'Summary of findings' tables in network meta-analysis (NMA)

Juan José Yepes-Nuñez

MD, MSc, PhD candidate

Department of Health Research Methods, Evidence, and Impact (HEI), McMaster University, Canada

Professor, School of Medicine, University of Antioquia, Colombia

Holger J Schünemann

MD, PhD, MSc, FRCP(C)

Chair, Department of Health Research Methods, Evidence, and Impact (*formerly "Clinical Epidemiology and Biostatistics"*)

Professor of Clinical Epidemiology and Medicine | Michael Gent Chair in Healthcare Research

Director, Cochrane Canada and McMaster GRADE Centre. McMaster University, Canada

Outline

Part 1. Learning objective and introduction to NMA

- Objective

- What is an NMA

- Ranking of treatments

- NMA GRADE certainty in evidence assessment

- Summary of Findings (SoF) tables in Systematic Reviews and Meta-analysis

Outline

Part 2. NMA-SoF table
Introduction to the NMA-SoF table project

Part 3. NMA-SoF table examples
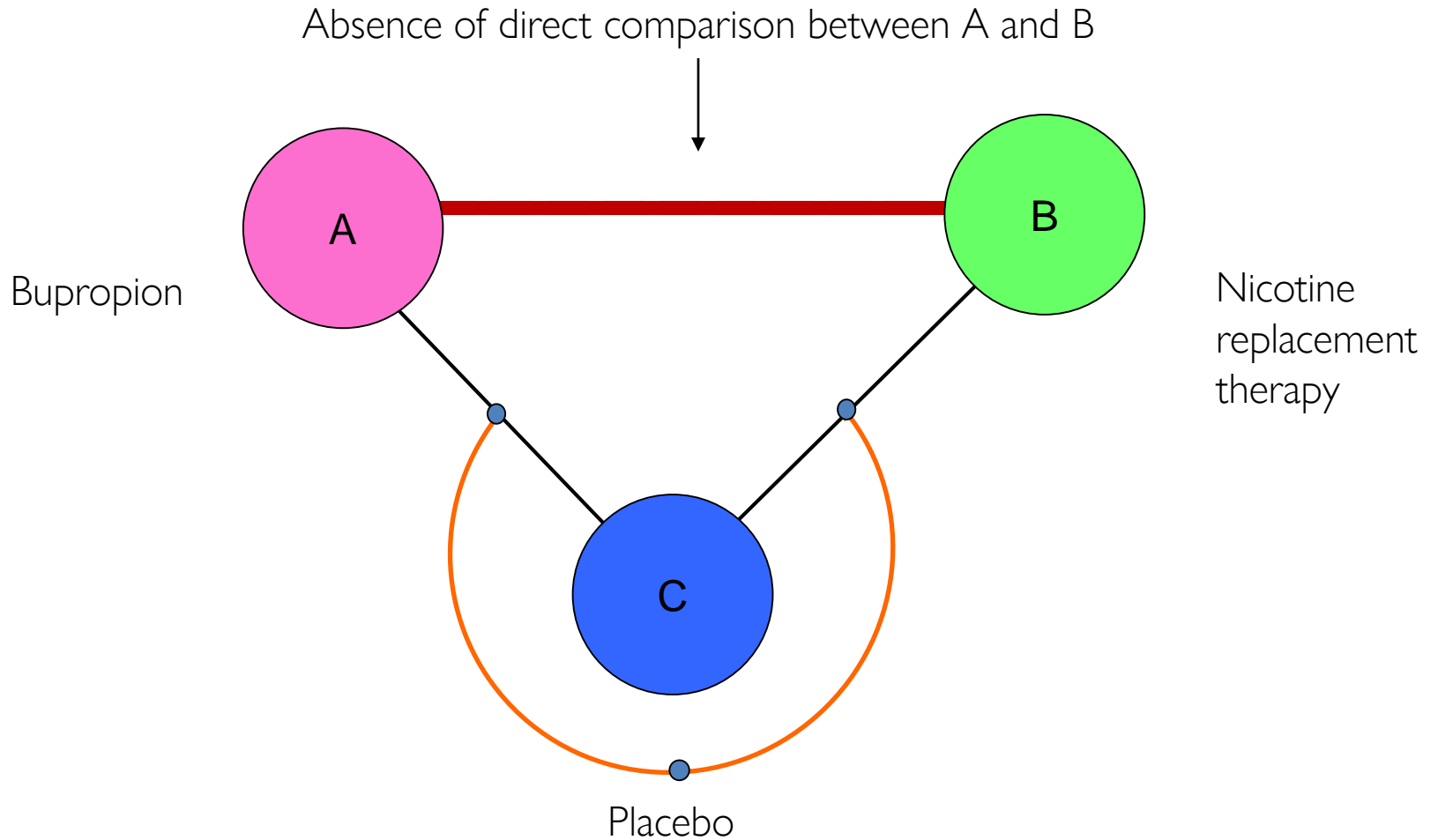
Part 4. Q&A

Part 1

LEARNING OBJECTIVE AND INTRODUCTION TO NMA
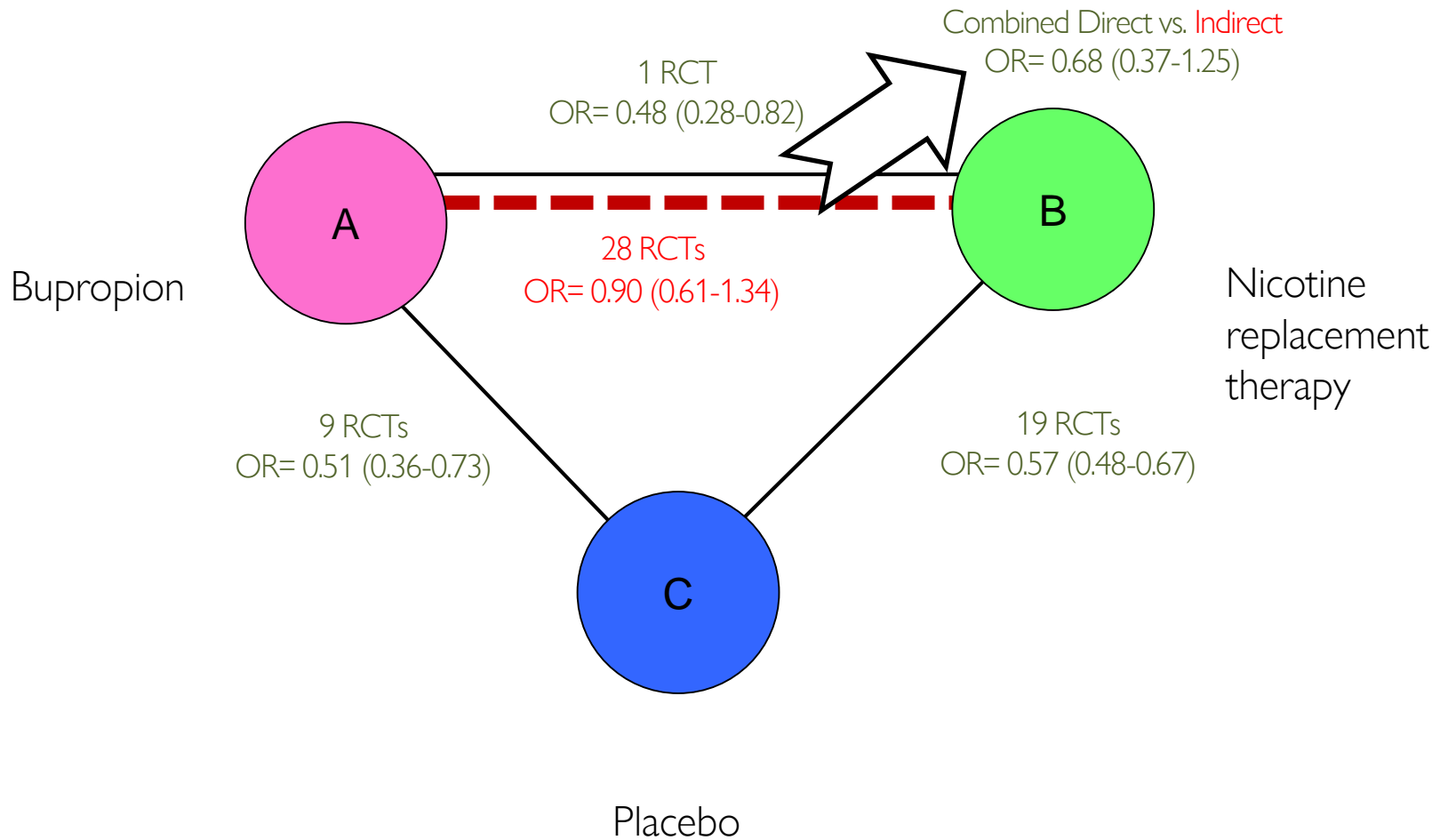
Learning objective

- To gain familiarity in interpreting findings of network meta-analysis (NMA) through NMA 'Summary of findings' (SoF) tables developed based on principles of the **GRADE** approach to rating certainty of evidence from NMAs
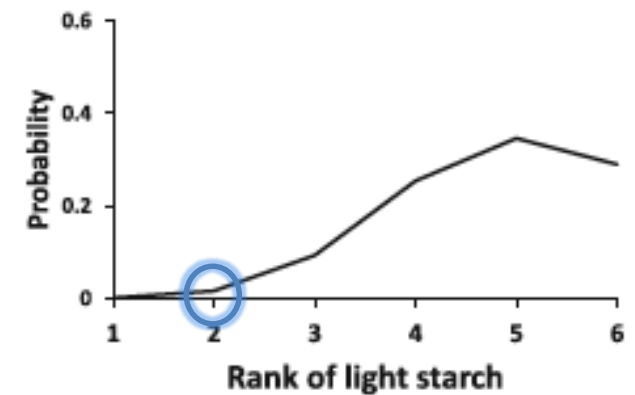
WHAT IS AN NMA?

