Electronic Journal of Statistics
Vol. 5 (2011) 83–101
ISSN: 1935-7524
DOI: 10.1214/11-EJS598

Comparison of methods for fixed effect meta-regression of standardized differences of means

Michael J. Malloy, Luke A. Prendergast and Robert G. Staudte

Department of Mathematics and Statistics
La Trobe University
Bundoora Victoria 3086, Australia
e-mail: M.Malloy@latrobe.edu.au; luke.prendergast@latrobe.edu.au; R.Staudte@latrobe.edu.au

Abstract: Given a number of different studies estimating the same effect size, it is often desired to explain heterogeneity of outcomes using concomitant covariates. For very large sample sizes, effect size estimates are approximately normally distributed and a straightforward application of weighted least squares is appropriate. However in practice within study sample variances are often small to moderate, casting doubt on the normality assumption for effect sizes and results based on weighted least squares. One can alternatively variance stabilize the effect size estimates and adopt a generalized linear model. Both methods are compared on two examples when effect sizes are the standardized difference of means. Then simulation studies are conducted to compare the coverage and width of confidence intervals for the meta-regression coefficients. These simulations show that the coverage probability associated with weighted least squares can be well below the nominated level for small to moderate sample sizes. Further empirical investigations reveal a bias in estimation due to using estimated weights which were assumed to be known. For these models, the generalized linear model approach resulted in much improved coverage probabilities.

AMS 2000 subject classifications: Primary 62J05; secondary 62J12.
Keywords and phrases: Weighted least squares, generalized linear model.

Received June 2010.

Contents

1. Introduction

The combination of results of several studies of a specific topic to increase the accuracy of findings and decisions is called meta-analysis. Meta-regression analysis investigates how the study outcomes in the form of estimated effect sizes are dependent on explanatory variables, often called *moderators* or *covariates* in the meta-analytic literature. Estimated effect sizes include standardized mean differences, log odds ratios, log relative risks and risk differences.

When explanatory variables are assumed to explain all the variation outside of random error in the effect sizes of the studies, the appropriate meta-regression model is one assuming fixed effects. The purpose of this paper is to analyze and compare two fixed effects meta-regression approaches for the investigation of explanatory variables when the effect sizes are standardized mean differences. The first approach is weighted least squares with inverse variance weights. The second approach is generalized linear models involving variance stabilization, first introduced in Chapter 14 of [11].

While there are a large number of meta-regression articles in the literature, see for example, [7–10] and [15], as well as overview publications about how to use meta-regression and interpret its results, see for example, [2, 14] and [17], there does not appear to be any until now that investigate how well these methods perform, even for the fixed effects model. The random effects model in meta-regression is a subject for further research.

In Section 2, we introduce the models and notation and derive a closed form expression for the Hessian matrix of the maximum likelihood estimates of the regression parameters following variance stabilization. Then we evaluate this expression for the standardized difference of means, which simplifies the application of generalized linear models.

In Section 3, two practical examples are analyzed by each method, and in Section 4, the methods are investigated through simulations from which it is shown that when within study samples are small to moderate in size, the generalized linear model approach is preferable as it maintains better coverage probability for most of the model coefficient estimates.

2. Fixed effects meta-regression for the standardized difference of means

2.1. The setting

Suppose we have a meta-analysis with K two-sample studies where each study consists of a control group (Group 1) and a treatment group (Group 2). Further assume that the populations are normal. Letting the sample size of each arm be denoted by $n_{ki}, (k = 1, \ldots, K), (i = 1, 2)$, denote the total sample size in the kth group by $N_k = n_{k1} + n_{k2}$ and the proportion of observations in the kth treatment group by $q_k = n_{k2}/N_k$.

In addition, let μ_{ki} and σ_{ki}^2 denote the kth population mean and variance for the ith arm, estimated by their sample counterparts $\bar{y}_{ki}$ and s_{ki}^2, respectively. The simplest case is to assume common population variances $\sigma_{k1}^2 = \sigma_{k2}^2 = \sigma_k^2$ within each study. The effect size of interest is the standardized mean difference $d_k = (\mu_{k2} - \mu_{k1})/\sigma_k$ which is typically estimated by $\hat{d}_k = (\bar{y}_{k2} - \bar{y}_{k1})/s_{k,pool}$ where $s_{k,pool} = \sqrt{\{(n_{k1} - 1)s_{k1}^2 + (n_{k2} - 1)s_{k2}^2\}/(N_k - 2)}$.

Let $\delta_k = \sqrt{q_k(1 - q_k)}d_k$ for which an estimator is $\hat{\delta}_k = \sqrt{q_k(1 - q_k)}\hat{d}_k$. Then the two-sample t-test statistic

$$t_{k,pool} = \sqrt{N_k}\hat{\delta}_k \sim t_{\nu_k}(\lambda_k) \tag{1}$$

which has a known non-central t-distribution where $\nu_k = N_k - 2$ is the degrees of freedom and λ_k the non-centrality parameter.

2.2. The model and estimation

Consider the following model

$$\mathrm{E}\left(\hat{d}_k|\mathbf{x}_k\right) = \boldsymbol{\beta}^\top \mathbf{x}_k \tag{2}$$

where $\mathbf{x}_k = [1, x_{k1}, \ldots, x_{kp}]^\top \in \mathbb{R}^{p+1}$ is a vector of study level explanatory variables for the kth study, $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_p]^\top \in \mathbb{R}^{p+1}$ is a vector of unknown explanatory variable coefficients and $\hat{d}_k = (\bar{y}_{k2} - \bar{y}_{k1})/s_{k,pool}$ is the estimated standardized difference in mean for the kth study. Two methods are considered for the fitting of the proposed model in (2), weighted least squares and the generalized linear model following variance stabilization.

2.2.1. The weighted least squares approach

To apply weighted least squares (WLS) directly, the non-central t effect sizes are assumed to be approximately normally distributed with ‘known’ weights $w_k = 1/[\mathrm{se}(\hat{d}_k)]^2$, where

$$\mathrm{se}(\hat{d}_k) = \sqrt{N_k/(n_{k1}n_{k2}) + \hat{d}_k^2/(2N_k)} \tag{3}$$

[6, p.86]. Equation (3) is an approximate large sample estimator of the standard error of $\hat{d}_k$. Although there are other estimation formulae for $[\mathrm{se}(\hat{d}_k)]^2$, see for example [18], we will be assuming (3) for this procedure as it is commonly found in the literature and used in practice.

In this setting, a WLS approach assumes that, approximately,

$$\hat{d}_k = \boldsymbol{\beta}^\top \mathbf{x}_k + \varepsilon_k$$

where $\varepsilon_k \sim N(0, [\mathrm{se}(\hat{d}_k)]^2)$ and the $[\mathrm{se}(\hat{d}_k)]^2$, given in (3), is assumed to be known, even though it is estimated.

