Generalised Confidence Intervals in Meta Regression

Friedrich, Thomas
Fakultät Statistik, Technische Universität Dortmund, 44221 Dortmund, Germany
friedrich@statistik.tu-dortmund.de

Knapp, Guido
Fakultät Statistik, Technische Universität Dortmund, 44221 Dortmund, Germany
knapp@statistik.tu-dortmund.de

Introduction

The explanation of heterogeneity that occurs when combining different studies is an important issue in meta analysis. Besides including a heterogeneity parameter in the analysis, it is also important to understand the possible causes of heterogeneity. One possibility is to incorporate study-specific covariates in the analysis that account for such between-trial variability. This yields the random effects meta regression model. We examine commonly used tests on the regression coefficients and propose a new method for constructing confidence intervals for the regression coefficients based on principles from generalised inference. The proposed method will be compared by simulation studies with respect to coverage probability and average length.

The random effects meta regression model

Before we start, let us fix some notation. Let $k \in \mathbb{N}$, $n = (n_1, \ldots, n_k)' \in \mathbb{N}^k$, $\tau \in \mathbb{R}_{>0}$, $\sigma = (\sigma_1, \ldots, \sigma_k)' \in \mathbb{R}^k_{>0}$ and $\delta = (\delta_1, \ldots, \delta_k)' \in \mathbb{R}^k_{>0}$. For any vector $c \in \mathbb{R}^k$, let $[c]$ denote the $k \times k$ matrix having c on the diagonal, in signs $\mathrm{diag}[c] = c$. If c is a scalar, i. e. $c \in \mathbb{R}$, then let $[c]_k$ denote the $k \times k$ matrix having c on each of its diagonal elements. Hence,

$$[\tau]_k = \begin{pmatrix} \tau & & \\ & \ddots & \\ & & \tau \end{pmatrix}, \qquad [\delta] = \begin{pmatrix} \delta_1 & & \\ & \ddots & \\ & & \delta_k \end{pmatrix}, \qquad ([\tau]_k + [\delta])^{-1} = \begin{pmatrix} \frac{1}{\tau+\delta_1} & & \\ & \ddots & \\ & & \frac{1}{\tau+\delta_k} \end{pmatrix}.$$

We want to perform a meta analysis on k studies. We denote the summary response of the jth study by Y_j. For example, one may think of Y_j as the mean response of n_j patients having participated in the jth study, in signs: $Y_j = \frac{1}{n_j}\sum_i Y_{ji}$ and Y_{ji} is the individual response of patient i in the jth study. Let θ_j denote the expected value and δ_j denote the variance of Y_j. The variance δ_j will usually be a function of the jth population variance σ_j and its sample size n_j. For example, in case of Y_j denoting a mean response, it is $\delta_j = \frac{\sigma_j}{n_j}$. Each study comes with an estimate D_j of δ_j. Let S_j denote the sum of squares of the jth study, in signs: $S_j = \sum_i (Y_{ji} - Y_j)^2$. If Y_j denotes the mean response, we may define $D_j := \frac{S_j}{n_j(n_j-1)}$, since $\mathbb{E}(S_j) = \sigma_j(n_j - 1)$.

Let $D = (D_1, \ldots, D_k)$, $S = (S_1, \ldots, S_k)$ and $Y = (Y_1, \ldots, Y_k)$. Consider the Gaußian-Gaußian hierarchical model ($j = 1, \ldots, k$):

$$(1) \quad Y_j|\theta_j \sim \mathfrak{N}(\theta_j, \delta_j),$$
$$(2) \quad \theta_j \sim \mathfrak{N}(x_j\beta, \tau),$$

where x_j denotes the jth row of a $k \times p$ matrix X with $\mathrm{rank}(X) = p < k - 1$ and $\beta \in \mathbb{R}^p$. The matrix X is referred to as the *design matrix* of the model. The vector β is a vector of parameters and called the vector of *regression coefficients*. The parameter τ stands for the variability between the studies. If we assume that $(Y, \theta_1, \ldots, \theta_k)$ has a multivariate Gaußian distribution, the marginal model for Y

yields the *random effects meta regression model*. In matrix notation,

$$Y \sim \mathfrak{N}_k\left(X\beta, [\tau]_k + [\delta]\right). \tag{3}$$

In particular, $Y_j \sim \mathfrak{N}\left(x_j\beta, \tau + \delta_j\right)$ for $j = 1, \ldots, k$. Since $\mathfrak{N}$ is infinitely divisible, we have

$$\frac{S_j}{\sigma_j} \sim \mathfrak{X}^2_{n_j-1}, \quad \text{for } j = 1, \ldots, k. \tag{4}$$