Introduction to NMA

Absence of direct comparison between A and B

Bupropion

Nicotine replacement therapy

Placebo

Introduction to NMA

Combined Direct vs. Indirect
OR= 0.68 (0.37-1.25)

1 RCT
OR= 0.48 (0.28-0.82)

A

B

Bupropion

28 RCTs
OR= 0.90 (0.61-1.34)

Nicotine replacement therapy

9 RCTs
OR= 0.51 (0.36-0.73)

19 RCTs
OR= 0.57 (0.48-0.67)

C

Placebo

WHAT ARE RANKING TREATMENTS?

Ranking Treatments

Graphical
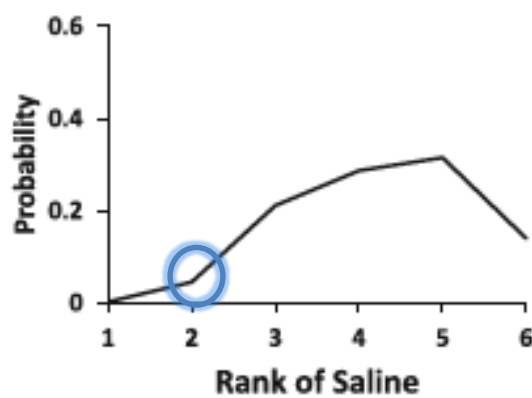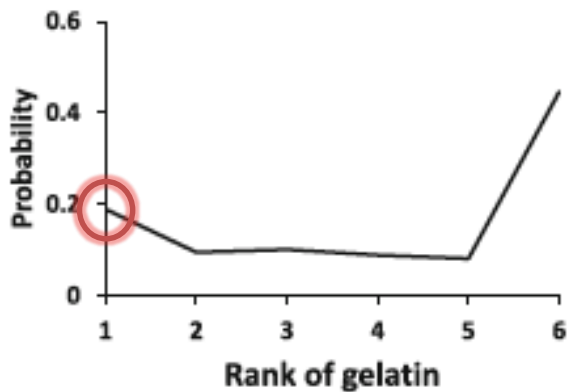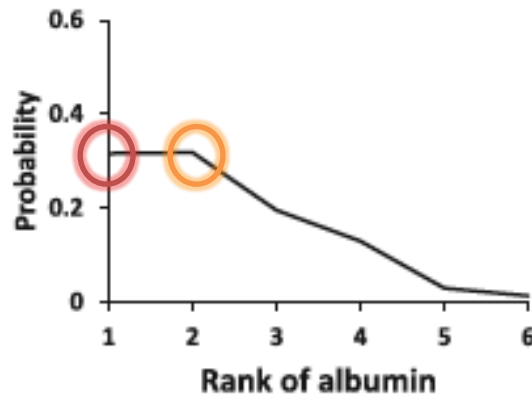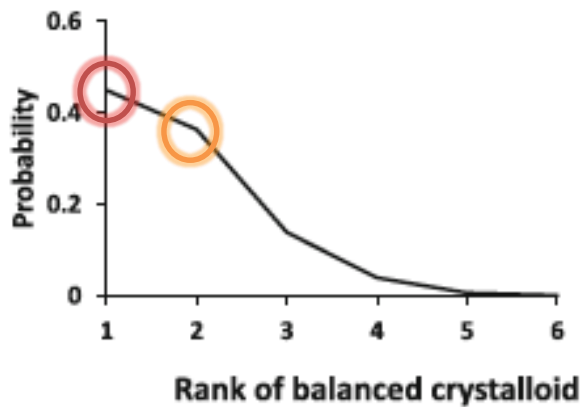
Ranking Treatments

Numerical

SUCRA

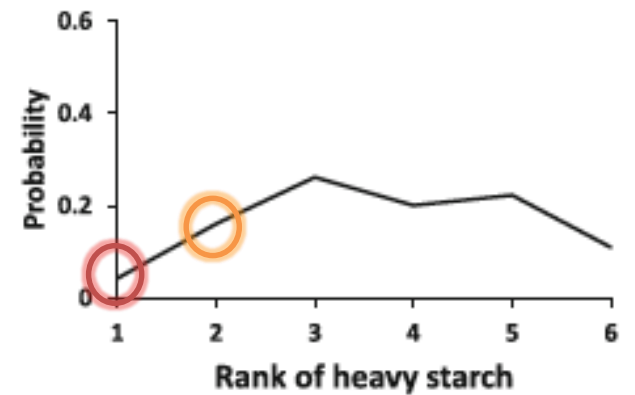| Rank	Treatment	SUCRA
1	Balanced crystalloid	84.1%
2	Albumin	74.5%
3	Heavy starch	45.4%
4	Gelatin	37.7%
5	Saline	34.2%
6	Light starch	24.0%

SUCRA surface under the cumulative ranking curve

Median and 95% CrI for the rank of each treatment

Nodes and Fluids   **Median ranks (95%CrI)**

Six node analysis

Treatment	Median ranks (95%CrI)
Balanced crystalloid	2.00 (1.00, 4.00)
Albumin	2.00 (1.00, 5.00)
Heavy starch	4.00 (2.00, 6.00)
Gelatin	4.00 (1.00, 6.00)
Saline	5.00 (3.00, 6.00)
Light starch	5.00 (1.00, 6.00)
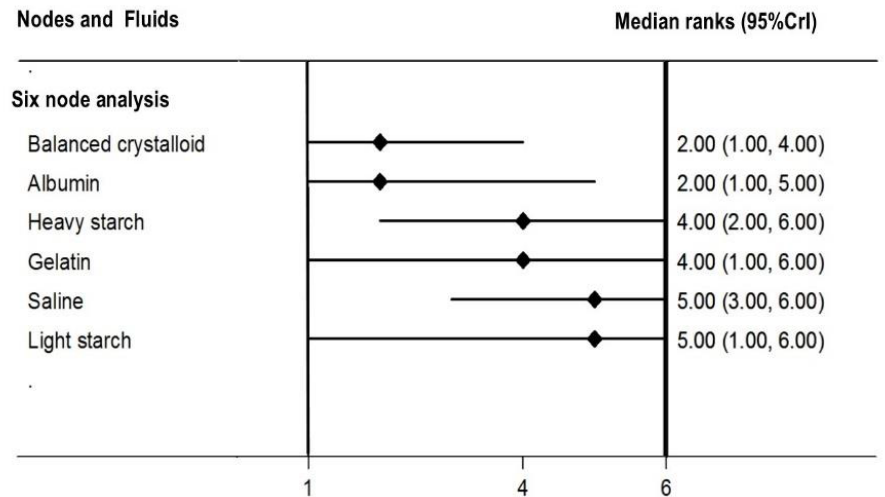
HOW TO ASSESS NMA CERTAINTY (QUALITY) IN EVIDENCE WITH **GRADE**?