In typical application of the WLS approach, the variances of the dependent variables are unknown and assumed to be of the form σ^2/w where σ^2 needs to be estimated and w is some fixed weight. Standard WLS estimates σ^2 via the mean square error (MSE) and the standard errors of the coefficient estimates are a scalar multiple of $\sqrt{\mathrm{MSE}}$. In our case, there is no unknown component in the variance and the equivalence here is fixing $\sigma^2 = 1$. As such, the correct standard errors to use are the traditional WLS standard errors divided by $\sqrt{\mathrm{MSE}}$. Also, since there is no variance component to estimate, $(1-\alpha/2)100\%$ confidence intervals for β_j $(j = 0, 1, \ldots, p)$ are then of the form $\widehat{\beta}_j \pm z_{1-\alpha/2}\mathrm{se}(\widehat{\beta}_j)$ where the $\mathrm{se}(\widehat{\beta}_j) = [(\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1}]^{1/2}_{(j+1)(j+1)}$, $\mathbf{X} = [\mathbf{x}_k]$ which is a $K \times (p+1)$ design matrix of full rank with kth row equal to $x_k^\top$ and $\mathbf{W}$ is a $K \times K$ constant diagonal matrix whose elements are $w_1, \ldots, w_K$. Here $z_{1-\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution. For more about WLS, see p.176 of [13].

2.2.2. The generalized linear model approach

The generalized linear model (GLM) approach cannot be applied directly to the $\hat{d}_k$'s, because their distributions do not belong to an exponential family. By application of a variance stabilizing transformation one obtains approximately normal transformed effect sizes. The transformation we apply, from [1] [also see 11, Ch.20], is

$$\mathcal{K}(\delta_k) = \sqrt{2}\sinh^{-1}(\delta_k/\sqrt{2}) \tag{4}$$

where $\sinh^{-1}(x) = \ln(x + \sqrt{x^2+1})$. This produces approximately normal transformed effect sizes $Y_k = \mathcal{K}(\hat{\delta}_k) \overset{a}{\sim} N(\mathcal{K}(\delta_k), N_k^{-1})$ where $\overset{a}{\sim}$ is the symbol for 'approximately distributed as' and the observed transformed effect sizes are denoted by $y_1, \ldots, y_K$.

Now consider the following model

$$Y_k = \mu_{\boldsymbol{\beta}}(\mathbf{x}_k) + \varepsilon_k \tag{5}$$

where $\mu_{\boldsymbol{\beta}}(\mathbf{x}_k)$ is a function of the unknown $\boldsymbol{\beta}$ and the fixed $\mathbf{x}_k$, $\varepsilon_k \sim N(0, 1/N_k)$ and, subsequently, $Y_k \sim N(\mu_{\boldsymbol{\beta}}(\mathbf{x}_k), 1/N_k)$. Under the setting described in Section 2.1 and an appropriate choice of $\mu_{\boldsymbol{\beta}}$ in (5), we propose that this model holds approximately for our transformed effect sizes. We now consider the Hessian matrix associated with parameter $\boldsymbol{\beta}$ in the following proposition. A simple proof can be found in Appendix A or one may utilize, for example, Chapter 3 of [5] for general results with the distribution of Y_k belonging to the exponential family. For more about GLM's, see for example, [12].

Proposition 2.1. *Under the model in (5), the Hessian matrix associated with the vector of explanatory variable coefficients $\boldsymbol{\beta}$ is*

$$\mathbf{H}(\boldsymbol{\beta}) = \mathbf{X}^\top \mathbf{D}\mathbf{X}$$

where $\mathbf{X}$ *is the design matrix whose* k*th row is* $\mathbf{x}_k^\top$ *and* $\mathbf{D}$ *is the diagonal matrix with* k*th diagonal element*

$$\mathbf{D}_{kk} = -N_k \left\{ \left(\frac{\partial \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)}{\partial \boldsymbol{\beta}^\top \mathbf{x}_k} \right)^2 - [y_k - \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)] \frac{\partial^2 \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)}{\partial (\boldsymbol{\beta}^\top \mathbf{x}_k)^2} \right\}.$$

Consequently, for $\widehat{\boldsymbol{\beta}}$ denoting the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ in (5), we have that

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}) \approx -\mathbf{H}^{-1}(\widehat{\boldsymbol{\beta}})$$

for $\mathbf{H}(\cdot)$ given in Proposition 2.1.

Restricting attention to the setting presented in Section 2.1, the generalized linear model of (5) will require the link function $g(y) = \sqrt{2}\sinh(y/\sqrt{2})/\sqrt{q(1-q)}$ where q is the proportion of cases in the second arm of the study. Then the inverse link function is

$$g^{-1}(x) = \sqrt{2}\sinh^{-1}\left\{ [q(1-q)/2]^{1/2} x \right\}$$

which in turn gives, for the kth study

$$\mu_{\boldsymbol{\beta}}(\mathbf{x}_k) = \sqrt{2}\sinh^{-1}\left\{ [q_k(1-q_k)/2]^{1/2} \boldsymbol{\beta}^\top \mathbf{x}_k \right\}. \tag{6}$$

From (6), we have that

$$\mu_{\boldsymbol{\beta}}(\mathbf{x}_k)/\sqrt{2} = \sinh^{-1}\left\{ [q_k(1-q_k)/2]^{1/2} \boldsymbol{\beta}^\top \mathbf{x}_k \right\}$$

so that

$$\frac{\sinh\left(\mu_{\boldsymbol{\beta}}(\mathbf{x}_k)/\sqrt{2}\right)}{\sqrt{q_k(1-q_k)/2}} = \boldsymbol{\beta}^\top \mathbf{x}_k. \tag{7}$$

From (7)

$$\frac{\partial \boldsymbol{\beta}^\top \mathbf{x}_k}{\partial \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)} = \frac{1}{\sqrt{q_k(1-q_k)}} \cosh(\mu_{\boldsymbol{\beta}}(\mathbf{x}_k)/\sqrt{2})$$

which, since $\cosh^2(\mu_{\boldsymbol{\beta}}(\mathbf{x}_k)/\sqrt{2}) = 1 + \sinh^2(\mu_{\boldsymbol{\beta}}(\mathbf{x}_k)/\sqrt{2})$ and using (7), becomes

$$\begin{aligned}\frac{\partial \boldsymbol{\beta}^\top \mathbf{x}_k}{\partial \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)} &= \frac{1}{\sqrt{q_k(1-q_k)}} \sqrt{1 + \frac{q_k(1-q_k)}{2} \left(\boldsymbol{\beta}^\top \mathbf{x}_k\right)^2} \\ &= \sqrt{\frac{1}{q_k(1-q_k)} + \frac{1}{2}\left(\boldsymbol{\beta}^\top \mathbf{x}_k\right)^2}.\end{aligned}$$

Consequently,

$$\frac{\partial \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)}{\partial \boldsymbol{\beta}^\top \mathbf{x}_k} = \frac{1}{\sqrt{\frac{1}{q_k(1-q_k)} + \frac{1}{2}\left(\boldsymbol{\beta}^\top \mathbf{x}_k\right)^2}}. \tag{8}$$

For the second derivative, from (8) we have

$$\frac{\partial^2 \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)}{\partial(\boldsymbol{\beta}^\top \mathbf{x}_k)^2} = \frac{-\sqrt{2}(\boldsymbol{\beta}^\top \mathbf{x}_k)}{\left(\frac{2}{q_k(1-q_k)} + (\boldsymbol{\beta}^\top \mathbf{x}_k)^2\right)^{3/2}}. \tag{9}$$

Hence, from (8) and (9), $\mathbf{D}$ in Proposition 2.1 for the standardized difference of means effect size is the matrix whose kth diagonal element is

$$\mathbf{D}_{kk} = -\frac{N_k}{c_k}\left\{2 + \frac{[y_k - \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)]\boldsymbol{\beta}^\top \mathbf{x}_k}{\sqrt{c_k/2}}\right\}$$

where

$$c_k = \frac{2}{q_k(1-q_k)} + (\boldsymbol{\beta}^\top \mathbf{x}_k)^2.$$

Given a maximum likelihood estimate for $\boldsymbol{\beta}$, the unknown $\boldsymbol{\beta}$ in (8) and (9) can be replaced with this estimate to obtain an estimate of the associated Hessian matrix.

