Kenward-Roger–type corrections for inference methods of network meta-analysis and meta-regression

Hisashi Noma, PhD*.†
Department of Data Science, The Institute of Statistical Mathematics, Tokyo, Japan
ORCID: http://orcid.org/0000-0002-2520-9949

Yasuyuki Hamura, PhD †
Graduate School of Economics, Kyoto University, Kyoto, Japan

Masahiko Gosho, PhD
Department of Biostatistics, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan

Toshi A. Furukawa, MD, PhD
Departments of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan

*Corresponding author: Hisashi Noma
Department of Data Science, The Institute of Statistical Mathematics
10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan
TEL: +81-50-5533-8440
e-mail: noma@ism.ac.jp

† The two authors equally contributed to this work.
Abstract

Network meta-analysis has been an essential methodology of systematic reviews for comparative effectiveness research. The restricted maximum likelihood (REML) method is one of the current standard inference methods for multivariate, contrast-based meta-analysis models, but recent studies have revealed the resultant confidence intervals of average treatment effect parameters in random-effects models can seriously underestimate statistical errors; i.e., the actual coverage probability of a true parameter cannot retain the nominal level (e.g., 95%). In this article, we provided improved inference methods for the network meta-analysis and meta-regression models using higher-order asymptotic approximations based on the approach of Kenward and Roger (Biometrics 1997;53:983–997). We provided two corrected covariance matrix estimators for the REML estimator and improved approximations for its sample distribution using a $t$-distribution with adequate degrees of freedom. All of the proposed procedures can be implemented using only simple matrix calculations. In simulation studies under various settings, the REML-based Wald-type confidence intervals seriously underestimated the statistical errors, especially in cases of small numbers of trials meta-analyzed. By contrast, the proposed Kenward-Roger–type inference methods consistently showed accurate coverage properties under all the settings considered in our experiments. We also illustrated the effectiveness of the proposed methods through applications to two real network meta-analysis datasets.

Key words: network meta-analysis; meta-regression; Kenward-Roger–type corrections; higher-order approximations; inconsistency.
1. Introduction

Meta-analysis has played an important role in evidence-based medicine to assemble and summarize evidence from multiple clinical trials. One limitation of the conventional methodology is that it is restricted to direct pairwise comparisons and cannot be used for simultaneous comparisons of multiple available treatments. To overcome this limitation, network meta-analysis has been developed that can synthesize both direct and indirect evidence for all treatment comparisons of interest, and provide estimates of comparative efficacies among treatments, even when no direct comparison evidence exists for some of the included treatments.\textsuperscript{1,2}

In network meta-analyses,\textsuperscript{, heterogeneity in effect sizes from different studies is a common relevant issue as in conventional pairwise meta-analyses and random-effects models are widely adopted to address it.\textsuperscript{3} In pairwise meta-analyses, the validity of standard statistical inference methods for average treatment effect parameter in random-effects models (e.g., the DerSimonian-Laird method)\textsuperscript{4} is well known to be violated under practical situations (e.g., coverage probabilities of confidence intervals cannot retain their nominal confidence levels) because most of these methods depend on large sample approximations for the number of synthesized trials.\textsuperscript{5-7} Recent studies have revealed that the same problem can occur in network meta-analysis when the number of trials is moderate or small and that this invalidity can seriously influence the overall conclusions of network meta-analyses.\textsuperscript{8-10}

To address this issue, several alternative improved methods have been developed, but most of these methods adopt Monte Carlo inferences and require enormous computational burdens.\textsuperscript{8-10} In modern computational environments, although the computational costs are not so problematic, analytical approaches would be alternative effective options that are easy-to-handle for practitioners and do not require heavy
computational burdens. In this article, we propose Kenward-Roger–type improved inference methods for network meta-analysis and meta-regression. The Kenward-Roger methods have been standard inference methods for linear mixed-effects models that improve the conventional restricted maximum likelihood (REML) method, and rich numerical evidence supporting their relatively preferable performances in simulation studies under various settings has been reported. We have derived standard error estimators, confidence intervals, and significance tests for the random-effects models of network meta-analysis using the same principle of Kenward and Roger. The proposed methods can be implemented via simple matrix computations if the REML estimates of the model parameters are obtained. Also, simulation studies have demonstrated that the proposed methods have preferable performances compared with the current standard methods of network meta-analysis under various conditions. We also illustrate the effectiveness of the proposed methods by applying them to two real data examples of network meta-analyses reported by Siontis et al. and Ades et al. This numerical evidence clearly indicates that the current standard methods can markedly underestimate uncertainties of the statistical inferences, and alternative improved methods would be recommended in practice.

2. Methods

2.1 Random-effects model for network meta-analysis and meta-regression

We consider the contrast-based random-effects model, which is one of the representative approaches for network meta-analysis. We consider the synthesis of $N$ trials and a comparison of $p + 1$ treatments. Let $Y_{ij}$ denote an estimator of a treatment effect in contrast to a reference treatment (e.g., placebo) for the $j$th treatment in the $i$th trial ($i = 1, 2, ..., N; j = 1, 2, ..., p$). Commonly used effect measures are mean difference,
standardized mean difference, risk difference, risk ratio, odds ratio, and hazard ratio; the ratio measures are usually transformed on a logarithmic scale. For the contrast-based network meta-regression, we consider a multivariate meta-analysis model for the outcome variable $Y_i = (Y_{i1}, Y_{i2}, ..., Y_{ip})^T$,

$$Y_i = \theta_i + e_i$$

(*)

$$\theta_i = X_i^T \beta + \epsilon_i$$

where $\theta_i = (\theta_{i1}, \theta_{i2}, ..., \theta_{ip})^T$ and $\beta = (\beta_1^T, ..., \beta_p^T)^T$. The regression function model $X_i^T \beta$ involves a design matrix,

$$X_i = \begin{pmatrix} x_{i1} & 0 & \cdots & 0 \\ 0 & x_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{ip} \end{pmatrix}$$

where $x_{ij}$ is a $q_j \times 1$ covariate vector for $Y_{ij}$ ($q_j$ is the number of the covariates) and $\beta_j$ is its $q_j \times 1$ regression coefficient vector. For standard network meta-analysis without modeling any covariates, $x_{ij} = (1)$, a vector of ones. Also, $e_i$ and $\epsilon_i$ are independent random variation terms within and across studies ($p \times 1$ random vectors), assumed to be distributed $e_i \sim \text{MVN}(0, S_i)$ and $\epsilon_i \sim \text{MVN}(0, \Sigma)$. $S_i$ (a $p \times p$ matrix) is the within-study covariance matrix,