- Grading system in health-care to assess the quality (or certainty) of evidence and strength of recommendations

Systematic Reviews

Clinical practice guidelines

Determinants of certainty in a body of evidence GRADE

- A body of evidence starts as: high | ⊕⊕⊕⊕

- 5 factors that can lower quality

  1. Risk of bias criteria

     - Lack of randomization (non-randomized or observational studies) lowers confidence to low

  2. Inconsistency (or heterogeneity)

  3. Indirectness (PICO and applicability)
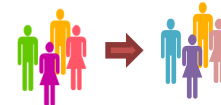
  4. Imprecision

  5. Publication bias

Determinants of certainty in a body of evidence: GRADE

- **3 factors can increase quality**
  1. large magnitude of effect
  2. opposing plausible residual bias or confounding
  3. dose-response gradient

NMA certainty in evidence

High certainty and *direct* evidence contributes as much as indirect evidence

Rate CiE *direct* estimates	Rate CiE *indirect* estimates	Rate CiE *NMA* estimates
• Risk of Bias • Inconsistency • Indirectness • Publication bias	• Transitivity • Lowest of the ratings of the two direct comparisons forming the most dominant first-order loop	• Incoherence • Imprecision
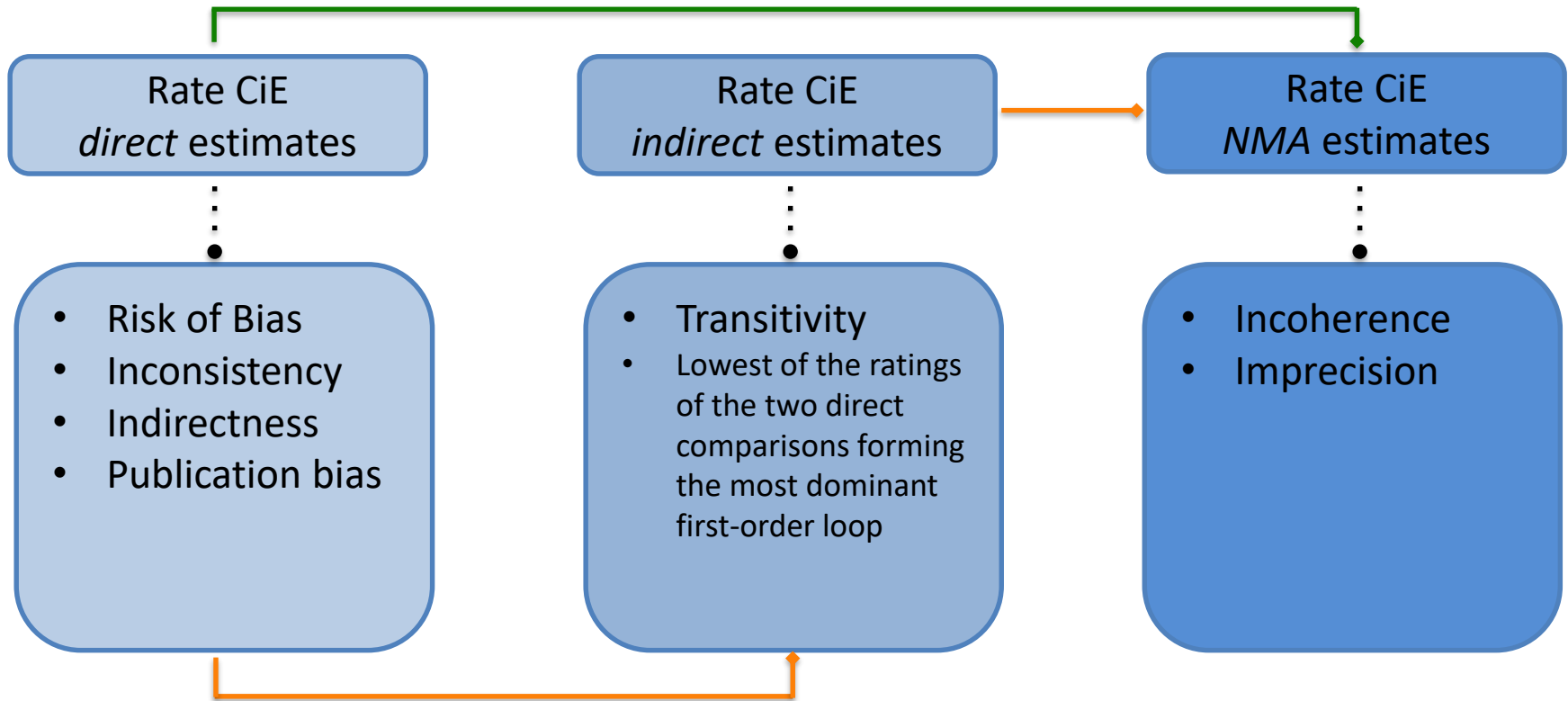
Not sufficient evidence, moderate, low or very low certainty

SUMMARY of FINDINGS (SoF) TABLES IN SYSTEMATIC REVIEWS AND META-ANALYSES

SoF tables in Systematic Reviews and Meta-analysis

Elements of a GRADE SoF table

Probiotics compared to no probiotics in INFANTS for the prevention of allergies

Patient or population: INFANTS for the prevention of allergies
Setting: outpatient
Intervention: probiotics
Comparison: no probiotics

Outcome № of participants (studies)	Relative effect (95% CI)	Anticipated absolute effects (95% CI)		Difference	Certainty	What happens
Asthma / wheezing - infants follow up: range 6 to 24 months to № of participants: 412 (3 RCTs)	**RR 1.04** (0.63 to 1.70)	12.1%	**12.6%** (7.6 to 20.6)	**0.5% more** (4.5 fewer to 8.5 more)	⊕◯◯◯ VERY LOW a,b,c	
Adverse effects follow up: range 6 to 24 months to № of participants: 187 (2 RCTs)	**RR 1.27** (0.51 to 3.18)	53.2%	**67.6%** (27.1 to 100.0)	**14.4% more** (26.1 fewer to 116 more)	⊕◯◯◯ VERY LOW b,c,d	

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio; **SMD:** Standardised mean difference

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

a. Concerns with high risk of bias for allocation concealment and blinding. Unclear risk of bias in random allocation and adequate follow-up
b. Clinical heterogeneity due to high risk vs average risk of allergies and different probiotics among studies.
c. Confidence interval does not exclude appreciable benefit or harm
d. One study with unclear description of randomization process, allocation concealment, blinding and follow-up

SoF tables in Systematic Reviews and Meta-analysis

Elements of a GRADE SoF table

Probiotics compared to no probiotics in INFANTS for the prevention of allergies

Bibliography: WAO systematic review

Outcomes	№ of participants (studies) Follow-up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with no probiotics	Risk difference with probiotics
Asthma / wheezing - infants follow up: range 6 to 24 months to	412 (3 RCTs)	⊕◯◯◯ VERY LOW a,b,c	RR 1.04 (0.63 to 1.70)	121 per 1,000	5 more per 1,000 (45 fewer to 85 more)
Adverse effects follow up: range 6 to 24 months to	187 (2 RCTs)	⊕◯◯◯ VERY LOW b,c,d	RR 1.27 (0.51 to 3.18)	532 per 1,000	144 more per 1,000 (261 fewer to 1,160 more)

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the __relative effect__ of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio; **SMD:** Standardised mean difference

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

a. Concerns with high risk of bias for allocation concealment and blinding. Unclear risk of bias in random allocation and adequate follow-up
b. Clinical heterogeneity due to high risk vs average risk of allergies and different probiotics among studies.
c. Confidence interval does not exclude appreciable benefit or harm
d. One study with unclear description of randomization process, allocation concealment, blinding and follow-up

SoF tables in Systematic Reviews and Meta-analysis

Elements of a **GRADE** SoF table

Probiotics compared to no probiotics in INFANTS for the prevention of allergies

Patient or population: INFANTS for the prevention of allergies
Setting: outpatient
Intervention: probiotics
Comparison: no probiotics

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with no probiotics	**Risk with probiotics**				
Asthma / wheezing - infants follow up: range 6 to 24 months to	121 per 1,000	**126 per 1,000** (76 to 206)	**RR 1.04** (0.63 to 1.70)	412 (3 RCTs)	⊕◯◯◯ VERY LOW a,b,c	
Adverse effects follow up: range 6 to 24 months to	532 per 1,000	**676 per 1,000** (271 to 1,000)	**RR 1.27** (0.51 to 3.18)	187 (2 RCTs)	⊕◯◯◯ VERY LOW b,c,d	

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio; **SMD:** Standardised mean difference

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

a. Concerns with high risk of bias for allocation concealment and blinding. Unclear risk of bias in random allocation and adequate follow-up
b. Clinical heterogeneity due to high risk vs average risk of allergies and different probiotics among studies.
c. Confidence interval does not exclude appreciable benefit or harm
d. One study with unclear description of randomization process, allocation concealment, blinding and follow-up

Part 2

NMA-SOF TABLE

Introduction to the NMA-SoF table project

NMA GRADE SoF table format

NMA-SoF TABLE: WHY?

