Interpreting results of network meta-analyses: GRADE's minimally contextualized approach

Gordon Guyatt
McMaster University

First, Questions for you

- How familiar are you with GRADE's approach to rating of certainty of evidence in pairwise MA?
- How familiar are you with the basic principles of NMA?
- Have you looked at this paper?

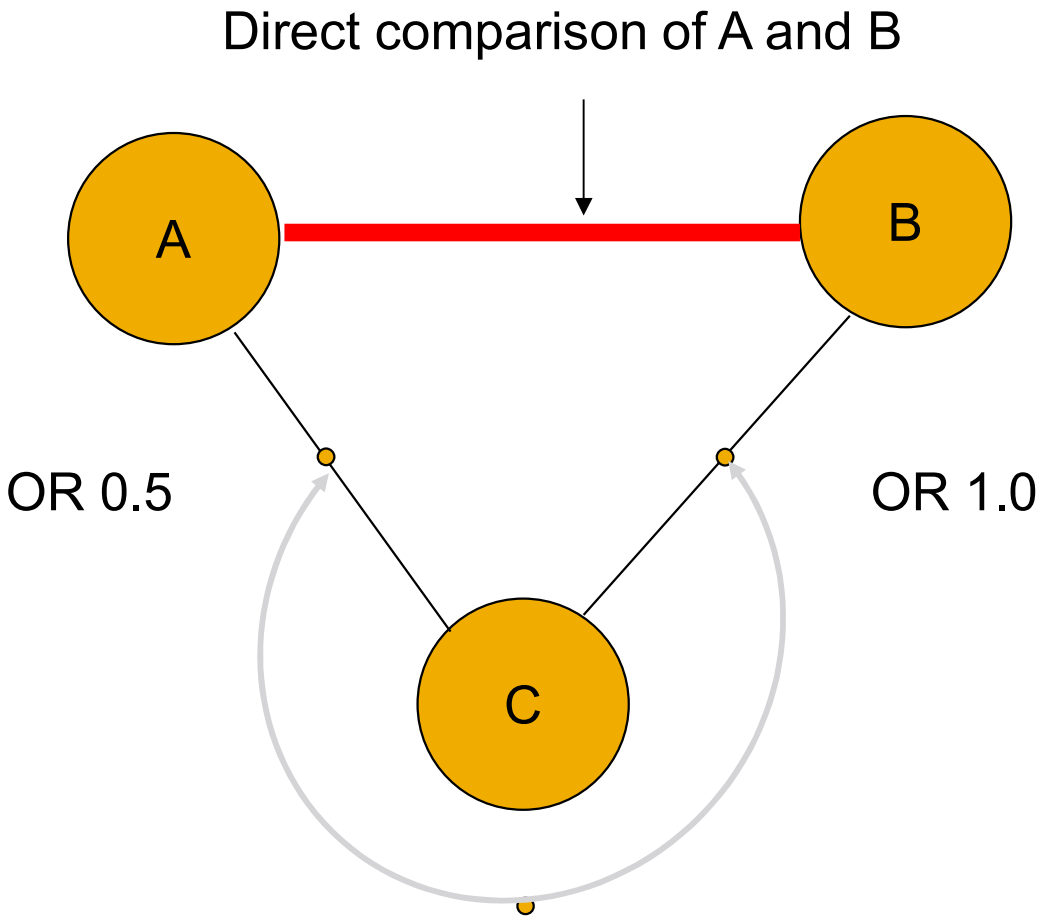
GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework

Romina Brignardello-Petersen,[1] Ivan D Florez,[1,2] Ariel Izcovich,[3] Nancy Santesso,[1] Glen Hazlewood,[4] Waleed Alhazanni,[1] Juan José Yepes-Nuñez,[5] George Tomlinson,[6,7] Holger J Schünemann,[1] Gordon H Guyatt,[1] on behalf of the GRADE working group

Quality/Certainty/Confidence criteria

Study Design	Confidence in estimates	Lower if	Higher if
Randomized trials	High	Risk of bias -1 Serious -2 Very serious Inconsistency -1 Serious -2 Very serious Indirectness -1 Serious -2 Very serious Imprecision -1 Serious -2 Very serious Publication bias -1 Likely -2 Very likely	Large Effect + 1 Large + 1 Very large Dose response +1 Evidence of a gradient All plausible confounding +1 Would reduce a demonstrated effect or +1 would suggest a spurious effect when results show no effect
	Moderate		
Observational studies	Low		
	Very Low		

