Improving the Accuracy of Confidence Intervals and Regions in Multivariate Random-effects Meta-analysis

Tsubasa Ito\textsuperscript{1} and Shonosuke Sugasawa\textsuperscript{2}

\textsuperscript{1}Department of Statistical Thinking, The Institute of Statistical Mathematics
\textsuperscript{2}Center for Spatial Information Science, The University of Tokyo

Abstract

Multivariate random-effects meta-analyses have been widely applied in evidence synthesis for various types of medical studies. However, standard inference methods usually underestimate statistical errors and possibly provide highly overconfident results under realistic situations since they ignore the variability in the estimation of variance parameters. In this article, we propose new improved inference methods without any repeated calculations such as Bootstrap or Monte Carlo methods. We employ distributional properties of ordinary least squares and residuals under random-effects models, and provide relatively simple and closed form expressions of confidence intervals and regions whose coverages are more accurate than conventional methods such as restricted maximum likelihood methods. As specific applications, we consider two multivariate meta-analysis: bivariate meta-analysis of diagnostic test accuracy and multiple treatment comparisons via network meta-analysis. We also illustrate the practical effectiveness of these methods via real data applications and simulation studies.

Key words: Asymptotic expansion; Bias correction; Confidence interval; Second order accuracy
In evidence-based medicine, meta-analysis is recognized as an essential tool for quantitatively summarizing multiple studies and producing integrated evidence. In general, the treatment effects from different sources for studies are heterogeneous due to various factors, which should be adequately addressed to avoid underestimation of statistical errors and misleading conclusions [Higgins and Green 2011]. To this end, random-effects models are widely adopted in most medical meta-analyses. The applications cover various types of systematic reviews, for example, conventional univariate meta-analysis [DerSimonian and Laird 1986, Whitehead and Whitehead 1991], bivariate meta-analysis of diagnostic test accuracy [Reitsma et al. 2005], network meta-analysis for comparing the effectiveness of multiple treatments [Salanti 2012], and individual participant meta-analysis [Riley et al. 2010].

In random-effects meta-analyses, standard inference methods depend on large sample approximations for the number of studies synthesized, e.g., the DerSimonian-Laird methods [DerSimonian and Laird 1986] and its extension [Chen et al. 2012, Jackson et al. 2010, 2018] and restricted maximum likelihood (REML) estimation [Jackson et al. 2011], but the numbers of trials are often moderate or small. In this situation, validities of the inference methods can be violated, which would lead over-confidence results, that is, coverage probabilities of the confidence regions or intervals cannot retain their nominal confidence levels and also the type-I error probabilities of the corresponding tests can be inflated. Such problem with random-effects models was well recognized in the context of both univariate and multivariate meta-analysis, even when the models are completely specified [Veroniki et al. 2019]. Recently, several refined methods have been proposed to overcome this issue (e.g., Jackson and Riley 2014, Hartung and Knapp 2001, Noma et al. 2018, 2019, Sugawara and Noma 2019), but most existing methods are computationally intensive based on Monte Carlo or Bootstrap methods or lack of theoretical justification. On the other hand, Noma (2011) proposed Bartlett-type corrections and provided closed form expressions of refined confidence intervals whose coverage is second order accurate. However,
this method is only applicable to the univariate meta-analysis and the derivation of such confidence intervals or regions under more complicated multivariate random-effects models would be quite difficult. Although Zucker et al. (2000) proposed an improved likelihood test in linear mixed models based on asymptotic expansions of the (restricted) maximum likelihood estimators, the expression of the test statistics includes tedious algebraic expressions, which is not useful in practice.

In this paper, we consider a new improved inference method in general multivariate meta-analysis without using repeated calculations such as Monte Carlo and Bootstrap methods, so that the method has low computational complexity. Moreover, we pursue two additional properties, that is, the method has theoretical justification and has relatively simple expressions for confidence intervals or regions. To this end, we employ the distributional properties between the ordinary least squares estimator and residuals, and define a class of estimators of variance parameters in random-effects models. Then, we find a relatively simple formula for asymptotic approximation of the coverage probability of the crude Wald-type confidence intervals and regions, and construct a second order accurate confidence intervals and regions. For specific applications, we provide refined confidence regions in bivariate meta-analysis for diagnostic test accuracy and refined confidence intervals in network meta-analysis for multiple treatment comparison. In both cases, we carry out simulation studies to evaluate the finite sample performance of the proposed methods compared with the existing standard methods, and demonstrate the practical usefulness using datasets.

This paper is set out as follows. In Section 2 we describe the proposed confidence intervals and regions under general multivariate random-effects models. In Sections 3 and 4, the general results are applied to bivariate meta-analysis and network meta-analysis, in which simulation studies and applications to real detests are given. We conclude with a short discussion in Section 5.
2 Improved Inference in Random-effects Meta-analysis

2.1 Multivariate Random-effects models

We first consider the general random-effects model for multivariate meta-analysis. Let $y_{ir}$ denote an estimator of the $r$th outcome measure in the $i$th study ($i = 1, \ldots, n$; $r = 1, \ldots, p$). Typically, mean difference, risk difference, odds ratios and hazard ratios are used for effect measures, and the ratio measures are usually log-transformed to allow normal approximations. Let $\beta$ be the $p$-dimensional vector of true overall average effects, and $u_i$ be the corresponding random effects. In practical situation, each study often reports only a subset of outcomes, that is, a subset of $(y_{i1}, \ldots, y_{ip})$ is observed. Let $p_i (\leq p)$ be the number of observed outcomes and $y_i$ be the $p_i$-dimensional vector of observed outcomes. Further, we define two design matrices $X_i$ and $Z_i$ showing which outcomes are observed in the $i$th study. The details will be discussed in specific applications given in Sections 3 and 4.

Here we consider the following multivariate random-effects model:

$$y_i = X_i \beta + Z_i u_i + \epsilon_i, \quad i = 1, \ldots, n,$$

where $u_i \sim N_p(0, \Sigma(\psi))$, $\epsilon_i \sim N_p(0, S_i)$ with known $S_i$, and $\psi$ is a vector of unknown parameters in $\Sigma$. Throughout the paper, we assume that the elements of $S_i$’s are uniformly bounded. The model (1) can be expressed as the following matrix form $y = X \beta + Zu + \epsilon$, where $y = (y_1^t, \ldots, y_n^t)^t$, $X = (X_1^t, \ldots, X_n^t)^t$, $Z = \text{block diag}(Z_1, \ldots, Z_m)$ and $u$ and $\epsilon$ are defined in the same way.

Under the formulation, one is interested in inference on $\mu = c^t \beta$ with a $p$-dimensional vector $c$ or $\eta = C \beta$ with $k \times p$ matrix $C$. For instance, when $c = (1, 0, \ldots, 0)$, $c^t \beta = \beta_1$, which is the average effect in the first element of $\beta$. Since $\mu$ is a scalar and $\eta$ is a vector, confidence intervals and regions, respectively, are typically used for statistical inference on these parameters. We consider Wald-Type confidence intervals and regions with higher order accuracy.

The variance parameter $\psi$ in $\Sigma$ is a nuisance parameter, so that the variability
of the estimation should be adequately taken into account for valid inference on \( \mu \) and \( \eta \). Here we restrict the class for estimators of \( \psi \) that satisfies the following three conditions:

(C1) \( \hat{\psi} \) is an even function of \( y \) and translation invariant, that is, \( \hat{\psi}(y) = \hat{\psi}(-y) \), and 
\[
\hat{\psi}(y + XT) = \hat{\psi}(y) \quad \text{for any} \quad T \in \mathbb{R}^p.
\]

(C2) \( \hat{\psi} \) is \( \sqrt{n} \)-consistent and \( \Sigma(\hat{\psi}) \) is second-order unbiased, namely 
\[
\hat{\psi} - \psi = O(n^{-1/2}) \quad \text{and} \quad E[\Sigma(\hat{\psi})] = \Sigma(\psi) + o(n^{-1}).
\]

(C3) \( \hat{\psi} \) is a function of \( Py \) with \( P = I_n - X^tX)^{-1}X^t \).

The first condition (C1) is typically satisfied by typical estimators including (restricted) maximum likelihood estimator and moment-based estimators. The \( \sqrt{n} \)-consistency in (C2) is also a standard condition, but second order unbiasedness of \( \hat{\psi} \) is not always satisfied. For example, the maximum likelihood (ML) estimator does not hold the property. The condition (C3) requires that the estimator should be function of residuals based on ordinary least squares estimator of \( \beta \), which is a key assumption of the proposed confidence intervals and regions in this paper. The condition enables to get a relatively simple form of corrected confidence intervals and regions. Note that the typical estimators (e.g. REML) does not satisfy the condition (C3). The detailed estimator \( \hat{\psi} \) which satisfies all the conditions are discussed later, which leads to a new estimator of \( \psi \). For given \( \psi \), \( \beta \) can be estimated by the generalized lead squares estimator \( \hat{\beta}(\psi) = (X^t\{Z(I_n \otimes \Sigma(\psi))Z^t + S\}^{-1}X)^{-1}X^t\{Z(I_n \otimes \Sigma(\psi))Z^t + S\}^{-1}y \). We consider inference on \( \mu \) or \( \eta \) based on the estimator \( \hat{\mu}(\hat{\psi}) = c^t\hat{\beta}(\hat{\psi}) \) and \( \hat{\eta}(\hat{\psi}) = C\hat{\beta}(\hat{\psi}) \) with \( \hat{\psi} \) satisfying conditions (C1)∼(C3).