Introduction NMA-SoF table project

- No standardized Network metanalysis (NMA) Summary of Findings (SoF) table format

Presentational approaches used in the UK for reporting evidence synthesis using indirect and mixed treatment comparisons

Sze Huey Tan[1], Sylwia Bujkiewicz[2], Alexander Sutton[3], Pascale Dequen[4] and Nicola Cooper[5]

Reporting of results from network meta-analyses: methodological systematic review

OPEN ACCESS

Aïda Bafeta *PhD student*[1], Ludovic Trinquart *postdoctoral research fellow*[1234], Raphaèle Seror *associate professor of rheumatology*[13], Philippe Ravaud *professor of epidemiology and director*[1234]

What Guidance Are Researchers Given on How to Present Network Meta-Analyses to End-Users such as Policymakers and Clinicians? A Systematic Review

Shannon M. Sullivan[1]*, Doug Coyle[2], George Wells[1,2]

1. University of Ottawa Heart Institute, Ottawa, Ontario, Canada, 2. University of Ottawa, Department of Epidemiology and Community Medicine, Ottawa, Ontario, Canada

CrossMark

Characteristics and knowledge synthesis approach for 456 network meta-analyses: a scoping review

Wasifa Zarin[1], Areti Angeliki Veroniki[1], Vera Nincic[1], Afshin Vafaei[1], Emily Reynen[1], Sanober S. Motiwala[1], Jesmin Antony[1], Shannon M. Sullivan[1], Patricia Rios[1], Caitlin Daly[1], Joycelyne Ewusie[1], Maria Petropoulou[2], Adriani Nikolakopoulou[2,3], Anna Chaimani[2], Georgia Salanti[2,3,4], Sharon E. Straus[1,5] and Andrea C. Tricco[1,6]*

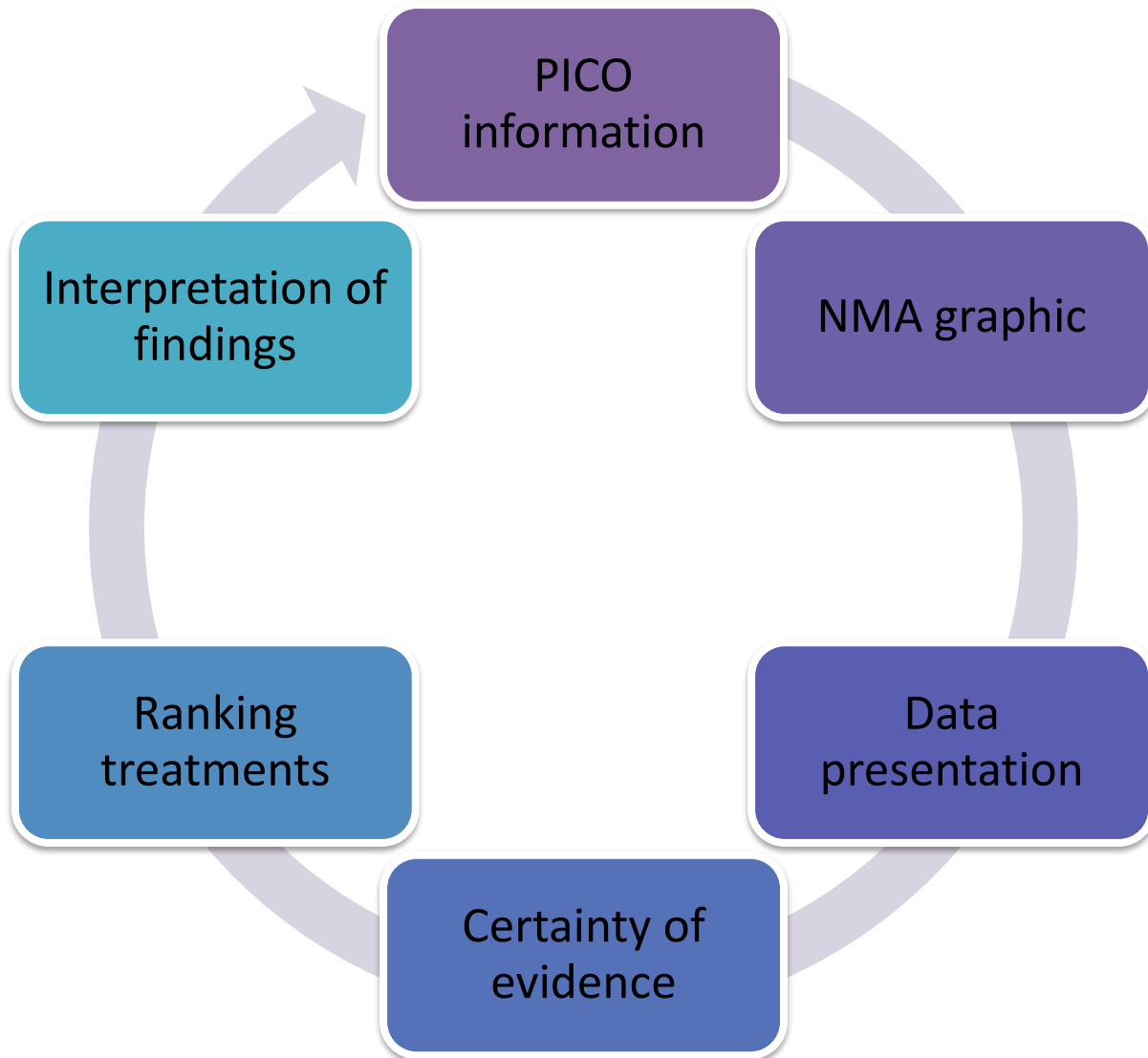# Introduction NMA-SoF table project

Initial development

Brainstorm meeting .... Input from interviews, and Input from advisory group

New version NMA-SoF table

Round 1 and 2

Input from interviews, and Input from advisory group

New version NMA-SoF table

Round 3

Brainstorm meeting .... Input from interviews, and Input from advisory group

New version NMA-SoF table

Round 4

Input from interviews, and Input from advisory group

FINAL version NMA-SoF table

WHAT IS THE OPTIMAL PRESENTATION OF RESULTS OF NMA REPORTS?

NMA-SoF TABLE FORMAT

NMA-SoF table example 1

Estimates of effects, credible intervals, and certainty of the evidence for comparison fluid resuscitation in patients with sepsis
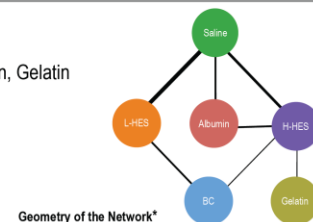
Bayesian NMA-SoF table

Patient or population: Critically ill patients with severe sepsis or septic shock

Interventions: Balanced crystalloid (BC), Albumin, High-molecular-weight hydroxyethyl starch (H-HES), Saline solution, Gelatin

Comparator (reference): Low-molecular weight hydroxyethyl starch (L- HES)

Outcome: Mortality; range of follow up between 24 hours to 90 days

Setting(s): Inpatient

Geometry of the Network*

Total studies: 6 RCT Total Participants: 8308		Relative effect** (95% CrI)	Anticipated absolute effect*** (95% CrI)			Certainty of evidence	Ranking**** (95% CrI)	Interpretation of Findings
			Without intervention	*With* intervention	Difference			
●	Balanced crystalloid (2 RCT; 846 participants)	**0.75** (0.58 to 0.97) Network estimate	180 per 1000[1]	141 per 1000	39 per 1000 fewer (from 67 fewer to 5 fewer)	⊕⊕⊕◯ **Moderate** Due to Indirectness[2]	**2.00** (1.00 to 4.00)	Probably superior
●	Albumin (No direct evidence, Indirect evidence only)	**0.79** (0.59 to 1.06) Network estimate	180 per 1000[1]	148 per 1000	32 per 1000 fewer (from 65 fewer to 88 more)	⊕⊕◯◯ **Low** Due to Imprecision[3], and Indirectness[4]	**2.00** (1.00 to 5.00)	Probably inferior
●	H-HES (No direct evidence, Indirect evidence only)	**0.91** (0.63 to 1.33) Network estimate	180 per 1000[1]	164 per 1000	16 per 1000 fewer (from 59 fewer to 46 more)	⊕⊕◯◯ **Low** Due to Imprecision[3], and Indirectness[4]	**4.00** (2.00 to 6.00)	Probably superior
●	Saline solution (4 RCT; 7642 participants)	**1.04** (0.87 to 1.25) Network estimate	180 per 1000[1]	186 per 1000	6 per 1000 more (from 20 fewer to 35 more)	⊕⊕⊕◯ **Moderate** Due to Imprecision[4], Indirectness[6], and Inconsistency[5]	**4.00** (1.00 to 6.00)	Probably superior
●	Gelatin (No direct evidence, Indirect evidence only)	**1.00** (0.44 to 2.21) Network estimate	180 per 1000[1]	180 per 1000	0 per 1000 fewer (from 92 fewer to 146 more)	⊕◯◯◯ **Very Low** Due to Imprecision[3], and Indirectness[2]	**5.00** (3.00 to 6.00)	Definitely inferior
●	L-HES	Reference Comparator	No estimable	No estimable	No estimable	Reference Comparator	**5.00** (1.00 to 6.00)	Reference comparator

NMA-SoF table definitions

* Solid lines represent direct comparisons

** Network Metanalysis (NMA) estimates are reported as odds ratio. CrI: credible interval. Results are expressed in credible intervals as opposed to the confidence intervals (CI) since a Bayesian analysis has been conducted.