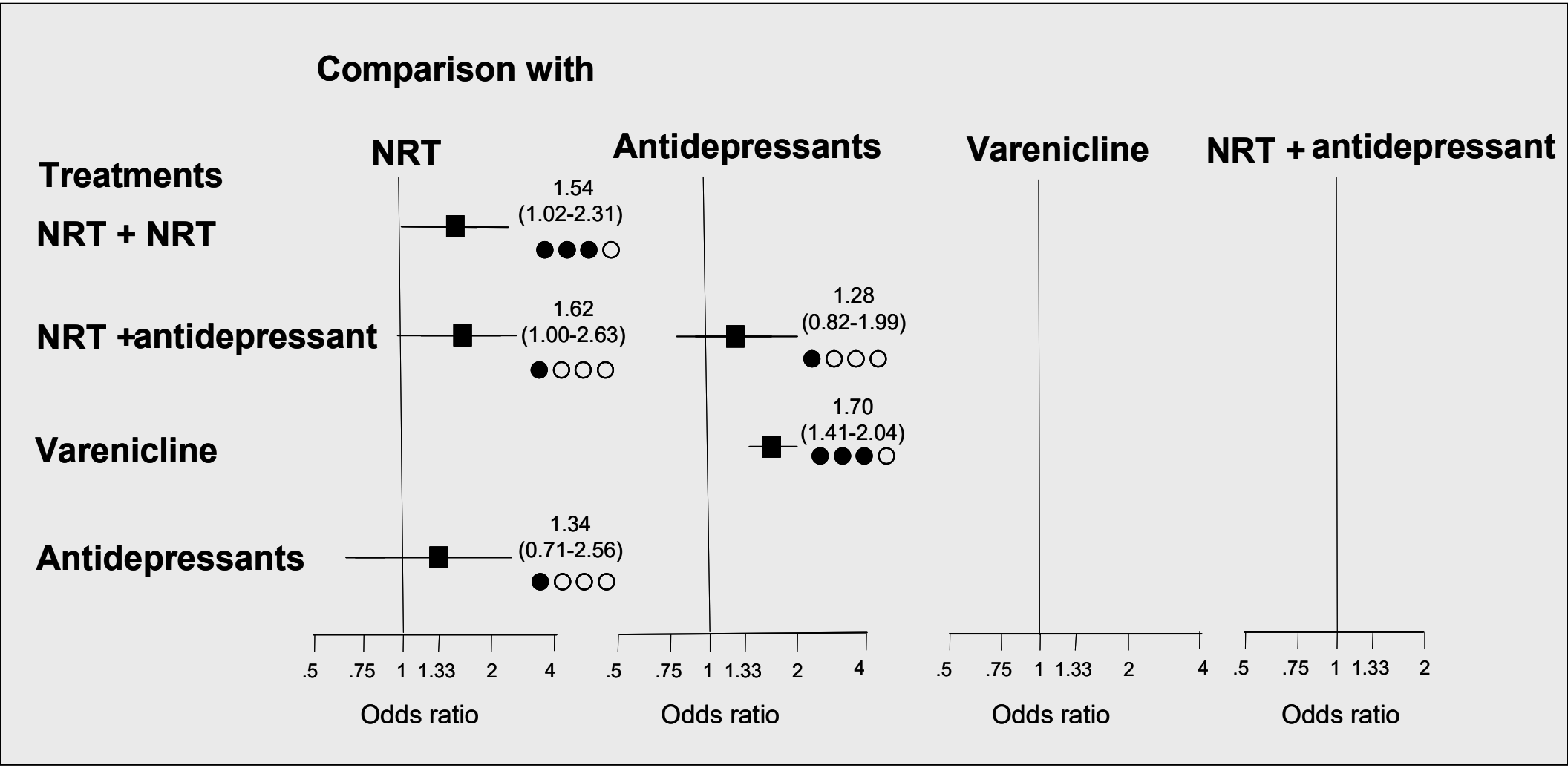
Indirect Comparisons

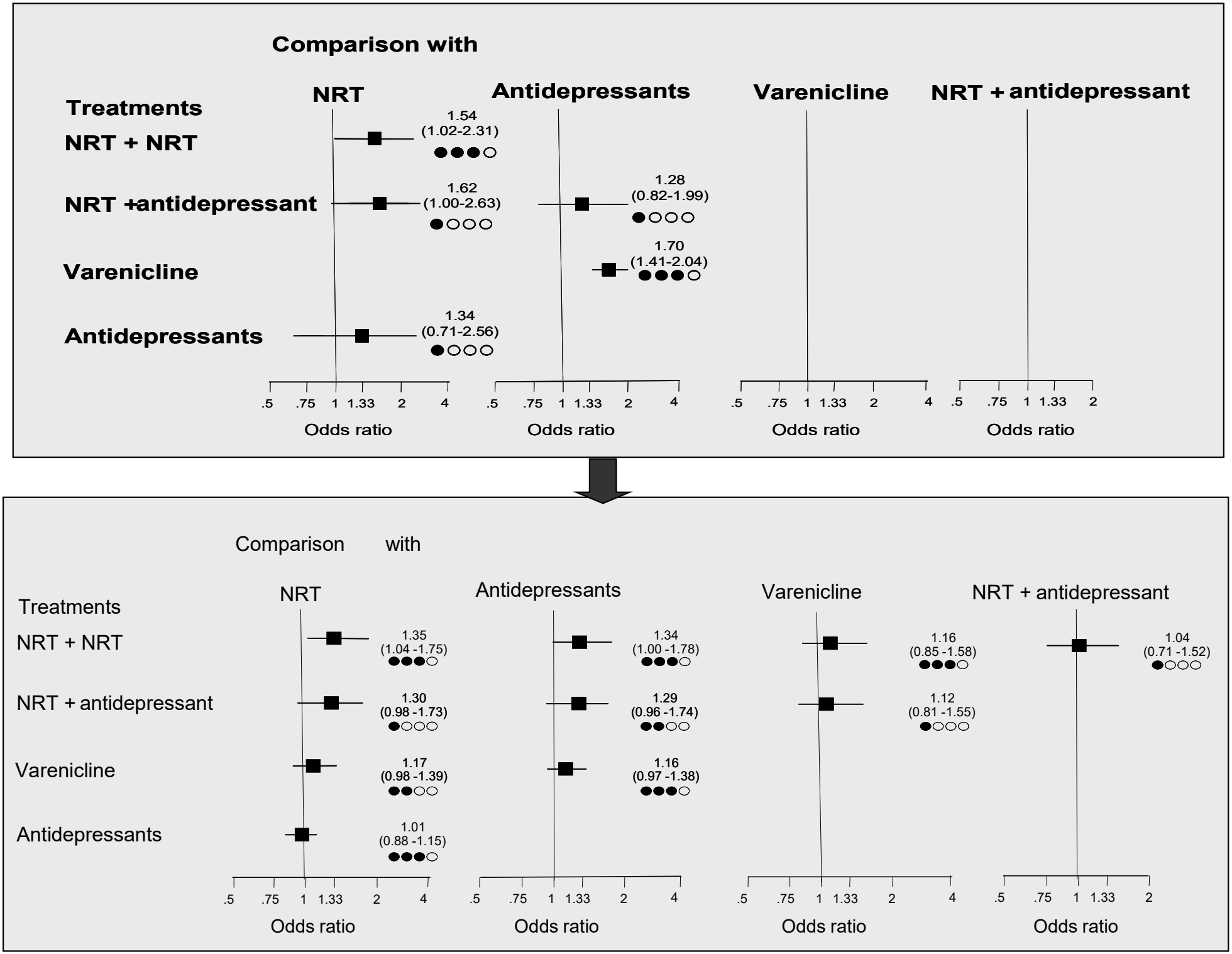


Which is a better treatment A or B?

What is your best guess as to OR for A versus B

Network Meta-analysis Case Study: Which Approach to Nicotine Addiction Works Best




Comparison with
NRT
Antidepressants
Varenicline
NRT + antidepressant
Treatments
NRT + NRT
NRT +antidepressant
Varenicline
Antidepressants
1.54 (1.02-2.31)
1.62 (1.00-2.63)
1.34 (0.71-2.56)
1.28 (0.82-1.99)
1.70 (1.41-2.04)
.5 .75 1 1.33 2 4
Odds ratio
Comparison with
NRT
Antidepressants
Varenicline
NRT + antidepressant
Treatments
NRT + NRT
NRT + antidepressant
Varenicline
Antidepressants
1.35 (1.04 -1.75)
1.30 (0.98 -1.73)
1.17 (0.98 -1.39)
1.01 (0.88 -1.15)
1.34 (1.00 -1.78)
1.29 (0.96 -1.74)
1.16 (0.97 -1.38)
1.16 (0.85 -1.58)
1.12 (0.81 -1.55)
1.04 (0.71 -1.52)
.5 .75 1 1.33 2 4
Odds ratio

Evaluating the Quality of Evidence from a Network Meta-Analysis

Georgia Salanti[1], Cinzia Del Giovane[2], Anna Chaimani[1], Deborah M. Caldwell[3], 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

Abstract

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 network meta-analysis. Judgements about evidence from a network meta-analysis can be different from those made about evidence from pairwise meta-analyses.

PLOS | ONE

Received January 13, 2014; **Accepted** May 18, 2014; **Published** July 3, 2014

GRADE

thebmj

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

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 Puhan[1], Holger J Schünemann[2], Mohammad Hassan Murad[3], Tianjing Li[4], Romina Brignardello-Petersen[5], Jasvinder A Singh[6], Alfons G Kessels[7], Gordon H Guyatt[2], for the GRADE Working Group

	HIGH
	MODERATE
	LOW
	VERY LOW

ELSEVIER

CrossMark

Journal of Clinical Epidemiology 93 (2018) 36–44

Journal of Clinical Epidemiology

Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis

Romina Brignardello-Petersen[a,b], Ashley Bonner[a], Paul E. Alexander[a,c], Reed A. Siemieniuk[a,d], Toshi A. Furukawa[e,f], Bram Rochwerg[a,g], Glen S. Hazlewood[h,i], Waleed Alhazzani[a,g], Reem A. Mustafa[a,j], M. Hassan Murad[k], Milo A. Puhan[l,m], Holger J. Schünemann[a], Gordon H. Guyatt[a,*], For the GRADE Working Group