2.2 Improved confidence intervals and regions

We first consider confidence intervals of \( \mu \). Let \( V(\psi) \) be the variance of \( \hat{\beta}(\psi) \), that is, 
\[
V(\psi) = (X^t\{Z(I_n \otimes \Sigma(\psi))Z^t + S\}X)^{-1},
\]
noting that \( V(\psi) = O(n^{-1}) \). Then, the Wald-type confidence interval of \( \mu \) is given by 
\[
(\hat{\mu}(\hat{\psi}) - \{c^tV(\hat{\psi})c\}^{1/2}z, \hat{\mu}(\hat{\psi}) + \{c^tV(\hat{\psi})c\}^{1/2}z)
\]
for some \( z \), which we call naive confidence interval. For example, if \( z = z_{\alpha/2} \) with \( z_q \)
being the upper 100\%\%-quantile of $N(0,1)$, the coverage probability of the confidence intervals is approximately $1-\alpha$, so that the confidence interval is asymptotically valid. However, the coverage error is $O(n^{-1})$ as demonstrated in the proof of Theorem 1, which cannot be negligible when $n$ is small or moderate as common in practical meta-analysis. To overcome the difficulty, we modify the confidence intervals by adequately calibrating $z$. Such calibration can be done based on the asymptotic expansion of the coverage probability of the naive confidence interval, which usually has a tedious algebraic expression. On the other hand, it is shown that $\hat{\beta}(\psi)$ and $\hat{\psi}$ are mutually independent under (C3) and the asymptotic bias of $V(\psi)$ is negligible under condition (C2), both of which make the asymptotic formula for the coverage probability considerably simple. Based on the asymptotic expansion, we can define the corrected confidence interval

$$\text{CCI}_\alpha : (\hat{\mu}(\hat{\psi}) - \{c^tV(\hat{\psi})c\}^{1/2}z^*(\hat{\psi}), \hat{\mu}(\hat{\psi}) + \{c^tV(\hat{\psi})c\}^{1/2}z^*(\hat{\psi})),$$  

(2)

where $z^*(\psi) = z_{\alpha/2} + (z_{\alpha/2}^3 + z_{\alpha/2})A(\psi)/(8\{c^tV(\psi)c\}^2)$ and $A(\psi) = E[\{c^tV(\hat{\psi})c - c^tV(\psi)c\}^2]$, noting that $A(\psi) = O(n^{-3})$. The coverage probability of the corrected confidence interval is correct up to the second order as shown in the following Theorem, where the proof is given in the Supplementary Material.

**Theorem 1.** Under the conditions (C1)\textasciitilde(C3), $P(\mu \in \text{CCI}_\alpha) = 1 - \alpha + o(n^{-1})$.

For using the corrected confidence interval, we need to derive the analytical expression of $A(\psi)$, which can be calculated based on the (asymptotic) expression of $\hat{\psi}$ up to $O_p(n^{-1/2})$. The detailed derivation is demonstrated in network meta-analysis in Section 4.

We next consider confidence regions of $\eta$. The naive confidence region of $\eta$ can be defined as $\{\tilde{\eta}(\hat{\psi}) - \eta\}^{1\{CV(\hat{\psi})C^t\}^{-1}\{\tilde{\eta}(\hat{\psi}) - \eta\} \leq x, and x = \chi^2_k(\alpha)$, the upper 100\%\%-quantile of $\chi^2_k$ distribution, leads to the approximate confidence region with the nominal level $1 - \alpha$. However, it has considerable coverage error of $O(n^{-1})$ as demonstrated in the proof of Theorem 2, we consider corrected coverage regions. To
this end, we define the following three quantities:

\[ B_1(\psi) = E\left[ \text{tr}(K(\hat{\psi}))^2 \right], \quad B_2(\psi) = \text{tr}\left( E[K(\hat{\psi})^2] \right), \quad B_3(\psi) = \text{tr}\left( E[K(\hat{\psi})] \right), \quad (3) \]

where \( K(\hat{\psi}) = C\{V(\hat{\psi}) - V(\psi)\}C^t\{CV(\hat{\psi})C^t\}^{-1} \). Note that \( B_1, B_2 \) and \( B_3 \) are \( O(n^{-1}) \). Then, the proposed corrected confidence region is given by

\[ \text{CCR}_\alpha : \{\hat{\eta}(\hat{\psi}) - \eta\}^t\{CV(\hat{\psi})C^t\}^{-1}\{\hat{\eta}(\hat{\psi}) - \eta\} \leq \chi^2_k(\alpha)(1 + h(\hat{\psi})), \quad (4) \]

where \( k \) is the dimension of \( \eta \) and

\[ h(\psi) = -\frac{B_1(\psi) - 2B_2(\psi) + 8B_3(\psi)}{4k} + \frac{B_1(\psi) + 2B_2(\psi)}{4k(k + 2)} \chi^2_k(\alpha). \]

The coverage accuracy of the corrected confidence region is given in the following Theorem.

**Theorem 2.** Under the conditions \((C1)\sim(C3)\), \( P(\eta \in \text{CCR}_\alpha) = 1 - \alpha + o(n^{-1}) \).

For using the confidence region \( \text{CCR}_\alpha \), we need to find the approximation formula for \( B_1, B_2 \) and \( B_3 \), which can be done in specific settings. We demonstrate the proposed confidence region in bivariate meta-analysis in Section 3 and provide detailed approximated expressions of the three quantities.

### 3 Bivariate meta-analysis for diagnostic test accuracy

#### 3.1 Model settings

There has been increasing interest in systematic reviews and meta-analyses of data from diagnostic accuracy studies. For this purpose, a bivariate random-effect model \[(\text{Reitsma et al.}, 2005) \text{ and } \text{Harbord et al.}, 2007\] is widely used. Following \text{Reitsma et al.} (2005), we define \( \mu_{Ai} \) and \( \mu_{Bi} \) as the logit-transformed true sensitivity and specificity, respectively, in the \( i \)th study. Let \( y_{Ai} \) and \( y_{Bi} \) be the observed logit-transformed sensitivity and specificity, and \( s_{Ai} \) and \( s_{Bi} \) are associated standard errors. The bivariate model assumes that \( \mu_i = (\mu_{Ai}, \mu_{Bi})^t \) and \( y_i = (y_{Ai}, y_{Bi})^t \) follow bivariate normal
distributions:

\[ y_i | \mu_i \sim N_2(\mu_i, S_i), \quad \mu_i \sim N_2(\beta, \Sigma), \quad i = 1, \ldots, n, \]  

(5)

where \( \beta = (\beta_A, \beta_B)^t \) is a vector of the average logit-transformed sensitivity and specificity, and \( S_i = \text{diag}(s_{A_i}, s_{B_i}) \). The model (5) is a special case with \( X_i = Z_i = I_2 \) in (1). Here \( \Sigma \) is unstructured, so that it allows correlation between \( \mu_{A_i} \) and \( \mu_{B_i} \).

The unknown parameters are \( \beta \) and \( \Sigma \).

### 3.2 Confidence region

For summarizing the results of the meta-analysis, we consider CRs of \( \beta \) since sensitivity and specificity might be highly correlated. Reitsma et al. (2005) suggested the 100(1 - \( \alpha \))% joint CR for \( \mu \) as the interior points of the ellipse defined as

\[ \left\{ \beta : (\hat{\beta}(\hat{\Sigma}) - \beta)^t V(\hat{\Sigma})^{-1} (\hat{\beta}(\hat{\Sigma}) - \beta) \leq \chi_2^2(\alpha) \right\}, \]  

(6)

where \( \hat{\beta}(\Sigma) = (\sum_{i=1}^{m} D_i^{-1})^{-1} \sum_{i=1}^{m} D_i^{-1} y_i \) for \( D_i(\Sigma) = D_i = \Sigma + S_i \) is the generalized least squares estimator of \( \beta \), \( V = V(\Sigma) = (\sum_{i=1}^{n} D_i^{-1})^{-1} \) is the variance-covariance matrix of \( \hat{\beta} \), \( \hat{\Sigma} \) is the restricted maximum likelihood estimator and \( \chi_2^2(\alpha) \) is the upper 100\( \alpha \)% point of the \( \chi^2 \) distribution with 2 degrees of freedom. The joint CR (6) is approximately valid, that is, the coverage error converges to \( 1 - \alpha \) as the number of studies \( n \) goes to infinity. However, when \( n \) is not sufficiently large, the coverage error is not negligible, and the region (6) would under-cover the true \( \mu \).

We consider the following moment-based estimator of \( \Sigma \):

\[ \hat{\Sigma}_0 = \frac{1}{n} \sum_{i=1}^{n} \left\{ (y_i - X_i \hat{\beta}^{OLS}) (y_i - X_i \hat{\beta}^{OLS})^t - S_i \right\}, \]

for \( \hat{\beta}^{OLS} = (X^t X)^{-1} X^t y \). Let \( \hat{\Sigma} \) be the non-negative definite bias corrected moment estimator, that is, \( \hat{\Sigma} = \hat{\Sigma}_0 - \text{Bias}_{\Sigma_0}(\hat{\Sigma}) \) with \( \text{Bias}_{\Sigma_0}(\Sigma) = -n^{-2} \sum_{i=1}^{n} D_i(\Sigma) \), which satisfies all the conditions (C1)~(C3). Based on the simple form of \( \hat{\Sigma} \), we can derive the approximation of \( B_1, B_2 \) and \( B_3 \) in (3), that is, we can obtain \( B_\ell^*, \ \ell = 1, 2, 3 \)
which satisfy $B^*_\ell = B_\ell + o(n^{-1})$, where

$$B^*_1 = \frac{2}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \text{tr} \left( VU_{ijk} VU_{kij} \right)$$

$$B^*_2 = \frac{1}{n^2} \sum_{i=1}^n \text{tr} \left( V \sum_{j=1}^n U_{jij}^2 \right) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \text{tr}^2 \left( VU_{ijk} \right)$$

$$B^*_3 = B_2 - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{tr} \left( VU_{ijij} D_j D_j^{-1} \right) - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{tr} \left( D_j^{-1} D_j \right) \text{tr} \left( VU_{ijij} \right),$$

for $U_{ijk} = D_j^{-1} D_j D_k^{-1}$. The detailed derivation is given in the Supplementary Material. Using $B^*_\ell$, $\ell = 1, 2, 3$, the corrected confidence region is obtained from (4) with $B_\ell = B^*_\ell$ and $\Sigma = \hat{\Sigma}$.

