Evidence Synthesis / Meta-analysis

Session 4, Lecture 7: Network Meta-Analysis

Guido Knapp[1], Gerta Rücker[2], Guido Schwarzer[2]

[1] Department of Statistics, TU Dortmund University
[2] Center for Medical Biometry and Medical Informatics, University of Freiburg

ruecker@imbi.uni-freiburg.de

DAGStat 2016, Göttingen, 14 March 2016

Overview Lecture 7

Multivariate meta-analysis

- Univariate meta-analysis
  - Consider each outcome separately
  - Example: efficacy, safety, quality of life ⇒ at least three meta-analyses
- Multivariate meta-analysis
  - Outcomes are correlated (e.g., a more efficacious treatment may yield more adverse events)
  - Consider outcomes jointly
  - Problem: within-study correlations often unknown

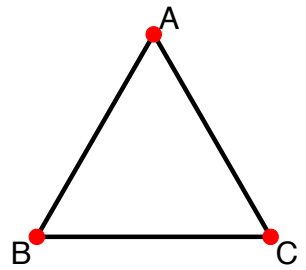
Multivariate meta-analysis

- Special case 1: Meta-analysis of diagnostic accuracy studies (not covered in this tutorial)
  - Outcomes sensitivity and specificity ⇒ Bivariate model (Reitsma et al., 2005; Chu and Cole, 2006)
  - Sensitivity and specificity not correlated within study, as coming from independent groups (diseased, non-diseased)
  - Correlation across studies often (not always!) negative because of threshold effect
- Special case 2: Network meta-analysis
  - Outcomes correspond to different pairwise comparisons of treatments
  - If only two-arm studies: diagonal variance-covariance matrix
  - In the presence of multi-arm studies: correlation structure can be modelled
- Schwarzer et al. (2015)
  - Ch. 7: Multivariate meta-analysis
  - Ch. 8: Network meta-analysis
  - Ch. 9: Meta-analysis of diagnostic accuracy studies

Pairwise meta-analysis and network meta-analysis

- Pairwise meta-analysis
  - Consider two treatments *A* and *B* for a given condition
  - Include studies comparing *A* and *B* with respect to one or more outcomes of interest
  - Based on a suitable effect measure, compare *A* and *B* at the study level
  - Pool treatment effect estimates across studies
- Network meta-analysis
  - Consider a number of treatments $A, B, C, \ldots$ for a given condition
  - Include studies comparing two or more of these treatments with respect to the outcomes of interest
  - Look at contrasts *A* vs *B*, *A* vs *C*, etc. at the study level
  - Summarise the evidence–**but how?**

Network meta-analysis



- Network meta-analysis
  - Key tool for evidence-based medicine (Nietert et al., 2013)
  - Active research area (Salanti et al., 2008; Salanti, 2012; Lee, 2014; Achana et al., 2013)
- Other names
  - Multiple treatment comparison
  - Mixed treatment comparison
    - 'Mixed' means combining direct and indirect evidence

Network meta-analysis

- Questions leading to network meta-analysis
  - ? "I want to compare *A* with *B*, but there are no studies comparing *A* directly to B. There are only studies comparing *A* to *C* and others comparing *B* to *C*"
  - ? "How do existing treatments $A, B, C, \ldots$ for patient with a given condition compare to each other?"
  - ? "Which is the best treatment for my patient?"
- Challenges: Similar to those known from pairwise meta-analysis
  - Clinical heterogeneity between patient populations in different studies
  - Inconsistency between direct and indirect evidence
  - Limited power for assessing heterogeneity
  - Reporting issues (Mills et al., 2012; Hutton et al., 2014)
  - Grading the evidence (Puhan et al., 2014)
- Methods and overviews of methods
  - Salanti et al. (2008); Salanti (2012); Bafeta et al. (2013); Donegan et al. (2013); Efthimiou et al. (2015)

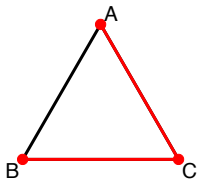
Network meta-analysis: Methods and software

- Bayesian approach [1]
  - WinBUGS (Dias et al., 2013c,a,b,d)
  - GeMTC/BUGS/JAGS (van Valkenhoef et al., 2012)
  - Excel/WinBUGS (Brown et al., 2014)
- Frequentist approach [2]
  - Stata: mvmeta (White, 2011)
  - SAS (Jones et al., 2011; Piepho, 2014)
  - R (Viechtbauer, 2015; Rücker et al., 2015)
  - Overview to R packages (Neupane et al., 2014)
  - In this talk we use R package **netmeta** (Rücker et al., 2015)

[1] Lu and Ades (2006); Dias et al. (2010)

[2] Lu et al. (2011); White et al. (2012); Senn et al. (2013); Rücker (2012); Krahn et al. (2013); Jackson et al. (2014)

Example with 3 treatments



- Given three treatments A, B, C
  - Each study having arms A and C provides a *direct* estimate $\hat{\theta}_{AC}^{direct}$
  - Each study having arms B and C provides a *direct* estimate $\hat{\theta}_{BC}^{direct}$.
- Obtain an *indirect* estimate for comparison $A - B$ from the treatment difference $A - C$ and $B - C$:

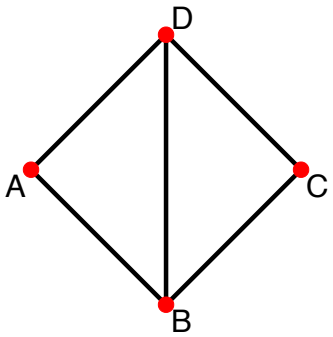
$$\hat{\theta}_{AB}^{indirect} = \hat{\theta}_{AC}^{direct} - \hat{\theta}_{BC}^{direct}$$

- with variance

$$\widehat{\text{Var}}(\hat{\theta}_{AB}^{indirect}) = \widehat{\text{Var}}(\hat{\theta}_{AC}^{direct}) + \widehat{\text{Var}}(\hat{\theta}_{BC}^{direct})$$

More complex example with 4 treatments

Network with four treatments A, B, C, and D

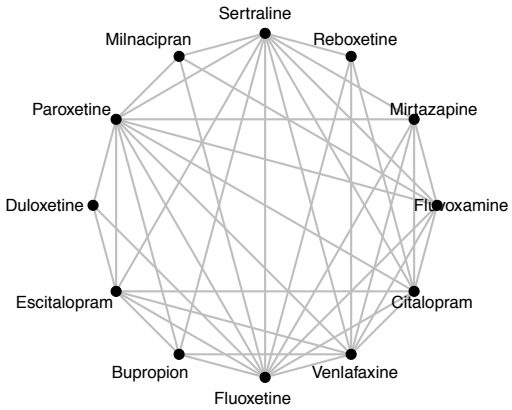


Direct evidence is available for all comparisons except $A - C$

- Comparison $A - C$
  - indirect information via $A - B - C$ and $A - D - C$
  - Different ways may lead to different estimates
- Comparison $A - D$
  - There is direct information $A - D$
  - and indirect information via $A - B - D$
  - and indirect information via $A - B - C - D$

Very complex example with 12 treatments

Cipriani et al. (2009)



Network meta-analysis

- Aim
  - To combine direct and indirect evidence to get the most precise estimates of the treatment differences and associated standard errors
- Assumptions
  - The studies are independent
  - The underlying effects are consistent. This statistical property follows from exchangeability (Dias et al., 2013a), also termed transitivity (Salanti, 2012; Veroniki et al., 2013) or similarity (Donegan et al., 2013)
- Consistency
  - The sum of direct treatment effects over all closed circuits in the graph is zero
  - Equivalent: The indirect evidence for the difference between any two treatments does not differ from the direct evidence
  - Example three treatments A, B, C:

$$\theta_{AB}^{direct} = \theta_{AB}^{indirect} = \theta_{AC}^{direct} - \theta_{BC}^{direct}$$

Estimation under consistency assumption

- Given
  - n treatments, denoted by a vector $\boldsymbol{\theta}^{treat}$ of length n
  - Data from m pairwise comparisons, denoted by $\hat{\boldsymbol{\theta}}$, with associated standard errors $\mathbf{s} = (s_1, s_2, \ldots, s_m)$
  - If there are K two-arm trials, $\hat{\boldsymbol{\theta}}$ is a vector of length K
  - In general, if there are also multi-arm trials, $\hat{\boldsymbol{\theta}}$ has length $m \geq K$ with m denoting the total number of pairwise comparisons
- Model

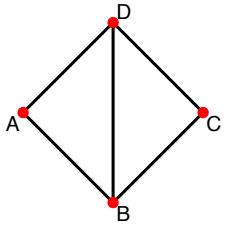
$$\hat{\boldsymbol{\theta}} = \mathbf{X}\boldsymbol{\theta}^{treat} + \epsilon, \qquad \epsilon \sim N(\mathbf{0}, \boldsymbol{\Sigma}),$$

  where $\boldsymbol{\Sigma}$ is a diagonal matrix whose i^{th} entry is s_i^2.
- Network structure defined by the $m \times n$ design matrix $\mathbf{X}$

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

- $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_m)^T$ vector of observed treatment differences
- Design matrix $\mathbf{X}$ (dimension $m \times n$) defines the network structure
- $\mathbf{W}$ diagonal matrix (dimension $m \times m$) containing the 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 $\mathbf{H}$.

