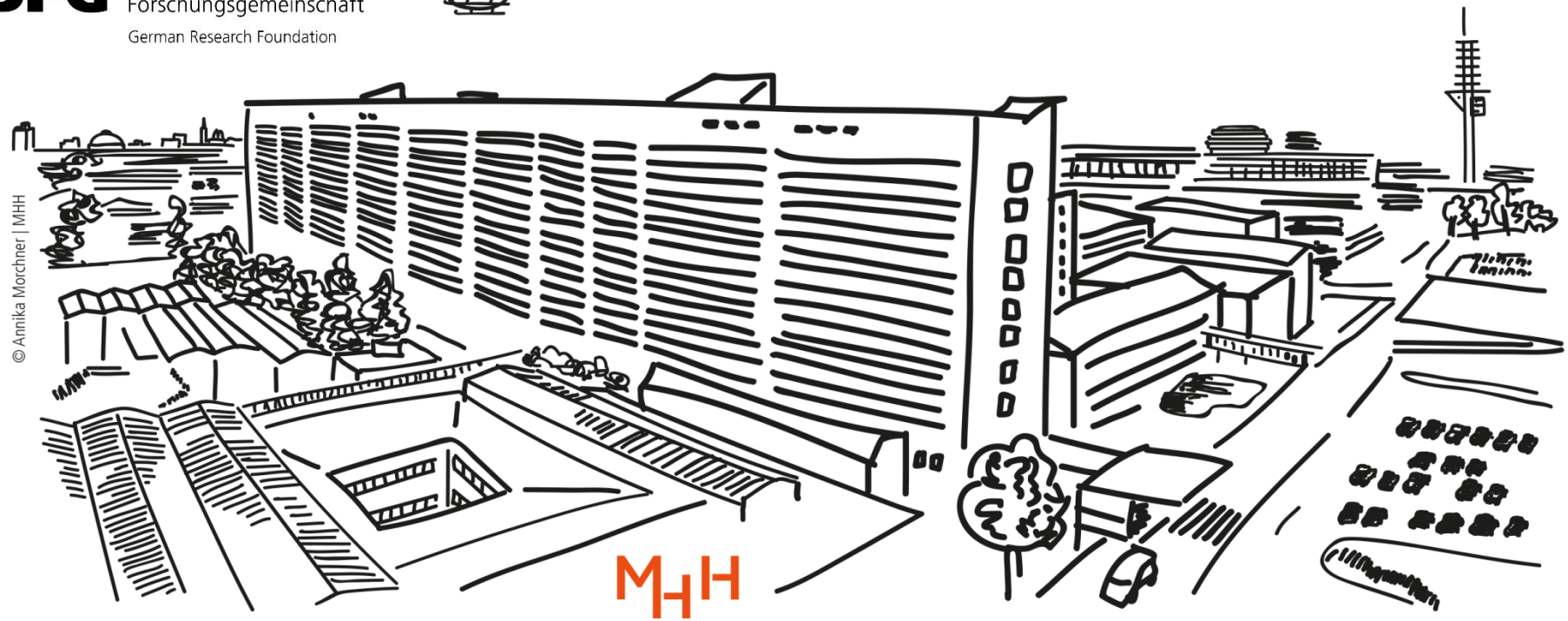
Hannover Medical School

Hierarchical clustering for the evaluation of transitivity assumption in a network of interventions

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation



Loukia M. Spineli (Spineli.Loukia@mh-hannover.de)

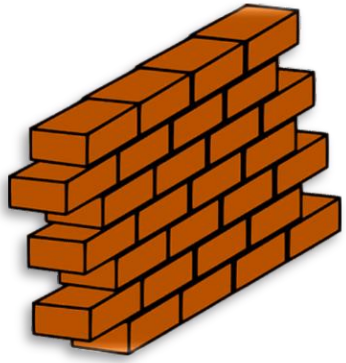
Disclaimer: The work has been presented to the "44th Annual Conference of the International Society for Clinical Biostatistics"

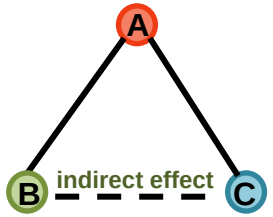
Why is transitivity assumption important

Network meta-analysis

- The **cornerstone** of network meta-analysis (NMA)
- **Extension of the similarity assumption** in pairwise meta-analysis
  - ➢ Clinical and methodological similarity **across** the comparisons