### 3.3 Simulation study

We assessed the finite sample performance of the proposed confidence region (4) together with the approximate confidence region (6) by Reitsma et al. (2005). We set $\mu_A = \mu_B = 0$ and $\tau_A = \tau_B (= \tau)$. We used the between study variances $\tau^2$ of 0.5, 0.75 and 1, and the between study correlations $\rho$ of 0, 0.4 and 0.8. Following Jackson and Riley (2014), for each simulation, two sets of $k$ within-study variances were simulated from a scaled chi-squared distribution with 1 degree of freedom, multiplied by 0.25, and truncated to lie within the interval $[0.009, 0.6]$. We changed the number of studies $n$ over 8, 12 and 16, and set the nominal level $\alpha$ to 0.05. In the 1000 simulations, we evaluated empirical coverage probabilities for 95% confidence regions of the true parameters. For simplicity, we evaluated coverage rates assessing rejection rates of the test of null hypothesis for the true parameters. Since areas of the corrected confidence region is approximately $1 + h(\hat{\psi})$ times larger than those of naive ones, we also computed median values of $h(\hat{\psi})$.

The results of the simulations are shown in Table 1. The simulated coverage probabilities of the standard method, ACR, are seriously smaller than the nominal level (95%), especially in the case with the small number of studies ($n = 8$). Such undesirable results would come from the crude approximation in (6). On the other
hand, the simulated coverage probabilities of the proposed CCR are around the nominal level in all the scenarios. Under high correlation \((\rho = 0.8)\), the proposed CCR tends to be conservative, but it provides reasonable credible regions judging from the reported values of \(h\). Also it is reasonable to observe that \(h\) decreases as the number of samples increases.

Table 1: Simulated coverage probabilities (%) the proposed corrected confidence region (CCR), and naive confidence region (NCR) with 95% confidence level.

| \(\tau^2\) | \(n\) | \(\rho = 0\) | \(\rho = 0.4\) | \(\rho = 0.8\) |
|-----|-----|---------|---------|---------|
|     | NCR | CCR     | \(h\)   | NCR     | CCR     | \(h\)   | NCR     | CCR     | \(h\)   |
| 8   | 74.8 | 97.2    | 1.23    | 74.2    | 98.8    | 1.39    | 75.1    | 99.5    | 1.98    |
| 12  | 77.1 | 96.8    | 0.65    | 76.9    | 96.2    | 0.76    | 76.3    | 97.9    | 1.53    |
| 16  | 80.4 | 95.4    | 0.46    | 80.0    | 96.0    | 0.55    | 77.6    | 98.1    | 1.29    |
| 8   | 85.8 | 97.1    | 0.81    | 85.3    | 97.3    | 0.90    | 82.3    | 99.3    | 1.62    |
| 0.75 | 12 | 88.5    | 96.1    | 0.49    | 85.7    | 94.9    | 0.53    | 82.8    | 98.8    | 1.20    |
| 16  | 87.3 | 94.7    | 0.35    | 88.0    | 95.4    | 0.38    | 85.4    | 98.3    | 0.80    |
| 8   | 89.0 | 95.8    | 0.67    | 87.0    | 96.2    | 0.74    | 85.2    | 98.3    | 1.34    |
| 12  | 91.1 | 95.7    | 0.42    | 88.4    | 94.6    | 0.45    | 88.4    | 98.3    | 0.81    |
| 16  | 91.2 | 95.3    | 0.31    | 89.5    | 94.7    | 0.33    | 90.4    | 98.0    | 0.58    |

3.4 Example: screening test accuracy for alcohol problems

Here we provide a re-analysis of the dataset given in Kriston et al. (2008), including \(n = 14\) studies regarding a short screening test for alcohol problems. Following Reitsma et al. (2005), we used logit-transformed values of sensitivity and specificity, denoted by \(y_{Ai}\) and \(y_{Bi}\), respectively, and associated standard errors \(s_{Ai}\) and \(s_{Bi}\). For the bivariate summary data, we fitted the bivariate models \(5\) and computed 95% CRs of \(\beta\) based on NCR \(6\) given in Reitsma et al. (2005) and the proposed CCR. Following Reitsma et al. (2005), the obtained two CRs of \(\beta\) were transformed to the scale \((\logit(\beta_A), 1 - \logit(\beta_B))\), where \(\logit(\beta_A)\) and \(1 - \logit(\beta_B)\) are the sensitivity and false positive rate, respectively. The obtained two CRs are presented in Figure 1 with a plot of the observed data, summary points \(\hat{\beta}\), and the summary receiver operating curve. The approximate CR is smaller than the proposed CR, which may indicate that the approximation method underestimates the variability of estimating nuisance variance parameters.
Figure 1: The approximate and corrected CRs and summary receiver operating characteristics (SROC) curve.

4 Network meta-analysis

4.1 Model settings

Suppose there are $p$ treatments in contract to a reference treatment, and let $y_{ir}$ be an estimator of relative treatment effect for the $r$th treatment in the $i$th study. In network meta-analysis, each study contains only $p_i (< p)$ treatments ($p_i$ usually ranges from 2 to 5); thereby, several components in $y_i = (y_{i1}, \ldots, y_{ip})^t$ cannot be obtained. When the corresponding treatments are not involved in the $i$th study, the corresponding components in $y_i$ and the within-study variance-covariance matrix $S_i$ are shrunk to the sub-vector and sub-matrix, respectively. Moreover, when the references treatment is not involved in the $i$th study, we can adopt the data argumentation approach of White et al. (2012), in which a quasi-small data set is added to the reference arm, e.g. 0.001 events for 0.01 patients. To clarify the setting in which $y_i$ and $S_i$ are shrunk to
the sub-vector and sub-matrix, respectively, we introduce an index $a_{ij} \in \{1, \ldots, p\}$, $j = 1, \ldots, p_i$, representing the treatment estimates that are available in the $i$th study, and define the $p$-dimensional vector $x_{ij}$ of $0$’s, excluding the $a_{ij}$th element that is equal to $1$. Moreover, we define $X_i = (x_{i1}, \ldots, x_{ip_i})^t$, and $y_i$ and $S_i$ are the shrunken $p_i$-dimensional vector and $p_i \times p_i$ matrix of $y_i$ and $S_i$, respectively. Then the model can be expressed as the multivariate random effects model given in Eq. (1). This model is known as the contrast-based model (Salanti et al., 2008; Dias and Ades, 2016), which is commonly used in practice. Regarding the structure of between study variance $\Sigma$, since there are rarely enough studies to identify the unstructured model of $\Sigma$, the compound symmetry structure $\Sigma = \tau^2 P(0.5)$ is used in most cases (White, 2015), where $P(\rho)$ is a matrix with all diagonal elements equal to 1 and all off-diagonal elements equal to $\rho$.

4.2 Confidence interval

Since $E[y_i] = X_i \beta$ and $\text{Var}(y_i) = S_i + \tau^2 Q_i$ with $Q_i = Z_i P(0.5) Z_i^t$. Then the crude moment estimator of $\tau^2$ is given by $\hat{\tau}^2 = \text{tr}(Q^2)^{-1} \text{tr}\{Q(y^t P y - S)\}$, where $Q = \text{diag}(Q_1, \ldots, Q_n)$, $S = \text{diag}(S_1, \ldots, S_n)$ and $P = I_N - M$ with $N = \sum_{i=1}^n p_i$ and $M = X(X^t X)^{-1} X^t$. When ith study contains the quasi-small data, diagonal elements of $S_i$ have very large values, so that the corresponding diagonal element of $y^t P y - S$ is very likely to be negative and it would produce a negative value of $\hat{\tau}^2$. To avoid this problem, we introduce $\text{tr}+(W) = \sum_{\ell=1}^L \max(0, w_{\ell})$ for $L \times L$ matrix $W$ with $w_{\ell}$ being $\ell$’s diagonal element. Then, we use the following estimator:

$$\hat{\tau}^2_0 = \text{tr}(Q^2)^{-1} \text{tr}+\{Q(y^t P y - S)\}, \quad (8)$$

Note that $\hat{\tau}^2_0$ is a function of $P y$, so that $\hat{\tau}^2_0$ satisfies the condition (C3). On the other hand, the second-order bias of $\hat{\tau}^2_0$ is given by

$$\text{Bias}_{\hat{\tau}^2_0}(\tau^2) = \text{tr}(Q^2)^{-1} \text{tr}\{Q(VM - VM - MV)\},$$

12
where \( V \equiv V(\tau^2) = \{X^\top(S+\tau^2Q)^{-1}X\}^{-1} \). Then the bias corrected estimator of \( \tau^2 \) is given by \( \hat{\tau}^2 = \hat{\gamma}^2 - \text{Bias}(\hat{\gamma}^2) \), which satisfies all the conditions (C1)\( \sim \) (C3). Using the expression of \( \hat{\tau}^2 \), we can derive the following approximation \( A^*(\tau^2) \) of \( A(\tau^2) \) which satisfies \( A^*(\tau^2) = A(\tau^2) + o(n^{-1}) \):

\[
A^*(\tau^2) = \frac{2}{\text{tr}(Q^2)^2} \text{tr} \left\{ \left( \frac{QV}{\sigma^2} \right)^2 \right\} \left\{ c^\top VXV^{-1}QV^{-1}XHVc \right\}^2.
\]

(9)

The details of the derivation is given in the Supplementary Material. Using the expression (9), we can employ the corrected confidence interval of \( \mu = c^\top \beta \) developed in Section 2.2.

