NMA: Network meta-analysis based on multivariate meta-analysis and meta-regression models in R

Hisashi Noma, PhD*
Department of Interdisciplinary Statistical Mathematics, The Institute of Statistical Mathematics, Tokyo, Japan
The Graduate Institute for Advanced Studies, The Graduate University for Advanced Studies (SOKENDAI), Tokyo, Japan
ORCID: http://orcid.org/0000-0002-2520-9949

Kazushi Maruo, PhD
Department of Biostatistics, Institute of Medicine, University of Tsukuba, Tsukuba, Japan

Shiro Tanaka, PhD
Department of Clinical Biostatistics, Kyoto University Graduate School of Medicine, Kyoto, Japan

Toshi A. Furukawa, MD, PhD
Kyoto University Office of Institutional Advancement and Communications, Kyoto, Japan

*Corresponding author: Hisashi Noma
Department of Interdisciplinary Statistical Mathematics
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

Abstract

Network meta-analysis has become an established methodology within systematic reviews for comparing the effectiveness of multiple treatments, and it has been now a standard approach in comparative effectiveness research. However, the underlying statistical methods are often highly technical for non-statisticians in practice, and no freely available software package has been developed that can handle a general framework based on the multivariate meta-analysis and meta-regression models. To address these issues, we developed `NMA`, a comprehensive and user-friendly R package that covers extensive analysis and graphical tools of network meta-analysis with simple commands. The `NMA` package provides generic functional tools for evidence synthesis based on the multivariate meta-analysis models, network meta-regression, assessment of heterogeneity and inconsistency, comparative effectiveness analyses, and a range of graphical tools. In addition, `NMA` includes data-handling functions that facilitate the integration of both arm-level data and summary statistics easily. In this article, we provide a gentle introduction to the `NMA` package and illustrate its application through a case study of a network meta-analysis of antihypertensive drugs.

Key words: network meta-analysis; comparative effectiveness research; computational package; frequentist tool; multivariate meta-regression.

Highlights

What is already known?

- Several freely computational packages are available for network meta-analysis, but no general frequentist tool based on the multivariate meta-analysis and meta-regression models, introduced by White et al. [1], has been developed.

What is new?

- We developed `NMA`, a comprehensive R package for network meta-analysis based on multivariate meta-analysis and meta-regression models with frequentist approach.
- The `NMA` package provides a broad range of functions for evidence synthesis, heterogeneity and inconsistency assessment, comparative effectiveness analysis, and graphical visualization.
- Key analytical tools—such as Higgins' global inconsistency test [2], network meta-regression, and advanced inferential and prediction methods to address invalidity issues of the ordinary approaches [3,4] —are fully implemented.
- Generic data-handling tools can now integrate arm-level data with summary statistics. This provides greater flexibility in network meta-analysis, making it especially useful for studies of survival outcomes.

Potential impact for readers

- The package facilitates the practical use of network meta-analysis for a broad range of researchers, including non-statisticians, thereby enhancing the accessibility of systematic reviews on important clinical and public health questions.
- By comprehensively covering standard analyses and graphical tools, the `NMA` package is also valuable for educational purposes, serving as a practical resource for students, researchers, and clinicians to learn the research methods through real-world case studies.

1. Introduction

In almost all disease areas today, multiple treatment options exist, and the question of which treatment constitutes the best option for each patient has become a central issue [5]. At the same time, with the profound impact of population aging in developed countries, this question has also become increasingly important from the perspectives of health economics and public health policy [6,7]. Conventional pairwise meta-analysis, which has been developed since the 1980s, has substantial limitations for addressing this purpose, and network meta-analysis has become the new standard method for comparative effectiveness research [8,9]. Naci and O'Connor [10] have suggested that network meta-analysis can also be applied at the stage of regulatory approval of new drugs, in order to compare their efficacy and safety with those of existing therapies. The necessity of such approaches is expected to increase substantially in the future.

Several excellent free computational tools for network meta-analysis have already been developed in R (R Foundation for Statistical Computing, Vienna, Austria). For Bayesian approaches, which have been widely adopted since their early development, `gemtc` [11] and `pcnetmeta` [12] packages are available. For frequentist approaches, the contrast-based approach is comprehensively supported by the `netmeta` package [13]. While each of these packages implements sophisticated functions reflecting the latest methodological developments, the available methods are restricted by the underlying statistical models. In particular, the global inconsistency test proposed by Higgins et al. [2] is based on the multivariate meta-regression framework introduced by White et al. [1]. In Stata, the `network` package [14] provides a general computational tool for this framework, but no such tool has been available in R. Moreover, generic implementations of network meta-regression are still lacking in existing R packages. In addition, none of the current network meta-analysis software accommodates methods to overcome fundamental invalidity problems in the inferences of treatment effects, where confidence intervals obtained by standard methods can severely underestimate statistical error [3,15,16].

To address these gaps, we have developed the R package `NMA` as an extensive computational tool for practice of network meta-analysis, designed so that all functions can be easily used by non-statisticians. In addition to the functions of Stata's `network` package, the `NMA` package implements various methods based on multivariate meta-regression models, such as Jackson's random inconsistency model [17,18] and the multivariate I^2 statistic [19]. Furthermore, it incorporates new inference and prediction methods for treatment effects based on higher-order asymptotic theory. [3,4] A wide range of graphical tools have also been implemented, including those for assessing transitivity [20]. The package provides enhanced data-handling functionality, with tools that allow integration of both arm-level data and summary-level data. This provides greater flexibility in network meta-analysis, making it especially useful for studies of survival outcomes.

In this paper, we provide an overview of the functions implemented in the NMA package and illustrate its use with example programs and case analyses. In Section 2, we briefly describe the example data used in this article. Section 3 presents a summary of the functions implemented in the NMA package, and Sections 4 through 8 demonstrate various applications of the package using the network meta-analysis of antihypertensive drugs reported by Sciarretta et al. [21] as a case study with binary outcome. Additional analytical tools that cannot be fully described in the main text are detailed in the Supplementary Materials.

Due to space limitations, the R output has been selectively presented. Complete R programs reproducing all case analyses are available on the authors' GitHub page (https://github.com/nomahi/NMA). In addition, further examples with continuous and survival outcomes are given in e-Appendices A and B, respectively.

2. Illustrative example: Network meta-analysis of antihypertensive drugs

In this article, we present a case analysis of NMA using the network meta-analysis of antihypertensive drugs reported by Sciarretta et al. [21] as an example of a binary outcome. Sciarretta et al. [21] conducted a network meta-analysis comparing seven classes of antihypertensive drugs—α-blockers (AB), angiotensin-converting enzyme inhibitors (ACEIs), angiotensin II receptor blockers (ARBs), β-blockers (BBs), calcium channel blockers (CCBs), conventional treatment (CT), and diuretics (DDs)—with placebo. The outcome of interest was the occurrence of heart failure, treated as a binary variable. The dataset included 26 randomized controlled trials conducted between 1997 and 2009, comprising a total of 223,313 participants.

Within the NMA package, this is provided as a dataset heatfailure and contains arm-level data, including sample sizes (n) and event counts (d) for each group. In addition, arm-level covariate data are available, including mean baseline systolic and diastolic blood pressure (SBP, DBP), as well as year of publication (pubyear).

```
> install.packages("NMA")              # Install from CRAN
> library("NMA")
> data(heartfailure)
> head(heartfailure, 6)
  study      trial     trt    n  d pubyear   SBP  DBP
1     1   Syst-Eur     CCB 2398 37    1997 173.8 85.5
2     1   Syst-Eur Placebo 2297 49    1997 173.9 85.5
3     2 Syst-China     CCB 1253  4    1998 170.5 86.0
4     2 Syst-China Placebo 1141  8    1998 170.5 86.0
5     3      UKPDS     ACE  400 12    1998 159.0 94.0
```

```
6     3      UKPDS      BB  358  9    1998 159.0 93.0
```

3. General overview of the NMA package

3.1 Functionalities and features

The R package `NMA` was developed as a collection of generic computational tools for network meta-analysis. It is based on the multivariate meta-regression framework proposed by White et al. [1] and is designed to apply effective statistical methods available within this model. Table 1 provides an overview of the functions implemented in the package.

The package incorporates numerous features that have not been available in existing R packages, including network meta-regression, Higgins's global inconsistency test [2], Jackson's random inconsistency model [17], inference and prediction methods based on higher-order asymptotics developed by Noma et al. [3,4], and tools for assessing transitivity. In addition, the package offers enhanced data-handling functionality, implementing editing tools that allow the seamless integration of arm-based data (e.g., means, standard deviations) with trials reporting only summary statistics (e.g., mean differences).

To ensure accessibility for non-statisticians, the package has been designed with a user-friendly interface. Once an analysis dataset is generated using the `setup` function, all analytical tools can be applied with simple commands using the generated object. In the following sections, we provide an overview of the purpose of each function, how to use them, and how to interpret their results.

3.2 Supported data types and setup

The `NMA` package accommodates three types of outcomes—continuous, binary, and survival—as summarized in Table 2. For each outcome type, the package enables estimation using several effect measures, including the mean difference, standardized mean difference, risk difference, risk ratio, odds ratio, hazard ratio, and survival probability difference.

All analyses require that data be prepared in *arm-level* format. These arm-level datasets are edited by the `setup` function, which edits and creates contrast-based statistics suitable for the analyses using multivariate meta-analysis models. As in conventional meta-analysis, however, mixed data of arm-level and summary-level statistics are often created in practice [22]. The `NMA` package therefore provides dedicated functions for handling such cases. For example, in continuous outcomes, some trials may report mean, standard deviation, and sample size for each arm, whereas others provide only summary statistics such as the mean difference with a standard error (which can be reconstructed from confidence intervals or *p*-values). In the latter case, the function `trans.armdata` can be used to back-calculate (pseudo) arm-level data. Note that the reconstructed values

will not necessarily reproduce the original arm-level data exactly. However, in practice, the `setup` function transforms all the (pseudo) arm-level data into contrast-based summary statistics for analysis, and the resulting summary measures are designed to coincide closely with the originally reported summary statistics. Thus, the arm-level data generated by `trans.armdata` should be viewed as "working" pseudo-data, serving as a bridge to construct contrast-based summaries.

For survival outcomes, hazard ratios are frequently reported [23]. These summary data can also be passed to `trans.armdata`. Under the proportional hazards assumption, the function applies the complementary log–log transformation to invert the hazard ratio formula [24] [25], thereby generating pseudo-data in the form of reconstructed event counts. Also, arm-specific survival probability data can sometimes be obtained from a Kaplan–Meier plot [23]. These survival probabilities can be converted into (pseudo) event counts and sample sizes using the `trans.armdataP` function. These pseudo-data can then be combined with trials reporting only event counts and sample sizes, producing a dichotomized arm-level dataset across all studies. Passing this dataset to the `setup` function yields contrast-based summary statistics expressed as hazard ratios, which reproduce the original estimates.

Once arm-level data from different-type data are merged into a single dataset, the `setup` function can produce contrast-based summary statistics. All subsequent analyses can then be performed using the resulting object with simple commands. If arm-level covariates are available—for example, summary baseline characteristics that can be used in meta-regression or for evaluating transitivity—these variables can be added directly to the dataset in advance. Such data editing may conveniently be carried out in spreadsheet software such as Microsoft Excel (Microsoft, Redmond, WA.).