$$S_i = \begin{pmatrix} s_{i1}^2 & \rho_{i12}s_{i1}s_{i2} & \cdots & \rho_{i1p}s_{i1}s_{ip} \\ \rho_{i21}s_{i2}s_{i1} & s_{i2}^2 & \cdots & \rho_{i2p}s_{i2}s_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{ip1}s_{ip}s_{i1} & \rho_{ip2}s_{ip}s_{i2} & \cdots & s_{ip}^2 \end{pmatrix}$$

which is usually assumed to be known and fixed to its valid estimate; several suitable estimators that can be obtained from summary statistics have been proposed. In addition, $\Sigma$ is the between-studies variance–covariance matrix:
for $\tau^2 > 0$. Note that the correlation structure of $\Sigma$ can be assumed to be unstructured, but there are rarely a sufficient number of studies to identify all of the variance–covariance parameters. Thus, most network meta-analyses adopt the equal-variance assumption for the $p$ components of $\varepsilon_i$; then, all the pairwise correlation coefficients should be equal to 0.50 because of the consistency assumption \textsuperscript{19,20}. In the present work, we adopt the equal-variance assumption as a standard assumption of this model, similar to the approach of Jackson et al. \textsuperscript{21}, consistently. For trials that do not include a reference treatment, the data augmentation approach of White et al. \textsuperscript{22} can be adopted; a quasi-small dataset is added into the reference arm (e.g., 0.001 events for 0.01 patients for a binary outcome). Also, most individual clinical trials typically involve only two to three or four arms; thus, many of the components of $Y_i$ and $X_i$ are usually undefined. We formally regard these undefined components as “missing.” The frequentist-based direct likelihood inference methods provide valid results, supported by extensive theoretical and numerical evidence \textsuperscript{8-10}. We denote the inverse of the marginal covariance matrix of $Y_i$ as $W_i = (\Sigma + S_i)^{-1}$.

### 2.2 Standard inference methods

To estimate the model parameters, the REML estimation is the most widely adopted method in practice. The log-likelihood function of the multivariate meta-regression model is given as

$$
\ell(\beta, \tau^2) = \text{const} \cdot -\frac{1}{2} \sum_{i=1}^{N} \left\{ \log(\det(W_i^{-1})) + (y_i - X_i^T \beta)^T W_i (y_i - X_i^T \beta) \right\}
$$
where $Y_i$, $X_i$, and $S_i$ involve missing components, as previously mentioned. In the log-likelihood function, these are shrunk to the sub-vector and sub-matrix for the observed components. The log-restricted likelihood function is then written as

$$\ell_{RL}(\beta, \tau^2) = \text{const.} + \ell(\beta, \tau^2) - \frac{1}{2} \log \left( \det \left( \sum_{i=1}^{N} X_i W_i X_i^T \right) \right)$$

The REML estimators of $\{\hat{\beta}, \hat{\tau}^2\}$ are obtained by maximizing $\ell_{RL}(\beta, \tau^2)$ \cite{16, 23}. We denote the covariance matrix of $\hat{\beta}$ as $\Phi$, and the ordinary estimator is given as $\hat{\Phi} = \left( \sum_{i=1}^{N} X_i \tilde{W}_i X_i^T \right)^{-1}$, where $\tilde{W}_i = (\tilde{\Sigma} + S_i)^{-1}$. The resultant REML-based Wald-type tests and confidence intervals are assured to be valid and efficient under large sample settings. Although several other alternative effective inference methods involving the method-of-moment are available, the REML-based methods are well known to have favorable properties in general \cite{8, 9}. The validity and optimality of the REML-based inference methods are assured under large sample settings; however, as shown in the simulation studies in Section 3, they can be violated under realistic, moderate sample settings. To resolve the invalidity problem, we provide improved inference methods based on higher-order asymptotic approximations.

### 2.3 Corrected covariance matrix estimators

Kenward and Roger \cite{11} attempted to improve the approximation of the covariance matrix of regression coefficients estimator under finite sample settings via higher-order asymptotic approximations. The main motivation for the improvement was that the ordinary asymptotic covariance estimator $\hat{\Phi}$ underestimates the actual variability of the REML estimator $\hat{\beta}$ because the uncertainty caused by the variability of the variance component estimator $\hat{\tau}^2$ is not considered; thus, the underestimation bias should be corrected. That is, the asymptotic covariance matrix of the $\hat{\beta}$ is decomposed into two
Kenward and Roger \textsuperscript{11} derived adequate estimators of the adjustment component $\Lambda$ on the basis of the Kackar-Harville-type approximation \textsuperscript{24} and provided an adjusted covariance matrix estimator for linear mixed models.

Here, we apply the approach of Kenward and Roger \textsuperscript{11} to the multivariate random-effects model of network meta-analysis and derive the adjusted covariance matrix estimator. First, following the theoretical results by Kacker and Harville \textsuperscript{24}, we can obtain an estimator of the correction term $\Phi$:

$$\hat{\Phi} = 2\Phi\left(\tilde{\omega}(M_{xpx} - M_{xpx}M_{xx}^{-1}M_{xpx})\right)\Phi$$

where

$$M_{xx} = \sum_{i=1}^{N} X_iW_iX_i^T, M_{xpx} = \sum_{i=1}^{N} X_iW_iPW_iX_i^T,$$

$$M_{xppx} = \sum_{i=1}^{N} X_iW_iPW_iPW_iX_i^T$$

We denote the variance of $\hat{\tau}^2$ as $\omega$, and $\tilde{\omega}$ is an adequate estimator of $\omega$.

