NMA Package: Advanced Network Meta-Analysis Made Easy with R

Hisashi Noma
The Institute of Statistical Mathematics, Tokyo, Japan

1

The Importance of Comparative Effectiveness

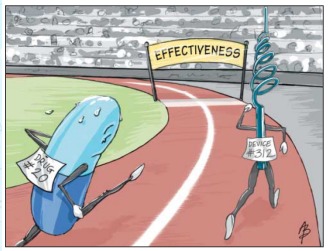
- In almost all areas of medicine today, multiple treatment options exist. A central question in both clinical epidemiology and health technology assessment is:
  - *"Which treatment provides the best benefit for each patient?"*
- This is not only a methodological issue but also a matter of public health and healthcare policy.
- With rapid population aging and the increasing burden of healthcare costs, answering this question has become more important than ever.

Del Fiol et al. (2014), Caldwell et al. (2015), Egger et al. (2022)

2

2

Comparative Effectiveness Research (CER)



Comparative effectiveness research allows investigators to determine the superiority, inferiority, or equivalence of various interventions when pitted against each other.

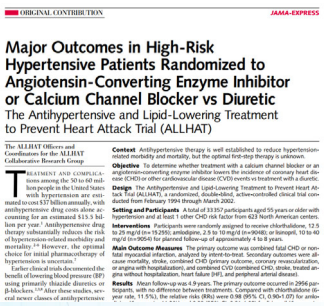
Mitka, M. (2010). US Government Kicks Off Program for Comparative Effectiveness Researches. JAMA 304: 2230-1.

- Institute of Medicine (IOM) defines CER as: *"The generation and synthesis of evidence comparing the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor clinical conditions or improve healthcare delivery. The purpose is to support informed decisions by patients, clinicians, and policymakers to make informed decisions that will improve health care at both the individual and population levels."*

3

3

The ALLHAT Trial (1990s, USA)



ORIGINAL CONTRIBUTION — JAMA-EXPRESS

Major Outcomes in High-Risk Hypertensive Patients Randomized to Angiotensin-Converting Enzyme Inhibitor or Calcium Channel Blocker vs Diuretic
The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT)

The ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group

Context Antihypertensive therapy is well established to reduce hypertension-related morbidity and mortality, but the optimal first-step therapy is unknown.

Objective To determine whether treatment with a calcium channel blocker or an angiotensin-converting enzyme inhibitor lowers the incidence of coronary heart disease (CHD) or other cardiovascular disease (CVD) events vs treatment with a diuretic.

Design The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT), a randomized, double-blind, active-controlled clinical trial conducted from February 1994 through March 2002.

Setting and Participants A total of 33 357 participants aged 55 years or older with hypertension and at least 1 other CHD risk factor from 623 North American centers.

Interventions Participants were randomly assigned to receive chlorthalidone, 12.5 to 25 mg/d (n= 15 255); amlodipine, 2.5 to 10 mg/d (n=9048); or lisinopril, 10 to 40 mg/d (n=9054) for planned follow-up of approximately 4 to 8 years.

Main Outcome Measures The primary outcome was combined fatal CHD or non-fatal myocardial infarction, analyzed by intent-to-treat. Secondary outcomes were all-cause mortality, stroke, combined CHD (primary outcome, coronary revascularization, or angina with hospitalization), and combined CVD (combined CHD, stroke, treated angina without hospitalization, heart failure (HF), and peripheral arterial disease).

Results Mean follow-up was 4.9 years. The primary outcome occurred in 2956 participants, with no difference between treatments. Compared with chlorthalidone (6-year rate, 11.5%), the relative risks (RRs) were 0.98 (95% CI, 0.90-1.07) for amlo-

https://doi.org/10.1001/jama.288.23.2981

- Large-scale RCT comparing four major antihypertensive drug classes: Diuretics, β-blockers, ACE inhibitors, Calcium channel blockers
- Planned sample size: 40,000 participants—an unprecedented scale at the time.
- Shows that head-to-head RCTs designed to provide definitive evidence on comparative effectiveness require enormous resources, time, and effort.
- Such studies are rare in practice, highlighting the need for alternative evidence synthesis methods.

4

4

Network Meta-Analysis (NMA)

- NMA is a statistical method that allows for the simultaneous comparison of multiple treatments.
- Unlike traditional pairwise meta-analysis, it can integrate evidence from 3 or more treatments at once.
- It combines both direct comparisons (head-to-head trials) and indirect comparisons (via common comparators).
- Earlier literature referred to NMA as "*multiple treatment comparison meta-analysis*" or "*mixed treatment comparison*."

Caldwell et al. (2005), Salanti (2012)

5

5

Applications of NMA

https://iris.who.int/handle/10665/271991

- NMA has become increasingly influential in evidence-based medicine.
- In 2016, the World Health Organization incorporated NMA evidence into its guidelines on HIV treatment.
- Since then, NMA has been routinely used in WHO guideline development.
- The method has also been widely adopted by Cochrane and NICE (UK), reflecting its suitability for clinical guideline development and health policy.

6

6

NMA in Clinical Guidelines

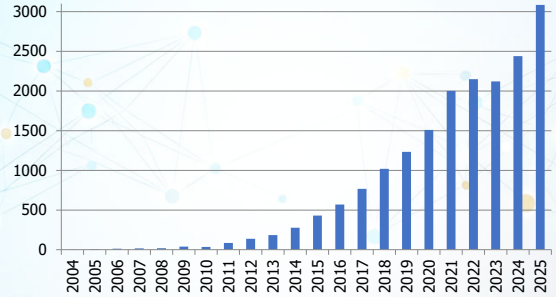
https://doi.org/10.1371/journal.pone.0315563

- Review of Cochrane NMAs (*PLOS One*, 2024): *Among 60 NMAs, 43% were cited in 89 different guidelines.*
- Each NMA was cited in 1–13 guidelines.
- Compared with traditional pairwise meta-analyses: *NMAs were 1.5 times more frequently cited.*
- NMAs influenced recommendations 4.4 times more often.
- NMAs were almost twice as likely to be included in NICE guidelines.
- These findings emphasize the practical importance of NMA for shaping medical recommendations.

7

7

Publication Trends

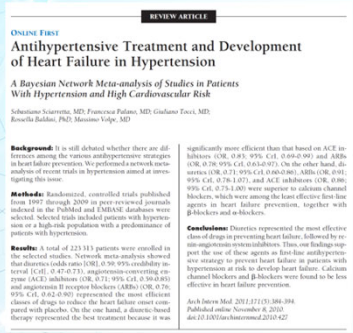


Pubmed search: ("network meta-analysis" OR "multiple treatment comparison meta-analysis" OR "mixed comparison meta-analysis")

8

8

Case Study: Antihypertensive Drugs



REVIEW ARTICLE

ONLINE FIRST

Antihypertensive Treatment and Development of Heart Failure in Hypertension

A Bayesian Network Meta-analysis of Studies in Patients With Hypertension and High Cardiovascular Risk

Sebastiano Sciarretta, MD; Francesca Palano, MD; Giuliano Tocci, MD; Rossella Baldini, PhD; Massimo Volpe, MD

Background: It is still debated whether there are differences among the various antihypertensive strategies in heart failure prevention. We performed a network meta-analysis of recent trials in hypertension aimed at investigating this issue.

Methods: Randomized, controlled trials published from 1997 through 2009 in peer-reviewed journals indexed in the PubMed and EMBASE databases were selected. Selected trials included patients with hypertension or a high-risk population with a predominance of patients with hypertension.

Results: A total of 223 313 patients were enrolled in the selected studies. Network meta-analysis showed that diuretics (odds ratio [OR], 0.59; 95% credibility interval [CrI], 0.47-0.73), angiotensin-converting enzyme (ACE) inhibitors (OR, 0.71; 95% CrI, 0.59-0.85) and angiotensin II receptor blockers (ARBs) (OR, 0.76; 95% CrI, 0.62-0.90) represented the most efficient classes of drugs to reduce the heart failure onset compared with placebo. On the one hand, a diuretic-based therapy represented the best treatment because it was significantly more efficient than that based on ACE inhibitors (OR, 0.83; 95% CrI, 0.69-0.99) and ARBs (OR, 0.78; 95% CrI, 0.63-0.97). On the other hand, diuretics (OR, 0.71; 95% CrI, 0.60-0.86), ARBs (OR, 0.91; 95% CrI, 0.78-1.07), and ACE inhibitors (OR, 0.86; 95% CrI, 0.73-1.00) were superior to calcium channel blockers, which were among the least effective first-line agents in heart failure prevention, together with β-blockers and α-blockers.

Conclusions: Diuretics represented the most effective class of drugs in preventing heart failure, followed by renin-angiotensin system inhibitors. Thus, our findings support the use of these agents as first-line antihypertensive strategy to prevent heart failure in patients with hypertension at risk to develop heart failure. Calcium channel blockers and β-blockers were found to be less effective in heart failure prevention.

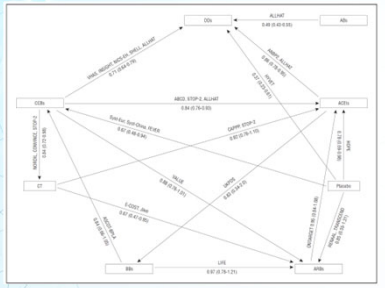
Arch Intern Med. 2011;171(5):384-394.
Published online November 8, 2010.
doi:10.1001/archinternmed.2010.427

- Example study: Sciarretta et al. (2011).
- Conducted an NMA of antihypertensive drug classes: α-blocker (AB), angiotensin-converting enzyme inhibitor (ACE), angiotensin II receptor blocker (ARB), β-blocker (BB), calcium channel blocker (CCB), conventional treatment (CT), diuretic (DD)] and placebo
- Outcome: incidence of heart failure.

Sciarretta et al. (2011)

9

Network of the evidence



Sciarretta et al. (2011)

- A network meta-analysis of antihypertensive drugs based on 26 randomized controlled trials (n = 223,313)
- No single randomized controlled trial compared all drug classes directly.
- Network meta-analysis enabled systematic evaluation by integrating both direct and indirect comparisons.
- This illustrates how NMA can address critical evidence gaps in clinical decision-making.