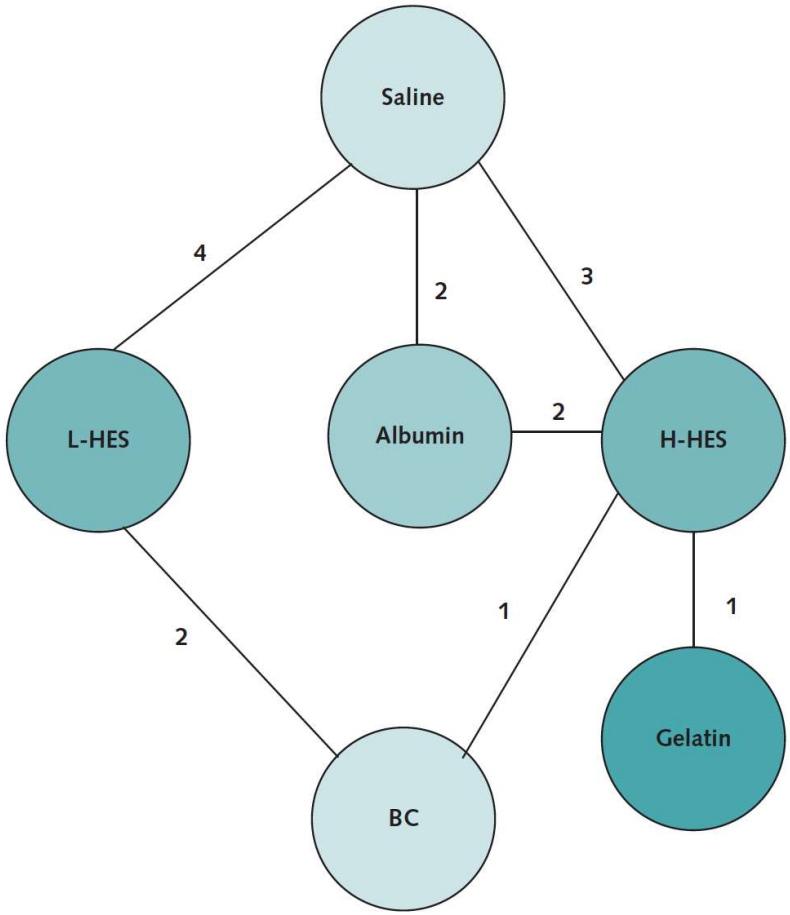
Annals of Internal Medicine

Review

Fluid Resuscitation in Sepsis

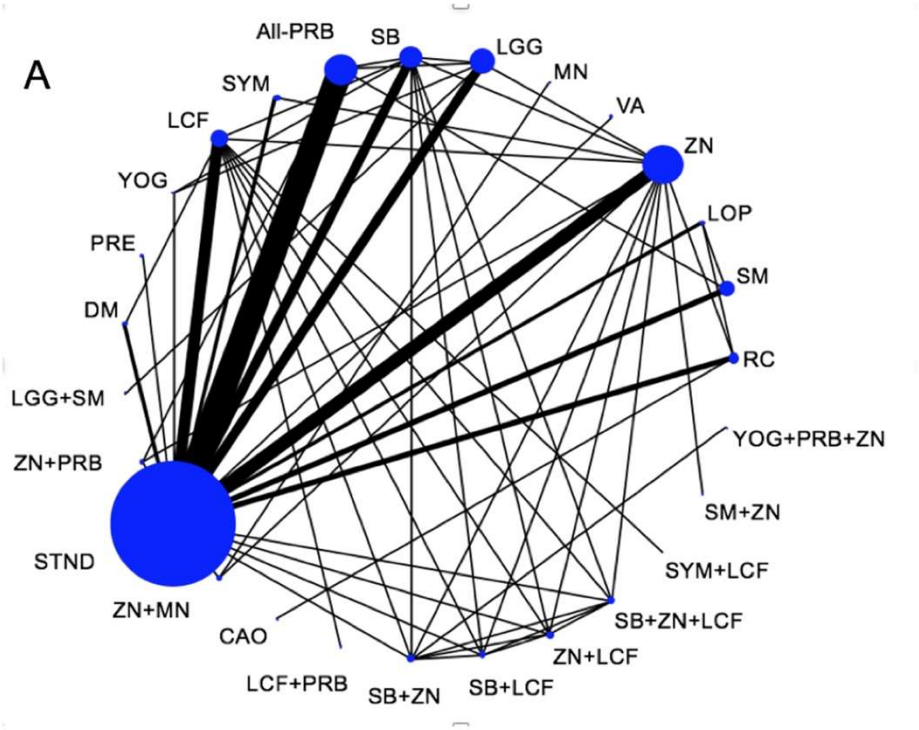
A Systematic Review and Network Meta-analysis

Bram Rochwerg, MD; Waleed Alhazzani, MD; Anees Sindi, MD; Diane Heels-Ansdell, MSc; Lehana Thabane, PhD; Alison Fox-Robichaud, MD; Lawrence Mbuagbaw, MSc; Wojciech Szczeklik, MD; Fayez Alshamsi, MD; Sultan Altayyar, MD; Wang-Chun Ip, MD; Guowei Li, MSc; Michael Wang, MD; Anna Władarczyk, MD; Qi Zhou, PhD; Gordon H. Guyatt, MD; Deborah J. Cook, MD; Roman Jaeschke, MD; and Djillali Annane, MD, PhD, for the Fluids in Sepsis and Septic Shock Group



BC = balanced crystalloid; H-HES = high-molecular-weight hydroxyethyl starch; L-HES = low-molecular-weight hydroxyethyl starch.

Diarrhea duration



- 27 interventions
- 138 studies
- 20,256 participants
- 62 direct comparisons
- 351 pairwise comparisons

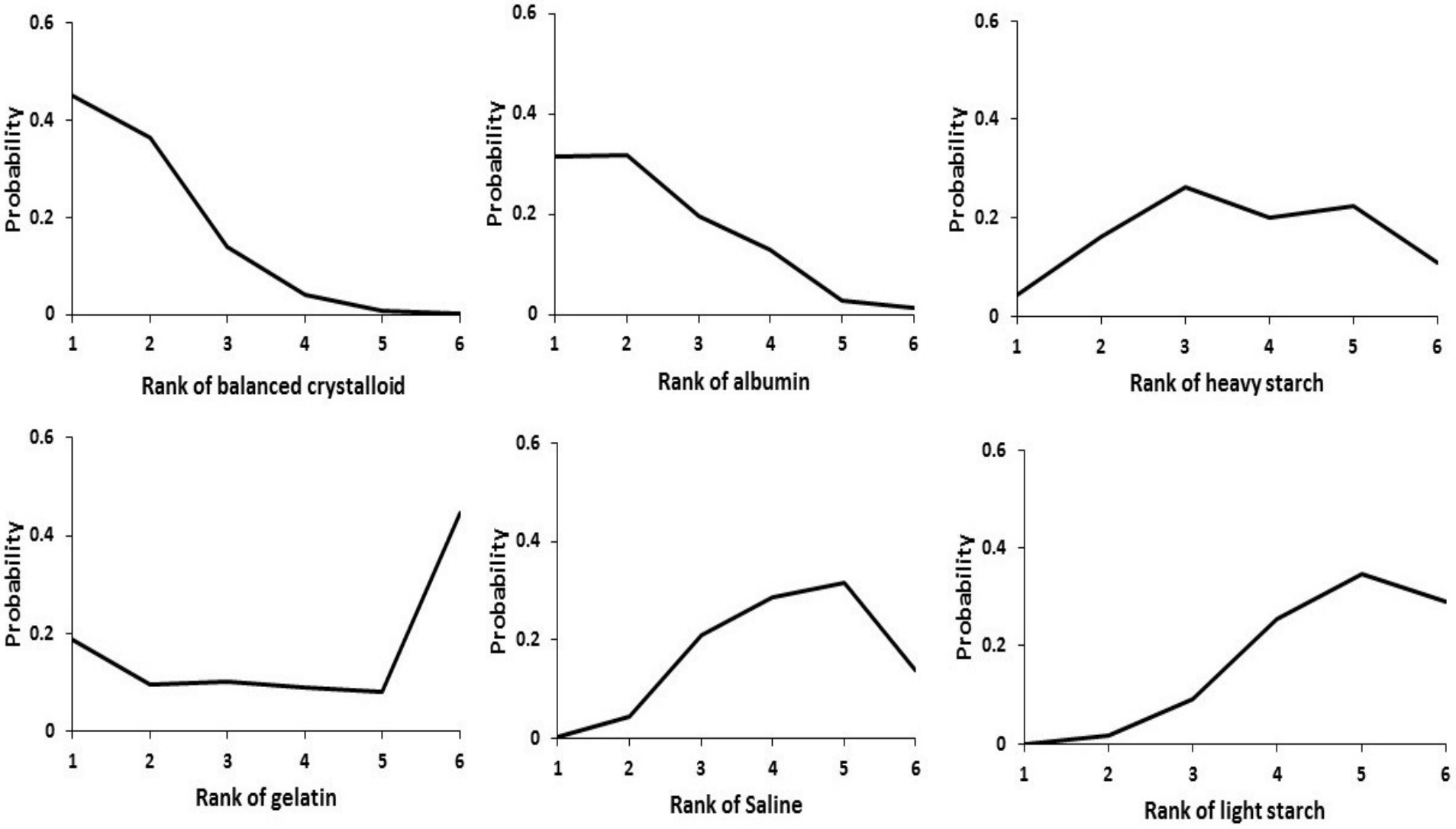
Table 4. NMA Results of 6-Node Analysis, Including Confidence Assessments