Transitivity assumption





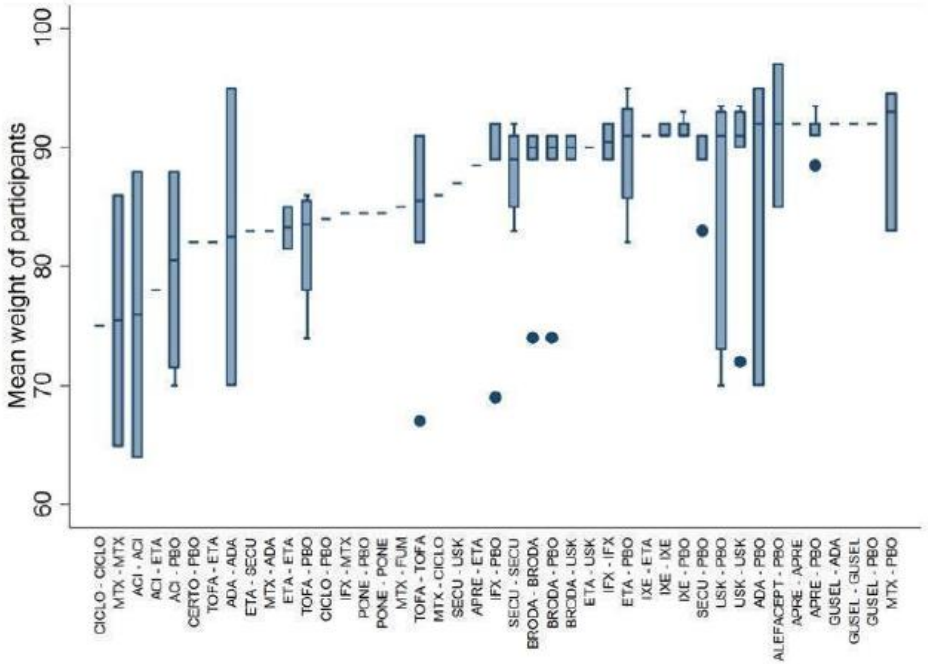
- Necessary to obtain **credible indirect estimates**
- **Violation** of transitivity compromises the NMA results

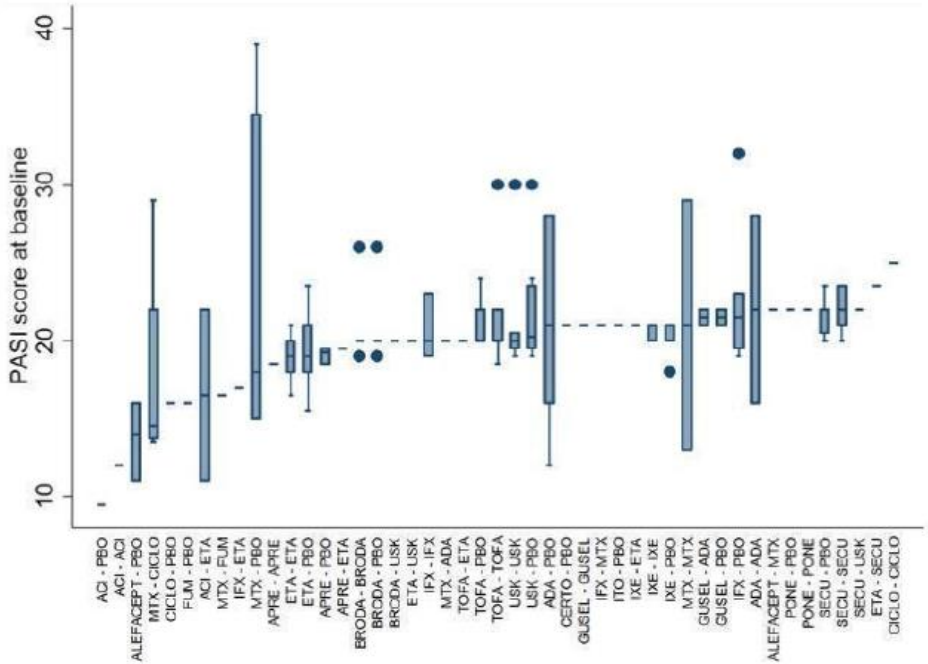
- **Validity** of transitivity rests on clinical and epidemiological grounds
  - ➢ … and the reporting quality of the included trials

Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. BMC Med 2013;11

What motivated this study (1)

- **Graphical evaluation** of each characteristic individually:





Sbidian E, Chaimani A, Garcia-Doval I, et al. Systemic pharmacological treatments for chronic plaque psoriasis: a network meta-analysis. Cochrane Database Syst Rev. 2017;12(12):CD011535.