10

Results from the NMA

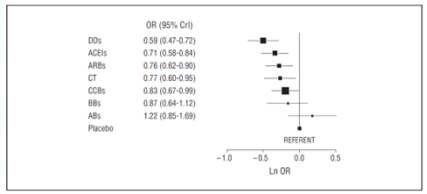


Figure 3. Results of network meta-analysis with placebo considered as a referent treatment. ABs indicates α-blockers; ACEIs, angiotensin converting enzyme inhibitors; ARBs, angiotensin II receptor blockers; BBs, β-blockers; CCBs, calcium channel blockers; CrI, credibility interval; CT, conventional treatment; DDs, diuretics; and OR, odds ratio.

- Comparative treatment effects can be displayed using forest plots or network plots, which make the findings more intuitive.
- Such results allow clinicians, researchers, and policymakers to evaluate the relative effectiveness of multiple interventions simultaneously, something not possible with conventional pairwise meta-analysis.

Sciarretta et al. (2011)

11

NMA: Network Meta-Analysis Based on Multivariate Meta-Analysis and Meta-Regression Models

Network meta-analysis tools based on contrast-based approach using the multivariate meta-analysis and meta-regression models (Noma et al. (2025) <doi:10.1101/2025.09.15.25335823>). Comprehensive analysis tools for network meta-analysis and meta-regression (e.g., synthesis analysis, ranking analysis, and creating league table) are available through simple commands. For inconsistency assessment, the local and global inconsistency tests based on the Higgins' design-by-treatment interaction model are available. In addition, the side-splitting methods and Jackson's random inconsistency model can be applied. Standard graphical tools for network meta-analysis, including network plots, ranked forest plots, and transitivity analyses, are also provided. For the synthesis analyses, the Noma-Hamura's improved REML (restricted maximum likelihood)-based methods (Noma et al. (2023) <doi:10.1002/jrsm.1652> <doi:10.1002/jrsm.1651>) are adopted as the default methods.

Version:	3.1-1
Depends:	R (≥ 3.5.0)
Imports:	stats, grid, MASS, ggplot2, metafor, stringr, forestplot, nleqslv
Published:	2026-01-28
DOI:	10.32614/CRAN.package.NMA
Author:	Hisashi Noma [aut, cre], Kazushi Maruo [aut], Shiro Tanaka [aut], Toshi A. Furukawa [aut]
Maintainer:	Hisashi Noma <noma at ism.ac.jp>
License:	GPL-3
URL:	https://github.com/nomahi/NMA
NeedsCompilation:	no
Materials:	NEWS
In views:	MetaAnalysis
CRAN checks:	NMA results

Documentation:

Reference manual: NMA.html , NMA.pdf

Downloads:

Package source:	NMA_3.1-1.tar.gz
Windows binaries:	r-release: NMA_3.1-1.zip, r-oldrel: NMA_3.1-1.zip
macOS binaries:	r-release (arm64): NMA_3.1-1.tgz, r-oldrel (arm64): NMA_3.1-1.tgz, r-release (x86_64): NMA_3.1-1.tgz, r-oldrel (x86_64): NMA_3.1-1.tgz
Old sources:	NMA archive

https://cran.r-project.org/web/packages/NMA/

12

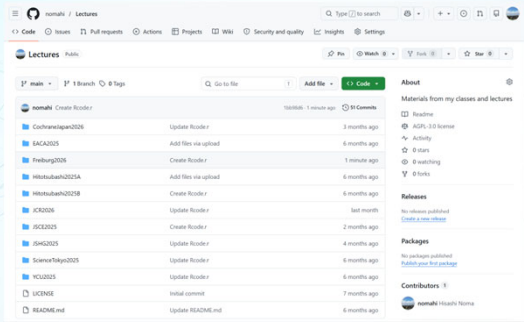
Today's Agenda

- Introduction to frequentist statistical methods for network meta-analysis.
- Demonstration using Sciarretta et al. (2011) antihypertensive trial data.
- Hands-on practice with the R package "NMA".
- Discussion of data preparation, modeling options, and diagnostics.
- Interpretation of results and implications for evidence-based medicine.

13

13

R Example Code



- Example R code is available for self-learning and replication. Repository: GitHub (https://github.com/nomahi/Lectures/tree/main/Freiburg2026).

14

14

Two Main Approaches in NMA

- Arm-based approach: models outcomes at the trial-arm level (events, means, standard deviations).
  - Primarily used in Bayesian frameworks, implemented in software such as R package "gemtc".
- Contrast-based approach: models effect measures such as odds ratios or hazard ratios.
  - Extends conventional pairwise meta-analysis models to networks with many treatments.
  - The contrast-based framework is most common in frequentist analyses.

White et al. (2019)

15

15

Contrast-Based Approach

- Treats study-specific effect estimators as correlated observations in a multivariate framework.
- Direct extension of DerSimonian–Laird type models from pairwise to multiple treatments.
- Compatible with both fixed-effect and random-effects assumptions.
- Provides tools for heterogeneity assessment such as Q test, I^2 statistic, and prediction intervals.
- Formally described in the literature as the main frequentist foundation for NMA.

White et al. (2012)

16

16

https://doi.org/10.1101/2025.09.15.25335823

17

17

R package “NMA”

- The “NMA” R package was originally developed for courses on Network Meta-Analysis at Kyoto University.
- It is a free, open-source package that provides a complete set of standard frequentist tools for NMA, including statistical analyses and graphical visualizations.
- All major functions are accessible through simple commands, lowering the barrier for applied researchers.
- The package also includes data-handling tools that allow users to integrate arm-level data (sample sizes and event counts) as well as summary statistics such as hazard ratios or mean differences, with minimal effort.

18

18

R language

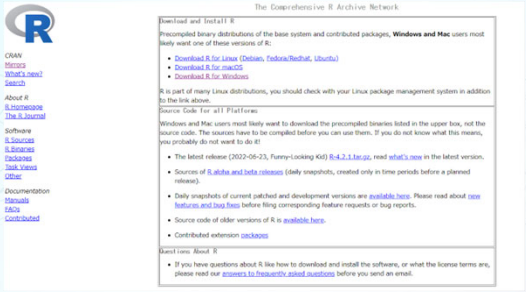
- R language is an open-source and freely available programming language and environment for statistical computing.
- It was originally derived from the S language developed at AT&T Bell Laboratories.
- Nearly all standard statistical methods used in clinical research have been implemented in R modules, making it a reliable software that is widely accepted and cited in medical publications.
- The vibrant open-source community ensures continuous development, testing, and validation of methods.

CRAN (The Comprehensive R Archive Network) https://cran.r-project.org/

19

19

CRAN (Comprehensive R Archive Network)



- All R packages, including those for advanced meta-analysis, are distributed through CRAN (https://cran.r-project.org/).

20

20

Installing and Loading the “NMA” Package

```
> pkgCheck <- function(pkg){
>       if (!requireNamespace(pkg, quietly = TRUE)) install.packages(pkg)
>       library(pkg, character.only = TRUE)
> }
>
> pkgCheck("NMA")
```

- This script first checks whether the NMA package is already installed. If not present, it downloads the package from CRAN and installs it automatically. If already installed, it simply loads the package into the current R session, ensuring efficiency and avoiding redundant installations.

21

21

Constructing a Dataset (1): Data Structure

- Network meta-analysis usually involves comparing three or more treatments.
- For multi-arm trials, each study contributes one row per treatment arm in the dataset.
- For example, a 3-arm trial will be represented by 3 rows.
- A unique study identifier (“`study`”) must be included for each trial.
- It is recommended to prepare datasets in Excel or similar tools and then import them into R for analysis.
- This structure ensures compatibility with functions in the NMA package.

22

22

Constructing a Dataset (2): Treatment Variable

- A dedicated variable is required to identify the treatment group in each trial.
- In the example dataset, this variable is named “`trt`”.
- Example coding:
  - `AB`: α-blocker, `ACE`: ACE inhibitor, `ARB`: Angiotensin receptor blocker, `BB`: β-blocker, `CCB`: Calcium channel blocker, `CT`: Conventional treatment, `DD`: Diuretic, `Placebo`: Placebo group.

23

23

Constructing a Dataset (3): Outcome Variable

- For binary outcomes, the dataset requires:
  - the number of events (e.g., cases of heart failure) per arm
  - the sample size per arm
- In the example dataset, these are coded as “`d`” (events) and “`n`” (sample size).
- Sometimes trial reports provide only odds ratios or risk ratios with confidence intervals.
- The NMA package includes functions that can integrate such summary statistics seamlessly.

24

24

Example Dataset: heartfailure

```
> data(heartfailure)
> print(heartfailure)
   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
7      4       ABCD      ACE   235   5    1998 155.0  98.0
8      4       ABCD      CCB   235   6    1998 156.0  98.0
9      5       VHAS      CCB   707   2    1997 169.1 102.0
10     5       VHAS       DD   707   0    1997 168.8 102.3
11     6      CAPPP      ACE  5492  75    1999 161.8  99.8
12     6      CAPPP       CT  5493  66    1999 159.6  98.1
13     7    NICS-EH      CCB   204   0    1999 171.9  94.2
14     7    NICS-EH       DD   210   3    1999 172.6  93.4
15     8     STOP-2      CCB  2196 186    1999 194.0  98.0
16     8     STOP-2       CT  2213 177    1999 194.0  98.0
17     8     STOP-2      ACE  2205 149    1999 194.0  98.0
```

- The dataset contains 26 clinical trials comparing antihypertensive treatments for prevention of heart failure. Each row corresponds to an arm-level observation.
- study: study identifier
- trt: treatment type
- n: sample size
- d: number of heart failure events
- SBP, DBP: baseline systolic/diastolic blood pressure
- pubyear: publication year

Outcome Variables by Data Type

- Required outcome variables depend on the type of data, similar to standard meta-analysis:
- Continuous outcomes: mean, standard deviation, and sample size.
- Binary outcomes: number of events and sample size.
- Time-to-event outcomes: hazard ratio (HR) and standard error, or survival rates at specific times.
- In the heart failure dataset (binary), event counts (`d`) and sample sizes (`n`) are provided.
- For survival data, functions are available to back-calculate group-level outcomes from published HRs or mean differences.

setup: Preprocessing Function

- Conducting a full NMA requires many complex calculations, but the NMA package automates preprocessing.
- The setup function takes arm-level data and creates a standardized R object that is then used by all downstream functions.
- This design ensures consistency across analyses and prevents redundant coding.
- By preparing only a clean dataset, researchers can perform the entire analysis pipeline with minimal commands.

setup: Basic Syntax

- Basic Syntax:
- `hf3 <- setup(study=study, trt=trt, d=d, n=n, z=c(SBP,DBP,pubyear), measure="RR", ref="Placebo", data=heartfailure)`
- `study`: study identifier, `trt`: treatment variable, `d`: number of events, `n`: sample size, `z`: optional covariates for meta-regression, `measure`: effect measure (`RR`, `RD`, `OR`), `ref`: reference treatment (here, Placebo), `data`: dataset object

Options for Effect Measures

- Depending on outcome type, the following measures can be used:
  - Binary outcomes: risk ratio (`RR`), risk difference (`RD`), odds ratio (`OR`)
  - Continuous outcomes: mean difference (`MD`), standardized mean difference (`SMD`)
  - Survival outcomes: hazard ratio (`HR`), differences in survival rates
- In this workshop we will focus on binary outcomes.
- For continuous and survival data, refer to Noma et al. (2025, medRxiv tutorial).