3. Examples

In this section, we will consider application of both the WLS and GLM approaches to two data sets. The first is an example consisting of just the one explanatory variable (i.e. $p = 1$) and the second extends this to consider an example with $p = 3$ explanatory variables. Throughout this section the freely available statistical software package R [see 16] was used. For the minimization required for the maximum likelihood estimates, we have used the R optimization function `nlminb` which utilizes PORT routines [see for example, 4].

3.1. Effect of open versus traditional education on student creativity

Data with effect size estimates from $K = 10$ studies of the effect of open versus traditional education on students creativity are given in Table 1. This example has previously been considered by [6] and more recently by [11] from which Table 1 was reproduced.

The question of interest is whether there is a significant difference between younger and older grade levels in regards to their impact on student creativity when open education is the learning mechanism. The explanatory variable is the grade level for which the study was conducted and we want to include an intercept coefficient so that each explanatory variable vector is equal to $[1, x_k]^\top$ for $k = 1, \ldots, K$.

Note that, each study has the same number of cases in both arms (i.e. $n_{k1} = n_{k2}$) so that $q_k = 1/2$ for all k. Equal variances in both arms is assumed and the variance stabilizing transformation (vst) given in (4) was used to produce

TABLE 1
Effect size ($\hat{d}_k$) and transformed effect size (y_k) estimates from $K = 10$ studies investigating the effect of open versus traditional education on student creativity

Study	$n_1 = n_2$	Grade Level	$\hat{d}_k$	y_k
1	90	6	−0.581	−0.288
2	40	5	0.530	0.263
3	36	3	0.771	0.381
4	20	3	1.031	0.505
5	22	2	0.553	0.275
6	10	4	0.295	0.147
7	10	8	0.078	0.039
8	10	1	0.573	0.284
9	39	3	−0.176	−0.088
10	50	5	−0.232	−0.116

Note: n_1 – sample size of the students undertaking open education;
n_2 – sample size of the students undertaking traditional education.

TABLE 2
Meta-regression results of the WLS and GLM approaches when fitting model (10) to the open versus traditional education on student creativity data

Method	RSS	β_j	$\hat{\beta}_j$	se($\hat{\beta}_j$)	95% CI	P-value
WLS	27.254	β_0	1.023	0.237	$[0.558, 1.487]$	0.000
		β_1	−0.218	0.050	$[-0.317, -0.120]$	0.000
GLM	27.742	β_0	1.053	0.238	$[0.587, 1.519]$	0.000
		β_1	−0.224	0.051	$[-0.323, -0.125]$	0.000

Note: The P-values are z-test p-values; RSS – residual sum of squares;
β_j – jth coefficient; $\hat{\beta}_j$ – jth coefficient estimate;
se($\hat{\beta}_j$) – standard error of jth coefficient estimate; CI – confidence interval.

the y_k's, that is, the estimated transformed effect sizes. The linear regression model fitted to this data set, from (2), is

$$\mathrm{E}(\hat{d}_k|x_k) = \beta_0 + \beta_1 x_k \tag{10}$$

where x_k is the grade level of interest.

Table 2 gives the results of the data analysis when fitting (10) using both the traditional WLS technique and the GLM approach. As can be seen in Table 2, both methods produce very similar parameter estimates and standard errors with the GLM method estimates being marginally larger. As a result, the 95% confidence intervals for β_0 and β_1 were also alike.

Furthermore, it is clear the parameter estimates of both methods are significant at the 0.05 level as all were found to have z-test p-values < 0.001. Also, when comparing the residual sum of squares (RSS) for each approach (both of which have 8 degrees of freedom), we can see that they are very similar with the WLS value being marginally smaller. Consequently, the null hypothesis of both procedures that all the model coefficients are equal to zero is rejected in favor of the alternative that at least one coefficient is non-zero at the 0.01 level using a chi-squared distribution with 8 degrees of freedom.

The question this example leaves us with is which methods results are preferable in this instance? Simulations based on this example, which are discussed in Section 4, were used to investigate this question.

3.2. The neuropsychological impact of sports-related concussion on "exposed" athletes

Data with effect size estimates from $K = 9$ studies of the neuropsychological impact on athletes which participate in a sport where the head is "exposed" to knocks are given in Table 3. This table was reproduced from [3].

The questions of interest are firstly whether there is a significant difference between soccer and boxing on the neuropsychological function of a participating athlete, and secondly, does the standard at which the athlete competes have a significant affect on their neuropsychological function as well.

The sport the athletes participated in and the standard or level of play at which they compete at are the explanatory variables investigated. The levels of play considered in the studies are amateur (A) and professional (P) with some studies considering both (A,P). Athletes that participated in sports where the head is generally not "exposed" to knocks, such as track and field, were the control sample. Furthermore, sample sizes are not equal in every study so the q_k are calculated individually. The assumption of equal variances in each arm is made. The 9 studies results in effect sizes $\hat{d}_k$ are transformed by (4) to obtain approximately normal y_k's. As in the previous example, we include an intercept coefficient.

The linear regression model fitted to this data set, from (2), is

$$\mathrm{E}(\hat{d}_k|x_k) = \beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2} + \beta_3 x_{k3} \tag{11}$$

where x_{k1} is the level of play which is equal to 1 when the standard is amateur and 0 otherwise, x_{k2} is also the level of play which is equal to 1 when the standard

TABLE 3

Effect size ($\hat{d}_k$) and transformed effect size (y_k) estimates from $K = 9$ studies of the neuropsychological impact on athletes which participate in a sport where the head is "exposed" to knocks

Study	Athletes n_1	Controls n_2	Sport	Level of Play	$\hat{d}_k$	y_k
1	31	31	Soccer	A	−0.18	−0.090
2	29	19	Boxing	A	0.41	0.200
3	32	29	Soccer	A,P	0.39	0.194
4	19	10	Boxing	P	1.08	0.503
5	10	10	Boxing	A,P	0.31	0.155
6	25	25	Boxing	A	0.22	0.110
7	37	20	Soccer	P	0.49	0.233
8	60	20	Soccer	A,P	0.21	0.091
9	21	12	Soccer	A	0.27	0.130

Note: A – amateur level of play; P – professional level of play;
A,P – combination of both amateur and professional level of play.

