC      BIC
-7.69139 25.38278 26.8957

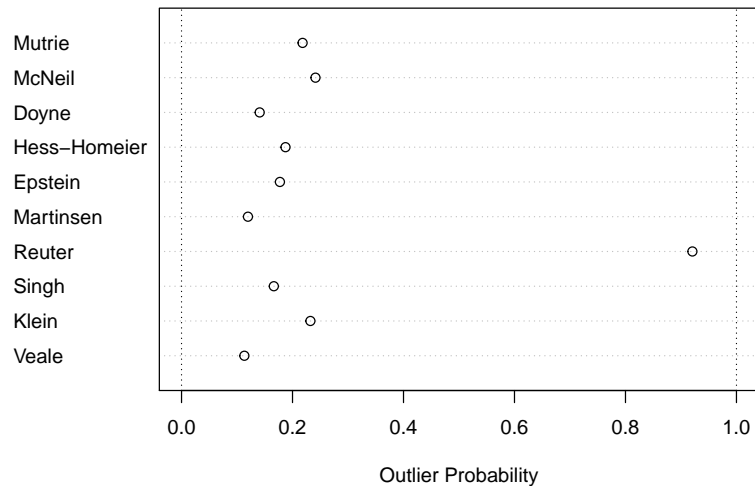
> exercise.testOutliers <- testOutliers(exercise.mix)
> summary(exercise.testOutliers)

Observed LRT statistic 0.9 p value 0.075

> exercise.outlierProbs <- outlierProbs(exercise.mix)
```

The test for outliers was close to being significant ( $p = 0.075$ ); however a conservative approach seems appropriate, by using the robust model where the presence of outliers is not conclusive but there is a reasonable amount of evidence that there are outliers, as in this case. Note also that the  $p$ -value is different from that obtained in [Beath \(2014\)](#), due to the use of randomly generated data in the parametric bootstrap. Running the parametric bootstrap with a large number of simulations showed that the  $p$ -value was actually near 0.04. Using `plot(exercise.outlierProbs)` the outlier probabilities are shown in [Figure 5](#) where the study by Reuter is an obvious outlier with a posterior probability greater than 0.9. This study is a dissertation and was not published in a peer-reviewed journal, and was not included in a later meta-analysis by [Krogh et al. \(2011\)](#). There is also strong evidence of the effect of trial duration.

As `metaplust` does not currently have a `predict` method, the alternative to calculate the effect at each of Weeks 4, 8 and 12 is to centre the data at those times and fit a meta-regression for each ([Johnson](#)



**Figure 5:** Outlier probabilities for depression versus exercise from the robust mixture random effect model.

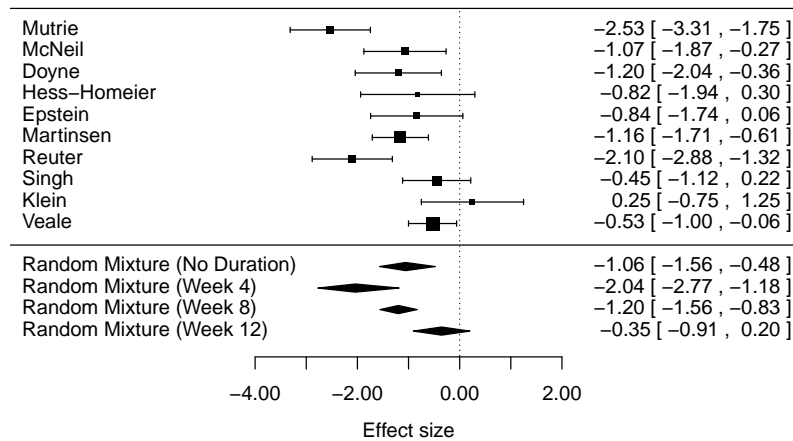
and Huedo-Medina, 2011). The intercept for each meta-regression will then be the estimated mean effect at that time. A model without including the covariate for study duration is also fitted. The forest plot is shown in Figure 6. This shows that the effect of exercise decreases rapidly the longer the trial runs, possibly indicating a placebo effect that rapidly wears off. It would also be possible to include the results from the standard random effect models on the plot.