What motivated this study (2)

- **Statistical testing** each characteristic individually:

Table 1—Summary of trial characteristics by type of treatment comparison.

	CPAP vs IC (n = 4)	ASV vs IC (n = 7)	ASV vs CPAP (n = 3)	O_2 vs IC (n = 1)	ASV vs O_2 (n = 1)	*P* (ANOVA)
Age (years)	62.8 (1.0)	68.5 (1.3)	64.3 (2.3)	64.1	68.5	.10
BMI (kg/m^2)	28.7 (0.5)	25.3 (0.8)	26.6 (0)	–	23.8	.31
AHI (events/h)	39.9 (0.3)	28.5 (2.4)	47.0 (2.9)	19.4	35.6	.002
LVEF (%)	23.9 (1.0)	33.7 (6.0)	32.7 (1.0)	33.7	35.6	.18
Length of follow-up (months)	3 (0)	4.8 (0.9)	3 (0)	3	3	.43
ACE (%)	79.6 (3.0)	85.5 (5.0)	91.7 (0.6)	62.5	78.2	.17
Beta blocker (%)	76.5 (4.1)	94.0 (3.1)	75.2 (3.8)	–	84.7	.42
Ischemic heart disease	66.5	46.9 (12.5)	26.2 (0.9)	–	48.4	.16

Data are reported as weighted mean (standard error) with weights equal to sample size. Standard deviation is not reported if data from only one study is available. AHI = apnea-hypopnea index, ANOVA = analysis of variance, ASV = adaptive servoventilation, BMI = body mass index, CPAP = continuous positive airway pressure, IC = inactive control, LVEF = left ventricular ejection fraction, O_2 = nocturnal oxygen.

Schwarz EI, Scherff F, Haile SR, Steier J, Kohler M. Effect of Treatment of Central Sleep Apnea/Cheyne-Stokes Respiration on Left Ventricular Ejection Fraction in Heart Failure: A Network Meta-Analysis. J Clin Sleep Med. 2019;15(12):1817-1825.

What motivated this study (3)

Depend on trial and characteristic availability

Graphical evaluation is overly subjective

❖ **Problems** with graphical evaluation and multiple testing!

Multiplicity due to multiple testing

Low power when insufficient trials and/or high missingness

Aim of the study

- **Hierarchical agglomerative clustering** for transitivity evaluation.
  - Do comparisons of interventions **form one cluster (evidence of transitivity)** or several clusters?
- **Exploring** plausibility of transitivity based on **semi-objective grounds** and without multiple statistical testing.
  - Well-established algorithms;
  - Visualisation toolkit of clustering.

And how is it supposed to work? (1)

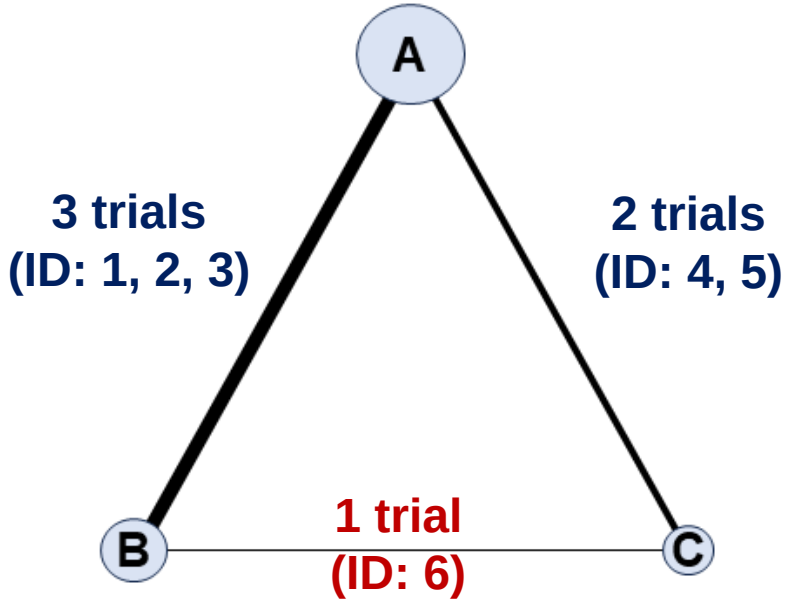


Table of Characteristics for the analysed network

Trial	Arm 1	Arm 2	Sample	Duration	Randomisation	Blinding	...
1	A	B	64	3	Yes	Yes	...
2	A	B	389	6	Yes	Yes	...
3	A	B	266	6	No	Yes	...
4	A	C	148	13	No	Yes	...
5	A	C	525	24	No	Yes	...
6	**B**	**C**	**153**	**6**	**No**	**Yes**	**...**

