

Lecture 8

CINeMA: a framework and software to evaluate Confidence in Network Meta-Analysis

Adriani Nikolakopoulou

None of the 456 NMAs published until 3/2015 attempted to evaluate the confidence in NMA results!

BMJ 2014;349:g5630 doi: 10.1136/bmj.g5630 (Published 24 September 2014)

Page 1 of 10

RESEARCH METHODS & REPORTING

A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis

Network meta-analysis (NMA), combining direct and indirect comparisons, is increasingly being used to examine the comparative effectiveness of medical interventions. Minimal guidance exists on how to rate the quality of evidence supporting treatment effect estimates obtained from NMA. We present a four-step approach to rate the quality of evidence in each of the direct, indirect, and NMA estimates based on methods developed by the GRADE working group. Using an example of a published NMA, we show that the quality of evidence supporting NMA estimates varies from high to very low across comparisons, and that quality ratings given to a whole network are uninformative and likely to mislead.

Milo A Puhan¹, Holger J Schünemann², Mohammad Hassan Murad³, Tianjing Li⁴, Romina Brignardello-Petersen⁵, Jasvinder A Singh⁶, Alfons G Kessels⁷, Gordon H Guyatt², for the GRADE Working Group

None of the 456 NMAs published until 3/2015 attempted to evaluate the confidence in NMA results!

OPEN & ACCESS Freely available online



BMJ 2014;349:g5

Evaluating the Quality of Evidence from a Network Meta-Analysis

Georgia Salanti¹, Cinzia Del Giovane², Anna Chaimani¹, Deborah M. Caldwell³, Julian P. T. Higgins^{3,4}*

1 Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, 2 Statistics Unit, Department of Clinical and Diagnostic Medicine and Public Health, University of Modena and Reggio Emilia, Modena, Italy, 3 School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom, 4 Centre for Reviews and Dissemination, University of York, York, United Kingdom

A GRA quality meta-a

Network me used to exa on how to ra We present NMA estima a published to very low a and likely to

Milo A Puha Brignardello - - -Working Group Systematic reviews that collate data about the relative effects of multiple interventions via network meta-analysis are highly informative for decision-making purposes. A network meta-analysis provides two types of findings for a specific outcome: the relative treatment effect for all pairwise comparisons, and a ranking of the treatments. It is important to consider the confidence with which these two types of results can enable clinicians, policy makers and patients to make informed decisions. We propose an approach to determining confidence in the output of a network meta-analysis. Our proposed approach is based on methodology developed by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group for pairwise meta-analyses. The suggested framework for evaluating a network meta-analysis acknowledges (i) the key role of indirect comparisons (ii) the contributions of each piece of direct evidence to the network meta-analysis; estimates of effect size; (iii) the importance of the transitivity assumption to the validity of network meta-analysis; and (iv) the possibility of disagreement between direct evidence and indirect evidence. We apply our proposed strategy to a systematic review comparing topical antibiotics without steroids for chronically discharging ears with underlying eardrum perforations. The proposed framework can be used to determine confidence in the results from a

Abstract

CINeMA framework

Consider the **network estimates**

Study limitations
Indirectness
Inconsistency (heterogeneity, incoherence)
Imprecision
Publication bias

Rate each network
estimate
No concerns
Some concerns
Major concerns

Network estimate	Study limitations	Indirectness	Inconsistency		Imprecision	Publication bias	Confidence
			Heterogeneity	Incoherence			
A vs B	Some concerns	Some concerns	Major concerns	Some concerns	Some concerns	undetected	Very low
A vs C	No concerns	No concerns	No concerns	Major concerns	No concerns	suspected	Low

Welcome to CINEMA!

CINEMA (Confidence in Network Meta-Analysis) is a web application that simplifies the evaluation of confidence in the findings from network meta-analysis.

It is based on a framework described in [1] which considers the five GRADE domains: study limitations, indirectness, inconsistency, imprecision and publication bias. The framework combines judgments about direct evidence with their statistical contribution to network meta-analysis results, enabling evaluation of the credibility of NMA treatment effects.

[1. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JPT. Evaluating the quality of evidence from a network meta-analysis. PloS One. 2014;9(7):e99682.]

→ To browse your projects or upload a new one go to

MY PROJECTS

Methods developed by:

O CONFIGURATION 1 STUDY LIMITATIONS 2 IMPRECISION 3 INCONSISTENCY 4 INDIRECTNESS 5 PUBLICATION BIAS

Georgia Salanti Julian Higgins Anna Chaimani Adriani Nikolakopoulou

Web developer:

Theodore Papakonstantinou



UNIVERSIT. BERN

Institute of Social and Preventive Medicine (ISPM)





CINEMA is distributed, in the hope that it will be useful but with the SUBERING CINEMA you accept the following DISCLARM:

Matthias Egger

Project



Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis

William J Elliott, Peter M Meyer

Summary

Background The effect of different classes of antihypertensive drugs on incident diabetes mellitus is controversial because traditional meta-analyses are hindered by heterogeneity across trials and the absence of trials comparing angiotensin-converting-enzyme (ACE) inhibitors with angiotensin-receptor blockers (ARB). We therefore undertook a network meta-analysis, which accounts for both direct and indirect comparisons to assess the effects of antihypertensive agents on incident diabetes.