For the variance estimator $\hat{\omega}$, two types of estimators are available: the inverses of the expected and observed information. These quantities are defined by the log-likelihood function

$$\ell(\tau^2) = -\frac{1}{2} \left( \sum_{i=1}^{N} \log|W_i^{-1}| + \log|M_{xx}| + M_{yy} - M_{yx}M_{xx}^{-1}M_{xy} \right)$$

where

$$M_{yy} = \sum_{i=1}^{N} Y_i^TW_iY_i, M_{xy} = \sum_{i=1}^{N} X_iW_iY_i$$

and the expected information and observed information are respectively given as
\[ I_E(\tau^2) = -E \left[ \frac{\partial^2 \ell(\tau^2)}{\partial \tau^4} \right], \quad I_O(\tau^2) = -\frac{\partial^2 \ell(\tau^2)}{\partial \tau^4} \]

We can obtain two variance estimators \( \hat{\sigma}_E = I_E(\tau^2)^{-1} \) and \( \hat{\sigma}_O = I_O(\tau^2)^{-1} \). Through several matrix calculations, the two quantities are provided as

\[
I_E(\tau^2) = \frac{1}{2} \left\{ \sum_{i=1}^{N} \text{tr}(W_iPW_iP) - \text{tr}\left(2M^{-1}_xM_{xpp} - M^{-1}_{xpp}M_{xpx}M^{-1}_{xpx}\right) \right\}
\]

\[
I_O(\tau^2) = \frac{1}{2} \left( -\sum_{i=1}^{N} \text{sum}\left\{(W_iPW_i) \circ P\right\} \right.
\]

\[
+ \text{sum}\left\{\left(\left(-M^{-1}_{xpp}M_{xpx}\right) \circ M_{xpx}\right) + \left(M^{-1}_x \circ (2M_{xpp})\right)\right\}
\]

\[
+ 2\left(M_{xpp} - M^{-1}_{xpx}M^{-1}_xM_{xpp} - 2M_{xy}M^{-1}_{xx}M_{xpp}\right)
\]

\[
+ 2M^T_{xy}M^{-1}_{xx}M_{xpx}M^{-1}_{xpx} + M^T_{xy}M^{-1}_{xx}M_{xpp}M^{-1}_{xx}M_{xy}
\]

\[
- M^T_{xy}M^{-1}_{xx}M_{xpx}M^{-1}_{xpx}M^{-1}_{xx}M_{xy}\right) \]

where \( \text{sum}\{a_{i,j}\} = \sum_i \sum_j a_{i,j}, \quad (a_{i,j}) \circ (a'_{i,j}) = (a_{i,j}a'_{i,j}) \) and

\[
M_{xpy} = \sum_{i=1}^{N} X_iW_iPW_iY_i, \quad M_{xpp} = \sum_{i=1}^{N} X_iW_iPW_iPW_iY_i,
\]

\[
M_{ypyp} = \sum_{i=1}^{N} Y_i^T W_iPW_iPW_iY_i
\]

The details of derivations of the expected and observed information are presented in e-Appendix A of Supporting Information.

Substituting these two variance estimators, we can obtain two adjusted covariance matrix estimators of \( \hat{\boldsymbol{\beta}} \). Based on (1) and (2), the covariance matrix estimators are summarized as

\[
\hat{\Phi}_E = \hat{\Phi} + 2\hat{\Phi}\left[\hat{\sigma}_E\left(M_{xpp} - M_{xpx}M^{-1}_xM_{xpx}\right)\right]\hat{\Phi}
\]

\[
\hat{\Phi}_O = \hat{\Phi} + 2\hat{\Phi}\left[\hat{\sigma}_O\left(M_{xpp} - M_{xpx}M^{-1}_xM_{xpx}\right)\right]\hat{\Phi}
\]
2.4 Inferences of the regression parameters

Kenward and Roger \(^{11}\) also considered approximating the sampling distributions of the regression coefficients of linear mixed models by \(t\)-distribution for constructing significance tests and confidence intervals. We adopt similar approximations using a \(t\)-distribution for the inferences of \(\beta\) of the multivariate meta-analysis model (*)}. For the significance test of \(j\)th element of \(\beta\), \(H_0: \beta_j = 0\) vs. \(H_1: \beta_j \neq 0\), the test statistic is given as

\[
T_{E}^{(j)} = \hat{\beta}_j / \sqrt{\hat{\phi}_{Ejj}} \quad \text{or} \quad T_{O}^{(j)} = \hat{\beta}_j / \sqrt{\hat{\phi}_{Ojj}}
\]

where \(\hat{\phi}_{Ejj}\) and \(\hat{\phi}_{Ojj}\) are \(j\)th diagonal elements of \(\hat{\Phi}_E\) and \(\hat{\Phi}_O\), respectively \((j = 1, 2, \ldots, p)\). Based on the Kenward-Roger-type higher order asymptotic approximation that generalized the Satterthwaite approximation \(^{25}\), the reference distributions of the two test statistics can be approximated by the \(t\)-distributions with approximate degrees of freedom,

\[
m_{E,j} = \frac{2\kappa_{jj}^2}{\omega_{E}\eta_{jj}^2}, m_{O,j} = \frac{2\kappa_{jj}^2}{\omega_{O}\eta_{jj}^2}
\]

where \(\kappa_{jj}\) and \(\eta_{jj}\) are the \(j\)th diagonal elements of \(M_{xx}^{-1}\) and \(M_{xx}^{-1}M_{xx}M_{xx}^{-1}\), respectively \((j = 1, 2, \ldots, p)\). Similarly, the \(100 \times (1 - \alpha)\%\) confidence intervals of \(\beta_j\) are calculated by

\[
\left(\hat{\beta}_j - t_{m_{E,j},1-\alpha/2} \sqrt{\hat{\phi}_{Ejj}}, \hat{\beta}_j + t_{m_{E,j},1-\alpha/2} \sqrt{\hat{\phi}_{Ejj}}\right)
\]

\[
\left(\hat{\beta}_j - t_{m_{O,j},1-\alpha/2} \sqrt{\hat{\phi}_{Ojj}}, \hat{\beta}_j + t_{m_{O,j},1-\alpha/2} \sqrt{\hat{\phi}_{Ojj}}\right)
\]

where \(t_{m,q}\) is the \(q\)th quantile of the \(t\)-distribution with \(m\) degrees of freedom. To distinguish the two proposed methods, we refer to the inference method that uses \(\hat{\Phi}_E\) as KR(E) method and that uses \(\hat{\Phi}_O\) as KR(O) method.
3. Simulations

We performed simulation studies to evaluate the performance of the proposed methods under practical situations of network meta-analyses, especially under moderate numbers of synthesized studies. We compared the two proposed Kenward-Roger–type methods with the current standard inference methods of network meta-analyses: the maximum likelihood (ML) and REML estimation methods. We considered a hexagonal network comparing six treatments (A, B, C, D, E, and F, where A is formally regarded as a reference). The simulation settings were based on the network meta-analysis example of Siontis et al.\textsuperscript{14} (Section 4.1). The number of combined trials (N) was set to 10, 14, 18, 22, 26, or 30 and the standard synthesis analyses of network meta-analysis without covariates were considered, i.e., $x_{ij} = (1)$ and denote $μ_j = x_{ij}β_j$ ($j = 1,2,\ldots,p$). The network plot of the simulated data and the designs of the constitutional studies are summarized in Figure 1.