Estimation under the fixed effect model

- Standard errors for the network estimates calculated from the variance-covariance matrix of

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

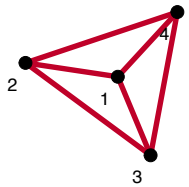
- Heterogeneity/inconsistency measured by generalised Q_{total} statistic

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

(Jackson et al., 2012; Rücker, 2012)

Multi-arm studies: Need to account for correlation

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



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)

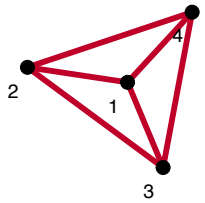
- Based on standard regression methodology
- For each multi-arm study, choose a study-specific reference treatment
- Consider only comparisons to the reference treatment ('basic parameters')

Alternative approach: Reduce weights
(Rücker, 2012; Rücker and Schwarzer, 2014)

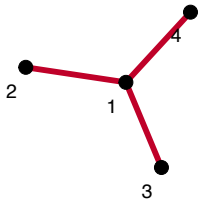
- Implemented in R package **netmeta**
- Based on electrical network methodology
- For each multi-arm study, adjust the standard errors appropriately

Comparison of the approaches

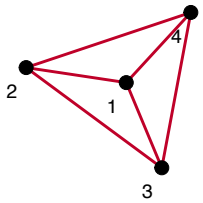
Given a four-arm study with six comparisons,



we may cut off three of six comparisons:



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



Data sets

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., 2013d)

Both data sets are part of R package **netmeta**

How to use R package netmeta: Diabetes data

```
# Make R package netmeta available
library(netmeta)

 ## Loading required package:  meta
 ## Loading 'meta' package (version 4.3-0).
 ## Loading 'netmeta' package (version 0.8-0).
```

```
# 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
```


How to use R package netmeta: Diabetes data

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

```
# Look at result
print(net1, digits=2)

## Original data (with adjusted standard errors for multi-arm studies):
##
##                   treat1 treat2    TE seTE seTE.adj narms multiarm
## DeFronzo1995        metf   plac -1.90 0.14     0.14     2
## Lewin2007           metf   plac -0.82 0.10     0.10     2
## Willms1999          acar   metf  0.20 0.36     0.39     3        *
## Davidson2007        plac   rosi  1.34 0.14     0.14     2
## Wolffenbuttel1999   plac   rosi  1.10 0.11     0.11     2
## Kipnes2001          piog   plac -1.30 0.13     0.13     2
*** (Output truncated) ***
## Willms1999          metf   plac -1.20 0.38     0.41     3        *
## Willms1999          acar   plac -1.00 0.47     0.82     3        *
```


How to use R package netmeta: Diabetes data

```
# Look at result, continued

## Number of treatment arms (by study):
##                   narms
## Alex1998              2
## Baksi2004             2
## Costa1997             2
*** (Output truncated) ***

## Willms1999            3
*** (Output truncated) ***

## Results (fixed effect model):
##
##                   treat1 treat2    MD          95%-CI     Q leverage
## DeFronzo1995        metf   plac -1.11  [-1.23; -1.00] 30.89     0.18
## Lewin2007           metf   plac -1.11  [-1.23; -1.00]  8.79     0.36
## Willms1999          acar   metf  0.29  [ 0.06;  0.51]  0.05     0.09
## Davidson2007        plac   rosi  1.20  [ 1.11;  1.30]  0.93     0.11
*** (Output truncated) ***
## Willms1999          metf   plac -1.11  [-1.23; -1.00]  0.04     0.02
## Willms1999          acar   plac -0.83  [-1.04; -0.61]  0.04     0.02
```


How to use R package netmeta: Diabetes data

```
# Look at result, continued

## Number of studies: k=26
## Number of treatments: n=10
## Number of pairwise comparisons: m=28
##
## Fixed effect model
##
## Treatment estimate (sm='MD'):
##       acar  benf  metf  migl  piog  plac  rosi  sita  sulf  vild
## acar     .  0.08  0.29  0.12  0.24 -0.83  0.37 -0.26 -0.39 -0.13
## benf -0.08     .  0.21  0.04  0.16 -0.91  0.30 -0.34 -0.47 -0.21
## metf -0.29 -0.21     . -0.17 -0.05 -1.11  0.09 -0.54 -0.67 -0.41
## migl -0.12 -0.04  0.17     .  0.12 -0.94  0.26 -0.37 -0.50 -0.24
## piog -0.24 -0.16  0.05 -0.12     . -1.07  0.14 -0.50 -0.63 -0.37
## plac  0.83  0.91  1.11  0.94  1.07     .  1.20  0.57  0.44  0.70
## rosi -0.37 -0.30 -0.09 -0.26 -0.14 -1.20     . -0.63 -0.76 -0.50
## sita  0.26  0.34  0.54  0.37  0.50 -0.57  0.63     . -0.13  0.13
## sulf  0.39  0.47  0.67  0.50  0.63 -0.44  0.76  0.13     .  0.26
## vild  0.13  0.21  0.41  0.24  0.37 -0.70  0.50 -0.13 -0.26     .
```