And how is it supposed to work? (2)

- Obtain **Gower's dissimilarity** **per comparison** (ideal for mixed-type variables)

B vs A	1	2	3
1	0	0	0
2	0.50	0	0
3	0.62	0.58	0

C vs A	4	5
4	0	0
5	0.30	0

C vs B	6	6_min	6_max
6	0	0	0
6_min	0.40	0	0
6_max	0.40	0.50	0

Completely similar

Gower $\in [0, 1]$

Completely dissimilar

Gower between two trials for characteristic *i*

- $d_{xy}^{i} = \frac{|x_i - y_i|}{R_i}$ (metric characteristic)
- $d_{xy}^{i} = \begin{cases} 1, & \text{if } x_i \neq y_i \\ 0, & \text{if } x_i = y_i \end{cases}$ (unordered categorical)
- $d_{xy}^{i} = \frac{|rank(x_i) - rank(y_i)|}{RR_i}$ (ordered categorical)

Gower J. General Coefficient of Similarity and Some of Its Properties. Biometrics 1971;27:857–71.

And how is it supposed to work? (3)

- **Pseudostudies** for single-trial comparisons!

C vs B	6	6_min	6_max
6	0	0	0
6_min	0.40	0	0
6_max	0.40	0.50	0

- A conservative **tentative** suggestion;
- **Map the dispersion** of the observed evidence in each characteristic;
- **Prevent** spurious low dissimilarity;

Trial	Arm 1	Arm 2	Sample	Duration	Randomisation	Blinding	...
6	**B**	**C**	**153**	**6**	**No**	**Yes**	**...**
6_min	B	C	64	3	Yes	Yes	...
6_max	B	C	525	24	No	Yes	...

6_min (6_max): Add pseudostudy with minimum (maximum) value of each metric characteristic and the least (most) frequent category of each categorical characteristic based on the remaining studies.

And how is it supposed to work? (4)

- **Total dissimilarity** (the **between-trial non-statistical heterogeneity**)!

Square root of mean square around zero dissimilarity of the lower off-diagonal elements

B vs A	1	2	3
1	0	0	0
2	0.50	0	0
3	0.62	0.58	0

$$\sqrt{\frac{(0.50-0)^2+(0.62-0)^2+(0.58-0)^2}{3}} = \mathbf{0.57}$$

C vs A	4	5
4	0	0
5	0.30	0

$$\sqrt{\frac{(030-0)^2}{1}} = \mathbf{0.30}$$

C vs B	6	6_min	6_max
6	0	0	0
6_min	0.40	0	0
6_max	0.40	0.50	0

$$\sqrt{\frac{(0.40-0)^2+(0.40-0)^2+(0.50-0)^2}{3}} = \mathbf{0.44}$$

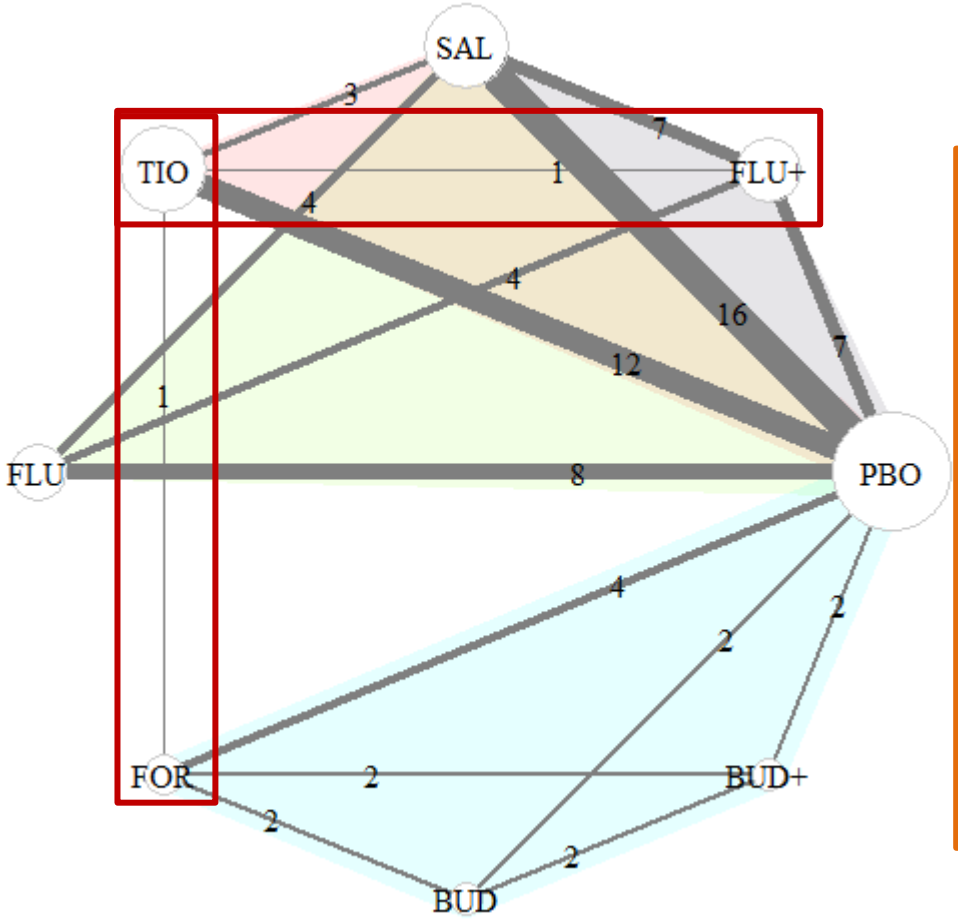
And how is it supposed to work? (5)

