Frequentist network meta-analysis using the R package netmeta

Gerta Rücker

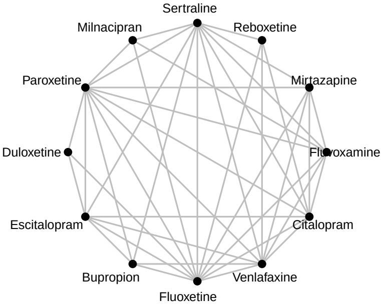
Institute for Medical Biometry and Statistics and Cochrane Germany
Medical Center - University of Freiburg

ruecker@imbi.uni-freiburg.de

European Statistical Meeting on Evidence Synthesis
Bruxelles, November 22, 2016

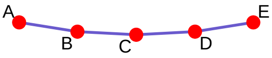
Outline

Graph-theoretical methods for network meta-analysis



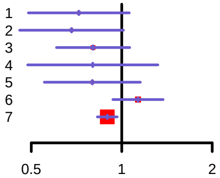
- Networks are graphs
  - Nodes are treatments
  - Edges are comparisons between treatments, based on studies
- 'Variances combine like electrical resistances' (Bailey, 2007)
- It is possible to apply methods from electrical network theory to network meta-analysis (Rücker, 2012)

Variances combine like electrical resistances



- Connection in series Variances in a chain of $n - 1$ independent comparisons of successive treatments $A, B, C, \ldots$ add:

$$V_{A-E} = V_{A-B} + V_{B-C} + V_{C-D} + V_{D-E}$$



- Parallel connection For a pairwise meta-analysis with parallel comparisons, inverse variances add:

$$\frac{1}{V(\bar{x})} = \sum_k \frac{1}{V_k}$$

Terminology in meta-analysis and electrical networks

Meta-analytic network		Electrical network
Treatments $i = 1, \ldots, n$	$\Longleftrightarrow$	Nodes $i = 1, \ldots, n$
Existing comparisons $e = 1, \ldots, m$	$\Longleftrightarrow$	Edges $e = 1, \ldots, m$
Variance V_e	$\Longleftrightarrow$	Resistance R_e
Inverse variance weight $w_e = 1/V_e$	$\Longleftrightarrow$	Conductance $1/R_e$
Outcome of treatment i	$\Longleftrightarrow$	Potential at node i
Treatment effect $i – j$	$\Longleftrightarrow$	Voltage at edge $i – j$
Weighted treatment effect $i – j$	$\Longleftrightarrow$	Current flow at edge $i – j$

- Ohm's law relates treatment effects and weights
- Kirchhoff's current law says how to combine the observed effects
- Kirchhoff's potential law guarantees consistency of the estimated treatment effects over closed circuits
  - Consistency means that the difference between two treatments is always the same, whatever (direct or indirect) path is chosen

Statistical model

Model

$$\hat{\theta} = \mathbf{X}\theta^{treat} + \epsilon, \qquad \epsilon \sim N(\mathbf{0}, \mathbf{\Sigma}),$$

where

- $\hat{\theta}$ is a vector of m observed pairwise comparisons with known standard errors $\mathbf{s} = (s_1, s_2, \dots, s_m)$
- $\mathbf{X}$ is the $m \times n$ design matrix defining the network structure
- θ^{treat} a vector of length n (number of treatments)
- $\mathbf{\Sigma}$ is a diagonal matrix whose i^{th} entry is s_i^2.

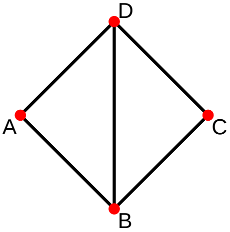
Note:

- If there are K two-arm trials, $\hat{\theta}$ has length K
- If there are also multi-arm trials, $\hat{\theta}$ has length $m \geq K$ with m denoting the total number of pairwise comparisons