Here, $\mathfrak{X}^2_\nu$ denotes the distribution of a chi-square random variable with ν degrees of freedom. Let y be the observed value of Y, let s be the observed value of S and d the observed value of D. From (4) follows

$$D_j = \frac{S_j}{n_j(n_j-1)} = \frac{\sigma_j}{n_j} \cdot \frac{S_j}{\sigma_j(n_j-1)} = \frac{\delta_j}{K_j(n_j-1)}, \quad \text{where } K_j^{-1} \sim \mathfrak{X}^2_{n_{j-1}}. \tag{5}$$

In this paper, we will look at different constructions of confidence intervals for the regression coefficients β_i, $i = 1, \ldots, p$, in the presence of the nuisance parameters τ and δ. For notational brevity, define $\Omega_{\tau\delta} = \left([\tau]_k + [\delta]\right)^{-1}$. Then $\mathbb{V}(Y) = \Omega^{-1}_{\tau\delta}$ and $Y \sim \mathfrak{N}_k(X\beta, \Omega^{-1}_{\tau\delta})$. Let

$$\begin{aligned}
V_{\tau\delta} &= \left(X'\Omega_{\tau\delta}X\right)^{-1}, \\
B_{\tau\delta} &= V_{\tau\delta}X'\Omega_{\tau\delta} = \left(X'\Omega_{\tau\delta}X\right)^{-1}X'\Omega_{\tau\delta}, \\
H_{\tau\delta} &= XB_{\tau\delta} = X\left(X'\Omega_{\tau\delta}X\right)^{-1}X'\Omega_{\tau\delta}, \\
E_{\tau\delta} &= I_k - H_{\tau\delta} = I_k - X\left(X'\Omega_{\tau\delta}X\right)^{-1}X'\Omega_{\tau\delta}.
\end{aligned}$$

Then $V_{\tau\delta}$ is a $p \times p$ matrix, $B_{\tau\delta}$ is a $p \times k$ matrix and $H_{\tau\delta}$ and $E_{\tau\delta}$ are $k \times k$ matrices. In particular, $B_{\tau\delta}$, $H_{\tau\delta}$ and $E_{\tau\delta}$ are linear operators acting on $\mathbb{R}^k$, the image space of Y. For simplicity write: $V = V_{\tau\delta}$, $B = B_{\tau\delta}$, $H = H_{\tau\delta}$, $E = E_{\tau\delta}$ and $\Omega = \Omega_{\tau\delta}$ whenever τ, δ are equal to the *true* τ, δ in (3). Since $\mathbb{E}(BY) = \beta$ and $\mathbb{V}(BY) = V$, this yields an estimator of β, namely BY, with

$$BY \sim \mathfrak{N}_p\left(\beta, V\right).$$

The observed value of BY is By. In fact, if we assume that τ, δ were known, BY is the maximum likelihood estimator of β.

Now, let $b_i = (b_{i1}, \ldots, b_{ik})$ denote the ith row of B. Then $b_i y$ is an estimate of β_i, since $b_i Y \sim \mathfrak{N}\left(\beta_i, V_{ii}\right)$. Say, $\hat\tau$ and $\hat\delta$ are consistent estimators of τ and δ. Let $\hat V = V_{\hat\tau\hat\delta}$, $\hat B = B_{\hat\tau\hat\delta}$, $\hat H = H_{\hat\tau\hat\delta}$, $\hat E = E_{\hat\tau\hat\delta}$ and $\hat\Omega = \Omega_{\hat\tau\hat\delta}$. To keep notation consistent, we denote the $i$the row of $\hat B$ by $\hat b_i$. Since $\hat\tau$ and $\hat\delta$ are consistent, also $\hat b_i \xrightarrow{P} b_i$ and $\hat V_{ii} \xrightarrow{P} V_{ii}$ and, hence, $\frac{\hat b_i Y - \beta_i}{\sqrt{\hat V_{ii}}} \rightsquigarrow \mathfrak{N}(0,1)$, which yields the approximate $(1-\gamma)$-confidence interval for β_i and fixed $i = 1, \ldots, p$

