GRADE approach to rate the certainty of evidence from Network Meta-Analysis and Summary of Findings Tables

Cochrane NMA Learning Live Webinar series

Romina Brignardello-Petersen, Holger Schünemann

February 11, 2020

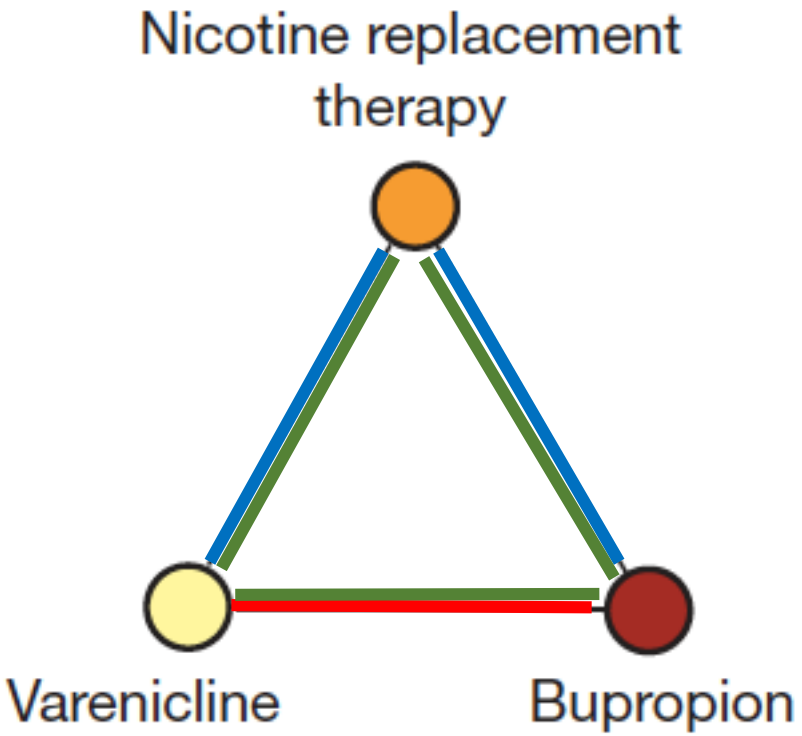
Conflicts of interest

- No financial conflicts of interest
- Members of GRADE working group

Outline

- Where is the guidance available
- General process
- Summary of findings tables

Network meta-analysis



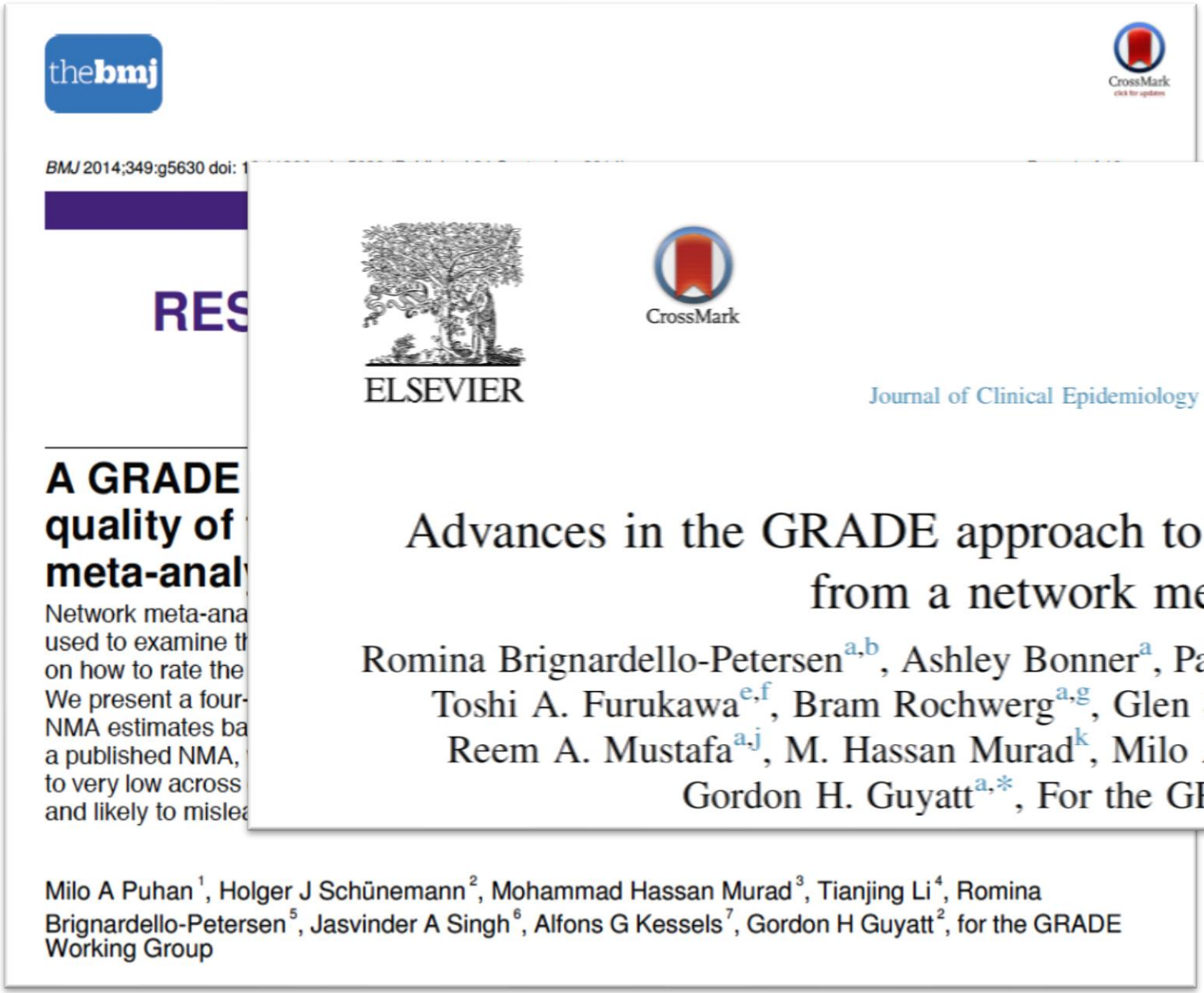
- For the Vareniciline-Bupropion comparison:
  - Direct evidence
  - Indirect evidence (via NRT)
  - Network evidence

GRADE
Network Meta-Analysis

I figure there's a 40% chance of showers and a 10% chance we know what we are talking about.

High
Moderate
Low
Very low

GRADE and NMA


thebmj

BMJ 2014;349:g5630 doi: 1

RES

A GRADE
quality of
meta-anal

Network meta-ana
used to examine t
on how to rate the
We present a four-
NMA estimates ba
a published NMA,
to very low across
and likely to mislea

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

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

ELSEVIER

GRADE appr

Romina Brig
Hassan Murad[c]

[a]Department of Health Rese
[b]Division of Nephrology and H

[d]Division of General Internal M

[e]Int

ELSEVIER

Journal of Clinical Epidemiology 105 (2019) 60–67

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

GRADE approach to rate the certainty from a network meta-analysis: avoiding spurious judgments of imprecision in sparse networks

Romina Brignardello-Petersen[a], M. Hassan Murad[b,*], Stephen D. Walter[a], Shelley McLeod[a,c], Alonso Carrasco-Labra[a,d], Bram Rochwerg[a,e], Holger J. Schünemann[a], George Tomlinson[f,g], Gordon H. Guyatt[a], for the GRADE Working Group

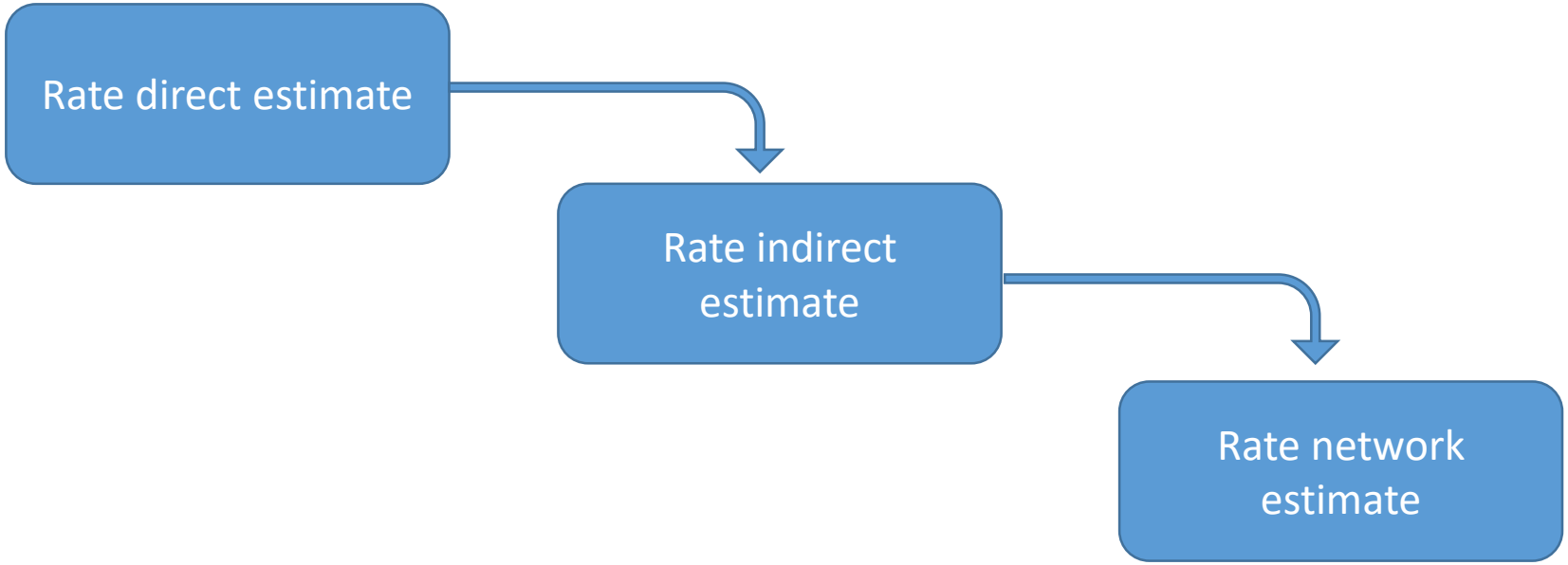
[a]*Department of Health Research Methods, Evidence and Impact, McMaster University, 1280 Main St W, Hamilton, ON L8S 48L, Canada*
[b]*Evidence-Based Practice Center, Mayo Clinic, 200 1st Street SW, Rochester, MN 55905, USA*
[c]*Department of Family and Community Medicine, Schwartz/Reisman Emergency Medicine Institute, University of Toronto, 200 Elizabeth Street, Toronto, ON M5G 2C4, Canada*
[d]*Evidence-Based Dentistry Unit, Faculty of Dentistry, Universidad de Chile, 200 1st Street SW, Rochester, MN 55905, USA*
[e]*Department of Medicine, McMaster University, 1280 Main St W, Hamilton, ON L8S 48L, Canada*
[f]*Department of Medicine, UHN and Mt Sinai Hospital, 200 Elizabeth Street, Toronto, ON M5G 2C4, Canada*
[g]*Institute of Health Policy, Management and Evaluation, University of Toronto, 4th Floor, 155 College St, Toronto, ON M5T 3M6, Canada*