Example network with $n = 4$ arms

Example network with $n = 4$ arms

- $\boldsymbol{\theta}^{treat} = (\theta_A, \theta_B, \theta_C, \theta_D)^T$
- $K = 5$ studies each providing a single pairwise treatment comparison
- $m = 5$ pairwise treatment comparisons
- Model:



$$\begin{pmatrix} \hat{\theta}_1^{AB} \\ \hat{\theta}_2^{BC} \\ \hat{\theta}_3^{CD} \\ \hat{\theta}_4^{AD} \\ \hat{\theta}_5^{BD} \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \theta_A \\ \theta_B \\ \theta_C \\ \theta_D \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{pmatrix}$$

$$= \mathbf{X}\boldsymbol{\theta}^{treat} + \epsilon$$

Estimation under the fixed effect model

- $\mathbf{W} = \text{diag}(1/s_1^2, \ldots, 1/s_m^2)$ diagonal matrix (dimension $m \times m$) of inverse variance weights
- Network estimates $\hat{\boldsymbol{\theta}}^{nma}$ estimated by

$$\hat{\boldsymbol{\theta}}^{nma} = \mathbf{H}\hat{\boldsymbol{\theta}}$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^+\mathbf{X}^T\mathbf{W}$ is known as the *hat matrix* in regression.
- Interpretation: The network estimates are weighted sums of the observed estimates with weights coming from the rows of **H**.
- Standard errors calculated from the variance-covariance matrix

$$\widehat{\text{Cov}}\,(\hat{\boldsymbol{\theta}}^{nma}) = \mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^+\mathbf{X}^T$$

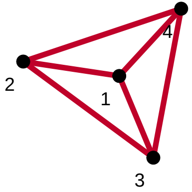
- Heterogeneity/inconsistency measured by generalised Q_{total} statistic

$$Q_{total} = (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{nma})^\top \mathbf{W}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{nma})$$

(Jackson et al., 2012; Rücker, 2012; Krahn et al., 2013)

Multi-arm studies: Need to account for correlation

- A study with k arms contributes $\binom{k}{2}$ pairwise comparisons
- Note: These are correlated, as there are only k treatments
  - $k-1$ independent comparisons
  - $k-1$ degrees of freedom (df)
- Example $k = 4$: $df = 3$



Adjustment for correlation within multi-arm studies

Standard approach: Reduce dimension

(Lu et al., 2011; Higgins et al., 2012; White et al., 2012; König et al., 2013)

- Based on standard regression methodology
- For each multi-arm study, choose a study-specific reference treatment
- Consider only comparisons to the reference treatment ('basic parameters')

Alternative approach: Reduce weights

(Rücker, 2012; Rücker and Schwarzer, 2014)

- Based on electrical network methodology
- For each multi-arm study, reduce all 'conductances' (weights) by specific factors that must be calculated
- Implemented in the R package **netmeta** (Rücker et al., 2016)

Comparison of the approaches

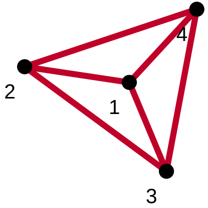
Standard approach

- Natural for statisticians with a background in regression analysis

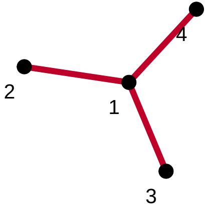
Alternative approach

- Natural for scientists coming from graph theory and its applications

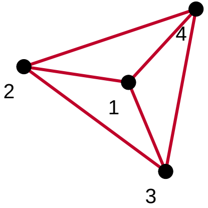
Given a four-arm study with six comparisons,



we may cut off three of six comparisons:



or reduce all weights by $1/2$ (in average):



Examples

1. Diabetes data
   Network of 10 diabetes treatments including 26 studies, where the outcome was HbA1c (measured as mean change or mean post treatment value) (Senn et al., 2013)
2. Smoking cessation data
   Network of four interventions for smoking cessation (binary outcome) (Higgins et al., 2012; Dias et al., 2013)

Both examples are part of R package **netmeta**

How to use R package netmeta: Diabetes data

```
# Make R package netmeta available
install.packages("netmeta")
library(netmeta)
```

```
# Load diabetes data (Senn 2013), included in R package netmeta
data(Senn2013)
# Look at first 5 lines: data are in contrast-based format
head(Senn2013, 5)

##      TE   seTE treat1 treat2          studlab
## 1 -1.90 0.1414   metf   plac     DeFronzo1995
## 2 -0.82 0.0992   metf   plac        Lewin2007
## 3 -0.20 0.3579   metf   acar       Willms1999
## 4 -1.34 0.1435   rosi   plac     Davidson2007
## 5 -1.10 0.1141   rosi   plac Wolffenbuttel1999
```

```
# Network meta-analysis of diabetes data
net1 <- netmeta(TE, seTE, treat1, treat2, studlab, data = Senn2013, sm = "MD",
        comb.fixed=FALSE, comb.random=TRUE)
```