Lancet 2007; 369: 201-07

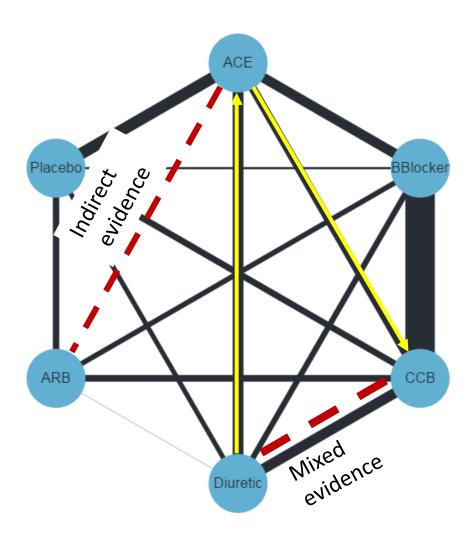
Department of Preventive Medicine, Rush Medical College of Rush University at Rush University Medical Center, Chicago, IL 60612, USA

Number of studies 22

Number of treatment nodes 6

Primary outcome Effect of antihypertensives on incidence diabetes mellitus - proportion of patients who developed diabetes

Measurement Binary



				Incons	sistency			
Comparison	Number of Studies	Study Limitations	Imprecision	Heterogeneity	Incoherence	Indirectness	Publication bias	CONFIDENCE
			Mixed evidence					
ACE vs BBlocker	3	No concerns	No concerns	Some concerns	Some concerns	No concerns	Undetected	MODERATE
ACE vs CCB	3	No concerns	Some concerns	Some concerns	No concerns	No concerns	Undetected	MODERATE
ACE vs Diuretic	2	No concerns	No concerns	Some concerns	No concerns	No concerns	Undetected	MODERATE
ACE vs Placebo	3	No concerns	Some concerns	Some concerns	No concerns	No concerns	Suspected	LOW
ARB vs BBlocker	1	No concerns	No concerns	Some concerns	No concerns	No concerns	Undetected	MODERATE
ARB vs CCB	1	Some concerns	No concerns	Some concerns	No concerns	No concerns	Undetected	LOW
ARB vs Diuretic	1	No concerns	No concerns	Some concerns	No concerns	No concerns	Undetected	MODERATE
ARB vs Placebo	2	Some concerns	Some concerns	No concerns	No concerns	No concerns	Suspected	VERY LOW
BBlocker vs CCB	5	No concerns	Some concerns	Some concerns	No concerns	No concerns	Undetected	MODERATE
BBlocker vs Diuretic	2	No concerns	Some concerns	Some concerns	No concerns	No concerns	Undetected	MODERATE
BBlocker vs Placebo	1	No concerns	Some concerns	Some concerns	Some concerns	No concerns	Suspected	VERY LOW
CCB vs Diuretic	2	No concerns	Some concerns	No concerns	No concerns	No concerns	Undetected	MODERATE
CCB vs Placebo	1	No concerns	Some concerns	Some concerns	No concerns	No concerns	Suspected	LOW
Diuretic vs Placebo	3	No concerns	No concerns	Some concerns	No concerns	No concerns	Suspected	LOW
			Indirect evidence	•				_
ACE vs ARB		No concerns	Some concerns	Some concerns	No concerns	No concerns	Undetected	LOW

				Inco	onsistency			
Comparison	Number of Studies	Study Limitations	Imprecision	Heterogeneity	Incoherence	Indirectness	Publication bias	CONFIDENC
			Mixed evidence	е				
ACE vs BBlocker	3	No concerns	No concerns	Some concerns	Some concerns	No concerns	Undetected	MODERATE
ACE vs CCB	3	No concerns	Some concerns	Some concerns	No concerns	No concerns	Undetected	MODERATE
ACE vs Diuretic	2		Se	mi-autom	ated proc	ess		MODERATE
ACE vs Placebo	3							LOW
ARB vs BBlocker	1	- 1.					, .	MODERATE
ARB vs CCB	1	Explic	cit rules tha	•			anaiysis	LOW
ARB vs Diuretic	1		eff	fect for ea	ch domain	to		MODERATE
ARB vs Placebo	2	N	No concerns	, Some co	ncerns, M	ajor conc	erns	VERY LOW
BBlocker vs CCB	5				ne docume			MODERATE
BBlocker vs Diuretic	2		as acse	indea in ti	ic docume	intation		MODERATE
BBlocker vs Placebo	1							VERY LOW
CCB vs Diuretic	2		<u>The </u>	<u>rules can l</u>	<u>oe overwri</u>	itten!		MODERATE
CCB vs Placebo	1	No concerns	Some concerns	Some concerns	No concerns	No concerns	Suspected	LOW
Diuretic vs Placebo	3	No concerns	No concerns	Some concerns	No concerns	No concerns	Suspected	LOW
			Indirect evidence	e				
ACE vs ARB		No concerns	Some concerns	Some concerns	No concerns	No concerns	Undetected	LOW