Accepted 17 August 2018; Published online 22 September 2018

GRADE
Network Meta-Analysis

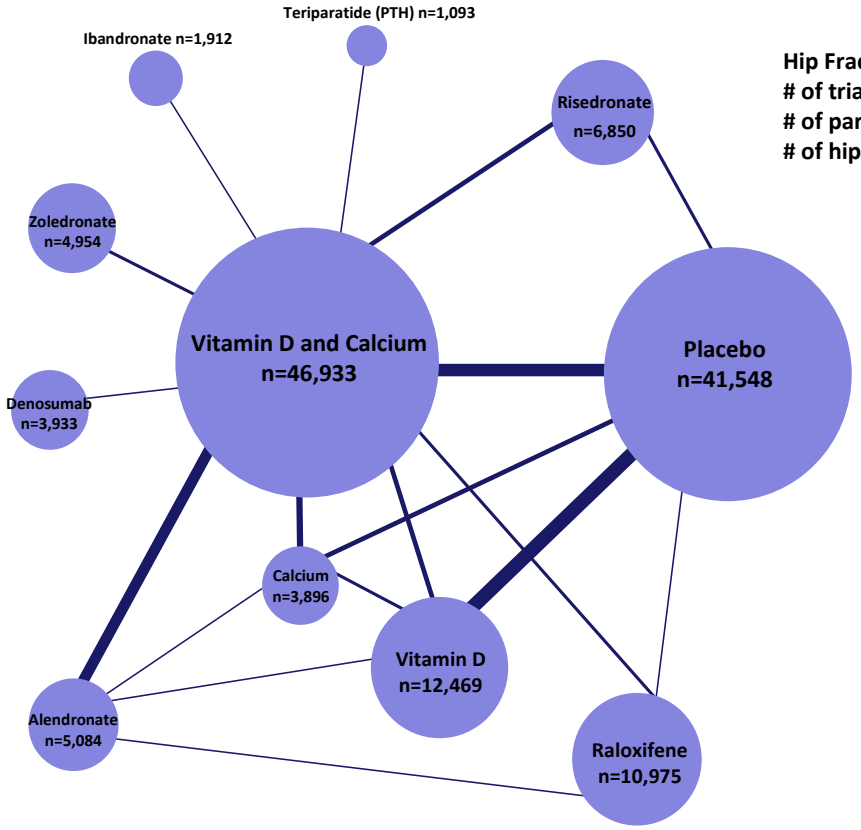
Rating the certainty of estimates from NMA

- Rating informed by the certainty of the pieces of information contributing to the NMA estimate
- Done for each comparison and outcome



GRADE
Network Meta-Analysis

NMA: treatments for preventing hip fractures

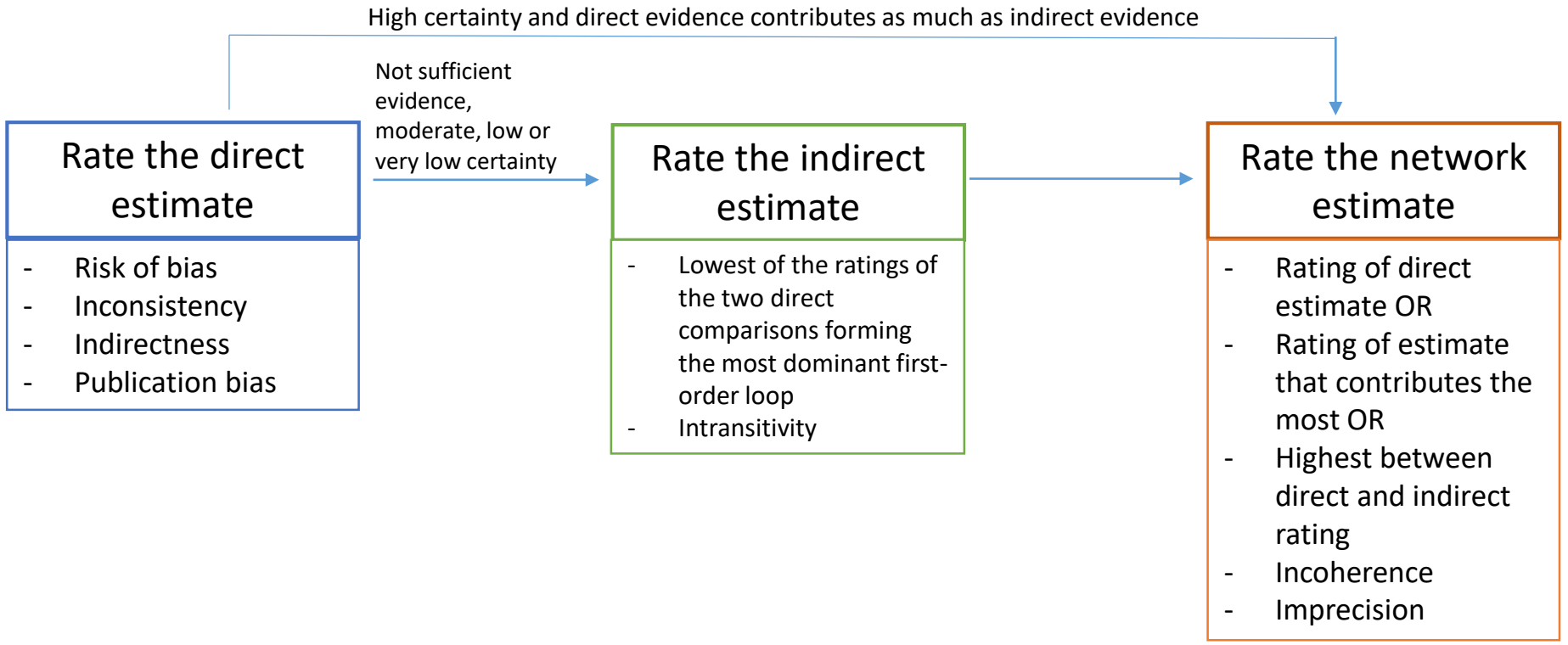


Hip Fractures
# of trials =40
# of participants =139,647
# of hip fracture =2,567

Comparison	Network OR (95% credible interval)	Network confidence in estimates
Teriparatide vs. placebo	0.42 (0.10-1.82)	very low
Denosumab vs. placebo	0.50 (0.27-0.86)	high
Raloxifene vs. placebo	0.87 (0.63-1.22)	moderate
Zoledronate vs. placebo	0.50 (0.34-0.73)	high
Risedronate vs. placebo	0.48 (0.31-0.66)	moderate
Ibandronate vs. placebo	0.49 (0.21-1.20)	very low
Alendronate vs. placebo	0.45 (0.27-0.68)	moderate
Vitamin D vs. placebo	1.13 (0.94-1.34)	low
Vitamin D+Calcium vs. placebo	0.81 (0.68-0.96)	moderate
Calcium vs. placebo	1.14 (0.82-1.59)	moderate
Denosumab vs. Teriparatide	1.17 (0.24-5.54)	low
Raloxifene vs. Teriparatide	2.05 (0.47-9.47)	very low
Zoledronate vs. Teriparatide	1.18 (0.26-5.30)	low
Risedronate vs. Teriparatide	1.12 (0.25-4.98)	very low
Ibandronate vs. Teriparatide	1.11 (0.22-6.42)	very low
Alendronate vs. Teriparatide	1.02 (0.24-4.82)	very low
Vitamin D vs. Teriparatide	2.67 (0.63-11.97)	very low
Vitamin D+Calcium vs. Teriparatide	1.92 (0.45; 8.42)	low
Calcium vs. Teriparatide	2.69 (0.63-12.23)	very low
Raloxifene vs. Denosumab	1.76 (0.95-3.41)	low
Zoledronate vs. Denosumab	1.02 (0.54-1.93)	moderate
Risedronate vs. Denosumab	0.96 (0.50-1.78)	very low
Ibandronate vs. Denosumab	0.98 (0.36-2.79)	low
Alendronate vs. Denosumab	0.90 (0.45-1.78)	low
Vitamin D vs. Denosumab	2.28 (1.28-4.16)	moderate
Vitamin D+Calcium vs. Denosumab	1.64 (0.97-2.87)	high
Calcium vs. Denosumab	2.33 (1.25-4.40)	moderate
Zoledronate vs. Raloxifene	0.57 (0.35-0.93)	moderate
Risedronate vs. Raloxifene	0.55 (0.31-0.84)	low
Ibandronate vs. Raloxifene	0.55 (0.23-1.42)	very low
Alendronate vs. Raloxifene	0.51 (0.29- 0.87)	moderate
Vitamin D vs. Raloxifene	1.30 (0.89-1.86)	low