*** Anticipated absolute effect. Anticipated absolute effect compares two risks by calculating the difference between the risk of the intervention group with the risk of the control group.

**** Median and credible intervals are presented. Rank statistics is defined as the probabilities that a treatment out of *n* treatments in a network meta-analysis is the best, the second, the third and so on until the least effective treatment.

† Information is reported from studies included in the network metanalysis for the comparison displays.

GRADE Working Group grades of evidence (or certainty in the evidence)

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanatory Footnotes

[1] Mortality is reported from a large randomized control trail where critically ill patients admitted to an intensive care unit (ICU) required fluid resuscitation with hydroxyethyl starch (HES).

[2] Serious indirectness. The indirect evidence for this comparison goes through a second order loop via heavy starch and saline.

[3] Serious imprecision. Due to wide confidence intervals in the indirect estimate.

[4] Serious indirectness. The indirect evidence for this comparison goes through a first order loop via saline and saline vs. light starch.

[5] Serious inconsistency. Due to there was significant heterogeneity in the direct comparison of light starch vs. balanced crystalloid.

[6] Serious indirectness. The indirect evidence for this comparison goes through a second order loop via balance crystalloid and heavy starch.

NMA-SoF table example 1

Estimates of effects, credible intervals, and certainty of the evidence for comparison fluid resucitation in patients with sepsis
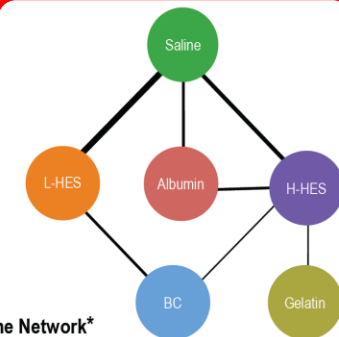
Bayesian NMA-SoF table

Patient or population: Critically ill patients with severe sepsis or septic shock

Interventions: Balanced crystalloid (BC), Albumin, High-molecular-weight hydroxyethyl starch (H-HES), Saline solution, Gelatin

Comparator (reference): Low-molecular weight hydroxyethyl starch (L-HES)

Outcome: Mortality; range of follow up between 24 hours to 90 days

Setting(s): Inpatient

Geometry of the Network*

Total studies: 6 RCT Total Participants: 8308		Relative effect** (95% CrI)	Anticipated absolute effect*** (95% CrI)			Certainty of evidence	Ranking**** (95% CrI)	Interpretation of Findings
			Without intervention	*With* intervention	Difference			
●	Balanced crystalloid (2 RCT; 846 participants)	**0.75** (0.58 to 0.97) Network estimate	180 per 1000[1]	141 per 1000	39 per 1000 fewer (from 67 fewer to 5 fewer)	⊕⊕⊕◯ **Moderate** Due to Indirectness[2]	**2.00** (1.00 to 4.00)	Probably superior
●	Albumin (No direct evidence, Indirect evidence only)	**0.79** (0.59 to 1.06) Network estimate	180 per 1000[1]	148 per 1000	32 per 1000 fewer (from 65 fewer to 88 more)	⊕⊕◯◯ **Low** Due to Imprecision[3], and Indirectness[4]	**2.00** (1.00 to 5.00)	Probably inferior
●	H-HES (No direct evidence, Indirect evidence only)	**0.91** (0.63 to 1.33) Network estimate	180 per 1000[1]	164 per 1000	16 per 1000 fewer (from 59 fewer to 46 more)	⊕⊕◯◯ **Low** Due to Imprecision[3], and Indirectness[4]	**4.00** (2.00 to 6.00)	Probably superior

NMA-SoF table example 1

●	Saline solution (4 RCT; 7642 participants)	**1.04** (0.87 to 1.25) Network estimate	180 per 1000[1]	186 per 1000	6 per 1000 more (from 20 fewer to 35 more)	⊕⊕⊕○ **Moderate** Due to Imprecision[4], Indirectness[6], and Inconsistency[5]	**4.00** (1.00 to 6.00)	Probably superior
●	Gelatin (No direct evidence, Indirect evidence only)	**1.00** (0.44 to 2.21) Network estimate	180 per 1000[1]	180 pe r 1000	0 per 1000 fewer (from 92 fewer to 146 more)	⊕○○○ **Very Low** Due to Imprecision[3], and Indirectness[2]	**5.00** (3.00 to 6.00)	Definitely inferior
●	L-HES	Reference Comparator	No estimable	No estimable	No estimable	Reference Comparator	**5.00** (1.00 to 6.00)	Reference comparator

NMA-SoF table definitions

* Solid lines represent direct comparisons

** Network Metanalysis (NMA) estimates are reported as odds ratio. CrI: credible interval. Results are expressed in credible intervals as opposed to the confidence intervals (CI) since a Bayesian analysis has been conducted.

*** Anticipated absolute effect. Anticipated absolute effect compares two risks by calculating the difference between the risk of the intervention group with the risk of the control group.

**** Median and credible intervals are presented. Rank statistics is defined as the probabilities that a treatment out of *n* treatments in a network meta-analysis is the best, the second, the third and so on until the least effective treatment.

† Information is reported from studies included in the network metanalysis for the comparison displays.

GRADE Working Group grades of evidence (or certainty in the evidence)

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanatory Footnotes

[1] Mortality is reported from a large randomized control trail where critically ill patients admitted to an intensive care unit (ICU) required fluid resuscitation with hydroxyethyl starch (HES).

[2] Serious indirectness. The indirect evidence for this comparison goes through a second order loop via heavy starch and saline.

[3] Serious imprecision. Due to wide confidence intervals in the indirect estimate.

[4] Serious indirectness. The indirect evidence for this comparison goes through a first order loop via saline and saline vs. light starch.

[5] Serious inconsistency. Due to there was significant heterogeneity in the direct comparison of light starch vs. balanced crystalloid.

[6] Serious indirectness. The indirect evidence for this comparison goes through a second order loop via balance crystalloid and heavy starch.

NMA-SoF table example 2

Estimates of effects, credible intervals, and certainty of the evidence for chemoprevention of colorectal cancer in individuals with previous colorectal neoplasia

Bayesian NMA-SoF table

BENEFITS

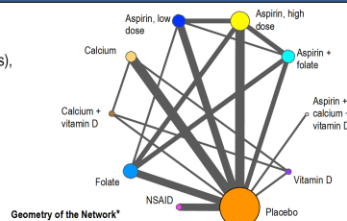**Patient or population:** Individuals with previous colorectal neoplasia

Interventions: Low and high dose aspirin, nonaspirin non-steroidal anti-inflammatory drugs (NSAIDs), calcium, vitamin D, folic acid

Comparator (reference): Placebo

Outcome: Prevention of advanced neoplasia; range of follow up between three to five years

Setting: Outpatient

Geometry of the Network*

Total studies: 21 RCT Total Participants: 12088	Relative effect** (95% CrI)	Anticipated absolute effect*** (95% CrI)			Certainty of evidence	Ranking**** (95% CrI)	Interpretation of Findings
		Without intervention	With intervention	Difference			
Aspirin + calcium + vitamin D (1 RCT; 427 participants)	0.71 (0.18 to 2.49) Network estimate	74 per 1000[1]	53 per 1000	21 fewer per 1000 (61 fewer to 110 more)	⊕⊕◯◯ Low Due to Imprecision[2,3]	3 (1 to 10)	Probably inferior
Calcium + vitamin D (1 RCT; 1028 participants)	0.91 (0.52 to 1.63) Network estimate	74 per 1000[1]	67 per 1000	7 fewer per 1000 (36 fewer to 47 more)	⊕⊕◯◯ Low Due to Imprecision[2,3]	6 (1 to 10)	Probably inferior
Aspirin + folate (2 RCT; 916 participants)	0.73 (0.43 to 1.19) Network estimate	74 per 1000[1]	54 per 1000	20 fewer per 1000 (42 fewer to 14 more)	⊕⊕◯◯ Low Due to Imprecision[2,3]	4 (2 to 8)	Probably inferior
Aspirin, high dose (3 RCT; 917 participants)	0.81 (0.50 to 1.28) Network estimate	74 per 1000[1]	60 per 1000	14 fewer per 1000 (37 fewer to 21 more)	⊕⊕◯◯ Low Due to Imprecision[2,3]	5 (2 to 9)	Probably inferior
Aspirin, low dose (3 RCT; 823 participants)	0.71 (0.41 to 1.23) Network estimate	74 per 1000[1]	53 per 1000	21 fewer per 1000 (44 fewer to 17 more)	⊕⊕◯◯ Low Due to Imprecision[2,3]	3 (2 to 9)	Probably inferior
Nonaspirin NSAIDs (4 RCT; 3486 participants)	0.37 (0.24 to 0.53) Network estimate	74 per 1000[1]	27 per 1000	47 fewer per 1000 (56 fewer to 35 fewer)	⊕⊕⊕⊕ High[1]	1 (1 to 2)	Definitely superior
Vitamin D (1 RCT; 764 participants)	1.19 (0.65 to 2.15) Network estimate	74 per 1000[1]	88 per 1000	14 more per 1000 (26 fewer to 85 more)	⊕⊕◯◯ Low Due to Imprecision[2,3]	9 (3 to 10)	Probably inferior
Calcium (3 RCT; 2503 participants)	1.00 (0.66 to 1.52) Network estimate	74 per 1000[1]	74 per 1000	0 fewer per 1000 (25 fewer to 38 more)	⊕⊕◯◯ Low Due to Imprecision[4,5]	7 (3 to 10)	Probably inferior
Folate (3 RCT; 1224 participants)	1.32 (0.85 to 2.00) Network estimate	74 per 1000[1]	51 per 1000	23 more per 1000 (11 fewer to 74 more)	⊕⊕◯◯ Low Due to Imprecision[2,3]	9 (5 to 10)	Probably inferior
Placebo	Reference comparator	No estimable	No estimable	No estimable	Reference comparator	7 (4 to 9)	Reference comparator