- Comparison-specific **total dissimilarities** are **the input for clustering**:

  1) Dissimilarity matrix among the comparisons (e.g., using Euclidean distance)
  2) Cophenetic coefficient to select the proper linkage method (e.g., average, single, etc)
  3) Optimal partitioning using three internal measures: connectivity, Silhouette width and Dunn index.
  4) Dendrogram with coloured branches based on optimal partitioning.

Handl J, Knowles J, Kell DB. Computational cluster validation in post-genomic data analysis. Bioinformatics 2005;21:3201–12

Using a real systematic review

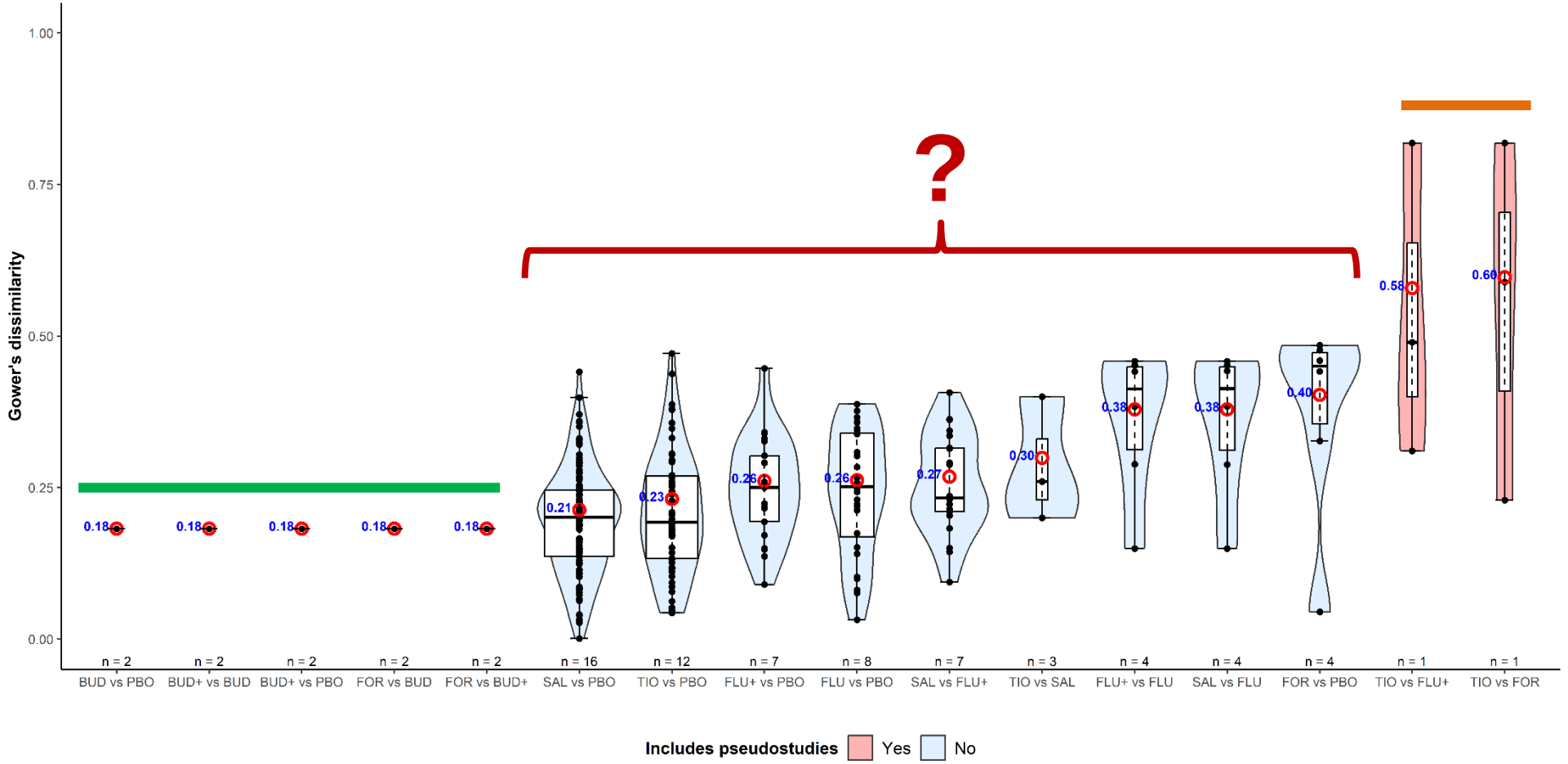


- Seven interventions for COPD
- **Two singe-trial comparisons**
- Seven metric characteristics
- Three categorical characteristics
- Coloured loops refer to multi-arm trials
- *rnmamod* R-package for analysis

Baker W, Baker E, Coleman C. Pharmacologic treatments for chronic obstructive pulmonary disease: a mixed-treatment comparison meta-analysis. Pharmacotherapy 2009;29:891–905
Spineli L. rnmamod: Bayesian Network Meta-Analysis with Missing Participants. R package version 0.3.0. 2022. https://CRAN.R-project.org/package=rnmamod