TABLE 4
Meta-regression results of the WLS and GLM approaches when fitting model (11) to the neuropsychological impact on athletes which participate in a sport where the head is "exposed" to knocks data

Method	RSS	β_j	$\hat{\beta}_j$	$\mathrm{se}(\hat{\beta}_j)$	95% CI	P-value
WLS	2.186	β_0	0.577	0.256	$[0.074, 1.079]$	0.024
		β_1	-0.277	0.238	$[-0.744, 0.190]$	0.245
		β_2	0.319	0.291	$[-0.251, 0.888]$	0.273
		β_3	-0.321	0.224	$[-0.760, 0.119]$	0.153
GLM	2.192	β_0	0.580	0.258	$[0.075, 1.085]$	0.024
		β_1	-0.279	0.239	$[-0.747, 0.190]$	0.243
		β_2	0.319	0.291	$[-0.251, 0.889]$	0.273
		β_3	-0.323	0.224	$[-0.763, 0.117]$	0.150

Note: The P-values are z-test p-values; RSS – residual sum of squares;
β_j – jth coefficient; $\hat{\beta}_j$ – jth coefficient estimate;
$\mathrm{se}(\hat{\beta}_j)$ – standard error of jth coefficient estimate; CI – confidence interval.

is professional and 0 otherwise, and x_{k3} which is the sport played which is equal to 1 when the sport is soccer and 0 when the sport is boxing. Table 4 lists the results of using both the WLS and GLM approaches when fitting (11) to the data.

It is clear from Table 4 that both meta-regression procedures produce very similar parameter, standard error and 95% confidence interval estimates. When comparing the p-values of each coefficient from both approaches, they were also alike, of which only β_0 was found to be significant at the 0.05 level for both procedures.

Furthermore, when considering the RSS of WLS and GLM (both of which have 5 degrees of freedom), the WLS value was found to be slightly smaller. However, the null hypothesis of both procedures is not rejected at the 0.01 level using a chi-squared distribution with 5 degrees of freedom.

4. Simulations

In this section, we consider simulations for the comparison between the WLS and GLM approaches. Certain subtleties require mention here with regards to simulating data according to (2). Firstly, from (1) we have that $\sqrt{N_k q_k(1-q_k)}\hat{d}_k \sim t_{\nu_k}(\lambda_k)$ so that (2) can be rewritten as

$$\mathrm{E}(\sqrt{N_k q_k(1-q_k)}\hat{d}_k|\mathbf{x}_k) = \sqrt{N_k q_k(1-q_k)}(\boldsymbol{\beta}^\top \mathbf{x}_k). \tag{12}$$

Using the fact that, for $T \sim t_\nu(\lambda)$, $\mathrm{E}(T) = \lambda\sqrt{(\nu/2)}\Gamma((\nu-1)/2)/\Gamma(\nu/2)$, we have, from (12), that

$$\sqrt{N_k q_k(1-q_k)}\boldsymbol{\beta}^\top \mathbf{x}_k = \lambda_k \sqrt{\frac{\nu_k}{2}}\frac{\Gamma((\nu_k-1)/2)}{\Gamma(\nu_k/2)}$$

which then gives us

$$\lambda_k = \sqrt{\frac{2}{\nu_k}}\frac{\sqrt{N_k q_k(1-q_k)}\boldsymbol{\beta}^\top \mathbf{x}_k}{\Gamma((\nu_k-1)/2)/\Gamma(\nu_k/2)}. \tag{13}$$

Using (13), we are therefore able to simulate data under the model in (2) for fixed q_k, N_k, $\mathbf{x}_k$ and a proposed $\boldsymbol{\beta}$, by randomly generating values for each $\sqrt{N_k}\hat{\delta}_k$ from the $t_{\nu_k}(\lambda_k)$ distribution where $\nu_k = N_k - 2$ and the non-centrality parameter λ_k, which is provided in (13).

Let us consider again the example described in Section 3.1. Simulations applying the WLS and GLM methods to the proposed model in (2) were conducted for comparative purposes with the parameter of interest being β_1. In all the simulations β_0 was fixed at 1.05, the estimated value from the example of Section 3.1, and β_1 was selected over the range -3 to 3 in increments of 0.02.

For each value of β_1, 10000 trials were performed from which coverage probabilities and average confidence interval widths were estimated. Coverage probability is the proportion of times that the true parameter value of interest (in this simulation, the β_1 parameter) falls within the estimated confidence interval bounds. The nominal confidence level for these estimated intervals was set at the usual 95%.

Recall from Section 2.2.1 that the approximate confidence interval formula for the model coefficients is $\hat{\beta}_j \pm z_{1-\alpha/2}\mathrm{se}(\hat{\beta}_j)$. The first simulations performed looked at the effect that varying K in a meta-analysis would have on the model estimates of each procedure. For these simulations, K was set equal to 10, 20, 30 and 50. As $K = 10$ in the example, both the within study sample sizes and the explanatory variable vector from the example were simply replicated to account for the changes in K. The results of each K are plotted in Figure 1.

It is clear from the output in Figures 1 (a) and (b) that for small to moderate size meta-analyses, the estimated coverage probabilities for both methods are similar. We can see that for small to moderate K, the estimated coverage probability for GLM approach remains very stable and consistently ranges between 94% and 95% for majority of the β_1 values.

When considering the WLS approach, the coverage also remains very consistent for small to moderate K. We can see in Figures 1 (a) and (b) that its coverage probability for negative values of β_1 performs as well as the GLM's and even slightly better for positive values.

However, when meta-analyses contain larger K, the WLS coverage fluctuates a little. As can be seen in Figures 1 (c) and (d), its coverage tends to be slightly lower than that of the GLM method for negative values of β_1, dropping as low as 93% for some β_1 values when $K = 50$. Whereas for positive values of β_1, the coverage remains between the 94% and 95% range.

In regard to the average confidence interval widths estimated, it can clearly be seen in Figure 1 that despite the size of the meta-analysis, there is no significant difference in the interval widths produced by both methods. As expected, the width of the intervals increase as the value of β_1 becomes larger in the negative and positive directions.

Although there are marginal differences in the coverage between the methods, the results of the average confidence interval widths and the coverage indicates that as K increases, the use of either approach is appropriate under the assumption that (2) is the true underlying model.

