Cluster-robust estimators for multivariate mixed-effects meta-regression

Computational Statistics and Data Analysis 179 (2023) 107631

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Cluster-robust estimators for multivariate mixed-effects meta-regression

Thilo Welz [a,*], Wolfgang Viechtbauer [b], Markus Pauly [a,c]

[a] TU Dortmund University, Vogelpothsweg 87, 44221, Dortmund, Germany
[b] Maastricht University, Vijverdalseweg 1, 6226 NB, Maastricht, the Netherlands
[c] UA Ruhr, Research Center Trustworthy Data Science and Security, 44227, Dortmund, Germany

ARTICLE INFO

Article history:
Received 22 April 2022
Received in revised form 1 October 2022
Accepted 2 October 2022
Available online 13 October 2022

Keywords:
Meta-regression
Multivariate analysis
Cluster-robust estimators
Mixed-effects models

ABSTRACT

Meta-analyses frequently include trials that report multiple outcomes based on a common set of study participants. These outcomes will generally be correlated. Cluster-robust variance-covariance estimators are a fruitful approach for synthesizing dependent outcomes. However, when the number of studies is small, state-of-the-art robust estimators can yield inflated Type 1 errors. Therefore, two new cluster-robust estimators are presented, in order to improve small sample performance. For both new estimators the idea is to transform the estimated variances of the residuals using only the diagonal entries of the hat matrix. The proposals are asymptotically equivalent to previously suggested cluster-robust estimators such as the bias reduced linearization approach. The methods are applied to a dataset of 81 trials examining overall and disease-free survival in neuroblastoma patients with amplified versus normal MYC-N genes. Furthermore, their performance is compared and contrasted in an extensive simulation study. The focus is on bivariate meta-regression, although the approaches can be applied more generally.

1. Introduction

In psychometric and medical research, studies frequently report multiple dependent outcomes. These outcomes can be synthesized across studies, while incorporating study level moderators, via multivariate meta-regression (Berkey et al., 1998; Mavridis and Salanti, 2013). This is a more sophisticated approach than averaging the effects within studies to create aggregate outcomes, which are then synthesized. A fruitful approach to achieve reliable inference in the case of a multivariate meta-regression is to use a cluster-robust (CR) variance-covariance estimator (Hedges et al., 2010). Robust estimators are designed to account for potential model misspecification. They can handle dependent effect size estimates and heteroscedastic model errors. A frequent problem in multivariate meta-analysis models is that it is difficult to compute the variance-covariance matrix of the vector of effect estimates. This is because trials (studies) frequently report neither the sampling covariances between study outcomes nor individual patient data (IPD). This is where CR estimators come into play: They have multiple advantages, such as providing consistent standard errors and asymptotically valid tests without requiring restrictive assumptions regarding the (correlation) structure of the model errors. One example for utilization of robust estimators and a working model for the correlation structure is a pseudo-likelihood approach for multivariate meta-analysis of test accuracy studies (Guolo and To, 2021).

* Corresponding author.
E-mail address: thilo.welz@tu-dortmund.de (T. Welz).

https://doi.org/10.1016/j.csda.2022.107631
0167-9473/

Cluster-robust estimators are an extension of heteroscedasticity consistent (HC) estimators. HC estimators, proposed by White (1980) and later extended in Cribari-Neto (2004) and Cribari-Neto et al. (2007), were first proposed in the meta-analytic literature by Sidik and Jonkman (2005). They have been examined and applied for use in ANCOVA (Zimmermann et al., 2019), ordinary least squares regression (Hayes and Cai, 2007) and mixed-effect meta-regression (Hedges et al., 2010; Viechtbauer et al., 2015; Welz and Pauly, 2020). When trials report multiple effects stemming from the same study participants, their clustered, i.e. correlated nature should be accounted for. This is where CR estimators come in. The original formulations of both HC and CR estimators have been shown to possess a downward bias for variance components, as well as yielding highly inflated Type 1 errors of respective test procedures in case of a small number of studies/clusters (Viechtbauer et al., 2015; Tipton and Pustejovsky, 2015; Welz and Pauly, 2020). Therefore it is recommended to instead use one of various improvements that have been suggested. We discuss some of these, such as the bias reduced linearization approach and CR_3 as introduced in Bell and McCaffrey (2002), as well as two new proposals in Section 3. These can be applied generally for multivariate meta-regression, but we focus specifically on the bivariate case. In the supplement we also provide some simulation results for the setting of three correlated effect sizes.

First, we present the statistical model, as well as tests and confidence regions for the model coefficients in Section 2. In Section 3, we describe multiple CR estimators, including two new suggestions CR_3^* and CR_4^*. In Section 4, we conduct a real world data analysis. Section 5 describes the design and results of our simulation study. We close with a discussion of the results and an outlook for future research (Section 6).

2. The set-up

The usual multivariate mixed-effects meta-regression model (Jackson et al., 2011; Mavridis and Salanti, 2013) is given by

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\beta + \boldsymbol{u}_i + \boldsymbol{\varepsilon}_i,\ i = 1, \dots, k, \tag{1}$$

where k is the number of independent studies, $\boldsymbol{\beta} \in \mathbb{R}^q$ is a vector of coefficients and $\boldsymbol{X}_i$ a $p_i \times q$ design matrix of study-level covariates. In the following we will assume that there are p effects (outcomes) of interest per study, but only $p_i \leq p$ effects are observed (reported) in study i, i.e. $\boldsymbol{Y}_i \in \mathbb{R}^{p_i}$. Furthermore, $\boldsymbol{u}_i$ is a random effect that is typically assumed to be multivariate normally distributed with $\boldsymbol{u}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{T}_i)$ and $\boldsymbol{\varepsilon}_i$ is the *within-study* error with $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{V}_i)$. With $\boldsymbol{T}_i$ we refer to the $p_i \times p_i$ submatrix of the matrix $\boldsymbol{T}$ denoting the $p \times p$ *between*-study variance-covariance matrix (under complete data). $\boldsymbol{V}_i$ refers to the corresponding $p_i \times p_i$ *within*-study variance-covariance matrix. In practice the sampling variance-covariance matrices $\boldsymbol{V}_i$ can be difficult to construct, as studies rarely report covariances of effect size measures and information needed to compute them is often unavailable. This would require the formulation of a working model (Hedges et al., 2010). In our simulation study, presented in Section 5, we utilize estimates of the sampling covariances. We also present simulation results based on a working model in the supplementary materials. We rewrite model (1) in matrix notation as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} + \boldsymbol{\varepsilon}, \tag{2}$$

with $\boldsymbol{\beta} \in \mathbb{R}^q$, $\boldsymbol{Y} = (\boldsymbol{Y}_1', \dots, \boldsymbol{Y}_k')'$, and design matrix $\mathbf{X}$. Assuming that we have a block diagonal matrix of weights $\widehat{\boldsymbol{W}} = \operatorname{diag}(\widehat{\boldsymbol{W}}_1, \dots, \widehat{\boldsymbol{W}}_k)$, usually corresponding to the inverse variance weights with $\widehat{\boldsymbol{W}}_{\boldsymbol{i}} = \left(\widehat{\boldsymbol{T}}_{\boldsymbol{i}} + \boldsymbol{V}_{\boldsymbol{i}}\right)^{-1}$, then the weighted least squares estimator for $\boldsymbol{\beta}$ is given by (Mavridis and Salanti, 2013)

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{W}}\boldsymbol{Y}. \tag{3}$$

With $\widehat{\boldsymbol{T}}_{\boldsymbol{i}}$ we refer to an estimate of the heterogeneity variance matrix $\boldsymbol{T}_{\boldsymbol{i}}$. In our implementations we use restricted maximum likelihood estimation, as is the default in the `rma.mv` function of the `metafor R` package.