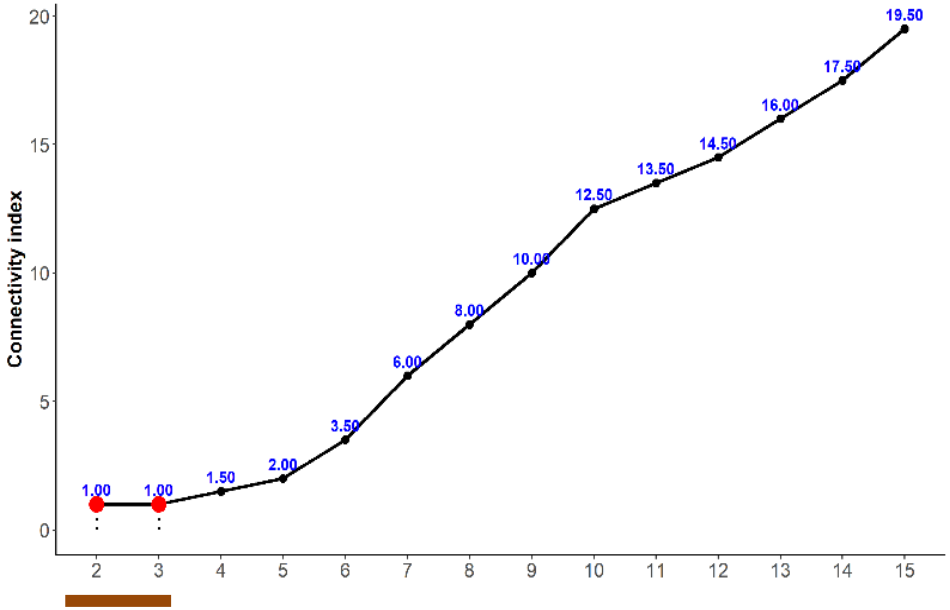
Using a real systematic review

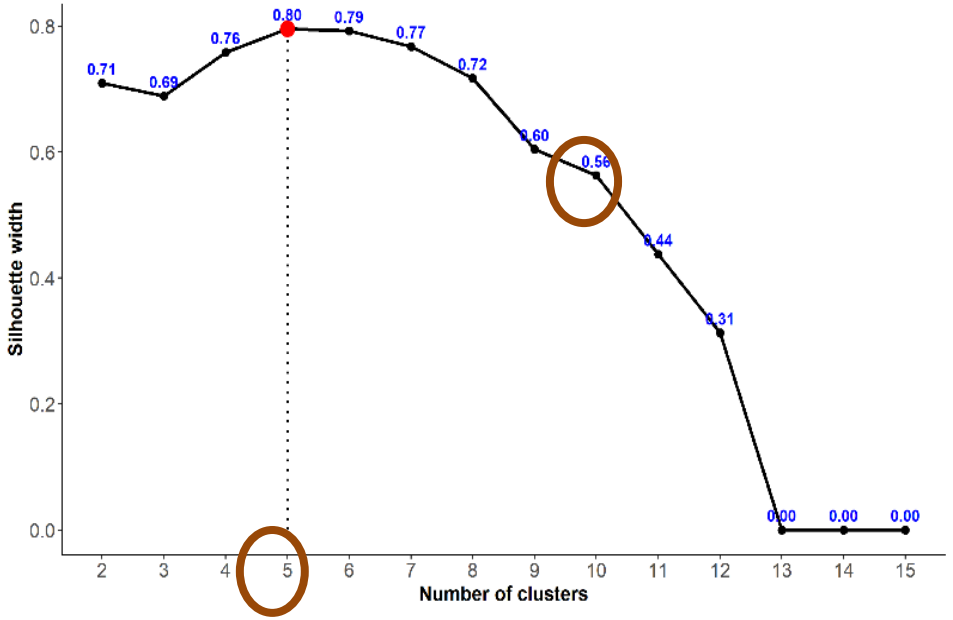
Distribution of Gower‘s dissimilarity (black points) and **total dissimilarity (red points)**