Study limitation

- Major concerns
- Some concerns
- No concerns

ACE BBlocker Placebo ARB CCB Diuretic Plot direct comparison

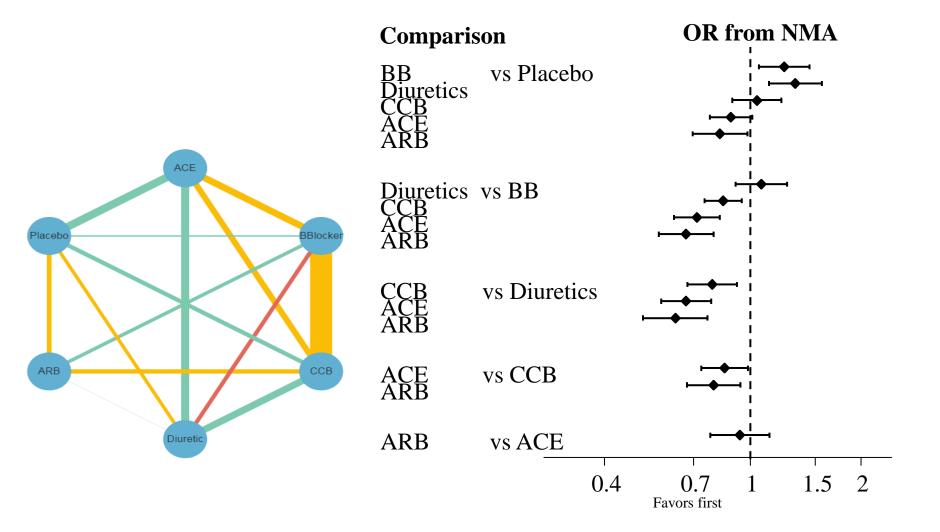
in green

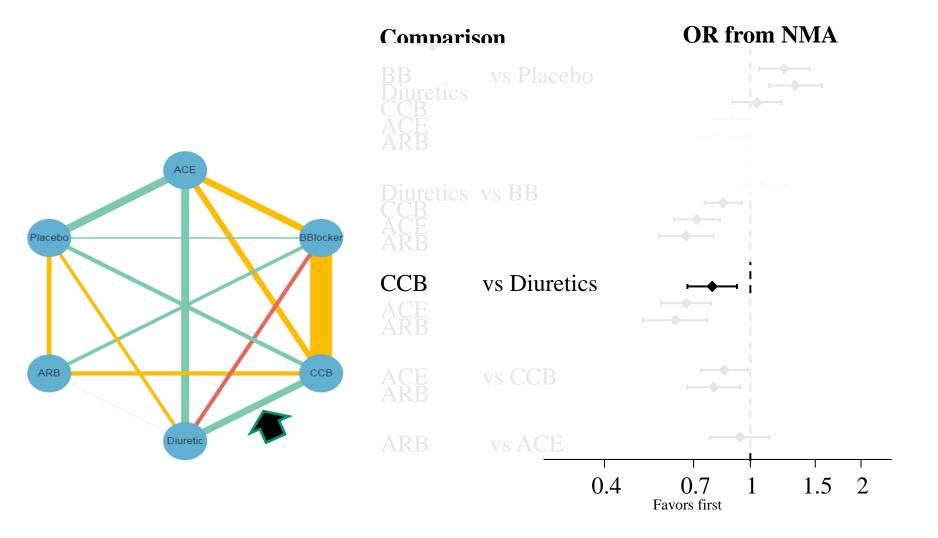
Form risk of bias judgements for each study

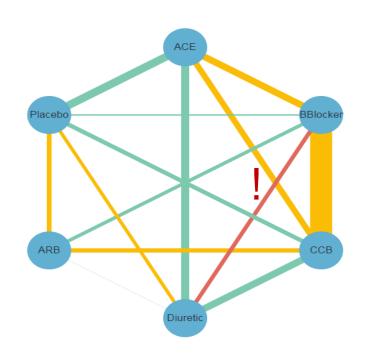
Consider selection, performance, attrition, detection and reporting bias Study name Risk of Bias

CCB vs Diuretics: overall low risk of bias

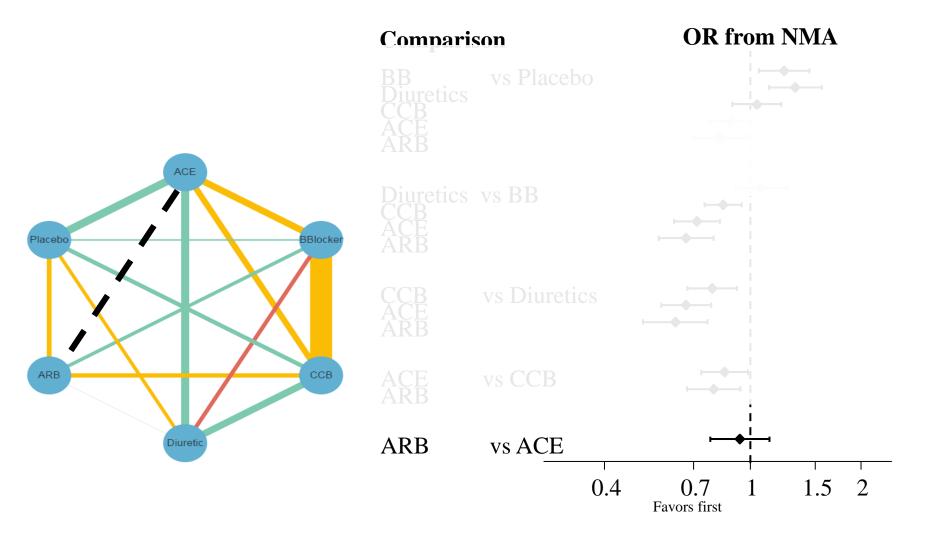
Study name	Risk of Bias
AASK	LOW
ALLHAT	LOW
ALPINE	LOW
ANBP-2	LOW
ASCOT	LOW
CAPPP	MODERATE
CHARM	LOW
DREAM	LOW
EWPHE	MODERATE
FEVER	LOW
HAPPHY	HIGH
HOPE	LOW
INSIGHT	LOW
INVEST	LOW
LIFE	LOW
MRC	LOW
NORDIL	LOW
PEACE	LOW
SCOPE	MODERATE
SHEP	LOW
STOP-2	MODERATE
VALUE	MODERATE