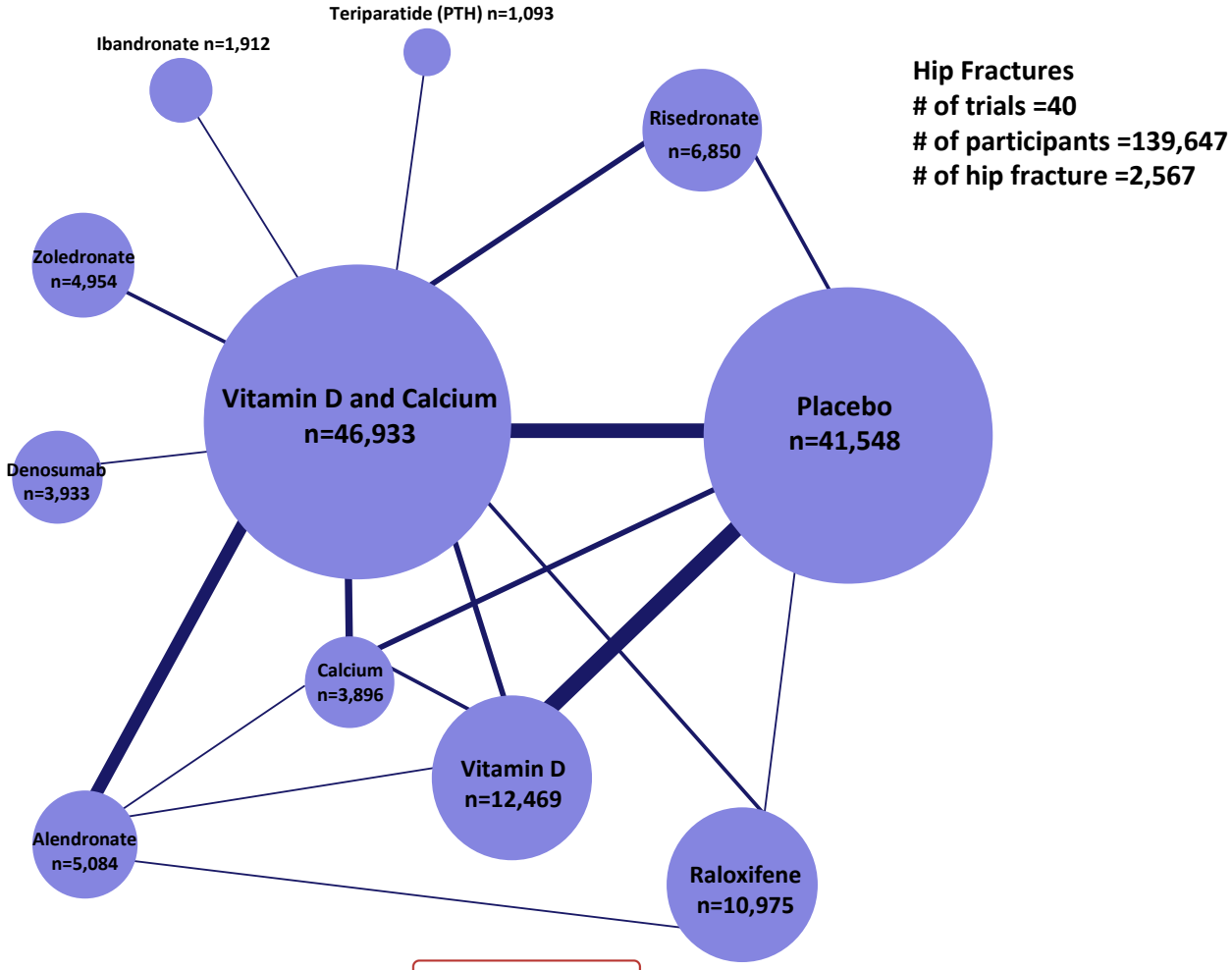
GRADE
Network Meta-Analysis

Rating the certainty of evidence from NMA



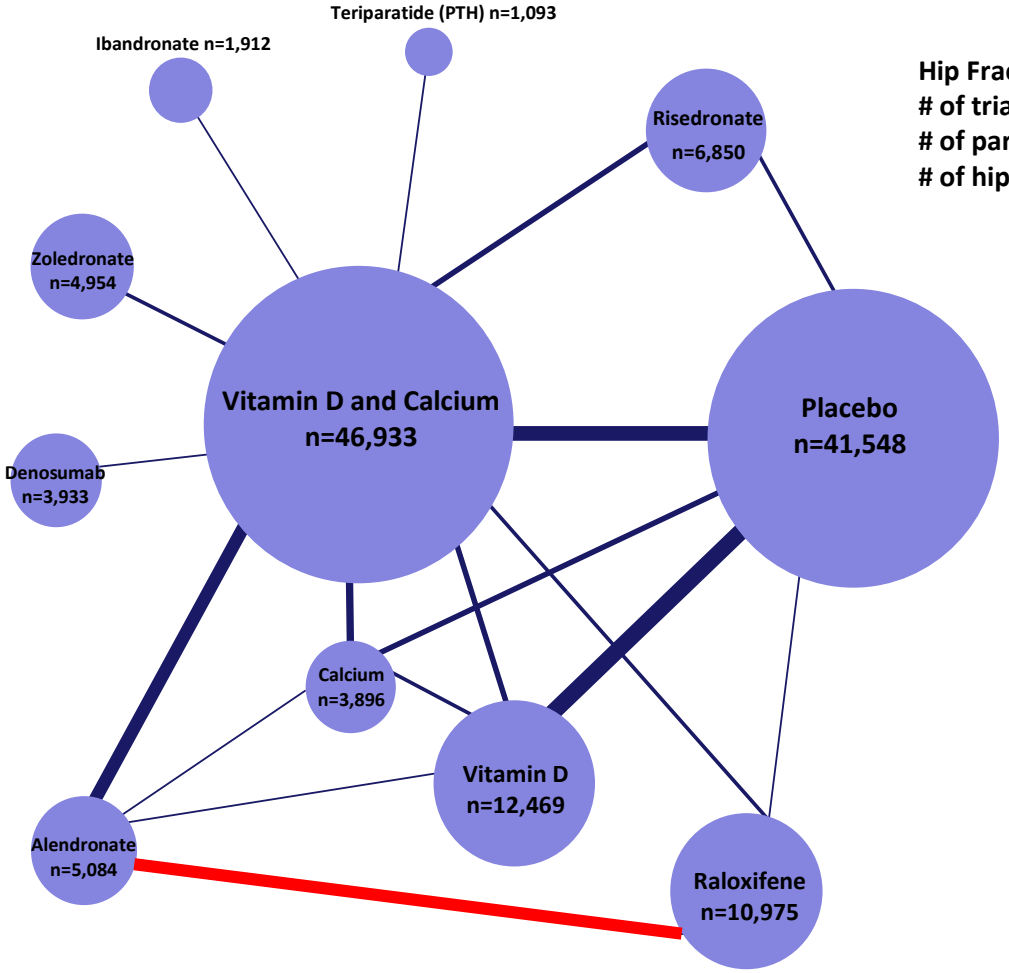
GRADE
Network Meta-Analysis

NMA: treatments for preventing hip fractures



GRADE
Network Meta-Analysis

Example: Alendronate versus Raloxifene



Hip Fractures
# of trials =40
# of participants =139,647
# of hip fracture =2,567

GRADE
Network Meta-Analysis

Alendronate versus Raloxifene: Rating direct estimate

1. Assess risk of bias
2. Assess inconsistency
3. Assess indirectness
4. Assess publication bias

GRADE
Network Meta-Analysis

Alendronate versus Raloxifene: Rating direct estimate

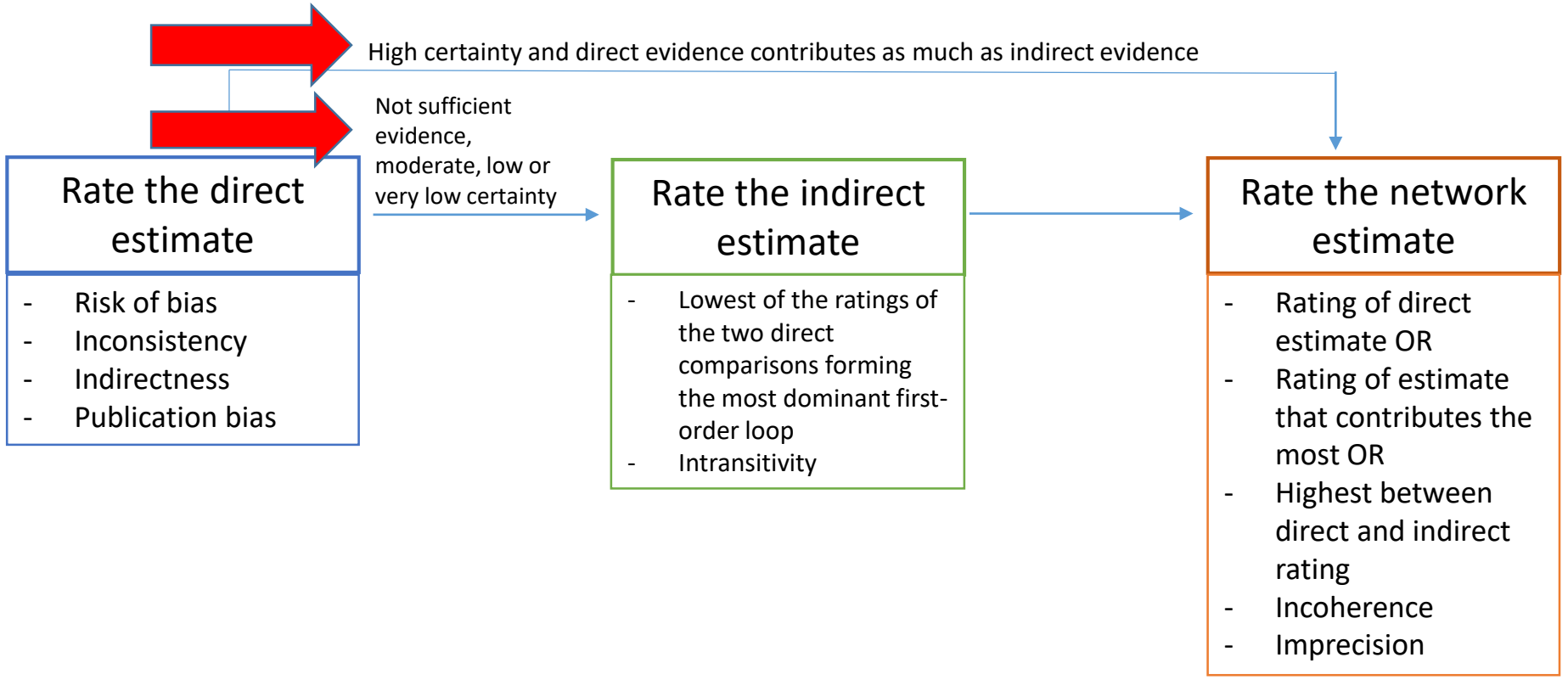
- Estimate: OR 0.49, 95% CI 0.04; 5.45

1. Risk of bias: not serious
2. Inconsistency: not serious (only one study)
3. Indirectness: not serious
4. Publication bias: undetected

Rating: High ⨁⨁⨁⨁

GRADE
Network Meta-Analysis

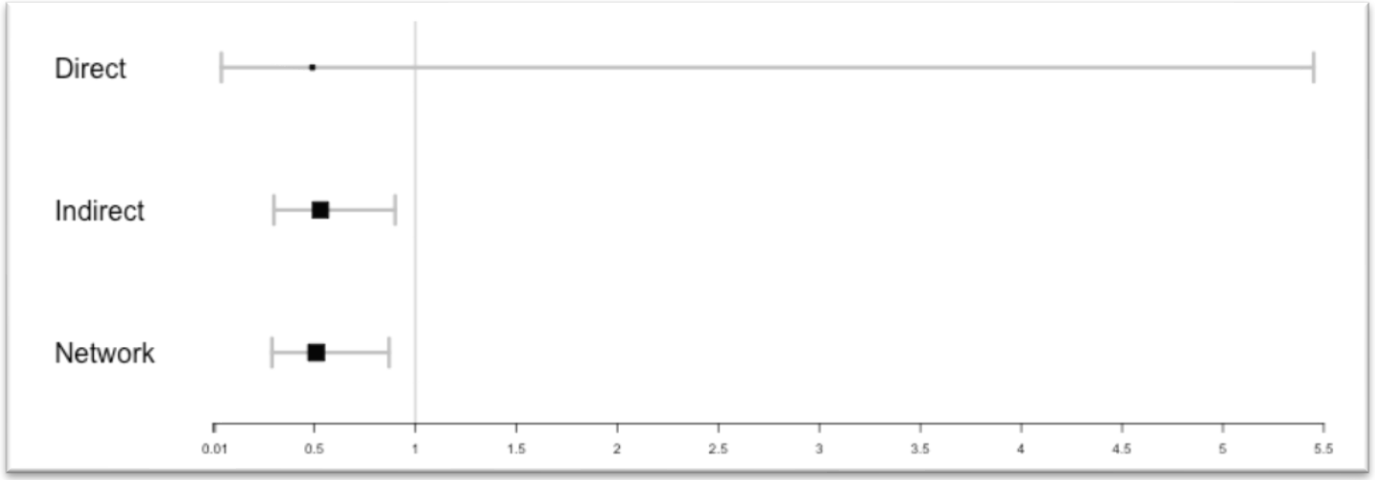
Rating the certainty of evidence from NMA



GRADE
Network Meta-Analysis

Alendronate versus Raloxifene: Direct estimate dominant?

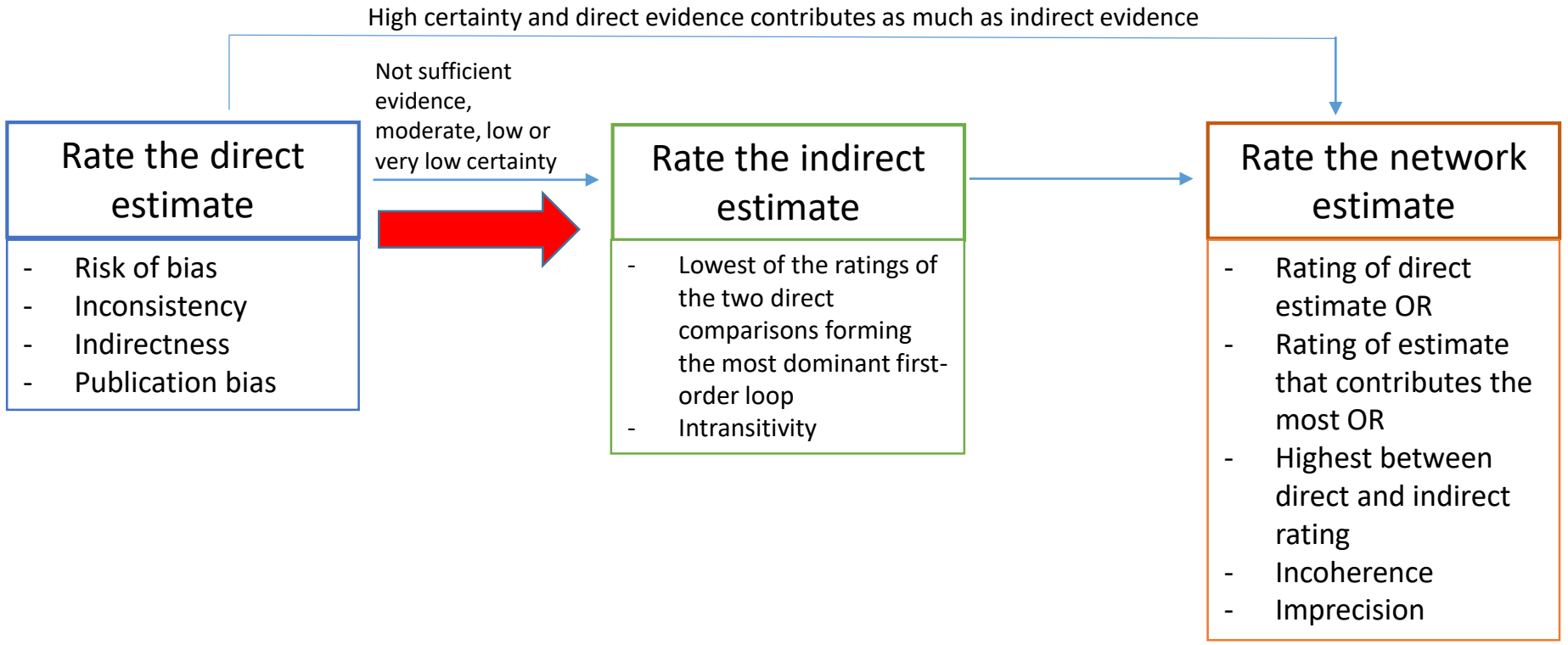
- Does the direct estimate seem to be contributing at least as much as the indirect estimate to the network estimate?
- Indirect estimate obtained using the "node splitting approach"