```
> exercise$duration4 <- exercise$duration - 4
> exercise$duration8 <- exercise$duration - 8
> exercise$duration12 <- exercise$duration - 12
> exercise.nodurn <- metaplust(smd, sqrt(varsm),
+ label = "Random Mixture (No Duration)", slab = study,
+ random = "mixture", data = exercise)
> exercise.wk4 <- metaplust(smd, sqrt(varsm),
+ mods = duration4, label = "Random Mixture (Week 4)",
+ slab = study, random = "mixture", data = exercise)
> exercise.wk8 <- metaplust(smd, sqrt(varsm),
+ mods = duration8, label = "Random Mixture (Week 8)",
+ slab = study, random = "mixture", data = exercise)
> exercise.wk12 <- metaplust(smd, sqrt(varsm),
+ mods = duration12, label = "Random Mixture (Week 12)",
+ slab = study, random = "mixture", data = exercise)
> plot(exercise.nodurn, extrameta = list(exercise.wk4, exercise.wk8,
+ exercise.wk12), xlab = "Effect size")
```

## Conclusions and future developments

The capabilities of the **metaplust** package have been presented for fitting both standard normal random effect and robust random effect models. Using three examples it has been shown how it can test for the presence of outliers and compare the results of the robust and standard methods for both meta-analysis and meta-regression. The package has also been successfully applied to meta-analyses with larger number of studies, for example [Marinho et al. \(2009\)](#) with 70 studies and 3 definite outliers, and simulated data with 200 studies. One difficulty with large number of studies is the increasing computation time, especially for `testOutliers`. This will be improved by the use of parallel processing as a future enhancement.

The design of the package allows for expansion in other areas. A planned future functionality is to fit binary data, using likelihood methods based on the distribution of the binomial responses, rather than the log odds ratios fitted using a normal distribution which is the method currently used. The robust methods can then be applied in a similar way to the current models. A possible future expansion is to allow for other robust distributions although this doesn't seem necessary given the similarity of the results obtained in [Baker and Jackson \(2008\)](#) to those using the  $t$ -distribution.



**Figure 6:** Forest plot for exercise versus depression studies (effect size) with summaries. Studies are sorted by increasing duration.

## Acknowledgements

The author would like to thank the two anonymous reviewers for their comments and suggestions that have greatly improved the quality of the manuscript and the `metapplus` package, and provided suggestions for future enhancements.

## Bibliography

- A. C. Atkinson. Masking unmasked. *Biometrika*, 73(3):533–541, 1986. [p5]
- R. Baker and D. Jackson. A new approach to outliers in meta-analysis. *Health Care Management Science*, 11(2):121–131, 2008. [p5, 6, 14]
- K. Beath, B. Bolker, and R Core Team. *metapplus: Robust Meta-Analysis and Meta-Regression*, 2016. URL <https://CRAN.R-project.org/package=metapplus>. R package version 0.7-7. [p5]
- K. J. Beath. A finite mixture method for outlier detection and robustness in meta-analysis. *Research Synthesis Methods*, 5:285–293, 2014. [p5, 6, 13]
- B. Bolker and R Core Team. *bbmle: Tools for General Maximum Likelihood Estimation*, 2014. URL <http://CRAN.R-project.org/package=bbmle>. R package version 1.0.17. [p7, 9]
- A. J. Branscum and T. E. Hanson. Bayesian nonparametric meta-analysis using Polya tree mixture models. *Journal of the American Statistical Association*, 64(3):825–833, 2008. [p5]
- D. Burr and H. Doss. A Bayesian semiparametric model for random-effects meta-analysis. *Journal of the American Statistical Association*, 100(469):242–251, 2005. [p5]
- W. Chang. *extrafont: Tools for Using Fonts*, 2014. URL <https://CRAN.R-project.org/package=extrafont>. R package version 0.17. [p10]
- E. Demidenko. *Mixed Models: Theory and Applications with R*. Wiley, Hoboken, 2nd edition, 2013. [p5]
- D. Downing. Is ISIS-4 research misconduct? *Journal of Nutritional & Environmental Medicine*, 9:5–13, 1999. [p8]