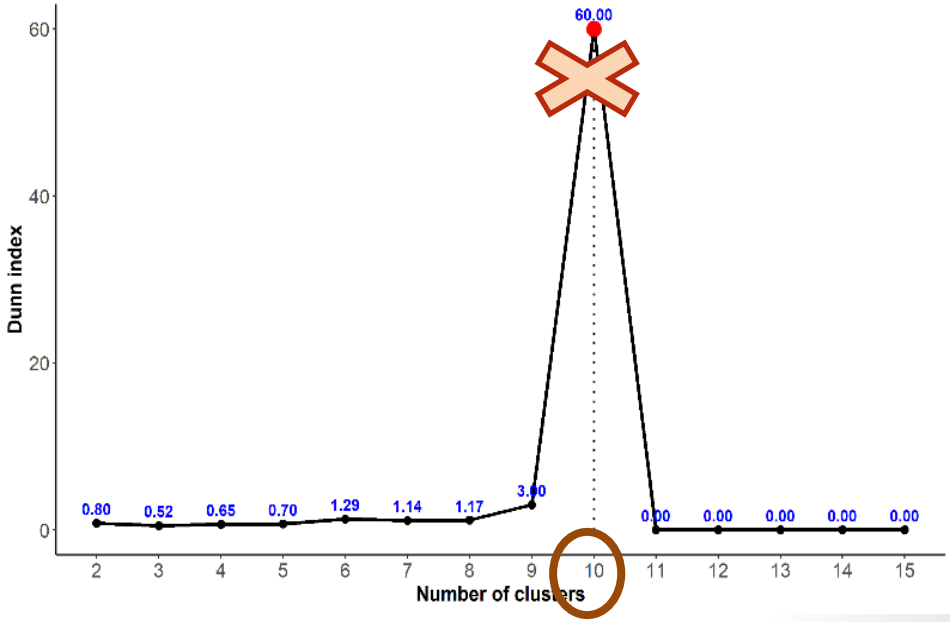


Baker W, Baker E, Coleman C. Pharmacologic treatments for chronic obstructive pulmonary disease: a mixed-treatment comparison meta-analysis. Pharmacotherapy 2009;29:891–905