As an illustrative example for the binary outcome case, the required variables include study identifier (`study`), treatment (`trt`), number of events (`d`), sample size (`n`), and covariates (`z`; optional, multiple allowed). Users must also specify the effect measure of interest (risk difference, risk ratio, or odds ratio) and designate a reference treatment (`ref`).

```
> hf2 <- setup(study=study,trt=trt,d=d,n=n,z=c(SBP,DBP,pubyear),
  measure="OR",ref="Placebo",data=heartfailure)
```

By using the output object generated by the `setup` function as an argument, all subsequent analyses can be executed with simple commands.

4. Summarizing basic information on the network of evidence

4.1 Network diagram

In network meta-analysis, network diagrams [26] are presented in most studies to visualize both the amount and structure of the evidence being synthesized. The `NMA` package facilitates this process through the `netplot` function, which enables users to generate such diagrams in a straightforward manner. The function can be executed using the syntax illustrated below, and its specifications can be flexibly modified according to user preferences.

```
> netplot(hf2)
```

(The figure created is shown in Figure 1.)

4.2 Pairwise meta-analysis for all treatment pairs with direct comparisons

As part of the basic information, the results of pairwise meta-analyses for all direct comparisons in the network can be summarized using the `pairwise` function. This output includes the results of meta-analyses based on the DerSimonian–Laird-type random-effects model [27], along with estimates of τ^2, I^2, and H^2 statistics, and the Egger test [28]. By default, the Hartung-Knapp-Sidik-Jonkman (HKSJ) method [15,29,30] is applied; however, when only a single trial is available for a given direct comparison, the standard restricted maximum likelihood (REML) method is used instead.

```
> pairwise(hf2)
$coding
  code treatment
1    1   Placebo
2    2        AB
3    3       ACE
4    4       ARB
5    5        BB
6    6       CCB
7    7        CT
8    8        DD

$`Summary effect measures`
        N  estimate      95%CL      95%CU      P-value
1 vs. 3 1 0.7589762 0.66310730  0.8687054 6.256620e-05
1 vs. 4 2 0.8455378 0.05065789 14.1129905 5.873462e-01
1 vs. 6 3 0.6749003 0.45918502  0.9919541 4.811086e-02
1 vs. 8 1 0.3746546 0.22813398  0.6152790 1.049431e-04
```

```
2 vs. 8 1 0.4940656 0.43255714  0.5643204 2.635675e-25
3 vs. 4 1 1.0521863 0.92870506  1.1920856 4.244710e-01
3 vs. 5 1 0.8338109 0.34714117  2.0027604 6.843627e-01
3 vs. 6 3 1.1898514 1.06490416  1.3294588 2.130285e-02
3 vs. 7 1 0.8783785 0.62964150  1.2253779 4.452120e-01
3 vs. 8 1 1.1357796 0.81826423  1.5765022 4.466252e-01
4 vs. 5 1 1.0582300 0.84487651  1.3254609 6.222499e-01
4 vs. 6 1 1.1454889 0.98920803  1.3264598 6.952727e-02
4 vs. 7 2 1.5065824 0.10166720 22.3256921 3.041121e-01
5 vs. 6 1 0.8386887 0.66530182  1.0572625 1.365580e-01
6 vs. 7 3 0.8624067 0.66631548  1.1162060 1.322648e-01
6 vs. 8 5 0.7112912 0.25578025  1.9780072 4.074070e-01

$`Heterogeneity measures`
        N       tau^2        I^2      H^2
1 vs. 3 1 0.000000000 0.00000000 1.000000
1 vs. 4 2 0.070811602 0.78793578 4.715553
1 vs. 6 3 0.008587957 0.07155368 1.077068
1 vs. 8 1 0.000000000 0.00000000 1.000000
2 vs. 8 1 0.000000000 0.00000000 1.000000
3 vs. 4 1 0.000000000 0.00000000 1.000000
3 vs. 5 1 0.000000000 0.00000000 1.000000
3 vs. 6 3 0.000289373 0.01800004 1.018330
3 vs. 7 1 0.000000000 0.00000000 1.000000
3 vs. 8 1 0.000000000 0.00000000 1.000000
4 vs. 5 1 0.000000000 0.00000000 1.000000
4 vs. 6 1 0.000000000 0.00000000 1.000000
4 vs. 7 2 0.036694439 0.34928395 1.536769
5 vs. 6 1 0.000000000 0.00000000 1.000000
6 vs. 7 3 0.002581315 0.11647850 1.131834
6 vs. 8 5 0.679782905 0.86271763 7.284256
```

5. Network meta-analysis using multivariate meta-analysis models

5.1 Multivariate meta-analysis models and synthesis methods

The contrast-based approach is one of the principal methods used for evidence synthesis in network meta-analysis [20,31]. In this framework, summary statistics such as odds ratios, hazard ratios, or mean differences are modeled as outcome measures. This approach can be interpreted as a generalization of the conventional DerSimonian–Laird method for

pairwise meta-analysis [27], whereby study-specific outcomes are treated as multivariate outcomes and modeled within a multivariate meta-analysis framework [32].

White et al. [1] demonstrated that such multivariate meta-regression models allow for the implementation of contrast-based network meta-analysis, and further showed that this framework provides a general means to evaluate inconsistency through the design-by-treatment interaction model [2]. In this section, we outline the underlying statistical model and corresponding analytical procedures.

Specifically, we consider the synthesis of N trials comparing $p+1$ treatments. Let Y_{ij} denote the estimator of the treatment effect of the *j*th treatment relative to a designated reference treatment (e.g., placebo) in the *i*th trial $(i = 1, 2, \dots, N; j = 1,2, \dots, p)$. Commonly used effect measures include the mean difference, standardized mean difference, risk difference, risk ratio, odds ratio, and hazard ratio, with ratio measures typically modeled on the logarithmic scale. For the contrast-based network meta-regression, we adopt a multivariate meta-analysis model for the outcome variable $\boldsymbol{Y}_i = \left(Y_{i1}, Y_{i2}, \dots, Y_{ip}\right)^T$,

$$\begin{aligned} \boldsymbol{Y}_i &= \boldsymbol{\theta}_i + \boldsymbol{e}_i \\ \boldsymbol{\theta}_i &= \boldsymbol{X}_i^T \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \end{aligned} \qquad (*)$$

where $\boldsymbol{\theta}_i = \left(\theta_{i1}, \theta_{i2}, \dots, \theta_{ip}\right)^T$ and $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_p^T\right)^T$. The regression function model $\boldsymbol{X}_i^T \boldsymbol{\beta}$ involves a design matrix,

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{x}_{i1} & 0 & \cdots & 0 \\ 0 & \boldsymbol{x}_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{x}_{ip} \end{pmatrix}$$

where $\boldsymbol{x}_{ij}$ is a $q_j \times 1$ covaraiate vector for Y_{ij} (q_j is the number of the covariates) and $\boldsymbol{\beta}_j$ is its $q_j \times 1$ regression coefficient vector. Also, $\boldsymbol{e}_i$ and $\boldsymbol{\varepsilon}_i$ are independent random variation terms within and across studies ($p \times 1$ random vectors), assumed to be distributed $\boldsymbol{e}_i \sim \mathrm{MVN}(\boldsymbol{0}, \boldsymbol{S}_i)$ and $\boldsymbol{\varepsilon}_i \sim \mathrm{MVN}(\boldsymbol{0}, \boldsymbol{\Sigma})$. $\boldsymbol{S}_i$ (a $p \times p$ matrix) is the within-study covariance matrix,

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

which is usually assumed to be known and fixed to its valid estimate [33]. In addition, $\boldsymbol{\Sigma}$ is the between-studies variance–covariance matrix:

$$\boldsymbol{\Sigma} = \tau^2 \boldsymbol{P} = \tau^2 \begin{pmatrix} 1 & 1/2 & \cdots & 1/2 \\ 1/2 & 1 & \dots & 1/2 \\ \vdots & \vdots & \ddots & \vdots \\ 1/2 & 1/2 & \cdots & 1 \end{pmatrix}$$

for $\tau^2 > 0$. Note that the correlation structure of $\boldsymbol{\Sigma}$ can be assumed to be unstructured, however, there are rarely enough studies to allow identification of all variance–covariance parameters. Consequently, most network meta-analyses adopt the equal-variance assumption for the p components of $\boldsymbol{\varepsilon}_i$. Under this assumption, and by virtue of the consistency property, all pairwise correlation coefficients are set to 0.50 [34]. In the `NMA` package, we likewise adopt the equal-variance assumption as the standard specification of the model.

For trials that do not include a reference treatment, the data augmentation strategy proposed by White et al. [1] can be applied; in this approach, a quasi-negligible dataset is added to the reference arm (e.g., 0.001 events for 0.01 patients in the case of a binary outcome). In addition, because most individual clinical trials typically involve only two to four arms, we formally employ subvectors or submatrices in the likelihood function [1].

Standard network meta-analysis without covariates and under the consistency assumption can be expressed as a special case of the model with $\boldsymbol{x}_{ij} = (\mathbf{1})$, a vector of ones. Meta-regression analyses incorporating covariates can then be performed within the same framework. It should be noted that the regression function is modeled separately for each outcome. Furthermore, when $\tau^2 = 0$, the model reduces to a fixed-effect model. In the `NMA` package, standard network meta-analyses can be conducted using the `nma` function, while meta-regression can be performed with the `nmareg` function.

Model parameters can be estimated using the standard REML method. Under a random-effects model, as in pairwise meta-analyses, the confidence intervals for $\boldsymbol{\beta}$ may be inaccurate when the number of studies N is small [3,16,35]. The Cochrane Handbook [22] recommends the use of the HKSJ method [15,29,30] for pairwise meta-analyses. For network meta-analysis, Noma et al. [3] also proposed a higher-order asymptotic correction, which has been adopted as the default method in the `NMA` package.

5.2 Assessment of heterogeneity

The issue of between-trial heterogeneity in treatment effects is also of central importance in network meta-analysis. The conventional measures used in pairwise meta-analysis, such as Cochrane's Q-test [36] and Higgins' I^2 statistic [37], have been extended to the multivariate meta-analysis framework by Jackson et al. [19], and can be directly computed from model (*). The parameter τ can be interpreted as the heterogeneity standard deviation. In addition, prediction intervals [38], which represent the expected range of treatment effects in future populations while accounting for between-study heterogeneity, have become widely used in recent meta-analyses. Although prediction intervals have

commonly been calculated using approximations based on the *t*-distribution, concerns about their accuracy have been raised [35,39]. To address this issue, Noma et al. [4] proposed several improved methods. The NMA package adopts, as the default method, an approach based on higher-order asymptotic approximations developed by Noma et al. [4], and these results are reported collectively in the output of the nma function.