29

29

Network Plot

- A network plot provides a graphical overview of the treatment evidence network.
- Nodes represent treatments, with size proportional to cumulative sample size.
- Edges represent direct head-to-head comparisons, with thickness proportional to sample size.
- This visualization makes it easy to see where evidence is dense and where gaps remain.

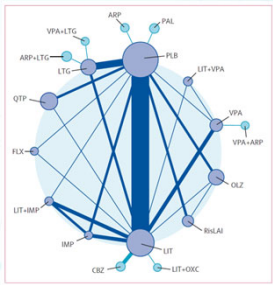


Figure 2: Network of all eligible comparisons for the network meta-analysis

Miura, Noma, Furukawa et al. (2014)

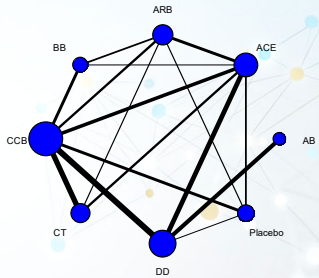
30

30

netplot: Creating Network Plots

```
> netplot(hf3)
```

- Generates a network diagram from the dataset prepared using setup().



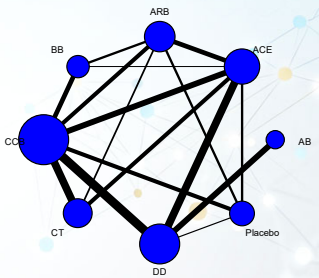
31

31

netplot: Customizing Node and Edge Size

```
> netplot(hf3, base.lwd = 1.5,
base.cex = 1.5)
```

- `base.lwd`: adjusts the thickness of edges (comparisons).
- `base.cex`: adjusts the size of nodes (treatments).



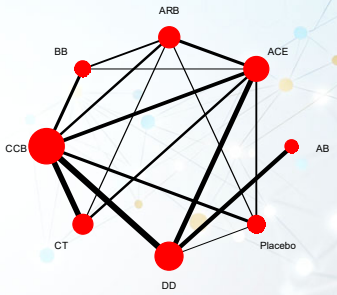
32

32

netplot: Customizing Colors

```
> netplot(hf3, col = "red", bg = "red")
```

- `col`: outline color of nodes.
- `bg`: fill color of nodes.



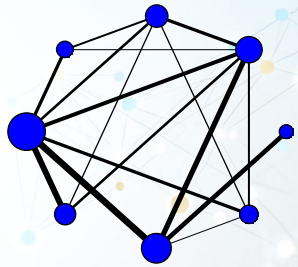
33

33

netplot: Removing Labels

```
> netplot(hf3, text = FALSE)
```

- Removes node labels from the plot.
- Useful when exporting figures for external editing (e.g., adding labels manually in PowerPoint).
- Reduces visual clutter when the number of treatments is large.



34

34

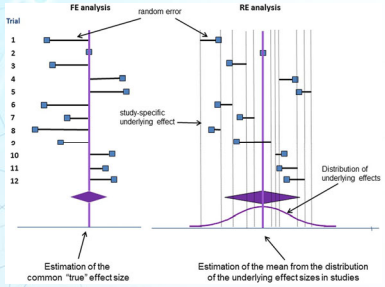
Pairwise Meta-Analysis of Direct Comparisons

- NMA assumes consistency between direct and indirect evidence.
- To check assumptions, it is necessary to run pairwise meta-analyses first.
- Each available comparison is analyzed individually to reveal potential heterogeneity.
- This provides the foundation for the subsequent network analysis.

35

35

Fixed-Effect vs Random-Effects Models



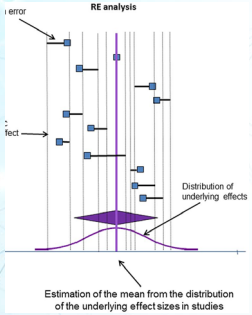
Nikolakopoulou, Mavridis and Salanti (2014)

- Fixed-effect models assume one true effect size across all studies.
- Random-effects models assume that study-specific effects follow a normal distribution around an average effect.
- Random-effects are more realistic in clinical settings with heterogeneous study designs.

36

36

Random-Effects Models



- In practice, treatment effects are rarely identical across trials.
- The random-effects model explicitly accounts for this between-study variability (τ^2).
- Provides wider confidence intervals, reflecting greater uncertainty.
- Better suited for decision-making when evidence is diverse.

Nikolakopoulou, Mavridis and Salanti (2014)

37

37

Cochran's Q Test

- A classical method to assess heterogeneity in meta-analysis.
- Null hypothesis: all studies estimate the same true effect.
- A large Q statistic ($p < 0.05$) indicates significant heterogeneity.
- Widely used but known to have low power with small numbers of studies.

Cochran (1954)

38

38

Higgins' I^2 Statistic

- Clinical and methodological heterogeneity is essentially inevitable, even before formal statistical testing.
- Under a random-effects model, the variability of estimates can be decomposed into between-study heterogeneity, and within-study sampling error.
- The proportion attributable to between-study heterogeneity is quantified by the I^2 statistic, a standard measure of heterogeneity.

Higgins and Thompson (2002)

39

39

Guidance from the Cochrane Handbook

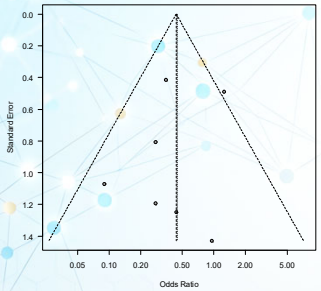
- The Cochrane Handbook cautions that strict thresholds for I^2 can be misleading, since the importance of inconsistency depends on multiple factors. Suggested rough guide for interpretation:
  - 0%–40%: might not be important
  - 30%–60%: may represent moderate heterogeneity
  - 50%–90%: may represent substantial heterogeneity
  - 75%–100%: considerable heterogeneity
- Note: Importance depends on: magnitude and direction of effects, and strength of evidence for heterogeneity (e.g., *p*-value from Q test, or CI for I^2).

Higgins and Thomas (2019)

40

40

Funnel Plot



- Funnel plots assess potential small-study effects and publication bias.
- Plotting study effect sizes against precision (e.g., standard error) should form a symmetrical “funnel.”
- Asymmetry suggests bias or heterogeneity.
- Provides a visual but subjective assessment.

Egger et al. (1997)

41

41

SumPMA: Meta-Analysis of All Direct Comparisons

```
> pairwise(hf3)
Call:
SumPMA(x = hf3, method = "REML",
test = "z")

Coding:
 code treatment
    1   Placebo
    2        AB
    3       ACE
    4       ARB
    5        BB
    6       CCB
    7        CT
    8        DD

Measure: RR
```

- The function pairwise performs meta-analyses for every direct comparison.
- Allows side-by-side inspection of all treatment pairs.
- Summarizes pooled effects and heterogeneity statistics.
- Forms the empirical basis for evaluating consistency in the network.

42

42

SumPMA: Example Output

```
Summary effect measures:
       N Estimate Lower Upper Pr(>|z|)
1 vs. 3 1    0.781 0.691 0.881    0.000
1 vs. 4 2    0.864 0.593 1.260    0.448
1 vs. 6 3    0.682 0.490 0.950    0.024
1 vs. 8 1    0.382 0.234 0.622    0.000
2 vs. 8 1    0.508 0.447 0.577    0.000
3 vs. 4 1    1.049 0.933 1.179    0.424
3 vs. 5 1    0.838 0.357 1.965    0.684
3 vs. 6 3    1.174 1.070 1.288    0.001
3 vs. 7 1    0.880 0.633 1.222    0.445
3 vs. 8 1    1.132 0.822 1.559    0.447
4 vs. 5 1    1.056 0.850 1.313    0.622
4 vs. 6 1    1.138 0.990 1.308    0.070
4 vs. 7 2    1.483 0.983 2.238    0.060
5 vs. 6 1    0.841 0.669 1.056    0.137
6 vs. 7 3    0.874 0.756 1.009    0.067
6 vs. 8 5    0.727 0.662 0.799    0.000
```

- Output includes forest plots for each comparison.
- Estimates are pooled across studies with confidence intervals.
- Heterogeneity statistics (Q, I^2, τ^2) are also reported.
- Enables quick identification of comparisons with unstable results.

43

43

SumPMA: Example Output (cont’d)

```
Heterogeneity measures:
       N   tau tau^2   I^2   H^2
1 vs. 3 1 0.000 0.000 0.000 1.000
1 vs. 4 2 0.242 0.059 0.791 4.783
1 vs. 6 3 0.000 0.000 0.000 1.000
1 vs. 8 1 0.000 0.000 0.000 1.000
2 vs. 8 1 0.000 0.000 0.000 1.000
3 vs. 4 1 0.000 0.000 0.000 1.000
3 vs. 5 1 0.000 0.000 0.000 1.000
3 vs. 6 3 0.000 0.000 0.000 1.000
3 vs. 7 1 0.000 0.000 0.000 1.000
3 vs. 8 1 0.000 0.000 0.000 1.000
4 vs. 5 1 0.000 0.000 0.000 1.000
4 vs. 6 1 0.000 0.000 0.000 1.000
4 vs. 7 2 0.159 0.025 0.279 1.387
5 vs. 6 1 0.000 0.000 0.000 1.000
6 vs. 7 3 0.000 0.000 0.000 1.000
6 vs. 8 5 0.003 0.000 0.000 1.000
```

- Results are automatically formatted in a consistent style.
- Visual clarity ensures results can be directly included in reports or presentations.
- Facilitates reproducibility and transparency.