4.3 Simulation study

We investigate the performance of the proposed Monte Carlo (MC) method under practical network meta-analysis scenarios. We compared the coverage probabilities of the MC method with those of widely used standard methods: the Wald-type confidence intervals based on REML estimates, the LR-based confidence interval. Throughout the experiments, we set the nominal level \( \alpha \) to 0.05. Following Noma et al. (2018), we considered a quadrangular network comparing 4 treatments (A, B, C, and D, regarding A as a reference). The numbers of trials \( n \) were set to 8, 12, and 16 and the detailed designs of trials are presented in Table 2. To approximate practical situations of medical meta-analyses, we mimicked the simulation settings considered by Sidik and Jonkman (2007). We first generated binomial data from \( X_{ir} \sim \text{Binomial}(n_{ir}, p_{ir}) \), \((i = 1, \ldots, k)\), where \( r = 0, 1, 2, \) and 3 corresponds to the treatments A, B, C, and D, respectively. The response rate of treatment A, \( p_{i0} \), was generated from a continuous uniform distribution on \([0.095, 0.65]\) and we set \( p_{ir} = p_{i0} \exp(\theta_{ir})/\{1 - p_{i0} + p_{i0} \exp(\theta_{ir})\} \) for \( r = 1, 2, \) and 3, which means that \( \theta_{ir} \) is odds ratio (ORs) to the reference treatment A, i.e. \( \theta_{ir} = \text{legit}(p_{ir}) - \text{legit}(p_{i0}) \). Also, the OR parameters \((\theta_{i1}, \theta_{i2}, \theta_{i3})\) were generated from a multivariate normal distribution \( N(\mu, \tau^2P(0.5)) \), where \( \mu = (\mu_1, \mu_2, \mu_3) \) is a vector of the true average treatment effects set to \( \mu = (0.4, 0.7, 1.0) \). The sample sizes were set to equal one
Table 2: Numbers of trials for each study design included in the simulation studies in Section 4.3

|                   | $n = 8$ | $n = 12$ | $n = 16$ |
|-------------------|--------|---------|---------|
| A vs. B           | 1      | 2       | 2       |
| A vs. C           | 3      | 4       | 6       |
| A vs. D           | 1      | 2       | 3       |
| B vs. C           | –      | –       | 1       |
| B vs. D           | –      | 1       | 1       |
| C vs. D           | 1      | 1       | 1       |
| A vs. C vs. D     | 1      | 1       | 1       |
| B vs. C vs. D     | 1      | 1       | 1       |

Another, $n_{i0} = n_{i1} = n_{i2} = n_{i3}$ for any $i$ and were drawn from a discrete uniform distribution on 20 and 200. From the generated binomial data $X_{ir}$’s, we calculated trial-specific summary statistics $y_i$ and $S_i$ in the standard manner (Higgins and Green, 2011). In the 5000 simulations, we evaluated empirical coverage probabilities and average lengths of 95% confidence intervals of the true parameters.

The results of the simulations are shown in Table 3. In general, the coverage probabilities of the REML confidence intervals are slightly better than the LR confidence intervals. However, they showed under-coverage properties under moderate number of studies ($n = 8, 12$) and large heterogeneity ($\tau = 0.4$). On the other hand, the coverage probabilities of the proposed MC method were generally around the nominal level (95%) in most cases. Under the small number of studies $k = 8$ and large heterogeneity ($\tau = 0.4$), the coverage rates were relatively low, but even under these scenarios, they performed better than the ML and REML methods.

4.4 Example: Smoking cessation data

We considered network meta-analysis based on 24 studies that compared 4 smoking cessation counseling programs (no contact, self-help, individual counseling, and group counseling), which was used in Lu and Ades (2006) and Noma et al. (2018). The outcome was successful smoking cessation at 6 to 12 months, and the comparative efficacy was assessed using odds ratios for the response rates based on each treatment arm. We applied the standard methods using ML and REML as well as the proposed
Table 3: Simulated coverage probabilities (%) and average length of 95% confidence intervals from the proposed CCI, REML and LR methods.

| n | \( \tau \) | Coverage Probability | Average Length |
|---|---|---|---|
|   |   | REML | LR | CCI | REML | LR | CCI |
| 0.3 | \( \mu_1 \) | 92.3 | 90.8 | 97.8 | 1.22 | 1.14 | 1.79 |
| 0.3 | \( \mu_2 \) | 92.0 | 90.3 | 97.7 | 0.76 | 0.71 | 1.13 |
| 0.3 | \( \mu_3 \) | 92.5 | 90.7 | 97.8 | 0.92 | 0.86 | 1.37 |
| 0.4 | \( \mu_1 \) | 90.9 | 88.8 | 96.4 | 1.41 | 1.30 | 1.88 |
| 8 | 0.4 | \( \mu_2 \) | 90.9 | 89.2 | 96.7 | 0.88 | 0.82 | 1.19 |
| 0.4 | \( \mu_3 \) | 90.8 | 88.9 | 96.8 | 1.07 | 0.99 | 1.43 |
| 0.6 | \( \mu_1 \) | 90.8 | 88.3 | 93.6 | 1.85 | 1.69 | 2.09 |
| 0.6 | \( \mu_2 \) | 90.8 | 88.5 | 93.7 | 1.17 | 1.07 | 1.33 |
| 0.6 | \( \mu_3 \) | 90.4 | 88.1 | 93.7 | 1.41 | 1.29 | 1.60 |
| 12 | 0.4 | \( \mu_2 \) | 90.7 | 89.5 | 96.3 | 0.79 | 0.75 | 1.02 |
| 0.4 | \( \mu_3 \) | 92.6 | 91.2 | 96.8 | 0.88 | 0.84 | 1.15 |
| 0.6 | \( \mu_1 \) | 92.3 | 90.9 | 94.0 | 1.39 | 1.31 | 1.51 |
| 0.6 | \( \mu_2 \) | 92.4 | 91.1 | 94.1 | 1.04 | 0.99 | 1.14 |
| 0.6 | \( \mu_3 \) | 92.9 | 91.3 | 94.2 | 1.17 | 1.11 | 1.28 |
| 16 | 0.4 | \( \mu_2 \) | 93.0 | 92.2 | 97.9 | 0.81 | 0.78 | 1.14 |
| 0.4 | \( \mu_3 \) | 92.9 | 92.3 | 98.3 | 0.68 | 0.65 | 0.96 |
| 0.6 | \( \mu_1 \) | 92.2 | 90.8 | 96.8 | 0.95 | 0.91 | 1.20 |
| 0.6 | \( \mu_2 \) | 92.6 | 91.9 | 97.0 | 0.67 | 0.64 | 0.84 |
| 0.6 | \( \mu_3 \) | 92.9 | 91.8 | 97.1 | 0.80 | 0.77 | 1.01 |

methods with multivariate random-effects model. The reference treatment was set to be “no contact.” In Table 4, we present 95% confidence intervals based on the proposed method (denoted by CCI) as well as the REML method and the likelihood ratio (LR) method. It is observed that the proposed methods produce uniformly wider confidence intervals than the other methods, which shows that the proposed CCI method would adequately quantify the statistical errors as confirmed in the simulation study.
Table 4: 95% confidence intervals based on the proposed method (denoted by CCI) and REML and likelihood ratio (LR) methods in the application to network-meta analysis of smoking cessation data.

| no contact v.s. | REML      | LR        | CCI        |
|----------------|-----------|-----------|------------|
| self-help      | (-0.383, 0.951) | (-0.365, 0.939) | (-0.556, 1.21) |
| individual counseling | (0.288, 1.09) | (0.297, 1.09) | (0.217, 1.27) |
| group counseling | (-0.174, 1.35) | (-0.157, 1.34) | (-0.297, 1.65) |

5 Discussion

In this paper, we presented new inference methods for multivariate meta-analysis without using repeated calculations such as Monte Carlo or Bootstrap methods. The proposed confidence intervals and regions have relatively simple form and they are shown to have second order accurate coverage probabilities while the standard inference methods (e.g. REML and LR) have significant coverage errors. In simulation studies, we demonstrated that possible undercoverage properties of the standard methods under the small number of studies to be synthesized while the proposed method provides reasonable coverage properties.

Although we provided a general expression of the refined confidence intervals and regions in section 2, the expressions depend on the specific forms of estimators of variance parameters satisfying three conditions given in section 2. In Sections 3 and 4, we employed simple bias corrected estimators of variance parameters, but there might be another useful estimator satisfying the three conditions. However, the detailed investigation would extend the scope of this paper and we left it for a future study.

A possible limitation of the proposed method might be that the coverage accuracy still depends on the number of studies to be synthesized. On the other hand, inference methods that does not rely on large sample approximation have been recently proposed (e.g. Noma et al. 2019; Sugasawa and Noma 2019), which are computationally intensive, so they would not be necessarily practical. Then, the proposed method would be regarded as a reasonable compromise between methods with exact empirical coverage and computational efficiency.
Acknowledgments

This research was supported by Japan Society of Promotion of Science KAKENHI (grant number: 18K12757).

References

Chen, H., A. K. Manning, and J. Dupuis (2012). A method of moments estimator for random effect multivariate meta-analysis. *Biometrics 68*, 1278–1284.

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

Dias, S. and A. E. Ades (2016). Absolute or relative effects? arm-based synthesis of trial data. *Research Synthesis Methods 7*, 23–28.

Harbord, R. M., J. J. Deeks, H. Egger, P. Whiting, and J. A. C. Sterne (2007). A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics 8*, 239–251.

Hartung, J. and G. Knapp (2001). A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine 20*, 3875–3889.

Higgins, J. P. T. and S. Green (2011). *Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0*. The Cochrane Collaboration.

Jackson, D., S. Bujkiewicz, M. Law, R. D. Riley, and I. R. White (2018). A matrix-based method of moments for fitting multivariate network meta-analysis models with multiple outcomes and random inconsistency effects. *Biometrics 74*, 548–556.

Jackson, D., R. Riley, and I. R. White (2011). Multivariate meta-analysis: potential and promise. *Statistics in Medicine 30*, 2481–2498.