We will focus on constructing (multivariate) confidence regions for $\boldsymbol{\beta}$ and confidence intervals for the individual coefficients $\beta_j,\ j = 1, \dots, q$ based on testing the hypotheses $H_0 : \{\boldsymbol{\beta} = \boldsymbol{\beta}_0\}$ vs. $H_1 : \{\boldsymbol{\beta} \neq \boldsymbol{\beta}_0\}$. We set $\boldsymbol{\Sigma} = \operatorname{Cov}(\hat{\boldsymbol{\beta}})$ and denote estimates thereof by $\widehat{\boldsymbol{\Sigma}}$. We discuss specific choices for estimating $\boldsymbol{\Sigma}$ in Section 3.

Neglecting multiplicity, we note that a commonly used confidence interval for $\beta_j,\ j = 1, \dots, q$ is given by

$$\hat{\beta}_j \pm \sqrt{\widehat{\boldsymbol{\Sigma}}_{jj}}\, z_{1-\alpha/2}. \tag{4}$$

Here $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution and $\widehat{\boldsymbol{\Sigma}}_{jj}$ denotes the j^{th} diagonal element of $\widehat{\boldsymbol{\Sigma}} = (\boldsymbol{X}'\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}$. A confidence interval with better small sample performance that is asymptotically equivalent for $k \to \infty$ is given by using the $t_{p(k)-q,1-\alpha/2}$ quantile instead, which refers to the $1-\alpha/2$ quantile of the t-distribution with $p(k)-q$ degrees of freedom. Here $p(k) := \sum_{i=1}^{k} p_i$ is the total number of observed effects, which is equal to the number of studies k in the univariate setting (Viechtbauer et al., 2015). Alternatively the degrees of freedom of the t distribution can be estimated via a Satterthwaite approximation, as suggested by Bell and McCaffrey (2002).

In order to construct a $(1-\alpha)$ confidence region for $\boldsymbol{\beta}$ we consider the usual Wald-type test-statistic (Tipton and Pustejovsky, 2015)

$$Q = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \widehat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0). \tag{5}$$

Under the null hypothesis Q is approximately χ^2_q-distributed, assuming $\boldsymbol{\Sigma}$ is positive definite. However, it is known that tests based on this approximation can perform poorly for small to moderate values of k (Tipton and Pustejovsky, 2015). An arguably better alternative is the F-test

$$\mathbb{1}\left\{Q > qF_{q,k-q,1-\alpha}\right\}, \tag{6}$$

where $F_{q,k-q,1-\alpha}$ denotes the $1-\alpha$ quantile of an F-distribution with q and $k-q$ degrees of freedom. This is analogous to the t-tests for univariate coefficients and is superior to the test based on the asymptotic χ^2-approximation (Tipton and Pustejovsky, 2015). However, the F-test has been criticized for only performing well in certain scenarios (Tipton, 2015). As a remedy for smaller k, Tipton and Pustejovsky (2015) proposed to approximate Q by a Hotelling's T^2 distribution with parameters q and (degrees of freedom) η, such that

$$\frac{\eta - q + 1}{\eta q} Q \sim F(q, \eta - q + 1). \tag{7}$$

They discuss different approaches for estimating the degrees of freedom η. Based on their research, they recommend an estimation approach, which they call "HTZ". We briefly summarize this estimator, originally proposed by Zhang (2012) for heteroscedastic one-way MANOVA, and refer to their paper for details.

First note that the statistic in (5) can also be written as $Q = \boldsymbol{z}'\boldsymbol{S}^{-1}\boldsymbol{z}$ with $\boldsymbol{z} = \boldsymbol{\Sigma}^{-1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ and $\boldsymbol{S} = \boldsymbol{\Sigma}^{-1/2}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1/2}$. Under H_0, $\boldsymbol{z}$ is normally distributed with mean $\boldsymbol{0}$ and covariance $\boldsymbol{I}$ (Tipton and Pustejovsky, 2015). Moreover, if $\boldsymbol{S}$ is a random $q \times q$ matrix such that $\eta\boldsymbol{S}$ follows a Wishart distribution with η degrees of freedom and scale matrix $\boldsymbol{I}_q$, the estimator is given by

$$\hat{\eta}_Z = \frac{q(q+1)}{\sum_{a=1}^{q}\sum_{b=1}^{q} \mathrm{Var}(s_{ab})}.$$

Here s_{ab} denotes the entry (a, b) of $\boldsymbol{S}$. This approach corresponds to setting the total variation in $\boldsymbol{S}$ equal to the total variation in a Wishart distribution (Tipton and Pustejovsky, 2015).

However, our own simulations showed that there are situations when $\hat{\eta}_Z < q-1$ and therefore $\hat{\eta}_Z - q + 1 < 0$. Specifically this frequently happened in cases with a small number of studies ($k \leq 5$). As the degrees of freedom in an F distribution cannot be negative the HTZ approach is not applicable here. Therefore we will stick to the classical F-test (6), although we propose a small sample adjustment. In our simulations the F-test (6) leads to very liberal or conservative results, depending on the variance-covariance estimator used, in settings with $k = 5$ studies. We therefore propose to truncate the denominator degrees of freedom at the value two, i.e. we consider the F-test

$$\mathbb{1}\left\{Q > qF_{q,\max(2,k-q),1-\alpha}\right\}. \tag{8}$$

The simple motivation behind this adjustment is that for an $F_{m,n}$ distribution with degrees of freedom m and n the expected value $\frac{n}{n-2}$ only exists when $n > 2$. We also tested a truncation of the denominator degrees of freedom at three. However, simulations indicate superior coverage of respective confidence intervals for a truncation at two.

Confidence regions for $\boldsymbol{\beta}$ can be derived via test inversion. For example, if (8) is a test for $H_0 : \{\boldsymbol{\beta} = \boldsymbol{\beta_0}\}$ vs. $H_1 : \{\boldsymbol{\beta} \neq \boldsymbol{\beta_0}\}$, then the set

$$\Lambda := \left\{\boldsymbol{\beta} \in \mathbb{R}^q : (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \widehat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq qF_{q,\max(2,k-q),1-\alpha}\right\} \tag{9}$$

is a corresponding confidence region for $\boldsymbol{\beta}$.

A confidence ellipsoid can be obtained following Johnson et al. (2014), based on the eigenvalues $\hat{\lambda}_j$ and eigenvectors $\hat{\boldsymbol{e}}_j$ of $\widehat{\boldsymbol{\Sigma}}$. This means Λ is an ellipsoid centered around $\hat{\boldsymbol{\beta}}$, whose axes are given by

$$\hat{\boldsymbol{\beta}} \pm \sqrt{\hat{\lambda}_j q F_{q,\max(2,k-q),1-\alpha}}\,\hat{\boldsymbol{e}}_j, \quad j = 1, \ldots, q.$$

This means Λ extends for $\sqrt{\hat{\lambda}_j q F_{q,\max(2,k-q),1-\alpha}}$ units along the estimated eigenvector $\hat{\boldsymbol{e}}_j$ for $j = 1, \ldots, q$. Since the volume of an n-dimensional ellipsoid with axis lengths $a_1, \ldots, a_n$ is given by (Wilson, 2010)

$$V = \frac{2\pi^{n/2}}{n\Gamma(n/2)} \prod_{i=1}^{n} a_i,$$

the volume of the confidence ellipsoid Λ is equal to

$$V_\Lambda = \frac{2\pi^{q/2}}{q\Gamma(q/2)} \prod_{i=1}^{q} \sqrt{\hat{\lambda}_i q F_{q,\max(2,k-q),1-\alpha}}.$$

We consider these multivariate confidence regions Λ in our simulation study in Section 5, as this reduces the number of simulation results and yields concise evaluation. In practice, researchers would more likely consider individual confidence intervals. However, if the confidence set Λ yields poor coverage for $\boldsymbol{\beta}$, then the individual confidence intervals for the components of $\boldsymbol{\beta}$ cannot all perform well.