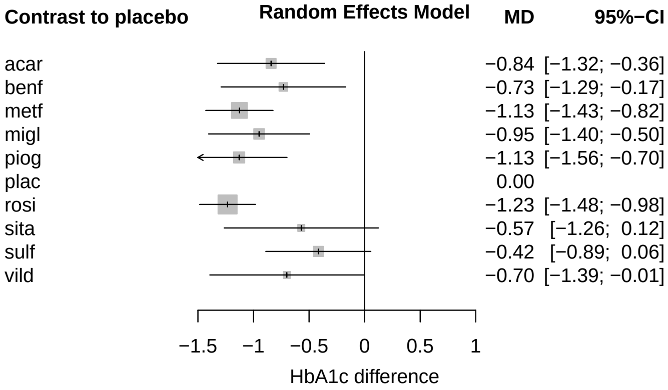
Summary output of diabetes data

```
# Summarize results
summary(net1)

## Number of studies: k=26
## Number of treatments: n=10
## Number of pairwise comparisons: m=28
##
## Random effects model
##
## Treatment estimate (sm='MD'):
##         acar    benf    metf    migl    piog    plac    rosi    sita
## acar       . -0.1106  0.2850  0.1079  0.2873 -0.8418  0.3917 -0.2718
## benf  0.1106       .  0.3956  0.2186  0.3979 -0.7311  0.5023 -0.1611
## metf -0.2850 -0.3956       . -0.1770  0.0023 -1.1268  0.1067 -0.5568
## migl -0.1079 -0.2186  0.1770       .  0.1794 -0.9497  0.2837 -0.3797
*** Output truncated ***
##
## Quantifying heterogeneity/inconsistency:
## tau^2 = 0.1087; I^2 = 81.4%
##
## Test of heterogeneity/inconsistency:
##      Q d.f.  p-value
##  96.99   18 < 0.0001
```

Forest plot of diabetes data

```r
# Look at result
forest(net1, ref = "plac",
       pooled = "random", digits=2,
       smlab = "Random effects model",
       xlab = "HbA1c difference",
       leftlabs = "Contrast to placebo")
```



Smoking cessation data

```
# Load diabetes data (Senn 2013)
data(smokingcessation)
```

```
# Look at first lines: data are in arm-based format
head(smokingcessation)

##   event1  n1 event2   n2 event3  n3 treat1 treat2 treat3
## 1      9 140     23  140     10 138      A      C      D
## 2     11  78     12   85     29 170      B      C      D
## 3     75 731    363  714     NA  NA      A      C
## 4      2 106      9  205     NA  NA      A      C
## 5     58 549    237 1561     NA  NA      A      C
## 6      0  33      9   48     NA  NA      A      C

# The first two trials are three-arm trials
```

Smoking cessation data

```
# Transform data from arm-based format to contrast-based format
p2 <- pairwise(treat = list(treat1, treat2, treat3),
               event = list(event1, event2, event3),
               n = list(n1, n2, n3),
               data = smokingcessation, sm = "OR")
head(p2, 9)

##             TE      seTE studlab treat1 treat2 event1  n1 event2   n2
## 1 -1.051293027 0.4132432       1      A      C      9 140     23  140
## 2 -0.128527575 0.4759803       1      A      D      9 140     10  138
## 3  0.922765452 0.3997972       1      C      D     23 140     10  138
## 4 -0.001244555 0.4504070       2      B      C     11  78     12   85
## 5 -0.225333286 0.3839393       2      B      D     11  78     29  170
## 6 -0.224088731 0.3722995       2      C      D     12  85     29  170
## 7 -2.202289286 0.1430439       3      A      C     75 731    363  714
## 8 -0.870353637 0.7910933       4      A      C      2 106      9  205
## 9 -0.415648522 0.1557329       5      A      C     58 549    237 1561

# Note the two three-arm studies 1 and 2, now each filling three data lines
```