Jackson, D. and R. D. Riley (2014). A refined method for multivariate meta-analysis and meta-regression. *Statistics in Medicine 33*, 541–554.
Jackson, D., I. R. White, and S. G. Thompson (2010). Extending Dersimonian and Laird’s methodology to perform multivariate random effects meta-analyses. *Statistics in Medicine* 29, 1282–1297.

Kriston, L., L. Höelzel, A. Weiser, M. Berner, and M. Haerter (2008). Meta-analysis: Are 3 questions enough to detect unhealthy alcohol use? *Annals of Internal Medicine* 149, 879–888.

Lu, G. and A. E. Ades (2006). Assessing evidence consistency in mixed treatment comparisons. *Journal of the American Statistical Association* 101, 447–459.

Noma, H. (2011). Confidence intervals for a random-effects meta-analysis based on Bartlett-type corrections. *Statistics in Medicine* 30, 3304–3312.

Noma, H., K. Nagashima, and T. Furukawa (2019). Permutation inference methods for multivariate meta-analysis. *arXiv:1809.03935*.

Noma, H., K. Nagashima, K. Maruo, M. Gosho, and T. A. Furukawa (2018). Bartlett-type corrections and bootstrap adjustments of likelihood-based inference methods for network meta-analysis. *Statistics in Medicine* 37, 1178–1190.

Reitsma, J. B., A. S. Glas, A. W. S. Rutjes, R. J. P. M. Scholten, P. M. Bossuyt, and A. H. Zwinderman (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 58, 982–990.

Riley, R. D., P. C. Lambert, and G. Abo-Zaid (2010). Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 340, c221.

Salanti, G. (2012). Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research Synthesis Methods* 3, 80–97.

Salanti, G., J. P. Higgins, A. Ades, and J. P. Ioannidis (2008). Evaluation of networks of randomized trials. *Statistical Methods in Medical Research* 17, 279–301.
Sidik, K. and J. N. Jonkman (2007). A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine* 26, 1964–1981.

Sugasawa, S. and H. Noma (2019). A unified method for improved inference in random effect meta-analysis. *Biostatistics to appear*.

Veroniki, A. A., D. Jackson, R. Bender, O. Kuss, D. Langan, J. P. T. Higgins, G. Knapp, and G. Salanti (2019). Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. *Research Synthesis Methods* 10, 23–43.

White, I. R. (2015). Network meta-analysis. *The State Journal* 15, 951–985.

White, I. R., J. K. Barrett, D. Jackson, and J. P. Higgins (2012). Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Research Synthesis Methods* 3, 111–125.

Whitehead, A. and J. Whitehead (1991). A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine* 10, 1665–1677.

Zucker, D. M., O. Lieberman, and O. Manor (2000). Improved small sample inference in the mixed linear model: Bartlett correction and adjusted likelihood. *Journal of the Royal Statistical Society: Series B* 62, 827–838.
Supplementary Material for “Improving the Accuracy of Confidence Intervals and Regions in Multivariate Random-effects Meta-analysis”

Tsubasa Ito¹ and Shonosuke Sugasawa²

¹Department of Statistical Thinking, The Institute of Statistical Mathematics
²Center for Spatial Information Science, The University of Tokyo

This supplementary material provides the proofs and the detailed derivations of Theorem 1, Theorem 2, Equation (7) and Equation (9). In what follows, we denote \( \Sigma_n(\psi) = I_n \otimes \Sigma(\psi) \) and write \( \hat{\beta}(\psi) \) as \( \tilde{\beta} \), \( \hat{\beta}(\hat{\psi}) \) as \( \hat{\beta} \), \( \Sigma(\psi) \) as \( \tilde{\Sigma} \), \( \Sigma_n(\psi) \) as \( \tilde{\Sigma}_n \), \( V(\psi) \) as \( V \) and \( V(\hat{\psi}) \) as \( \hat{V} \), for notational simplicity.

S1 Key lemmas

We first introduce lemmas which play important roles in the proofs of Theorems 1 and Theorem 2. The first lemma is used for deriving the conditional distribution of \( \mu(\hat{\psi}) \) and \( \eta(\hat{\psi}) \).

Lemma S1. Under the conditions (C1)-(C3) given in the main document, \( \tilde{\beta} \) is independent of \( Py \) for \( P = I_n - X(X^tX)^{-1}X^t \). Also, \( \hat{\beta} - \tilde{\beta} \) is a function of \( Py \), and independent of \( \tilde{\beta} \).

Proof. The covariance of \( Py \) and \( \tilde{\beta} \) is

\[
E[Py(\tilde{\beta} - \beta)^t]V^{-1} = E[(y - X\hat{\beta}_{OLS})(y - X\beta)^t](Z\Sigma_nZ^t + S)^{-1}X \\
= [Z\Sigma_nZ^t + S - X(Z\Sigma_nZ^t + S)](Z(I_n \otimes \Sigma)Z^t + S)^{-1}X \\
= 0,
\]

which implies that \( \tilde{\beta} \) is independent of \( Py \) from the normality assumption.
Now, we write $\tilde{\beta}$ as $\tilde{\beta}(\psi, y)$ and $\beta$ as $\tilde{\beta}(\psi(y), y)$. Since $\tilde{\beta}(\psi, y + XT) = \tilde{\beta}(\psi, y) + T$ and $\tilde{\beta}(\psi(y + XT), y + XT) = \tilde{\beta}(\psi(y), y) + T$ from (C3), we have

$$\tilde{\beta}(\psi(y + XT), y + XT) - \tilde{\beta}(\psi, y + XT) = \tilde{\beta}(\psi(y), y) - \tilde{\beta}(\psi, y),$$

which implies that $\tilde{\beta} - \beta$ is invariance with respect to the translation $y \rightarrow y + XT$. Moreover, $Py$ is maximal invariant with respect to the translation $y \rightarrow y + XT$ since $P(y + XT) = Py$ and $P_{y_1} = P_{y_2}$ implies that $y_1 = y_2 + XT'$ for $T' = (X'X)^{-1}X'(y_1 - y_2)$. Then, $\tilde{\beta} - \beta$ is a function of $Py$ from Theorem 2 in [Berger (1985), p.403].

In the next lemma, we show the first order bias of the plug-in estimator $\hat{V}$ is approximately the same as the negative covariance of $\tilde{\beta} - \beta$.

**Lemma S2.** Under the conditions (C1)-(C3), it holds that

$$E[\hat{V}] - V = -E[(\hat{\beta} - \beta)(\tilde{\beta} - \beta)^t] + O(n^{-5/2}).$$

**Proof.** We will show the Lemma by directly comparing both sides of the equation in the Lemma. Noting that $V = \{X'((Z\Sigma_nZ^t + S)X)^{-1} and A^{-1} - B^{-1} = -A^{-1}(A - B)B^{-1}$ for some non-singular matrices $A$ and $B$, we have

$$\hat{V} - V = -\hat{V}X'\{(Z\hat{\Sigma}_nZ^t + S)^{-1} - (Z\Sigma_nZ^t + S)^{-1}\}XV$$

$$= \hat{V}X'(Z\hat{\Sigma}_nZ^t + S)^{-1}Z(\hat{\Sigma}_n - \Sigma_n)Z'(Z\Sigma_nZ^t + S)^{-1}XV$$

$$= VX'(Z\hat{\Sigma}_nZ^t + S)^{-1}Z(\hat{\Sigma}_n - \Sigma_n)Z'(Z\Sigma_nZ^t + S)^{-1}XV$$

$$+ (\hat{V} - V)X'(Z\hat{\Sigma}_nZ^t + S)^{-1}Z(\hat{\Sigma}_n - \Sigma_n)Z'(Z\Sigma_nZ^t + S)^{-1}XV$$

$$\equiv I_1 + I_2.$$  

Since $V = O(n^{-1})$ and $\hat{V} - V = O_p(n^{-1/2})$ from the condition (C2), we have

$$I_1 = VX'(Z\Sigma_nZ^t + S)^{-1}Z(\hat{\Sigma}_n - \Sigma_n)Z'(Z\Sigma_nZ^t + S)^{-1}XV$$

$$- VX'(Z\Sigma_nZ^t + S)^{-1}Z(\hat{\Sigma}_n - \Sigma_n)Z'(Z\Sigma_nZ^t + S)^{-1}Z(\hat{\Sigma}_n - \Sigma_n)Z'(Z\Sigma_nZ^t + S)^{-1}XV$$

$$+ O_p(n^{-5/2}),$$

$$- Vx'(Z\Sigma_nZ^t + S)^{-1}Z(\hat{\Sigma}_n - \Sigma_n)Z'(Z\Sigma_nZ^t + S)^{-1}Z(\hat{\Sigma}_n - \Sigma_n)Z'(Z\Sigma_nZ^t + S)^{-1}XV$$
and

\[ I_2 = V X^t (Z \Sigma_n Z^t + S)^{-1} Z (\hat{\Sigma}_n - \Sigma_n) Z^t (Z \Sigma_n Z^t + S)^{-1} X V X^t \]
\[ \times (Z \Sigma_n Z^t + S)^{-1} Z (\hat{\Sigma}_n - \Sigma_n) Z^t (Z \Sigma_n Z^t + S)^{-1} X V + O_p(n^{-5/2}). \]

Then, for \( R = X^t (Z \Sigma_n Z^t + S)^{-1} Z \) we have

\[
E[\hat{\Sigma} - \Sigma] = E[V R(\hat{\Sigma}_n - \Sigma_n) R^t V] + O(n^{-5/2})
\]
\[
= E[V R(\hat{\Sigma}_n - \Sigma_n) R^t V] + O(n^{-5/2})
\]
\[
= E[V R(\hat{\Sigma}_n - \Sigma_n) R^t V] + O(n^{-5/2}),
\]

(S1)

where the last equality holds since \( \hat{\Sigma} \) is a second-order unbiased estimator of \( \Sigma \).

Next, we evaluate the first term of the right side of the equation in the Lemma.