NMA-SoF table definitions
* Lines represent direct comparisons
** Estimates are reported as odds ratio. CrI: credible interval. Results are expressed in credible intervals as opposed to the confidence intervals (CI) since a Bayesian analysis has been conducted.
*** Anticipated absolute effect. Anticipated absolute effect compares two risks by calculating the difference between the risks of the intervention group with the risk of the control group.
**** Surface under the cumulative (SUCRA) ranking and credible intervals for efficacy are presented. Rank statistics is defined as the probabilities that a treatment out of n treatments in a network meta-analysis is the best, the second, the third and so on until the least effective treatment.

GRADE Working Group grades of evidence (or certainty in the evidence)
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanatory Footnotes
[1] Baseline risks (assumed control risk) obtained from the National Cancer Institute pooling project
[2] Very serious imprecision since 95% CrI crosses unity, and with wide credible intervals suggesting high possibility of harm.
[3] Very serious imprecision since RR>1 (suggesting greater likelihood of harm than benefit), and with wide credible intervals).
[4] Very serious imprecision since RR is one (suggesting no evidence of benefit) and wide credible intervals suggesting high possibility of harm.
[5] Conceptually, there was no significant intransitivity, with comparable distribution of plausible effect modifiers across trials of different chemopreventive agents.

31

NMA-SoF table example 2

Estimates of effects, credible intervals, and certainty of the evidence for chemoprevention of colorectal cancer in individuals with previous colorectal neoplasia
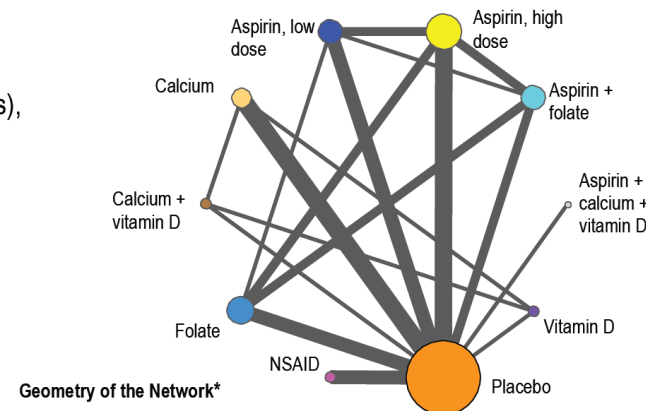
BENEFITS

Patient or population: Individuals with previous colorectal neoplasia

Interventions: Low and high dose aspirin, nonaspirin non-steroidal anti-inflammatory drugs (NSAIDs), calcium, vitamin D, folic acid

Comparator (reference): Placebo

Outcome: Prevention of advanced neoplasia; range of follow up between three to five years

Setting: Outpatient

Geometry of the Network*

Total studies: 21 RCT Total Participants: 12088		Relative effect** (95% CrI)	Anticipated absolute effect*** (95% CrI)			Certainty of evidence	Ranking**** (95% CrI)	Interpretation of Findings
			Without intervention	*With* intervention	Difference			
⚪	Aspirin + calcium + vitamin D (1 RCT; 427 participants)	**0.71** (0.18 to 2.49) Network estimate	74 per 1000[1]	53 per 1000	21 fewer per 1000 (61 fewer to 110 more)	⊕⊕◯◯ **Low** Due to Imprecision[2,5]	3 (1 to 10)	Probably inferior
🟤	Calcium + vitamin D (1 RCT; 1028 participants)	**0.91** (0.52 to 1.63) Network estimate	74 per 1000[1]	67 per 1000	7 fewer per 1000 (36 fewer to 47 more)	⊕⊕◯◯ **Low** Due to Imprecision[2,5]	6 (1 to 10)	Probably inferior
🔵	Aspirin + folate (2 RCT; 916 participants)	**0.73** (0.43 to 1.19) Network estimate	74 per 1000[1]	54 per 1000	20 fewer per 1000 (42 fewer to 14 more)	⊕⊕◯◯ **Low** Due to Imprecision[2,5]	4 (2 to 8)	Probably inferior
🟡	Aspirin, high dose (3 RCT; 917 participants)	**0.81** (0.50 to 1.28) Network estimate	74 per 1000[1]	60 per 1000	14 fewer per 1000 (37 fewer to 21 more)	⊕⊕◯◯ **Low** Due to Imprecision[2,5]	5 (2 to 9)	Probably inferior

NMA-SoF table example 2

	Intervention	Relative effect (95% CrI)**	Assumed risk (control)	Corresponding risk (intervention)	Anticipated absolute effect***	Certainty of evidence (GRADE)	SUCRA rank****	Interpretation
● (blue)	Aspirin, low dose (3 RCT; 823 participants)	**0.71** (0.41 to 1.23) Network estimate	74 per 1000[1]	53 per 1000	21 fewer per 1000 (44 fewer to 17 more)	⊕⊕○○ **Low** Due to Imprecision[2, 5]	3 (2 to 9)	Probably inferior
● (magenta)	Nonaspirin NSAIDs (4 RCT; 3486 participants)	**0.37** (0.24 to 0.53) Network estimate	74 per 1000[1]	27 per 1000	47 fewer per 1000 (56 fewer to 35 fewer)	⊕⊕⊕⊕ **High**[, 5]	1 (1 to 2)	Definitely superior
● (purple)	Vitamin D (1 RCT; 764 participants)	**1.19** (0.65 to 2.15) Network estimate	74 per 1000[1]	88 per 1000	14 more per 1000 (26 fewer to 85 more)	⊕⊕○○ **Low** Due to Imprecision[3, 5]	9 (3 to 10)	Probably inferior
● (tan)	Calcium (3 RCT; 2503 participants)	**1.00** (0.66 to1.52) Network estimate	74 per 1000[1]	74 per 1000	0 fewer per 1000 (25 fewer to 38 more)	⊕⊕○○ **Low** Due to Imprecision[4, 5]	7 (3 to 10)	Probably inferior
● (blue)	Folate (3 RCT; 1224 participants)	**1.32** (0.85 to 2.00) Network estimate	74 per 1000[1]	51 per 1000	23 more per 1000 (11 fewer to 74 more)	⊕⊕○○ **Low** Due to Imprecision[2, 5]	9 (5 to 10)	Probably inferior
● (orange)	Placebo	Reference comparator	No estimable	No estimable	No estimable	Reference comparator	7 (4 to 9)	Reference comparator

NMA-SoF table definitions

* Lines represent direct comparisons

** Estimates are reported as odds ratio. CrI: credible interval. Results are expressed in credible intervals as opposed to the confidence intervals (CI) since a Bayesian analysis has been conducted.

*** Anticipated absolute effect. Anticipated absolute effect compares two risks by calculating the difference between the risks of the intervention group with the risk of the control group.

**** Surface under the cumulative (SUCRA) ranking and credible intervals for efficacy are presented. Rank statistics is defined as the probabilities that a treatment out of *n* treatments in a network meta-analysis is the best, the second, the third and so on until the least effective treatment.

GRADE Working Group grades of evidence (or certainty in the evidence)
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanatory Footnotes
[1] Baseline risks (assumed control risk) obtained from the National Cancer Institute pooling project
[2] Very serious imprecision since 95% CrI crosses unity, and with wide credible intervals suggesting high possibility of harm.
[3] Very serious imprecision since RR>1 (suggesting greater likelihood of harm than benefit), and with wide credible intervals).
[4] Very serious imprecision since RR is one (suggesting no evidence of benefit) and wide credible intervals suggesting high possibility of harm.
[5] Conceptually, there was no significant intransitivity, with comparable distribution of plausible effect modifiers across trials of different chemopreventive agents.