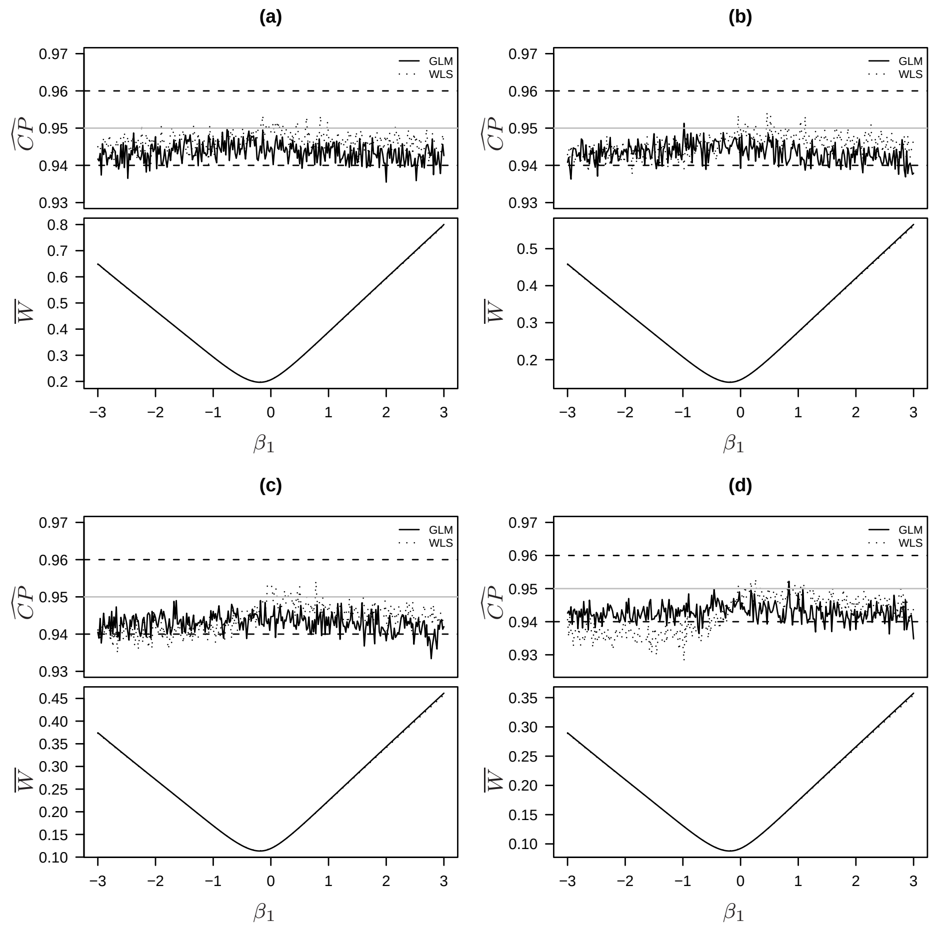

FIG 1. *Estimated coverage probabilities* ($\widehat{\mathrm{CP}}$) *and average* 95% *confidence interval widths* ($\overline{\mathrm{W}}$) *for the GLM and WLS approaches where (a)* $K = 10$, *(b)* $K = 20$, *(c)* $K = 30$ *and (d)* $K = 50$ *with the sample sizes from the study used,* β_0 *fixed to* 1.05 *and* $\beta_1 \in [-3, 3]$.

It is important to note that the within study sample sizes from the example used in the above simulation varied a bit in size. Some of the sample sizes where quite big, for example, both arms in study one had sample sizes of 90. Whereas, some of the studies sample sizes were quite small. For example, both arms in studies six, seven and eight had sample sizes of 10. Noting this, an obvious question that needed investigation was how do these methods perform when all the within study sample sizes in each arm of a meta-analysis are small?

To understand the effect that small within study sample sizes may have on each approaches parameter estimates, the above simulation was conducted again except the size of the samples for both arms were set to 10. The results of each K are displayed in Figure 2.

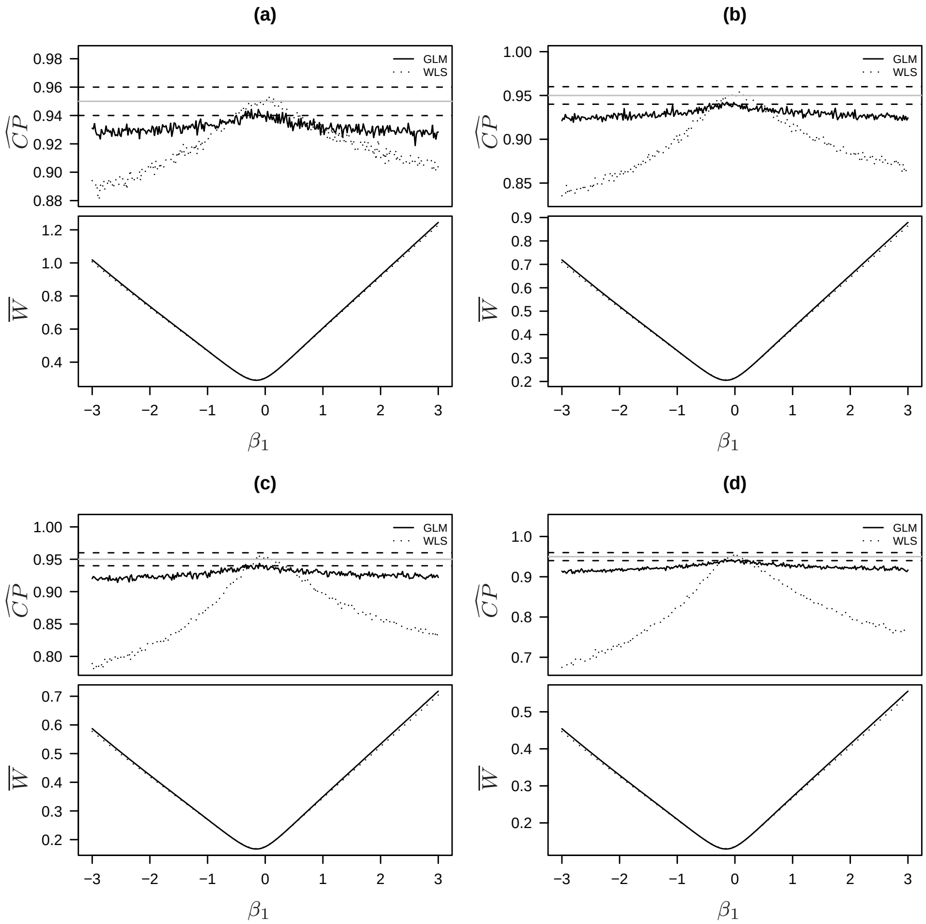

Fig 2. *Estimated coverage probabilities* ($\widehat{\mathrm{CP}}$) *and average* 95% *confidence interval widths* ($\overline{\mathrm{W}}$) *for the GLM and WLS approaches where (a)* $K = 10$*, (b)* $K = 20$*, (c)* $K = 30$ *and (d)* $K = 50$ *with* $n_1 = n_2 = 10$*,* β_0 *fixed to* 1.05 *and* $\beta_1 \in [-3, 3]$.

From the output in Figure 2 (a), it is clear that when within study sample sizes are small with small K, the GLM approach for majority of the choices of β_1 outperforms WLS in terms of coverage. Its coverage seems to consistently range between 92% and 94%. On the other hand, the WLS coverage is not as stable. Its coverage continually decreases as β_1 becomes larger in both the positive or negative directions. Although, for a small range of β_1 values around 0, the coverage of WLS is as good as and even outperforms the GLM method.

Similarly, the results in Figures 2 (b), (c) and (d) show that as K increases, the difference in performance in terms of coverage probability increases significantly in favor of the GLM procedure. For example, when K was increased into 50, the coverage probability when applying the GLM approach remains above 90% across the range of β_1 whereas the WLS coverage drops to below 70% for some values of β_1. However, for a small range of β_1 values, again around 0, the coverage of WLS is comparable with the GLM methods.

When considering the average confidence interval widths, it is clear from Figure 2 that even with small within study sample sizes, there is no significant difference in the interval widths produced by both methods.

Thus, regardless of the size of the meta-analysis, when the within study sample sizes are all small in both arms, the GLM method may be the preferable approach when (2) is the assumed true underlying model as better coverage is maintained for most values of the estimated model coefficient.

Meta-analyses with moderate sample sizes were also considered. Setting the size of the samples for both arms to 30, simulations for varying K were conducted with the results shown in Figure 3.