44

44

SumPMA: Example Output (cont'd)

```
Egger test:
        N Pr(>|z|)
1 vs. 3 1       NA
1 vs. 4 2       NA
1 vs. 6 3    0.065
1 vs. 8 1       NA
2 vs. 8 1       NA
3 vs. 4 1       NA
3 vs. 5 1       NA
3 vs. 6 3    0.672
3 vs. 7 1       NA
3 vs. 8 1       NA
4 vs. 5 1       NA
4 vs. 6 1       NA
4 vs. 7 2       NA
5 vs. 6 1       NA
6 vs. 7 3    0.526
6 vs. 8 5    0.952
```

- Each comparison can be examined individually.
- Differences in study size, outcome definition, and follow-up duration may explain variability.
- Detailed examination supports more credible network integration.

45

45

Settings for Multivariate Meta-Analysis

- We consider synthesizing N trials and comparing $p+1$ treatments. Let Y_{ij} denote an estimator of treatment effect in contrast to a reference treatment (e.g., placebo) for the jth treatment in the ith trial ($i = 1, 2, \ldots, N; j = 1,2, \ldots, p$).
- Commonly used effect measures include mean difference, standardized mean difference, risk difference, risk ratio, odds ratio, and hazard ratio; the ratio measures are usually transformed on a logarithmic scale.

White et al. (2012), Noma et al. (2023ab)

46

46

Settings for Multivariate Meta-Analysis (cont'd)

- For appropriate modelling of within- and between-studies correlations, we consider multivariate outcome variable $\boldsymbol{Y}_i = \left(Y_{i1}, Y_{i2}, \ldots, Y_{ip}\right)^T$. Also, let $\boldsymbol{S}_i$ (a $p \times p$ matrix) be the within-study covariance matrix, which is assumed to be known and fixed to its valid estimate.
- Note that for trials that do not include a reference treatment, the data augmentation approach of White et al. (2012) can be adopted; a quasi-small dataset is added into the reference arm, e.g., 0.001 events for 0.01 patients.

White et al. (2012), Noma et al. (2023ab)

47

47

Multivariate Pooling Model

- The multivariate extended DerSimonian-Laird model:
  - $\boldsymbol{Y}_i \sim \mathrm{MVN}(\boldsymbol{\theta}_i, \boldsymbol{S}_i)$
  - $\boldsymbol{\theta}_i \sim \mathrm{MVN}(\boldsymbol{\mu}, \boldsymbol{\Psi}(\tau^2))$
- where $\boldsymbol{\theta}_i = \left(\theta_{i1}, \theta_{i2}, \ldots, \theta_{ip}\right)^T$ and $\boldsymbol{\mu} = \left(\mu_1, \ldots, \mu_p\right)^T$. $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N$ are the random effects and $\boldsymbol{\mu}$ is the grand mean that expresses an average treatment effect for p comparisons. $\boldsymbol{S}_i$ (a $p \times p$ matrix) is the within-study covariance matrix.

Jackson et al. (2011) , Noma et al. (2023ab)

48

48

Within-Study Covariance Matrix

- Within-study covariance matrix:
  - $\boldsymbol{S}_i = \begin{pmatrix} s_{i1}^2 & \rho_{i12}s_{i1}s_{i2} & \cdots & \rho_{i1p}s_{i1}s_{ip} \\ \rho_{i21}s_{i2}s_{i1} & s_{i2}^2 & \cdots & \rho_{i2p}s_{i2}s_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{ip1}s_{ip}s_{i1} & \rho_{ip2}s_{ip}s_{i2} & \cdots & s_{ip}^2 \end{pmatrix}$
- which is usually assumed to be known and fixed to its valid estimate. For multi-arm trials (≥ 3 arms), the correlations across multiple arms are modelled in the within-study covariance matrix. For the correlation coefficients $\rho_{i12}, \ldots, \rho_{ip(p-1)}$, several suitable estimators are provided in Noma (2024b).

Between-Studies Covariance Matrix

- Between-studies covariance matrix:
  - $\boldsymbol{\Psi}(\tau^2) = \tau^2\boldsymbol{P} = \tau^2 \begin{pmatrix} 1 & 1/2 & \cdots & 1/2 \\ 1/2 & 1 & \ldots & 1/2 \\ \vdots & \vdots & \ddots & \vdots \\ 1/2 & 1/2 & \cdots & 1 \end{pmatrix}$
- where $\tau^2 (> 0)$ is the heterogeneity variance across the N studies.

Note that the correlation structure of $\boldsymbol{\Psi}$ can be assumed to be unstructured; however, in the general practice of network meta-analysis, the number of studies involved is rarely sufficient to estimate all of the variance and covariance parameters. Thus, the equal-variance assumption for the p components (compound symmetry assumption) is standardly adopted. Under this assumption, all the pairwise correlation coefficients are equal to 0.50 because of the consistency assumption for the network (Higgins and Whitehead, 1996).

Analysis of Heterogeneity

- Generalization of classic metrics to the multivariate setting (Jackson et al., 2012) allows direct use of familiar measures.
- Cochran's Q test and Higgins' I^2 statistic are available for network contrasts.
- Prediction intervals: in a random-effects model, the pooled mean is only a mean; a 95% prediction interval shows where the true effect in a future population is likely to lie, providing an interpretable summary of heterogeneity.

Higgins et al. (2009), Nagashima et al. (2019)

Research methods
Research

Plea for routinely presenting prediction intervals in meta-analysis

Joanna IntHout [1], John P A Ioannidis [2, 3, 4, 5], Maroeska M Rovers [1], Jelle J Goeman [1]
Correspondence to Dr Joanna IntHout; joanna.inthout@radboudumc.nl

Abstract

Objectives Evaluating the variation in the strength of the effect across studies is a key feature of meta-analyses. This variability is reflected by measures like τ^2 or I^2, but their clinical interpretation is not straightforward. A prediction interval is less complicated: it presents the expected range of true effects in similar studies. We aimed to show the advantages of having the prediction interval routinely reported in meta-analyses.

Design We show how the prediction interval can help understand the uncertainty about whether an intervention works or not. To evaluate the implications of using this interval to interpret the results, we selected the first meta-analysis per intervention review of the Cochrane Database of Systematic Reviews Issues 2009–2013 with a dichotomous (n=2009) or continuous (n=1254) outcome, and generated 95% prediction intervals for them.

Results In 72.4% of 479 statistically significant (random-effects p<0.05) meta-analyses in the Cochrane Database 2009–2013 with heterogeneity I^2>0, the 95% prediction interval suggested that the intervention effect could be null or even be in the opposite direction. In 20.3% of those 479 meta-analyses, the prediction interval showed that the effect could be completely opposite to the point estimate of the meta-analysis. We demonstrate also how the prediction interval can be used to calculate the probability that a new trial will show a negative effect and to improve the calculations of the power of a new trial.

Conclusions The prediction interval reflects the variation in treatment effects over different settings, including what effect is to be expected in future patients, such as the patients that a clinician is interested to treat. Prediction intervals should be routinely reported to allow more informative inferences in meta-analyses.

https://doi.org/10.1136/bmjopen-2015-010247

Google Scholar Citation: 1815 times!! https://bmjopen.bmj.com/content/6/7/e010247

nma: Conducting a Network Meta-Analysis

- Primary function for network synthesis: Performs the integrated analysis for a treatment network using a multivariate meta-analysis framework in the frequentist setting.
- Modeling approach: Fits a random-effects NMA by default to account for between-study heterogeneity (a fixed-effect option is also available), and reports standard heterogeneity indices.
- Effect measures relative to a reference: Estimates effect measures (e.g., risk ratio, risk difference, etc.) for all treatments vs the prespecified reference treatment (set in setup()).

nma: Example Output

```
> nma(hf3, eform=TRUE)
Call:
nma(x = hf3, eform = TRUE)

Coding:
 code treatment
    1   Placebo
    2        AB
    3       ACE
    4       ARB
    5        BB
    6       CCB
    7        CT
    8        DD

Reference: Placebo

Number of studies: 26

Method: Noma-Hamura's improved REML-based inference and prediction methods
```

- By simply inputting the object created by setup, the nma function performs the entire network meta-analysis as well as computes heterogeneity indices in one step.
- The Cochrane Handbook recommends HKSJ-type adjustments because DerSimonian-Laird and REML underestimate error. The same issue exists in NMA, and the Noma-Hamura method (Noma et al., 2023a,b) is implemented as the default remedy.

nma: Example Output (cont'd)

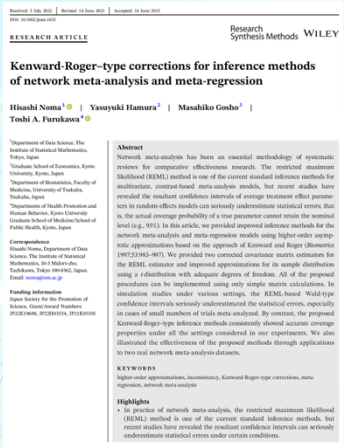
```
Coef. (vs. treat 1):
        Est.    SE Lower Upper Pr(>|z|)
2: cons 1.230 0.149 0.846 1.788    0.220
3: cons 0.738 0.078 0.614 0.887    0.006
4: cons 0.781 0.079 0.655 0.931    0.010
5: cons 0.899 0.123 0.689 1.171    0.400
6: cons 0.861 0.086 0.712 1.042    0.111
7: cons 0.795 0.104 0.637 0.991    0.043
8: cons 0.625 0.104 0.491 0.795    0.002

tau (Between-studies_SD) estimate: 0.085

tau2 (Between-studies_variance) estimate: 0.007
```

- Provides pooled risk ratio estimates with CIs.
- Also outputs heterogeneity parameters (τ and τ^2) for the network.

nma: Example Output (cont'd)

```
Multivariate H2-statistic: 1.607

Multivariate I2-statistic: 0.564

Test for Heterogeneity: Q(df = 21) = 33.74, p-value = 0.039

95%PI (vs. treat 1):
       Lower Upper
2: cons 0.778 1.945
3: cons 0.557 0.978
4: cons 0.601 1.014
5: cons 0.649 1.243
6: cons 0.657 1.130
7: cons 0.596 1.059
8: cons 0.454 0.860
```

- The Multivariate I^2 is interpreted in the same way as pairwise meta-analysis (proportion of variance due to heterogeneity).
- Multivariate Q test provides a formal heterogeneity test.
- 95% prediction intervals show the range of true treatment effects expected in a new study population.