$$\left[\hat b_i Y + z_{\frac{\gamma}{2}} \cdot \sqrt{\hat V_{ii}},\ \hat b_i Y + z_{1-\frac{\gamma}{2}} \cdot \sqrt{\hat V_{ii}}\right] \tag{6}$$

for any $\gamma \in (0, 1]$. Here, z_γ denotes the γ-quantile of the standard Gaußian distribution $\mathfrak{N}(0,1)$.

The confidence interval in (6) is based on approximations of the nuisance parameters τ and δ. If we can assume that D is consistent for δ, we simply use $\hat\delta := D$. In the case of τ, we are spoilt for choice. For an extensive list of possible estimators and a study of their performances see [5].

Generalised confidence intervals for the regression coefficients

The concepts of generalised inference were developed by Weerahandi, see [9]. His interest was in constructing *exact* confidence sets for all parameters of interest, in our case β, in the presence of

nuisance parameters, here τ and δ. To construct such exact confidence sets, it is necessary to find a (generalised) pivotal quantity for β which distribution is free of the nuisance parameters τ and δ. First, though, we will need to construct pivotal quantities for τ and δ.

A pivotal quantity for δ is quickly constructed. Recall that under model (3) the estimator D_j in each study is distributed via (5). Let

$$\tilde{D}_j := \delta_j \cdot \frac{d_j}{D_j} = \delta_j \cdot d_j \cdot \frac{K_j(n_j-1)}{\delta_j} = d_j(n_j-1)K_j, \quad \text{where } K_j^{-1} \sim \mathfrak{X}^2_{n_j-1}. \tag{7}$$

The distribution of $\tilde{D}_j$ is independent of unknowns. The observed value of $\tilde{D}_j$ is $\tilde{d}_j = \delta_j$ which is monotone in δ_j. Hence, $\tilde{D}_j$ is a pivotal quantity for δ_j.

Unfortunately, the construction of a pivotal quantity for τ is not as straight forward as for δ. The following construction is based on a quadratic form used to construct confidence intervals for τ, see [2, 4]. We follow an idea of [3] to extract a generalised pivotal quantity from this quadratic. Applying the linear operator E to the random vector Y yields the *residuals* EY of (3). In standard linear regression analysis $\|EY\|^2$ is frequently used as a measure of *model fit*. When dealing with heteroscedasticity, one works with a weighted quadratic instead:

$$Q_\delta(\tau) := (E_{\tau\delta}Y)'\Omega_{\tau\delta}(E_{\tau\delta}Y) = Y'\left(E'_{\tau\delta}\Omega_{\tau\delta}E_{\tau\delta}\right)Y = \left\|\mathbb{V}(E_{\tau\delta}Y)^{-\frac{1}{2}} \cdot E_{\tau\delta}Y\right\|^2. \tag{8}$$

Let q_δ denote the observed version of Q_δ. That is $q_\delta(\tau) = y'\left(E'_{\tau\delta}\Omega_{\tau\delta}E_{\tau\delta}\right)y$. As the above notation already suggests, we want to consider Q_δ as a function in τ. Recall that $\Omega_{\tau\delta}$ denotes a diagonal matrix. Note that for all $j = 1, \ldots, k$ each $(\operatorname{diag}\Omega_{\tau\delta})_j$ is strictly monotone decreasing in τ. Moreover, $\operatorname{diag}\Omega_{\tau\delta}$ converges uniformly to 0, in signs: $\lim_\tau \|\operatorname{diag}\Omega_{\tau\delta}\|_\infty = 0$. Hence, $Q_\delta(\tau)$ is strictly monotone decreasing in τ with $\lim_\tau Q_\delta(\tau) = 0$. This enables us to define an inverse function to Q_δ, namely