Studies with high risk of bias contribute to the estimation of the OR CCB vs Diuretics!

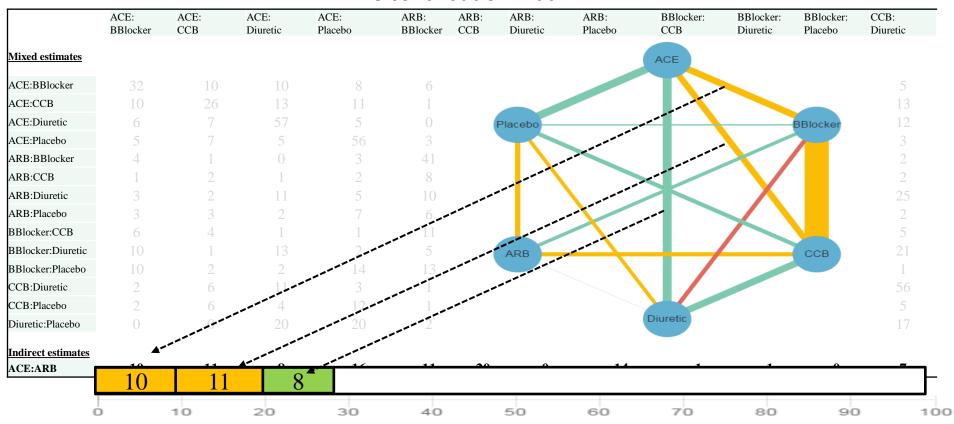


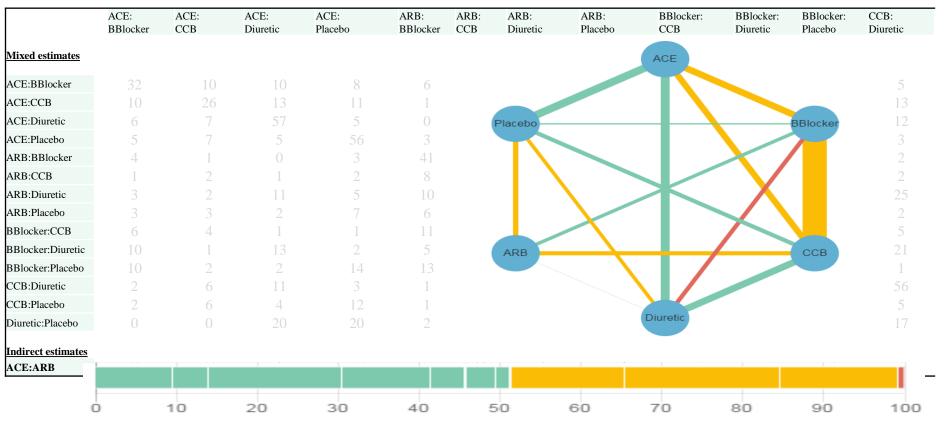
An indirect or mixed treatment effect is a combination of the available direct treatment effects