Comparison	Trials With Direct Comparisons, *n*	Direct Estimate (95% CI); Quality of Evidence	Indirect Estimate (95% CrI); Quality of Evidence	NMA Estimate (95% CrI)*; Quality of Evidence
L-HES vs. saline	4	1.07 (0.89–1.29); Moderate†	0.59 (0.25–1.35); Very low†‡§	1.04 (0.87–1.25); Moderate
H-HES vs. saline	3	0.64 (0.30–1.37); Moderate†	1.13 (0.71–1.80); Very low†‡	0.95 (0.64–1.41); Moderate
Albumin vs. saline	2	0.81 (0.64–1.03); Moderate†	0.96 (0.14–6.31); Very low‡‖	0.82 (0.65–1.04); Moderate
Balanced crystalloid vs. saline	0	–	0.78 (0.58–1.05); Low†‡	0.78 (0.58–1.05); Low
Gelatin vs. saline	0	–	1.04 (0.46–2.32); Very low†‡	1.04 (0.46–2.32); Very low
H-HES vs. L-HES	0	–	0.91 (0.63–1.33); Low†‡	0.91 (0.63–1.33); Low
Albumin vs. L-HES	0	–	0.79 (0.59–1.06); Low†‡	0.79 (0.59–1.06); Low
Balanced crystalloid vs. L-HES	2	0.80 (0.61–1.04); Moderate§	0.44 (0.19–0.97); Moderate‡	0.75 (0.58–0.97); Moderate
Gelatin vs. L-HES	0	–	1.00 (0.44–2.21); Very low†‡	1.00 (0.44–2.21); Very low
Albumin vs. H-HES	2	1.40 (0.35–5.56); Low‖	0.83 (0.52–1.33); Low†‡	0.87 (0.55–1.36); Low
Balanced crystalloid vs. H-HES	1	0.74 (0.52–1.05); Moderate†	1.35 (0.63–2.92); Very low‡‖	0.82 (0.60–1.13); Moderate
Gelatin vs. H-HES	1	1.09 (0.55–2.19); Low‖	–	1.10 (0.54–2.21); Low
Balanced crystalloid vs. albumin	0	–	0.95 (0.65–1.38); Very low†‡	0.95 (0.65–1.38); Very low
Gelatin vs. albumin	0	–	1.26 (0.55–2.90); Very low‡‖	1.26 (0.55–2.90); Very low
Gelatin vs. balanced crystalloid	0	–	1.34 (0.61–2.89); Very low‡‖	1.34 (0.61–2.89); Very low

How to decide what is best?

Familiarity with sucra?

Sucra good solution?



Surface under the cumulative ranking curve (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%

Reasons to be skeptical

- Chance may explain the differences
  - 0.8 looks a lot bigger than 0.4 but…
- What is the magnitude of difference?
  - 0.8 looks a lot bigger than 0.4 but…
- Quality of evidence may be low for higher Sucra

Table 4. NMA Results of 6-Node Analysis, Including Confidence Assessments

Comparison	Trials With Direct Comparisons, *n*	Direct Estimate (95% CI); Quality of Evidence	Indirect Estimate (95% CrI); Quality of Evidence	NMA Estimate (95% CrI)*; Quality of Evidence
L-HES vs. saline	4	1.07 (0.89–1.29); Moderate†	0.59 (0.25–1.35); Very low†‡§	1.04 (0.87–1.25); Moderate
H-HES vs. saline	3	0.64 (0.30–1.37); Moderate†	1.13 (0.71–1.80); Very low†‡	0.95 (0.64–1.41); Moderate
Albumin vs. saline	2	0.81 (0.64–1.03); Moderate†	0.96 (0.14–6.31); Very low‡‖	0.82 (0.65–1.04); Moderate
Balanced crystalloid vs. saline	0	–	0.78 (0.58–1.05); Low†‡	0.78 (0.58–1.05); Low
Gelatin vs. saline	0	–	1.04 (0.46–2.32); Very low†‡	1.04 (0.46–2.32); Very low
H-HES vs. L-HES	0	–	0.91 (0.63–1.33); Low†‡	0.91 (0.63–1.33); Low
Albumin vs. L-HES	0	–	0.79 (0.59–1.06); Low†‡	0.79 (0.59–1.06); Low
Balanced crystalloid vs. L-HES	2	0.80 (0.61–1.04); Moderate§	0.44 (0.19–0.97); Moderate‡	0.75 (0.58–0.97); Moderate
Gelatin vs. L-HES	0	–	1.00 (0.44–2.21); Very low†‡	1.00 (0.44–2.21); Very low
Albumin vs. H-HES	2	1.40 (0.35–5.56); Low‖	0.83 (0.52–1.33); Low†‡	0.87 (0.55–1.36); Low
Balanced crystalloid vs. H-HES	1	0.74 (0.52–1.05); Moderate†	1.35 (0.63–2.92); Very low‡‖	0.82 (0.60–1.13); Moderate
Gelatin vs. H-HES	1	1.09 (0.55–2.19); Low‖	–	1.10 (0.54–2.21); Low
Balanced crystalloid vs. albumin	0	–	0.95 (0.65–1.38); Very low†‡	0.95 (0.65–1.38); Very low
Gelatin vs. albumin	0	–	1.26 (0.55–2.90); Very low‡‖	1.26 (0.55–2.90); Very low
Gelatin vs. balanced crystalloid	0	–	1.34 (0.61–2.89); Very low‡‖	1.34 (0.61–2.89); Very low

Objective

- To provide guidance on how to draw conclusions regarding which treatments are more likely to be superior or inferior to others considering the estimates of effects, QoE, and rankings

Guiding principles

Based on our experience, very seldom, if ever, will an NMA establish a single treatment clearly superior to all others