$$P_\delta(\eta) = \begin{cases} Q_\delta^{-1}(\eta) & : \quad 0 < \eta < Q_\delta(0) \\ 0 & : \quad \text{otherwise} \end{cases}, \qquad p_\delta(\eta) = \begin{cases} q_\delta^{-1}(\eta) & : \quad 0 < \eta < q_\delta(0) \\ 0 & : \quad \text{otherwise} \end{cases}$$

for $\eta > 0$. The observed value of $P_\delta(\eta)$ is $p_\delta(\eta)$. Let us now investigate the distribution of $Q_\delta(\tau)$. Since $E = E_{\tau\delta}$ is a linear operator, EY is multivariate Gaußian distributed with $\mathbb{E}(EY) = 0$ and $\mathbb{V}(EY) = \Omega^{-1} - X(X'\Omega X)^{-1}X'$. Hence, $\Omega \cdot \mathbb{V}(EY) = I_k - H$, which is idempotent with rank $k - p > 1$. Thus $Q_\delta(\tau) \sim \mathfrak{X}^2_{k-p}$. Define,

$$\tilde{T} := p_{\tilde{D}}\left(Q_\delta(\tau)\right) = p_{dK}(Q), \quad \text{where } K = (K_1, \ldots, K_p),\ K_j^{-1} \sim \mathfrak{X}^2_{n_j-1} \text{ and } Q \sim \mathfrak{X}^2_{k-p}.$$

We claim that $\tilde{T}$ is a generalised pivotal quantity of τ. The distribution of $\tilde{T}$ is free of unknowns and, since q_δ is monotone decreasing and $\tau > 0$, the observed value of $\tilde{T}$ is $\tilde{t} = p_\delta(q_\delta(\tau)) = q_\delta^{-1}(q_\delta(\tau)) = \tau$, which is monotone in τ.

Let $\tilde{V} = V_{\tilde{T}\tilde{D}}$, $\tilde{B} = B_{\tilde{T}\tilde{D}}$. Also, let $\tilde{b}_i$ denote the ith row of $\tilde{B}$. The observed values of $\tilde{V}, \tilde{B}$ and $\tilde{b}_i$ are V, B and b_i respectively. Note that $\frac{1}{\sqrt{V_{ii}}}(b_iY - \beta_i) \sim \mathfrak{N}(0,1)$. Consider

$$L_i := \tilde{b}_i y - \frac{b_iY - \beta_i}{\sqrt{V_{ii}}} \cdot \sqrt{\tilde{V}_{ii}}. = \tilde{b}_i y - N \cdot \sqrt{\left(X'\left([p_{dK}(Q)]_k + [dK]\right)^{-1}X\right)^{-1}_{ii}}, \tag{9}$$

where $K = (K_1, \ldots, K_k)$, $K_j^{-1} \sim \mathfrak{X}^2_{n_j-1}$, $Q \sim \mathfrak{X}^2_{k-p}$ and $N \sim \mathfrak{N}(0,1)$. Thus, the distribution of L_i only depends on the observed data and is free of the nuisance parameters τ and δ. The observed value of L_i is β_i which is monotone in β_i. Hence, L_i is a generalised pivotal quantity for β_i. Let $\mathfrak{L}_i$ denote the distribution of L_i and $\mathfrak{l}_{i,\gamma}$ its γ-quantile for any $\gamma \in (0,1]$. Then for any fixed $\gamma \in (0,1]$ and fixed $i = 1, \ldots, p$

$$[\mathfrak{l}_{i,\frac{\gamma}{2}}, \mathfrak{l}_{i,1-\frac{\gamma}{2}}] \tag{10}$$

is an exact $(1-\gamma)$-confidence interval for β_i.

The question remains how to obtain the necessary quantiles $\mathfrak{l}_{i,\gamma}$. For this, we generate random variables $K_1, \dots K_k, Q$ and N with $K_j^{-1} \sim \mathfrak{X}^2_{n_j-1}$ and $Q \sim \mathfrak{X}^2_{k-p}$ and $N \sim \mathfrak{N}(0, I)$ and define L_i such as in (9). Then L_i has the distribution $\mathfrak{L}_i$. Let $(L_{il})_{l\in\mathbb{N}}$ be a sequence of independent draws from $\mathfrak{L}_i$ and let $\mathfrak{l}_{i,\gamma,m}$ denote the empirical γ-quantile of $(L_{il})_{l\leq m}$. Then $\mathfrak{l}_{i,\gamma,m} \xrightarrow{P} \mathfrak{l}_{i,\gamma}$ for $m \to \infty$.