Smoking cessation data

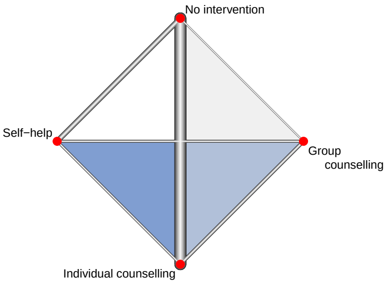
```
net2 <- netmeta(TE, seTE, treat1, treat2, studlab, data = p2,
                comb.fixed = FALSE, comb.random = TRUE)
summary(net2)

## Number of studies: k=24
## Number of treatments: n=4
## Number of pairwise comparisons: m=28
##
## Random effects model
##
## Treatment estimate (sm='OR'):
##        A      B      C      D
## A      . 0.6595 0.4803 0.4056
## B 1.5162      . 0.7282 0.6150
## C 2.0822 1.3732      . 0.8446
## D 2.4653 1.6259 1.1840      .
*** (Output truncated) ***

## Quantifying heterogeneity/inconsistency:
## tau^2 = 0.5989; I^2 = 88.6%
## Test of heterogeneity/inconsistency:
##       Q d.f.  p.value
##  202.62   23 < 0.0001
```

Smoking cessation data

```
# Define treatment names
tname <- c("No intervention","Self-help","Individual counselling","Group
            counselling")
# Produce network graph
# Transparent coloured areas correspond to three-arm studies
netgraph(net2, points=TRUE, cex.points=3, cex=1.25, labels=tname)
```

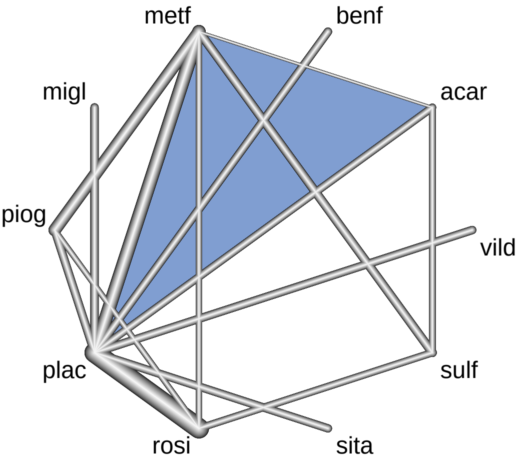


Drawing the network with **netmeta**

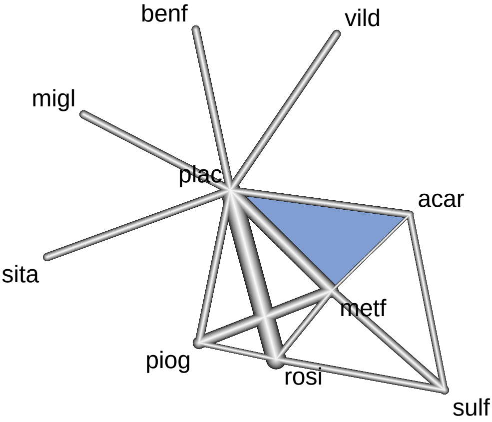
For network visualisation, use function `netgraph`

- Iteration method implemented in **netmeta**: Stress algorithm (Kamada and Kawai, 1989; Hu, 2012, related to multi-dimensional scaling)
- Various starting (also random) layouts available
- Iteration steps visible/printable, if desired
- Variable choice of scale, node size, line width, colours, highlighting
- Coloured polygons may represent multiarm studies (where transparent colours are available)

Drawing the network with netmeta: Diabetes data



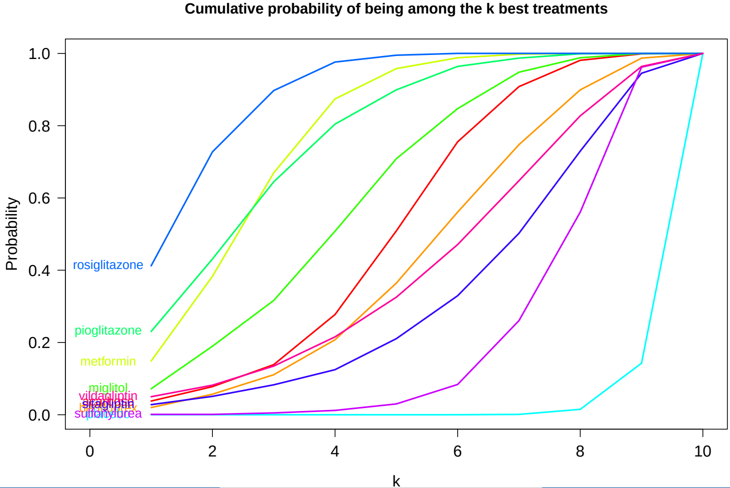
Drawing the network with netmeta: Diabetes data