The outcome data of individual studies were simulated as binomial data, $X_{ij} \sim \text{Binomial} (n_{ij}, p_{ij})$ ($i = 1,\ldots,N; j = 0,\ldots,5$, corresponding to treatments A–F). We adopted odds ratios (ORs) as the outcome measure (i.e., $θ_{ij} = \log \left[ p_{ij}(1 - p_{ij}) / \{p_{i0} \cdot (1 - p_{0j})\} \right]$ ($j = 1,2,\ldots,5$). The response rate of the reference treatment group $p_{i0}$ was randomly simulated from a continuous uniform distribution on $[0.020, 0.436]$. In addition, the log OR, $θ_{ij}$, was generated from a normal distribution $N(μ_j, τ_j^2)$. For the heterogeneity variance–covariance matrix $Σ$, we considered a compound symmetry structure that assumes the equivalent variances $τ^2 = τ_1^2 = \cdots = τ_5^2$ with correlation coefficients ($κ_{ij}s$) of 0.50. The degree of heterogeneity was varied among $τ = 0.20, 0.40, 0.60, 0.80, 1.00, 1.20, \text{ and } 1.40$. In addition, the average treatment effects were set to be $μ_1 = 1.057$, $μ_2 = -0.362$, $μ_3 = -0.508$, $μ_4 = 1.415$, and $μ_5 = -1.317$, consistent with the REML estimate of the dataset of Siontis et al.\textsuperscript{14} (Section 4.1). All of the response-rate parameters $p_{i0}/s$ were
uniquely specified from $\theta_{ij}$ and $p_{00}$. The sample sizes were set to be equal among all the arms of individual trials and were assigned random integer values between 30 and 1126. We calculated trial-specific log OR estimate $Y_i$ and its covariance matrix estimate $S_i$ on the basis of the generated binomial data $X_{ij}$. For each scenario, we replicated the simulations 10,000 times. Among the 10,000 simulations, we calculated the empirical estimates of coverage probabilities and expected widths for the 95% confidence intervals of the grand mean parameters.

The main results of the simulations are presented in Figure 2; the empirical coverage probabilities of OR for E vs. A. The corresponding results for other pairwise ORs are presented in e-Appendix B of the Supporting Information. In general, the ML and REML methods showed undercoverage performance, especially for small $N$. In particular, the REML method, which is one of the current standard inference methods of network meta-analysis, could seriously underestimate the actual statistical errors.

For the proposed Kenward-Roger–type inference methods, the coverage probabilities were approximately the nominal level (95%) irrespective of the degrees of heterogeneity ($\tau$) and the number of trials synthesized ($N$). Under a few scenarios, the coverage probabilities were relatively too large ($>0.95$), especially for small $N$ and small $\tau$ settings; however, even under these scenarios, they performed well compared with the ML and REML methods, which were seriously undercovered under these settings. In addition, even when the number of trials $N$ was too large, the Kenward-Roger–type inference methods provided accurate confidence intervals compared with the REML method and the coverage probabilities were $\sim0.95$ under all the scenarios considered. The two proposed methods were generally comparable. These results possibly involve Monte Carlo errors, but in the 10,000 replications, the Monte Carlo standard error was controlled to less than 0.0022 for estimating a binomial probability of 0.95. The simulation results
clearly demonstrate the effectiveness of the proposed methods.

Additional simulation studies based on the schizophrenia study in Section 4.2 and for continuous outcome data based on a diabetes study of Phung et al. 26 are presented in e-Appendix C of Supporting Information. We could confirm the practical effectiveness of the proposed methods consistently.

4. Applications

4.1 Noninvasive diagnostic modalities for detecting coronary artery disease

Siontis et al. 14 performed network meta-analyses to evaluate the comparative effectiveness of noninvasive diagnostic modalities for detecting coronary artery disease. Six treatments were compared using a dataset of 18 clinical trials ($n = 11,329$) for patients with low-risk acute coronary syndrome: coronary computed tomographic angiography (CCTA), cardiovascular magnetic resonance (CMR), exercise echocardiography (ECG), single-photon-emission computed tomography–myocardial perfusion imaging (SPECT-MPI), standard care, and stress echocardiography. Here, we considered the downstream testing as the outcome and adopted OR as the effect measure. The network diagram is presented in Figure 3(A).

We analyzed this network meta-analysis dataset for illustration of the proposed methods, compared with the current standard inference methods (i.e., ML and REML methods). We adopted the multivariate random-effects model with a compound symmetric within-studies correlation matrix. The reference treatment was “CCTA.” The estimates of $\tau$, the between-studies standard deviation (SD), were 0.89 for ML and 1.10 for REML estimation (log OR scale), and substantial heterogeneity was indicated. The point estimates of the Kenward-Roger–type inference methods were formally set to the REML estimates.
The comparative OR estimates and their 95% confidence intervals are presented in Table 1. Overall, the widths of the confidence intervals of the OR estimates obtained by the Kenward-Roger–type methods were larger than those obtained by the ML and REML methods. In particular, the upper bounds of the confidence intervals obtained by the Kenward-Roger–type methods were obviously higher than those of the confidence intervals obtained by the ML and REML methods. The lower bounds of the confidence intervals were not markedly different; however, those obtained by the Kenward-Roger–type methods were slightly smaller, i.e., the confidence intervals were wider. The two Kenward-Roger–type methods were comparable and provided similar confidence intervals. According to the simulation results in Section 3, the ML and REML methods can underestimate the statistical errors; the proposed methods are expected to correct the undercoverage properties of the conventional methods.

4.2 Pharmacotherapy of schizophrenia

Ades et al. 15 conducted a network meta-analysis of pharmacotherapy for preventing relapses of schizophrenia. Their analysis involved 15 trials (n = 3,533) that compared eight drugs (amisulpride, aripiprazole, haloperidol, olanzapine, paliperidone, risperidone, ziprasidone, and zotepine) and a placebo. The network diagram is presented in Figure 3(B). The following four outcomes were available at the end of follow-up: relapse, discontinuation of treatment due to intolerable side effects and other reasons, not reaching any of the three endpoints, and still in remission. Here, we considered the last outcome and adopted OR as the effect measure.