1. We should consider **categories of interventions**,

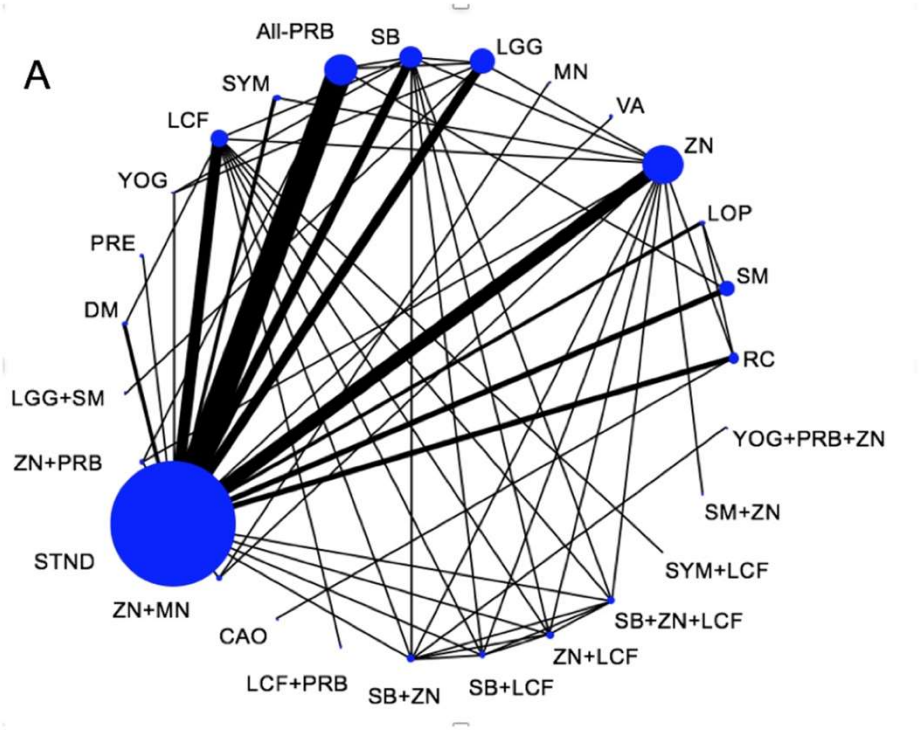
e.g., those **clearly superior**; those with **intermediate effectiveness**; and those that are **inferior**

2. The judgements that place interventions in categories will rely on the **estimates of effect**, and the **certainty of evidence** supporting those estimates, and the rankings

Approach: Example

- **Network Meta-Analysis of the interventions for Acute Diarrhea and Gastroenteritis in Children** (Florez et al. 2018, Submitted for publication)
- **Population**: Children with acute diarrhea and gastroenteritis
- **Interventions/Comparisons:** Pharmacological and nutritional interventions, including Placebo and standard treatment
- **Main Outcome:** Diarrhea Duration in hours (mean difference): Negative value, means a reduction in the duration of the diarrhea in hours; Positive value means an increase in the duration of the diarrhea in hours

Diarrhea duration



- 27 interventions
- 138 studies
- 20,256 participants
- 62 direct comparisons
- 351 pairwise comparisons

II. Approach

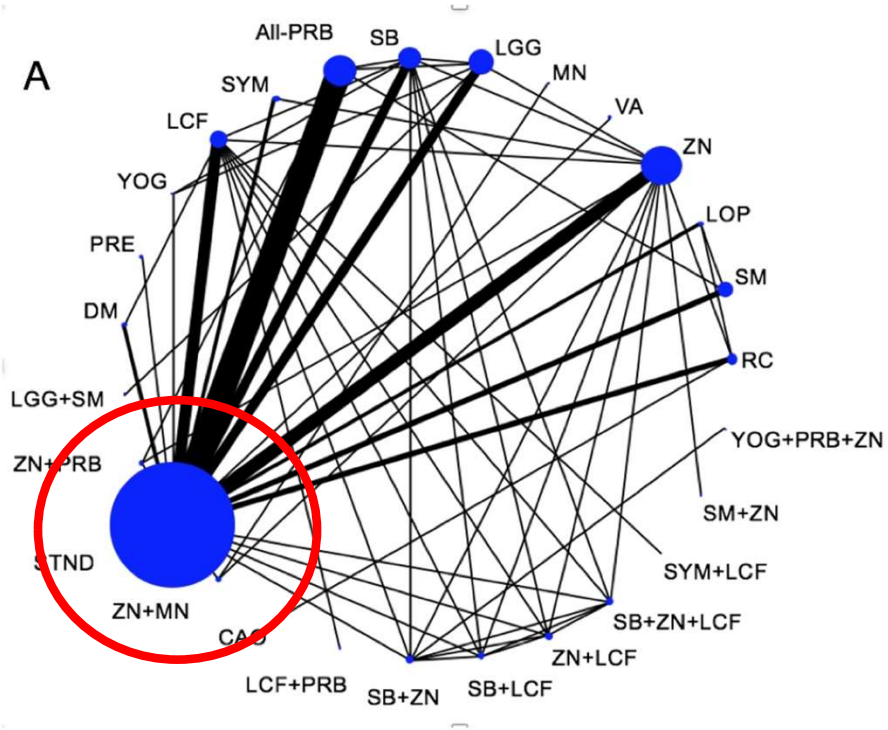
1. Choice of reference treatment and a decision threshold
2. First classification based on comparison with reference
3. Second classification based on pairwise comparisons
4. Separation in 2 main groups according to quality of evidence
5. Checking consistency with pairwise comparisons and rankings

II. Steps

1. **Choice of reference treatment and decision threshold**
2. First classification based on comparison with reference
3. Second classification based on pairwise comparisons
4. Separation in 2 main groups according to quality of evidence
5. Checking consistency with pairwise comparisons and rankings

1. Reference and decision threshold

- Reference should be the treatment most connected to others in the network
  - More direct evidence -> more likely to be higher quality
  - More appropriate to make conclusions based on higher quality
- If more than one treatment highly connected
  - Choose the one for which there is the highest quality when compared to others
- Nice if it's placebo/standard care



1. Reference and decision threshold

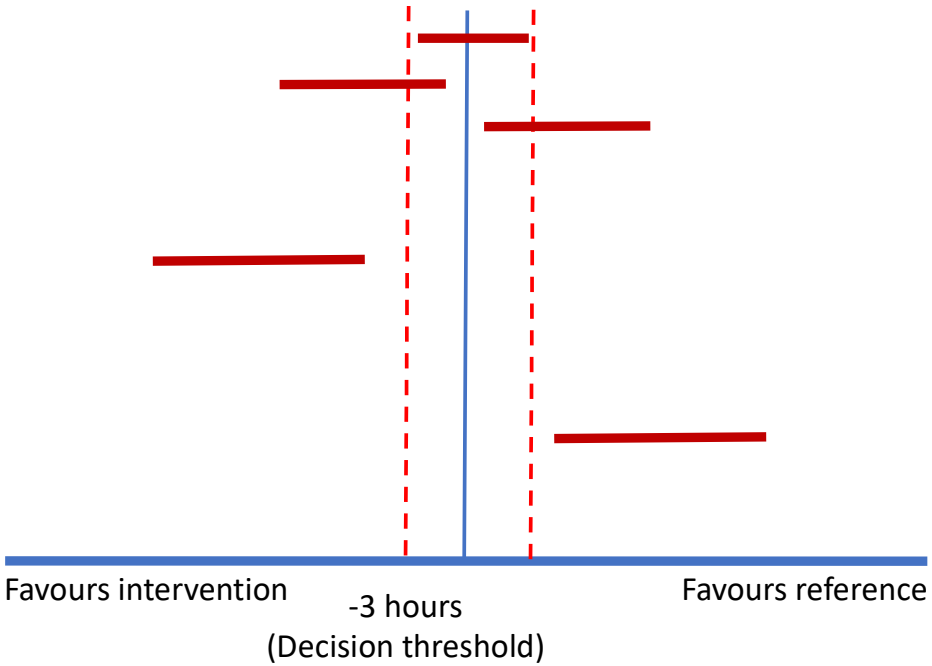
- Decision threshold value that represents reasonablec criterion to claim one treatment is better than another
  - Chance unlikely explanation (Confidence interval excludes null)
  - Difference greater than the MID (CI excludes MID)
    - Should be established considering absolute effects
- Original paper, this presentation, threshold exclude chance
- In BMJ paper, difference 3 hours of diarrhea threshold