3. Cluster-robust covariance estimators

Robust variance-covariance estimators, also known as sandwich estimators or Huber-White estimators, have been recommended as a promising alternative in the context of meta-regression (Hedges et al., 2010; Tipton, 2015; Welz and Pauly, 2020). Robust estimators are designed to account for potential model misspecification. They have many desirable properties, such as consistency under heteroscedasticity or asymptotic normality (Hedges et al., 2010) without making restrictive assumptions about the specific form of the effect sizes' sampling distributions.

The reliability of confidence regions based on the statistic (5) depends on the quality of the estimator $\widehat{\boldsymbol{\Sigma}}$ for $\boldsymbol{\Sigma} = \text{Cov}(\hat{\boldsymbol{\beta}})$. The standard (Wald-type) estimator, which we will refer to as ST, is given by $(\boldsymbol{X'\widehat{W}X})^{-1}$. The motivation behind this estimator is that the true covariance matrix of $\widehat{\boldsymbol{\beta}}$ (given correct weights) is equal to $\boldsymbol{\Sigma} = (\boldsymbol{X'WX})^{-1}$ with $\boldsymbol{W} = \text{diag}(\boldsymbol{W_1}, \ldots, \boldsymbol{W_K})$ and $\boldsymbol{W_i} = \boldsymbol{T_i} + \boldsymbol{V_i}$. However, this ignores the imprecision in the estimation of $\boldsymbol{T}, \boldsymbol{V}$ and therefore in the estimation of $\mathbf{W}$. In fact, if $\boldsymbol{T}$ is estimated poorly, this may lead to deviations from nominal Type 1 error and coverage of corresponding confidence regions (Sidik and Jonkman, 2005).

In the case of univariate meta-analysis and meta-regression heteroscedasticity-consistent (HC) estimators can be applied (Sidik and Jonkman, 2005; Viechtbauer et al., 2015; Welz and Pauly, 2020). There are various HC estimators that can be used for ordinary least squares regression, see Cribari-Neto et al. (2007). For multivariate meta-regression however, the correlated nature of the study effects needs to be taken into account. We therefore consider cluster-robust (CR) estimators. A selection of CR estimators is, e.g., implemented in the `R` package `clubSandwich` (Pustejovsky, 2021). The package recommendation is the "bias reduced linearization" approach CR_2, which is discussed in detail in Tipton and Pustejovsky (2015); Pustejovsky and Tipton (2018). Sandwich estimators (of HC- as well as CR-type) are all of the general form

$$\widehat{\boldsymbol{\Sigma}} = (\boldsymbol{X'\widehat{W}X})^{-1}\boldsymbol{X'\widehat{W}\widehat{\Omega}\widehat{W}X}(\boldsymbol{X'\widehat{W}X})^{-1}, \tag{10}$$

with the differences lying in the central "meat" matrix $\widehat{\boldsymbol{\Omega}}$, surrounded by the "bread". This form motivates the name "sandwich" estimator. $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{HC_1}} = \frac{k}{k-q}(\boldsymbol{X'\widehat{W}X})^{-1}\left(\sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \boldsymbol{X_i'\widehat{W}_i\hat{e}_i^2\widehat{W}_iX_i}\right)(\boldsymbol{X'\widehat{W}X})^{-1}$ is arguably the best known sandwich estimator in the context of univariate meta-regression (Hedges et al., 2010; Viechtbauer et al., 2015; Tipton and Pustejovsky, 2015). However, the extensions HC_3 and HC_4 are frequently recommended as superior alternatives in the non meta-analytic literature, see Cribari-Neto et al. (2007) for details, and have been shown to be superior to HC_1 (Long and Ervin, 2000; Hayes and Cai, 2007; Zimmermann et al., 2019). A natural extension of HC_1 for the multivariate setting and what we will refer to as CR_1^* is defined as

$$\widehat{\boldsymbol{\Sigma}}_{CR_1^*} = \frac{k}{k-q}(\boldsymbol{X'\widehat{W}X})^{-1}\left(\sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \boldsymbol{X_i'\widehat{W}_i\widehat{\Omega}_i\widehat{W}_iX_i}\right)(\boldsymbol{X'\widehat{W}X})^{-1}, \tag{11}$$

where $\widehat{\boldsymbol{\Omega}}_i = \boldsymbol{E_iE_i'}$ with $\boldsymbol{E_i} = \boldsymbol{Y_i - X_i\hat{\beta}}$ and $\frac{k}{k-q}$ is a correction factor that converges to 1 as k goes to infinity. The motivation for this factor is to correct for a liberal behavior in case of few studies/clusters k; see the `clubSandwich` package for similar choices.

However, as our simulation study below will show, tests based on CR_1^* are still quite liberal when k is small. An alternative is to instead use a bias reduced linearization approach, which was originally proposed by Bell and McCaffrey (2002) and further developed by Pustejovsky and Tipton (2018). This estimator, called CR_2, is designed to be exactly unbiased under the correct specification of a working model. This is achieved via a clever choice of adjustment matrices in the formulation of the estimator, see Tipton and Pustejovsky (2015); Pustejovsky and Tipton (2018) for details. This is the recommended approach in the `clubSandwich` package (Pustejovsky, 2021). Another alternative is the CR_3 estimator, which is a close approximation of the leave-one-(cluster)-out Jackknife variance-covariance estimator. CR_3 is also implemented in the `clubSandwich` package.

However, all of the estimators above can be unsatisfactory for small k, as our simulations will show. Therefore, in addition to these CR-estimators, we propose two others, which are extensions of the HC_3 and HC_4 estimators. Since HC_3 and HC_4 often outperform both HC_1 and HC_2 in the univariate regression setting (Long and Ervin, 2000; Cribari-Neto, 2004; Welz and Pauly, 2020), one would suspect their respective cluster-robust extensions to outperform in the case of multivariate regression. We therefore define CR_3^* and CR_4^* via

$$\widehat{\boldsymbol{\Sigma}}_{CR_3^*} = (\boldsymbol{X'\widehat{W}X})^{-1}\left(\sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \boldsymbol{X_i'\widehat{W}_i\widehat{\Omega}_{3i}\widehat{W}_iX_i}\right)(\boldsymbol{X'\widehat{W}X})^{-1}, \tag{12}$$

$$\widehat{\boldsymbol{\Sigma}}_{CR_4^*} = (\boldsymbol{X}'\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}\left(\sum_{i=1}^{k}\boldsymbol{X}_i'\widehat{\boldsymbol{W}}_i\widehat{\boldsymbol{\Omega}}_{4i}\widehat{\boldsymbol{W}}_i\boldsymbol{X}_i\right)(\boldsymbol{X}'\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}. \tag{13}$$