So, instead of obtaining approximate confidence intervals for β_i such as in (6), we obtain exact confidence interval by approximating the quantiles $\mathfrak{l}_{i,\gamma}$. By doing so, we have shifted the approximation problem from a statistical one – that depends on sample size – to a numerical one – that may be solved up to any required precision.

Performance study of the confidence intervals based on a real world example

We studied the performance of the above confidence intervals with respect to coverage probability and average length. For comparison, we turned to a data set that has already been discussed in the literature, e. g. in [1, 5]. This data set is indeed quite predesignated for such an analysis. It consists

Table 1: Data of 13 clinical trials of the Bacillus Calmette-Guérin vaccine efficacy.

Trial	Author	Year	Vaccinated		Not vaccinated		Absolute
			Disease	No disease	Disease	No disease	Latitude
A	Aronson	1948	4	119	11	128	44
B	Ferguson & Simes	1949	6	300	29	274	55
C	Rosenthal et al	1960	3	228	11	209	42
D	Hart & Sutherland	1977	62	13536	248	12619	52
E	Frimodt-Moller et al	1973	33	5036	47	5761	13
F	Stein & Aronson	1953	180	1361	372	1079	44
G	Vandiviere et al	1973	8	2537	10	619	19
H	TPT Madras	1980	505	87886	499	87892	13
I	Coetzee & Berjak	1968	29	7470	45	7232	27
J	Rosenthal et al	1961	17	1699	65	1600	42
K	Comstock et al	1974	186	50448	141	27197	18
L	Comstock & Webster	1969	5	2493	3	2338	33
M	Comstock et al	1976	27	16886	29	17825	33

of a combination of 13 clinical trials which evaluated the efficacy of the Bacillus Calmette-Guérin vaccine for the prevention of tuberculosis. The distance of a trial to the equator may serve as a potential influential covariate here. One can think of this distance as a surrogate for the presence of environmental mycrobacteria that provide a certain level of natural immunity against tuberculosis.

The data set is put together in Table 1. It can also be found in the `metafor` package of the statistical software environment `R`, see [7, 8]. Table 1 and Figure 1 show how unbalanced the data

Figure 1: Barplots of study sizes and subject assignments in each study. The *measure of unbalance* is $\frac{v-\neg v}{v+\neg v}$ where v is the number of vaccinated and $\neg v$ is the number of non-vaccinated subjects in a study.

are when it comes to study size. Study sizes range from a minimum of 262 in trial A to 176782 in trial H. But the clinical trials not only differ in total sample size. Within each study the proportion of vaccinated and non-vaccinated subjects is quite different. Whereas most trials are relatively balanced between the two groups, Figure 1 shows that in some trials the vaccinated group is considerably larger

than the non-vaccinated, see trials K and G for example. All this makes the data difficult to access for methods relying on good statistical approximations.

Following [1] and [5], we use the *logarithm of the relative risk* as outcome measure in the analysis. The model of choice is (3) with one covariate. The parameters of interest are the regression coefficients β_0, β_1 for the intercept and slope of the regression line respectively. In the analysis and in the simulation study we choose the restricted maximum-likelihood estimator $\hat{\tau}$ for τ, see [6]. Besides the above confidence intervals (6) and (10), we have also included an adjustment to (6) by Knapp and Hartung in our analysis, see [5]. We used the statistical software environment `R` for analysis and simulation study, see [7].

Table 2: Different 95% confidence intervals for β_0 and β_1 respectively. The exact confidence intervals are based on 1000 draws from $\mathfrak{L}_i$, $i = 1, 2$ respectively.

for β_0	lower bound	upper bound
Approximate	-0.24	0.74
Adjusted	-0.37	0.88
Exact	-0.77	0.98

for β_1	lower bound	upper bound
Approximate	-0.04	-0.01
Adjusted	-0.05	-0.01
Exact	-0.05	-0.00

Figure 2: Different 95% confidence intervals for β_0 and β_1 respectively. The dotted lines correspond to the point estimates $\hat{b}_0 y$ and $\hat{b}_1 y$ respectively. The exact confidence interval is based on 1000 draws from $\mathfrak{L}_i$, $i = 1, 2$.