II. Steps

1. Choice of reference treatment and decision threshold
2. **First classification based on comparison with reference**
3. Second classification based on pairwise comparisons
4. Separation in 2 main groups according to quality of evidence
5. Checking consistency with pairwise comparisons and rankings

2. First classification based on the comparison with the reference

- Use CI of relative estimate comparing each treatment versus reference



Not convincingly different from reference

Intervention more effective than reference

Intervention less effective than reference

2. First classification based on the comparison with the reference

- Use CI comparing each treatment versus reference
- Classify based on decision threshold
  - Kaolin pectin vs placebo -5.32 (-33.76; 22.83) → Kaolin pectin not different than placebo
  - Symbiotics vs placebo -26.26 (-36.14; -16.22) → Symbiotics better than placebo (category 1)

2. First classification based on the comparison with the reference

Not convincingly different from placebo
Prebiotics
S. boulardii + Zinc + LCF
Yogurt + Probiotics + Zinc
LCF + Probiotics
S. boulardii + LCF
Vitamin A
Kaolin-Pectin
Micronutrients
Standard treatment/placebo
Diluted milk
Yogurt

Category 1: better than placebo
S. boulardii + Zinc
Smectite + Zinc
LGG + Smectite
Zinc + Probiotics
Symbiotics
Zinc + LCF
Zinc (All)
Loperamide
Zinc + Micronutrients
Symbiotics + LCF
Smectite
LGG (All)
All Probiotics
Racecadotril
S. boulardii
LCF

II. Steps

1. Choice of reference treatment and decision threshold
2. First classification based on comparison with reference
3. **Second classification based on pairwise comparisons**
4. Separation in 2 main groups according to quality of evidence
5. Checking consistency with pairwise comparisons and rankings

3. Second classification based on pairwise comparisons

- Aim to differentiate among those different than reference (category 1)
- If one treatment is better than any of the others, it moves up

Comparison	Estimate	Decision
All probiotics vs. racecadotril	-2.18 (-10.66; 6.32)	Stays in the same group
Zinc+ probiotics vs. zinc	-10.96 (-21.65; -0.39)	Zinc + probiotics moves up
Zinc+ micronutrients vs. zinc	0.63 (-13.2: 14.56)	Stays in the same group
S.Boulardii + Zinc vs zinc	-21.55 (-33.66; -9.38)	S.Boulardii + Zinc moves up

Category 1: better than placebo
S. boulardii + Zinc
Smectite + Zinc
LGG + Smectite
Zinc + Probiotics
Symbiotics
Zinc + LCF
Zinc (All)
Loperamide
Zinc + Micronutrients
Symbiotics + LCF
Smectite
LGG (All)
All Probiotics
Racecadotril
S. boulardii
Yogurt
LCF

3. Second classification based on pairwise comparisons

Not apparently different than placebo
Prebiotics
S. boulardii + Zinc + LCF
Yogurt + Probiotics + Zinc
LCF + Probiotics
S. boulardii + LCF
Vitamin A
Kaolin-Pectin
Micronutrients
Standard treatment/placebo
Diluted milk
Yogurt

Category 1
Symbiotics
Zinc + LCF
Zinc (All)
Loperamide
Zinc + Micronutrients
Symbiotics + LCF
Smectite
LGG (All)
All Probiotics
Racecadotril
S. boulardii
LCF

Category 2
S. boulardii + Zinc
Smectite + Zinc
LGG + Smectite
Zinc + Probiotics

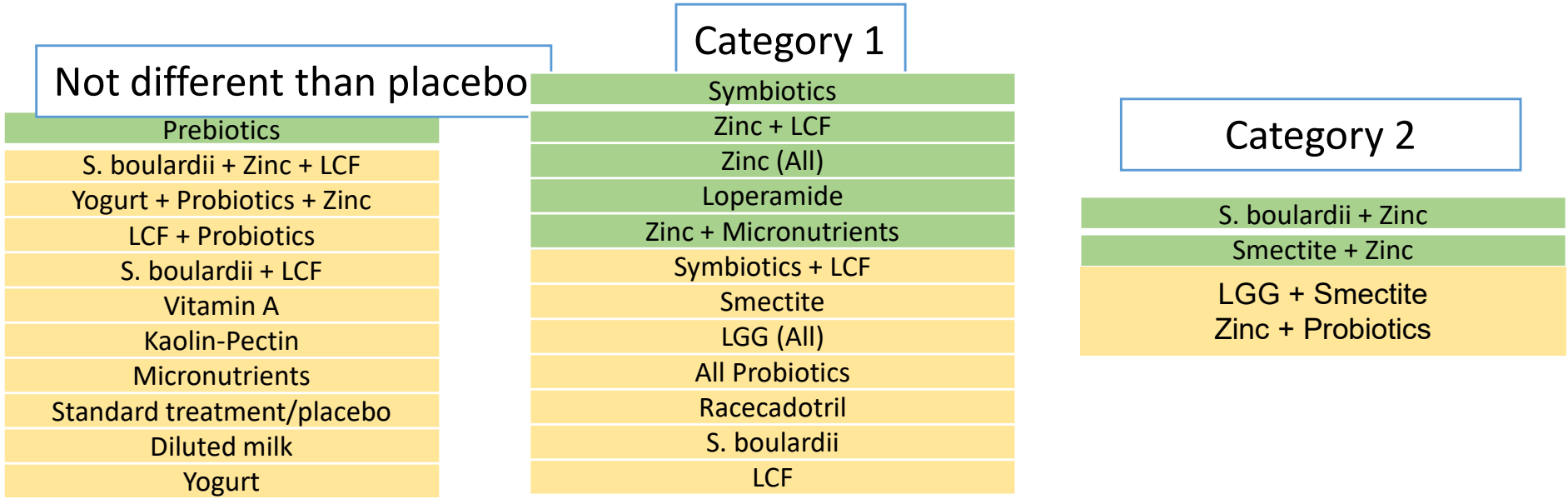
II. Steps

1. Choice of reference treatment and decision threshold
2. First classification based on comparison with reference
3. Second classification based on pairwise comparisons
4. **Separation in 2 main groups according to quality of evidence**
5. Checking consistency with pairwise comparisons and rankings

4. Separation in two groups based on QoE

- QoE when compared to the reference treatment
- High/ moderate vs low/very low

4. Separation in two groups based on QoE



II. Steps

1. Choice of reference treatment and acceptable effect size
2. First classification based on comparison with reference
3. Second classification based on pairwise comparisons
4. Separation in 2 main groups according to quality of evidence
5. **Checking consistency with pairwise comparisons and rankings**

5. Checking consistency with pairwise comparisons and rankings

- Strong evidence contradicting classification would be
  - Relative effect shows that a treatment is better than another classified as worse, with high/moderate quality, AND/OR
  - Ranking shows that a treatment should have been classified in a better category than it is
- Decision based on a case-by-case analysis
- Have not seen a situation like this