Here $\widehat{\boldsymbol{\Omega}}_{3i}$ is defined as

$$\widehat{\boldsymbol{\Omega}}_{3i} = \widehat{\boldsymbol{\Omega}}_i - \Delta + \Delta\cdot\left(\boldsymbol{I}_{p_i} - \text{diag}(\boldsymbol{H}_i)\right)^{-2}, \tag{14}$$

where $\boldsymbol{H}_i$ refers to the submatrix of $\boldsymbol{H}$ with entries pertaining to study i, p_i is the number of observed effects in study i and $\Delta = \text{diag}(\boldsymbol{E}_i\boldsymbol{E}_i')$. $\boldsymbol{H}$ refers to the hat matrix $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{W}}$. Furthermore, $\widehat{\boldsymbol{\Omega}}_{4i}$ is equal to (14) except Δ is multiplied with $\left(\boldsymbol{I}_{p_i} - \text{diag}(\boldsymbol{H}_i)\right)^{-\delta_i}$, where $\delta_i = \min\left\{4, h_{ii}/\bar{h}\right\}$ with h_{ii} denoting the i-th diagonal element of $\boldsymbol{H}$ and $\bar{h}$ is the average of the values in the diagonal of the hat matrix. This data-dependent exponent stems from the HC_4 suggestion by Cribari-Neto (2004). HC_4 performs well in univariate meta-regression (Welz and Pauly, 2020) and therefore motivates an extension to the cluster-robust context.

We highlight that our proposed estimator CR_3^* is different from the estimator CR_3 implemented in the `R` package `clubSandwich` as proposed by Bell and McCaffrey (2002). Whereas the latter uses the entire hat matrix for each cluster, we propose to use just the diagonal elements. In contrast, the "meat" matrix for CR_3 is given by $\sum_{i=1}^{K}\boldsymbol{X}_i'\widehat{\boldsymbol{W}}_i(\boldsymbol{I}-\boldsymbol{H}_i)^{-1}\widehat{\boldsymbol{\Omega}}_i(\boldsymbol{I}-\boldsymbol{H}_i)^{-1}\widehat{\boldsymbol{W}}_i\boldsymbol{X}_i$. Furthermore note that CR_3^* is not even equal to the estimator with meat matrix given by

$$\sum_{i=1}^{K}\boldsymbol{X}_i'\widehat{\boldsymbol{W}}_i(\boldsymbol{I}-\text{diag}(\boldsymbol{H}_i))^{-1}\widehat{\boldsymbol{\Omega}}_i(\boldsymbol{I}-\text{diag}(\boldsymbol{H}_i))^{-1}\widehat{\boldsymbol{W}}_i\boldsymbol{X}_i$$

because $\widehat{\boldsymbol{\Omega}}_i$ is in general not a diagonal matrix (only block-diagonal), due to the clustered nature of the data.

For univariate regression we were able to prove the asymptotic equivalence of all HC estimators, which is formulated in the supplement of Welz and Pauly (2020). Under some weak regularity conditions it follows that the leverages asymptotically converge to zero, as the number of studies k goes to infinity. Therefore, we expected similar results to hold for CR estimators with analogous arguments. A theorem regarding the asymptotic equivalence of CR estimators under regularity conditions is given in the supplement of this paper, along with a proof.

4. Data analysis

We exemplify the methods presented in this manuscript with the analysis of a dataset containing 81 trials examining overall (OS) and/or disease-free survival (DFS) in neuroblastoma patients with amplified (extra copies) versus normal MYC-N genes. The dataset is contained in the `R` package `metafor` under the name `dat.riley2003` and was extracted from Riley (2011). The data were also analyzed in Riley et al. (2004, 2007). Amplified MYC-N levels are associated with poorer outcomes. The effect measures are log hazard ratios with positive values indicating an increased risk of death or relapse/death for patients with higher MYC-N levels as compared to patients with lower levels. 17 studies reported both outcomes, 25 studies only reported DFS and 39 studies only reported OS.

The dataset contains the log hazard ratios and the corresponding sampling variances. However, since no information is available on the sampling covariances between OS and DFS we must make some assumptions with regard to our working model. Emura et al. (2021) estimated a Kendall's tau of 0.9 between DFS and OS for a sample of 14 randomized clinical trials on resectable gastric cancers. Although using Kendall's rank correlation coefficient is more common for meta-analyses of the survival endpoints DFS and OS (Burzykowski et al., 2001), we stick to the Pearson correlation coefficient (in alignment with our simulation study). We thus formulate a working model with a correlation of 0.9 between DFS and OS. In the spirit of a sensitivity analysis we will also assume a weaker correlation of $\varrho_1 = 0.5$ in addition to the stronger correlation of $\varrho_2 = 0.9$ and then compare the results. This means for a hypothetical study i that reports log hazard ratios for OS and DFS, $y_{i,OS}$ and $y_{i,DFS}$, with an assumed correlation of 0.5 along with respective sampling variances $\sigma^2_{i,OS}$ and $\sigma^2_{i,DFS}$, we have the sampling variance-covariance matrix $V_i = \begin{pmatrix}\sigma^2_{i,OS} & 0.5\cdot\sigma_{i,OS}\sigma_{i,DFS}\\ 0.5\cdot\sigma_{i,OS}\sigma_{i,DFS} & \sigma^2_{i,DFS}\end{pmatrix}$.

We assume a multivariate meta-regression model that includes a random effect as in Section 2 as well as an unstructured (but positive definite) variance-covariance matrix. In the following we are interested in testing whether both pooled effects are different from zero. When the full dataset is analyzed, the Wald-test for $H_0 : \{\boldsymbol{\beta} = \boldsymbol{0}\}$ vs. $H_1 : \{\boldsymbol{\beta} \neq \boldsymbol{0}\}$ returns a p-value < 0.001 for all CR estimators and for both ϱ_1 and ϱ_2.

However, in order to demonstrate differences in small sample behavior, let us assume we only had data from a subset of 5 studies. Such a situation is not unrealistic, considering the median number of studies per meta-analysis in a sample of 22,453 published meta-analyses from the Cochrane Database was three (Davey et al., 2011). We assume this subset contains at least two results for both OS and DFS. An example subset is shown in Table 1. The p-values and volumes of confidence ellipsoids (9) at level 95% for the estimators CR_1^*, CR_3^*, CR_4^*, CR_2 and ST for assumed correlations ϱ_1, ϱ_2 are displayed in Table 2.

Table 1
Sample of five studies containing log hazard ratios (y_i) for disease-free and overall survival and their respective sampling variances (v_i).

study	y_i	v_i	outcome
1	−0.11	0.45	DFS
1	−0.14	0.66	OS
2	0.30	0.07	DFS
2	0.67	0.08	OS
3	0.41	0.77	DFS
3	0.43	0.66	OS
4	0.47	0.29	DFS
4	2.08	0.45	OS
5	0.76	0.24	DFS
5	0.70	0.31	OS

Table 2
p-values and volumes of 95% confidence ellipsoids for Wald-tests based on CR estimators and the standard variance-covariance estimator $(\boldsymbol{X'\widehat{W}X})^{-1}$ for assumed correlations of $\varrho_1 = 0.5$ and $\varrho_2 = 0.9$.