Indirect estimate is contributing more the network estimate

GRADE
Network Meta-Analysis

Rating the certainty of evidence from NMA

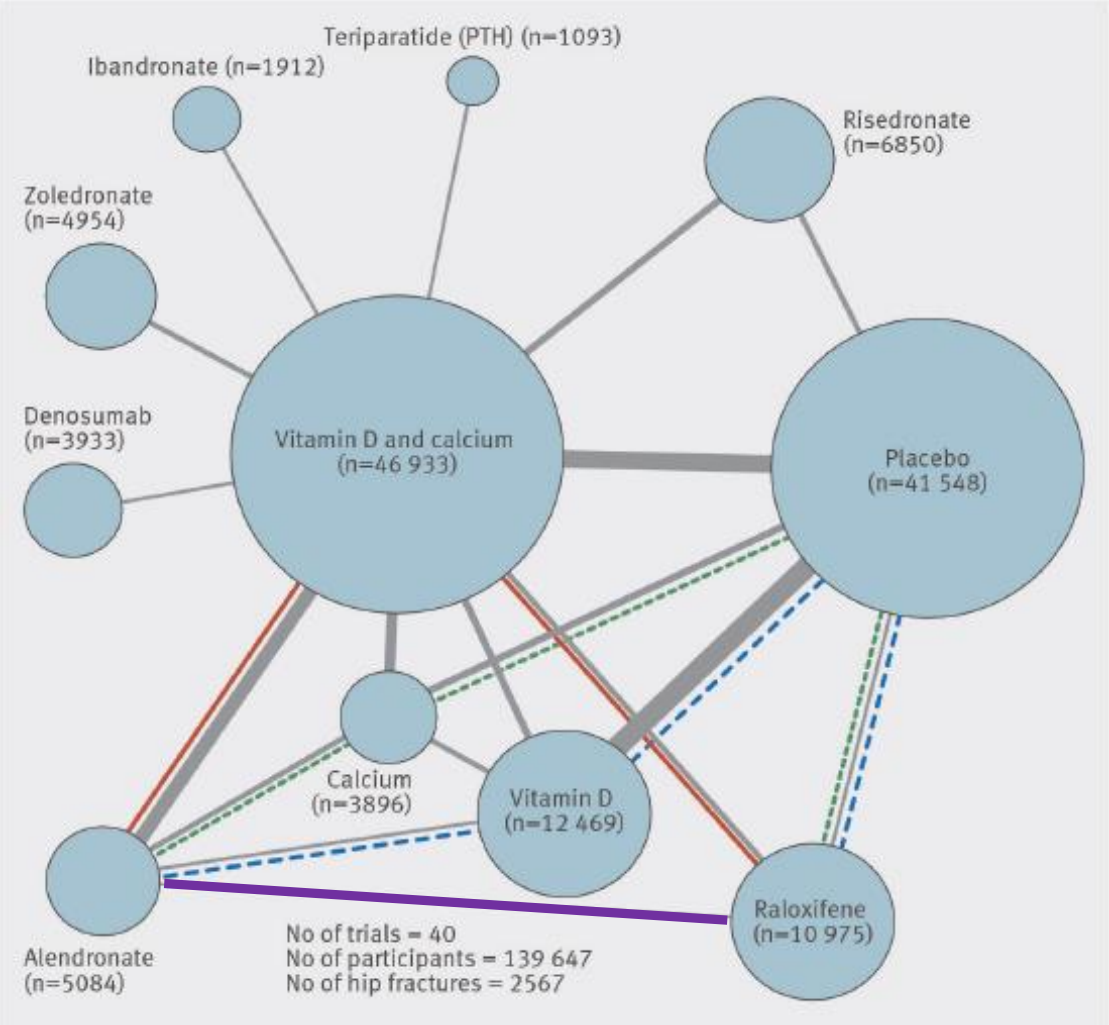


GRADE
Network Meta-Analysis

Alendronate versus Raloxifene: Rating indirect estimate

1. Choose the most dominant first-order loop
2. Look at the rating of each of the direct estimates from that loop
3. Choose the lowest of the two ratings
4. Examine for intransitivity

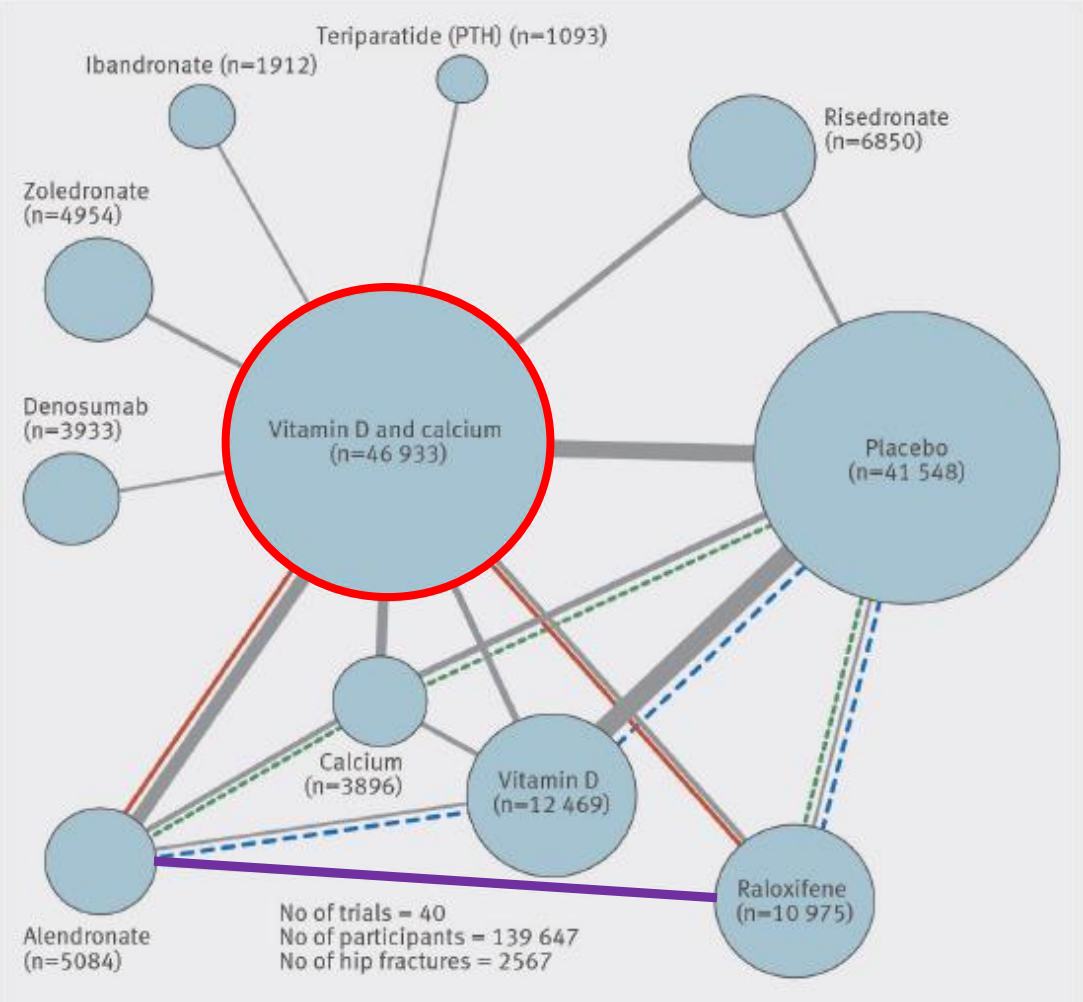
1. Choosing the most dominant first order loop- Loops in NMA



Alendronate versus raloxifene

- First order via vitamin D+ calcium
- Second order via calcium - placebo
- Second order via vitamin D – placebo
- Third order via vitamin D+ calcium – risendronate - placebo

1. Choosing the most dominant first order loop



- In this example there is only one first order loop
- If there is more than one:
  - Larger number of trials and participants

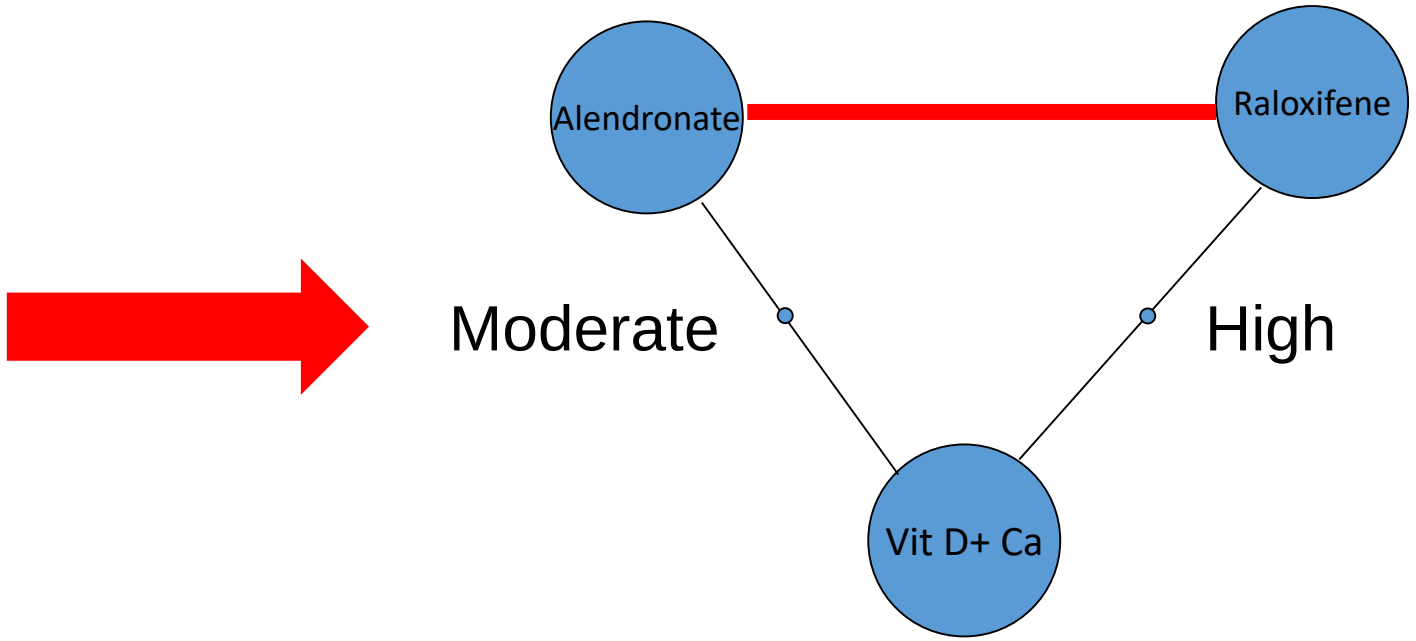
2. Rating of each of the direct estimates