Certainty on the evidence	Classification	Intervention	Intervention vs. Standard/placebo MD (95%CrI)	SUCRA
High Certainty (Moderate- to High-quality evidence)	Category 2: Among the most effective	S. boulardii + Zinc (M)	-39.45 (-52.5; -26.7)	0.92
		Smectite + Zinc (M)	-35.63 (-57.6; -13.2)	0.88
	Category 1: Inferior to the most effective / superior to the least effective	Symbiotics (H)	-26.26 (-36.1; -16.2)	0.77
		Zinc + LCF (M)	-21.37 (-36.5; -6.1)	0.61
		Zinc (All) (M)	-18.38 (-23.4; -13.5)	0.50
		Loperamide (M)	-17.79; (-30.4; -5.7)	0.46
		Zinc + Micronutrients (M)	-17.76 (-31.8; -4.1)	0.46
	Category 0: Among the least effective	Prebiotics (M)	-15.32 (-42.8; 12.0)	0.38
Low Certainty (Low- to Very Low-quality evidence)	Category 2: May be among the most effective	LGG + Smectite (VL)	-51.08 (-64.3; -37.9)	1.00
		Zinc + Probiotics (L)	-29.39 (-40.3; -18.6)	0.81
	Category 1: May be inferior to the most effective / superior than the least effective	Symbiotics + LCF (VL)	-32.11 (-53.0; -11.3)	0.85
		Smectite (VL)	-23.90 (-30.8; -17.0)	0.69
		LGG (All) (L)	-22.74 (-28.8; -16.7)	0.65
		All Probiotics (L)	-19.36 (-23.7; -15.1)	0.54
		Racecadotril (L)	-17.19 (-24.7; -9.8)	0.46
		S. boulardii (L)	-16.48 (-23.3; -9.7)	0.42
		LCF (VL)	-12.50 (-19.0; -6.0)	0.31
	Category 0: May be among the least effective	S. boulardii + Zinc + LCF (L)	-16.74 (-36.1; 2.7)	0.42
		Yogurt (VL)	-16.43 (-30.5; -2.1)	0.42
		Yogurt + Probiotics + Zinc (VL)	-15.63 (-56.8; 26.6)	0.38
		LCF + Probiotics (VL)	-13.27 (-36.0; 9.2)	0.31
		S. boulardii + LCF (VL)	-12.32 (-30.0; 6.0)	0.27
		Vitamin A (VL)	-5.95 (-21.4; 9.3)	0.19
		Kaolin-Pectin (VL)	-5.32 (-33.8; 22.8)	0.15
		Micronutrients (L)	-0.68 (-33.3; 32.8)	0.08
		Standard treatment/placebo	--	0.08
		Diluted milk (VL)	3.02 (-14.3; 8.4)	0.04

What about Harms?

- For harms same process, but opposite ratings
- Best would be no worse than reference if placebo/standard care
  - If alternative treatment best could be less harms than reference
- Inferior to best but not worst would be more harms than reference
  - Category 1
- Worst would be more harms than at least 1 category 1

BMJ Open

Development and design validation of a novel network meta-analysis presentation tool for multiple outcomes: a qualitative descriptive study

Mark R Phillips ,[1] Behnam Sadeghirad ,[1,2] Jason W Busse ,[1,2] Romina Brignardello-Petersen,[1] Carlos A Cuello-Garcia,[1] Fernando Kenji Nampo,[3] Yu Jia Guo,[4] Sofia Bzovsky,[5] Raveendhara R Bannuru,[6] Lehana Thabane ,[1,7] Mohit Bhandari,[1,8] Gordon H Guyatt[1]

- Qualitative study develop optimal presentation multiple outcomes
- Seven-member steering committee
  - Oversaw process
- Choice of NMA:
  - Variability certainty/magnitude; benefits/harms; continuous/binary; minimum 5 intervention, five outcomes
- Management of acute musculoskeletal injuries in ER

User testing

- Developed two initial possible formats
- Feedback on initial in two large group sessions
  - Methodologists, graduate students, statisticians, pain researchers
- Modified formats and conducted one-to-one interviews
  - 20 Academic and non clinicians, 3 residents, 3 methodologists
  - Four rounds of interviews, revision after each

Intervention	BENEFIT OUTCOMES					ADVERSE EVENTS		
	Pain ≤ 2 h post-tx	Pain 1 to 7 d post-tx	Physical function	Treatment satisfaction	Symptom relief	GI-related AE's	Neurologic AE's	Dermatologic AE's
	MD (95% CI)	MD (95% CI)	MD (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Topical NSAID	-1.02 (-1.64,-0.39)	-1.08 (-1.40,-0.75)	1.66 (1.16,2.16)	5.20 (2.03,13.33)	6.39 (3.48,11.75)	1.14 (0.65,2.01)	1.18 (0.51,2.74)	0.78 (0.52,1.15)
Oral NSAID	-0.93 (-1.49,-0.37)	-0.99 (-1.46,-0.52)	0.73 (0.17,1.30)	3.24 (0.43,24.70)	3.10 (1.39,6.91)	1.77 (1.33,2.35)	1.02 (0.65,1.59)	1.33 (0.43,4.09)
Acetaminophen	-1.03 (-1.82,-0.24)	-1.07 (-1.89,-0.24)	0.90 (-0.27,2.61)	2.43 (0.18,32.70)	2.73 (0.90,8.27)	0.50 (0.06,4.38)	-	-
Acetaminophen + Diclofenac	-1.11 (-2.00,-0.21)	-1.09 (-2.20,0.01)	-	3.45 (0.18,66.96)	3.72 (1.02,13.52)	-	-	-
Topical NSAID + Menthol Gel	-1.68 (-0.27,-3.09)	-0.89 (-2.33,0.54)	-	-	13.34 (3.30,53.92)	2.35 (0.04,124.85)	1.22 (0.02,69.98)	0.53 (0.05,6.29)
TENS	-1.94 (-2.90,-0.98)	-1.18 (-2.09,-0.28)	0.68 (-0.20,1.57)	-	6.00 (0.78,46.36)	1.25 (0.14,11.01)	1.12 (0.13,9.98)	1.18 (0.13,11.03)
Specific acupressure	-1.59 (-2.52,-0.66)	-2.09 (-3.86,-0.32)	1.51 (0.42,2.61)	0.50 (0.04,6.49)	2.54 (0.52,12.38)	0.80 (0.02,41.67)	0.80 (0.01,42.60)	0.80 (0.01,45.60)
Manipulation	-1.75 (-2.68,-0.81)	0.40 (-1.71,2.51)	0.09 (-1.06,0.87)	-	167.71 (6.67,4217.10)	0.50 (0.01,31.30)	1.41 (0.03,78.76)	-
Acetaminophen + Chlorzoxazone	-	-2.92 (-5.41,-0.43)	-	-	-	0.35 (0.01,10.59)	-	-
Laser therapy	-	-1.04 (-2.28,0.19)	-	-	32.08 (4.93,208.60)	0.49 (0.01,24.85)	0.49 (0.01,25.41)	0.49 (0.01,27.21)
Mobilization	-	3.40 (-0.05,6.85)	0.12 (-0.59,0.83)	2.07 (0.07,58.49)	7.99 (1.29,49.41)	0.93 (0.02,47.12)	0.93 (0.02,48.18)	0.93 (0.02,51.60)
Acetaminophen + Opioid	-0.52 (-1.47,0.43)	-1.71 (-2.97, -0.46)	-	2.50 (0.14,44.86)	1.47 (0.55,3.91)	5.63 (2.84,11.16)	3.53 (1.92,6.49)	-