Estimators	p-values		Ellipsoid Volume	
	ϱ_1	ϱ_2	ϱ_1	ϱ_2
CR1*	0.073	0.075	0.870	1.103
CR3*	0.069	0.077	0.924	1.178
CR4*	0.076	0.090	0.995	1.114
CR2	0.054	0.055	0.662	0.854
ST	0.138	0.206	1.459	1.797

The results show that when the number of studies is small the p-values can vary substantially, depending on the choice of estimator. Furthermore, the results based on CR estimators appear to be more stable and depend much less on the underlying $\boldsymbol{V}$ matrix i.e. the assumed correlation between OS and DFS than the standard estimator $(\boldsymbol{X'\widehat{W}X})^{-1}$. Additionally, the CR estimators yield smaller confidence regions. This motivates the use of a CR approach over the standard variance-covariance estimator.

5. Simulation study

5.1. Simulation design

In order to assess the performance of the previously discussed methods, we conducted a Monte Carlo simulation. We considered $k \in \{5, 10, 20, 40\}$ studies, average study sizes $N \in \{40, 100\}$ with balanced treatment and control groups, coefficient vectors $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)' \in \{(0,0,0,0)', (0.2,0.2,0.1,0.1)', (0.4,0.4,0.2,0.3)'\}$, correlations $\varrho \in \{0, 0.3, 0.7\}$ and missing data ratios from $\{0, 0.1, 0.2, 0.3, 0.4\}$. The latter refers to the number of studies that only report one of the two effects of interest and ϱ refers to the IPD correlations between the two observed outcomes. In the coefficient vector $\boldsymbol{\beta}$ the first two entries refer to the population means of the two effects of interest and the other two represent the effect of the study-level moderator on each effect respectively. Study sizes were varied, such that for an average study size N, 20% of studies had size $0.8N, 0.9N, \ldots, 1.2N$ respectively. Datasets with missing data were generated by first simulating complete data and then removing entries completely at random.

The simulated study-level effects are (correlated) standardized mean differences (SMD). We estimated these SMDs via the adjusted Hedges' g (Hedges, 1981)

$$g := \frac{\Gamma(m/2)}{\sqrt{(m/2)}\Gamma((m-1)/2)} d$$

with $m = n_T + n_C - 2$ and where n_T and n_C refer to the treatment and control group sizes. Hedges' g is defined as $d = (\bar{x}_T - \bar{x}_C)/s^*$, with a pooled standard deviation $s^* = \sqrt{\frac{(n_T-1)s_T^2 + (n_C-1)s_C^2}{m}}$, where s_T^2, s_C^2 refer to the variances in the treatment and control groups respectively (Hedges, 1981). This adjustment to Hedges' g yields an unbiased effect estimator (Lin and Aloe, 2021). We generated the SMDs by first simulating individual participant data (IPD). The treatment and control group IPD observations Y_{ij}^T and Y_{ij}^C were drawn from bivariate normal distributions respectively. More precisely, for study $i = 1, \ldots, k$ and participant $j = 1, \ldots, N_i/2$ the observations are drawn from $Y_{ij}^T \sim \mathcal{N}(\theta_i, P)$ and $Y_{ij}^C \sim \mathcal{N}(\boldsymbol{0}, P)$ with $\theta_i = \boldsymbol{X\beta + u_i}$ and $P = \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix}$ is the population correlation matrix of the outcomes in study i. $\boldsymbol{X}$ is a $2 \times q$ design matrix of covariates. In

our specific simulation design of a single study-level covariate x with potentially different influence on the two study effects we have $\boldsymbol{X} = \begin{pmatrix} 1 & 0 & x & 0 \\ 0 & 1 & 0 & x \end{pmatrix}$.

For the heterogeneity matrix $\boldsymbol{T}$ we consider the two settings

$$\begin{pmatrix} \tau^2 & 0.2\tau^2 \\ 0.2\tau^2 & \tau^2 \end{pmatrix} \text{and} \begin{pmatrix} \tau^2 & 0.4\tau^2 \\ 0.4\tau^2 & 2\tau^2 \end{pmatrix}.$$

For $M = N/2$ (average size of the treatment and control groups), we set $\tau^2 := \frac{2}{M} + \frac{\beta_0^2}{4M} = \frac{4}{N} + \frac{\beta_0^2}{2N}$, which is approximately equal to the sampling variance of the standardized mean difference (Borenstein et al., 2021). This corresponds to an I^2 value of 0.5. Here, I^2 refers to the percentage of the total variation across studies that is due to heterogeneity rather than sampling variation (Higgins and Thompson, 2002). We estimate $\boldsymbol{T}$ via restricted maximum likelihood (REML) estimation, which is the default setting in the `rma.mv` function of the `metafor R` package.

We briefly discuss the covariance between two SMDs in the setting where we have a single treatment and control group but with different outcome measures. The resulting effect sizes will be correlated because the outcomes are collected from the same study participants. Olkin and Gleser (2009) showed that a large sample estimate for the covariance between two SMDs d_1 and d_2 with estimated (raw data) correlation $\hat{\varrho}$ is given by

$$\widehat{\mathrm{Cov}}(d_1, d_2) = \hat{\varrho}\left(\tfrac{1}{n_T} + \tfrac{1}{n_C}\right) + \frac{\hat{\varrho}^2 d_1 d_2}{m}. \tag{15}$$

Thus we obtain

$$\widehat{\mathrm{Cov}}(g_1, g_2) = \left(\frac{\Gamma(m/2)}{\sqrt{(m/2)}\Gamma((m-1)/2)}\right)^2 \left(\hat{\varrho}\left(\frac{1}{n_T} + \frac{1}{n_C}\right) + \frac{\hat{\varrho}^2 d_1 d_2}{m}\right). \tag{16}$$

All results are based on a nominal significance level $\alpha = 0.05$. For each scenario we performed $N = 5000$ simulation runs. The primary focus was on comparing empirical coverage of the confidence regions (9) with nominal coverage being $1 - \alpha = 0.95$. For 5000 iterations, the Monte Carlo standard error of the simulated coverage will be approximately $\sqrt{\frac{0.95 \times 0.05}{5000}} \approx 0.31\%$ and assuming a power of 80% the Monte Carlo standard error of the simulated power will be approximately $\sqrt{\frac{0.8 \times 0.2}{5000}} \approx 0.57\%$ (Morris et al., 2019).

All simulations were performed using the open-source software `R`. The `R` scripts written by the first author especially make use of the `metafor` package for meta-analysis (Viechtbauer, 2010) as well as James Pustejovsky's `clubSandwich` package.

5.2. Results

Figs. 1–4 display the empirical coverage based on the adjusted F-test (8) and estimators CR_1^*, CR_3^*, CR_4^*, CR_2 and ST. The figures summarize the results for all combinations of N, β, ϱ and missing data ratios. CR_1^* and CR_2 yield much less than nominal coverage 95% in all settings, but especially for $k < 40$. CR_2 gives around 50% coverage for five studies, between 70-80% for ten, 82-87% for twenty and 88-91% coverage for forty studies. The CR_1^* estimator yields between 25-50% coverage for five studies, 65-75% for ten, 80-86% for twenty and 87-91% for forty studies. It is interesting to observe a clustering of coverage results for the estimator CR_1^* and $k = 5$ (depending on the inter-study correlation of effects) that cannot be observed for any other setting or estimator. The standard estimator ST gives approximately correct coverage for $k \geq 20$ but is highly conservative for $k \leq 10$ studies, especially for five. CR_3^* very consistently yields slightly more coverage than CR_4^* in all settings except for $k = 40$ where the difference between the two is negligible. For $k = 5$ coverage based on CR_4^* is approximately nominal and when based on CR_3^* slightly conservative. For $k = 10$ and $k = 20$ CR_4^* gives coverage around 91-92% and CR_3^* around 93-94%. For $k = 40$ both yield coverage around 92-94%.