How to use R package netmeta: Diabetes data

```
# Look at result, continued

## Lower 95%-confidence limit:
##       acar  benf  metf  migl  piog  plac  rosi  sita  sulf  vild
## acar     . -0.25  0.06 -0.21 -0.01 -1.04  0.15 -0.59 -0.61 -0.46
## benf -0.41     . -0.07 -0.31 -0.13 -1.15  0.03 -0.69 -0.77 -0.56
## metf -0.51 -0.48     . -0.44 -0.18 -1.23 -0.04 -0.82 -0.85 -0.69
## migl -0.44 -0.39 -0.10     . -0.17 -1.19 -0.01 -0.73 -0.81 -0.60
## piog -0.49 -0.45 -0.09 -0.41     . -1.22 -0.02 -0.79 -0.84 -0.66
## plac  0.61  0.66  1.00  0.70  0.92     .  1.11  0.32  0.26  0.45
## rosi -0.60 -0.56 -0.22 -0.52 -0.30 -1.30     . -0.90 -0.94 -0.77
## sita -0.07 -0.02  0.27  0.02  0.20 -0.82  0.36     . -0.44 -0.23
## sulf  0.17  0.16  0.50  0.20  0.42 -0.62  0.58 -0.18     . -0.05
## vild -0.20 -0.15  0.14 -0.11  0.08 -0.95  0.24 -0.49 -0.57     .
##
## Upper 95%-confidence limit:
##       acar  benf  metf  migl  piog  plac  rosi  sita  sulf  vild
## acar     .  0.41  0.51  0.44  0.49 -0.61  0.60  0.07 -0.17  0.20
## benf  0.25     .  0.48  0.39  0.45 -0.66  0.56  0.02 -0.16  0.15
## metf -0.06  0.07     .  0.10  0.09 -1.00  0.22 -0.27 -0.50 -0.14
## migl  0.21  0.31  0.44     .  0.41 -0.70  0.52 -0.02 -0.20  0.11
## piog  0.01  0.13  0.18  0.17     . -0.92  0.30 -0.20 -0.42 -0.08
## plac  1.04  1.15  1.23  1.19  1.22     .  1.30  0.82  0.62  0.95
```

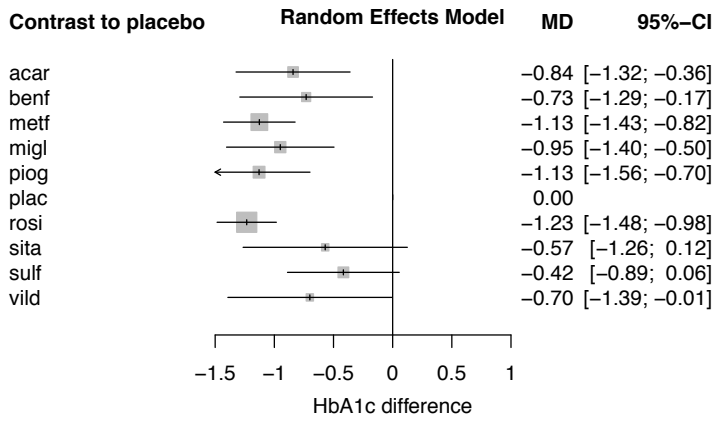

How to use R package netmeta: Diabetes data

```
# Look at result, continued and last part

## 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
```

```
# Create forest plot (now use random effects model!)
forest(net1, ref = "plac", pooled = "random", digits=2,
       smlab = "Random effects model",
       xlab = "HbA1c difference",
       leftlabs = "Contrast to placebo")
```


Forest plot of diabetes data



Summary output of diabetes data

```
# Summary of comparison to reference treatment placebo, random effects model only
summary(net1, ref = "plac", comb.fixed=FALSE, comb.random=TRUE)

## Number of studies: k=26
## Number of treatments: n=10
## Number of pairwise comparisons: m=28
##
## Random effects model
##
## Treatment estimate (sm='MD', reference.group='plac'):
##           MD              95%-CI
## acar -0.8418  [-1.3236; -0.3600]
## benf -0.7311  [-1.2918; -0.1705]
## metf -1.1268  [-1.4291; -0.8244]
## migl -0.9497  [-1.4040; -0.4955]
## piog -1.1291  [-1.5596; -0.6986]
## plac       .                  .
## rosi -1.2335  [-1.4839; -0.9830]
## sita -0.5700  [-1.2640;  0.1240]
## sulf -0.4166  [-0.8887;  0.0556]
## vild -0.7000  [-1.3927; -0.0073]
##
*** (Output truncated) ***
```

Smoking cessation data

```
# Load smoking cessation data (Dias 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
```

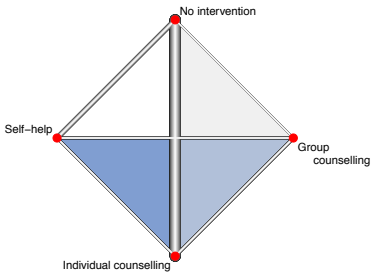

Drawing the network with **netmeta**

For network visualisation, use function `netgraph`

- Various starting (also random) layouts available, with or without iteration
- Iteration steps visible/printable, if desired
- Variable choice of scale, point size, line width, colours, etc.
- Highlighting of single comparisons available
- Coloured polygons may represent multiarm studies (transparent colours available)

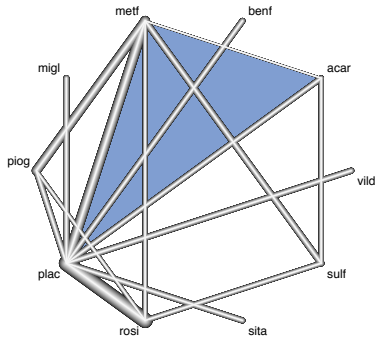
Smoking cessation data

```
# 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)
```