Research Synthesis Methods WILEY

RESEARCH ARTICLE

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

Hisashi Noma | Yasuyuki Hamura | Masahiko Gosho | Toshi A. Furukawa

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; that is, 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.

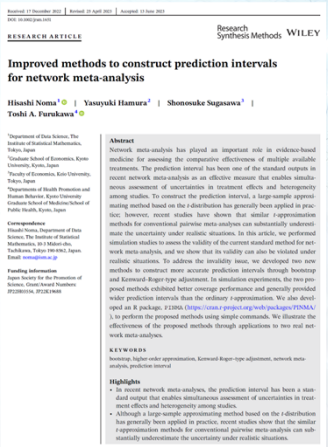
KEYWORDS

higher-order approximations, inconsistency, Kenward-Roger–type corrections, meta-regression, network meta-analysis

Highlights

- In practice of network meta-analysis, the restricted maximum likelihood (REML) method is one of the current standard inference methods, but recent studies have revealed the resultant confidence intervals can seriously underestimate statistical errors under certain conditions.

https://doi.org/10.1002/jrsm.1652



Research Synthesis Methods WILEY

RESEARCH ARTICLE

Improved methods to construct prediction intervals for network meta-analysis

Hisashi Noma | Yasuyuki Hamura | Shonosuke Sugasawa | Toshi A. Furukawa

Abstract

Network meta-analysis has played an important role in evidence-based medicine for assessing the comparative effectiveness of multiple available treatments. The prediction interval has been one of the standard outputs in recent network meta-analysis as an effective measure that enables simultaneous assessment of uncertainties in treatment effects and heterogeneity among studies. To construct the prediction interval, a large-sample approximating method based on the *t*-distribution has generally been applied in practice; however, recent studies have shown that similar *t*-approximation methods for conventional pairwise meta-analyses can substantially underestimate the uncertainty under realistic situations. In this article, we performed simulation studies to assess the validity of the current standard method for network meta-analysis, and we show that its validity can also be violated under realistic situations. To address the invalidity issue, we developed two new methods to construct more accurate prediction intervals through bootstrap and Kenward–Roger–type adjustment. In simulation experiments, the two proposed methods exhibited better coverage performance and generally provided wider prediction intervals than the ordinary *t*-approximation. We also developed an R package, PIMA (https://cran.r-project.org/web/packages/PIMA/), to perform the proposed methods using simple commands. We illustrate the effectiveness of the proposed methods through applications to two real network meta-analyses.

KEYWORDS

bootstrap, higher-order approximation, Kenward–Roger–type adjustment, network meta-analysis, prediction interval

Highlights

- In recent network meta-analyses, the prediction interval has been a standard output that enables simultaneous assessment of uncertainties in treatment effects and heterogeneity among studies.
- Although a large-sample approximating method based on the *t*-distribution has generally been applied in practice, recent studies show that the similar *t*-approximation methods for conventional pairwise meta-analysis can substantially underestimate the uncertainty under realistic situations.

https://doi.org/10.1002/jrsm.1651

57

Ranked Forest Plot

- Combines ranking information with forest plot presentation.
- Treatments are ordered from most to least effective.
- Confidence intervals are preserved, ensuring uncertainty is visible.
- Enhances interpretability for non-statistical audiences.

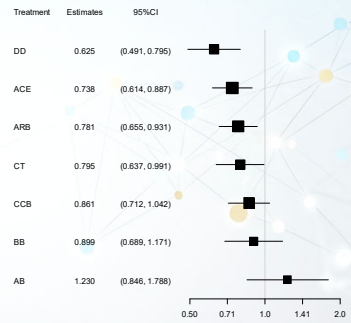
Nikolakopoulou et al. (2020)

58

nmaforest: Forest Plot Function

```
> nmaforest(hf3)
```

- nmaforest provides customizable forest plots.
- Allows adjustment of plot aesthetics (colors, fonts, axis labels).
- Useful for preparing figures for publication.
- Ensures results are communicated clearly and consistently.

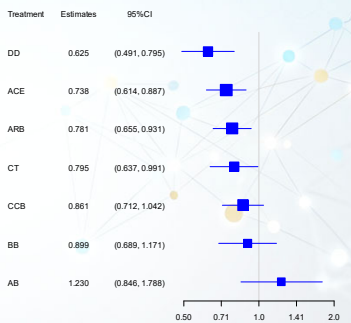


59

nmaforest: Example Output (Color Changed)

```
> nmaforest(hf3, col.plot="blue")
```

- Option `col.plot` allows you to change the color of the plot (here: blue).

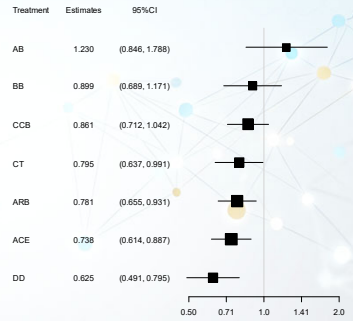


60

nmaforest: Example Output (Order Changed)

```
> nmaforest(hf3, ascending=FALSE)
```

- The option ascending=TRUE/FALSE lets you sort treatments in ascending or descending order of effect.
- Fine-tuning options (e.g., x-axis scale, tick marks, display settings) are demonstrated in the example programs on GitHub.



61

61

Treatment Ranking

Salanti (2012)

- In comparative analyses, ranking treatments is a central question: *Which treatment is best?*
- However, the distinction between 1st and 2nd place can be misleading:If the difference is very small, the ranking may easily change as more evidence accumulates. If the difference is large, the top rank is unlikely to be overturned.
- Integer ranks (1st, 2nd, 3rd, …) are often misleading, because they imply equal spacing between treatments regardless of actual evidence strength.
- This motivates defining continuous ranking measures, which capture not only order but also the magnitude of separation between treatments.

62

62

Ranking Metrics

- SUCRA: A 0–1 continuous index of relative ranking; values closer to 1 indicate treatments nearer the top rank. It reflects how far apart treatments are in ranking, not just their order.
- MeanRank: Expected rank on a 1–p scale (p = number of treatments); numerically consistent with SUCRA and often reported together.
- Rank probabilities: For each treatment, the probability of being 1st, 2nd, 3rd, …; useful for conveying uncertainty around the top ranks.

Salanti et al. (2011), Chaimani et al. (2013)

63

63

netrank: Ranking Statistics

```
> nmarank(hf3)
SUCRA:
          SUCRA
DD        0.993
ACE       0.795
ARB       0.661
CT        0.613
CCB       0.406
BB        0.347
Placebo   0.166
AB        0.018
```

- Here, DD's SUCRA ≈ 0.99, very close to 1, indicating it is highly likely to be the top-ranked treatment.
- ACE and ARB are ranked 2nd and 3rd, but their SUCRA values are considerably lower than DD, suggesting a large relative gap.
- AB has a SUCRA close to 0, indicating it is very likely to be the lowest-ranked treatment.

64

64

netrank: Mean Rank

```
MEANRANK:
          MEANRANK
DD           1.048
ACE          2.434
ARB          3.371
CT           3.706
CCB          5.155
BB           5.570
Placebo      6.841
AB           7.875
```

- The MeanRank table expands SUCRA onto a 1-p scale, where 1 = best.
- Reporting both SUCRA and MeanRank can aid readers unfamiliar with SUCRA while preserving comparability across reviews.

65

65

netrank: Rank Probabilities

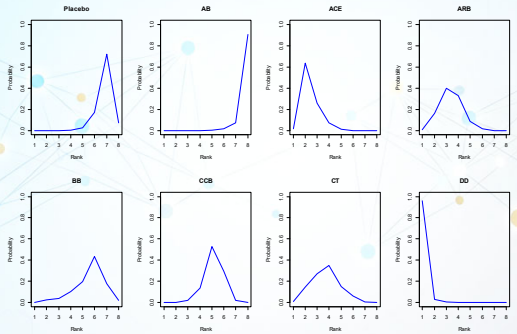
```
RANKPROB:
          Probability of rank 1 Probability of rank 2 Probability of rank 3 Probability of rank 4
DD                        0.965                 0.027                 0.006                 0.002
ACE                       0.017                 0.641                 0.260                 0.070
ARB                       0.007                 0.160                 0.400                 0.331
CT                        0.009                 0.149                 0.272                 0.348
CCB                       0.000                 0.002                 0.020                 0.138
BB                        0.002                 0.022                 0.042                 0.106
Placebo                   0.000                 0.000                 0.000                 0.005
AB                        0.000                 0.000                 0.000                 0.001
          Probability of rank 5 Probability of rank 6 Probability of rank 7 Probability of rank 8
DD                        0.000                 0.000                 0.000                 0.000
ACE                       0.011                 0.001                 0.000                 0.000
ARB                       0.088                 0.016                 0.000                 0.000
CT                        0.150                 0.064                 0.007                 0.000
CCB                       0.533                 0.289                 0.019                 0.000
BB                        0.189                 0.445                 0.174                 0.020
Placebo                   0.027                 0.169                 0.724                 0.076
AB                        0.002                 0.016                 0.076                 0.904
```

- Rank-probability matrix: rows = treatments; columns = probability of being rank 1, 2, 3, … .

66

66

netrank: Ranking Statistics



67

67

League Table

- A League Table summarizes all pairwise relative effects (e.g., risk ratios) in one matrix.
- It allows a quick, study-integrated view of how each treatment compares with every other option.
- Frequently presented in NMA papers to support transparent, side-by-side interpretation.

Salanti et al. (2011)

68

68

Figure 3: Efficacy (any mood episode relapse or recurrence) and tolerability (discontinuation due to adverse event) according to the network meta-analysis

Comparisons between treatments should be read from left to right and the estimates are in the cell in common between the column-defining treatment and the row-defining treatment. Drugs are reported in order of efficacy (any mood episode relapse or recurrence) ranking estimated by SUCRA (surface under the cumulative ranking curve). For tolerability, a risk ratio (RR) lower than 1·00 favours the row-defining treatment. For any mood episode relapse or recurrence, a RR lower than 1·00 favours the column-defining treatment. Significant results are in bold. The RR of drug B over drug A can be obtained by calculating the inverse of the RR of drug A over drug B. ARP=aripiprazole. CBZ=carbamazepine. CrI=credible interval. FLX=fluoxetine. IMP=imipramine. LIT=lithium. LTG=lamotrigine. OLZ=olanzapine. QTP=quetiapine. OXC=oxcarbazepine. PAL=paliperidone. PLB=placebo. RisLAI= risperidone longacting injection. VPA=valproate.