We can write \( \hat{\beta} - \tilde{\beta} \) as

\[
\hat{\beta} - \tilde{\beta} = (V - V) X^t (Z \hat{\Sigma}_n Z^t + S)^{-1} (y - X \beta)
\]
\[
+ V X^t ((Z \hat{\Sigma}_n Z^t + S)^{-1} - (Z \Sigma_n Z^t + S)^{-1})(y - X \beta)
\]
\[
= J_1 + J_2.
\]

In order to approximate the covariance of \( \hat{\beta} - \tilde{\beta} \) up to the order \( O(n^{-5/2}) \), we expand \( J_1 \) and \( J_2 \) as

\[
J_1 = V R(\hat{\Sigma}_n - \Sigma_n) R^t V X^t (Z \Sigma_n Z^t + S)^{-1} (y - X \beta) + O_p(n^{-1}),
\]
\[
J_2 = - V R(\hat{\Sigma}_n - \Sigma_n) Z^t (Z \Sigma_n Z^t + S)^{-1} (y - X \beta) + O_p(n^{-1}).
\]
The straightforward calculation shows that
\[
E[J_1J_1'] = E[VR(\hat{\Sigma}_n - \Sigma_n)R'R(\hat{\Sigma}_n - \Sigma_n)R'V] + O(n^{-5/2}),
\]
\[
E[J_2J_2'] = E[VR(\hat{\Sigma}_n - \Sigma_n)Z(\Sigma_nZ' + S)^{-1}Z(\hat{\Sigma}_n - \Sigma_n)R'V] + O(n^{-5/2}),
\]
\[
E[J_1J_2'] = E[J_2J_1'] = -E[VR(\hat{\Sigma}_n - \Sigma_n)R'R(\hat{\Sigma}_n - \Sigma_n)R'V] + O(n^{-5/2}),
\]
thereby we have
\[
-E[(\hat{\beta} - \bar{\beta})(\hat{\beta} - \bar{\beta})'] = E[VR(\hat{\Sigma}_n - \Sigma_n)R'R(\hat{\Sigma}_n - \Sigma_n)R'V]
- E[VR(\hat{\Sigma}_n - \Sigma_n)Z(\Sigma_nZ' + S)^{-1}Z(\hat{\Sigma}_n - \Sigma_N)R'V] + O(n^{-5/2}),
\]
which has the same expression as (S1).

S2 Proof of Theorem 1

We denote $H$ as the variance of $\hat{\mu}(\psi)$, that is, $H = c'Vc$. From Lemma S1, the conditional distribution of $c'(\hat{\beta} - \beta)$ given $Py$ is $N_k(c'(\hat{\beta} - \bar{\beta}), H)$. Since $\hat{\psi}$ is a function of $Py$, this implies that
\[
P\left(\frac{c'(\hat{\beta} - \beta)}{H^{1/2}} \leq z\right) = E\left[P\left(\frac{c'(\hat{\beta} - \beta) - c'(\hat{\beta} - \bar{\beta})}{H^{1/2}} \leq \frac{\hat{H}^{1/2}z - c'(\hat{\beta} - \bar{\beta})}{H^{1/2}}\right|Py\right]
= E\left[\Phi\left(\frac{\hat{H}^{1/2}z - c'(\hat{\beta} - \bar{\beta})}{H^{1/2}}\right)\right].
\]

Thus, it is observed that
\[
P\left(c'\hat{\beta} - \hat{H}^{1/2}z \leq c'\beta \leq c'\hat{\beta} + \hat{H}^{1/2}z\right)
= E\left[\Phi\left(\frac{\hat{H}^{1/2}z - c'(\hat{\beta} - \bar{\beta})}{H^{1/2}}\right)\right] - E\left[\Phi\left(-\frac{\hat{H}^{1/2}z - c'(\hat{\beta} - \bar{\beta})}{H^{1/2}}\right)\right]
= E\left[\Phi(r_1) - \Phi(-r_2)\right] = E\left[\Phi(r_1 + r_2) + \Phi(r_1 - r_2)\right] - 1,
for $r_1 = \hat{H}^{1/2} z / H^{1/2}$ and $r_2 = c^t (\hat{\beta} - \tilde{\beta}) / H^{1/2}$. By the Taylor series expansion, for $r_1^* \in (r_1, r_1 + r_2)$ and $r_1^{**} \in (r_1, r_1 - r_2)$, we have

$$\Phi(r_1 + r_2) + \Phi(r_1 - r_2) = 2\Phi(r_1) + r_2^2 \phi^{(1)}(r_1) + \frac{1}{24} r_2^4 (\phi^{(3)}(r_1^*) + \phi^{(3)}(r_1^{**})), \quad (S2)$$

where $\phi^{(1)}(\cdot)$ and $\phi^{(3)}(\cdot)$ are the first and third derivatives of the standard normal density $\phi(\cdot)$. We evaluate the expectation of the first term $2\Phi(r_1)$ in the right side of (S2). By Taylor series expansion, for $z^* \in (z, r_1)$, we have

$$\Phi(r_1) - \Phi(z) = (r_1 - z)\phi(z) + \frac{(r_1 - z)^2}{2} \phi^{(1)}(z) + \frac{(r_1 - z)^3}{6} \phi^{(2)}(z) + \frac{(r_1 - z)^4}{24} \phi^{(1)}(z^*). \quad (S3)$$

Since $\hat{H}^{1/2}$ can be expanded as

$$\hat{H}^{1/2} = H^{1/2} \left( 1 + \frac{\hat{H} - H}{2H} - \frac{(\hat{H} - H)^2}{8H^2} + (1 + c)^{-5/2} \frac{(\hat{H} - H)^3}{16H^3} \right), \quad (S4)$$

for $c \in [0, (\hat{H} - H)/H]$, we can obtain the following expression of $E[r_1]$:

$$E[r_1] = z \left\{ 1 + \frac{E[\hat{H} - H]}{2H} - \frac{E[(\hat{H} - H)^2]}{8H^2} + \frac{E[(1 + c)^{-5/2}(\hat{H} - H)^3]}{16H^3} \right\}. \quad (S5)$$

Since $\hat{H} - H = O_p(n^{-3/2})$, the forth term in the right side of the above equation is $O(n^{-3/2})$. Then, for $A = E[(\hat{H} - H)^2]$ we have

$$E[r_1] - z = \left( \frac{E[\hat{H} - H]}{2H} - \frac{A}{8H^2} \right) z + O(n^{-3/2}). \quad (S6)$$

Moreover, since $E[(r_1 - z)^2] = E[r_1^2] - 2zE[r_1] - z^2$, it is observed that

$$E[(r_1 - z)^2] = A \frac{z^2}{4H^2} + O(n^{-3/2}), \quad (S7)$$

which implies that $r_1 - z = O_p(n^{-1/2})$. Then, the expectation of the third and forth terms in the right side of (S3) is $O(n^{-3/2})$.

We next evaluate the expectation of the second term in the right side of (S2).
Since $E[r_2^2] = O(n^{-1})$ from Lemma S2 and $E[r_1] - z = O(n^{-1/2})$, we have

$$E[r_2^2 \phi(1)(r_1)] = E \left[ \frac{c'(\hat{\beta} - \beta)(\hat{\beta} - \beta)'c}{H} \right] \phi(1)(z) + O(n^{-3/2}).$$  \hfill (S8)

Since $r_2^2 = O_p(n^{-2})$ from Lemma S2 and $E[r_1^2] - z = O(1)$, the expectation of the third term in the right side of (S2) is $O(n^{-2})$.

Combining (S2), (S3), (S6), (S7) and (S8) gives the following asymptotic expansion of the coverage probability of the naive confidence interval:

$$P \left( c'\beta - H^{1/2}z \leq c'\beta \leq c'\beta + \hat{H}^{1/2}z \right) = 2\Phi(z) + \frac{A}{4H^2} \left( \frac{E[H - H]}{H} + \frac{E[c'(\hat{\beta} - \beta)(\hat{\beta} - \beta)'c]}{H} \right) z\phi(z) + O(n^{-3/2})$$

where the last equality holds from Lemma S2 and the fact that $\phi^{(1)}(z) = -z\phi(z)$.

Note that the third term in the above expression is $O(n^{-1})$ since $A = O(n^{-3})$ and $A/H^2 = O(n^{-1})$. In order to calibrate the suitable $z$ under the nominal level $1 - \alpha$, we consider the following equation:

$$2\Phi(z) - 1 - \frac{A}{4H^2} (z^3 + z)\phi(z) = 1 - \alpha,$$

which gives an approximate solution $z^*$ of $z$ given by $z^* = z_{\alpha/2} + (z_{\alpha/2}^3 + z_{\alpha/2}) A/(8H^2)$.

Therefore, the improved confidence interval with coverage probability $1 - \alpha + o(n^{-1})$ is given by (2), which completes the proof.