Diabetes data

```
# Draw network
# The blue triangle represents the three-arm trial Willms 1999
netgraph(net1)
```



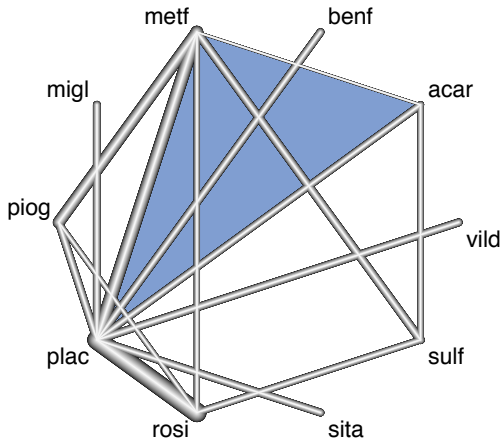
Drawing the network with netmeta: Diabetes data

```
# Draw network
# The blue triangle represents the three-arm trial Willms 1999
netgraph(net1)
```

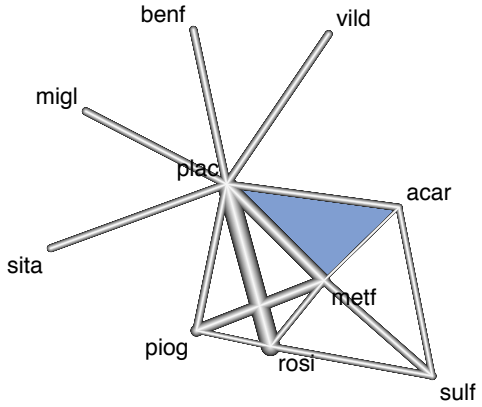
```
netgraph(net1, plastic = TRUE, iterate = TRUE, thickness = "se.fixed")
```

```
netgraph(net1, plastic = TRUE, iterate = TRUE, thickness = "se.fixed",
              allfigures = TRUE)
```


Drawing the network with netmeta: Diabetes data



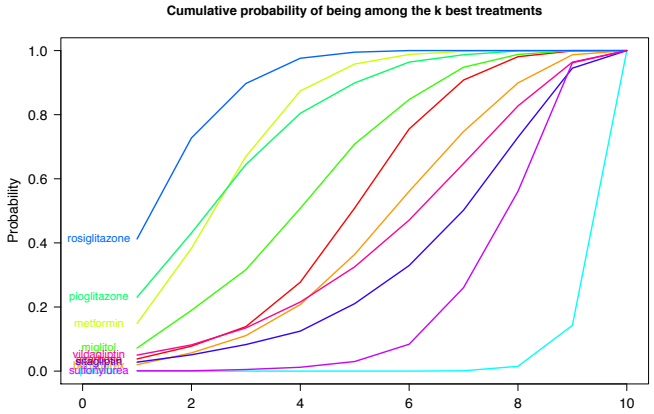
Drawing the network with netmeta: Diabetes data



Ranking treatments

Bayesian framework:

Treatments may be ranked by the surface under the cumulative ranking curve (SUCRA) (Salanti et al., 2011)



P-scores

Frequentist framework:

We introduced a quantity, called P-score, as an analogue to SUCRA (Rücker and Schwarzer, 2015b)

- P-scores allow ranking the treatments on a continuous 0-1 scale
- Derived from the p-values of all pairwise comparisons
- Obtained without resampling methods
- P-score $\bar{P}_i$ represents the continuous rank of treatment i within the given range of treatments (1 = best, 0 = worst)
- $\bar{P}_i$ is the expected proportion of treatments worse than treatment i
- Proposition: If probabilities are assumed as known, P-scores and SUCRA values are equal