5.3 Computational tools for the primary synthesis methods

The synthesis analysis based on the consistency assumption can be performed by the nma function, which adopts the higher-order asymptotic correction methods proposed by Noma et al. [3,4] as default (named as "Noma-Hamura" methods). The output also provides various heterogeneity statistics, including τ, τ^2, the H^2 statistic, the I^2 statistic, the Q-test, and the prediction interval. By setting eform = TRUE, the estimated effect measures are exponentiated, thereby allowing direct interpretation of effect sizes such as odds ratios, risk ratios, and hazard ratios, which are synthesized on the logarithmic scale under normality assumptions.

```
> nma(hf2, eform=TRUE)
$coding
  code treatment
1    1   Placebo
2    2        AB
3    3       ACE
4    4       ARB
5    5        BB
6    6       CCB
7    7        CT
8    8        DD

$reference
[1] "Placebo"

$`number of studies`
[1] 26

$method
[1] "Noma-Hamura's improved REML-based inference and prediction methods"

$`Coef. (vs. treat 1)`
            Coef.         SE     95%CL     95%CU     P-value
```

```
2: cons 1.2141695 0.16494000 0.8109284 1.8179257 0.283952099
3: cons 0.7141049 0.08604287 0.5859275 0.8703225 0.004313309
4: cons 0.7585898 0.08783881 0.6260565 0.9191799 0.008727958
5: cons 0.8776217 0.13324667 0.6585468 1.1695750 0.344713352
6: cons 0.8432971 0.09325718 0.6885251 1.0328601 0.092099553
7: cons 0.7745084 0.11211384 0.6106908 0.9822700 0.036691187
8: cons 0.5998794 0.11334983 0.4647085 0.7743678 0.001368851

$`tau (Between-studies_SD) estimate`
[1] 0.09876971

$`tau2 (Between-studies_variance) estimate`
[1] 0.009755455

$`Multivariate H2-statistic`
[1] 1.604202

$`Multivariate I2-statistic`
[1] 0.5939454

$`Test for Heterogeneity`
[1] "Q(df = 21) = 33.6882, p-value = 0.0391"

$`95%PI (vs. treat 1)`
               pl        pu
2: cons 0.7406591 1.9903998
3: cons 0.5245222 0.9722102
4: cons 0.5664416 1.0159185
5: cons 0.6120679 1.2583895
6: cons 0.6258326 1.1363263
7: cons 0.5632973 1.0649140
8: cons 0.4249319 0.8468542
```

Similarly, network meta-regression can be conducted using the `nmareg` function. The example below illustrates an analysis of effect modification for the odds ratio of ACE versus placebo with systolic blood pressure (SBP) included as a covariate. It should be noted that in multivariate meta-regression, a different regression function is specified for each component of the outcome vectors $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N$. The variable to be evaluated as an effect modifier is specified via the argument `z`, and the corresponding outcome components are designated using the argument `treats`. For instance, to perform a meta-

regression analysis assessing effect modification of SBP on the odds ratio for treatment 3 versus 1, the following command can be applied. Both `z` and `treats` can be specified as vectors to allow for the inclusion of multiple covariates and outcomes.

```
> nmareg(hf2,z=SBP,treats=3)
$Covariates
[1] "SBP"

$`Outcomes evaluated the effect modifications`
[1] "3 vs. 1"

$Coefficients
               Coef.          SE       95%CL        95%CU      P-value
2: cons  0.169720006 0.164951411 -0.15357882  0.493018832 3.035224e-01
3: cons  0.141231688 0.467850573 -0.77573858  1.058201961 7.627485e-01
3: SBP  -0.003243991 0.003111771 -0.00934295  0.002854968 2.971845e-01
4: cons -0.288428570 0.086905936 -0.45876108 -0.118096065 9.038598e-04
5: cons -0.154108754 0.134032892 -0.41680839  0.108590887 2.502332e-01
6: cons -0.203094458 0.094855696 -0.38900821 -0.017180710 3.226689e-02
7: cons -0.305617661 0.119742428 -0.54030851 -0.070926814 1.070167e-02
8: cons -0.535366925 0.110231127 -0.75141596 -0.319317885 1.193177e-06

$`Between-studies_SD`
[1] 0.1022552
```

In addition, to estimate the coefficients of each regression function, a sufficient number of direct comparison trials are required. As in ordinary least squares regression, the number of direct comparisons between the corresponding treatments must exceed the number of explanatory variables. If this condition is not met, the estimates from the meta-regression become indeterminate and an error will be produced.

6. Analyses of ranking and comparative effectiveness

6.1 Ranking probability and SUCRA

As an analytic approach to comparative effectiveness, treatment ranking methods are widely used [20]. Based on the estimates from model (*), the probabilities of each treatment being ranked best, second, …, or worst can be estimated using a parametric bootstrap procedure [40]. In addition, the bootstrap method can be used to estimate the surface under the cumulative ranking curve (SUCRA) [41]. SUCRA is a continuous score ranging from 0

to 1 that indicates the relative position of each treatment within the overall distribution (analogous to a quantile of the distribution). These quantities can be computed using the `nmarank` function, which also provides plots of the ranking probabilities.

```
> nmarank(hf2)
$SUCRA
             [,1]
DD      0.99172857
ACE     0.79718571
ARB     0.66017143
CT      0.61078571
CCB     0.40621429
BB      0.35471429
Placebo 0.15463571
AB      0.02456429
```

(The ranking probabilities of each treatment are presented in Figure 2.)

6.2 League table

As another type of output for comparative effectiveness, a league table [41] can also be generated easily using the `nmaleague` function. The league table can additionally be exported in CSV format.

Based on the results of the synthesis analysis, a league table can be created with user-specified options. The number of decimal places can be set via the argument `digits`, and the output file name for CSV export can be specified using `out.csv`. For example:

```
> nmaleague(hf2, eform=TRUE, digits=2, out.csv="nmaleague_out.csv")
```

(The output is omitted here due to space limitations.)

6.3 Ranked forest plot

The results of the synthesis analysis are often presented graphically using a forest plot. For comparative effectiveness analyses, a ranked forest plot [42], in which the estimated effect measures are ordered from largest to smallest, is frequently employed. Such plots can be easily generated using the `nmaforest` function.

```
> nmaforest(hf2, col.plot="blue")
```

(The forest plot created is shown in Figure 3.)

7. Evaluation of inconsistency

7.1 Local inconsistency test

Various statistical methods have been proposed for evaluating inconsistency, and several representative approaches are implemented in the `NMA` package. The first is the local inconsistency test (`local.ict` function). This method extracts triangular loops of three treatments from the evidence network and tests whether the treatment effects obtained from the direct and indirect comparison paths within each loop are consistent [43,44].

```
> local.ict(hf2)
$`loop inconsistency tests`
            N         tau X2-statistic df    P-value
loop: 1-3-4 4 0.281448603   0.01474229  1 0.90336010
loop: 1-3-6 7 0.007420627   2.21766577  1 0.13643921
loop: 1-3-8 4 0.179208394   2.62675447  1 0.10507630
loop: 1-4-6 6 0.202504385   1.16102588  1 0.28125288
loop: 1-6-8 9 0.008641779   0.66447229  1 0.41498548
loop: 3-4-5 3 0.008996435   0.38478474  1 0.53505358
loop: 3-4-6 5 0.008421575   0.01475838  1 0.90330762
loop: 3-4-7 5 0.163069127   1.64440130  1 0.19972266
loop: 3-5-6 5 0.007324249   1.30054438  1 0.25411381
loop: 3-6-7 6 0.008217668   3.30405421  3 0.34707877
loop: 3-6-8 8 0.006484820   3.53846185  3 0.31580790
loop: 4-5-6 3 0.007014825   1.97822111  1 0.15957793
loop: 4-6-7 6 0.008820186   3.66875023  1 0.05544174
```

7.2 Global inconsistency test

The second method is the global inconsistency test (`global.ict` function). While inconsistency within a triangular loop can be readily understood as the discrepancy between direct and indirect comparisons along the loop, it is less straightforward to define inconsistency across more complex networks. Higgins et al. [2] formalized this by defining inconsistency as "differences in treatment effects depending on which treatments are being compared," and expressed it through a design-by-treatment interaction framework (where "design" refers to the set of treatments being compared) [45]. In this approach, all possible design-by-treatment interactions in the network are modeled within a meta-regression framework, and a test of the global null hypothesis regarding these interactions was proposed as the global inconsistency test.

```
> global.ict(hf2)
$`Coefficients of the design-by-treatment interaction model`
```

```
                   Coef.        SE       95%CL        95%CU      P-value
2: cons      -0.27666415 0.3427377 -0.94841776  0.395089454 0.4195407483
3: cons      -0.27578681 0.1707338 -0.61041900  0.058845381 0.1062452021
4: cons      -0.15952053 0.1470291 -0.44769233  0.128651272 0.2779399668
4: des_3-4   -0.06539617 0.2814742 -0.61707551  0.486283172 0.8162782248
5: cons      -0.45753924 0.5034281 -1.44424025  0.529161770 0.3634307976
5: des_4-5    0.35461626 0.5591588 -0.74131478  1.450547305 0.5259529069
6: cons      -0.40088070 0.2009496 -0.79473469 -0.007026721 0.0460503444
6: des_3-6    0.31177011 0.6850657 -1.03093398  1.654474210 0.6490407487
6: des_3-6-7  0.36952186 0.3271175 -0.27161671  1.010660431 0.2586320498
6: des_3-6-8  0.27989308 0.3118060 -0.33123545  0.891021611 0.3693714463
6: des_4-6    0.37719158 0.3033194 -0.21730351  0.971686683 0.2136667518
6: des_5-6   -0.23257448 0.5763566 -1.36221264  0.897063667 0.6865618008
7: cons      -0.09380816 0.2586352 -0.60072379  0.413107467 0.7168255825
7: des_3-7   -0.31165678 0.3016609 -0.90290122  0.279587666 0.3015401478
7: des_4-7    0.34052387 0.3659629 -0.37675028  1.057798021 0.3521188091
7: des_6-7   -0.52971412 0.3634510 -1.24206494  0.182636693 0.1449903040
8: cons      -0.98175109 0.2974292 -1.56470171 -0.398800476 0.0009641348
8: des_3-6-8  0.52475844 0.3807743 -0.22154538  1.271062268 0.1681615188
8: des_3-8    0.83328328 0.4123163  0.02515811  1.641408450 0.0432817831
8: des_6-8    0.15579853 0.4383138 -0.70328067  1.014877732 0.7222526593

$`Between-studies_SD`
[1] 0.156216

$`P-value`
[1] 0.4586125
```

As a supplementary tool, the `nmaQ` function can be used to decompose the *Q*-statistic, providing *Q*-statistics for each design as well as separating the within-design and between-design components. This allows the results of heterogeneity tests to be obtained for each component.

```
> nmaQ(hf2)
$`Within designs (individual designs)`
   design N     n          Q df    P-value
1     1-3 1  9297 0.00000000  0 1.00000000
2     1-4 2  7439 5.15663031  1 0.02315779
3     1-6 3 16800 0.50202309  2 0.77801339
4     1-8 1  3845 0.00000000  0 1.00000000
```

```
5      2-8 1 24335 0.00000000  0 1.00000000
6      3-4 1 17118 0.00000000  0 1.00000000
7      3-5 1   758 0.00000000  0 1.00000000
8      3-6 1   470 0.00000000  0 1.00000000
9    3-6-7 1  6614 0.00000000  0 1.00000000
10   3-6-8 1 33357 0.00000000  0 1.00000000
11     3-7 1 10985 0.00000000  0 1.00000000
12     3-8 1  6083 0.00000000  0 1.00000000
13     4-5 1  9193 0.00000000  0 1.00000000
14     4-6 1 15245 0.00000000  0 1.00000000
15     4-7 2  5129 1.33869867  1 0.24726359
16     5-6 1 19257 0.00000000  0 1.00000000
17     6-7 2 27357 0.07400339  1 0.78559461
18     6-8 4 10031 4.58073728  3 0.20520078

$`Q-statistics`
                       Q df    P-value
Within designs  11.65209  8 0.16741465
Between designs 22.03614 13 0.05480433
Total           33.68823 21 0.03912123
```