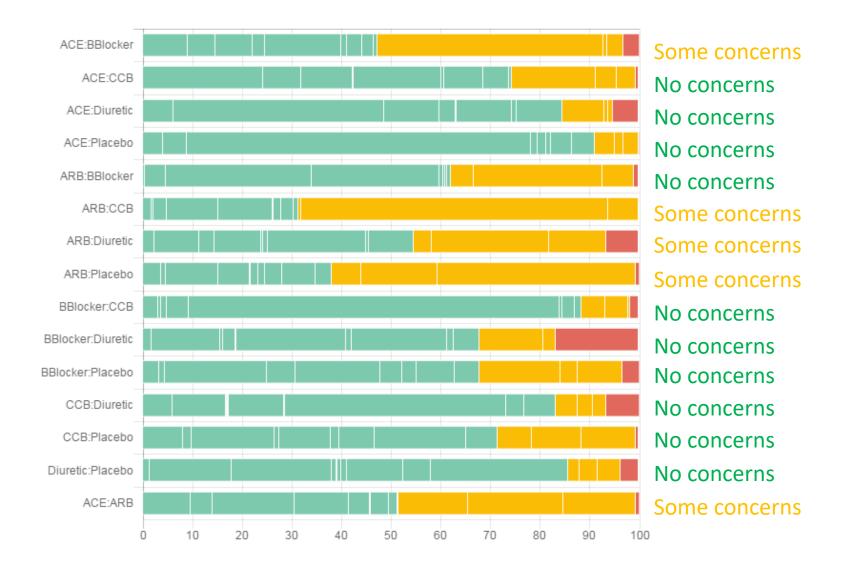
	ACE: BBlocker	ACE: CCB	ACE: Diuretic	ACE: Placebo	ARB: BBlocker	ARB: CCB	ARB: Diuretic	ARB: Placebo	BBlocker: CCB	BBlocker: Diuretic	BBlocker: Placebo	CCB: Diuretic	CCB: Placebo	Diuretic: Placebo
Mixed estimates														
ACE:BBlocker	32	10	10	8	6	1	0	4	15	6	2	5	2	0
ACE:CCB	10	26	13	11	1	6	0	4	9	1	0	13	6	0
ACE:Diuretic	6	7	57	5	0	2	0	2	1	5	0	12	2	2
ACE:Placebo	5	7	5	56	3	3	0	6	1	0	2	3	8	2
ARB:BBlocker	4	1	0	3	41	21	0	5	19	2	2	2	1	0
ARB:CCB	1	2	1	2	8	67	0	6	8	1	0	2	4	0
ARB:Diuretic	3	2	11	5	10	27	0	8	0	7	0	25	0	2
ARB:Placebo	3	3	2	7	6	15	0	49	0	1	2	2	10	1
BBlocker:CCB	6	4	1	1	11	12	0	0	53	4	2	5	2	0
BBlocker:Diuretic	10	1	13	2	5	3	0	2	19	20	2	21	0	2
BBlocker:Placebo	10	2	2	14	13	3	0	16	16	4	8	1	11	2
CCB:Diuretic	2	6	11	3	1	3	0	2	7	6	0	56	3	2
CCB:Placebo	2	6	4	12	1	15	0	16	6	0	2	5	28	2
Diuretic:Placebo	0	0	20	20	2	7	0	9	0	5	2	17	11	7
Indirect estimates														
ACE:ARB	10	11	8	16	11	20	0	14	1	1	0	7	2	0

An indirect or mixed treatment effect is a combination of the available direct treatment effects

	ACE: BBlocker	ACE: CCB	ACE: Diuretic	ACE: Placebo	ARB: BBlocker	ARB: CCB	ARB: Diuretic	ARB: Placebo	BBlocker: CCB	BBlocker: Diuretic	BBlocker: Placebo	CCB: Diuretic	CCB: Placebo	Diuretic: Placebo
Mixed estimates														
ACE:BBlocker	32	10	10	8	6	1	0	4	15	6	2	5	2	0
ACE:CCB	10	26	13	11	1	6	()	4	9	1	0	13	6	0
ACE:Diuretic	6	7	57	5	()	2	()	2	1	5	0	12	2	2
ACE:Placebo	5	7	5	56	3	3	()	6	1	()	2	3	8	2
ARB:BBlocker	4	1	()	3	41	21	()	5	19	2	2	2	1	0
ARB:CCB	1	2	1	2	8	67	()	6	8	1	0	2	4	0
ARB:Diuretic	3	2	11	5	10	27	()	8	()	7	0	25	()	2
ARB:Placebo	3	3	2	7	6	15	()	49	()	1	2	2	10	1
BBlocker:CCB	6	4	1	1	11	12	()	()	53	4	2	5	2	0
BBlocker:Diuretic	10	1	13	2	5	3	()	2	19	20	2	21	()	2
BBlocker:Placebo	10	2	2	14	13	3	()	16	16	4	8	1	11	2
CCB:Diuretic	2	6	11	3	1	3	()	2	7	6	0	56	3	2
CCB:Placebo	2	6	4	12	1	15	()	16	6	()	2	5	28	2
Diuretic:Placebo	0	0	20	20	2	7	()	9	0	5	2	17	11	7
Indirect estimates	. <u> </u>													
ACE:ARB	10	11	8	16	11	20	0	14	1	1	0	7	2	0







Indirectness

- Major concerns
- Some concerns
- No concerns

Indirectness

- Considerations similar to those in a pairwise meta-analysis
- How relevant is the study PICO and setting to the research question?
- Score each study at 3 levels
 - Low indirectness to the research question
 - Moderate indirectness to the research question
 - High indirectness to the research question
- Then study-level judgements are summarized within pairwise comparisons and across the network using the contribution matrix exactly as with the Risk of Bias.
- This also addresses the condition of transitivity!
 - If the studies across comparisons have differences in important characteristics (e.g. effect modifiers) compared to the target population, then the transitivity assumption is challenged

Now it is time for....

