COMPARISON BETWEEN CINeMA AND GRADE TO ASSESS THE CERTAINTY OF EVIDENCE FROM NETWORK META-ANALYSIS

Silvia Minozzi, Michela Cinquini, Cinzia Del Giovane and the University of Milan Post Graduate Course on Network Meta-analysis Working Group
(Arienti C, Battain PC, Brigadoi G, Del Vicario M, Di Domenico G, Farma T, Federico S, Innocenti T, La Rosa G, Orlandi E, Piersanti A, Selvanetti A, Zanetta L)

Bias Methods Group annual meeting

Prague, September 10, 2024

Conflict of interest

Silvia Minozzi, Michela Cinquini, Cinzia Del Giovane

S. Minozzi, Joint Co-Ed Drugs and Alcohol Group, Department of Epidemiology, Lazio Regional Health Service, Rome, Italy; Laboratory of methodology of systematic reviews and guidelines production, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy

M. Cinquini, Head of Laboratory of methodology of systematic reviews and guidelines production, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy

C. Del Giovane, Senior Researcher at the University of Modena and Reggio Emilia

S. Minozzi and M. Cinquini have no actual or potential conflict of interest in relation to this presentation

C. Del Giovane is a co-author of the publications about CINeMA

Background

Two approaches are available to assess the confidence in NMA results:

- the Grading of Recommendations Assessment, Development and Evaluation (**GRADE**)[1],

- the Confidence in Network Meta-Analysis (**CINeMA**)[2].

Although they share many common aspects, their operationalization differs.

1.Brignardello-Petersen R et al. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. J Clin Epidemiol. 2018 Jan;93:36-44

2.Nikolakopoulou A, et al. (2020) CINeMA: An approach for assessing confidence in the results of a network meta-analysis. PLoS Med 17(4): e1003082.

Confidence in the evidence of each comparison

- Outcome and comparison –specific
- Subjective
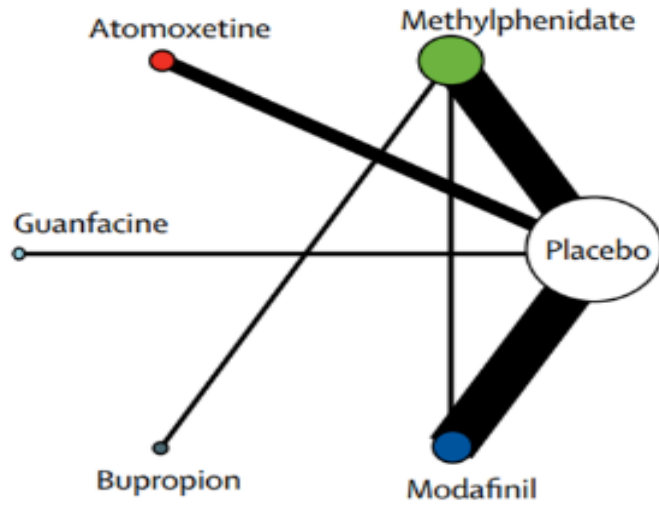
GRADE domains

1. Study limitations
2. Publication bias
3. Indirectness
4. Imprecision
5. Inconsistency
6. Incoherence

For each domain:
- Not serious
- Serious
- Very serious

CINeMA domains

1. Within-study bias
2. Reporting bias
3. Indirectness
4. Imprecision
5. Heterogeneity
6. Incoherence

For each domain:
- No concerns
- Some concerns
- Major concerns

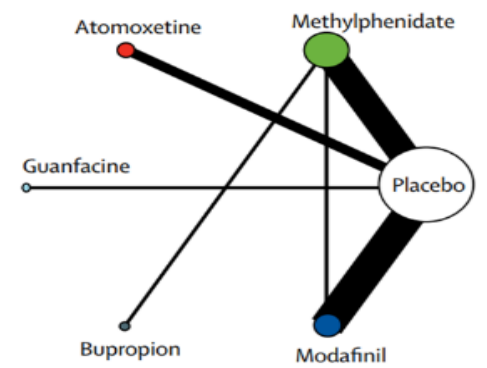## Confidence

⊕⊕⊕⊕ High

⊕⊕⊕◯ Moderate

⊕⊕◯◯ Low

⊕◯◯◯ Very low

- Not serious/No concerns: do not downgrade
- Serious/Some concerns: downgrade 1 level
- Very serious/Major concerns: downgrade of 2 levels

Main differences between GRADE and CINeMA

GRADE

- **All domains except imprecision and incoherence** assessed at the **pairwise levels**, both for direct and indirect evidence

- Imprecision assessed at the network estimate level

- Assess direct estimate at first; if high certainty and its contribution to the network estimate ≥ as that of the indirect evidence, rating the indirect estimates not needed

- Only 1st (or 2nd) order loop of the indirect evidence is considered

- **Manual Step by step** approach for each comparison at a time

- **Publication bias**: only funnel plot and qualitative consideration considered

CINeMA

- **All domains assessed at the network level**

- Within-study bias, indirectness and reporting bias assessed considering the entire network, with the contribution of each direct evidence as weights