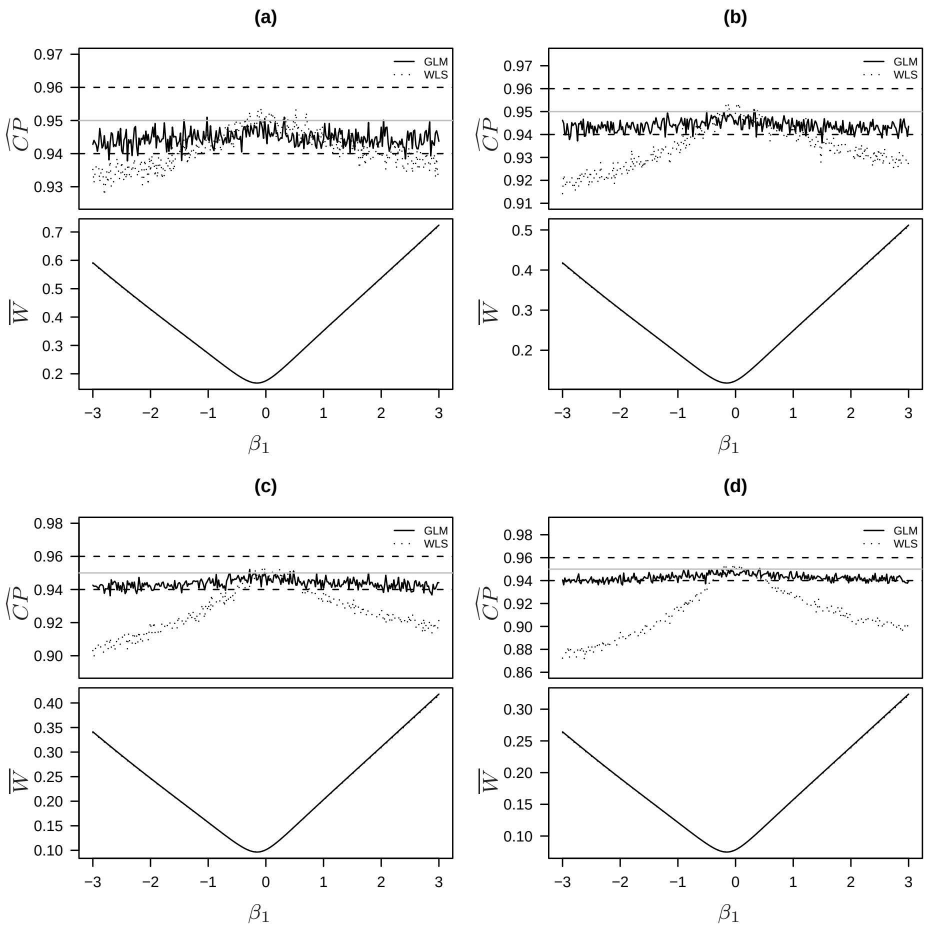

Fig 3. *Estimated coverage probabilities* ($\widehat{\mathrm{CP}}$) *and average* 95% *confidence interval widths* ($\overline{\mathrm{W}}$) *for the GLM and WLS approaches where (a)* $K = 10$*, (b)* $K = 20$*, (c)* $K = 30$ *and (d)* $K = 50$ *with* $n_1 = n_2 = 30$*,* β_0 *fixed to* 1.05 *and* $\beta_1 \in [-3, 3]$.

We can see from Figure 3 that the GLM approach for majority of the β_1 values outperforms WLS in terms of coverage, regardless of the choice of K. This difference tends to increase as K increases, with the GLM coverage consistently ranging between 94% and 95% for all choices of K. This is in contrast to WLS where for larger positive and negative values of β_1, its coverage drops when K becomes larger. For example, when $K = 10$, the larger positive and negative values of β_1 records a coverage that is only slightly lower than the GLM by about 1%. But when $K = 50$, this difference is quite significant with some choices of β_1 having a coverage around 6% lower than the GLM. However, for β_1 values around 0, the coverage of WLS is as good as and even outperforms the GLM method.

In addition, it is important to note that the difference in the methods coverage is not as significant as when the within study sample sizes were small. This indicates, and as would be expected, that the WLS method performs much better when sample sizes in studies are larger. Moreover, when looking at the average confidence interval widths for both approaches, it is evident that there is no significant difference in the interval widths produced by both methods.

Hence, with the increase in sample sizes in both arms of each study, the performance of WLS improves in terms of coverage. However, it is obvious that when dealing with a meta-analysis of any size K with small to moderate within study samples, the GLM approach may be the preferred method.

Table 5 lists estimated coverage probabilities and average confidence interval widths for three choices β_1, as well as standard deviation estimates for the average confidence interval widths, which are noted in parentheses. The three values of β_1 considered were $\beta_1 = -1, 0$ and 1, to which 10000 trials were performed for each value. The sample sizes in each study for both arms were set equal (i.e. $n_1 = n_2$), β_0 was fixed to 1.05 and the grade level again was our explanatory variable of interest.

We can see that for most choices of K and sample size, the average confidence interval widths and their respective standard deviation estimates of both methods are almost equivalent (at least to 3 decimal places) for each value of β_1. As well as providing more evidence that there is no significant difference in width of confidence intervals between these two approaches, the results in Table 5 further verifies that for larger meta-analyses, both methods maintain similar coverage probabilities for most β_1 values. However, it also shows that if a meta-analysis has small to moderate within study sample sizes, that regardless of the size of K, the GLM will have better coverage for most β_1 values.

The outcomes of these simulations raised the following question. Given that the average estimated confidence widths are similar, why is the estimated coverage probability for the WLS approach comparatively well below the nominated level of 95% as the parameter moves further from zero?

The results suggest that the confidence intervals are based on a biased estimate of the parameter and where this bias grows with $|\beta_1|$. When the weights for WLS are fixed (i.e. when the true d_k's are used in the weights) it is straightforward to show that the corresponding WLS estimate is unbiased.

TABLE 5

Estimated coverage probabilities ($\widehat{\mathrm{CP}}$) *and average* 95% *confidence interval widths* ($\overline{\mathrm{W}}$) *with standard deviation estimates in parentheses for the GLM and WLS approaches and* β_0 *fixed to* 1.05