Table 3: Performance of 95% confidence intervals for β_0, β_1 with respect to coverage probability and expected length based on 1000 random draws of (Y, D) via (3) and (5). The generalised confidence intervals for each (Y, D) were based on 120 draws from $\mathfrak{L}_i$, $i = 0, 1$ respectively.

β_0	Approximate	Adjusted	Exact
Length	0.8453	0.9386	1.3840
Covers	0.8720	0.8990	0.9640

β_1	Approximate	Adjusted	Exact
Length	0.0246	0.0273	0.0382
Covers	0.9020	0.9150	0.9580

Applying the methods to the data in Table 1 yield the results in Table 2 and Figure 2. The confidence intervals based on stochastic approximation are shorter than the exact one. However, as we shall see in the simulation study, these intervals are in fact underestimating the variability of $\hat{\beta}$ and the two confidence intervals based on (6) claim way more confidence than they can actually guarantee.

For the simulation, we fixed a set of parameters β and δ and generated responses Y and D from (3) and (5) respectively. We choose $\beta = \hat{\beta}$ and $\delta = \hat{\delta}$, where $\hat{\beta}$ and $\hat{\delta}$ correspond to the point

estimates based on the data set from Table 1. For each simulated pair (Y, D) we logged the lengths of each of the 95% confidence intervals and whether or not the interval was covering the *true* parameter β. The results of the simulation are put together in Table 3.

As we can see, the mean length of the approximate and the adjusted confidence intervals are smaller than the exact ones. However, the cost of their short mean length bears their coverage probability. The approximate confidence interval as defined in (6) is underestimating the true variability of $\hat{\beta}$, which leads to a short but poor performing confidence interval. The adjustment introduced in [5] accounts for the variability of $\hat{\tau}$ by introducing a correction factor to its estimated standard derivation. This makes the expected length of the confidence interval larger but brings its coverage probability closer to the requested significance bound. In both cases, though, the approximate confidence intervals do not fulfil the requested confidence bound. The exact confidence intervals, we have developed in this paper, are in fact the *only* ones to adhere the requested confidence bounds.

Conclusion

We have developed a new method for constructing confidence intervals for the regression coefficients in the random effects meta regression model. The method we have developed here is based on principles from generalised inference. In particular, the proposed intervals are exact. In simulation studies, we were able to show that the coverage probability of our proposed intervals not only meet predefined confidence bounds but also outperform classic approaches based on stochastic approximation.

References

[1] C. S. Berkey, D. C. Hoaglin, F. Mosteller, and C. A. Colditz, "A random-effects regression model for meta-analysis," *Statistics in Medicine*, vol. 14, pp. 395–411, 1995.

[2] J. Hartung and G. Knapp, "On confidence intervals for the among-group variance in the one-way random effects model with unequal error variances," *Journal of Statistical Planning and Inference*, vol. 127, pp. 157–177, 2005.

[3] H. K. Iyer, C. M. J. Wang, and T. Mathew, "Models and confidence intervals for true values in interlaboratory trials," *Journal of the American Statistical Association*, vol. 99, pp. 1060–1071, 2004.

[4] G. Knapp, B. Biggerstaff, and J. Hartung, "Assessing the amount of heterogeneity in random-effects meta-analysis," *Biometrical Journal*, vol. 48, pp. 271–285, 2006.

[5] G. Knapp and J. Hartung, "Improved tests for a random effects meta-regression with a single covariate," *Statistics in Medicine*, vol. 22, pp. 2693–2710, 2003.

[6] C. N. Morris, "Parametric empirical Bayes inference: Theory and applications (with discussion)," *Journal of the American Statistical Association*, vol. 78, pp. 47–65, 1983.

[7] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2010, ISBN 3-900051-07-0. [Online]. Available: http://www.R-project.org

[8] W. Viechtbauer, "Conducting meta-analyses in R with the metafor package," *Journal of Statistical Software*, vol. 36, pp. 1–48, 2010. [Online]. Available: http://www.jstatsoft.org/v36/i03/

[9] S. Weerahandi, *Exact Statistical Methods for Data Analysis*. New York: Springer, 1995.