We analyzed the network meta-analysis dataset using the random-effects multivariate meta-analysis model with the compound symmetry within-studies correlation matrix. The reference treatment was set to “placebo.” The between-studies SD estimates were 0.28
and 0.52 for the ML and REML estimation methods, respectively. The results of the estimations of comparative ORs are presented in Table 2. The confidence interval obtained by ML (ordinary Wald confidence interval) was much narrower than those obtained by the other methods, and several drugs were found to be significantly efficacious compared with placebo, reflecting the small ML estimate of the between-studies SD. In addition, the REML and Kenward-Roger–type methods provided wider confidence intervals and some of the drugs found to be significant by ML were not significantly different compared with placebo. In particular, the Kenward-Roger–type methods provided wider confidence intervals than the REML method and indicated greater uncertainty in the estimation of treatment effects. The two proposed Kenward-Roger–type methods provided comparable results, and similar confidence intervals were constructed. Compared with the simulation results, this case involves a relatively small number of trials ($N = 15$); the conventional inference methods possibly underestimate the statistical errors, and the proposed methods would adjust the possible invalidity.

5. Discussion

The evidence for comparative effectiveness of healthcare technologies from network meta-analysis has been widely used in public health, clinical practices, health technology assessment, and policy making. If misleading evidence has been reported as a result of the adoption of inadequate statistical methods, the influences in various applications might be enormous. The REML method has been the current standard method for the inferences of contrast-based models and has been adopted in numerous statistical packages for network meta-analysis (e.g., network of Stata and netmeta of R); however, our simulation results indicate that the REML method might be invalid under realistic situations in practice. Some effective alternative approaches have been
developed \textsuperscript{8-10}, however, they require enormous computational efforts involving Monte Carlo computations. In the present work, we developed the Kenward-Roger–type inference methods, which have been well investigated and are known to have favorable properties for analyses using mixed-effects models for longitudinal data \textsuperscript{12,13}. The main advantages of the proposed methods are their computational simplicity and the fact that they only require simple matrix computations. They can be implemented by standard statistical software via several matrix computations, and the computations can be rapidly implemented. In addition, the coverage probabilities accurately accorded to the nominal level and performed well compared with the REML method in the simulation studies. The proposed Kenward-Roger–type inference methods would be alternative standard methods for the analyses of network meta-analysis. Note that the inconsistency evaluations in network meta-analysis \textsuperscript{16,22,28} can be implemented within the meta-regression framework, and the proposed methods can also be simply applied to these analyses. Further, similar discussions have already performed for other applications of multivariate meta-analysis, e.g., bivariate meta-analysis for diagnostic test accuracy \textsuperscript{29}.

In addition, our findings suggest that statistical methods in network meta-analysis should be carefully selected in practice. Many discrepant results have been reported among meta-analyses \textsuperscript{30,31}, and meta-analyses have typically tended to provide liberal results in these cases. Various systematic biases such as small study effects might be important sources of these discrepancies; however, we should recognize the risks of providing overconfident and misleading interpretations caused by inaccurate inference methods. In particular, when the number of synthesized trials involving the cases of subgroup analyses is small, improved accurate inference methods are recommended. These analyses are at least worth conducting as alternatives to the ordinary REML method.
to check how the confidence limits might be altered. The overall conclusion might be changed by adopting these improved methods.

Comparing the proposed two methods, the standard error estimator of KR(O) was relatively unstable. Thus, in general practice, KR(E) would be recommended. Also, another relevant issue caused by the small-sample problem in meta-analysis is the invalidity of prediction intervals. Several studies have discussed the undercoverage properties of ordinary prediction intervals for conventional pairwise meta-analyses. In particular, Partlett and Riley discussed applying Kenward-Roger–type methods to the prediction intervals of pairwise meta-analysis and reported that the coverage performance was improved. Although the uncertainty caused by estimation of $\tau^2$ should be adequately addressed, the same problem might occur for network meta-analysis. The proposed methods in the present work might be useful to construct improved prediction intervals. These issues are relevant for future studies. In addition, our results could be similarly applied to other meta-analysis methods such as meta-analysis for diagnostic accuracy studies. The development of concrete computational methods and evaluations of their practical effectiveness are also topics for future studies. Accurate inference methods would certainly improve the statistical evaluations of these systematic reviews.

**Acknowledgements**

This study was supported by Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (grant numbers: JP21K03105, JP22H03554, JP22K19688).
References


Table 1. Results of the network meta-analysis for noninvasive diagnostic modalities for detecting coronary artery disease †,*.

<table>
<thead>
<tr>
<th></th>
<th>ML</th>
<th>REML</th>
<th>KR(E)</th>
<th>KR(O)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds ratios and 95% confidence intervals (vs. CCTA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMR</td>
<td>2.76 (0.40, 19.30)</td>
<td>2.88 (0.32, 25.94)</td>
<td>2.88 (0.27, 30.21)</td>
<td>2.88 (0.28, 29.90)</td>
</tr>
<tr>
<td>Exercise ECG</td>
<td>0.70 (0.19, 2.63)</td>
<td>0.70 (0.14, 3.44)</td>
<td>0.70 (0.11, 4.37)</td>
<td>0.70 (0.11, 4.30)</td>
</tr>
<tr>
<td>SPECT-MPI</td>
<td>0.60 (0.16, 2.22)</td>
<td>0.60 (0.12, 2.98)</td>
<td>0.60 (0.09, 3.95)</td>
<td>0.60 (0.09, 3.87)</td>
</tr>
<tr>
<td>Standard care</td>
<td>4.17 (2.04, 8.50)</td>
<td>4.12 (1.74, 9.74)</td>
<td>4.12 (1.53, 11.10)</td>
<td>4.12 (1.54, 11.00)</td>
</tr>
<tr>
<td>Stress echo</td>
<td>0.27 (0.06, 1.23)</td>
<td>0.27 (0.04, 1.68)</td>
<td>0.27 (0.03, 2.19)</td>
<td>0.27 (0.03, 2.15)</td>
</tr>
</tbody>
</table>

† ML: maximum likelihood estimate (Wald C.I.); REML: restricted maximum likelihood estimate (REML-based Wald-type C.I.); KR: Kenward-Roger–type inference methods.
Table 2. Results of the network meta-analysis for pharmacotherapy of schizophrenia †.