NMA-SoF table example 2

Estimates of effects, credible intervals, and certainty of the evidence for chemoprevention of colorectal cancer in individuals with previous colorectal neoplasia
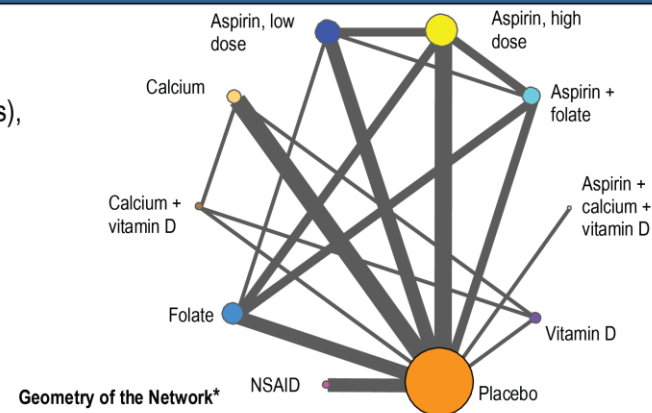
Bayesian NMA-SoF table

HARMS							

Patient or population: Individuals with previous colorectal neoplasia

Interventions: Low and high dose aspirin, nonaspirin non-steroidal anti-inflammatory drugs (NSAIDs), calcium, vitamin D, folic acid

Comparator (reference): Placebo

Outcome: Serious adverse events; range of follow up between three to five years

Setting: Outpatient

Geometry of the Network*

Total studies: 21 RCT Total Participants: 14135	Relative effect** (95% CrI)	Anticipated absolute effect*** (95% CrI)			Certainty of evidence	Ranking**** (95% CrI)	Interpretation of Findings
		Without intervention	*With* intervention	Difference			
● Aspirin + calcium + vitamin D (1 RCT; 714 participants)	0.90 (0.54 to1.51) Network estimate	187 per 1000[1]	89 per 1000	15 more per 1000 (71 more to 77 fewer)	⊕⊕◯◯ Low Due to Imprecision[2, 3]	4 (2 to 7)	Probably inferior
● Calcium + vitamin D (1 RCT; 1125 participants)	1.11 (0.76 to 1.70) Network estimate	187 per 1000[1]	203 per 1000	16 more per 1000 (38 fewer to 94 more)	⊕⊕◯◯ Low Due to Imprecision[2, 3]	2 (1 to 7)	Probably inferior
● Aspirin + folate (3 RCT; 1017 participants)	1.21 (0.83 to 1.77) Network estimate	187 per 1000[1]	218 per 1000	31 more per 1000 (27 fewer to 102 more)	⊕⊕◯◯ Low Due to Imprecision[2, 3]	10 (6 to 10)	Probably inferior
● Aspirin, high dose (3 RCT; 1507 participants)	1.06 (0.76 to 1.49) Network estimate	187 per 1000[1]	196 per 1000	9 more per 1000 (38 fewer to 68 more)	⊕⊕◯◯ Low Due to Imprecision[2, 3]	6 (1 to 10)	Probably inferior

NMA-SoF table example 2

	Intervention	Odds Ratio	Assumed risk	Corresponding risk	Anticipated absolute effect	Certainty	SUCRA rank	
●	Aspirin, low dose (2 RCT; 794 participants)	**0.78** (0.43 to 1.38) Network estimate	187 per 1000[1]	152 per 1000	35 fewer per 1000 (54 more to 97 fewer)	⊕⊕◯◯ **Low** Due to Imprecision[2,3]	8 (3 to 10)	Probably inferior
●	Nonaspirin NSAIDs (3 RCT; 3964 participants)	**1.23** (0.95 to 1.64) Network estimate	187 per 1000[1]	221 per 1000	34 more per 1000 (8 fewer to 87 more)	⊕⊕◯◯ **Low** Due to Imprecision[2,3]	2 (1 to 9)	Probably inferior
●	Vitamin D (1 RCT; 835 participants)	**1.10** (0.74 to 1.70) Network estimate	187 per 1000[1]	212 per 1000	25 more per 1000 (20 fewer to 78 more)	⊕⊕◯◯ **Low** Due to Imprecision[2,3]	5 (2 to 10)	Probably inferior
●	Calcium (4 RCT; 2669 participants)	**1.38** (1.07 to 1.89) Network estimate	187 per 1000[1]	238 per 1000	51 more per 1000 (22 more to 82 more)	⊕⊕⊕⊕ **High**[3]	8 (3 to 10)	Probably superior
●	Folate (3 RCT; 1511 participants)	**0.85** (0.59 to 1.22) Network estimate	187 per 1000[1]	165 per 1000	22 fewer per 1000 (21 more to 59 fewer)	⊕⊕◯◯ **Low** Due to Imprecision[2,3]	6 (2 to 10)	Probably inferior
●	Placebo	Reference comparator	No estimable	No estimable	No estimable	Reference comparator	3 (1 to 10)	Reference comparator

NMA-SoF table definitions

* Lines represent direct comparisons

** Estimates are reported as odds ratio. CrI: credible interval. Results are expressed in credible intervals as opposed to the confidence intervals (CI) since a Bayesian analysis has been conducted.

*** Anticipated absolute effect. Anticipated absolute effect compares two risks by calculating the difference between the risks of the intervention group with the risk of the control group.

**** Surface under the cumulative (SUCRA) ranking and credible intervals for harms are presented. Rank statistics is defined as the probabilities that a treatment out of *n* treatments in a network meta-analysis is the best, the second, the third and so on until the least effective treatment.

GRADE Working Group grades of evidence (or certainty in the evidence)
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect
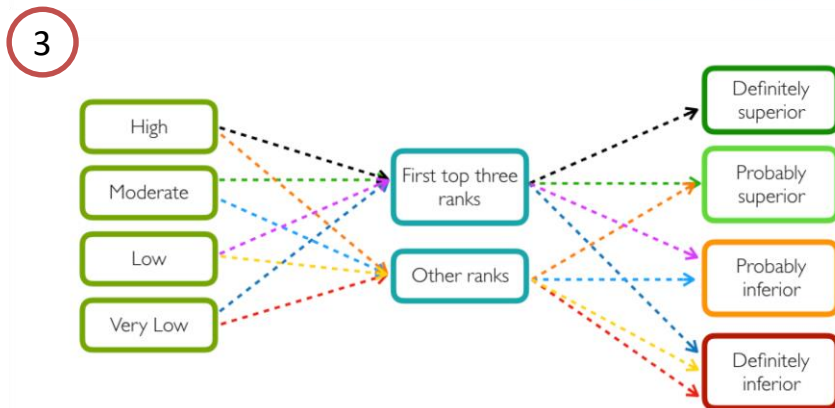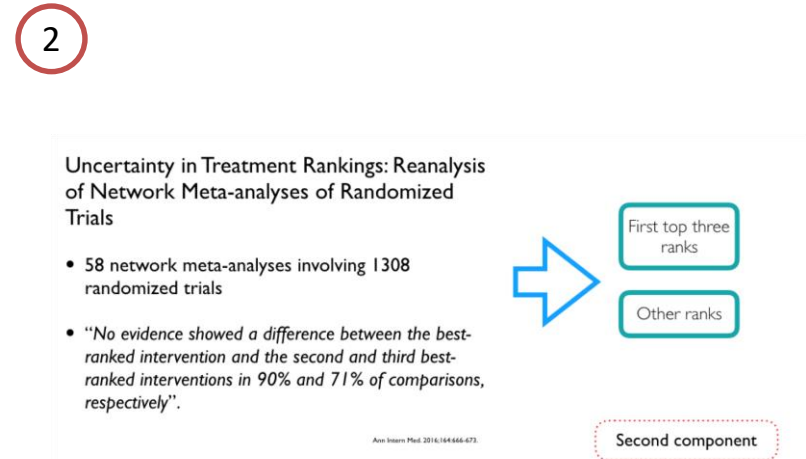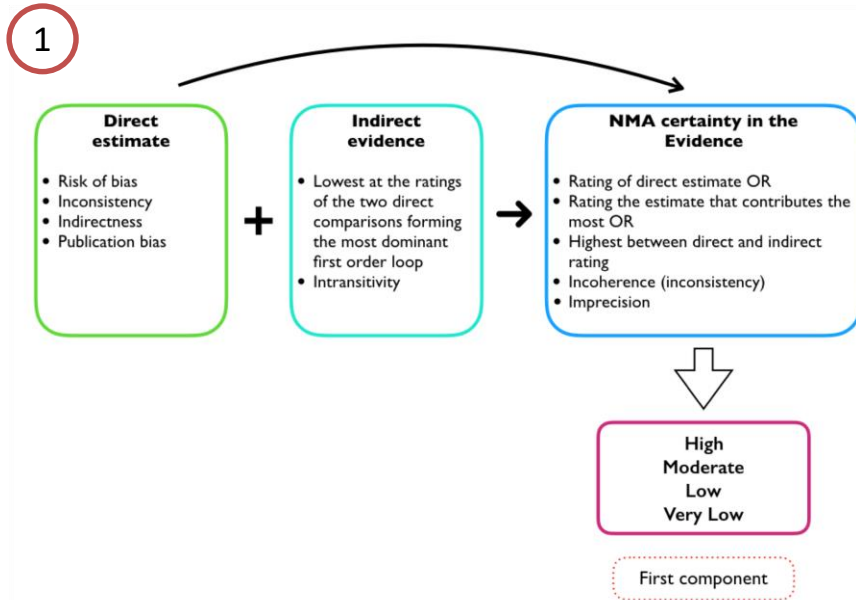
Explanatory Footnotes
[1] Based on assumed control risk of 18.7% (corresponding to pooled 18.7% risk of SAEs in placebo-treated patients of included trials)
[2] Very serious imprecision since 95% CrI crosses unity, and with wide credible intervals suggesting uncertainty in the estimate.
[3] Conceptually, there was no significant intransitivity, with comparable distribution of plausible effect modifiers across trials of different chemopreventive agents.