7.3 Sidesplitting

The third method is side-splitting (`sidesplit` function). This approach summarizes the direct and indirect evidence for a specific pair of treatments and tests whether their pooled quantities are in disagreement. When multi-arm trials with three or more arms are included, special considerations are required to avoid bias [46,47]. The `NMA` package adopts the approach proposed by Noma {Noma, 2024 #1780}, which is based on network meta-regression.

```
> sidesplit(hf2)
$`Direct evidence`
              Coef.        SE       95%CL        95%CU      P-value
1 vs. 3 -0.27578485 0.1431193 -0.55629345  0.004723748 5.398483e-02
1 vs. 4 -0.15528793 0.1305472 -0.41115580  0.100579939 2.342369e-01
1 vs. 6 -0.39198658 0.1797154 -0.74422231 -0.039750846 2.917220e-02
1 vs. 8 -0.98175078 0.2607196 -1.49275173 -0.470749835 1.661873e-04

$`Indirect evidence`
              Coef.          SE         95%CL        95%CU      P-value
```

```
1 vs. 3 -0.40692535  0.12589589   -0.65367677  -0.16017393 1.228252e-03
1 vs. 4 -0.43966819  0.13678629   -0.70776439  -0.17157199 1.307762e-03
1 vs. 6 -0.09977613  0.09199397   -0.28008099   0.08052874 2.781014e-01
1 vs. 8 -0.42782143  0.09339764   -0.61087744  -0.24476541 4.635446e-06

$Difference
              Coef.         SE         95%CL        95%CU    P-value
1 vs. 3  0.13114050  0.1906119   -0.24245197   0.50473298 0.49145440
1 vs. 4  0.28438026  0.1890848   -0.08621917   0.65497969 0.13258586
1 vs. 6 -0.29221045  0.2018923   -0.68791217   0.10349128 0.14779675
1 vs. 8 -0.55392936  0.2769437   -1.09672902  -0.01112969 0.04548390
```

(Due to space limitations, only the top four rows of results are presented.)

7.4 Random inconsistency model

The fourth method is the random inconsistency model proposed by Jackson et al. [17] (`random.icm` function). This approach extends the design-by-treatment interaction meta-regression model of Higgins et al. [2] by treating the design effects as random effects [17,18]. In this model, inconsistency can be evaluated by testing the variance component associated with the random design effects. In addition, this model provides I^2 statistics for heterogeneity, for inconsistency, and for the overall model that jointly accounts for both heterogeneity and inconsistency.

```
> random.icm(hf2)
$`number of designs`
[1] 18

$`Coef. (vs. treat 1)`
             Coef.         SE      95%CL        95%CU      P-value
2: cons  0.1876289 0.16906583 -0.1437341  0.518991807 2.670862e-01
3: cons -0.3423417 0.08765170 -0.5141359 -0.170547531 9.395095e-05
4: cons -0.2833837 0.09060082 -0.4609581 -0.105809374 1.761036e-03
5: cons -0.1385517 0.13637643 -0.4058446  0.128741183 3.096529e-01
6: cons -0.1770212 0.09351189 -0.3603012  0.006258688 5.835326e-02
7: cons -0.2596636 0.11418011 -0.4834525 -0.035874671 2.295651e-02
8: cons -0.5174581 0.11160982 -0.7362094 -0.298706888 3.546781e-06

$`Between-studies_SD`
[1] 0.09221444
```

```
$`Between-designs_SD`
[1] 0.05496757

$`Likelihood ratio tests for the variance components`
                              X2-statistic df   P-value
Heterogeneity                    0.4793887  1 0.4886993
Inconsistency                    0.0393532  1 0.8427506
Heterogeneity + Inconsistency    0.5187419  2 0.7715368

$`Heterogeneity and inconsistency statistics`
                              R-statistic I2-statistic
Heterogeneity                    1.523210   0.56899691
Inconsistency                    1.051398   0.09538123
Heterogeneity + Inconsistency    1.601500   0.61010652
```

8. Other graphical tools

8.1 Comparison-adjusted funnel plot

In network meta-analysis, the assessment of publication bias is challenging because multiple treatments are compared and the direction of publication bias may differ across trials. One approach is the comparison-adjusted funnel plot [40], in which the results of trials with direct comparisons against a specified reference treatment are adjusted to share a common center line, thereby allowing a descriptive evaluation of publication bias across the network. Such plots can be generated using the `nmafunnel` function.

```
> nmafunnel(wd4,legends="bottomright")
```

(The comparison-adjusted funnel plot created is shown in Figure 4.)

8.2 Study weights and contribution matrix

In network meta-analysis, it is important to assess how individual studies influence the synthesized results. Jackson et al. [48] and Noma et al. [47] demonstrated that contribution rates can be estimated using factorized information, and that contribution weight matrices can be derived accordingly. In the `NMA` package, the `nmaweight` function allows users to compute these contribution weight matrices and to visualize them graphically as a heatmap.

```
> nmaweight(wd4)
```

(*The heat plot created is shown in Figure 5.*)

8.3 Check of transitivity

In network meta-analysis, not only consistency but also transitivity is regarded as a fundamental and essential assumption [20], as it underpins the premise that indirect comparisons are theoretically valid even in the absence of direct evidence. To date, no standard statistical method has been firmly established for evaluating transitivity. However, Cipriani, Furukawa, Salanti et al. [49] illustrated an approach in which the distributions of key covariates are compared across designs using side-by-side boxplots, thereby enabling an assessment of heterogeneity in covariate distributions. In the `NMA` package, the `transitivity` function generates such side-by-side boxplots to facilitate the evaluation of transitivity.

```
> transitivity(hf2, SBP, yrange=c(100,220))
```

(*The side-by-side boxplot created is shown in Figure 6.*)

9. Concluding remarks

Network meta-analysis is now a well-established methodology within systematic reviews and has already become widely adopted as a standard approach for comparative effectiveness research [22,50]. However, the statistical methods are often highly technical for non-statisticians. In the practice of systematic reviews, statistical experts are not always involved, and many researchers lack either the time or the technical skills to implement their own programming. Therefore, the availability of user-friendly software with versatile functionality is of critical importance for enabling practitioners to conduct research on important clinical and public health questions. We hope that the `NMA` package will encourage more researchers and practitioners to undertake network meta-analysis and will contribute to the generation of new evidence that can ultimately improve clinical practice and public health.

Acknowledgements. This `NMA` package was originally developed for practical training in "Meta-analysis" class at the Kyoto University School of Public Health. We would like to express our gratitude to all those who contributed to this work.

Author contributions. Conceptualization: H.N., K.M.; Formal analysis: H.N.; Investigation: H.N., K.M.; Methodology: H.N., K.M., S.T., T.A.F.; Supervision: S.T., T.A.F.; Writing—original draft: H.N.; Writing—review and editing: H.N., K.M., S.T., T.A.F.

Competing interest statement. The authors declare no conflicts of interest.

Data availability statement. All datasets used in this article are available in `NMA` package on CRAN (https://doi.org/10.32614/CRAN.package.NMA). Also, complete R programs reproducing all case analyses are available on the authors' GitHub page (https://github.com/nomahi/NMA).

Funding statement. This study was supported by Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (grant number: JP23K24811).

References

[1] White IR, Barrett JK, Jackson D, Higgins JP. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Res Synth Meth*. 2012;3(2):111-125.

[2] Higgins JP, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Meth*. 2012;3(2):98-110.

[3] Noma H, Hamura Y, Gosho M, Furukawa TA. Kenward-Roger-type corrections for inference methods of network meta-analysis and meta-regression. *Res Synth Methods*. 2023;14(5):731-741.

[4] Noma H, Hamura Y, Sugasawa S, Furukawa TA. Improved methods to construct prediction intervals for network meta-analysis. *Res Synth Methods*. 2023;14(6):794-806.

[5] Del Fiol G, Workman TE, Gorman PN. Clinical questions raised by clinicians at the point of care: a systematic review. *JAMA Intern Med*. 2014;174(5):710-8.

[6] Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. *Ann Intern Med*. 2009;151(3):203-5.

[7] Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ*. 2005;331:897-900.

[8] Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Ann Intern Med*. 2013;159(2):130-7.

[9] Higgins JP, Welton NJ. Network meta-analysis: a norm for comparative effectiveness? *Lancet*. 2015;386(9994):628-30.

[10] Naci H, O'Connor AB. Assessing comparative effectiveness of new drugs before approval using prospective network meta-analyses. *J Clin Epidemiol*. 2013;66(8):812-6.

[11] van Valkenhoef G, Lu G, de Brock B, Hillege H, Ades AE, Welton NJ. Automating network meta-analysis. *Res Synth Methods*. 2012;3(4):285-99.

[12] Lin L, Zhang J, Hodges JS, Chu H. Performing arm-based network meta-analysis in R with the pcnetmeta package. *J Stat Softw*. 2017;80(2017):5.

[13] Balduzzi S, Rücker G, Nikolakopoulou A, et al. netmeta: An R Package for network meta-analysis using frequentist methods. *J Stat Softw*. 2023;106:1-40.
[14] White IR. Network meta-analysis. *Stata J*. 2015;15(4):951-985.
[15] IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol*. 2014;14:25.
[16] Noma H, Nagashima K, Maruo K, Gosho M, Furukawa TA. Bartlett-type corrections and bootstrap adjustments of likelihood-based inference methods for network meta-analysis. *Stat Med*. 2018;37(7):1178-1190.
[17] Jackson D, Barrett JK, Rice S, White IR, Higgins JP. A design-by-treatment interaction model for network meta-analysis with random inconsistency effects. *Stat Med*. 2014;33(21):3639-3654.
[18] Law M, Jackson D, Turner R, Rhodes K, Viechtbauer W. Two new methods to fit models for network meta-analysis with random inconsistency effects. *BMC Med Res Methodol*. 2016;16:87.
[19] Jackson D, White IR, Riley RD. Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. *Stat Med*. 2012;31(29):3805-3820.
[20] Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Meth*. 2012;3(2):80-97.
[21] Sciarretta S, Palano F, Tocci G, Baldini R, Volpe M. Antihypertensive treatment and development of heart failure in hypertension: a Bayesian network meta-analysis of studies in patients with hypertension and high cardiovascular risk. *Arch Intern Med*. 2011;171(5):384-94.
[22] Higgins JPT, Thomas J. *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd ed. Wiley-Blackwell; 2019.
[23] Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med*. 1998;17(24):2815-2834.
[24] Hedeker D, Siddiqui O, Hu FB. Random-effects regression analysis of correlated grouped-time survival data. *Stat Methods Med Res*. 2000;9(2):161-79.
[25] Salika T, Turner RM, Fisher D, Tierney JF, White IR. Implications of analysing time-to-event outcomes as binary in meta-analysis: empirical evidence from the Cochrane Database of Systematic Reviews. *BMC Med Res Methodol*. 2022;22(1):73.
[26] Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Annals of Internal Medicine*. 2015;162(11):777-784.
[27] DerSimonian R, Laird NM. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188.