CINeMA

cinema.ispm.ch

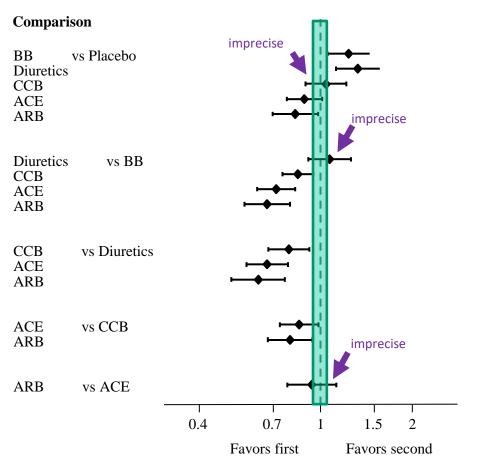
Imprecision

- Major concerns
- Some concerns
- No concerns

Imprecision

- Traditional GRADE considers, among others, the total sample size available and compares it with the Optimal Information Size
- The sample size in a NMA relative effect makes little sense (as studies in the network contribute direct and indirect information!)
- Imprecision relates to the width of the 95% confidence interval:
 - Does the 95% CI include values that lead to different clinical decisions?
- Set a "margin of equivalence"
 - The range of relative treatment effect around the no-effect line that do not signify important differences between the interventions
 - Could be set using the Minimum Clinically Important Difference

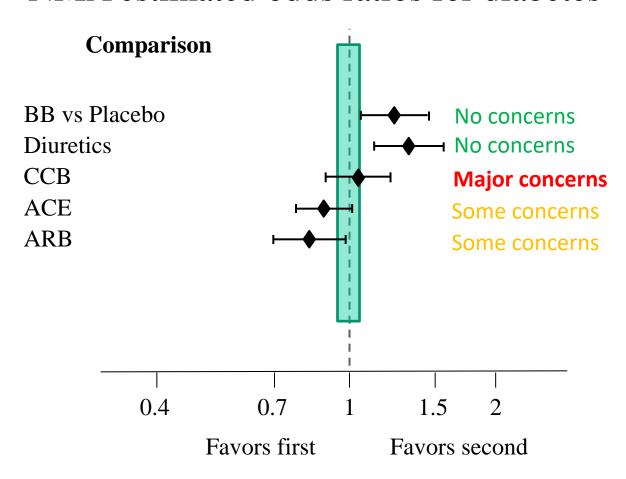
NMA estimated odds ratios for diabetes



Imprecision: Confidence intervals include values that lead into different clinical decisions

Margin of equivalence:
OR=1.05 in either
direction
Imprecision when the
confidence interval
crosses both 0.95 and
1.05

NMA estimated odds ratios for diabetes



Now it is time for....

CINeMA

HETEROGENEITY

- ☐ Major concerns
- Some concerns
- No concerns

INCOHERENCE

- Major concerns
- Some concerns
- No concerns

Heterogeneity

between-study wariance within a comparison evidence

Heterogeneity

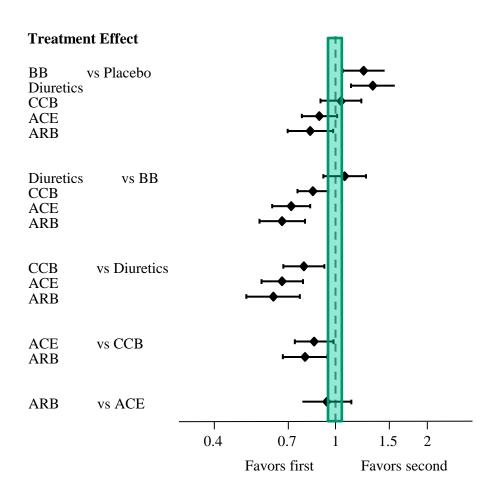
between-study variance within a comparison

Incoherence

disagreement between different sources of evidence

Inconsistency heterogeneity

- The major driver or our decisions is whether the heterogeneity impacts on clinical decisions
- Heterogeneity is represented by the predictive intervals: the intervals within which we expect to find the true effect size of a new study
- They are extensions of the confidence intervals