In addition to these empirical coverage results, we also consider the power related to the respective tests and confidence regions. The power plots are provided in Figs. 5 and 6 for $\boldsymbol{\beta} = (0.2, 0.2, 0.1, 0.1)'$ and $\boldsymbol{\beta} = (0.4, 0.4, 0.2, 0.3)'$ respectively. We show box plots to summarize the various simulation settings. For $\boldsymbol{\beta} = (0.4, 0.4, 0.2, 0.3)'$ power is monotone increasing in the number of studies k for all estimators. For $\boldsymbol{\beta} = (0.2, 0.2, 0.1, 0.1)'$ power is monotone increasing in k for CR_3^*, CR_4^* and ST, whereas for CR_1^* and CR_2 power decreases from a median of approximately 70% and 60% to 55% and 52% respectively, when going from five to ten studies and then increases in k beyond this point.

The differences in power between the considered estimators are small for a large number of studies and become more pronounced as the number of studies decreases. For forty studies the power based on all estimators is nearly identical for both choices of $\boldsymbol{\beta}$. For twenty studies power based on CR_1^* and CR_2 is slightly higher than for the other estimators. CR_3^*, CR_4^* and ST yield approximately the same power for both choices of $\boldsymbol{\beta}$ and twenty studies. For $k = 10$ and $\boldsymbol{\beta} = (0.2, 0.2, 0.1, 0.1)'$ the median power for CR_1^* and CR_2 is around 55% and 52% respectively, whereas for CR_3^*, CR_4^* and ST it is around 25%, 31% and 20% respectively. For $k = 10$ and $\boldsymbol{\beta} = (0.4, 0.4, 0.2, 0.3)'$ the median power for CR_1^* and CR_2 is around 87%, whereas for CR_3^*, CR_4^* and ST it is around 70%, 74% and 73% respectively. For $k = 5$ and $\boldsymbol{\beta} = (0.2, 0.2, 0.1, 0.1)'$ the median power

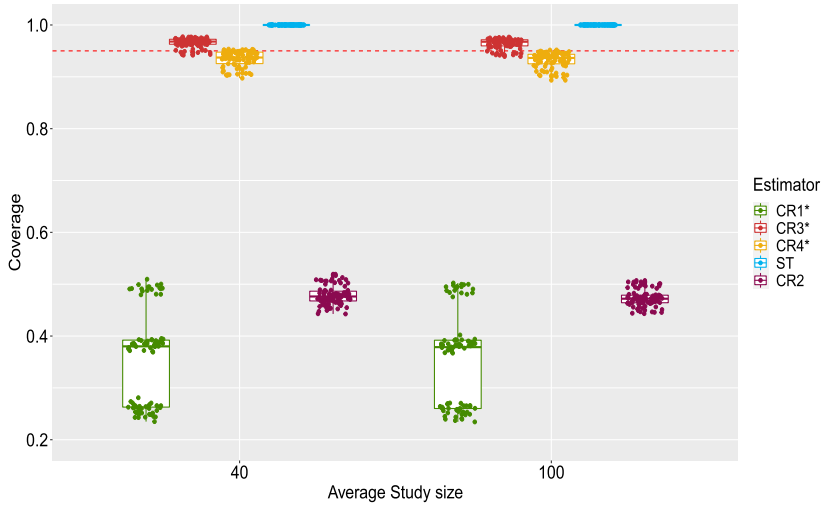

Fig. 1. Coverage of the confidence set (9) based on an inversion of the adjusted F-test for $k = 5$ studies.

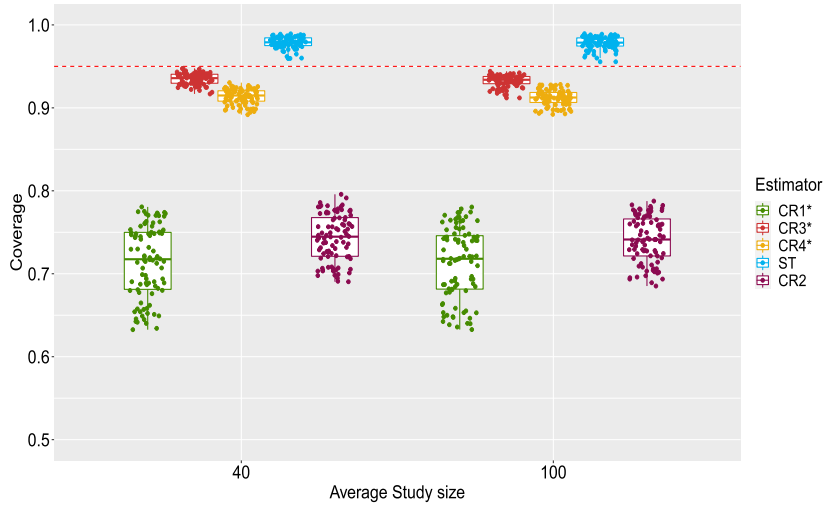

Fig. 2. Coverage of the confidence set (9) based on an inversion of the adjusted F-test for $k = 10$ studies.

for CR_1^* and CR_2 is around 70% and 60% respectively, whereas for CR_3^*, CR_4^* and ST it is only around 8%, 12% and 0% respectively. For $k = 5$ and $\boldsymbol{\beta} = (0.4, 0.4, 0.2, 0.3)'$ the median power for CR_1^* and CR_2 is around 83% and 70% respectively, whereas for CR_3^*, CR_4^* and ST it is around 13%, 24% and 1% respectively.

In the accompanying supplement we also present simulation results for the case $p = 3$ (see Section 2). In our main simulation, we assumed that estimation of the sampling covariances between study effects were available, implying approximately correct specification of the within-study variance-covariance matrix $\hat{\boldsymbol{V}}_i$ for $i = 1, \ldots, k$. In the extra simulations for dimension 3, we also check the effect of possible model misspecification, when only the sampling variances of the study effects are available but within-study correlations must be assumed, resulting in a (possibly incorrect) working model. Our results are similar to the bivariate case. The new estimators CR_3^* and CR_4^* outperform the other considered estimators with regard to empirical coverage of confidence regions for $\boldsymbol{\beta}$ and the results are stable under model misspecification. However, further work is necessary to check performance of the new estimators in even higher dimensions or when more extreme errors are made in the specification of the working model for the within-study variance-covariance matrix $\boldsymbol{V}_i$.