[28] Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629-34.
[29] Hartung J, Knapp G. On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Stat Med*. 2001;20(12):1771-82.
[30] Sidik K, Jonkman JN. A simple confidence interval for meta-analysis. *Stat Med*. 2002;21(21):3153-3159.
[31] Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23:3105-3124.
[32] Jackson D, Riley R, White IR. Multivariate meta-analysis: potential and promise. *Stat Med*. 2011;30(20):2481-98.
[33] Noma H. Within-study covariance estimators for network meta-analysis with contrast-based approach. *Jpn J Biometrics*. 2024;44(2):119-126.
[34] Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med*. 1996;15(24):2733-2749.
[35] Partlett C, Riley RD. Random effects meta-analysis: Coverage performance of 95% confidence and prediction intervals following REML estimation. *Stat Med*. 2017;36(2):301-317.
[36] Cochran WG. The combination of estimates from different experiments. *Biometrics*. 1954;10:101-129.
[37] Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539-1558.
[38] Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J Royal Stat Soc A*. 2009;172(1):137-159.
[39] Nagashima K, Noma H, Furukawa TA. Prediction intervals for random-effects meta-analysis: A confidence distribution approach. *Stat Methods Med Res*. 2019;28(6):1689-1702.
[40] Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PloS One*. 2013;8:e76654.
[41] Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol*. 2011;64(2):163-71.
[42] Nikolakopoulou A, Higgins JPT, Papakonstantinou T, et al. CINeMA: An approach for assessing confidence in the results of a network meta-analysis. *PLoS Med*. 2020;17(4):e1003082.
[43] Veroniki AA, Vasiliadis HS, Higgins JP, Salanti G. Evaluation of inconsistency in networks of interventions. *Int J Epidemiol*. 2013;42(1):332-45.
[44] Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol*. 1997;50(6):683-691.
[45] Jackson D, Boddington P, White IR. The design-by-treatment interaction model: a

unifying framework for modelling loop inconsistency in network meta-analysis. *Res Synth Methods*. 2016;7(3):329-32.

[46] Noma H. Sidesplitting using network meta-regression. *Jpn J Biometrics*. 2024;44(2):107-118.

[47] Noma H, Tanaka S, Matsui S, Cipriani A, Furukawa TA. Quantifying indirect evidence in network meta-analysis. *Stat Med*. 2017;36(6):917-927.

[48] Jackson D, White IR, Price M, Copas J, Riley RD. Borrowing of strength and study weights in multivariate and network meta-analysis. *Stat Methods Med Res*. 2017;26(6):2853-2868.

[49] Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet*. 2018;391(10128):1357-1366.

[50] Egger M, Higgins JP, Smith GD. *Systematic Reviews in Health Research: Meta-Analysis in Context*. 3rd ed. BMJ Books; 2022.

Table 1. Main functions of the `NMA` Package and their descriptions.

Function	Description	Section
`setup`	Transforming arm-level data into contrast-based summary statistics and creating data objects for all subsequent analyses.	3.2
`trans.armdata`	Transforming reported contrast-based summary statistics into arm-level data.	3.2
`trans.armdataP`	Transforming reported incidence proportion or survival probability data into arm-level data.	3.2
`netplot`	Generating a network diagram.	4.1
`pairwise`	Pairwise meta-analyses for all directly compared treatment pairs in the network.	4.2
`nma`	Performs a network meta-analysis using a multivariate meta-analysis model and computes heterogeneity measures.	5.3
`nmareg`	Performs a network meta-regression using a contrast-based approach.	5.3
`nmarank`	Calculates ranking statistics for the network meta-analysis.	6.1
`nmaleague`	Generates a league table.	6.2
`nmaforest`	Generates a ranked forest plot of the network meta-analysis results.	6.3
`local.ict`	Local inconsistency tests for all closed loops in the network.	7.1
`global.ict`	Higgins' global inconsistency test.	7.2
`nmaQ`	Computes the multivariate Q-statistic and its factorization.	7.2
`sidesplit`	Conducts sidesplitting to quantify direct and indirect evidence for all possible treatment pairs and to test inconsistency.	7.3
`random.icm`	Jackson's random inconsistency model.	7.4
`nmafunnel`	Generates a comparison-adjusted funnel plot.	8.1
`nmaweight`	Evaluates study weights and the contribution matrix.	8.2
`transitivity`	Assesses transitivity.	8.3

Table 2. Outcome types supported in the `NMA` package and their descriptions.

Type of outcome	Effect measure	Description
Continuous	Mean difference Standardized mean difference	For the included trials, arm-level data on the mean, standard deviation, and sample size are required. If some trials report only summary statistics, the `trans.armdata` function can be used to back-calculate the corresponding arm-level data (see e-Appendix A).
Binary	Risk difference Risk ratio Odds ratio	For the included trials, arm-level data on the number of events, and sample size are required. If some trials report only summary statistics, the `trans.armdata` function can be used to back-calculate the corresponding arm-level data (see Section 3.2).
Survival	Hazard ratio Survival probability difference	For the included trials, either the estimated log-hazard ratio with its standard error or arm-level data on the number of events and sample size are required. For convenience, the `trans.armdata` function can be used to convert the estimated log-hazard ratio, under the proportional hazards assumption, into arm-level data on the number of events and sample size that carry equivalent statistical information. Also, the arm-level data can be estimated from survival probability estimates using the `trans.armdataP` function. When the analysis is performed using the `setup` function, the original hazard ratio estimates are restored, enabling all subsequent analyses to be conducted (see e-Appendix B).

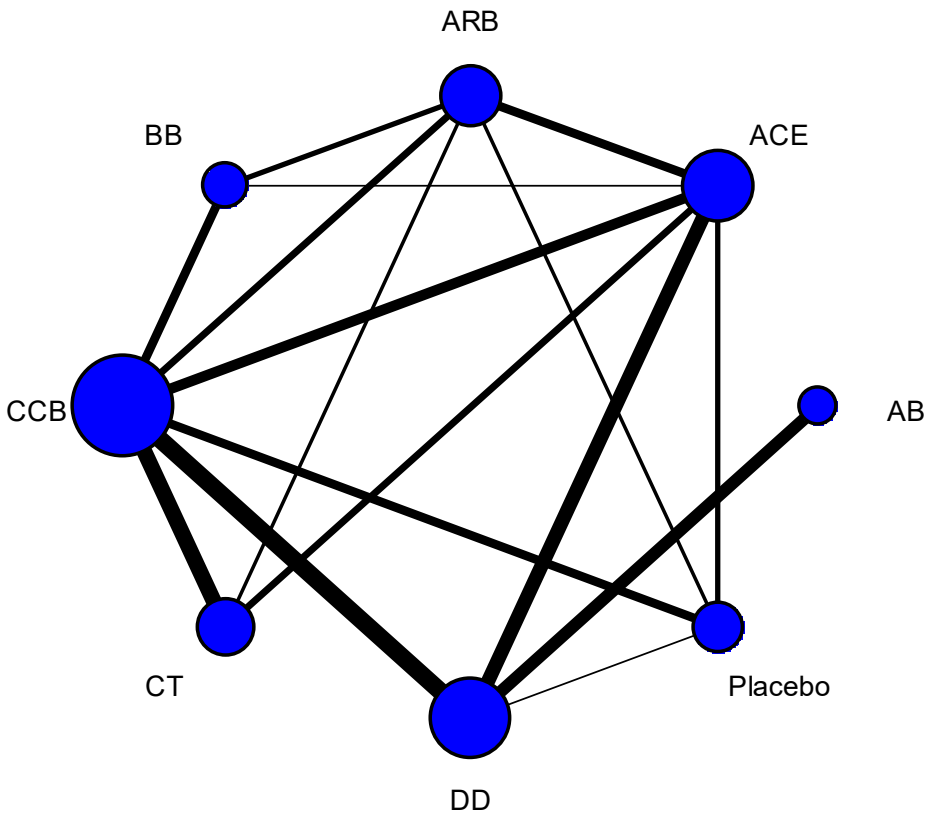

Figure 1. Network diagram for the antihypertensive drug study created by the `netplot` function.

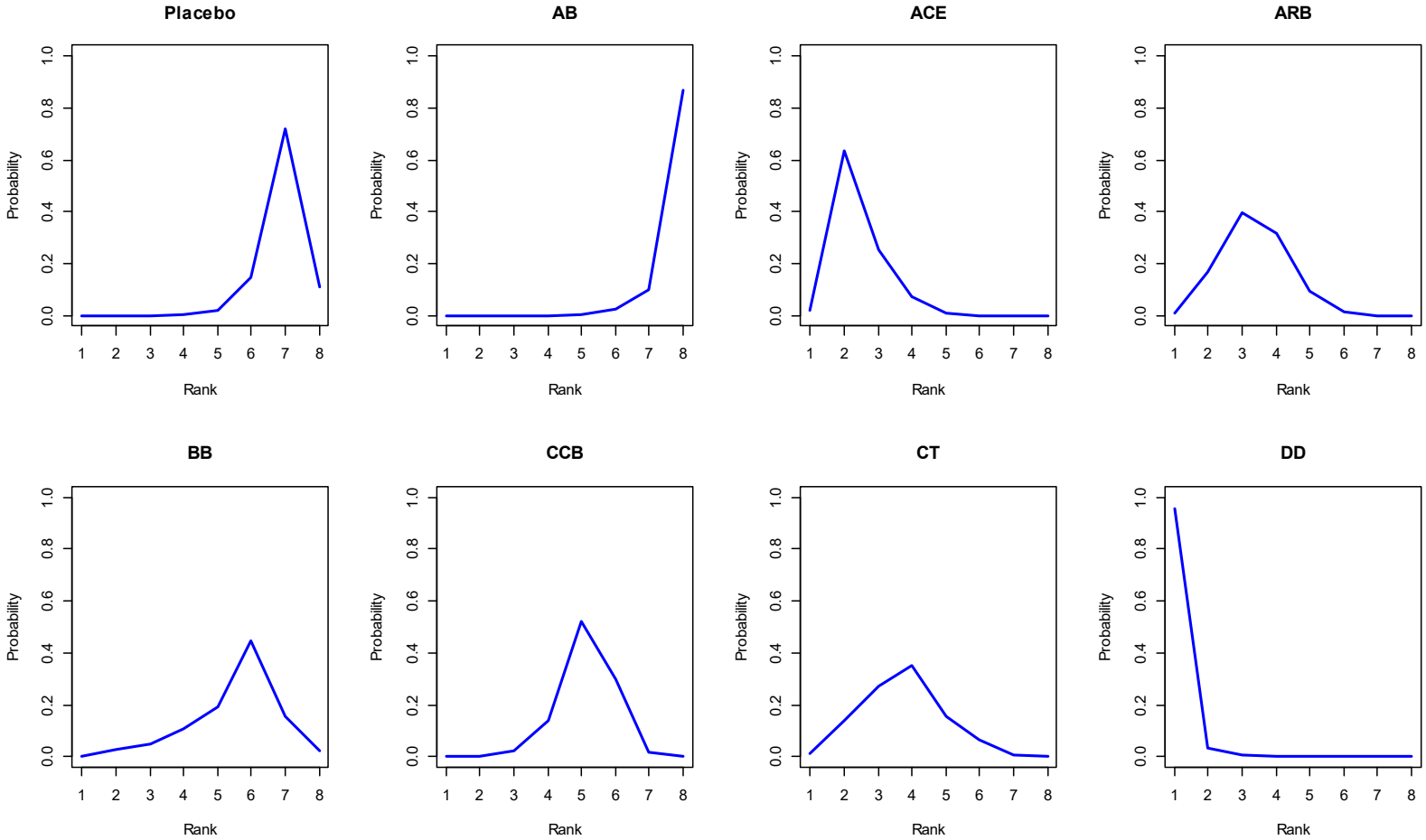

Figure 2. Ranking probability plots for the antihypertensive drug study created by the `nmarank` function.