<table>
<thead>
<tr>
<th></th>
<th>ML</th>
<th>REML</th>
<th>KR(E)</th>
<th>KR(O)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds ratios and 95% confidence intervals (vs. placebo)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amisulpride</td>
<td>3.38 (1.62, 7.07)</td>
<td>3.36 (1.22, 9.31)</td>
<td>3.36 (0.96, 11.73)</td>
<td>3.36 (1.02, 11.13)</td>
</tr>
<tr>
<td>Aripiprazole</td>
<td>2.07 (1.01, 4.23)</td>
<td>2.07 (0.67, 6.38)</td>
<td>2.07 (0.47, 9.07)</td>
<td>2.07 (0.47, 9.01)</td>
</tr>
<tr>
<td>Haloperidol</td>
<td>2.65 (1.38, 5.10)</td>
<td>2.36 (0.94, 5.95)</td>
<td>2.36 (0.74, 7.52)</td>
<td>2.36 (0.79, 7.12)</td>
</tr>
<tr>
<td>Olanzapine</td>
<td>4.91 (2.83, 8.52)</td>
<td>4.52 (2.15, 9.50)</td>
<td>4.52 (1.85, 11.05)</td>
<td>4.52 (1.91, 10.69)</td>
</tr>
<tr>
<td>Paliperidone</td>
<td>2.08 (0.95, 4.51)</td>
<td>2.08 (0.65, 6.67)</td>
<td>2.08 (0.45, 9.60)</td>
<td>2.08 (0.48, 9.02)</td>
</tr>
<tr>
<td>Risperidone</td>
<td>5.46 (2.59, 11.52)</td>
<td>5.05 (1.74, 14.64)</td>
<td>5.05 (1.31, 19.40)</td>
<td>5.05 (1.40, 18.18)</td>
</tr>
<tr>
<td>Ziprasidone</td>
<td>5.03 (2.41, 10.49)</td>
<td>4.90 (1.81, 13.26)</td>
<td>4.90 (1.48, 16.26)</td>
<td>4.90 (1.55, 15.53)</td>
</tr>
<tr>
<td>Zotepine</td>
<td>2.66 (0.94, 7.52)</td>
<td>2.66 (0.68, 10.33)</td>
<td>2.66 (0.55, 12.90)</td>
<td>2.66 (0.57, 12.28)</td>
</tr>
</tbody>
</table>

† ML: maximum likelihood estimate (Wald C.I.); REML: restricted maximum likelihood estimate (REML-based Wald-type C.I.); KR: Kenward-Roger–type inference methods.
Figure 1. Settings of the simulation studies: the network plot and details of the 42 scenarios.
Figure 2. Simulation results. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
Figure 3. Network plots for the network meta-analysis of (A) noninvasive diagnostic modalities for detecting coronary artery disease and (B) pharmacotherapy of schizophrenia.
SUPPORTING INFORMATION for
Kenward-Roger–type corrections for inference methods of network meta-analysis and meta-regression

Hisashi Noma1,†, Yasuyuki Hamura2,†, Masahiko Gosho3 and Toshi A. Furukawa4

1 Department of Data Science, The Institute of Statistical Mathematics, Tokyo, Japan
2 Graduate School of Economics, The University of Tokyo, Tokyo, Japan
3 Department of Biostatistics, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan
4 Departments of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan
† The two authors equally contributed to this work.

e-Appendix A: Derivations of the expected and observed information

The network meta-regression model (*) is expressed as

\[ Y_i \sim N_p \left( X_i^T \beta, S_i + \tau^2 P \right) \]

\((i = 1, 2, ..., N)\). For detailed mathematical evaluations, we use several different notations here:

\[ Y = (Y_i)_{i=1}^N, \quad X = (X_i^T)_{i=1}^N, \quad \Omega = \text{diag}(\tau^2 P + S_i), \quad \sigma = \tau^2 \]

Then, we can express

\[ Y \sim N_{N_p}(X\beta, \Omega) \]

and

\[ \Phi = (X^T \Omega^{-1} X)^{-1} \]

Next, we denote the dimension of \( \beta \) as \( q \) and we consider \( K \) as a \((Np - q) \times Np\) full rank matrix that satisfies \( KX = 0 \). Denoting \( M = (K^T X) \), the restricted likelihood function is expressed as

\[
L_{RL}(KY) = (2\pi)^{-(N-p)/2} \left( \frac{|M^T M|}{|X^T X|^2} \right)^{-1/2} |\Omega|^{-1/2} |X^T \Omega^{-1} X|^{-1/2} \\
\times \exp \left[ -\frac{1}{2} y^T \{ \Omega^{-1} - \Omega^{-1} X (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} \} y \right]
\]
\[
= (2\pi)^{-(N-p)/2} \left( \frac{|M^T M|}{|X^T X|^2} \right)^{-1/2} |\Omega|^{-1/2} |X^T \Omega^{-1} X|^{-1/2} \exp \left\{ -\frac{1}{2} y^T K^T (K \Omega K^T)^{-1} K y \right\}
\]

The expected information and observed information are then expressed as

\[
I_E = \frac{1}{2} \left[ \text{tr} \left( \frac{\partial (\Omega^{-1})}{\partial \sigma} \Omega \frac{\partial (\Omega^{-1})}{\partial \sigma} \Omega \right) - \text{tr} (2\Phi Q_{11} - \Phi P_1 \Phi P_1) \right],
\]

\[
I_0 = -\frac{\partial^2}{\partial \sigma \partial \sigma} \log(L_{RL}(KY)),
\]

where

\[
P_1 = X^T \frac{\partial (\Omega^{-1})}{\partial \sigma} X, \quad Q_{11} = X^T \frac{\partial (\Omega^{-1})}{\partial \sigma} \Omega \frac{\partial (\Omega^{-1})}{\partial \sigma} X
\]