6. Discussion

Multivariate Meta-Regression is an important tool for synthesizing and interpreting results from trials reporting multiple, correlated effects. However, information on these correlations is rarely available to analysts, making it difficult to construct the variance-covariance $\boldsymbol{V}$ matrix of the studies' sampling errors. Cluster-robust estimators allow for a correction of the standard errors, therefore enabling more reliable inference. In this paper we introduced two new proposals of CR estimators

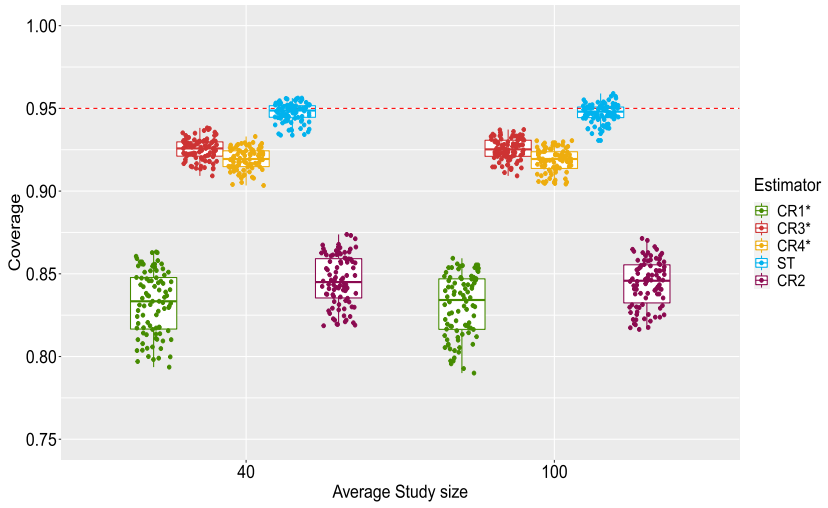

Fig. 3. Coverage of the confidence set (9) based on an inversion of the adjusted F-test for $k = 20$ studies.

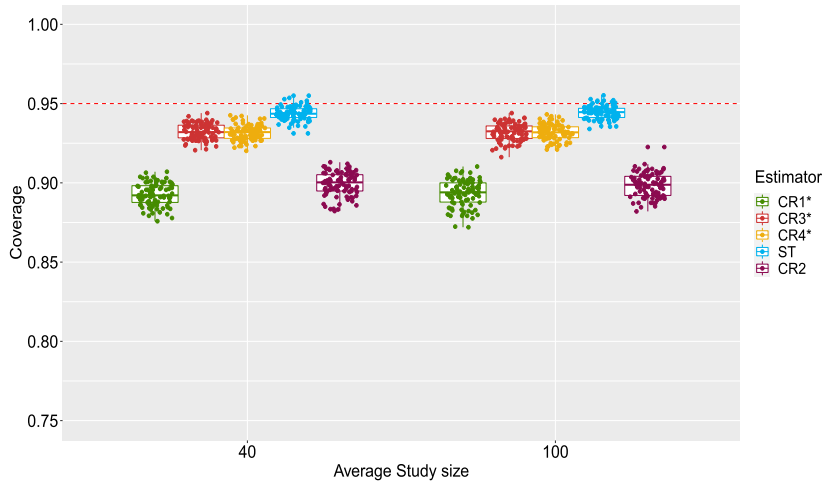

Fig. 4. Coverage of the confidence set (9) based on an inversion of the adjusted F-test for $k = 40$ studies.

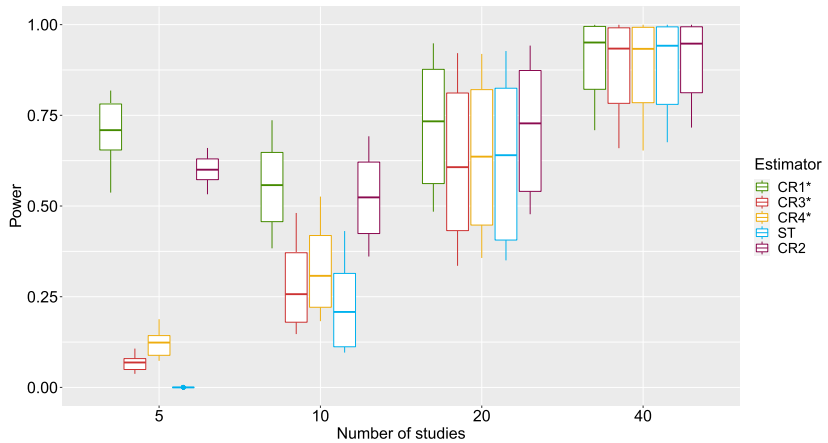

Fig. 5. Box plots of power based on adjusted F-test for all settings with $\boldsymbol{\beta} = (0.2, 0.2, 0.1, 0.1)'$.

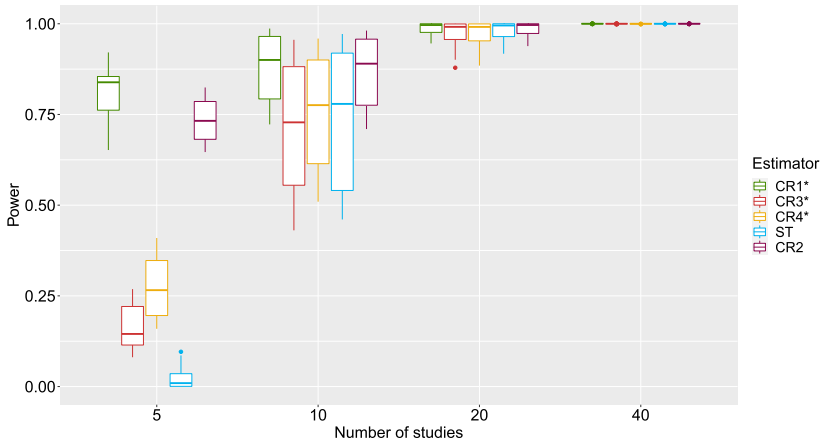

Fig. 6. Box plots of power based on adjusted F-test for all settings with $\boldsymbol{\beta} = (0.4, 0.4, 0.2, 0.3)'$.

for use in multivariate meta-regression. We performed a simulation study, comparing these estimators with results based on two alternative CR estimators and the standard variance-covariance estimator with a focus on coverage and power of confidence sets and tests, as well as an illustrative real life data analysis. In our manuscript we only investigated the bivariate meta-regression setting, although all methods discussed are also applicable in higher dimensions. Further work is necessary to assess the viability of our suggestions in other settings, such as when the number of effects per study is greater than two.

Our main findings can be summarized as follows: The Zhang estimator, discussed in Tipton and Pustejovsky (2015), can lead to a negative estimate of the denominator degrees of freedom in the F-distribution. This can occur when the number of studies is very small. The AHZ approach is therefore not recommendable for bivariate meta-regression if the number of studies is small ($k \leq 5$). Furthermore, when using the classical F-test in the bivariate setting, we recommend truncating the denominator degrees of freedom at two. The CR_1^* and CR_2 estimators yield an empirical coverage that lies far below the nominal level $1-\alpha$ and the coverage based on the other estimators, especially for smaller numbers of studies. On the flip side the tests based on these two CR-estimators unsurprisingly have superior power. The ST estimator has approximately correct coverage for $k \geq 20$ studies but is highly conservative for $k \leq 10$ studies. CR_3^* and CR_4^* yield approximately correct coverage for five studies. CR_3^* also gives nearly correct coverage for ten studies whereas CR_4^* becomes slightly liberal in this case.

Based on our results we recommend using either the CR_3^* or CR_4^* estimator for bivariate meta-regression if $k \leq 10$ with a very slight preference for CR_3^*. For an analysis with $k \geq 20$ studies the ST estimator seems to work best.

A limitation of our simulation study is that the sampling covariances between study-level effects were available for the construction of weight matrices. As mentioned in the introduction, this is often not feasible in practice, requiring analysts to calculate weights using a specified working model for the covariance structure. Hedges et al. (2010) provide possible working models likely to be found in meta-analyses. They propose the use of approximately inverse variance weights, based on these working models.