Internal measures

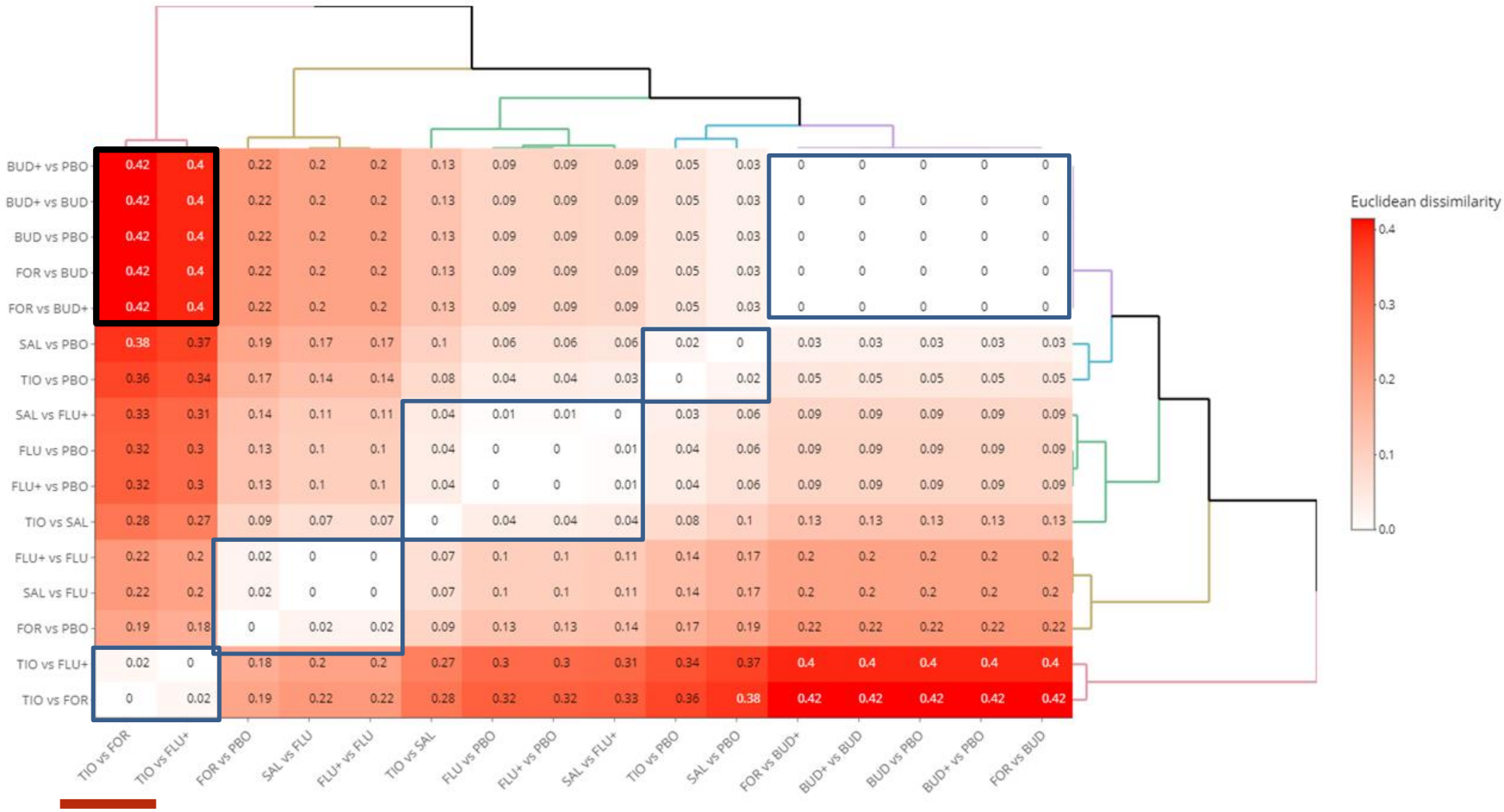






Baker W, Baker E, Coleman C. Pharmacologic treatments for chronic obstructive pulmonary disease: a mixed-treatment comparison meta-analysis. Pharmacotherapy 2009;29:891–905

Dendrogram with heatmap



Baker W, Baker E, Coleman C. Pharmacologic treatments for chronic obstructive pulmonary disease: a mixed-treatment comparison meta-analysis. Pharmacotherapy 2009;29:891–905

Discussion & conclusions (1)

- **Was the partitioning to several clusters justified?**
  - Total dissimilarities indicated **variable non-statistical heterogeneity** across comparisons (range: 0.18 to 0.60). **Evidence of possible intransitivity.**
  - Specifying the **extent of comparison dissimilarity** ➲ probaly less clusters.
  - Cluster based on the **extent of total dissimilarity** (informative clustering)
    - It **reminds of the I^2 statistic** and we could adopt its thresholds:

Non-statistical heterogeneity	Total dissimilarity
Low	$[0.00 - 0.25]$
Moderate	$(0.25 - 0.50]$
High	$(0.50 - 0.75]$
Very high	$(0.75 - 1.00]$

Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557–60.

Discussion & conclusions (2)

- **Comparisons did not form one cluster. And now what?**
  - Are there any **extraction errors**?
  - Are the **measure scales variable**?
    - *Mixing years and months regarding trial duration!*
  - Check violin plot with total dissimilarities and the **distibution of each characteristic for each cluster**.
  - **Split into subnetworks** based on the clusters.
  - Refrain from NMA if **subnetworks** are **not meaningful**
    - *Subnetworks are **disconnected or very sparse**!*

Linde K, Rücker G, Schneider A, Kriston L. Questionable assumptions hampered interpretation of a network meta-analysis of primary care depression treatments. J Clin Epidemiol 2016;71:86–96

Thank you for your attention!