The components of the expected information and observed information are calculated through the following matrix calculations:

\[
\Omega = \text{diag}(\sigma P + S_i),
\]

\[
\frac{\partial \Omega}{\partial \sigma} = \text{diag} P,
\]

\[
\Omega^{-1} = \text{diag} \left( \frac{1}{\sigma P + S_i} \right),
\]

\[
\frac{\partial (\Omega^{-1})}{\partial \sigma} = -\text{diag} \left( \left( \frac{1}{\sigma P + S_i} \right)^2 \frac{\partial (\sigma P + S_i)^{-1}}{\partial \sigma} \left( \sigma P + S_i \right)^{-1} \right),
\]

\[
\log|\Omega| = \sum_{i=1}^{N} \log |\sigma P + S_i|,
\]

\[
\frac{\partial}{\partial \sigma} \log|\Omega| = \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \sigma P + S_i \right)^{-1} \circ P
\]

\[
\frac{\partial^2}{\partial \sigma \partial \sigma} \log|\Omega| = -\sum_{i=1}^{N} \sum_{j=1}^{N} \left( \sigma P + S_i \right)^{-1} \circ P \left( \sigma P + S_i \right)^{-1} \circ P,
\]

\[
\log |X^T \Omega^{-1} X| = \log \left| \sum_{i=1}^{N} X_i (\sigma P + S_i)^{-1} X_i^T \right|
\]
\[
\frac{\partial}{\partial \sigma} \log |X' \Omega^{-1} X| = -\sum \left[ \left( \sum_{i=1}^{N} X_i (\sigma P + S_i)^{-1} X_i^T \right)^{-1} \right]
\]

\[
\times \sum_{i=1}^{N} X_i (\sigma P + S_i)^{-1} P (\sigma P + S_i)^{-1} X_i^T
\]

\[
\frac{\partial^2}{\partial \sigma \partial \sigma} \log |X' \Omega^{-1} X| = -\sum \left[ \left( \sum_{i=1}^{N} X_i (\sigma P + S_i)^{-1} X_i^T \right)^{-1} \right]
\]

\[
\times \sum_{i=1}^{N} X_i (\sigma P + S_i)^{-1} P (\sigma P + S_i)^{-1} X_i^T
\]

\[
\times \sum_{i=1}^{N} \left[\left( \sum_{i=1}^{N} X_i (\sigma P + S_i)^{-1} P (\sigma P + S_i)^{-1} X_i^T \right)^{-1} \right]
\]

\[
\times \sum_{i=1}^{N} \left[\sum_{i=1}^{N} X_i (\sigma P + S_i)^{-1} P (\sigma P + S_i)^{-1} X_i^T \right]
\]

\[
\frac{\partial^2}{\partial \sigma \partial \sigma} y' K' (K \Omega K')^{-1} Ky = \frac{\partial}{\partial \sigma} \left\{ -y' K' (K \Omega K')^{-1} K \left( \text{diagP} \right) K' (K \Omega K')^{-1} Ky \right\}
\]

\[
= 2y' K' (K \Omega K')^{-1} K \left( \text{diagP} \right) K' (K \Omega K')^{-1} K \left( \text{diagP} \right) K' (K \Omega K')^{-1} Ky
\]

\[
= 2y' \left( \Omega^{-1} - \Omega^{-1} X (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} \right) \left( \text{diagP} \right) \left\{ \Omega^{-1} \right\}
\]

\[
- \Omega^{-1} X (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} \right) \left( \text{diagP} \right) \left\{ \Omega^{-1} \right\}
\]

\[
- \Omega^{-1} X (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} \right) y
\]
Then, we can obtain

\[ \Phi = M_{xx}^{-1}, \]
\[ P_1 = -M_{xpx}, \]
\[ Q_{11} = M_{xppx}, \]

and the expected information is expressed as

\[ I_E = \frac{1}{2} \sum_{i=1}^{N} \text{tr}(W_i P W_i P) - \text{tr} \left( 2M_{xx}^{-1} M_{xpx} - M_{xx}^{-1} M_{xpx} M_{xx}^{-1} M_{xpx} \right) \]

Also, the observed information is expressed as

\[
I_O = \frac{1}{2} \frac{\partial^2 \log|\Omega|}{\partial \sigma \partial \sigma} + \frac{1}{2} \frac{\partial^2 \log|X^T \Omega^{-1} X|}{\partial \sigma \partial \sigma} + \frac{1}{2} \frac{\partial^2 \mathbf{y}^T K^T (K\Omega K^T)^{-1} K \mathbf{y}}{\partial \sigma \partial \sigma}
\]

\[
= \frac{1}{2} \left( - \sum_{i=1}^{N} \sum \{(W_i P W_i) \circ P \}
+ \sum \left\{ \left( -M_{xx}^{-1} M_{xpx} M_{xx}^{-1} \right) \circ M_{xpx} \right\} + \left\{ M_{xx}^{-1} \circ \left( 2M_{xpx} \right) \right\}
+ 2 \left( M_{ypp} - M_{xpy}^T M_{xx}^{-1} M_{xpy} - 2M_{xy}^T M_{xx}^{-1} M_{xpy} \right)
+ 2M_{xy}^T M_{xx}^{-1} M_{xpx} M_{xx}^{-1} M_{xpy} + M_{xy}^T M_{xx}^{-1} M_{xppx} M_{xx}^{-1} M_{xy}
- M_{xy}^T M_{xx}^{-1} M_{xpx} M_{xx}^{-1} M_{xpx} M_{xx}^{-1} M_{xy} \right) \right)
\]
e-Appendix B: All simulation results of Section 3

We present the remaining simulation results discussed in Section 3. The empirical coverage rates of OR for B vs. A, C vs. A, D vs. A, and F vs. A are presented in e-Figures 1-4. The overall results are similar to those discussed in Section 3. The coverage probabilities of the ML and REML methods were seriously lower than the nominal level (0.95). In addition, the Kenward-Roger–type inference methods clearly showed favorable coverage performances in general.
e-Figure 1. Simulation results for the comparison B vs. A. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
e-Figure 2. Simulation results for the comparison C vs. A. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
**e-Figure 3.** Simulation results for the comparison D vs. A. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
CONFIDENCE INTERVALS FOR $\text{OR}_F$ vs. $A$

**e-Figure 4.** Simulation results for the comparison $F$ vs. $A$. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
e-Appendix C: Additional simulations