			$n_1 = n_2$							
			10		20		50		100	
β_1	K	Method	$\widehat{\mathrm{CP}}$	$\overline{\mathrm{W}}$	$\widehat{\mathrm{CP}}$	$\overline{\mathrm{W}}$	$\widehat{\mathrm{CP}}$	$\overline{\mathrm{W}}$	$\widehat{\mathrm{CP}}$	$\overline{\mathrm{W}}$
−1	10	WLS	0.923	0.467	0.941	0.332	0.944	0.211	0.946	0.149
				(0.023)		(0.011)		(0.004)		(0.002)
		GLM	0.929	0.469	0.940	0.333	0.946	0.211	0.947	0.149
				(0.020)		(0.010)		(0.004)		(0.002)
	20	WLS	0.901	0.330	0.926	0.235	0.943	0.149	0.943	0.105
				(0.011)		(0.005)		(0.002)		(0.001)
		GLM	0.932	0.332	0.939	0.235	0.948	0.149	0.947	0.105
				(0.010)		(0.005)		(0.002)		(0.001)
	50	WLS	0.814	0.208	0.895	0.148	0.932	0.094	0.938	0.067
				(0.004)		(0.002)		(0.001)		(0.000)
		GLM	0.925	0.210	0.938	0.149	0.945	0.094	0.948	0.067
				(0.004)		(0.002)		(0.001)		(0.000)
0	10	WLS	0.947	0.306	0.949	0.215	0.949	0.136	0.952	0.096
				(0.009)		(0.004)		(0.002)		(0.001)
		GLM	0.941	0.304	0.949	0.215	0.953	0.136	0.950	0.096
				(0.007)		(0.003)		(0.001)		(0.001)
	20	WLS	0.954	0.216	0.951	0.152	0.946	0.096	0.947	0.068
				(0.004)		(0.002)		(0.001)		(0.000)
		GLM	0.935	0.215	0.944	0.152	0.949	0.096	0.957	0.068
				(0.004)		(0.002)		(0.001)		(0.000)
	50	WLS	0.954	0.137	0.951	0.096	0.957	0.061	0.951	0.043
				(0.002)		(0.001)		(0.000)		(0.000)
		GLM	0.940	0.136	0.943	0.096	0.946	0.061	0.952	0.043
				(0.001)		(0.001)		(0.000)		(0.000)
1	10	WLS	0.932	0.602	0.941	0.430	0.949	0.272	0.944	0.193
				(0.037)		(0.018)		(0.007)		(0.004)
		GLM	0.932	0.606	0.946	0.431	0.946	0.273	0.946	0.193
				(0.029)		(0.014)		(0.006)		(0.003)
	20	WLS	0.912	0.424	0.933	0.303	0.945	0.192	0.949	0.136
				(0.018)		(0.009)		(0.004)		(0.002)
		GLM	0.928	0.428	0.943	0.304	0.950	0.193	0.950	0.137
				(0.015)		(0.007)		(0.003)		(0.001)
	50	WLS	0.861	0.267	0.911	0.191	0.938	0.122	0.942	0.086
				(0.007)		(0.004)		(0.001)		(0.001)
		GLM	0.930	0.271	0.942	0.192	0.950	0.122	0.946	0.086
				(0.006)		(0.003)		(0.001)		(0.001)

Note: K – number of studies; n_1 – sample size of group 1; n_2 – sample size of group 2.

On the other hand, this is a totally impractical scenario since it is the d_k's themselves that need to be estimated and subsequently modeled. The bias term for the WLS without fixing the weights remains elusive due to the complexity of the WLS estimator that allows for random weights and responses. However, we shed some light on this in Figure 4 by considering estimated densities of GLM, WLS and WLS with true weights (which we will denote as WLS*) for three choices of β_1 which are $\beta_1 = -2, 0$ and 2. All the densities estimated were over 10000 trials and with β_0 fixed to 1.05.

Figure 4 (a)–(c) plots the results for $K = 50$ with $n_1 = n_2 = 10$. For $\beta_1 = -2$ we see that the estimated densities for the GLM and WLS* are approximately

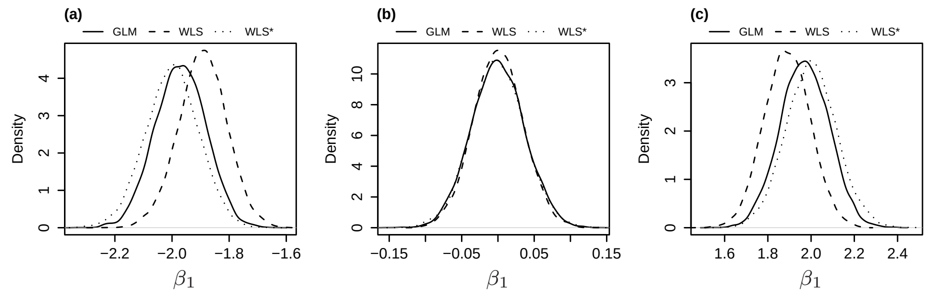

Fig 4. *Estimated density plots of the GLM, traditional WLS with estimated weights and WLS with the true weights (denoted by WLS*) for (a)* $\beta_1 = -2$, *(b)* $\beta_1 = 0$ *and (c)* $\beta_1 = 2$ *where* $K = 50$, $n_1 = n_2 = 10$ *and* β_0 *fixed to* 1.05.

centered around the true β_1. However, for the WLS approach, the estimated density is similar in shape but is shifted to the left indicating an obvious negative bias in estimation. When β_1 is set to zero this difference in estimated densities disappears and for $\beta_1 = 2$ the bias for the WLS is then positive. These findings support those shown in Figures 2 and 3 where the GLM and WLS approaches are comparable when β_1 is small, but where the performance of WLS declines when β_1 is chosen further from 0. In addition, it is clear from the earlier simulations that this estimation bias is a problem when a meta-analysis has small to moderate within study sample sizes. When the within study sample sizes are large, the bias is significantly reduced.

5. Discussion

Two meta-regression approaches (GLM and WLS) have been analyzed and compared under the assumption that a linear model explains the true relationship between the effect size of interest and the covariates under consideration. Traditionally, WLS has been used in this capacity, though the GLM approach has received some recent attention [11]. There has been very little in the way of comparisons made between them.

Through simulation we have shown that the amount of studies and their respective sample sizes in a meta-analysis plays a role in the effectiveness and accuracy of a meta-regression method when applied to an assumed underlying model. Both WLS and GLM were found to perform comparably in terms of their width of confidence intervals regardless of the number of studies or their within sample sizes. Likewise, the coverage probabilities of both methods β_1 were also found to be similar but only in the case where a meta-analysis contained large within study sample sizes. On the other hand, when within study sample sizes are small to moderate in size, significant differences in the estimation of the true β_1 are evident.

It was shown through the simulations that if within study sample sizes are small to moderate in size, WLS unfortunately contains an estimation bias of the model parameter and where this bias grows with $|\beta_1|$. As a result, its performance in terms of coverage probability significantly reduces the further β_1 is chosen from 0. On the other hand, the GLM approach was found to maintain significantly higher coverage as the $|\beta_1|$ increases, although for a small range β_1 values around 0, the coverage of WLS is comparable to GLM's.

It is clear from these results that one must take care in the method of choice when conducting a meta-regression analysis while assuming (2). Whilst the coverage probabilities for WLS were typically smaller than the GLM, with large enough study sample sizes, both methods would be appropriate when assuming (2) as their respective coverage probabilities are very close to the nominated coverage level. In contrast, if a meta-analyses contains small to moderate within study samples, the GLM approach may be the preferred option as it maintains a superior coverage for many parameter choices with respect to our chosen models.

Although the simulations conducted were for a linear model with one covariate, we did investigate a multiple covariate model based on the example in Section 3.2. The results of these simulations were very similar to the one covariate case already included and were thus not reported here. A natural extension to this paper that we are currently pursuing is the random effects model to which considerations for coverage probability and confidence interval width via example and simulation will be duly considered.