**S3  Proof of Theorem 2**

We denote $H$ as the covariance matrix of $\eta(\psi)$, that is, $H = CVC^t$. Note that we reuse the same notation to denote the different covariance from that of $\mu(\psi)$ in the proof of Thorem 1. From Lemma S1, the conditional distribution of $C(\hat{\beta} - \beta)$ given $Py$ is $N_k(C(\hat{\beta} - \beta), H)$. Let $w = H^{-1/2}\{C(\hat{\beta} - \beta) - C(\hat{\beta} - \beta)\}$. It is noted that $\hat{H} - H = \ldots$
because the Mahalanobis’ distance is approximated via Taylor series expansion as

\[(\hat{\beta} - \beta)^\dagger C^\dagger \hat{H}^{-1}C(\hat{\beta} - \beta)\]

\[= w^\dagger H^{1/2} \hat{H}^{-1}H^{1/2}w + 2(\hat{\beta} - \beta)^\dagger C^\dagger \hat{H}^{-1}H^{1/2}w + (\hat{\beta} - \beta)^\dagger C^\dagger H^{-1}C(\hat{\beta} - \beta)\]

\[= w^\dagger \left[ I_k - H^{-1/2}(\hat{H} - H)H^{-1/2} + H^{-1/2}(\hat{H} - H)H^{-1}(\hat{H} - H)H^{-1/2} \right] w\]

\[+ 2(\hat{\beta} - \beta)^\dagger C^\dagger \hat{H}^{-1}H^{1/2}w + (\hat{\beta} - \beta)^\dagger C^\dagger H^{-1}C(\hat{\beta} - \beta) + O_p(n^{-3/2})\]

\[= w^\dagger (I_k - G_1)w + 2g_2^\dagger w + g_3 + O_p(n^{-3/2}), \tag{S9}\]

where

\[G_1 = H^{-1/2}(\hat{H} - H)H^{-1/2} - H^{-1/2}(\hat{H} - H)H^{-1}(\hat{H} - H)H^{-1/2},\]

\[g_2 = H^{1/2} \hat{H}^{-1}C(\hat{\beta} - \beta),\]

\[g_3 = (\hat{\beta} - \beta)^\dagger C^\dagger \hat{H}^{-1}C(\hat{\beta} - \beta).\]

From [S9], the characteristic function \(\varphi(t) = \mathbb{E}[\exp\{it(\hat{\beta} - \beta)^\dagger C^\dagger \hat{H}^{-1}C(\hat{\beta} - \beta)\}]\) is approximated as

\[\varphi(t) = \mathbb{E}\exp\left(it\{w^\dagger (I_k - G_1)w + 2g_2^\dagger w + g_3\}\right) + O(n^{-3/2})\]

\[= \mathbb{E}\left[e^{itw^\dagger w}\left\{1 + it\{-w^\dagger G_1w + 2g_2^\dagger w + g_3\} - \frac{t^2}{2}\{-w^\dagger G_1w + 2g_2^\dagger w + g_3\}^2\}\right\} + O(n^{-3/2})\]

\[= \mathbb{E}\left[e^{itw^\dagger w}\left\{1 + it\{-w^\dagger G_1w + 2g_2^\dagger w + g_3\} - \frac{t^2}{2}\{(w^\dagger G_1w)^2 + 4w^\dagger g_2g_2^\dagger w - 4w^\dagger G_1wg_2^\dagger w\}\}\right\} + O(n^{-3/2}),\]

because \(G_1 = O_p(n^{-1/2}), g_2 = O_p(n^{-1/2})\) and \(g_3 = O_p(n^{-1})\). From the law of iterated expectations and the conditional normality of \(w\), the above equation reduces to

\[\varphi(t) = \mathbb{E}\left[e^{itw^\dagger w}\left\{1 + it\{-w^\dagger G_1w + g_3\} - \frac{t^2}{2}\{(w^\dagger G_1w)^2 + 4w^\dagger g_2g_2^\dagger w\}\}\right\} + O(n^{-3/2}).\]
For some deterministic matrix $A$ and $w \sim N_k(0, I_k)$, it holds that

$$
E\left[e^{itw^tAw}\right] = (2\pi)^{-k/2} \int \exp \left(-\frac{(1 - 2it)w^tAw}{2}\right) w^tAw \, dw
$$

$$
= (1 - 2it)^{-k/2 - 1} \text{tr}(A),
$$

$$
E\left[e^{itw^tA(w^tA)^2}\right] = (2\pi)^{-k/2} \int \exp \left(-\frac{(1 - 2it)w^tAw}{2}\right) (w^tAw)^2 \, dw
$$

$$
= (1 - 2it)^{-k/2 - 2}(\text{tr}^2(A) + 2\text{tr}(A^2)).
$$

Using these equalities, from the law of iterated expectations, we have

$$
\phi(t) = (1 - 2it)^{-k/2} \left[1 + it\left\{ - (1 - 2it)^{-1}\text{tr}(E[G_1]) + E[g_3] \right\} \right.
$$

$$
+ \frac{(it)^2}{2} \left\{ (1 - 2it)^{-2}\{E[\text{tr}^2(G_1)] + 2\text{tr}(E[G_1^2])\} \right. \right.
$$

$$
\left. \left. + (1 - 2it)^{-4}\text{tr}(E[g_2g_2'g_2]) \right\} \right\} + O(n^{-3/2}).
$$

For notational simplicity, let $J = E[\text{tr}^2(G_1)] + 2\text{tr}(E[G_1^2])$. Let $s = (1 - 2it)^{-1}$, or $it = (s - 1)/(2s)$. Then, $(1 - 2it)^{k/2}\phi(t) - 1$ can be written as

$$
\frac{it}{2s} \left\{ E[g_2g_2'] - E[g_3] \right\} + \left\{ \frac{1}{2}\text{tr}(E[G_1]) + \frac{1}{2}E[g_3] + \frac{J}{8} - E[g_2g_2'] \right\}
$$

$$
+ \left\{ - \frac{1}{2}\text{tr}(E[G_1]) - \frac{J}{4} + \frac{1}{2}E[g_2g_2'] \right\} s + \frac{J}{8}s^2
$$

We shall evaluate the moments in (S10). First, $G_1$ can be expressed as

$$
G_1 = H^{-1/2}(\hat{H} - H)H^{-1/2} - H^{-1/2}(\hat{H} - H)H^{-1}(\hat{H} - H)H^{-1/2},
$$

thereby it holds that

$$
\text{tr}(E[G_1]) = \text{tr}(E[K]) - \text{tr}(E[K^2]),
$$

for $K = H^{-1/2}(\hat{H} - H)H^{-1/2}$. Noting that the first term in (S11) is $O_p(n^{-1/2})$ and
the second term is $O(n^{-1})$, we can expand $G_1^2$ and $\text{tr}^2(G_1)$ as

\[
G_1^2 = H^{-1/2}(\hat{H} - H)H^{-1}(\hat{H} - H)H^{-1/2} + O_p(n^{-3/2}),
\]
\[
\text{tr}^2(G_1) = \text{tr}^2(H^{-1/2}(\hat{H} - H)H^{-1/2}) + O_p(n^{-3/2}),
\]

which lead to $E[G_1^2] = E[K^2] + O(n^{-3/2})$ and $E[\text{tr}^2(G_1)] = E[\text{tr}^2(K)] + O(n^{-3/2})$. Thus,

\[
J = E[\text{tr}^2(K)] + 2\text{tr}(E[K^2]) + O(n^{-3/2}). \tag{S13}
\]

It can be also observed that

\[
g_2^t g_2 = (\hat{\beta} - \tilde{\beta})^t C^t \hat{H}^{-1} C (\hat{\beta} - \tilde{\beta}) = (\hat{\beta} - \tilde{\beta})^t C^t \hat{H}^{-1} C (\hat{\beta} - \tilde{\beta}) + O_p(n^{-3/2}),
\]
\[
g_3 = (\hat{\beta} - \tilde{\beta})^t C^t \hat{H}^{-1} C (\hat{\beta} - \tilde{\beta}) = (\hat{\beta} - \tilde{\beta})^t C^t \hat{H}^{-1} C (\hat{\beta} - \tilde{\beta}) + O_p(n^{-3/2}).
\]

Then, from Lemma [S2] we have

\[
E[g_2^t g_2] = E[g_3] = -\text{tr}(E[K]) + O(n^{-3/2}). \tag{S14}
\]

Combining (S12), (S13) and (S14), we can see that the characteristic function of $(\hat{\beta} - \tilde{\beta})^t C^t \hat{H}^{-1} C (\hat{\beta} - \tilde{\beta})$ can be written as

\[
\varphi(t) = (1 - 2it)^{-k/2} \{1 + B_1/8 - B_2/4 + B_3 + (-B_1/4 - B_3)s + (B_1/8 + B_2/4)s^2\}
\]
\[
+ O(n^{-3/2}),
\]

for $B_1 = B_1(\psi)$, $B_2 = B_2(\psi)$ and $B_3 = B_3(\psi)$. From the fact that the characteristic function of the chi-squared distribution with degrees of freedom $k + 2h$ is given by $(1 - 2it)^{-k/2-h} = (1 - 2it)^{-k/2} s^h$, it follows that the asymptotic expansion of the cumulative distribution function of $(\hat{\beta} - \tilde{\beta})^t C^t \hat{H}^{-1} C (\hat{\beta} - \tilde{\beta})$ is

\[
F_k(x) + (B_1/8 - B_2/4 + B_3)F_k(x)
\]
\[
+ (-B_1/4 - B_3)F_{k+2}(x) + (B_1/8 + B_2/4)F_{k+4}(x) + O(n^{-3/2}),
\]

9
where $F_k(x)$ is the cumulative distribution function of the chi-squared distribution with degrees of freedom $k$. Note that $F_{k+r-2}(x) - F_{k+r}(x) = 2f_{k+r}(x)$, where $f_k(x)$ is the density function of the chi-squared distribution with degrees of freedom $k$. Then, it holds that

$$P((\hat{\beta} - \tilde{\beta})^tC\hat{H}^{-1}C(\hat{\beta} - \tilde{\beta}) \leq x)$$

$$= F_k(x) + 2\left(\frac{B_1}{8} - \frac{B_2}{4} + B_3\right) f_{k+2}(x) - \left(\frac{B_1}{4} + \frac{B_2}{2}\right) f_{k+4}(x) + O(n^{-3/2}),$$

which enables us to carry out the Bartlett-type correction. For a function $h = h(\psi)$ with order $O(n^{-1})$, we have

$$P\{((\hat{\beta} - \beta)^tC\hat{H}^{-1}C(\hat{\beta} - \beta) \leq x(1 + h)\}$$

$$= F_k(x) + hxf_k(x) + \left(\frac{B_1}{4} - \frac{B_2}{2} + 2B_3\right) f_{k+2}(x) - \left(\frac{B_1}{4} + \frac{B_2}{2}\right) f_{k+4}(x) + O(n^{-3/2}),$$

Note that the last three terms are $O(n^{-1})$. Thus, the coverage error of the second-order can be corrected if

$$hxf_k(x) + (B_1/4 - B_2/2 + 2B_3)f_{k+2}(x) - (B_1/4 + B_2/2)f_{k+4}(x) = 0.$$

(S15)

Since $\Gamma(x+1) = x\Gamma(x)$ for the gamma function $\Gamma(x)$, the approximate solution of the above equation on $h$ is

$$h^* = -(B_1/4 - B_2/2 + 2B_3)/k + (B_1/4 + B_2/2)x/k(k+2).$$

(S16)

For $h^*$ given in (S16), it holds that for any $x > 0$,

$$P\{(1 + h^*)^{-1}(\hat{\beta} - \beta)^tC\hat{H}^{-1}C(\hat{\beta} - \beta) \leq x\} = F_k(x) + O(n^{-3/2}),$$

which completes the proof.
Derivation of Equation (7)

We write functions given in Section 2 as functions of $\Sigma$ since the unknown parameter is $\Sigma$ in this example. For $H = \left( \sum_{i=1}^{n} D_i^{-1} \right)^{-1}$ and $D_i = \Sigma + S_i$, $\hat{H} - H$ can be expanded as

$$\hat{H} - H = H \left\{ \sum_{i=1}^{n} D_i^{-1} (\hat{\Sigma} - \Sigma) D_i^{-1} \right\} H + O_p(n^{-2}).$$

Since the first term on the right side of the above equation is of order $O_p(n^{-3/2})$, we only need to consider this term to evaluate the quantities (3).