Miura, Noma, Furukawa et al. (2014)

69

69

netleague: League Table

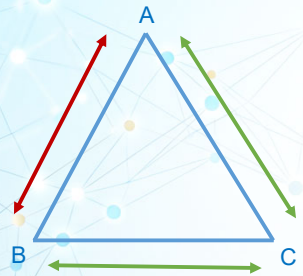
```
> nmaleague(hf3, eform=TRUE, digits=2)
League table:
Placebo          1.23 (0.85, 1.79)  0.74 (0.61, 0.89)  0.78 (0.65, 0.93)  0.90 (0.69, 1.17)  0.86 (0.71, 1.04)  0.79 (0.64, 0.99)  0.62 (0.49, 0.80)
0.81 (0.56, 1.18) AB                0.60 (0.42, 0.86)  0.63 (0.44, 0.92)  0.73 (0.49, 1.09)  0.70 (0.49, 1.00)  0.65 (0.45, 0.94)  0.51 (0.38, 0.68)
1.36 (1.13, 1.63)  1.67 (1.17, 2.38) ACE               1.06 (0.89, 1.26)  1.22 (0.95, 1.56)  1.17 (1.00, 1.36)  1.08 (0.90, 1.28)  0.85 (0.69, 1.04)
1.28 (1.07, 1.53)  1.58 (1.08, 2.29)  0.95 (0.80, 1.12) ARB                1.15 (0.92, 1.44)  1.10 (0.93, 1.30)  1.02 (0.84, 1.24)  0.80 (0.63, 1.02)
1.11 (0.85, 1.45)  1.37 (0.91, 2.05)  0.82 (0.64, 1.05)  0.87 (0.69, 1.09) BB                0.96 (0.76, 1.20)  0.88 (0.68, 1.15)  0.70 (0.52, 0.93)
1.16 (0.96, 1.40)  1.43 (1.00, 2.05)  0.86 (0.73, 1.00)  0.91 (0.77, 1.07)  1.04 (0.83, 1.31) CCB                0.92 (0.78, 1.09)  0.73 (0.59, 0.90)
1.26 (1.01, 1.57)  1.55 (1.07, 2.24)  0.93 (0.78, 1.11)  0.98 (0.81, 1.19)  1.13 (0.87, 1.47)  1.08 (0.92, 1.28) CT                0.79 (0.62, 1.00)
1.60 (1.26, 2.04)  1.97 (1.47, 2.63)  1.18 (0.96, 1.45)  1.25 (0.98, 1.59)  1.44 (1.08, 1.92)  1.38 (1.11, 1.71)  1.27 (1.00, 1.62) DD
```

- Exportable as CSV for reporting; ideal for guideline appendices and technical supplements.

70

70

Validity of Comparisons on the Network



- Consistency requirement: the direct A-B effect should agree with the indirect A-B effect via other paths (e.g., A-C-B).
- If these disagree, network comparisons become unreliable and must be investigated before ranking or recommending treatments.

Salanti (2012), Veroniki et al. (2013)

71

71

Inconsistency

- Consistency: Effect sizes along all paths connecting any treatment pair A-B are equal (consistent).
- Inconsistency: Lack of consistency; network comparisons become invalid.
- Conceptualized as design-by-treatment interaction (design = set of treatments compared within a trial).

Higgins et al. (2012), White et al. (2012)

72

72

Local Inconsistency Tests

- Extract each triangular loop in the network and test whether the treatment effects on the three edges are consistent.
- Known as the Bucher test or Loop Inconsistency test.

Bucher et al. (1997), Veroniki et al. (2013)

73

local.ict: Local Inconsistency Test

```
> local.ict(hf3)
Coding:
 code treatment
    1   Placebo
    2        AB
    3       ACE
    4       ARB
    5        BB
    6       CCB
    7        CT
    8        DD

Reference: Placebo
```

- The local.ict function automatically carries out the entire local inconsistency test, including the calculation of p-values, without requiring manual specification of loops or contrasts.

74

local.ict: Local Inconsistency Test (cont'd)

```
Local inconsistency tests:
           N   tau X2-statistic df Pr(>X2)
loop: 1-3-4 4 0.242       0.018  1   0.893
loop: 1-3-6 7 0.007       2.514  1   0.113
loop: 1-3-8 4 0.169       2.968  1   0.085
loop: 1-4-6 6 0.180       1.407  1   0.236
loop: 1-6-8 9 0.009       0.737  1   0.391
loop: 3-4-5 3 0.009       0.380  1   0.538
loop: 3-4-6 5 0.008       0.024  1   0.877
loop: 3-4-7 5 0.156       1.666  1   0.197
loop: 3-5-6 5 0.007       1.271  1   0.260
loop: 3-6-7 6 0.008       3.415  3   0.332
loop: 3-6-8 8 0.009       3.474  3   0.324
loop: 4-5-6 3 0.007       1.976  1   0.160
loop: 4-6-7 6 0.009       3.605  1   0.058
```

- Loops with $p < 0.05$ are suspected to exhibit inconsistency.

75

Global Inconsistency Test

- Tests whether any inconsistency exists in the entire network.
  - Null hypothesis: no inconsistent components.
  - Alternative hypothesis: at least one inconsistent component exists.
- Constructed via likelihood ratio test comparing the consistency model vs. design-by-treatment interaction model.

Higgins et al. (2012)

76

global.ict: Global Inconsistency Test

```
> global.ict(hf3)
Coding:
 code treatment
    1   Placebo
    2        AB
    3       ACE
    4       ARB
    5        BB
    6       CCB
    7        CT
    8        DD

Reference: Placebo

Number of studies: 26
```

77

77

global.ict: Global Inconsistency Test (cont'd)

```
Designs: 1-3 1-4 1-6 1-8 2-8 3-4 3-5 3-6 3-6-7 3-6-8 3-7 3-8 4-5 4-6 4-7 5-6 6-7 6-8

Coefficients of the design-by-treatment interaction model:
             Est.    SE  Lower  Upper Pr(>|z|)
2: cons    -0.286 0.327 -0.926  0.355    0.382
3: cons    -0.248 0.155 -0.552  0.057    0.111
4: cons    -0.143 0.134 -0.405  0.119    0.284
4: des_3-4 -0.057 0.257 -0.559  0.446    0.825
5: cons    -0.424 0.483 -1.372  0.523    0.380
5: des_4-5  0.336 0.533 -0.709  1.380    0.529
6: cons    -0.394 0.195 -0.775 -0.012    0.043
6: des_3-6  0.329 0.664 -0.972  1.629    0.620
6: des_3-6-7 0.372 0.306 -0.227  0.971    0.223
6: des_3-6-8 0.290 0.292 -0.282  0.862    0.320
6: des_4-6  0.380 0.285 -0.178  0.938    0.182
6: des_5-6 -0.204 0.552 -1.287  0.879    0.712
7: cons    -0.079 0.236 -0.542  0.384    0.738
7: des_3-7 -0.297 0.283 -0.851  0.258    0.295
7: des_4-7  0.328 0.339 -0.337  0.993    0.334
7: des_6-7 -0.535 0.341 -1.203  0.133    0.116
8: cons    -0.963 0.287 -1.525 -0.401    0.001
8: des_3-6-8 0.545 0.359 -0.159  1.250    0.129
8: des_3-8  0.840 0.391  0.072  1.607    0.032
8: des_6-8  0.151 0.424 -0.681  0.983    0.722
```

- Detailed output: coefficients of the design-by-treatment interaction model.
- Concept: inconsistency is modeled as effect modification by design.
- Practical note: usually, users only focus on the global p-value, not these details.

78

78

global.ict: Global Inconsistency Test (cont'd)

```
Between-studies SD: 0.142

Between-studies COR: 0.5

X2-statistic: 13.918

df: 13

P-value: 0.38
```

- Practical use: look mainly at p-value. If $p < 0.05$, inconsistency is significant at the network level.

79

79

Sidesplitting

- Inconsistency = considered as disagreement between direct and indirect evidence (in the early days, and it can be regarded as a specific version of design-by-treatment interaction).
- Sidesplitting method evaluates this for each treatment pair.
- It decomposes estimates into direct and indirect contributions, then compares them.

Dias et al. (2010), Noma et al. (2017), Noma (2024a)

80

80

sidesplit: Example

```
> sidesplit(hf2)
Coding:
 code treatment
    1   Placebo
    2        AB
    3       ACE
    4       ARB
    5        BB
    6       CCB
    7        CT
    8        DD

Reference: Placebo
```

- Output: decomposition of direct and indirect evidence, and *p*-values for inconsistency.

81

sidesplit: Example (cont'd)

```
Direct evidence:
          Est.    SE  Lower  Upper Pr(>|z|)
1 vs. 3 -0.248 0.131 -0.504  0.009    0.059
1 vs. 4 -0.142 0.123 -0.383  0.098    0.246
1 vs. 6 -0.385 0.173 -0.724 -0.045    0.026
1 vs. 8 -0.963 0.249 -1.451 -0.475    0.000
2 vs. 8 -0.677 0.107 -0.887 -0.467    0.000
3 vs. 4  0.048 0.135 -0.218  0.313    0.724
3 vs. 5 -0.177 0.444 -1.046  0.693    0.690
3 vs. 6  0.179 0.120 -0.056  0.414    0.135
3 vs. 7  0.052 0.137 -0.217  0.321    0.707
3 vs. 8 -0.087 0.106 -0.294  0.120    0.410
4 vs. 5  0.055 0.135 -0.210  0.319    0.685
4 vs. 6  0.129 0.137 -0.139  0.397    0.345
4 vs. 7  0.386 0.186  0.021  0.750    0.038
5 vs. 6 -0.173 0.135 -0.437  0.091    0.198
6 vs. 7 -0.149 0.105 -0.355  0.057    0.155
6 vs. 8 -0.341 0.126 -0.588 -0.094    0.007
```

- Direct evidence estimates with SE, CI, and *p*-values.
- Interpretation: provides effect sizes from studies directly comparing each pair.