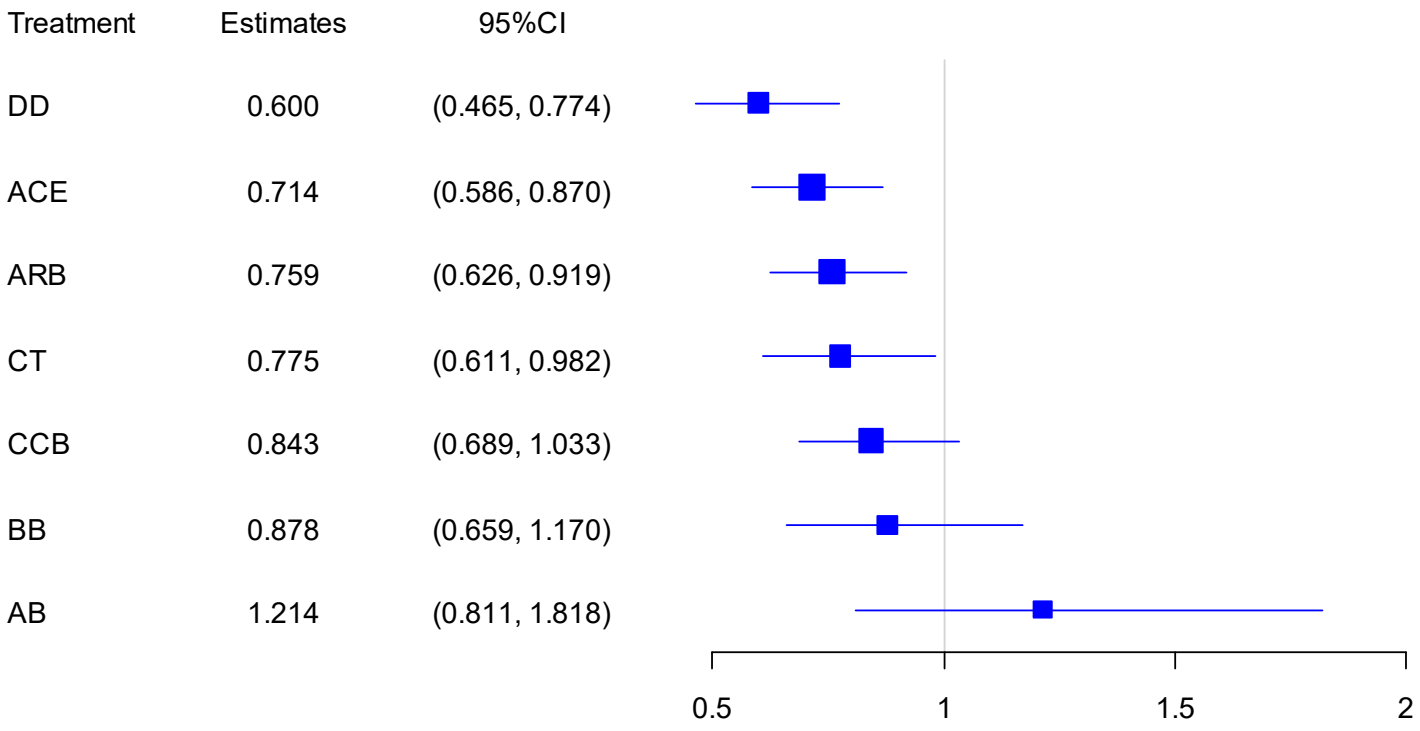

Figure 3. Ranked forest plot for the antihypertensive drug study created by the `nmaforest` function.

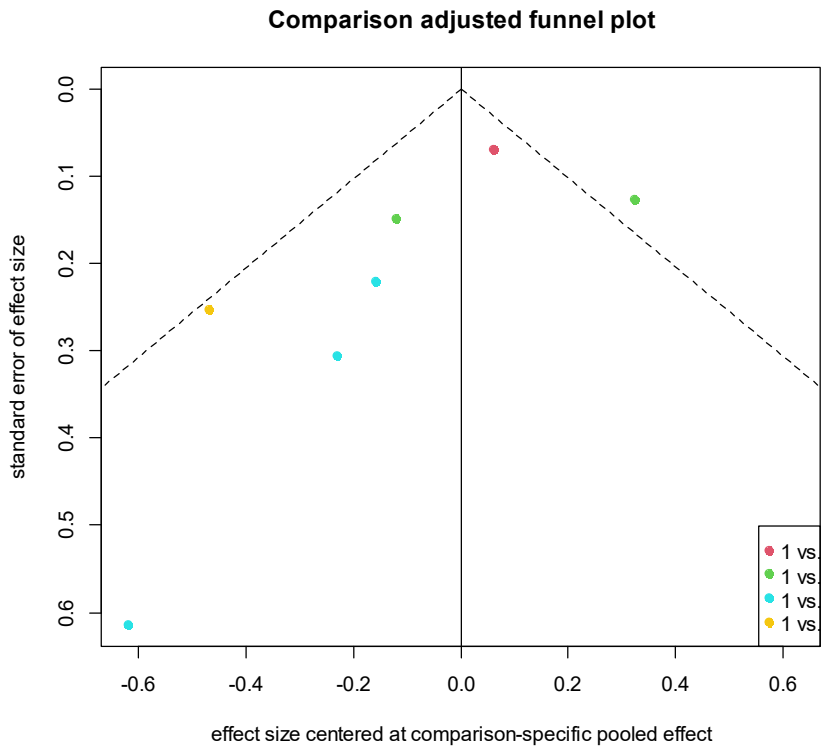

Figure 4. Comparison-adjusted funnel plot for the antihypertensive drug study created by the `nmafunnel` function.

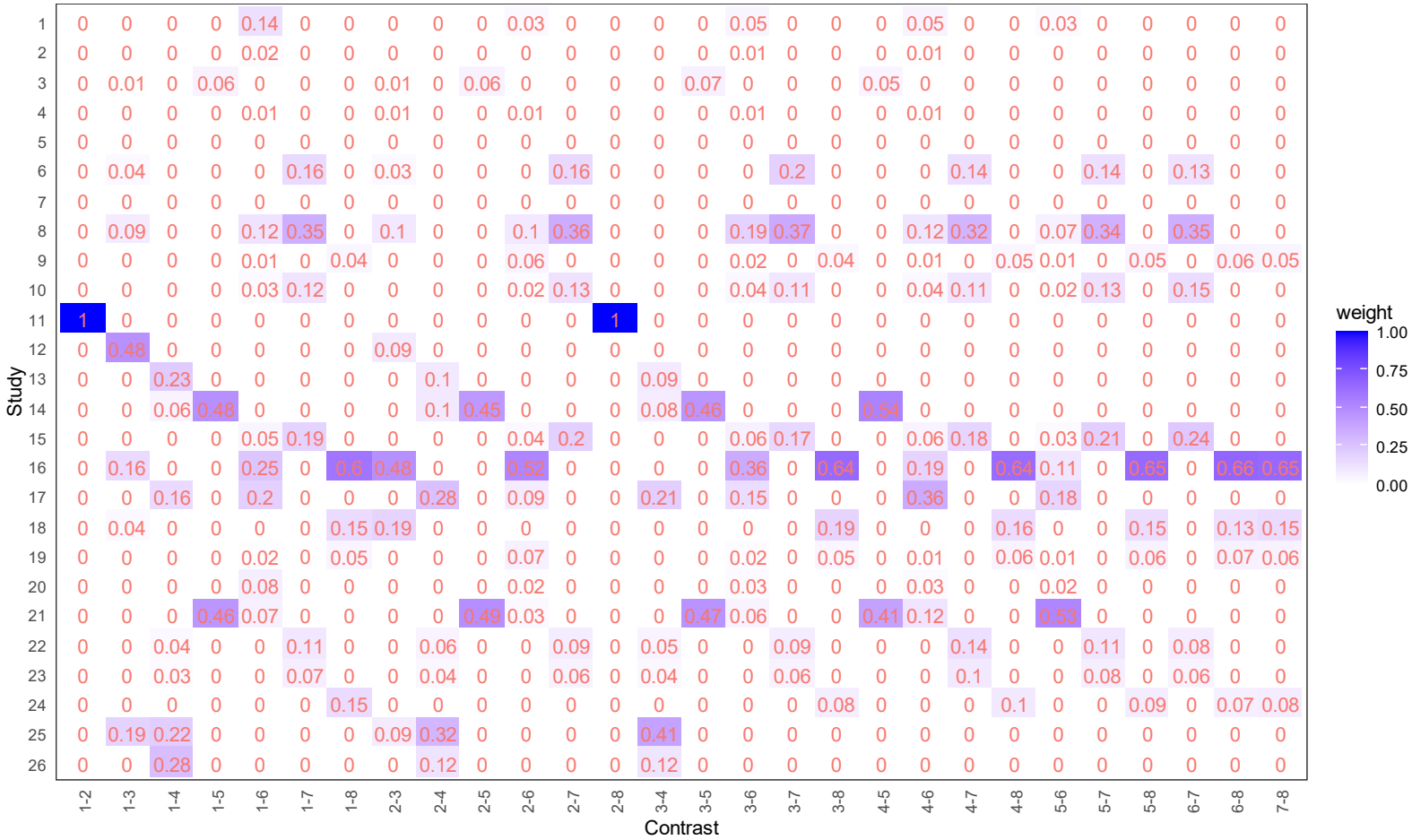

Figure 5. Heat plot for the contribution weights of individual studies for the antihypertensive drug study created by the `nmaweight` function.