Drawing conclusions from NMA

1

2

3

4

Estimates of effects, credible intervals, and certainty of the evidence for chemoprevention of colorectal cancer in individuals with previous colorectal neoplasia
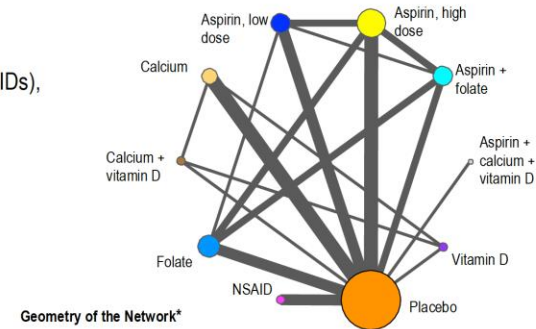
Bayesian NMA SoF table

Patient or population: Individuals with previous colorectal neoplasia

Interventions: Low and high dose aspirin, nonaspirin non-steroidal anti-inflammatory drugs (NSAIDs), calcium, vitamin D, folic acid

Comparator (reference): Placebo

Follow-up: range of follow up between three to five years

Setting: Outpatient

Geometry of the Network*

Prevention of advanced neoplasia

Total studies: 21 RCT Total Participants: 12088	Relative effect** (95% CrI)	Anticipated absolute effect*** (95% CrI)			Certainty of evidence	Ranking**** (95% CrI)	Interpretation of Findings
		Without intervention	*With* intervention	Difference			
Nonaspirin NSAIDs (4 RCT; 3486 participants)	0.37 (0.24 to 0.53) Network estimate	74 per 1000[1]	27 per 1000	47 fewer per 1000 (56 fewer to 35 fewer)	⊕⊕⊕⊕ High [5]	1 (1 to 2)	Definitely superior
Aspirin, low dose (3 RCT; 823 participants)	0.71 (0.41 to 1.23) Network estimate	74 per 1000[1]	53 per 1000	21 fewer per 1000 (44 fewer to 17 more)	⊕⊕○○ Low Due to Imprecision[2,5]	3 (2 to 9)	Probably inferior
Aspirin + calcium + vitamin D (1 RCT; 427 participants)	0.71 (0.18 to 2.49) Network estimate	74 per 1000[1]	53 per 1000	21 fewer per 1000 (61 fewer to 110 more)	⊕⊕○○ Low Due to Imprecision[2,5]	3 (1 to 10)	Probably inferior

Serious adverse events

Total studies: 21 RCT Total Participants: 14135	Relative effect** (95% CrI)	Anticipated absolute effect*** (95% CrI)			Certainty of evidence	Ranking**** (95% CrI)	Interpretation of Findings
		Without intervention	*With* intervention	Difference			
Calcium (4 RCT; 2669 participants)	1.38 (1.07 to 1.89) Network estimate	187 per 1000[1]	238 per 1000	51 more per 1000 (22 more to 82 more)	⊕⊕⊕⊕ High [3]	8 (3 to 10)	Probably superior
Calcium + vitamin D (1 RCT; 1125 participants)	1.11 (0.76 to 1.70) Network estimate	187 per 1000[6]	203 per 1000	16 more per 1000 (38 fewer to 94 more)	⊕⊕○○ Low Due to Imprecision[7,8]	2 (1 to 7)	Probably inferior
Nonaspirin NSAIDs (3 RCT; 3964 participants)	1.23 (0.95 to 1.64) Network estimate	187 per 1000[6]	221 per 1000	34 more per 1000 (8 fewer to 87 more)	⊕⊕○○ Low Due to Imprecision[7,8]	2 (1 to 9)	Probably inferior

Explanatory Footnotes

[1] Baseline risks (assumed control risk) obtained from the National Cancer Institute pooling project

[2] Very serious imprecision since 95% CrI crosses unity, and with wide credible intervals suggesting high possibility of harm.

[3] Very serious imprecision since RR>1 (suggesting greater likelihood of harm than benefit), and with wide credible intervals).

[4] Very serious imprecision since RR is one (suggesting no evidence of benefit) and wide credible intervals suggesting high possibility of harm.

[5] Conceptually, there was no significant intransitivity, with comparable distribution of plausible effect modifiers across trials of different chemopreventive agents.

[6] Based on assumed control risk of 18.7% (corresponding to pooled 18.7% risk of SAEs in placebo-treated patients of included trials)

[7] Very serious imprecision since 95% CrI crosses unity, and with wide credible intervals suggesting uncertainty in the estimate.

[8] Conceptually, there was no significant intransitivity, with comparable distribution of plausible effect modifiers across trials of different chemopreventive agents.

Wrapping up

- Our NMA-SoF table captures the complexity of the information reported in a NMA publication while maximizing simplicity to achieve a user-friendly presentation.

- In a single NMA-SoF table we report relevant information that the literature described as important for NMA findings, including certainty of evidence, and ranking.

- Further experience with *users* may result in modifications to the current table, or the development of alternative formats.

Learning objective

- To gain familiarity in interpreting findings of network meta-analysis (NMA) through NMA 'Summary of findings' (SoF) tables developed based on principles of the GRADE approach to rating certainty of evidence from NMAs

Part 4

QUESTIONS

Acknowledgements

- Holger Schünemann
- Shelly-Anne Li
- Susan M. Jack
- Gordon Guyatt
- Jan L. Brozek
- Joseph Beyene
- M. Hassan Murad

- Bram Rochwerg
- Lawrence Mbuagbaw
- Reem Mustafa
- Nancy Santesso

Acknowledgements

To all of participants of the interviews that with their knowledge and feedback helped us to create a GRADE NMA-SoF table

Contact:

Juan José Yepes-Nuñez

juanjosey@gmail.com and yepesnjj@mcmaster.ca

Holger J Schünemann

holger.schunemann@mcmaster.ca

References

- Puhan MA, Schunemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. Bmj. 2014;349:g5630.

- Brignardello-Petersen R, Bonner A, Alexander PE, Siemieniuk RA, Furukawa TA, Rochwerg B, et al. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. Journal of clinical epidemiology. 2018;93:36-44.

- Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, et al. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. Journal of clinical epidemiology. 2013;66(2):173-83.

- Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2011;14(4):417-28.

- Trinquart L, Attiche N, Bafeta A, Porcher R, Ravaud P. Uncertainty in Treatment Rankings: Reanalysis of Network Meta-analyses of Randomized Trials. Annals of internal medicine. 2016;164(10):666-73.

References

- Mbuagbaw L, Rochwerg B, Jaeschke R, Heels-Andsell D, Alhazzani W, Thabane L, et al. Approaches to interpreting and choosing the best treatments in network meta-analyses. Systematic reviews. 2017;6(1):79.

- Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Annals of internal medicine. 2015;162(11):777-84.

- Foote CJ, Chaudhry H, Bhandari M, Thabane L, Furukawa TA, Petrisor B, et al. Network Meta-analysis: Users' Guide for Surgeons: Part I - Credibility. Clinical orthopaedics and related research. 2015;473(7):2166-71.

- Salanti G. 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. 2012;3(2):80-97.

- Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. Journal of clinical epidemiology. 2011;64(2):163-71.

- Chaudhry H, Foote CJ, Guyatt G, Thabane L, Furukawa TA, Petrisor B, et al. Network Meta-analysis: Users' Guide for Surgeons: Part II - Certainty. Clinical orthopaedics and related research. 2015;473(7):2172-8.