Ranking treatments

- Bayesian framework:
  Derive ranking probabilities for each treatment from the posterior distributions
- Treatments may be ranked by the surface under the cumulative ranking curve (SUCRA) (Salanti et al., 2011)
- Frequentist framework:
  We introduced a quantity, called P-score, as an analogue to SUCRA (Rücker and Schwarzer, 2015)
- Example: Diabetes data

Surface under the cumulative ranking curve (SUCRA) for diabetes data (produced with WinBUGS and R)



Ranking treatments using P-scores: Diabetes data

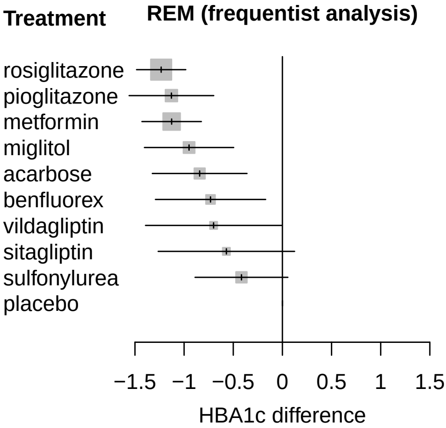
- P-scores allow ranking the treatments on a continuous 0-1 scale
- Based on frequentist point estimates and standard errors
- Frequentist analogue to SUCRA (Rücker and Schwarzer, 2015)

```
# Rank treatments
# Small values are "good" here (this is the default), otherwise "bad"
netrank(net1, small.values = "good")

##      P-score
## rosi  0.8934
## metf  0.7818
## piog  0.7746
## migl  0.6137
## acar  0.5203
## benf  0.4358
## vild  0.4232
## sita  0.3331
## sulf  0.2103
## plac  0.0139
```

Ranking treatments using P-scores: Diabetes data

Compare forest plot, point estimates, SUCRA values and P-scores



	Frequentist	SUCRA	P-score
rosi	-1.23	0.890	0.893
metf	-1.13	0.780	0.782
piog	-1.13	0.773	0.775
migl	-0.95	0.620	0.614
acar	-0.84	0.520	0.520
benf	-0.73	0.439	0.436
vild	-0.70	0.413	0.423
sita	-0.57	0.334	0.333
sulf	-0.42	0.213	0.210
plac	0	0.018	0.014

Inconsistency diagnostics

Designs in network meta-analysis

- A design is each combination of treatments within a study in a network meta-analysis
  - Example: For three treatments A, B, C, the possible designs are $A : B$, $A : C$, $B : C$, $A : B : C$
  - For n treatments the maximum number of designs is $2^n - n - 1$
  - Not all these need be present in a given network meta-analysis
  - In a pairwise meta-analysis, all trials have the same design $A : B$
- Clinical context
  - Example: Studies with design $A : C$ might differ to studies with design $A : B$ or $A : B : C$ in that they include patients who cannot be randomised to B
  - Heterogeneity between designs is plausible

Decomposition of the heterogeneity statistic

Total Q statistic

$$Q_{total} = (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{nma})^\top \mathbf{W} (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{nma})$$

Krahn et al. (2013):

- Q can be decomposed into
  - a part coming from within designs (heterogeneity between studies of the same design)
  - a part coming from between designs (inconsistency between studies of different designs)
- Q can be decomposed into parts coming from each design
- Q can be decomposed into parts coming from each study

Decomposition of Q: Diabetes data

```
# Decompose total Q statistics into parts from designs
decomp.design(net1)

## Q statistics to assess homogeneity / consistency
##
##                      Q df  p.value
## Whole network   96.99 18 < 0.0001
## Within designs  74.46 11 < 0.0001
## Between designs 22.53  7   0.0021
##
## Design-specific decomposition of within-designs Q statistic
##
##          Design     Q df  p.value
##       acar:plac  0.00  0       --
##       acar:sulf  0.00  0       --
##       benf:plac  4.38  1   0.0363
##       metf:piog  0.00  0       --
##       metf:plac 42.16  2 < 0.0001
##       metf:rosi  0.19  1   0.6655
##       metf:sulf  0.00  0       --
*** (Output truncated) ***
##  acar:metf:plac  0.00  0       --
```