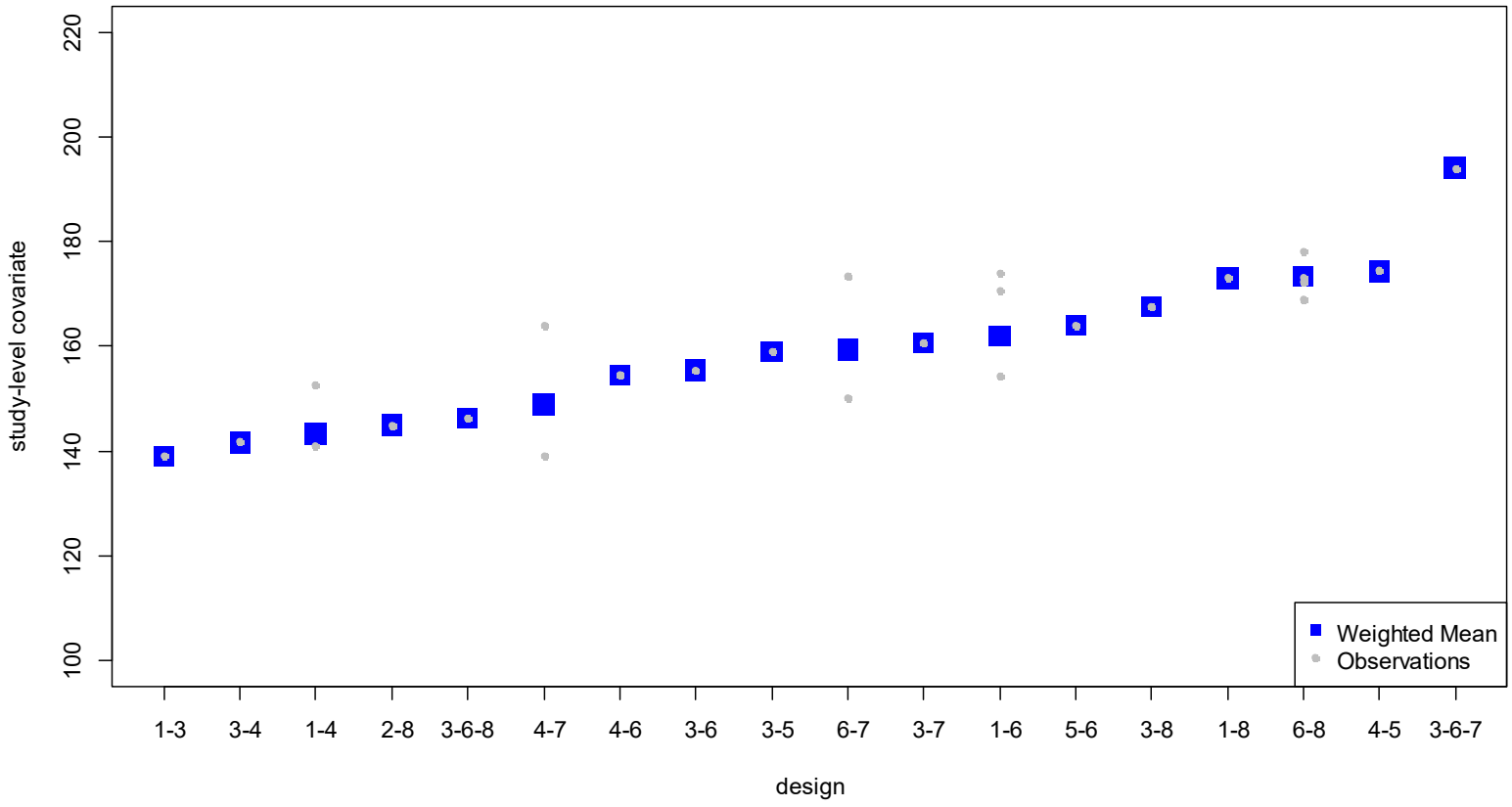

Figure 6. Side-by-side boxplot for assessing transitivity for the antihypertensive drug study created by the `transitivity` function.

SUPPLEMENTARY MATERIAL for

NMA: Network meta-analysis based on multivariate meta-analysis and meta-regression models in R

Hisashi Noma[1,2], Kazushi Maruo[3], Shiro Tanaka[4], and Toshi A. Furukawa[5]

[1] *Department of Interdisciplinary Statistical Mathematics, The Institute of Statistical Mathematics, Tokyo, Japan*

[2] *The Graduate Institute for Advanced Studies, The Graduate University for Advanced Studies (SOKENDAI), Tokyo, Japan*

[3] *Department of Biostatistics, Institute of Medicine, University of Tsukuba, Tsukuba, Japan*

[4] *Department of Clinical Biostatistics, Kyoto University Graduate School of Medicine, Kyoto, Japan*

[5] *Kyoto University Office of Institutional Advancement and Communications, Kyoto, Japan*

e-Appendix A: Illustrative example of continuous outcome case

A.1 Illustrative example: Network meta-analysis of antidiabetic drugs

In this section, we illustrate the application of network meta-analysis for continuous outcomes using the case study by Phung et al. [1] on antidiabetic drugs for type 2 diabetes. Phung et al. [1] conducted an network meta-analysis comparing six antidiabetic drug classes [α-glucosidase inhibitors (AGI), dipeptidyl peptidase-4 (DPP-4) inhibitors, glinides, glucagon-like peptide-1 (GLP-1) analogues, sulfonylureas, and thiazolidinediones] with placebo. The outcome of interest was the change in HbA1c. For all included trials, arm-level data were available in terms of the mean outcome (`y`), standard deviation (`sd`), and sample size (`n`). When such data are reported at the arm level, all analyses for continuous outcomes can in principle be conducted. In the `NMA` package, this dataset is provided as `antidiabetic`.

```
> data(antidiabetic)
> head(antidiabetic, 10)
   id               t     y   sd    n
1   1 DPP-4 inhibitor -0.69 0.95  186
2   1         Placebo  0.13 0.93  175
3   2 DPP-4 inhibitor -0.44 0.67 1118
4   2    Sulfonylurea -0.53 0.65 1072
5   3 DPP-4 inhibitor -0.66 1.20  119
6   3         Placebo  0.17 1.19  117
7   4 DPP-4 inhibitor -0.60 1.45  210
8   4         Placebo -0.10 1.02  104
9   5    GLP-1 analog -1.00 1.56  242
10  5    Sulfonylurea -1.00 1.56  242
```

A.2 Aggregating arm-level and summary level data

On the other hand, in conducting a systematic review, it is not always that arm-level data are reported for all eligible trials. In some studies, only summary statistics, such as an estimated mean difference and its confidence interval, may be available. Such trials cannot be directly incorporated into analyses using the NMA package. However, excluding these studies from the analysis would induce a bias analogous to reporting bias.

To address this issue, the NMA package provides the function `trans.armdata`, which reconstructs arm-level data from available summary statistics. As an illustration, the dataset `exdataMD` includes summary-level information from three trials in Phung et al. [1]. When these data are processed with the `trans.armdata` function, the following output is obtained.

```
> data(exdataMD)
> print(exdataMD)
  id            treat1  treat2    MD       seMD  n1  n2
1 10   DPP-4 inhibitor Placebo -0.51 0.09943817  91  88
2 10 Thiazolidinedione Placebo -0.57 0.09904131  87  88
3 14      GLP-1 analog Placebo -0.90 0.14102030 113 113
4 15      Sulfonylurea Placebo -0.47 0.14123580  61  61

> trans.armdata(study=id,treat1=treat1,treat2=treat2,n1=n1,n2=n2,y=MD,
SE=seMD,  measure="MD",data=exdataMD)
  study               trt     m         s   n
1     1   DPP-4 inhibitor -0.51 0.6651026  91
5     1           Placebo  0.00 0.6600943  88
2     1 Thiazolidinedione -0.57 0.6550860  87
3     2      GLP-1 analog -0.90 1.0600002 113
7     2           Placebo  0.00 1.0600002 113
8     3           Placebo  0.00 0.7800002  61
4     3      Sulfonylurea -0.47 0.7800002  61
```

By combining the reconstructed arm-level data with the remaining dataset, it becomes possible to generate an analysis-ready dataset for use in the NMA package. An important caveat, however, is that the reconstructed arm-level data may not coincide exactly with the original arm-level data. This discrepancy can arise because the back-calculation is based on large-sample approximations, or because the reported summary statistics were derived from adjusted analyses such as multivariable models.

Nevertheless, since all analyses in the NMA package are conducted within a contrast-

based framework, the reconstructed arm-level data are subsequently transformed back into summary statistics through the `setup` function prior to analysis. This transformation ensures that the data ultimately used in the network meta-analysis coincide, in principle, with the original reported summary statistics.

A.3 Data preparation and setup

Once arm-level data have been obtained for all eligible trials—whether directly reported or reconstructed using the methods described in the previous section—the `setup` function can be used to generate a dataset suitable for conducting a contrast-based network meta-analysis. The outcome measure may be expressed either as the mean difference or as the standardized mean difference. In the `NMA` package, the standardized mean difference is, by default, calculated as Cohen's d [2].

If summary data on arm-level covariates are available and intended for use in multivariable meta-regression or in assessing transitivity, these variables should be incorporated into the dataset prior to applying the `setup` function. Data editing itself can be conveniently performed using software such as Microsoft Excel before importing the dataset into the `NMA` package.

A.4 Network meta-analysis

Once the dataset object has been created using the `setup` function, all subsequent analyses can be carried out with simple commands, following the same procedures described in the main text. The illustrative program can be fully reproduced by running R script (B), which is provided at the GitHub page (https://github.com/nomahi/NMA).

As an example, the results of the integrated analysis using the `nma` function are shown below.

```
> nma(ad2)
$coding
  code         treatment
1    1           Placebo
2    2               AGI
3    3   DPP-4 inhibitor
4    4           Glinine
5    5      GLP-1 analog
6    6      Sulfonylurea
7    7 Thiazolidinedione

$reference
[1] "Placebo"
```

```
$`number of studies`
[1] 20

$method
[1] "Noma-Hamura's improved REML-based inference and prediction methods"

$`Coef. (vs. treat 1)`
             Coef.         SE      95%CL      95%CU      P-value
2: cons -0.6394145 0.17507898 -0.9915386 -0.2872905 6.476316e-04
3: cons -0.7535312 0.06470675 -0.8917243 -0.6153381 8.321657e-09
4: cons -1.0800000 0.36246816 -1.7942908 -0.3657092 3.205325e-03
5: cons -0.9264918 0.13321761 -1.2038307 -0.6491529 7.932971e-07
6: cons -0.7924970 0.08050693 -0.9652662 -0.6197278 1.197181e-07
7: cons -0.7853252 0.09838071 -0.9960916 -0.5745589 1.295332e-06

$`tau (Between-studies_SD) estimate`
[1] 0.1420544

$`tau2 (Between-studies_variance) estimate`
[1] 0.02017947

$`Multivariate H2-statistic`
[1] 2.410204

$`Multivariate I2-statistic`
[1] 0.6140172

$`Test for Heterogeneity`
[1] "Q(df = 16) = 38.5633, p-value = 0.0013"

$`95%PI (vs. treat 1)`
               pl         pu
2: cons -1.093121 -0.1857079
3: cons -1.089096 -0.4179664
4: cons -1.847206 -0.3127942
5: cons -1.333210 -0.5197735
6: cons -1.145481 -0.4395127
7: cons -1.158133 -0.4125171
```

e-Appendix B: Illustrative example of survival outcome case

B.1 Illustrative example: Network meta-analysis of COPD

As an example of a survival outcome, we introduce the case study by Woods et al. [3], which reported a network meta-analysis of treatments for chronic obstructive pulmonary disease (COPD). In this study, five trials comparing three active treatments—fluticasone, salmeterol, and the salmeterol–fluticasone combination (SFC)—with placebo were synthesized. The outcome of interest was overall survival. However, hazard ratio estimates were available from only two of the trials, whereas in the remaining three trials only the cumulative number of deaths and sample sizes in each arm were reported. These data are provided in the datasets `woods1` and `woods2`.

```
> data(woods1)
> print(woods1)
  studlab      treat1      treat2     TE       seTE   n1   n2
1       1 Fluticasone     Placebo -0.276 0.20300000  376  375
2       2         SFC     Placebo -0.209 0.09800000 1533 1524
3       2  Salmeterol     Placebo -0.154 0.09600000 1521 1524
4       2 Fluticasone     Placebo  0.055 0.09200000 1534 1524
5       2         SFC  Salmeterol -0.056 0.10000000 1533 1521
6       2         SFC Fluticasone -0.264 0.09600000 1533 1534
7       2  Salmeterol Fluticasone -0.209 0.09396776 1521 1534

> data(woods2)
> print(woods2)
  study         trt d   n
1     3  Salmeterol 1 229
2     3     Placebo 1 227
3     4 Fluticasone 4 374
4     4  Salmeterol 3 372
5     4         SFC 2 358
6     4     Placebo 7 361
7     5  Salmeterol 1 554
8     5     Placebo 2 270
```

In meta-analyses of time-to-event outcomes, it is common that some trials do not report an estimate of the effect measure of interest—most often the hazard ratio. Parmar et al. [4] described methods for deriving hazard ratio estimates from a variety of reported data. In principle, such methods form the basis for reconstructing trial-level information from published reports, and for organizing these into the two types of datasets required as a

prerequisite for conducting analyses using the NMA package.