C.1 Simulations based on the schizophrenia study

We implemented simulation studies based on another real dataset, the schizophrenia study presented in Section 4.2, to evaluate the practical effectiveness of the proposed methods under different settings. We considered a nonagonal network comparing nine treatments presented in e-Figure 5 (A is formally regarded as a reference). The outcome data of individual studies were simulated as binomial data, \( X_{ij} \sim \text{Binomial}(n_{ij}, p_{ij}) \) \((i = 1, \ldots, N; j = 0, 1, \ldots, 8, \text{corresponding to treatments A–I})\). We adopted ORs as the outcome measure (i.e., \( \theta_{ij} = \log \left[ \frac{p_{ij}(1 - p_{0})}{p_{0}(1 - p_{ij})} \right] \)). The response rate of the reference treatment group \( p_{0} \) was randomly simulated from a continuous uniform distribution on [0.140, 0.461]. In addition, the log OR, \( \theta_{ij} \), was generated from a normal distribution \( N(\mu_j, \tau_j^2) \). For the heterogeneity variance–covariance matrix \( \Sigma \), we considered a compound symmetry structure that assumes the equivalent variances \( \tau^2 = \tau_1^2 = \cdots = \tau_8^2 \). The degree of heterogeneity was varied among \( \tau \in \{0.20, 0.40, 0.60, \ldots, 1.40\} \). The number of combined trials \((N)\) was set to 10, 14, 18, 22, 26, or 30. In addition, the average treatment effects were set to be \( \mu_1 = 1.213, \mu_2 = 0.726, \mu_3 = 0.861, \mu_4 = 1.508, \mu_5 = 0.730, \mu_6 = 1.619, \mu_7 = 1.590, \text{and} \mu_8 = 0.977 \). All of the response-rate parameters \( p_{ij} \)s were uniquely specified from \( \theta_{ij} \) and \( p_{0} \). The sample sizes were set to be equal among all the arms of individual trials and were assigned random integer values between 13 and 627.

The same inference methods as Section 3 were considered. For each scenario, we replicated the simulations 10,000 times.

The results of the simulations are presented in e-Figures 6-13; the empirical coverage probabilities for B vs. A, C vs. A, \ldots, I vs. A are presented. In general, similar results were obtained as the simulation studies in Section 3. The ML and REML methods showed undercoverage performance, especially for small \( N \). Besides, the proposed Kenward-Roger–type inference methods, the coverage probabilities were approximately the nominal level (95%) irrespective of the degrees of heterogeneity \((\tau)\) and the number of
trials synthesized ($N$). Under a few scenarios, the coverage probabilities were relatively too large (> 0.95), especially for small $N$ and small $\tau$ settings; however, even under these scenarios, they performed well compared with the ML and REML methods, which were seriously undercoveraged under these settings. The simulation results also clearly demonstrate the effectiveness of the proposed methods.
**e-Figure 5.** Settings of the simulation studies: the network plot and details of the 42 scenarios.
e-Figure 6. Simulation results for the comparison B vs. A. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
e-Figure 7. Simulation results for the comparison C vs. A. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
e-Figure 8. Simulation results for the comparison D vs. A. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
Confidence intervals for OR$_E$ vs. A.

e-Figure 9. Simulation results for the comparison E vs. A. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
e-Figure 10. Simulation results for the comparison $F$ vs. $A$. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
e-Figure 11. Simulation results for the comparison G vs. A. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
**e-Figure 12.** Simulation results for the comparison H vs. A. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
e-Figure 13. Simulation results for the comparison I vs. A. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
C.2 Simulations for continuous outcome data

We also performed additional simulation studies that considered continuous outcome data based on a network meta-analysis for type-2 diabetes of Phung et al. (2010); the original data is involved network graph package of Stata (Chaimani and Salanti, 2015). We aimed to investigate the operating characteristics of the proposed methods for the continuous outcome settings. We considered a heptagonal network comparing seven treatments presented in e-Figure 14 (A is formally regarded as a reference). We adopted mean difference (MD) as the outcome measure. The arm-based outcome data of individual studies were simulated as Gaussian data, \( X_{ij} \sim N(\theta_{ij}, s_i^2) \) \((i = 1, \ldots, N; j = 0, 1, \ldots, 6, \text{ corresponding to treatments A–G})\); the mean difference estimate comparing treatment \( j \) and \( k \) is calculated as \( X_{ij} - X_{ik} \). The mean of the reference treatment group \( \theta_{00} \) was randomly simulated from a continuous uniform distribution on \([-0.33, 0.22]\). In addition, the MD parameter, \( \theta_{ij} \), was generated from a normal distribution \( N(\mu_j, \tau_j^2) \). For the heterogeneity variance–covariance matrix \( \Sigma \), we considered a compound symmetry structure that assumes the equivalent variances \( \tau^2 = \tau_1^2 = \cdots = \tau_6^2 \). The degree of heterogeneity was varied among \( \tau = 0.20, 0.40, 0.60, \ldots, 1.40 \). The number of combined trials \( N \) was set to 10, 14, 18, 22, 26, or 30. In addition, the average treatment effects were set to be \( \mu_1 = -1.002, \mu_2 = -0.852, \mu_3 = -0.888, \mu_4 = -0.986, \mu_5 = -0.637 \) and \( \mu_6 = -1.080 \). Also, study-specific standard deviation \( \sigma_i \) was generated from a continuous uniform distribution on \([0.93, 1.56]\); \( s_i^2 \) was calculated as the variance of \( X_{ij} (= \sigma_i^2/n_{ij}) \). The sample sizes were set to be equal among all the arms of individual trials and were assigned random integer values between 27 and 453. Under these settings, the same inference methods as Section 3 were considered. For each scenario, we replicated the simulations 10,000 times.

The results of the simulations are presented in e-Figures 15-20; the empirical coverage probabilities for B vs. A, C vs. A,…, G vs. A are presented. Overall, similar results were obtained as the simulation studies for the binomial outcome. The ML and REML methods
showed undercoverage performance generally. Besides, for the proposed methods, the coverage probabilities were approximately the nominal level (95%) irrespective of the degrees of heterogeneity ($\tau$) and the number of trials synthesized ($N$). For these simulations, the conservative properties under small $N$ and small $\tau$ settings were not observed. The accuracies of the proposed inference methods were consistently shown.
e-Figure 14. Settings of the simulation studies: the network plot and details of the 42 scenarios.
e-Figure 15. Simulation results for the comparison B vs. A. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
e-Figure 16. Simulation results for the comparison C vs. A. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
e-Figure 17. Simulation results for the comparison D vs. A. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
**e-Figure 18.** Simulation results for the comparison $E$ vs. $A$. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
e-Figure 19. Simulation results for the comparison $F$ vs. $A$. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
**e-Figure 20.** Simulation results for the comparison $G$ vs. $A$. The coverage probabilities of 95% confidence intervals provided by the four methods: ML, REML, KR(E), and KR(O).
References