Ranking treatments using P-scores: Diabetes data

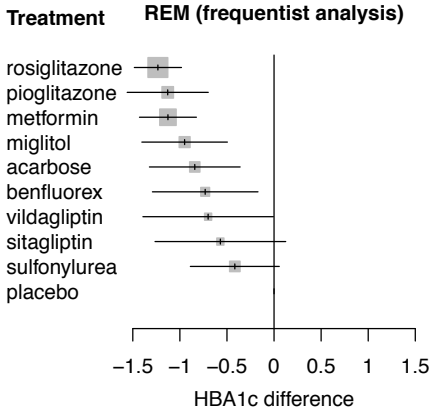
```
# Network meta-analysis of diabetes data
net1 <- netmeta(TE, seTE, treat1, treat2, studlab, data = Senn2013, sm = "MD")
```

```
# 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

Ranking treatments using P-scores

- P-scores allow ranking the treatments on a continuous 0-1 scale
- Based on frequentist point estimates and standard errors
- Frequentist analogue and numerically similar to SUCRA values
- Order of both rankings depends largely on point estimates
- Do not rank the P-scores (or SUCRA values) again
  Pay attention to their numerical values
- Never use ranking alone
  Always present confidence/credible intervals

Designs in network meta-analysis

- A design is each existing 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

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

- Measures the total squared deviation of the network meta-analysis estimates to the observed effects
- Measures total heterogeneity/inconsistency
- 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)
- Can be decomposed into parts coming from each design
- Can be decomposed into parts coming from each study

(Krahn et al., 2013)

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.

Decomposition of *Q*: Diabetes data

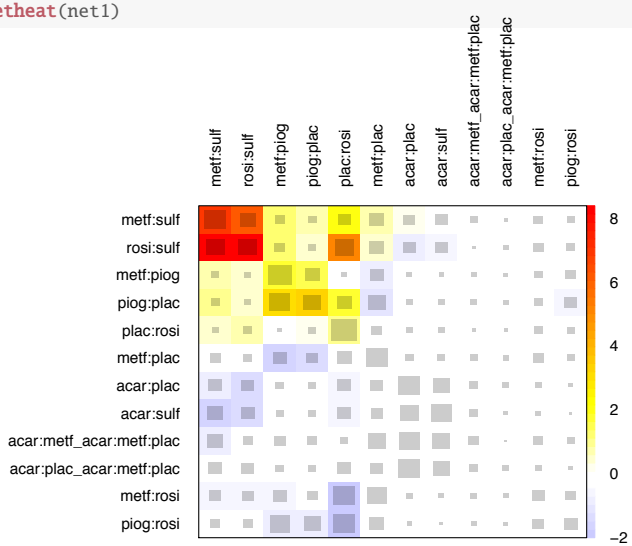
```
# Decompose total Q statistics into parts from designs
decomp.design(net1)

## Q statistic to assess consistency under the assumption of
## a full design-by-treatment interaction random effects model
##
##                    Q df  p.value tau.within tau2.within
## Between designs 2.19  7   0.9483     0.3797      0.1442
```

Explanation: If all designs are allowed to provide different treatment effects, residual inconsistency should be reduced. This is the case here.

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

```
netheat(net1)
```



Net heat plot (Krahn et al., 2013)

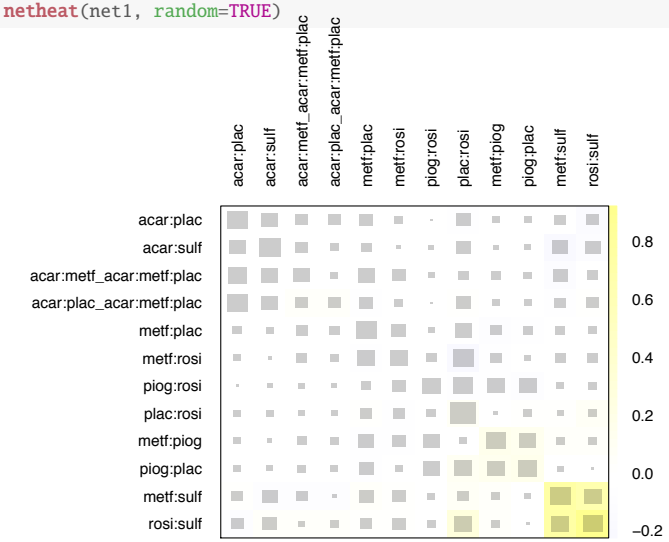
- Areas of grey squares ■: indicate the contribution from the treatment comparison in the column to the treatment comparison in the row
- Colours on the diagonal represent the inconsistency contribution of the corresponding design (red means large)
- Colours on the off-diagonal associated with the change 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)

Interpretation

- Heterogeneity in the network cannot be traced to one design
- Single largest reduction in the whole network inconsistency is achieved by removing the `rosi:sulf` design
- Random effects model more appropriate
- Corresponding net heat plot shows a marked reduction in inconsistency

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