- Heterogeneity: assess if the comparison between prediction interval and confidence interval and the margin of equivalence leads to different (clinical) conclusion

- **Automated (by software) overall approach**, rules implemented in the software

- **Reporting bias**: RoB MEN software used, which applies the RoB Missing Evidence approach to NMA

Aims

To compare the two approaches **on the overall judgments** in terms of:

➢ **inter-rater agreement** among assessors within each network and each approach

➢ similarities of results (**concordance**) between GRADE and CINeMA

➢ **time needed to apply** each approach within each network

Methods

Both approaches applied to 4 networks: antiplatelet treatments for secondary stroke prevention, pharmacotherapies for seizures, pharmacotherapies for ADHD, pharmacotherapies and surgery for obesity
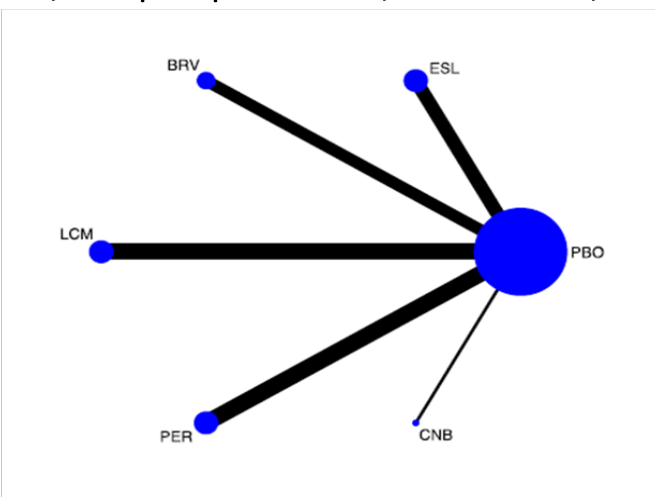
Ischemic stroke (dichotomous)
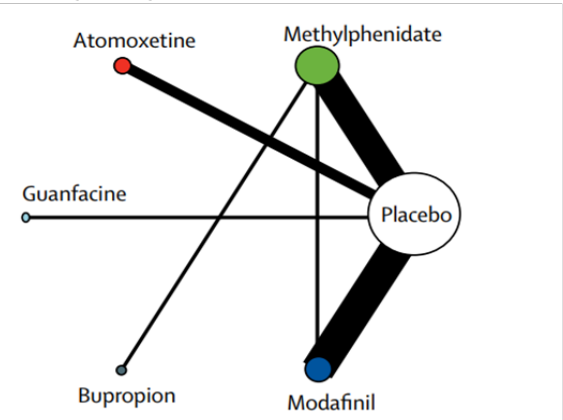n=39, treat=15, comp vs placebo=14, ind vs ref=10, loop vs ref=4
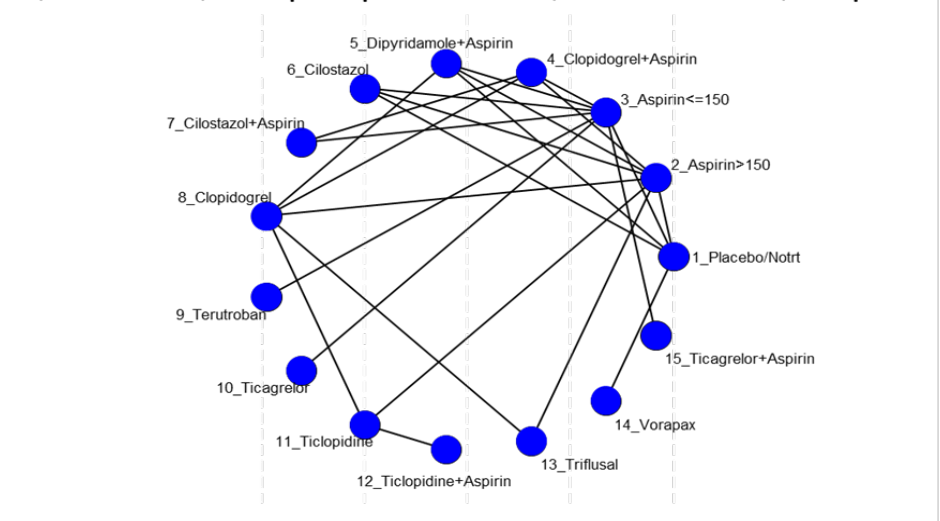
Seizure response (dichotomous)
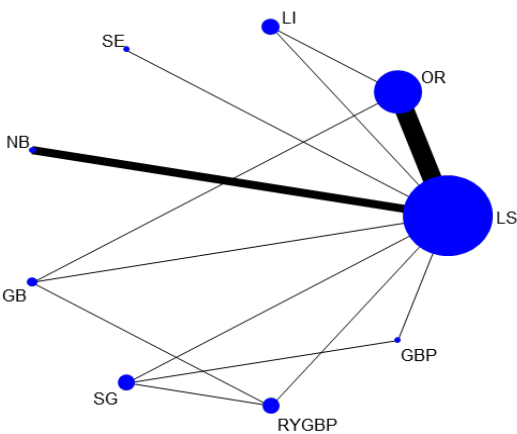n=16, treat=6, comp vs placebo=5, ind vs ref=0, loop vs ref=0