An open question that requires further research is what the best testing procedure is when the number of studies k is no greater than around five. Neither the adjusted Hotelling's T^2 approach in combination with Zhang's estimator for the degrees of freedom, which was recommended by Tipton and Pustejovsky (2015), nor the naive or adjusted F-tests used in our simulations seem to be the ideal approach. This requires more intensive work that is outside the scope of this manuscript. For a discussion of alternative estimation approaches for the degrees of freedom in the adjusted Hotelling approach, we refer to Tipton and Pustejovsky (2015). Another question for future research is whether other statistics or resampling approaches that have shown promising small sample approximations for heterogeneous MAN(C)OVA settings (Friedrich et al., 2017; Friedrich and Pauly, 2018; Zimmermann et al., 2020) can also help in multivariate meta-regression models.

Data availability Statement

The neuroblastoma dataset is contained in the `R` package `metafor`. All `R` scripts are publicly available at the online repository "Open Science Framework" under the address https://osf.io/sa6ey/.

Acknowledgements

This work was supported by the German Research Foundation (DFG) (Grant no. PA-2409 7-1). The authors gratefully acknowledge the computing time provided on the Linux HPC cluster at TU Dortmund University (LiDO3), partially funded in the course of the Large-Scale Equipment Initiative by the German Research Foundation as project 271512359.

We would also like to thank James Pustejowsky for his helpful comments during the research phase for this manuscript.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2022.107631.

References

Bell, R.M., McCaffrey, D.F., 2002. Bias reduction in standard errors for linear regression with multi-stage samples. Surv. Methodol. 28 (2), 169–182.
Berkey, C., Hoaglin, D., Antczak-Bouckoms, A., Mosteller, F., Colditz, G., 1998. Meta-analysis of multiple outcomes by regression with random effects. Stat. Med. 17 (22), 2537–2550.
Borenstein, M., Hedges, L.V., Higgins, J.P., Rothstein, H.R., 2021. Introduction to Meta-Analysis. John Wiley & Sons.
Burzykowski, T., Molenberghs, G., Buyse, M., Geys, H., Renard, D., 2001. Validation of surrogate end points in multiple randomized clinical trials with failure time end points. J. R. Stat. Soc., Ser. C, Appl. Stat. 50 (4), 405–422.
Cribari-Neto, F., 2004. Asymptotic inference under heteroskedasticity of unknown form. Comput. Stat. Data Anal. 45 (2), 215–233.
Cribari-Neto, F., Souza, T.C., Vasconcellos, K.L., 2007. Inference under heteroskedasticity and leveraged data. Commun. Stat., Theory Methods 36 (10), 1877–1888.
Davey, J., Turner, R.M., Clarke, M.J., Higgins, J.P., 2011. Characteristics of meta-analyses and their component studies in the Cochrane database of systematic reviews: a cross-sectional, descriptive analysis. BMC Med. Res. Methodol. 11 (1), 1–11.
Emura, T., Sofeu, C.L., Rondeau, V., 2021. Conditional copula models for correlated survival endpoints: individual patient data meta-analysis of randomized controlled trials. Stat. Methods Med. Res. 30 (12), 2634–2650.
Friedrich, S., Brunner, E., Pauly, M., 2017. Permuting longitudinal data in spite of the dependencies. J. Multivar. Anal. 153, 255–265.
Friedrich, S., Pauly, M., 2018. MATS: inference for potentially singular and heteroscedastic MANOVA. J. Multivar. Anal. 165, 166–179.
Guolo, A., To, D.-K., 2021. A pseudo-likelihood approach for multivariate meta-analysis of test accuracy studies with multiple thresholds. Stat. Methods Med. Res. 30 (1), 204–220.
Hayes, A.F., Cai, L., 2007. Using heteroskedasticity-consistent standard error estimators in ols regression: an introduction and software implementation. Behav. Res. Methods 39 (4), 709–722.
Hedges, L.V., 1981. Distribution theory for Glass's estimator of effect size and related estimators. J. Educ. Stat. 6 (2), 107–128.
Hedges, L.V., Tipton, E., Johnson, M.C., 2010. Robust variance estimation in meta-regression with dependent effect size estimates. Res. Synth. Methods 1 (1), 39–65.
Higgins, J.P., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21 (11), 1539–1558.
Jackson, D., Riley, R., White, I.R., 2011. Multivariate meta-analysis: potential and promise. Stat. Med. 30 (20), 2481–2498.
Johnson, R.A., Wichern, D.W., et al., 2014. Applied Multivariate Statistical Analysis, vol. 6. Pearson London, UK.
Lin, L., Aloe, A.M., 2021. Evaluation of various estimators for standardized mean difference in meta-analysis. Stat. Med. 40 (2), 403–426.
Long, J.S., Ervin, L.H., 2000. Using heteroscedasticity consistent standard errors in the linear regression model. Am. Stat. 54 (3), 217–224.
Mavridis, D., Salanti, G., 2013. A practical introduction to multivariate meta-analysis. Stat. Methods Med. Res. 22 (2), 133–158.
Morris, T.P., White, I.R., Crowther, M.J., 2019. Using simulation studies to evaluate statistical methods. Stat. Med. 38 (11), 2074–2102.
Olkin, I., Gleser, L., 2009. Stochastically dependent effect sizes. In: The Handbook of Research Synthesis and Meta-Analysis, pp. 357–376.
Pustejovsky, J., 2021. clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections. R package version 0.5.3.
Pustejovsky, J.E., Tipton, E., 2018. Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. J. Bus. Econ. Stat. 36 (4), 672–683.
Riley, R.D., 2011. Erratum: an evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. Stat. Med. 30 (4), 400.
Riley, R.D., Abrams, K., Lambert, P., Sutton, A., Thompson, J., 2007. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. Stat. Med. 26 (1), 78–97.
Riley, R.D., Sutton, A.J., Abrams, K.R., Lambert, P.C., 2004. Sensitivity analyses allowed more appropriate and reliable meta-analysis conclusions for multiple outcomes when missing data was present. J. Clin. Epidemiol. 57 (9), 911–924.
Sidik, K., Jonkman, J.N., 2005. A note on variance estimation in random effects meta-regression. J. Biopharm. Stat. 15 (5), 823–838.
Tipton, E., 2015. Small sample adjustments for robust variance estimation with meta-regression. Psychol. Methods 20 (3), 375.
Tipton, E., Pustejovsky, J.E., 2015. Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. J. Educ. Behav. Stat. 40 (6), 604–634.
Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Softw. 36 (3), 1–48.
Viechtbauer, W., López-López, J.A., Sánchez-Meca, J., Marín-Martínez, F., 2015. A comparison of procedures to test for moderators in mixed-effects meta-regression models. Psychol. Methods 20 (3), 360–374.
Welz, T., Pauly, M., 2020. A simulation study to compare robust tests for linear mixed-effects meta-regression. Res. Synth. Methods 11 (3), 331–342.
White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48 (4), 817–838.
Wilson, J., 2010. Volume of n-dimensional ellipsoid. Sciencia Acta Xaver. 1 (1), 101–106.
Zhang, J.-T., 2012. An approximate Hotelling T2-test for heteroscedastic one-way MANOVA. Open J. Stat. 2 (1), 1–11.
Zimmermann, G., Pauly, M., Bathke, A.C., 2019. Small-sample performance and underlying assumptions of a bootstrap-based inference method for a general analysis of covariance model with possibly heteroskedastic and nonnormal errors. Stat. Methods Med. Res. 28 (12), 3808–3821.
Zimmermann, G., Pauly, M., Bathke, A.C., 2020. Multivariate analysis of covariance with potentially singular covariance matrices and non-normal responses. J. Multivar. Anal. 177, 104594.