Decomposition of Q: Diabetes data

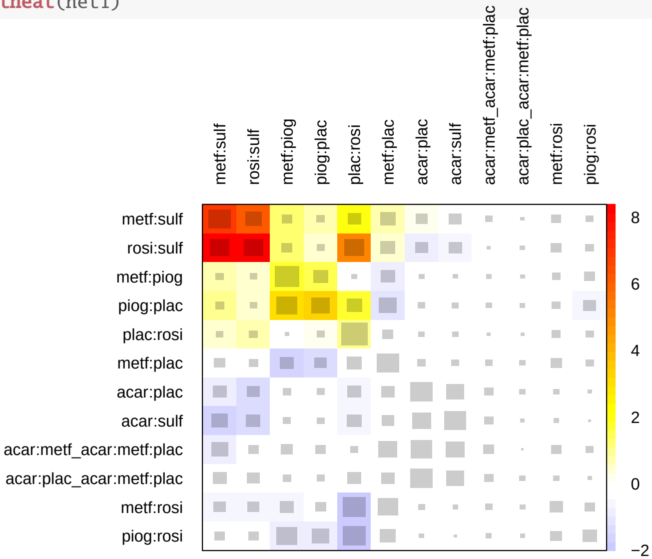
```
# Decompose total Q statistics into parts from designs
decomp.design(net1)

## Between-designs Q statistic after detaching of single designs
##
##  Detached design      Q  df  p.value
##        acar:plac 22.44  6   0.001
##        acar:sulf 22.52  6   0.001
##        metf:piog 17.13  6   0.0088
##        metf:plac 22.07  6   0.0012
##        metf:rosi 22.52  6   0.001
##        metf:sulf  7.51  6   0.276    ***
##        piog:plac 17.25  6   0.0084
##        piog:rosi 22.48  6   0.001
##        plac:rosi 16.29  6   0.0123
##        rosi:sulf  6.77  6   0.3425   ***
##   acar:metf:plac 22.38  5   0.0004
```

Explanation: Detaching a design means relaxing the consistency assumption for this design. If Q decreases markedly after detaching a design (*** added for the purpose of this talk), we conclude that this design contributed to between-design inconsistency. If Q does not decrease markedly, the design is not thought to contribute to between-design inconsistency.

Net heat plot (Krahn et al., 2013): Diabetes data

```r
netheat(net1)
```



Net heat plot (Krahn et al., 2013)

- Areas of grey squares ■: indicate the contribution from the treatment comparison in the column to the treatment comparison in the row
- Colours on the diagonal represent the inconsistency contribution of the corresponding design (red means large)
- Colours on the off-diagonal associated with the change in inconsistency between direct and indirect evidence in a network estimate in the row after relaxing the consistency assumption for the effect of one design in the column
  - Blue indicates that the evidence of the design in the column supports the evidence in the row
  - Red indicates that the evidence of the design in the column contrasts to the evidence in the row
- Largest inconsistency contribution by the `metf:sulf` and `rosi:sulf` designs (red squares in top left corner)

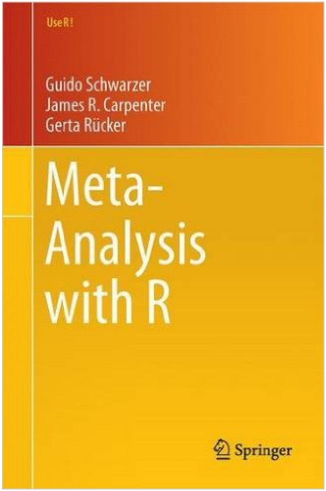
Summary

R package **netmeta** provides

- flexible data entry (`pairwise`)
- fixed / random effects model (`netmeta`)
- appropriate incorporation of multi-arm trials
- forest plots (`forest`)
- network graphs (`netgraph`)
- ranking of treatments (`netrank`)
- inconsistency diagnostics (`decomp.design`, `netheat`)

Currently not available:
Meta-regression

See book Schwarzer et al. (2015)



References

Bailey, R. A. (2007). Designs for two-colour microarray experiments. *Applied Statistics-journal of the Royal Statistical Society Series C*, 56(4):365–394.

Cipriani, A., Furukawa, T. A., Salanti, G., Geddes, J., Higgins, J., Churchill, R., Watanabe, N., Nakagawa, A., Omori, I., McGuire, H., Tansella, M., and Barbui, C. (2009). Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet*, 373(9665):746–758. doi: 10.1016/S0140-6736(09)60046-5.