- Vitamin D + calcium versus Alendronate
  - Moderate ⨁⨁⨁◯
  - Due to risk of bias

- Vitamin D + calcium versus Raloxifene
  - High ⨁⨁⨁⨁

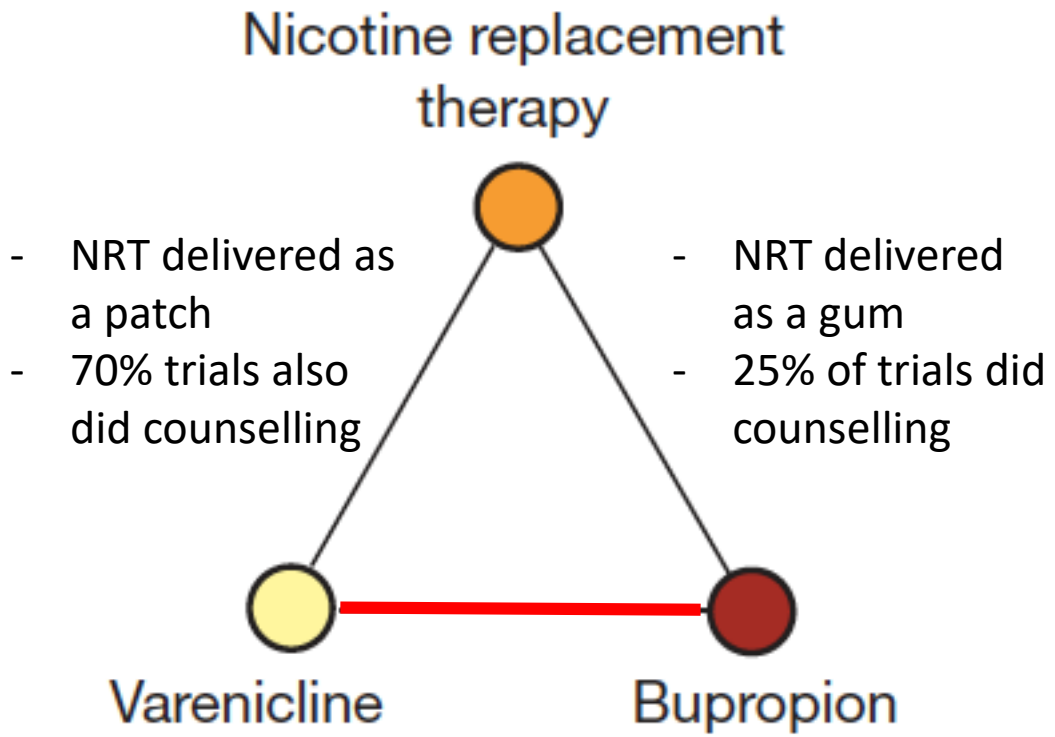
GRADE
Network Meta-Analysis

3. Choose the lowest of the two ratings



4. Examine for intransitivity

- Differences in study characteristics that may modify treatment effects on the direct comparisons that form the basis on an indirect estimate
- Consequence: biased indirect estimate
- It is evaluated conceptually (or it can be improved using a network meta-regression)



GRADE
Network Meta-Analysis

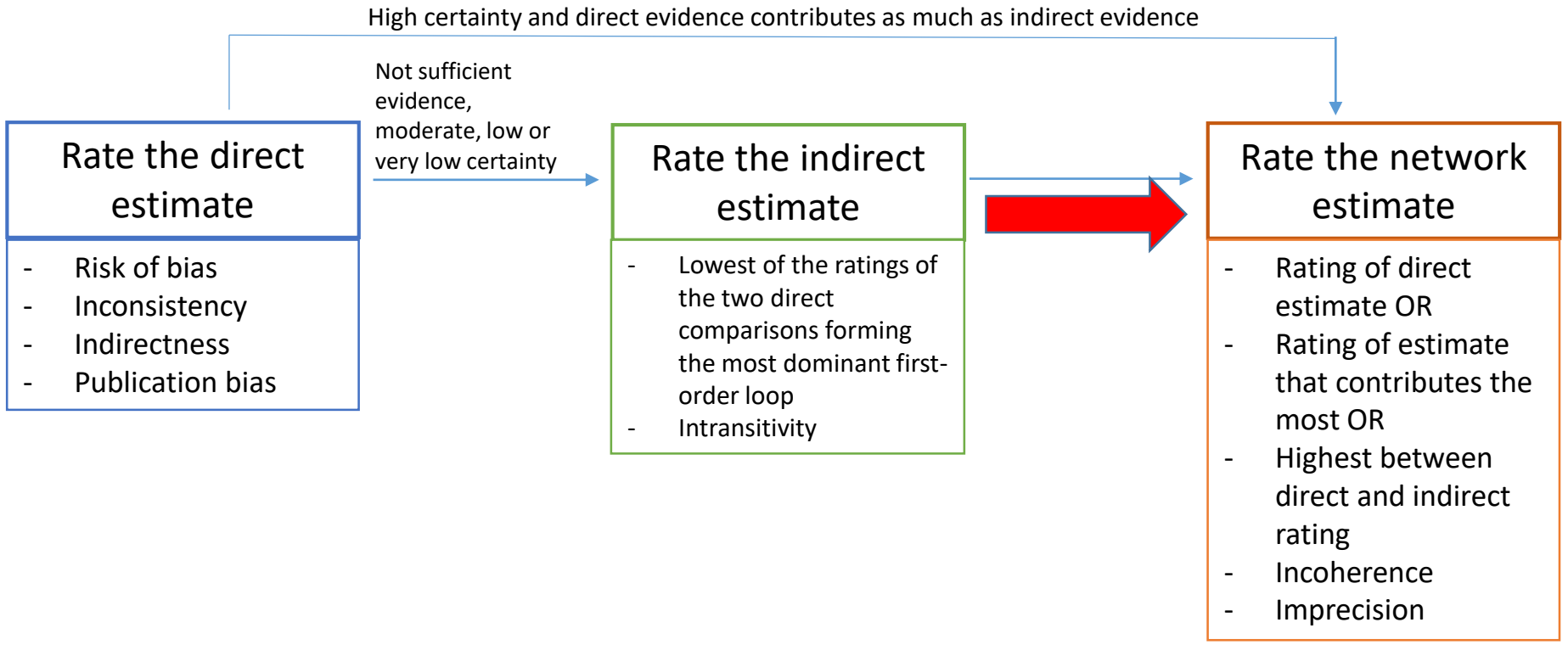
Alendronate versus Raloxifene: Rating indirect estimate

1. Most dominant first-order loop
   - Via Vitamin D+ calcium
2. Look at the rating of each of the direct estimates from that loop
   - High and moderate
3. Choose the lowest of the two ratings
   - Moderate
4. Assess intransitivity
   - Not serious

Rating: Moderate ⨁⨁⨁◯

GRADE
Network Meta-Analysis

Rating the certainty of evidence from NMA



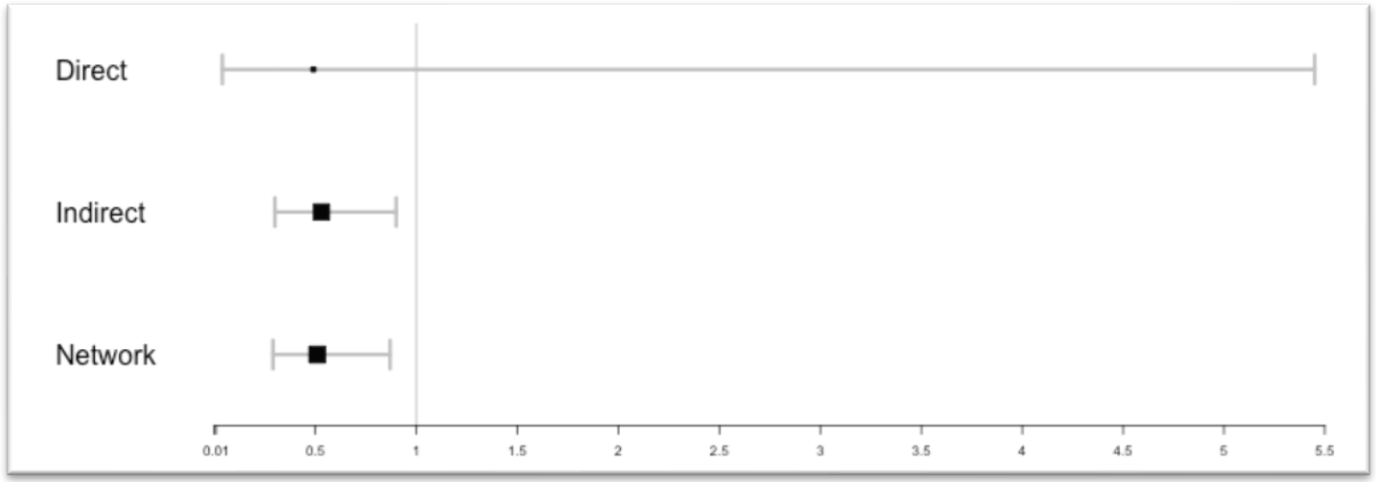
GRADE
Network Meta-Analysis

Alendronate versus Raloxifene: Rating network estimate

1. Choose the rating of the estimate that contributes the most
   - Or the highest if both contribute similarly and there is no incoherence
2. Examine for incoherence
3. Examine for imprecision

Alendronate versus Raloxifene: Estimates that contributes the most

- Indirect estimate obtained using the “node splitting approach”