Appendix A: Proof of Proposition 2.1

Let $\mathbf{X} \in \mathbb{R}^{K\times(p+1)}$ be the design matrix where the kth row of $\mathbf{X}$ is $\mathbf{x}_k$. Furthermore, let $\mathbf{y} = [y_1, \dots, y_K]^\top$ denote the observed transformed effects. Under the model in (5), we have that $Y_k \sim N(\mu_{\boldsymbol\beta}(\mathbf{x}_k), 1/N_k)$. Then a likelihood function for $\boldsymbol\beta$ with fixed $\mathbf{y}$ and $\mathbf{X}$ is, for $\mathbf{n} = [N_1, \dots, N_K]^\top$,

$$L(\boldsymbol\beta; \mathbf{y}, \mathbf{X}, \mathbf{n}) = \exp\left\{-\frac{1}{2}\sum_{k=1}^{K} N_k \left[y_k - \mu_{\boldsymbol\beta}(\mathbf{x}_k)\right]^2\right\}$$

and the corresponding log-likelihood function is

$$l(\boldsymbol\beta; \mathbf{y}, \mathbf{X}, \mathbf{n}) = \ln\ L(\boldsymbol\beta; \mathbf{y}, \mathbf{X}, \mathbf{n}) = -\frac{1}{2}\sum_{k=1}^{K} N_k \left[y_k - \mu_{\boldsymbol\beta}(\mathbf{x}_k)\right]^2. \tag{14}$$

Using the Chain Rule, the first derivative of (14) with respect to $\boldsymbol\beta$ is the vector

$$\frac{\partial}{\partial\boldsymbol\beta} l(\boldsymbol\beta; \mathbf{y}, \mathbf{X}, \mathbf{n}) = \sum_{k=1}^{K} N_k \left[y_k - \mu_{\boldsymbol\beta}(\mathbf{x}_k)\right] \frac{\partial}{\partial\boldsymbol\beta}\mu_{\boldsymbol\beta}(\mathbf{x}_k)$$

whose ith element is

$$\frac{\partial}{\partial\beta_i} l(\boldsymbol\beta; \mathbf{y}, \mathbf{X}, \mathbf{n}) = \sum_{k=1}^{K} N_k \left[y_k - \mu_{\boldsymbol\beta}(\mathbf{x}_k)\right] \frac{\partial}{\partial\beta_i}\mu_{\boldsymbol\beta}(\mathbf{x}_k).$$

It then follows that the Hessian matrix is equal to the $(p+1) \times (p+1)$ matrix

$$
\begin{aligned}
H\left[l(\boldsymbol{\beta};\mathbf{y},\mathbf{X},\mathbf{n})\right] &= \left[-\sum_{k=1}^{K} N_k \left\{ \frac{\partial \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)}{\partial \beta_i} . \frac{\partial \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)}{\partial \beta_j} - \left[y_k - \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)\right] \frac{\partial^2 \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)}{\partial \beta_i \partial \beta_j} \right\}\right]_{ij} \\
&= \left[-\sum_{k=1}^{K} N_k \left\{ \left(\frac{\partial \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)}{\partial \boldsymbol{\beta}^\top \mathbf{x}_k} \right)^2 x_{ki} x_{kj} \right.\right. \\
&\qquad \left.\left. - \left[y_k - \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)\right] \frac{\partial^2 \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)}{\partial (\boldsymbol{\beta}^\top \mathbf{x}_k)^2} x_{ki} x_{kj} \right\}\right]_{ij} \\
&= -\sum_{k=1}^{K} N_k \left\{ \left(\frac{\partial \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)}{\partial \boldsymbol{\beta}^\top \mathbf{x}_k} \right)^2 - \left[y_k - \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)\right] \frac{\partial^2 \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)}{\partial (\boldsymbol{\beta}^\top \mathbf{x}_k)^2} \right\} \mathbf{x}_k \mathbf{x}_k^\top \\
&= \mathbf{X}^\top \mathbf{D} \mathbf{X}
\end{aligned}
$$

where $\mathbf{D}$ is the diagonal matrix with kth diagonal element

$$
\mathbf{D}_{kk} = -N_k \left\{ \left(\frac{\partial \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)}{\partial \boldsymbol{\beta}^\top \mathbf{x}_k} \right)^2 - \left[y_k - \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)\right] \frac{\partial^2 \mu_{\boldsymbol{\beta}}(\mathbf{x}_k)}{\partial (\boldsymbol{\beta}^\top \mathbf{x}_k)^2} \right\}.
$$

Acknowledgements

The authors thank the editors and referees for their comments and suggestions which improved the content and presentation of this paper.

References

[1] Azorin, P.F. 1953. Sobre la distribución t no central i, ii. *Trabajos de Estadistica*, **4**, 173–198. MR0062989

[2] Baker, W.L., Michael White, C., Cappelleri, J.C., Kluger, J., & Coleman, C.I. 2009. Understanding heterogeneity in meta-analysis: the role of meta-regression. *The International Journal of Clinical Practice*, **63**(10), 1426–1434.

[3] Belanger, H.G., & Vanderploeg, R.D. 2005. The neuropsychological impact of sports-related concussion: A meta-analysis. *Journal of the International Neuropsychological Society*, **11**, 345–357.

[4] Gay, D.M. 1990. Usage summary for selected optimization routines. *Computing Science Technical Report*, **153**.

[5] Hardin, J.W., & Hilbe, J.M. 2007. *Generalized linear models and extensions*. Second edn. Stata Press.

[6] Hedges, L.V., & Olkin, I. 1985. *Statistical methods for meta-analysis.* Academic Press Inc. Ltd. MR0798597

[7] Hedges, L.V., Tipton, E., & Johnson, M.C. 2010. Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, **1**(1), 39–65.

[8] Hemming, K., Hutton, J.L., Maguire, M.G., & Marson, A.G. 2010. Meta-regression with partial information on summary trial or patient characteristics. *Statistics in Medicine*, **29**(12), 1312–1324.

[9] Jackson, D. 2008. The significance level of meta-regression's standard hypothesis test. *Communications in Statistics - Theory and Methods*, **37**, 1576–1590. MR2526473

[10] Knapp, G., & Hartung, J. 2003. Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, **22**(17), 2693–2710.

[11] Kulinskaya, E., Morgenthaler, S., & Staudte, R.G. 2008. *Meta analysis: A guide to calibrating and combining statistical evidence*. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons Ltd. MR2466371

[12] McCullagh, P., & Nelder, J.A. 1989. *Generalized linear models*. Second edn. Monographs on Statistics and Applied Probability. London: Chapman & Hall. MR0727836

[13] Montgomery, D.C., Peck, E.A., & Vining, G.G. 2006. *Introduction to linear regression analysis*. Fourth edn. Wiley-Interscience. MR1820113

[14] Roberts, C.J. 2005. Issues in meta-regression analysis: An overview. *Journal of Economic Surveys*, **19**(3), 295–298.

[15] Stanley, T.D., Doucouliagos, C., & Jarrell, S.B. 2008. Meta-regression analysis as the socio-economics of economics research. *Journal of Socio-Economics*, **37**(1), 276–292.

[16] Team, R Development Core. 2008. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

[17] Thompson, S.G., & Higgins, J.P.T. 2002. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, **21**, 1559–1573.

[18] Viechtbauer, W. 2007. Approximate confidence intervals for standardized effect sizes in the two-independent and two-dependent samples design. *Journal of Educational and Behavioral Statistics*, **32**(1), 39–60.