```
netheat(net1, random=TRUE)
```



Summary

- Network meta-analysis is a tool for using all the evidence in a particular area to estimate and compare treatment effects
- R package **netmeta** provides methods for network meta-analysis in a frequentist framework
  - fixed effect model and random effects model (`netmeta`) with 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

References

Achana, F., Hubbard, S., Sutton, A., Kendrick, D., and Cooper, N. (2013). An exploration of synthesis methods in public health evaluations of interventions concludes that the use of modern statistical methods would be beneficial. *Journal of Clinical Epidemiology*. doi: 10.1016/j.jclinepi.2013.09.018.

Bafeta, A., Trinquart, L., Seror, R., and Ravaud, P. (2013). Analysis of the systematic reviews process in reports of network meta-analysis: methodological systematic review. *BMJ*, 347:f3675.

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

Brown, S., Hutton, B., Clifford, T., Coyle, D., Grima, D., Wells, G., and Cameron, C. (2014). A Microsoft-Excel-based tool for running and critically appraising network meta-analyses–an overview and application of NetMetaXL. *Systematic Reviews*, 3:110. doi: 10.1186/2046-4053-3-110.

Chu, H. and Cole, S. R. (2006). Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed approach. *Journal of Clinical Epidemiology*, 59:1331–1333.

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., Sutton, A. J., Ades, A. E., and Welton, N. J. (2013a). Evidence synthesis for decision making 2: A generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Medical Decision Making*, 33:607–617. doi:10.1177/0272989X12458724.

Dias, S., Sutton, A. J., Welton, N. J., and Ades, A. E. (2013b). Evidence synthesis for decision making 3: Heterogeneity — subgroups, meta-regression, bias, and bias-adjustment. *Medical Decision Making*, 33:618–640.

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(7-8):932–944.

Dias, S., Welton, N. J., Sutton, A. J., and Ades, A. E. (2013c). Evidence synthesis for decision making 1: Introduction. *Medical Decision Making*, 33:597–606. doi:10.1177/0272989X13487604.

Dias, S., Welton, N. J., Sutton, A. J., Caldwell, D. M., Lu, G., and Ades, A. E. (2013d). 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.

Donegan, S., Williamson, P., D'Alessandro, U., and Smith, C. T. (2013). Assessing key assumptions of network meta-analysis: a review of methods. *Research Synthesis Methods*, 4:291–323.

Efthimiou, O., Debray, T. P. A., van Valkenhoef, G., Trelle, S., Panayidou, K., Moons, K. G. M., Reitsma, J. B., Shangg, A., Salanti, G., and GetReal Methods Review Group

(2015). Getreal in network meta-analysis: a review of the methodology. *Research Synthesis Methods*. doi: 10.1002/jrsm.1195.

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.

Hutton, B., Salanti, G., Chaimani, A., Caldwell, D. M., Schmid, C., Thorlund, K., Mills, E., Catalá-López, F., Turner, L., Altman, D. G., and Moher, D. (2014). The quality of reporting methods and results in network meta-analyses: an overview of reviews and suggestions for improvement. *PLoS One. 2014*, 9(3):e92508. doi: 10.1371/journal.pone.0092508. eCollection 2014.

Jackson, D., Barrett, J. K., Rice, S., White, I. R., and Higgins, J. P. (2014). A design-by-treatment interaction model for network meta-analysis with random inconsistency effects. *Statistics in Medicine*, 33(21):3639–3654. doi: 10.1002/sim.6188. Published online Apr 29, 2014.

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.

Jones, B., Roger, J., Lane, P. W., Lawton, A., Fletcher, C., Cappelleri, J. C., Tate, H., Moneuse, P., and PSI Health Technology Special Interest Group, Evidence Synthesis

sub-team (2011). Statistical approaches for conducting network meta-analysis in drug development. *Pharmaceutical Statistics*, 10(6):523–531. doi: 10.1002/pst.533.

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.

Krahn, U., Binder, H., and König, J. (2014). Visualizing inconsistency in network meta-analysis by independent path decomposition. *BMC Medical Research Methodology*, 14(1):131.

Lee, A. W. (2014). Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. *Journal of Clinical Epidemiology*, 67(2):138–143. Published online 3 October 2013.

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.

Lu, G. B. and Ades, A. E. (2006). Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association*, 101(474):447–459.

Mills, E. J., Ioannidis, J. P. A., Thorlund, K., Schünemann, H. J., Puhan, M. A., and Guyatt, G. H. (2012). How to use an article reporting a multiple treatment comparison meta-analysis. *JAMA-Journal of the American Medical Association*, 308(12):1246–1253. doi:10.1001/2012.jama.11228.

Neupane, B., Richer, D., Bonner, A., Kibret, T., and Beyene, J. (2014). Network meta-analysis using R: a review of currently available automated packages. *PLoS ONE*, 9(12):e115065. doi: 10.1371/journal.pone.0115065.

Nietert, P. J., Wahlquist, A. E., and Herbert, T. L. (2013). Characteristics of recent biostatistical methods adopted by researchers publishing in general/internal medicine journals. *Statistics in Medicine*, 32(1–10).

Piepho, H. P. (2014). Network-meta analysis made easy: detection of inconsistency using factorial analysis-of-variance models. *BMC Medical Research Methodology*, 14:61. doi: 10.1186/1471-2288-14-61.