Indirect estimate is contributing more the network estimate

GRADE
Network Meta-Analysis

1. Choose the rating of the evidence that contributes the most

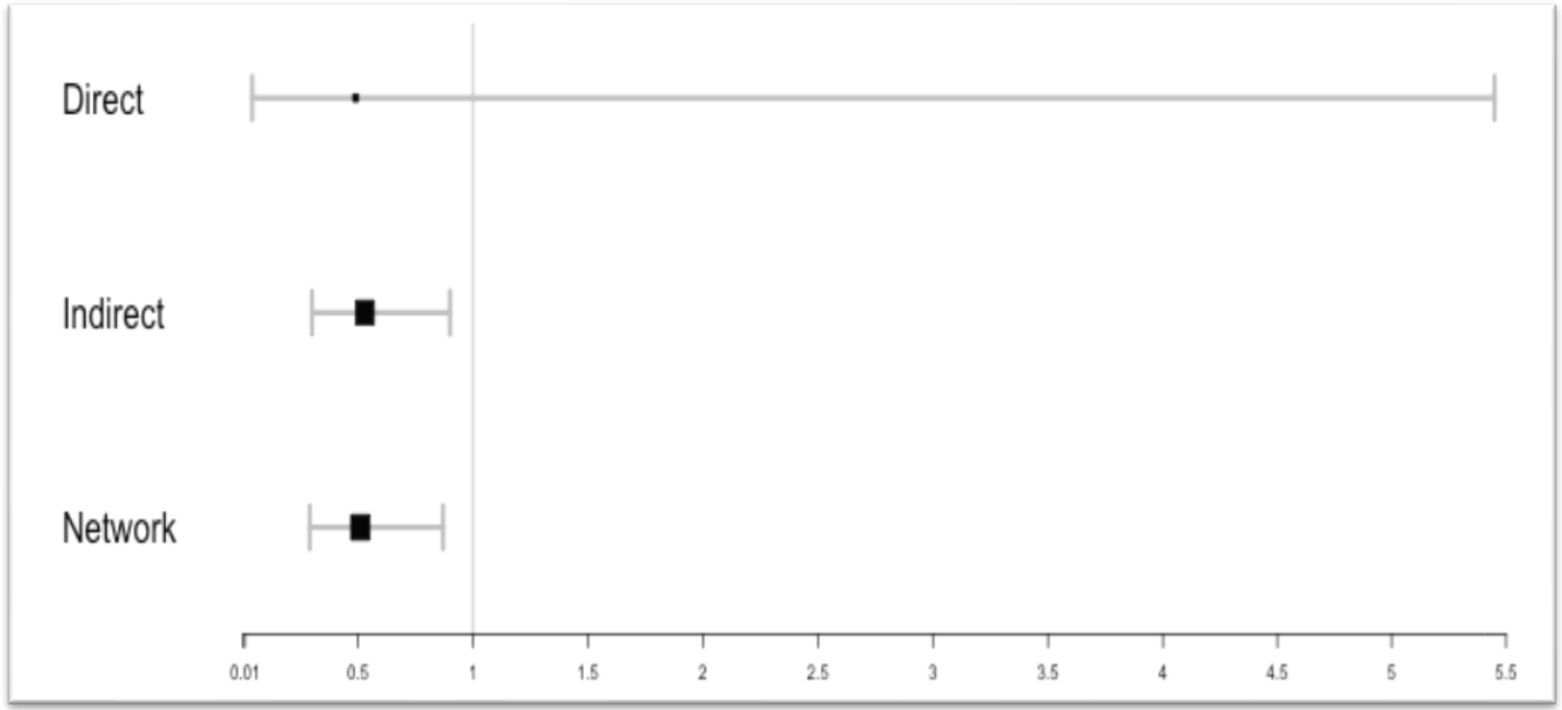
- Direct estimate: High ⨁⨁⨁⨁
- Indirect estimate: Moderate ⨁⨁⨁◯

GRADE
Network Meta-Analysis

2. Examine for incoherence

- Agreement between direct and indirect estimates
  - Similarity of point estimates
  - Overlap of confidence intervals
  - Statistical test

2. Examine for incoherence



Direct	0.49	0.04; 5.45
Indirect	0.53	0.30; 0.90

P-value test for incoherence= 0.97

2. Examine for incoherence

- Agreement between direct and indirect estimates
  - Similarity of point estimates: yes
  - Overlap of confidence intervals: yes
  - Statistical test: large p-value

Incoherence: Not serious

GRADE
Network Meta-Analysis

3. Examine for Imprecision

- Usual GRADE guidance
- Network estimate: 0.51, 95% CI 0.29; 0.87

Imprecision: Not serious

GRADE
Network Meta-Analysis

Alendronate versus Raloxifene: Rating network estimate

1. Choose between direct and indirect estimates ratings: Moderate
2. Incoherence: not serious
3. Imprecision: not serious

Final rating: Moderate ⨁⨁⨁◯

GRADE
Network Meta-Analysis

Presentation and interpretation of findings of NMA

ELSEVIER

Check for updates

Journal of Clinical Epidemiology 115 (2019) 1–13

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Development of the summary of findings table for network meta-analysis

Juan José Yepes-Nuñez[a,b], Shelly-Anne Li[c], Gordon Guyatt[a,d], Susan M. Jack[a,e], Jan L. Brozek[a,d], Joseph Beyene[a], M. Hassan Murad[f], Bram Rochwerg[a,d], Lawrence Mbuagbaw[a], Yuan Zhang[a], Ivan D. Flórez[a,g], Reed A. Siemieniuk[a], Behnam Sadeghirad[a], Reem Mustafa[a,h], Nancy Santesso[a], Holger J. Schünemann[a,d,*]

[a]*Department of Health Research Methods, Evidence, and Impact, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada*

[b]*School of Medicine, Universidad de los Andes, Carrera 7 No. 116 − 05, Bogotá, D.C., Colombia*

[c]*University of Toronto, 155 College Street, Toronto, Ontario, Canada*

[d]*Department of Medicine, McMaster University, Hamilton, Canada*

[e]*School of Nursing, McMaster University, Hamilton, Ontario, Canada*

[f]*Mayo Clinic, Evidence-Based Practice Center, 200 1st Street SW, Rochester, MN 55905, USA*

[g]*Department of Pediatrics, University of Antioquia, Calle 70 No. 52 − 21, Medellín, Colombia*

[h]*Division of Nephrology and Hypertension, University of Kansas Medical Center, Kansas City, KS, USA*

Accepted 24 April 2019; Published online 2 May 2019

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

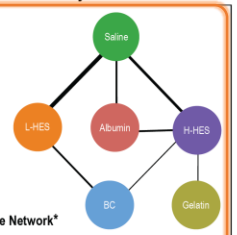
Patient or population: Critically ill patients with severe sepsis or septic shock

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 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 1

Estimates of effects, credible intervals, and certainty of the evidence for comparison fluid resuscitation in patients with sepsis

Bayesian NMA-SoF table

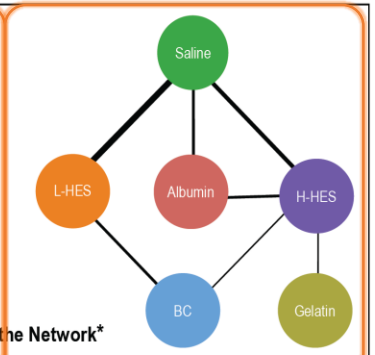
Patient or population: Critically ill patients with severe sepsis or septic shock

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

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

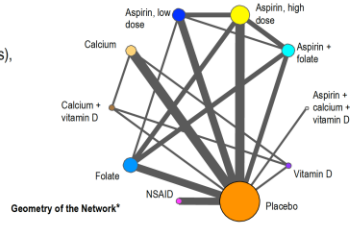
Patient or population: Individuals with previous colorectal neoplasia

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[2,5]	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
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
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
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
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
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
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

BENEFITS

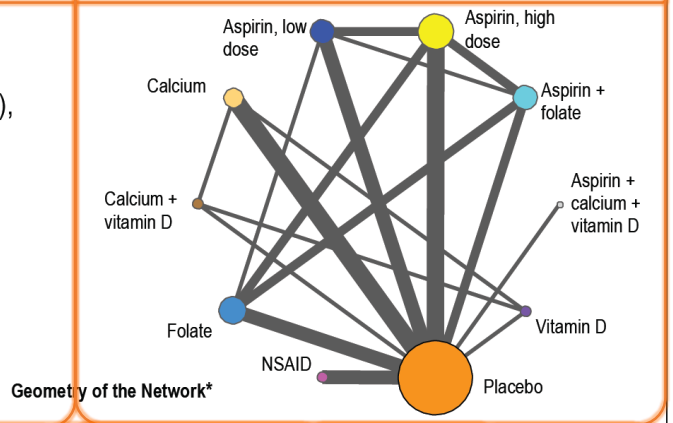
Patient or population: Individuals with previous colorectal neoplasia

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

●	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
●	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
●	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
●	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
●	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
●	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

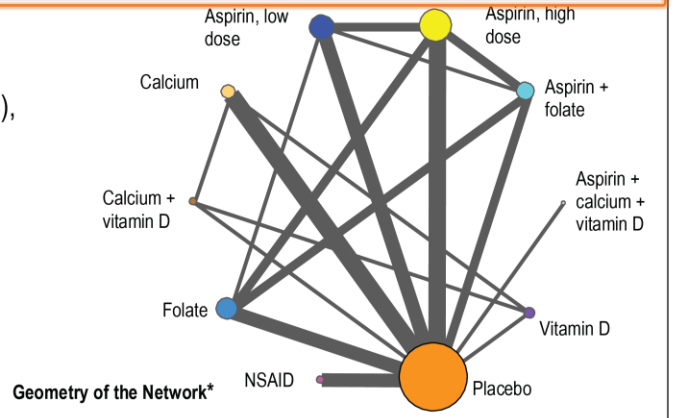
Patient or population: Individuals with previous colorectal neoplasia

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

●	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.

NM

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

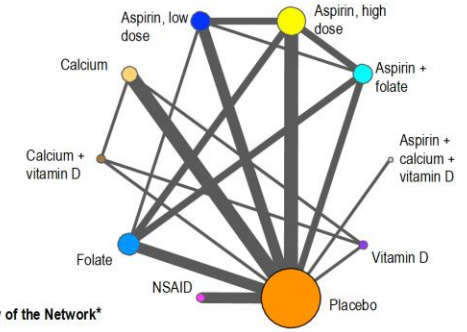
Patient or population: Individuals with previous colorectal neoplasia

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.

Partially contextualized framework for interpreting NMA

Considers the importance and the magnitude of the effects comparing the interventions without full regard for all outcomes in a PICO question

GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction

Pablo Alonso-Coello,[1,2] Holger J Schünemann,[2,3] Jenny Moberg,[4] Romina Brignardello-Petersen,[2,5] Elie A Akl,[2,6] Marina Davoli,[7] Shaun Treweek,[8] Reem A Mustafa,[2,9] Gabriel Rada,[10,11,12] Sarah Rosenbaum,[4] Angela Morelli,[4] Gordon H Guyatt,[2,3] Andrew D Oxman[4] the GRADE Working Group

For numbered affiliations see end of article.

Correspondence to:
A D Oxman oxmana@online.no

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2016;353:i2016
http://dx.doi.org/10.1136/bmj.i2016

Introduction

Healthcare decision making is complex. Decision-making processes and the factors (criteria) that decision makers should consider vary for different types of decisions, including clinical recommendations, coverage decisions, and health system or public health recommendations or decisions.[1-6] However, some criteria are relevant for all of these decisions, including the anticipated effects of the options being considered, the certainty of the evidence for those effects (also referred to as quality of evidence or confidence in effect estimates), and the costs and feasibility of the options. Decision makers must make judgments about each relevant factor, informed by the best evidence that is available to them.

Often, the processes that decision makers use, the criteria that they consider and the evidence that they use to reach their judgments are unclear.[5-8] They may omit important criteria, give undue weight to some criteria, or not use the best available evidence. Systematic and transparent systems for decision making can help to ensure that all important criteria are considered and that the best available research evidence informs decisions.

Clinicians depend on clinical practice guidelines. Rigorously developed guidelines synthesise the available relevant research, facilitating the translation of evidence into recommendations for clinical practice.[9] However, the quality of guidelines is often suboptimal.[10-13] If guidelines are not developed systematically and transparently, clinicians are not able to decide whether to rely on them or to explore disagreements when faced with conflicting recommendations.[12]

The GRADE (Grading of Recommendations Assessment, Development and Evaluation) Working Group has previously developed and refined a system to assess the certainty of evidence of effects and strength of recommendations.[13-15] More than 100 organisations globally, including the World Health Organization, the Cochrane Collaboration, and the National Institute for Health and Care Excellence (NICE) now use or have adopted the principles of the GRADE system. Recently, through the DECIDE (Developing and Evaluating Communication Strategies to Support Informed Decisions and Practice Based on Evidence) project (http://www.decide-collaboration.eu),[16] funded by the European Union, the GRADE Working Group has developed the Evidence to Decision (EtD) frameworks to support the process of moving from evidence to decisions. We have developed EtD frameworks for making clinical recommendations, coverage decisions, and health system or public health recommendations and decisions. The frameworks build on the GRADE approach to assessing the strength of recommendations.[17-19]

We developed EtD frameworks using an iterative process that is described in the project protocol.[16] The starting point for EtD frameworks was the GRADE Working Group's approach for moving from evidence to clinical recommendations.[17-19] We iteratively developed the frameworks based on reviews of relevant literature,[1-6] brainstorming, feedback from stakeholders,[20] application of EtD frameworks to a variety of recommendations and decisions, and user testing. We strove for consistency across EtD frameworks for different types of decisions, but, because of differences in the nature of the decisions, there are some differences in the frameworks. In appendix 1, we have provided a glossary of terms used in EtD frameworks, including certainty of the evidence, decisions, recommendations, and strength of recommendations.