- M. Fioravanti and M. Yanagi. Cytidinediphosphocholine (CDP choline) for cognitive and behavioural disturbances associated with chronic cerebral disorders in the elderly (review). *The Cochrane Database of Systematic Reviews*, 2005. URL <http://onlinelibrary.wiley.com/store/10.1002/14651858.CD000269.pub2/asset/CD000269.pdf>. [p11]
- M. Gordon and T. Lumley. *forestplot: Advanced Forest Plot Using 'grid' Graphics*, 2015. URL <https://CRAN.R-project.org/package=forestplot>. R package version 1.3. [p7]
- F. N. Gumedze and D. Jackson. A random effects variance shift model for detecting and accommodating outliers in meta-analysis. *BMC Medical Research Methodology*, 11, 2011. [p5, 11]

- R. J. Hardy and S. G. Thompson. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine*, 15:619–629, 1996. [p7]
- J. P. T. Higgins and D. J. Spiegelhalter. Being sceptical about meta-analyses: A Bayesian perspective on magnesium trials in myocardial infarction. *International Journal of Epidemiology*, 31(1):96–104, 2002. [p8]
- J. P. T. Higgins and S. G. Thompson. Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine*, 23(11):1663–82, 2004. [p12]
- ISIS-4: Collaborative Group. ISIS-4: A randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58 050 patients with suspected acute myocardial infarction. *The Lancet*, 345(8951):669–682, 1995. [p8]
- B. T. Johnson and T. B. Huedo-Medina. Depicting estimates using the intercept in meta-regression models: The moving constant technique. *Research Synthesis Methods*, 2(3):204–220, 2011. [p13]
- E. Kontopantelis and D. Reeves. Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: A simulation study. *Statistical Methods in Medical Research*, 21(4):409–426, 2012. [p5]
- J. Krogh, M. Nordentoft, J. A. C. Sterne, and D. A. Lawlor. The effect of exercise in clinically depressed adults: Systematic review and meta-analysis of randomized controlled trials. *The Journal of Clinical Psychiatry*, 72(4):529–38, 2011. [p13]
- D. A. Lawlor and S. W. Hopker. The effectiveness of exercise as an intervention in the management of depression: Systematic review and meta-regression analysis of randomised controlled trials. *British Medical Journal*, 322(31 March):1–8, 2001. [p12]
- K. J. Lee and S. G. Thompson. Flexible parametric models for random-effects distributions. *Statistics in Medicine*, 27:418–434, 2008. [p5, 6]
- J. Li, Q. Zhang, M. Zhang, and M. Egger. Intravenous magnesium for acute myocardial infarction (review). *The Cochrane Library*, 2009. URL <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD002755.pub2/pdf>. [p8]
- V. C. C. Marinho, J. P. T. Higgins, S. Logan, and A. Sheiham. Fluoride toothpastes for preventing dental caries in children and adolescents (review). *The Cochrane Database of Systematic Reviews*, 2009. URL <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD002278/pdf>. [p14]
- G. J. McLachlan. On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics*, 36(3):318–324, 1987. [p7]
- G. J. McLachlan and D. Peel. *Finite Mixture Models*. Wiley, New York, 2000. [p7]
- Y. Pawitan. *In All Likelihood: Statistical Modelling and Inference using Likelihood*. Clarendon Press, Oxford, 1st edition, 2001. [p7]
- J. A. C. Sterne, M. J. Bradburn, and M. Egger. Meta-analysis in Stata. In M. Egger, G. D. Smith, and D. G. Altman, editors, *Systematic Reviews in Health Care: Meta-Analysis in Context*, chapter 18, pages 347–369. BMJ Publishing Group, London, 2001. [p8]
- A. J. Sutton, K. R. Abrams, D. R. Jones, T. A. Sheldon, and F. Song. *Methods for Meta-Analysis in Medical Research*. Wiley, 2000. [p5, 6]
- W. Viechtbauer. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3):1–48, 2010. [p5]
- W. Viechtbauer and M. W.-L. Cheung. Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2):112–125, 2010. [p5]
- K. L. Woods. Commentary: Biostatistics, biological mechanisms and Bayes: Lessons from the magnesium trials. *International Journal of Epidemiology*, 31:105–106, 2002. [p8]

Ken J. Beath  
Department of Statistics  
Faculty of Science and Engineering  
Macquarie University NSW 2109  
Australia  
[ken.beath@mq.edu.au](mailto:ken.beath@mq.edu.au)