At first, we evaluate $B_1^*$. It is noted that we have

$$H^{-1/2} (\hat{H} - H) H^{-1/2} = \frac{1}{n} \sum_{j=1}^{n} H^{1/2} \left\{ \sum_{i=1}^{n} D_i^{-1} D_j^{1/2} \{ u_j u'_j - I_p \} D_j^{1/2} D_i^{-1} \right\} H^{1/2} + O_p(n^{-1}),$$

where $u_j$ are independently distributed as the standard normal distribution. Then, we have

$$B_1^* = E \left[ \text{tr}^2 \left\{ H^{-1/2} (\hat{H} - H) H^{-1/2} \right\} \right]$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} E \left[ \text{tr} \left\{ HD_j^{-1/2} D_i^{1/2} \{ u_j u'_j - I_p \} (\Sigma + S_i)^{1/2} D_j^{-1} \right\} \right.$$

$$\times \left. \text{tr} \left\{ HD_k^{-1/2} D_i^{1/2} \{ u_k u'_k - I_p \} (\Sigma + S_i)^{1/2} D_k^{-1} \right\} \right]$$

$$= \frac{2}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \text{tr} \left[ D_i^{1/2} D_j^{-1/2} H D_j^{-1} D_i^{1/2} D_j^{1/2} D_i^{-1} D_k^{-1} H D_k^{-1} D_i^{1/2} \right]. \ \ \ \ (S17)$$

Next, we evaluate $B_2^*$. It is noted that we have

$$(\hat{H} - H) H^{-1} (\hat{H} - H) H^{-1}$$

$$= H \left\{ \sum_{i=1}^{n} D_i^{-1} (\hat{\Sigma} - \Sigma) D_i^{-1} \right\} H \left\{ \sum_{i=1}^{n} D_i^{-1} (\hat{\Sigma} - \Sigma) D_i^{-1} \right\} + O_p(n^{-3/2}).$$
and that $\hat{\Sigma} - \Sigma$ can be written as

$$
\hat{\Sigma} - \Sigma = \frac{1}{n} \sum_{i=1}^{n} \{(y_i - \beta)(y_i - \beta)^\intercal - \Sigma - S_i\} + O_p(n^{-1}).
$$

Then, for $u_i$ for $i = 1, \ldots, n$ which are independently distributed as the multivariate standard normal distribution, it holds that for $\ell, m = 1, \ldots, n$,

$$
E \left[ (\hat{\Sigma} - \Sigma)D_{\ell}^{-1/2}H D_m^{-1}(\hat{\Sigma} - \Sigma) \right] \\
= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} D_i^{1/2} E \left[ (u_i u_j^\intercal - I_p)(\Sigma + S_i)^{1/2}D_{\ell}^{-1/2}H D_m^{-1}D_j^{1/2}(u_j u_i^\intercal - I_p) \right] D_j^{1/2} \\
= \frac{1}{n^2} \sum_{i=1}^{n} D_i^{1/2} E \left[ (u_i u_i^\intercal - I_p)(\Sigma + S_i)^{1/2}D_{\ell}^{-1/2}H(\Sigma)D_m^{-1/2}D_i^{1/2}(u_i u_i^\intercal - I_p) \right] D_i^{1/2} \\
= \frac{1}{n^2} \sum_{i=1}^{n} D_i^{1/2}(L_{i\ell m} + \text{tr}(L_{i\ell m})I_p)D_i^{1/2}
$$

for $L_{i\ell m} = D_i^{1/2}D_{\ell}^{-1/2}H D_m^{-1}D_i^{1/2}$. Then, we have

$$
B_2^* = E[\text{tr}((\hat{H} - H)H^{-1})^2] \\
= \frac{1}{n^2} \sum_{i,\ell,m=1}^{n} \text{tr} \left( D_i^{-1/2}D_{\ell}^{-1/2}L_{i\ell m}D_m^{1/2}D_m^{-1}H \right) + \frac{1}{n^2} \sum_{i,\ell,m=1}^{n} \text{tr}(L_{i\ell m})\text{tr} \left( D_i^{-1}D_iD_m^{-1}H \right) \\
= \frac{1}{n^2} \sum_{i=1}^{n} \text{tr} \left\{ H \sum_{j=1}^{n} \left( D_j^{-1}D_iD_j^{-1} \right)^2 \right\} + \frac{1}{n^2} \sum_{i,\ell,m=1}^{n} \text{tr}^2 \left( D_i^{-1}D_iD_m^{-1}H(\Sigma) \right).
$$

(S18)

Finally, we evaluate $B_3^*$. From the equation (S1), for $V = (\sum_{i=1}^{n} D_i^{-1} - 1)^{-1}$ we have

$$
E[K] = E \left[ V^{1/2} \left\{ \sum_{i=1}^{n} D_i^{-1}(\hat{\Sigma} - \Sigma)D_i^{-1} \right\} V^{1/2} \right] \\
- V^{1/2} \left\{ \sum_{i=1}^{n} D_i^{-1}(\hat{\Sigma} - \Sigma)D_i^{-1}(\hat{\Sigma} - \Sigma)D_i^{-1} \right\} V^{1/2} \\
+ O(n^{-3/2}).
$$

The trace of the first term in the above equation is exactly the same with $B_2^*$ and is
given in (S18). To evaluate the second term, it is noted that

\[
\begin{align*}
\mathbb{E}\left[ \sum_{i=1}^{n} D_i^{-1}(\hat{\Sigma} - \Sigma)D_i^{-1}(\hat{\Sigma} - \Sigma)D_i^{-1} \right] \\
= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} D_i^{-1} D_j^{-1} D_i D_j^{-1} D_i + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \text{tr}(D_i^{-1} D_j) D_i^{-1} D_i D_j^{-1}.
\end{align*}
\]

Then, the trace of the second term in the above equation is given by

\[
-\text{tr}\left( \mathbb{E}\left[ V^{1/2} \left\{ \sum_{i=1}^{n} D_i^{-1}(\hat{\Sigma} - \Sigma)D_i^{-1}(\hat{\Sigma} - \Sigma)D_i^{-1} \right\} V^{1/2} \right] \right)
= -\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \text{tr}\left( V D_i^{-1} D_j^{-1} D_i D_j^{-1} \right) - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \text{tr}\left( D_i^{-1} D_j \right) \text{tr}\left( V D_i^{-1} D_j D_i^{-1} \right).
\]

(S19)

Equation (2.2), (S17), (S18) and (S19) lead to the expression in the equation (7).

**S5 Derivation of (9)**

We first note that \( \hat{\tau}^2 - \tau^2 \) has the same asymptotic expansion with \( \hat{\tau}_0^2 - \tau^2 \), so that it suffices to consider \( \hat{\tau}_0^2 \). Remember that \( A = \mathcal{O}E[(\hat{V} - V)^2]c \). We derive the approximation \( A^* \) of \( A \) satisfying \( A^* = A + o(n^{-1}) \). It is noted that for \( \Sigma = \tau^2 P(0.5) \) we have

\[
\hat{H} - H = \mathcal{O}(VR(\hat{\Sigma}_n - \Sigma_n))R^tV)c + O_p(n^{-2}),
\]

and \( \hat{\Sigma} - \Sigma = (\hat{\tau}_0^2 - \tau^2)P(0.5) \). Then, we only need to evaluate the second moment of \( \hat{\tau}_0^2 - \tau^2 \). Since \( \hat{\tau}_0^2 - \tau^2 \) can be expanded as

\[
\hat{\tau}_0^2 - \tau^2 = \text{tr}^{-1}(Q) \text{tr} \left[ Q \left\{ (y - X\hat{\beta}_{\text{OLS}})(y - X\hat{\beta}_{\text{OLS}})^t - \tau^2 Q - S \right\} \right]
= \text{tr}^{-1}(Q^2) \text{tr} \left[ Q \left\{ (y - X\beta)(y - X\beta)^t - \tau^2 Q - S \right\} \right] + O_p(n^{-1/2}),
\]

13
for $\tau_0^2$ in (8), we have

$$E[(\hat{\tau}_0^2 - \tau^2)^2] = \text{tr}^{-2}(Q^2)E\left[\text{tr}^2 \left( Q \left\{ (y - X\beta)(y - X\beta)^t - \tau^2 Q - S \right\} \right) \right] + O(n^{-3/2})$$

$$= 2\text{tr}^{-2}(Q^2)\text{tr}^2 \left[ (Q \left\{ \tau^2 Q + S \right\})^2 \right] + O(n^{-3/2}),$$

which leads to the expression in the equation (9).

**References**

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, Springer, New York.