Treatment Effect

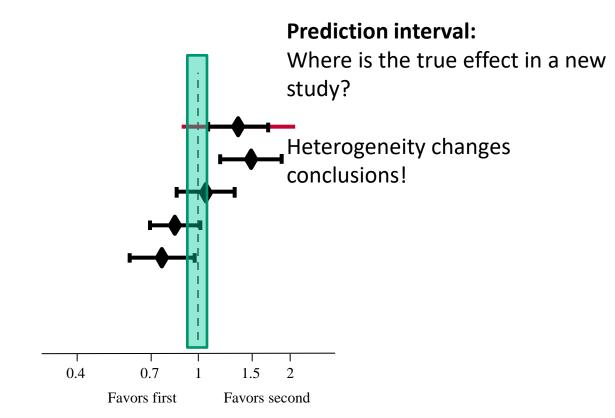
BB vs Placebo

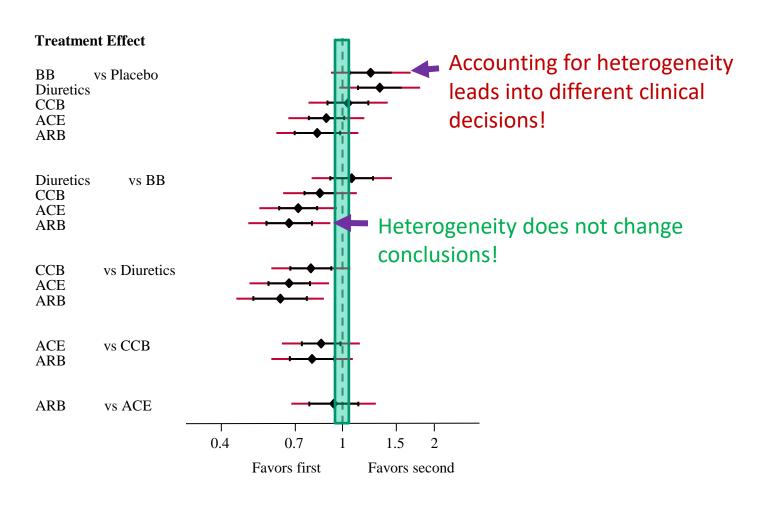
Diuretics

CCB

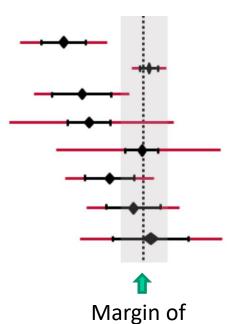
ACE

ARB





Rules implemented in the software



equivalence

No concerns: Confidence and prediction intervals agree in relation to clinically important effect No concerns: Confidence and prediction intervals agree in relation to clinically important effect Some concerns: Prediction interval extends into clinically important or unimportant effects Major concerns: Prediction interval extends into clinically important effects in both directions Major concerns: Prediction interval extends into clinically important effects in both directions No concerns: Confidence and prediction intervals agree in relation to clinically important effects Some concerns: Prediction interval extends into clinically important or unimportant effects No concerns: Confidence and prediction intervals agree in relation to clinically important effects

Inconsistency heterogeneity

- The major driver or our decisions is whether the heterogeneity impacts on clinical decisions
- Heterogeneity is represented by the **predictive intervals**: the intervals within which we expect to find the true effect size of a new study
- They are extensions of the confidence intervals
- Pairwise meta-analysis heterogeneity variances τ^2 can be estimated
 - But their estimation makes sense when you have enough studies
 - The <u>observed values</u> of τ² can be compared with the <u>expected values</u>
 from empirical evidence (*Turner et al Int J Epidemiol. 2012, Rhodes et al. J Clin Epidemiol. 2015*)
 - The expected values depend on the nature of the outcome and the treatments being compared

Inconsistency heterogeneity

Comparison Evidence: mixed	ACE:BBlocker
Between-study heterogodirect comparison	eneity for each
l ² :	49.8%
Estimated τ ² :	0.019
Reference Values for τ^2	
first quantile:	0.003
median:	0.014
third quantile:	0.061
95% intervals for NMA	estimate
Confidence interval:	(1.245,1.498)
Prediction interval:	(0.992,1.879)
Prediction interval extend	ls into clinically
important or unimporta	nt effects
Heterogeneity judgement	Serious 💠

Comparison Evidence: mixed	ARB:BBlocker
Between-study heterog direct comparison	eneity for each
l ² :	NA
Estimated τ^2 :	NA
Reference Values for τ ²	
first quantile:	0.003
median:	0.014
third quantile:	0.061
95% intervals for NMA	estimate
Confidence interval:	(1.372,1.657)
Prediction interval:	(1.094,2.077)
Confidence and predictio	n intervals agree in
relation to clinically impor	tant effect
Heterogeneity judgement	No serious 💠

Comparison Evidence: mixed	BBlocker:CCB
Between-study heterog	eneity for each
l ² :	62.5%
Estimated τ ² :	0.013
Reference Values for τ ²	
first quantile:	0.003
median:	0.014
third quantile:	0.061
95% intervals for NMA	estimate
Confidence interval:	(0.768,0.871)
Prediction interval:	(0.600,1.115)
Prediction interval extend important effects in both	•
Heterogeneity judgement	Very Serious \$

Heterogeneity

between-study variance within a comparison

Incoherence

disagreement between different sources of evidence

We consider prediction intervals for the **impact of heterogeneity** in clinical decision making

Heterogeneity

between-study variance within a comparison

We consider prediction intervals for the **impact of heterogeneity** in clinical decision making

Incoherence

disagreement between different sources of evidence

Separate Direct from Indirect Evidence test (node-splitting)

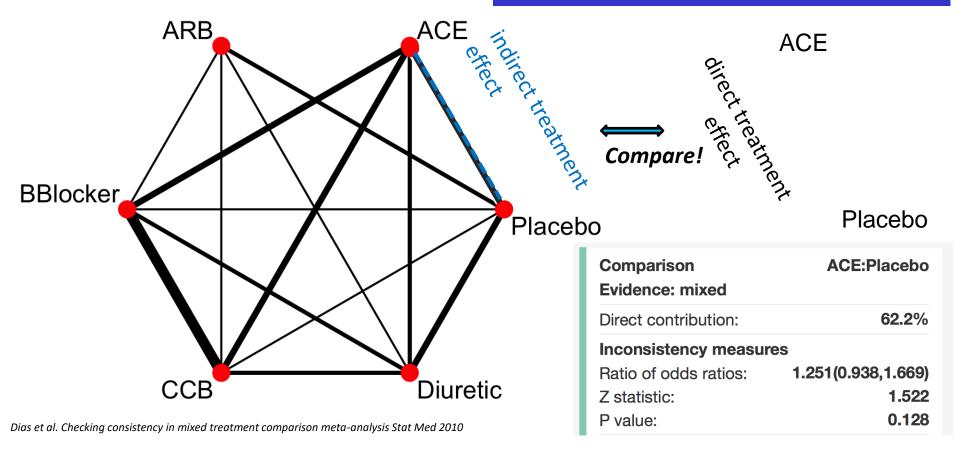
: Compare direct and indirect relative treatment effects using a Z-test