Change mean ADHD symptoms (continuous)
n=16, treat=6, comp vs placebo=5, ind vs ref=1, loop vs ref=1

Weight loss in Kg (continuous)
n=61, treat=9, comp vs LS=8, ind vs ref=0, loop vs ref=5

Methods

	Seizure response	Change in ADHD symptoms	Weight loss (Kg)	Ischemic stroke
Type of outcome	dichotomous	continuous	continuous	dichotomous
Number of studies	16	16	61	39
Number of treatments	6	6	9	15
Reference treatment	Placebo	Placebo	Life style modification	Placebo
Number of comparisons of interest vs reference	5	4	8	5
Number of indirect comparisons vs reference	-	1	-	10
Number of loops vs reference	-	1	5	4

	First evaluation		**Second evaluation**	
Group	**Method**	**Outcome**	**Method**	**Outcome**
A (n:4)	GRADE	Change in ADHD Symptoms (by teachers)	CINeMA	Ischemic stroke
B (n:3)	GRADE	Seizure response	CINeMA	Weight loss
C (n:3)	CINeMA	Change in ADHD Symptoms (by teachers)	GRADE	Ischemic stroke
D (n:3)	CINeMA	Seizure response	GRADE	Weight loss

Thirteen assessors, who never applied both approaches but were trained on NMA methodology and such evaluation methods, were randomly assigned to four groups applying first GRADE and then CINeMA or viceversa; each group applied both approaches on one continuous and one dichotomous outcome.

Methods

- **RoB, Indirectness**: utilized the results provided by reviews' authors

- **Imprecision**: margin of equivalence – MID defined at the protocol stage for each network.

GRADE: minimally contextualized approach; Brignardello 2021 GRADE Guidelines 33 and the excel worksheet provided with the paper

- **Publication/reporting bias**

GRADE: GRADE guidelines 5. JCE 2011

CINeMA: RoB MEN https://cinema.ispm.unibe.ch/rob-men/ ; burn in, n. iterations, thinning factor, assumption for treatment specific interaction: set by the statistician

Outcomes

- **Inter-Rater Reliability**: Gwet's AC kappa (linear weights) for overall judgments for each network (for ordinal data and multiple raters, accounts for the severity of disagreements between ordinal ratings)

- **Concordance** between the overall judgments of GRADE and CINeMA: Gwet's AC kappa (linear weights) for each network on medians of evaluators responses

- **Time to complete the evaluation**: median time (in minutes) and interquartile range (IQR) needed for each network with each approach, including time to consult the guidance papers.

Results -Inter-Rater Reliability
Gwet's AC (95% confidence interval)

	GRADE	CINeMA
Seizure response	0.49 (0.23-0.76)	0.66 (0.66-0.66)
Change in ADHD symptoms	0.61 (0.32-0.90)	0.73 (0.48-0.99)
Weight loss	0.55 (0.32-0.77)	0.61 (0.36-0.85)
Ischemic stroke	0.70 (0.52- 0.88)	0.02 (-0.12-0.16)

Results – Concordance between GRADE & CINeMA

	Gwet's AC (95% confidence interval)
Seizure response	1
Change in ADHD symptoms	0.90 (0.58-1)
Weight loss	0.68 (0.23-1)
Ischemic stroke	0.42 (0.12-0.71)

Results – Median Time to complete the evaluation (minutes)

	GRADE	CINeMA
Seizure response	160 (140-345)	150 (100-219)
Change in ADHD symptoms	316 (95-1300)	330 (169-370)
Weight loss	290 (140-310)	176 (160-200)
Ischemic stroke	481 (475-1253)	338 (191-720)

.

Preliminary Conclusions

- Moderate/Substantial agreement across raters in both methods, except for ischemic stroke with CINeMA

- The assessments obtained with the two approaches seem to agree for small networks with none or few indirect comparisons. Disagreements seems to increase as the number of comparisons and indirect evidence increase.

- When the networks become more complex and the evaluation assessment is performed on several comparisons the time spent to make the judgments seems longer with GRADE compared with CINeMA, concluding that GRADE seems more time consuming compared to CINeMA

Strenghts and Limitations

- **Strenghts**
  - Assessment of 4 different networks that enhance the generalizability of our results
  - Evaluators homogeneous about the knowledge and use of the methods
  - Evaluators representative of potential users of the methods

- **Limitations**
  - Small sample of evaluators: 3/4 evaluators per network
  - Different rating categories for publication bias/reporting bias (2 for GRADE vs 3 for CINeMA) and imprecision (4 for GRADE vs 3 for CINeMA) could impact the IRR and the concordance
  - Analysis per domain to be done to understand if there is one or more domains that explain the low values of IRR and concordance, especially for ischemic stroke