This series of two articles describing the EtD frameworks is targeted at guideline developers and users of guidelines. This first article introduces the frameworks. It describes their purpose, development, and structure. It also describes how different organisations can adapt the frameworks to their own contexts and decision-making processes. The second article presents the framework for clinical recommendations.[21]

SUMMARY POINTS

- Clinicians, guideline developers, and policymakers sometimes neglect important criteria, give undue weight to criteria, and do not use the best available evidence to inform their judgments
- Explicit and transparent systems for decision making can help to ensure that all important criteria are considered and that decisions are informed by the best available research evidence
- The purpose of Evidence to Decision (EtD) frameworks is to help people use evidence in a structured and transparent way to inform decisions in the context of clinical recommendations, coverage decisions, and health system or public health recommendations and decisions
- EtD frameworks have a common structure that includes formulation of the question, an assessment of the evidence, and drawing conclusions, though there are some differences between frameworks for each type of decision
- EtD frameworks inform users about the judgments that were made and the evidence supporting those judgments by making the basis for decisions transparent to target audiences
- EtD frameworks also facilitate dissemination of recommendations and enable decision makers in other jurisdictions to adopt recommendations or decisions, or adapt them to their context

Size of the effect estimate	Suggested statements *(replace X with intervention, replace 'reduce/increase' with direction of effect, replace 'outcome' with name of outcome, include 'when compared with Y' when needed)*
HIGH Certainty of the evidence	
Large effect	X results in a large reduction/increase in outcome
Moderate effect	X reduces/increases outcome X results in a reduction/increase in outcome
Small important effect	X reduces/increases outcome slightly X results in a slight reduction/increase in outcome
Trivial, small unimportant effect or no effect	X results in little to no difference in outcome X does not reduce/increase outcome
MODERATE Certainty of the evidence	
Large effect	X likely results in a large reduction/increase in outcome X probably results in a large reduction/increase in outcome
Moderate effect	X likely reduces/increases outcome X probably reduces/increases outcome X likely results in a reduction/increase in outcome X probably results in a reduction/increase in outcome
Small important effect	X probably reduces/increases outcome slightly X likely reduces/increases outcome slightly X probably results in a slight reduction/increase in outcome X likely results in a slight reduction/increase in outcome
Trivial, small unimportant effect or no effect	X likely results in little to no difference in outcome X probably results in little to no difference in outcome X likely does not reduce/increase outcome X probably does not reduce/increase outcome
LOW Certainty of the evidence	
Large effect	X may result in a large reduction/increase in outcome The evidence suggests X results in a large reduction/increase in outcom
Moderate effect	X may reduce/increase outcome The evidence suggests X reduces/increases outcome X may result in a reduction/increase in outcome The evidence suggests X results in a reduction/increase in outcome
Small important effect	X may reduce/increase outcome slightly The evidence suggests X reduces/increases outcome slightly X may result in a slight reduction/increase in outcome The evidence suggests X results in a slight reduction/increase in outcom

Chapter 15: Interpreting results and drawing conclusions

Holger J Schünemann, Gunn E Vist, Julian PT Higgins, Nancy Santesso, Jonathan J Deeks, Paul Glasziou, Elie A Akl, Gordon H Guyatt; on behalf of the Cochrane GRADEing Methods Group

Key Points:

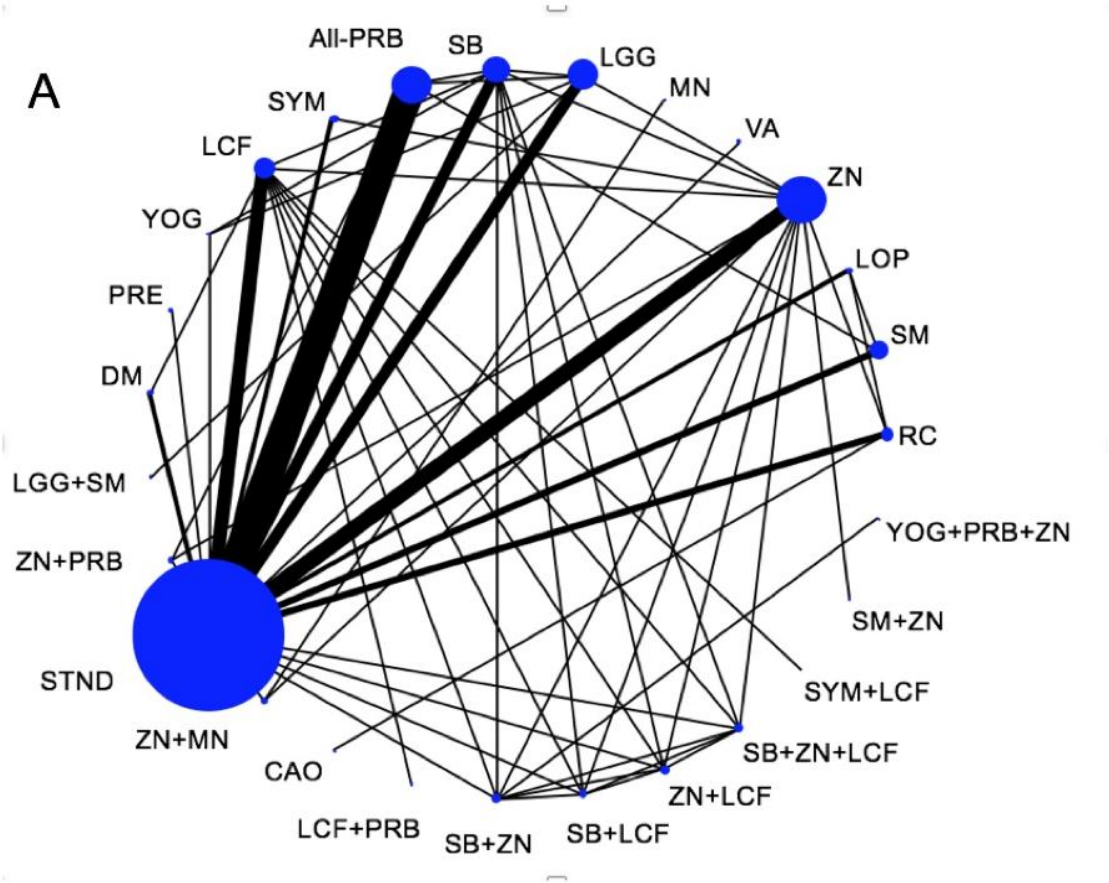
- This chapter provides guidance on interpreting the results of synthesis in order to communicate the conclusions of the review effectively.
- Methods are presented for computing, presenting and interpreting relative and absolute effects for dichotomous outcome data, including the number needed to treat (NNT).
- For continuous outcome measures, review authors can present summary results for studies using natural units of measurement or as minimal important differences when all studies use the same scale. When studies measure the same construct but with different scales, review authors will need to find a way to interpret the standardized mean difference, or to use an alternative effect measure for the meta-analysis such as the ratio of means.
- Review authors should not describe results as 'statistically significant', 'not statistically significant' or 'non-significant' or unduly rely on thresholds for P values, but report the confidence interval together with the exact P value.
- Review authors should not make recommendations about healthcare decisions, but they can – after describing the certainty of evidence and the balance of benefits and harms – highlight different actions that might be consistent with particular patterns of values and preferences and other factors that determine a decision such as cost.

Cite this chapter as: Schünemann HJ, Vist GE, Higgins JPT, Santesso N, Deeks JJ, Glasziou P, Akl EA, Guyatt GH. Chapter 15: Interpreting results and drawing conclusions. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

Example

- **NMA of the interventions for Acute Diarrhea and Gastroenteritis in Children** (Florez et al. 2019)
- **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. Steps

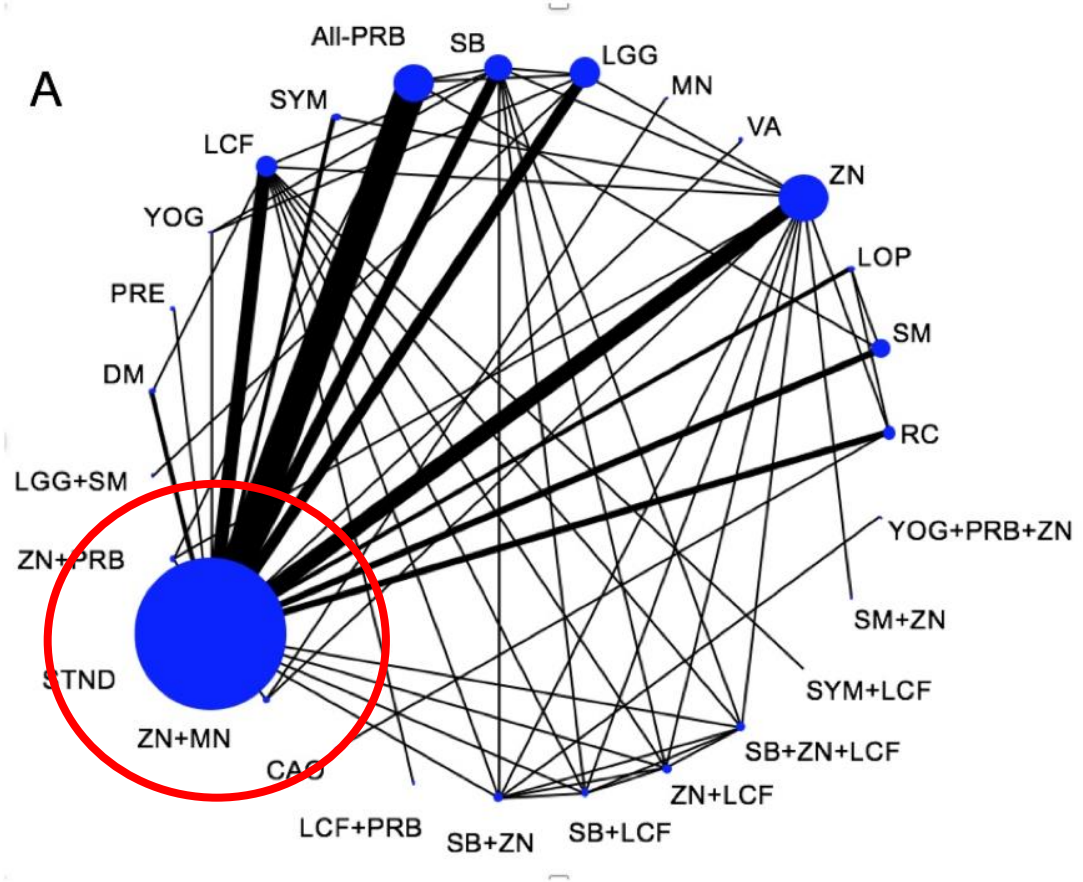
1. Choice of reference treatment and thresholds for effect sizes
2. Classification based on comparison with reference
3. Identification according to quality of evidence
4. Checking consistency with pairwise comparisons and rankings

II. Steps

1. **Choice of reference treatment and thresholds for effect sizes**
2. Classification based on comparison with reference
3. Identification according to quality of evidence
4. Checking consistency with pairwise comparisons and rankings

1. Reference and decision threshold

- Reference: treatment most connectec to others in the network
- Reference is for grouping treatments- other complementary comparators may be used for presentation
- If more than one treatment highly connected
  - Choose the one for which there is the highest quality when compared to others



1. Reference and thresholds for effect sizes

- Choose thresholds that represent
  - Small (but important) effect
  - Moderate effect
  - Large effect