82

sidesplit: Example (cont'd)

```
Indirect evidence:
          Est.     SE    Lower   Upper Pr(>|z|)
1 vs. 3 -0.379  0.118   -0.610  -0.147    0.001
1 vs. 4 -0.415  0.133   -0.676  -0.155    0.002
1 vs. 6 -0.085  0.076   -0.235   0.065    0.266
1 vs. 8 -0.401  0.066   -0.530  -0.272    0.000
2 vs. 8 -1.760 60.000 -119.358 115.838    0.977
3 vs. 4  0.059  0.105   -0.148   0.265    0.578
3 vs. 5  0.222  0.116   -0.005   0.449    0.055
3 vs. 6  0.118  0.127   -0.131   0.368    0.353
3 vs. 7  0.118  0.149   -0.173   0.409    0.428
3 vs. 8 -0.469  0.196   -0.853  -0.086    0.016
4 vs. 5  0.247  0.148   -0.044   0.538    0.096
4 vs. 6  0.083  0.099   -0.111   0.277    0.401
4 vs. 7 -0.087  0.099   -0.280   0.107    0.381
5 vs. 6  0.095  0.139   -0.178   0.367    0.497
6 vs. 7  0.073  0.163   -0.247   0.393    0.654
6 vs. 8 -0.295  0.195   -0.677   0.086    0.130
```

- Indirect evidence estimates with SE, CI, and *p*-values.
- Interpretation: estimates derived indirectly from the rest of the network.

83

sidesplit: Example (cont'd)

```
Difference between direct and indirect evidence with inconsistency test:
          Est.     SE    Lower   Upper Pr(>|z|)
1 vs. 3  0.131  0.176   -0.215   0.477    0.458
1 vs. 4  0.273  0.181   -0.081   0.628    0.131
1 vs. 6 -0.300  0.189   -0.671   0.072    0.114
1 vs. 8 -0.562  0.258   -1.067  -0.057    0.029
2 vs. 8  1.083 60.000 -116.515 118.681    0.986
3 vs. 4 -0.011  0.171   -0.347   0.325    0.950
3 vs. 5 -0.399  0.459   -1.298   0.500    0.384
3 vs. 6  0.061  0.175   -0.282   0.404    0.728
3 vs. 7 -0.066  0.202   -0.463   0.330    0.743
3 vs. 8  0.382  0.222   -0.053   0.818    0.085
4 vs. 5 -0.192  0.200   -0.585   0.200    0.337
4 vs. 6  0.046  0.169   -0.285   0.377    0.786
4 vs. 7  0.472  0.211    0.060   0.885    0.025
5 vs. 6 -0.268  0.194   -0.647   0.112    0.167
6 vs. 7 -0.222  0.194   -0.602   0.158    0.252
6 vs. 8 -0.046  0.232   -0.501   0.408    0.842
```

- Output: Difference (Direct - Indirect) with CI and *p*-value.
- If difference ≈ 0 and $p > 0.05$, direct and indirect evidence agree.
- Significant differences indicate possible inconsistency.

84

Diovan Scandal (Japan)

- Valsartan (brand name Diovan) trials in Japan manipulated by Novartis staff, who concealed involvement and altered analyses.
- Published in *The Lancet*, later retracted due to misconduct and conflict-of-interest concealment.
- In the NMA, the Jikei Heart Study (part of scandal) made ARB vs. CT comparisons appear spuriously strong, inconsistent with indirect evidence.

85

85

Comparison-adjusted Funnel Plot

- Publication bias assessment in NMA is challenging.
- Direct comparisons vary across different pairs.
- Sponsors differ, bias direction may even reverse.
- Many treatment comparisons have few studies.
- Proposed approach: overlay funnel plots against a common comparator (e.g., placebo) to assess asymmetry.

Chaimani et al. (2013)

86

86

nmafunnel: Funnel Plot

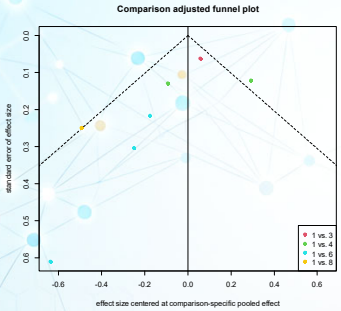
```
> nmafunnel(hf3,legends="bottomright")
# Comparison-adjusted funnel plot for placebo-controlled trials
Comparison adjusted funnel plot for the trials involving treatment 1 (as control)
Coding:
 code treatment
    1   Placebo
    2        AB
    3       ACE
    4       ARB
    5        BB
    6       CCB
    7        CT
    8        DD

Summary:
 design N     n
    1-3 1  9297
    1-4 2  7439
    1-6 3 16800
    1-8 1  3845
```

- Produces comparison-adjusted funnel plots for trials against placebo.

87

87

nmafunnel: Funnel Plot (cont'd)



- Example: placebo-controlled trials, funnel plots superimposed by design.
- If no publication bias: overall plot should be symmetric.
- Color-coding aids visual detection of asymmetry (requires enough studies).

88

88

Transitivity

- Transitivity assumption: indirect comparisons are valid only if patient populations and settings are sufficiently similar across trials.
- Example: for A vs B and B vs C trials to imply valid A vs C comparison, study characteristics must be comparable.
- Currently, no strict mathematical criteria; usually assessed by comparing distributions of effect modifiers.

Salanti (2012)

89

89

transitivity: Example

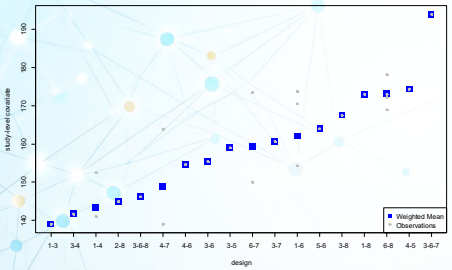
```
> transitivity(hf2, SBP)
Summary:
 design N     n wt.mean     min     max
    1-3 1  9297 139.000 139.000 139.000
    3-4 1 17118 141.750 141.750 141.750
    1-4 2  7439 143.340 141.001 152.504
    2-8 1 24335 145.000 145.000 145.000
  3-6-8 1 33357 146.254 146.254 146.254
    4-7 2  5129 148.940 139.000 163.895
    4-6 1 15245 154.649 154.649 154.649
    3-6 1   470 155.500 155.500 155.500
    3-5 1   758 159.000 159.000 159.000
    6-7 2 27357 159.387 150.100 173.450
    3-7 1 10985 160.700 160.700 160.700
    1-6 3 16800 162.072 154.300 173.849
    5-6 1 19257 164.000 164.000 164.000
    3-8 1  6083 167.500 167.500 167.500
    1-8 1  3845 173.000 173.000 173.000
    6-8 4 10031 173.365 168.950 178.150
    4-5 1  9193 174.400 174.400 174.400
  3-6-7 1  6614 194.000 194.000 194.000
```

- The transitivity function automatically produces summaries of covariates (here, systolic blood pressure: SBP) across study designs in the network.
- This allows you to compare the distributions of baseline covariates by design.
- The function also provides side-by-side boxplots of these distributions (as proposed by Cipriani, Furukawa, Salanti et al., 2018).
- Such plots help assess the heterogeneity of covariate distributions, offering a practical way to evaluate the plausibility of the transitivity assumption in network meta-analysis.

90

90

transitivity: Example (cont'd)



- Output: side-by-side boxplots of baseline SBP across designs.
- Interpretation: heterogeneity in covariate distributions provides insights for assessing the plausibility of transitivity.

91

91

Miscellaneous Functions

- Network meta-regression (nmareg): evaluates effect modification by covariates through multivariate meta-regression.
- Jackson's Random Inconsistency Model (random.icm): expresses Higgins' design-by-treatment interaction model as a random-effects model; inconsistency tested via variance components.

Noma et al. (2025)

92

92

nmareg: Network meta-regression

```
> nmareg(hf3,z=SBP,treats=3)
Call:
nmareg(x = hf3, z = SBP, treats = 3)

Coding:
 code treatment
    1   Placebo
    2        AB
    3       ACE
    4       ARB
    5        BB
    6       CCB
    7        CT
    8        DD

Covariates: SBP

Outcomes evaluated the effect modifications: 3 vs. 1
```

```
Coefficients:
         Coef.    SE  Lower  Upper Pr(>|z|)
2: cons  0.182 0.151 -0.113  0.477    0.227
3: cons  0.140 0.425 -0.692  0.972    0.741
3: SBP  -0.003 0.003 -0.009  0.003    0.284
4: cons -0.261 0.079 -0.415 -0.106    0.001
5: cons -0.131 0.125 -0.375  0.113    0.293
6: cons -0.182 0.088 -0.354 -0.010    0.038
7: cons -0.280 0.112 -0.499 -0.062    0.012
8: cons -0.496 0.102 -0.694 -0.297    0.000

Between-studies SD: 0.09

Between-studies COR: 0.5
```


93

random.icm: Random inconsistency model

```
> random.icm(hf3)
Number of studies: 26

Number of designs: 18

Designs:
1-3 1-4 1-6 1-8 2-8 3-4 3-5 3-6 3-6-7 3-6-8 3-7 3-8 4-5 4-6 4-7 5-6 6-7 6-8

Coef. (vs. treat 1):
         Coef.    SE  95%CL  95%CU P-value
2: cons  0.194 0.160 -0.119  0.508   0.224
3: cons -0.315 0.082 -0.476 -0.153   0.000
4: cons -0.260 0.085 -0.428 -0.093   0.002
5: cons -0.121 0.130 -0.376  0.134   0.351
6: cons -0.161 0.089 -0.335  0.013   0.070
7: cons -0.239 0.109 -0.452 -0.026   0.028
8: cons -0.483 0.106 -0.691 -0.275   0.000

Between-studies SD: 0.079

Between-designs SD: 0.061
```


94

random.icm: Random inconsistency model

```
Likelihood ratio tests for the variance components:
                            X2-statistic df P-value
Heterogeneity                      0.251  1   0.616
Inconsistency                      0.072  1   0.789
Heterogeneity + Inconsistency      0.323  2   0.851

Heterogeneity and inconsistency statistics:
                            R-statistic I2-statistic
Heterogeneity                     1.464        0.533
Inconsistency                     1.096        0.168
Heterogeneity + Inconsistency     1.605        0.612
```


95

Miscellaneous Functions (cont'd)

- Decomposition of Q statistic (nmaQ): partitions Q statistic into within-design and between-design components for detailed heterogeneity assessment.
- Contribution weight plot (nmaweight): quantifies each study's or comparison's contribution to the final synthesis; visualized as heatmaps for intuitive interpretation.