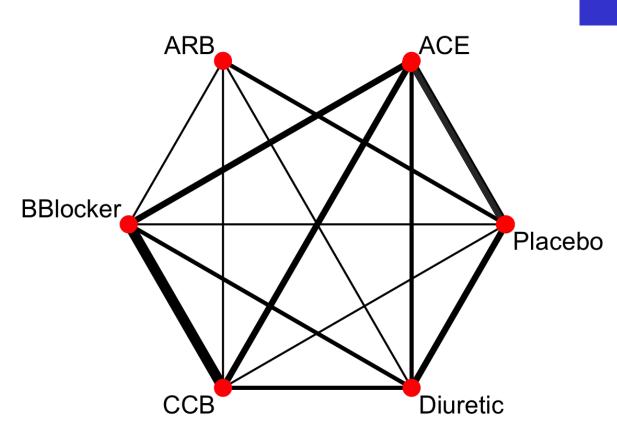
: one test for each treatment comparisons

Design-by-treatment test X²

: one test for the network

Separate Direct from Indirect Evidence test





Design-by-treatment X² test

Does the assumption of coherence hold for the entire network?

$$\chi^2 = 19.325 (13 df)$$

P-value=0.113

White et al. Consistency and inconsistency in network meta-analysis. Res Synth Meth 2012

Treatment comparisons that take at least 90% of the information from direct evidence have no concerns for incoherence

For comparisons with at least 10% of information derived from indirect evidence we use the following rule

Design-by-treatment interaction model

		p-value>0.1	0.01 <p-value<0.1< th=""><th>p-value<0.01</th></p-value<0.1<>	p-value<0.01
	p-value>0.1	No concerns	No concerns	Some concerns
SIDE	0.01 <p-value<0.1< th=""><th>Some concerns</th><th>Some concerns</th><th>Major concerns</th></p-value<0.1<>	Some concerns	Some concerns	Major concerns
	p-value <0.01	Some concerns	Major concerns	Major concerns

Comparison ACE:BBlocker

Evidence: mixed

Direct contribution: 51.4%

Inconsistency measures

Ratio of odds ratios: 0.719(0.533,0.969)

Z statistic: -2.165

P value: 0.030

Incoherence judgement

Some concerns \$

Comparison	ARB:CCB
Evidence: mixed	
Direct contribution:	41.7%
Inconsistency measure	s
Ratio of odds ratios:	1.012(0.709,1.444)
Z statistic:	0.066
P value:	0.948
Incoherence judgement	
No concerns ♣	

Comparison Evidence: mixed	BBlocker:Placebo
Direct contribution:	9.5%
Inconsistency measur	es
Ratio of odds ratios:	0.524(0.299,0.918)
Z statistic:	-2.261
P value:	0.024
Incoherence judgement	
Some concerns \$	

Comparison Evidence: mixed	ACE:CCB
Direct contribution:	41.5%
Inconsistency measure	es
Ratio of odds ratios:	1.099(0.810,1.490)
Z statistic:	0.605
P value:	0.545
Incoherence judgement	
No concerns 💠	

Comparison Evidence: mixed	ARB:Diuretic
Direct contribution:	1.0%
Inconsistency measu Ratio of odds ratios: Z statistic: P value:	res 5.247(0.634,43.445) 1.537 0.124
Incoherence judgement No concerns \$	rt

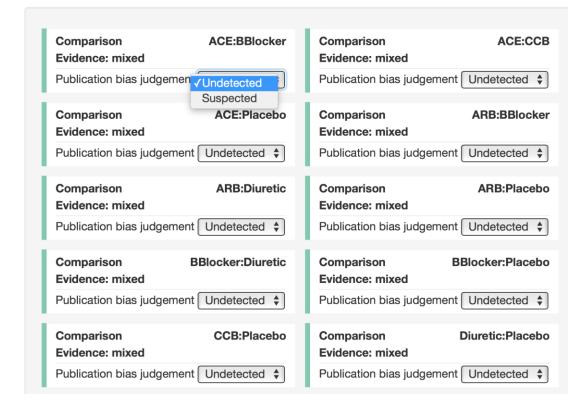
Comparison Evidence: mixed	CCB:Diuretic
Direct contribution:	48.0%
Inconsistency measure	es
Ratio of odds ratios:	0.932(0.676,1.286)
Z statistic:	-0.429
P value:	0.668
Incoherence judgement	
No concerns 💠	

Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis. Res Synth Meth 2012



publication bias

SuspectedUndetected



Now it is time for....

CINeMA

Disclaimer

You are welcome to use CINeMA with the understanding that it is still under development

- We will improve the data input module
- For some calculations CINeMA the netmeta package in R, so updates/debugging in netmeta affect CINeMA too
- Please notify us for any problems you come across <u>cinema.ispm@gmail.com</u>
- If you use it in a publication you can cite

CINeMA: Confidence in Network Meta-Analysis [Software]. University of Bern 2017. Available from cinema.ispm.ch