Dias, S., Welton, N. J., Sutton, A. J., Caldwell, D. M., Lu, G., and Ades, A. E. (2013). Evidence synthesis for decision making 4: Inconsistency in networks of evidence based on randomized controlled trials. *Medical Decision Making*, 33:641–656. doi:10.1177/0272989X12455847.

Higgins, J. P. T., 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(2):98–110.

Hu, Y. (2012). Algorithms for visualizing large networks. In Naumann, U. and Schenk, O., editors, *Combinatorial Scientific Computing*, pages 525–549. Chapman and Hall/CRC Computational Science, Boca Raton, London, New York. ISBN 9781439827352.

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(29):3805–3820.

Kamada, T. and Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1):7–15.

König, J., Krahn, U., and Binder, H. (2013). Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. *Statistics in Medicine*, 32(30):5414–5429. doi: 10.1002/sim.6001.

Krahn, U., Binder, H., and König, J. (2013). A graphical tool for locating inconsistency in network meta-analyses. *BMC Medical Research Methodology*, 13(1):35.

Lu, G., Welton, N. J., Higgins, J. P. T., White, I. R., and Ades, A. E. (2011). Linear inference for mixed treatment comparison meta-analysis: A two-stage approach. *Research Synthesis Methods*, 2(1):43–60.

Rücker, G. (2012). Network meta-analysis, electrical networks and graph theory. *Research Synthesis Methods*, 3(4):312–324.

Rücker, G. and Schwarzer, G. (2014). Reduce dimension or reduce weights? Comparing two approaches to multi-arm studies in network meta-analysis. *Statistics in Medicine*, 33:4353–4369. DOI: 10.1002/sim.6236.

Rücker, G. and Schwarzer, G. (2015). Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Medical Research Methodology*, 15(1):58. doi: 10.1186/s12874-015-0060-8.

Rücker, G., Schwarzer, G., Krahn, U., and König, J. (2016). netmeta: Network meta-analysis using frequentist methods. R package version 0.9-1.

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(2):163–171. doi: 10.1016/j.jclinepi.2010.03.016.

Schwarzer, G., Carpenter, J. R., and Rücker, G. (2015). *Meta-Analysis with R*. Use R! Springer International Publishing, Switzerland.

Senn, S., Gavini, F., Magrez, D., and Scheen, A. (2013). Issues in performing a network meta-analysis. *Statistical Methods in Medical Research*, 22(2):169–189. Epub 2012 Jan 3.

White, I. R., Barrett, J. K., Jackson, D., and Higgins, J. P. T. (2012). Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Research Synthesis Methods*, 3(2):111–125.

Appendix: A proof that SUCRA and P-score are the same

We assume the true probabilities as known. If $\mathrm{R}(i) = k$ means that treatment i has rank k, we have

$$P_{ij} = \sum_{k=1}^{n-1} \sum_{l=k+1}^{n} P(\mathrm{R}(i) = k \wedge \mathrm{R}(j) = l)$$

and

$$(n-1)SUCRA(i) = \sum_{r=1}^{n-1} F(i,r) = \sum_{r=1}^{n-1} \sum_{k=1}^{r} P(i,k) = \sum_{k=1}^{n-1} \sum_{r=k}^{n-1} P(i,k) = \sum_{k=1}^{n-1} (n-k) P(i,k)$$

It follows

$$\sum_{j=1}^{n} P_{ij} = \sum_{j=1}^{n} \sum_{k=1}^{n-1} \sum_{l=k+1}^{n} P(\mathrm{R}(i) = k \wedge \mathrm{R}(j) = l) = \sum_{k=1}^{n-1} \sum_{l=k+1}^{n} \sum_{j=1}^{n} P(\mathrm{R}(i) = k \wedge \mathrm{R}(j) = l)$$

$$= \sum_{k=1}^{n-1} \sum_{l=k+1}^{n} P(i,k) = \sum_{k=1}^{n-1} (n-k) P(i,k) = (n-1) SUCRA(i)$$

and thus

$$\bar{P}_i = \frac{1}{n-1} \sum_{j=1}^{n} P_{ij} = SUCRA(i)$$

which is what we wanted to prove. Note: For $n > 2$, neither ranking probabilities $P(i,k)$ nor probabilities P_{ij} can be uniquely determined from $\bar{P}_i$ or $SUCRA(i)$.