Noma et al. (2025)

96

nmaQ: Q-statistic decomposition

```
> nmaQ(hf3)

Within designs (individual designs):
      N     n     Q df Pr(>Q)
1-3   1  9297 0.000  0  1.000
1-4   2  7439 4.783  1  0.029
1-6   3 16800 0.515  2  0.773
1-8   1  3845 0.000  0  1.000
2-8   1 24335 0.000  0  1.000
3-4   1 17118 0.000  0  1.000
3-5   1   758 0.000  0  1.000
3-6   1   470 0.000  0  1.000
3-6-7 1  6614 0.000  0  1.000
3-6-8 1 33357 0.000  0  1.000
3-7   1 10985 0.000  0  1.000
3-8   1  6083 0.000  0  1.000
4-5   1  9193 0.000  0  1.000
4-6   1 15245 0.000  0  1.000
4-7   2  5129 1.387  1  0.239
5-6   1 19257 0.000  0  1.000
6-7   2 27357 0.072  1  0.788
6-8   4 10031 4.600  3  0.204
```

97

97

nmaQ: Q-statistic decomposition

```
Q-statistics:
                    Q df Pr(>Q)
Within designs  11.358  8  0.182
Between designs 22.383 13  0.050
Total           33.740 21  0.039
```

98

98

nmaweight: Contribution weight plot

```
> nmaweight(hf3)

Contribution of direct and indirect information:
          1-2 1-3   1-4 1-5   1-6 1-7   1-8 2-3 2-4 2-5 2-6 2-7 2-8   3-4   3-5   3-6   3-7   3-8
direct      0 0.5 0.516   0 0.205   0 0.135   0   0   0   0   0   1 0.415 0.062 0.584 0.535 0.764
indirect    1 0.5 0.484   1 0.795   1 0.865   1   1   1   1   1   0 0.585 0.938 0.416 0.465 0.236
            4-5   4-6   4-7 4-8   5-6 5-7 5-8   6-7  6-8 7-8
direct    0.547 0.373 0.224   0 0.519   0   0 0.703 0.74   0
indirect 0.453 0.627 0.776   1 0.481   1   1 0.297 0.26   1
```

Noma et al. (2017)

99

99

Research Article

Statistics in Medicine

Received: 28 February 2016, Accepted: 9 November 2016 Published online 4 December 2016 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.7187

Quantifying indirect evidence in network meta-analysis

Hisashi Noma,[a*†] Shiro Tanaka,[b] Shigeyuki Matsui,[c] Andrea Cipriani[d] and Toshi A. Furukawa[e]

Network meta-analysis enables comprehensive synthesis of evidence concerning multiple treatments and their simultaneous comparisons based on both direct and indirect evidence. A fundamental pre-requisite of network meta-analysis is the consistency of evidence that is obtained from different sources, particularly whether direct and indirect evidence are in accordance with each other or not, and how they may influence the overall estimates. We have developed an efficient method to quantify indirect evidence, as well as a testing procedure to evaluate their inconsistency using Lindsay's composite likelihood method. We also show that this estimator has complete information for the indirect evidence. Using this method, we can assess the degree of consistency between direct and indirect evidence and their contribution rates to the overall estimate. Sensitivity analyses can be also conducted with this method to assess the influences of potentially inconsistent treatment contrasts on the overall results. These methods can provide useful information for overall comparative results that might be biased from specific inconsistent treatment contrasts. We also provide some fundamental requirements for valid inference on these methods concerning consistency restrictions on multi-arm trials. In addition, the efficiency of the developed method is demonstrated based on simulation studies. Applications to a network meta-analysis of 12 new-generation antidepressants are presented. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: network meta-analysis; indirect evidence; likelihood factorization; composite likelihood methods; inconsistency; sensitivity analysis

https://doi.org/10.1002/sim.7187

100

100

nmaweight: Contribution weight plot

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

101

Summary

- Network meta-analysis (NMA) is expected to become even more widespread as a standard tool of systematic reviews, supported by the growth of Comparative Effectiveness Research.
- PRISMA-NMA extension (Hutton et al., 2015) provides reporting guidelines that address NMA-specific concerns.
- With the R package "NMA", even complex state-of-the-art analyses can be executed using simple commands.

102

References

- ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. (2002). Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA* **288**, 2981-2997.
- Bucher, H. C., Guyatt, G. H., Griffith, L. E., and Walter, S. D. (1997). The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology* **50**, 683-691.
- Caldwell, D. M., Ades, A. E., and Higgins, J. P. (2005). Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* **331**, 897-900.
- Chaimani, A., Higgins, J. P., Mavridis, D., Spyridonos, P., and Salanti, G. (2013). Graphical tools for network meta-analysis in STATA. *PLoS One* **8**, e76654.
- Cipriani, A., Furukawa, T. A., Salanti, G., *et al.* (2018). 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* **391**, 1357-1366.
- Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics* **10**, 101-129.
- Del Fiol, G., Workman, T. E., and Gorman, P. N. (2014). Clinical questions raised by clinicians at the point of care: a systematic review. *JAMA Internal Medicine* **174**, 710-718.

103

- Dias, S., Welton, N. J., Caldwell, D. M., and Ades, A. E. (2010). Checking consistency in mixed treatment comparison meta-analysis. *Statistics in Medicine* **29**, 932-944.
- Donegan, S., Connor, J., Alfirevic, Z., and Tudur-Smith, C. (2024). The impact of Cochrane Reviews that apply network meta-analysis in clinical guidelines: A systematic review. *PLoS One* **19**, e0315563.
- Egger, M., Davey Smith, G., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ* **315**, 629-634.
- Egger, M., Higgins, J. P., and Smith, G. D. (2022). *Systematic Reviews in Health Research: Meta-Analysis in Context*, 3rd edition. London: BMJ Books.
- Higgins, J. P., Jackson, D., Barrett, J. K., Lu, G., Ades, A. E., and White, I. R. (2012). Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Research Synthesis Methods* **3**, 98-110.
- Higgins, J. P., and Whitehead, A. (1996). Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* **15**, 2733-2749.
- Higgins, J. P. T., and Thomas, J. (2019). *Cochrane Handbook for Systematic Reviews of Interventions*, 2nd edition. Chichester: Wiley-Blackwell.
- Higgins, J. P. T., and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* **21**, 1539-1558.

104

- Higgins, J. P. T., Thompson, S. G., and Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society, Series A* **172**, 137-159.
- Hutton, B., Salanti, G., Caldwell, D. M., et al. (2015). The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Annals of Internal Medicine* **162**, 777-784.
- IntHout, J., Ioannidis, J. P., Rovers, M. M., and Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open* **6**, e010247.
- Jackson, D., Riley, R., and White, I. R. (2011). Multivariate meta-analysis: potential and promise. *Statistics in Medicine* **30**, 2481-2498.
- Jackson, D., White, I. R., and Riley, R. D. (2012). Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. *Statistics in Medicine* **31**, 3805-3820.
- Kanters, S., Ford, N., Druyts, E., Thorlund, K., Mills, E. J., and Bansback, N. (2016). Use of network meta-analysis in clinical guidelines. *Bulletin of the World Health Organization* **94**, 782-784.
- Mitka, M. (2010). US government kicks off program for comparative effectiveness research. *JAMA* **304**, 2230-2231.

105

- Miura, T., Noma, H., Furukawa, T. A., *et al.* (2014). Comparative efficacy and tolerability of pharmacological treatments in the maintenance treatment of bipolar disorder: a systematic review and network meta-analysis. *Lancet Psychiatry* **1**, 351-359.
- Nagashima, K., Noma, H., and Furukawa, T. A. (2019). Prediction intervals for random-effects meta-analysis: A confidence distribution approach. *Statistical Methods in Medical Research* **28**, 1689-1702.
- Nikolakopoulou, A., Higgins, J. P. T., Papakonstantinou, T., *et al.* (2020). CINeMA: An approach for assessing confidence in the results of a network meta-analysis. *PLoS Medicine* **17**, e1003082.
- Nikolakopoulou, A., Mavridis, D., and Salanti, G. (2014). How to interpret meta-analysis models: fixed effect and random effects meta-analyses. *Evidence-Based Mental Health* **17**, 64.
- Noma, H. (2024a). Sidesplitting using network meta-regression. *Japanese Journal of Biometrics* **44**, 107-118.
- Noma, H. (2024b). Within-study covariance estimators for network meta-analysis with contrast-based approach. *Japanese Journal of Biometrics* **44**, 119-126.
- Noma, H., Hamura, Y., Gosho, M., and Furukawa, T. A. (2023a). Kenward-Roger-type corrections for inference methods of network meta-analysis and meta-regression. *Research Synthesis Methods* **14**, 731-741.

106

- Noma, H., Hamura, Y., Sugasawa, S., and Furukawa, T. A. (2023b). Improved methods to construct prediction intervals for network meta-analysis. *Research Synthesis Methods* **14**, 794-806.
- Noma, H., Maruo, K., Tanaka, S., and Furukawa, T. A. (2025). NMA: Network meta-analysis based on multivariate meta-analysis and meta-regression models in R. *medRxiv*, 2025.2009.2015.25335823.
- Noma, H., Tanaka, S., Matsui, S., Cipriani, A., and Furukawa, T. A. (2017). Quantifying indirect evidence in network meta-analysis. *Statistics in Medicine* **36**, 917-927.
- Salanti, G. (2012). Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research Synthesis Methods* **3**, 80-97.
- Salanti, G., Ades, A. E., and Ioannidis, J. P. (2011). Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *Journal of Clinical Epidemiology* **64**, 163-171.
- Sciarretta, S., Palano, F., Tocci, G., Baldini, R., and Volpe, M. (2011). Antihypertensive treatment and development of heart failure in hypertension: a Bayesian network meta-analysis of studies in patients with hypertension and high cardiovascular risk. *Archives of Internal Medicine* **171**, 384-394.

107

- Veroniki, A. A., Vasiliadis, H. S., Higgins, J. P., and Salanti, G. (2013). Evaluation of inconsistency in networks of interventions. *International Journal of Epidemiology* **42**, 332-345.
- White, I. R., Barrett, J. K., Jackson, D., and Higgins, J. P. (2012). Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Research Synthesis Methods* **3**, 111-125.
- White, I. R., Turner, R. M., Karahalios, A., and Salanti, G. (2019). A comparison of arm-based and contrast-based models for network meta-analysis. *Statistics in Medicine* **38**, 5197-5213.

108