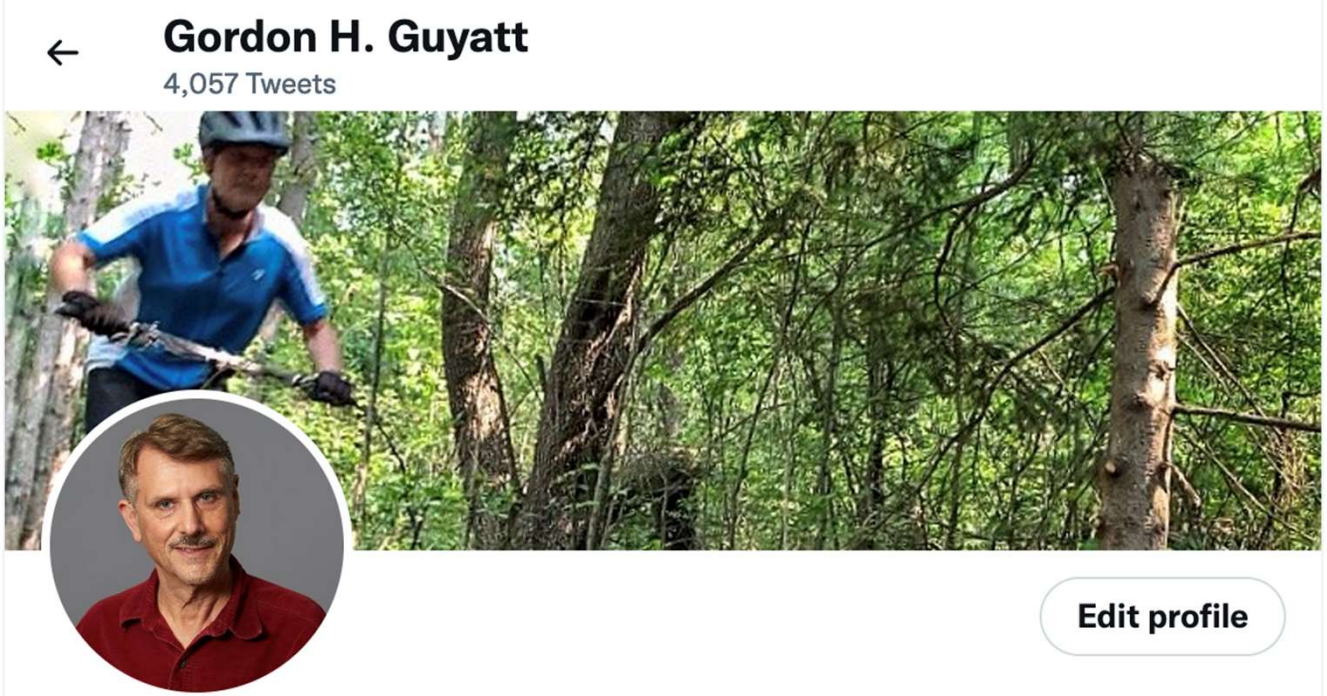
	BENEFIT OUTCOMES		ADVERSE EVENTS	
	High/Moderate Certainty Evidence	Low/Very Low Certainty Evidence	High/Moderate Certainty Evidence	Low/Very Low Certainty Evidence
AMONG THE BEST	Better than placebo and some other interventions	May be better than placebo and some alternatives	No more harmful than placebo	May be no more harmful than placebo
INTERMEDIATE	Better than placebo, but no better than any other interventions	May be better than placebo, but no better than other interventions	More harmful than placebo, but no worse than other interventions	May be more harmful than placebo, but no worse than other interventions
AMONG THE WORST	No better than placebo	May be no better than placebo	More harmful than placebo and some other interventions	May be more harmful than placebo and some alternatives

Intervention	BENEFIT OUTCOMES					ADVERSE EVENTS		
	Pain ≤ 2 h post-tx	Pain 1 to 7 d post-tx	Physical function	Treatment satisfaction	Symptom relief	GI-related AE's	Neurologic AE's	Dermatologic AE's
	MD (95% CI)	MD (95% CI)	MD (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Topical NSAID	**-1.02 (-1.64,-0.39)**	**-1.08 (-1.40,-0.75)**	**1.66 (1.16,2.16)**	**5.20 (2.03,13.33)**	**6.39 (3.48,11.75)**	1.14 (0.65,2.01)	1.18 (0.51,2.74)	0.78 (0.52,1.15)
Oral NSAID	**-0.93 (-1.49,-0.37)**	**-0.99 (-1.46,-0.52)**	**0.73 (0.17,1.30)**	3.24 (0.43,24.70)	**3.10 (1.39,6.91)**	**1.77 (1.33,2.35)**	1.02 (0.65,1.59)	1.33 (0.43,4.09)
Acetaminophen	**-1.03 (-1.82,-0.24)**	**-1.07 (-1.89,-0.24)**	0.90 (-0.27,2.61)	2.43 (0.18,32.70)	2.73 (0.90,8.27)	0.50 (0.06,4.38)	-	-
Acetaminophen + Diclofenac	**-1.11 (-2.00,-0.21)**	-1.09 (-2.20,0.01)	-	3.45 (0.18,66.96)	**3.72 (1.02,13.52)**	-	-	-
Topical NSAID + Menthol Gel	**-1.68 (-0.27,-3.09)**	-0.89 (-2.33,0.54)	-	-	**13.34 (3.30,53.92)**	2.35 (0.04,124.85)	1.22 (0.02,69.98)	0.53 (0.05,6.29)
TENS	**-1.94 (-2.90,-0.98)**	**-1.18 (-2.09,-0.28)**	0.68 (-0.20,1.57)	-	6.00 (0.78,46.36)	1.25 (0.14,11.01)	1.12 (0.13,9.98)	1.18 (0.13,11.03)
Specific acupressure	**-1.59 (-2.52,-0.66)**	**-2.09 (-3.86,-0.32)**	**1.51 (0.42,2.61)**	0.50 (0.04,6.49)	2.54 (0.52,12.38)	0.80 (0.02,41.67)	0.80 (0.01,42.60)	0.80 (0.01,45.60)
Manipulation	**-1.75 (-2.68,-0.81)**	0.40 (-1.71,2.51)	0.09 (-1.06,0.87)	-	**167.71 (6.67,4217.10)**	0.50 (0.01,31.30)	1.41 (0.03,78.76)	-
Acetaminophen + Chlorzoxazone	-	**-2.92 (-5.41,-0.43)**	-	-	-	0.35 (0.01,10.59)	-	-
Laser therapy	-	-1.04 (-2.28,0.19)	-	-	**32.08 (4.93,208.60)**	0.49 (0.01,24.85)	0.49 (0.01,25.41)	0.49 (0.01,27.21)
Mobilization	-	3.40 (-0.05,6.85)	0.12 (-0.59,0.83)	2.07 (0.07,58.49)	**7.99 (1.29,49.41)**	0.93 (0.02,47.12)	0.93 (0.02,48.18)	0.93 (0.02,51.60)
Acetaminophen + Opioid	-0.52 (-1.47,0.43)	**-1.71 (-2.97, -0.46)**	-	2.50 (0.14,44.86)	1.47 (0.55,3.91)	**5.63 (2.84,11.16)**	**3.53 (1.92,6.49)**	-

	BENEFIT OUTCOMES		ADVERSE EVENTS	
	High/Moderate Certainty Evidence	Low/Very Low Certainty Evidence	High/Moderate Certainty Evidence	Low/Very Low Certainty Evidence
AMONG THE BEST	**Better than placebo and some alternatives**	**May be better than placebo and some alternatives**	**No more harmful than placebo**	**May be no more harmful than placebo**
INTERMEDIATE	**Better than placebo, but no better than any alternatives**	**May be better than placebo, but no better than any alternatives**	**More harmful than placebo, but no worse than any alternatives**	**May be more harmful than placebo, but no worse than any alternatives**
AMONG THE WORST	**No better than placebo**	**May be no better than placebo**	**More harmful than placebo and some alternatives**	**May be more harmful than placebo and some alternatives**

Lessons Interpreting NMA Results

- If multiple treatments very hard to decide in league tables
  - Worse if multiple outcomes
- SUCRA bad solution to the problems
  - Precision, magnitude and certainty problems
- GRADE classifies among the best, intermediate, among worst
  - Focus on reference, but also intervention/intervention
  - Threshold for different from reference or one another either null or MID
- User tested presentation formats for multiple outcomes


Gordon H. Guyatt
4,057 Tweets
Edit profile
Gordon H. Guyatt
@GuyattGH
Promoting use of the best evidence and patient values and preferences to inform optimal clinical and health policy decisions
Hamilton, On. Canada
ebcp.mcmaster.ca
Joined June 2014
583 Following
7,194 Followers

@GuyattGH