Puhan, M. A., Schünemann, H. J., Murad, M. H., Li, T., Brignardello-Petersen, R., Singh, J. A., Kessels, A. G., Guyatt, G. H., and GRADE Working Group (2014). A GRADE working group approach for rating the quality of treatment effect estimates from network meta-analysis. *British Medical Journal*, 349:g5630. doi: 10.1136/bmj.g5630.

Reitsma, J., Glas, A., Rutjes, A., Scholten, R., Bossuyt, P., and Zwinderman, A. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*, 58(10):982–990.

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. (2015a). Automated drawing of network plots in meta-analysis. *Research Synthesis Methods*.

Rücker, G. and Schwarzer, G. (2015b). 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. (2015). netmeta: Network meta-analysis using frequentist methods. R package version 0.8-0.

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

Salanti, G., Higgins, J. P., Ades, A. E., and Ioannidis, J. P. (2008). Evaluation of networks of randomized trials. *Statistical Methods in Medical Research*, 17(3):279–301.

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.

van Valkenhoef, G., Lu, G., de Brock, B., Hillege, H., Ades, A. E., and Welton, N. J. (2012). Automating network meta-analysis. *Research Synthesis Methods*, 3(4):285–299.

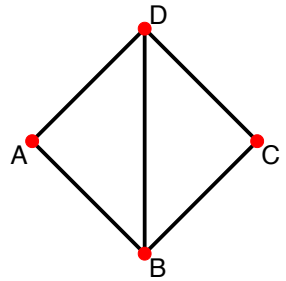
Veroniki, A. A., Vasiliadis, H. S., Higgins, J. P., and Salanti, G. (2013). Evaluation of inconsistency in networks of interventions. *International Journal of Epidemiology*, 42(1):332–345. doi: 10.1093/ije/dys222.

Viechtbauer, W. (2015). *metafor: Meta-Analysis Package for R*. R package version 1.9-8.

White, I. R. (2011). Multivariate random-effects meta-regression: Updates to mvmeta. *Stata Journal*, 11(2):255–270.

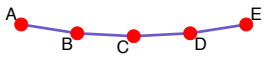
White, I. R., Barrett, J. K., Jackson, D., and Higgins, J. P. T. (2012). Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Research Synthesis Methods*, 3(2):111–125.

Appendix: Graph-theoretical methods for network meta-analysis



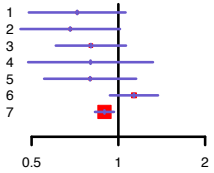
- Networks are graphs
  - Nodes are treatments
  - Edges are comparisons between treatments, based on studies
- 'Variances combine like electrical resistances' (Bailey, 2007)
- Exactly for this reason, it is possible to apply methods from electrical network theory to network meta-analysis (Rücker, 2012)

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 $i - j$

Meta-analytic networks and electrical networks

There is a complete correspondence

- 'Variances combine like electrical resistances' (Bailey, 2007)
- 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
- Implemented in the R package **netmeta** (Rücker et al., 2015)
  - Contrast-based or arm-based data entry

Drawing the network (Rücker and Schwarzer, 2015a)

In general, looking for a two-dimensional representation of the graph

- $\begin{pmatrix} x_i \\ y_i \end{pmatrix}$ point in the plane representing treatment i
- $\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} x_i, \ldots, x_n \\ y_i, \ldots, y_n \end{pmatrix}$ layout in the plane, representing all n treatments

- Find an 'optimal' layout, accounting for connectivity:
  - Avoid projecting different treatments on the same point in the plane
  - Treatments compared directly to each other should not be placed too widely distant from each other
  - Try to avoid unneccessary crossings
- Ideally, the layout should show some of the structure/symmetry of the network

Drawing the network

Method implemented in **netmeta**:
Stress algorithmus (Kamada and Kawai, 1989; Hu, 2012)

- Iterative procedure
- Specify 'ideal' distances d_{ij} between points i and j in $\mathbb{R}^2$
  - Our choice: For d_{ij}, take the graph distance (length of shortest path) between two points
- Minimise the objective function

$$\sum_{i \neq j} \left(\left\| \begin{pmatrix} x_i \\ y_i \end{pmatrix} - \begin{pmatrix} x_j \\ y_j \end{pmatrix} \right\| - d_{ij} \right)^2$$

That is, try to approximate the ideal distances!

Comparing two treatments

- Consider a network meta-analysis with n treatments and normal Bayesian posteriors Y_i with means μ_i and standard deviations σ_i

$$P(Y_i > Y_j) = \Phi\left(\frac{\mu_i - \mu_j}{\sqrt{\sigma_i^2 + \sigma_j^2}} \right) \quad (i, j = 1, \ldots, n)$$

- Frequentist version

$$P_{ij} = \Phi\left(\frac{\hat{\mu}_i - \hat{\mu}_j}{\sqrt{s_i^2 + s_j^2}} \right) = \begin{cases} p_{ij}/2, & \text{if } \hat{\mu}_i \leq \hat{\mu}_j \\ 1 - p_{ij}/2, & \text{if } \hat{\mu}_i > \hat{\mu}_j \end{cases}$$

- P_{ij} can be derived from the p-values of the comparisons μ_i versus μ_j
- The greater P_{ij}, the more certain we are that $\mu_i > \mu_j$, and vice versa.

P-scores

- P_{ij} derived from the p-values of the pairwise comparisons
- P_{ij} $(0 \leq P_{ij} \leq 1)$ is the probability that treatment i is better than treatment j
- P-score for treatment i defined as

$$\bar{P}_i = \frac{1}{n-1} \sum_{j,\, j \neq i}^{n} P_{ij}$$

- $\bar{P}_i$ is the mean extent of certainty that treatment i is better than treatment j, averaged over all competing treatments j $(j \neq i)$
- $\bar{P}_i$ represents the continuous rank of treatment i within the given range of treatments (1 = best, 0 = worst)
- Corresponds to the interpretation of the SUCRA values!

A proof that SUCRA and P-score are identical

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 $SUCRA(i)$.