B.2 Aggregating hazard ratio estimates and dichotomized data

For conducting an integrated network meta-analysis, it is necessary to combine these two types of datasets. As noted in the main text, the NMA package implements methods based on the contrast-based approach, and thus hazard ratio estimates are synthesized and analyzed using multivariate meta-analysis models. However, covariate summary data—used in meta-regression or transitivity analyses—are often reported at the arm level. To incorporate such covariates into the analysis, it is convenient first to convert the hazard ratio summary statistics into dichotomous data that carry equivalent information (a transformation that can be performed by the `trans.armdata` function).

As in Section A.2, the `trans.armdata` function can be applied to hazard ratio data to reconstruct estimated cumulative numbers of events in each arm, using back-calculation based on the complementary log-log transformation [5,6]. These reconstructed event counts may not coincide exactly with the original observed counts. Nevertheless, the purpose of these reconstructed data is to serve as "working data" for preprocessing. When passed to the `setup` function, they are transformed back into summary statistics using the same formulas, thereby returning to the original hazard ratio estimates. In this sense, the reconstructed event counts can be regarded as pseudo-data—intermediate quantities created solely for convenience in facilitating subsequent analyses.

When the `trans.armdata` function is applied to the `woods1` dataset, the following arm-level pseudo-data on event counts are obtained.

```
> woods3 <- trans.armdata(study=studlab,treat1=treat1,treat2=treat2,
n1=n1,n2=n2, y=TE,SE=seTE,measure="logHR",data=woods1)
study         trt         d    n
1     1 Fluticasone  43.15258  376
8     1     Placebo  55.65442  375
4     2 Fluticasone 247.88996 1534
9     2     Placebo 231.60891 1524
3     2  Salmeterol 204.67537 1521
2     2         SFC 193.52488 1533
```

The hazard ratio data are replaced by reconstructed estimates of arm-level event counts. As noted above, these reconstructed values may not exactly match the actual cumulative number of events. However, when processed with the `setup` function, they are transformed back into summary statistics using the same formulas, thereby recovering the original hazard ratio estimates; thus, this does not pose a problem.

By combining the reconstructed arm-level data with the remaining dichotomized

datasets, it is then possible to generate a dataset suitable for analysis using the NMA package.

```
> woods4 <- rbind(woods2,woods3)
> print(woods4)
   study         trt         d    n
1      3  Salmeterol   1.00000  229
2      3     Placebo   1.00000  227
3      4 Fluticasone   4.00000  374
4      4  Salmeterol   3.00000  372
5      4         SFC   2.00000  358
6      4     Placebo   7.00000  361
7      5  Salmeterol   1.00000  554
8      5     Placebo   2.00000  270
11     1 Fluticasone  43.15258  376
81     1     Placebo  55.65442  375
41     2 Fluticasone 247.88996 1534
9      2     Placebo 231.60891 1524
31     2  Salmeterol 204.67537 1521
21     2         SFC 193.52488 1533
```

In this way, a single combined dataset can be constructed in the format of event counts and sample sizes. By supplying data in this format to the `setup` function of the NMA package, hazard ratio estimates are calculated under the proportional hazards assumption [5,6], and a data object is automatically generated for analysis using multivariate meta-analysis models.

If summary data on arm-level covariates are available and intended for use in multivariable meta-regression or transitivity analysis, these variables can simply be added to the dataset object. For practical data handling, the dataset may first be exported as a CSV file and then edited in user-friendly software such as Microsoft Excel before being re-imported for analysis.

B.3 Data preparation and setup

Using the methods described in the previous sections, once dichotomized arm-level data have been obtained for all eligible trials, the `setup` function can be used to generate a dataset for network meta-analysis based on the contrast-based approach. The outcome measure in this context is the hazard ratio.

When original hazard ratio estimates are available from published reports, the data prior to transformation by the `trans.armdata` function are, in principle, restored during

processing by the `setup` function.

```
> wd4 <- setup(study=study,trt=trt,d=d,n=n,measure="HR",ref="Placebo",
data=woods4)
```

B.4 Network meta-analysis

Once the hazard ratio dataset object has been created using the `setup` function, all subsequent analyses can be executed with simple commands, following the same procedures as described in the main text. The illustrative program can be fully reproduced by running R script (C), which is provided at the GitHub page (https://github.com/nomahi/NMA).

As an example, the results of the integrated analysis using the `nma` function are shown below.

```
> nma(wd4, eform=TRUE)
$coding
  code   treatment
1    1     Placebo
2    2 Fluticasone
3    3  Salmeterol
4    4         SFC

$reference
[1] "Placebo"

$`number of studies`
[1] 5

$method
[1] "Noma-Hamura's improved REML-based inference and prediction methods"

$`Coef. (vs. treat 1)`
            Coef.        SE     95%CL    95%CU   P-value
2: cons 0.9005129 0.1973861 0.4804839 1.687722 0.6322839
3: cons 0.7605587 0.2379688 0.3566413 1.621937 0.3334454
4: cons 0.7139671 0.2273424 0.3463091 1.471948 0.2349524

$`tau (Between-studies_SD) estimate`
[1] 0.1880613
```

```
$`tau2 (Between-studies_variance) estimate`
[1] 0.03536705

$`Multivariate H2-statistic`
[1] 0.9550896

$`Multivariate I2-statistic`
[1] 0.8319671

$`Test for Heterogeneity`
[1] "Q(df = 6) = 5.7305, p-value = 0.454"

$`95%PI (vs. treat 1)`
               pl       pu
2: cons 0.2786405 2.910285
3: cons 0.2062360 2.804794
4: cons 0.2006095 2.541001
```

B.5 Handling the survival probability difference

In systematic reviews of survival outcomes, meta-analyses are sometimes conducted using the difference in survival probabilities as the effect measure [7,8], e.g., the difference in 5-year survival rates. Arm-specific event probabilities can be obtained from Kaplan–Meier curves (and their standard errors can be derived, for example, from reported confidence intervals). In network meta-analysis as well, it is possible to use survival probability differences as the effect measure.

The `NMA` package implements a data-processing function, `trans.armdataP`, which converts estimated arm-specific survival probabilities and their standard errors into equivalent information in the form of event counts and sample sizes. These estimates can then be used as working pseudo-data and processed through `setup` in the same manner as hazard ratios, thereby producing contrast-based statistics. Consequently, survival probability differences can be incorporated into all standard `NMA` analyses, alongside other effect measures.

An illustrative example of such a synthesis, using the dataset `exdataP`, is presented below. The illustrative program can be fully reproduced by running R script (D), which is provided at the GitHub page (https://github.com/nomahi/NMA).

```
> data(exdataP)
> print(exdataP)
```

```
   study        trt          y         se
9      1 Fluticasone 0.114361702 0.016412494
10     1    Placebo 0.149333333 0.018405281
11     2 Fluticasone 0.161668840 0.009399576
12     2    Placebo 0.152230971 0.009202330
13     2  Salmeterol 0.134779750 0.008756107
14     2        SFC 0.126549250 0.008491373
1      3  Salmeterol 0.004366812 0.004357267
2      3    Placebo 0.004405286 0.004395572
3      4 Fluticasone 0.010695187 0.005318920
4      4  Salmeterol 0.008064516 0.004637238
5      4        SFC 0.005586592 0.003939267
6      4    Placebo 0.019390582 0.007257547
7      5  Salmeterol 0.001805054 0.001803424
8      5    Placebo 0.007407407 0.005218393

> woods4 <- trans.armdataP(study=study,treat=trt,y=y,SE=se,
data=exdataP)
> print(woods4)             # "working" pseudo-data
   study        trt          d        n
1      1 Fluticasone  43.0000000  376.0000
2      1    Placebo  56.0000545  375.0004
3      2 Fluticasone 247.9723574 1533.8290
4      2    Placebo 231.9959209 1523.9731
5      2  Salmeterol 204.9984747 1520.9887
6      2        SFC 193.9993857 1532.9952
7      3  Salmeterol   1.0000000  229.0000
8      3    Placebo   1.0000000  227.0000
9      4 Fluticasone   3.9999937  373.9991
10     4  Salmeterol   3.0000001  372.0000
11     4        SFC   2.0000001  358.0000
12     4    Placebo   7.0000000  361.0000
13     5  Salmeterol   0.9999999  553.9999
14     5    Placebo   2.0000000  270.0000

> wd4 <- setup(study=study,trt=trt,d=d,n=n,measure="SPD",ref="Placebo",
data=woods4)
> nma(wd4)           # Survival probability difference
$coding
  code   treatment
```

```
1    1     Placebo
2    2 Fluticasone
3    3  Salmeterol
4    4        SFC

$reference
[1] "Placebo"

$`number of studies`
[1] 5

$method
[1] "Noma-Hamura's improved REML-based inference and prediction methods"

$`Coef. (vs. treat 1)`
              Coef.         SE      95%CL       95%CU   P-value
2: cons -0.002170323 0.007910307 -0.02098949 0.016648842 0.7919548
3: cons -0.007289116 0.005274460 -0.02320513 0.008626894 0.2527461
4: cons -0.014633489 0.007709072 -0.03520815 0.005941171 0.1232470

$`tau (Between-studies_SD) estimate`
[1] 0.006674468

$`tau2 (Between-studies_variance) estimate`
[1] 4.454852e-05

$`Multivariate H2-statistic`
[1] 1.425507

$`Multivariate I2-statistic`
[1] 0.5033585

$`Test for Heterogeneity`
[1] "Q(df = 6) = 8.553, p-value = 0.2003"

$`95%PI (vs. treat 1)`
                pl        pu
2: cons -0.02771307 0.02337243
3: cons -0.03949850 0.02492027
4: cons -0.04481170 0.01554472
```

References

[1] Phung OJ, Scholle JM, Talwar M, Coleman CI. Effect of noninsulin antidiabetic drugs added to metformin therapy on glycemic control, weight gain, and hypoglycemia in type 2 diabetes. *JAMA*. 2010;303(14):1410-8.

[2] Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Lawrence Erlbaum Associates; 1988.

[3] Woods BS, Hawkins N, Scott DA. Network meta-analysis on the log-hazard scale, combining count and hazard ratio statistics accounting for multi-arm trials: a tutorial. *BMC Med Res Methodol*. 2010;10:54.

[4] Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med*. 1998;17(24):2815-2834.

[5] Hedeker D, Siddiqui O, Hu FB. Random-effects regression analysis of correlated grouped-time survival data. *Stat Methods Med Res*. 2000;9(2):161-79.

[6] Salika T, Turner RM, Fisher D, Tierney JF, White IR. Implications of analysing time-to-event outcomes as binary in meta-analysis: empirical evidence from the Cochrane Database of Systematic Reviews. *BMC Med Res Methodol*. 2022;22(1):73.

[7] Maas C, Kent DM, Dinmohamed AG, van Klaveren D. Quantifying absolute treatment effect heterogeneity for time-to-event outcomes across different risk strata: divergence of conclusions with risk difference and restricted mean survival difference. *medRxiv*. 2024; DOI: https://doi.org/10.1101/2024.12.19.24319347.

[8] Higgins JPT, Thomas J. *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd ed. Wiley-Blackwell; 2019.