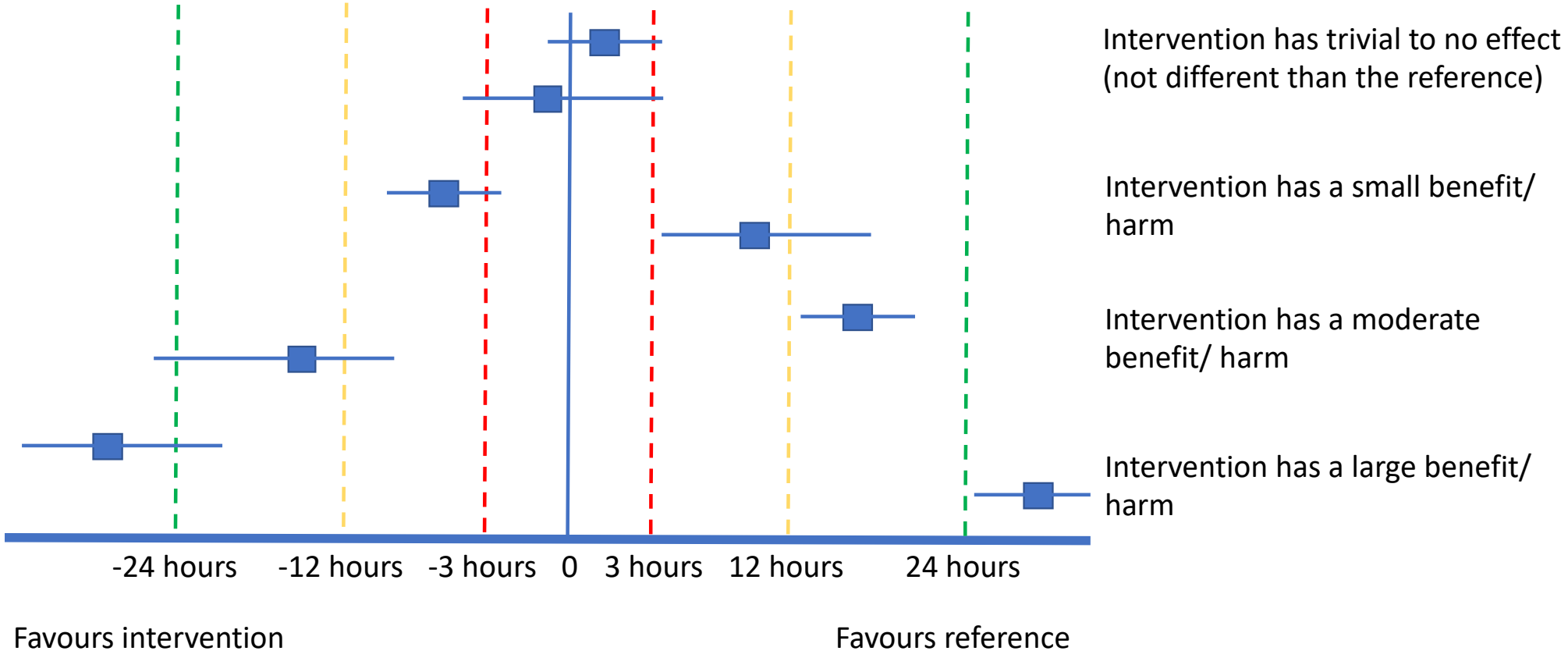
- Thresholds
  - Small but important effect: decrease or increase of 3 hours
  - Moderate effect: decrease or increase of 12 hours
  - Large effect: decrease or increase of 24 hours

II. Steps

1. Choice of reference treatment and threshold for effect sizes
2. **Classification based on comparison with reference**
3. Identification according to quality of evidence
4. Checking consistency with pairwise comparisons and rankings

2. Classification based on the comparison with the reference

- Use point estimate of relative estimate comparing each treatment versus reference



Less emphasis on imprecision

RR 0.8 CI (0.61 - 0.99)	Risk of bias -> moderate certainty
RR 0.8 CI (0.59 – 1.01)	Imprecision -> moderate certainty

2. Classification based on the comparison with the reference

- Use point estimate of the relative estimate of effect comparing each treatment versus reference
- Classify based on effect size
  - Micronutrients -0.68 → Trivial to no effect
  - Kaolin Pectin -5.32 → Small benefit
  - Zinc -18.38 → Moderate benefit
  - Zinc + probiotics -29.39 → Large benefit

Classification	Intervention	Effect on hours of diarrhea duration, MD (95%CI)
Large beneficial effect	LGG + Smectite (VL)	-51.08 (-64.30; -37.85)
	S. boulardii + Zinc (M)	-39.45 (-52.45; -26.73)
	Smectite + Zinc (M)	-35.63 (-57.57; -13.16)
	Symbiotics + LCF (VL)	-32.11 (-53.01; -11.33)
	Zinc + Probiotics (L)	-29.39 (-40.26; -18.57)
	Symbiotics (H)	-26.26 (-36.14; -16.22)
Moderate beneficial effect	Smectite (VL)	-23.90 (-30.80; -16.96)
	LGG (All) (L)	-22.74 (-28.81; -16.68)
	Zinc + LCF (M)	-21.37 (-36.54; -6.13)
	All Probiotics (L)	-19.36 (-23.66; -15.09)
	Zinc (All) (M)	-18.38 (-23.39; -13.45)
	Loperamide (M)	-17.79; (-30.35; -5.65)
	Zinc + Micronutrients (M)	-17.76 (-31.77; -4.13)
	Racecadotril (L)	-17.19 (-24.65; -9.76)
	S. boulardii + Zinc + LCF (L)	-16.74 (-36.05; 2.72)
	S. boulardii (L)	-16.48 (-23.33; -9.69)
	Yogurt (VL)	-16.43 (-30.49; -2.05)
	Yogurt + Probiotics + Zinc (VL)	-15.63 (-56.82; 26.63)
	Prebiotics (M)	-15.62 (-42.42; 11.28)
	LCF + Probiotics (VL)	-13.27 (-35.96; 9.19)
	LCF (VL)	-12.50 (-19.04; -5.99)
	S. boulardii + LCF (VL)	-12.32 (-30.01; 5.98)
Small beneficial effect	Vitamin A (VL)	-5.95 (-21.43; 9.32)
	Kaolin-Pectin (VL)	-5.32 (-33.76; 22.83)
Trivial to no effect (not different than placebo)	Micronutrients (L)	-0.68 (-33.29; 32.79)
Small harmful effect	Diluted milk (VL)	3.02 (-14.32; 8.41)

II. Steps

1. Choice of reference treatment and threshold for effect sizes
2. Classification based on comparison with reference
3. **Identification according to certainty of evidence**
4. Checking consistency with pairwise comparisons and rankings

3. Identification according to certainty of the evidence

- Use the CoE for the comparison between each intervention and the reference

Classification	Intervention	Effect on hours of diarrhea duration, MD (95%CI)	Certainty
Large beneficial effect	LGG + Smectite (VL)	-51.08 (-64.30; -37.85)	VERY LOW
	S. boulardii + Zinc (M)	-39.45 (-52.45; -26.73)	MODERATE
	Smectite + Zinc (M)	-35.63 (-57.57; -13.16)	MODERATE
	Symbiotics + LCF (VL)	-32.11 (-53.01; -11.33)	VERY LOW
	Zinc + Probiotics (L)	-29.39 (-40.26; -18.57)	LOW
	Symbiotics (H)	-26.26 (-36.14; -16.22)	HIGH
Moderate beneficial effect	Smectite (VL)	-23.90 (-30.80; -16.96)	VERY LOW
	LGG (All) (L)	-22.74 (-28.81; -16.68)	LOW
	Zinc + LCF (M)	-21.37 (-36.54; -6.13)	MODERATE
	All Probiotics (L)	-19.36 (-23.66; -15.09)	LOW
	Zinc (All) (M)	-18.38 (-23.39; -13.45)	MODERATE
	Loperamide (M)	-17.79; (-30.35; -5.65)	MODERATE
	Zinc + Micronutrients (M)	-17.76 (-31.77; -4.13)	MODERATE
	Racecadotril (L)	-17.19 (-24.65; -9.76)	LOW
	S. boulardii + Zinc + LCF (L)	-16.74 (-36.05; 2.72)	LOW
	S. boulardii (L)	-16.48 (-23.33; -9.69)	LOW
	Yogurt (VL)	-16.43 (-30.49; -2.05)	VERY LOW
	Yogurt + Probiotics + Zinc (VL)	-15.63 (-56.82; 26.63)	VERY LOW
	Prebiotics (M)	-15.62 (-42.42; 11.28)	VERY LOW
	LCF + Probiotics (VL)	-13.27 (-35.96; 9.19)	VERY LOW
	LCF (VL)	-12.50 (-19.04; -5.99)	VERY LOW
	S. boulardii + LCF (VL)	-12.32 (-30.01; 5.98)	VERY LOW
Small beneficial effect	Vitamin A (VL)	-5.95 (-21.43; 9.32)	VERY LOW
	Kaolin-Pectin (VL)	-5.32 (-33.76; 22.83)	VERY LOW
Trivial to no effect	Micronutrients (L)	-0.68 (-33.29; 32.79)	LOW
Small harmful effect	Diluted milk (VL)	3.02 (-14.32; 8.41)	VERY LOW

II. Steps

1. Choice of reference treatment and threshold for effect sizes
2. Classification based on comparison with reference
3. Identification according to quality of evidence
4. **Checking consistency with pairwise comparisons and rankings**

4. Checking consistency with pairwise comparisons and rankings

- Make sure that classification is consistent with pairwise comparisons between non-reference treatments (estimates and QoE)
- Smectite + Zinc → moderate QoE of large benefit
- Vit A → very low QoE small benefit
- Smectite + Zinc vs Vit A → MD, -29.54 (95% CI -56.09 to -2.84 , moderate quality evidence) → Smectite probably has a larger benefit than Vit A

4. Checking consistency with pairwise comparisons and rankings

- Make sure that classification is consistent with rankings

Classification	Intervention	Effect on hours of diarrhea duration, MD (95%CI)	SUCRA	Certainty
Large beneficial effect	LGG + Smectite (VL)	-51.08 (-64.30; -37.85)	1.00 (0.92; 1.00)	VERY LOW
	S. boulardii + Zinc (M)	-39.45 (-52.45; -26.73)	0.92 (0.77; 1.00)	MODERATE
	Smectite + Zinc (M)	-35.63 (-57.57; -13.16)	0.88 (0.35; 1.00)	MODERATE
	Symbiotics + LCF (VL)	-32.11 (-53.01; -11.33)	0.85 (0.27; 1.00)	VERY LOW
	Zinc + Probiotics (L)	-29.39 (-40.26; -18.57)	0.81 (0.5; 0.96)	LOW
	Symbiotics (H)	-26.26 (-36.14; -16.22)	0.77 (0.38; 0.92)	HIGH
Moderate beneficial effect	Smectite (VL)	-23.90 (-30.80; -16.96)	0.69 (0.42; 0.88)	VERY LOW
	LGG (All) (L)	-22.74 (-28.81; -16.68)	0.65 (0.38; 0.85)	LOW
	Zinc + LCF (M)	-21.37 (-36.54; -6.13)	0.61 (0.19; 0.92)	MODERATE
	All Probiotics (L)	-19.36 (-23.66; -15.09)	0.54 (0.31; 0.73)	LOW
	Zinc (All) (M)	-18.38 (-23.39; -13.45)	0.50 (0.27; 0.69)	MODERATE
	Loperamide (M)	-17.79; (-30.35; -5.65)	0.46 (0.15; 0.85)	MODERATE
	Zinc + Micronutrients (M)	-17.76 (-31.77; -4.13)	0.46 (0.15; 0.85)	MODERATE
	Racecadotril (L)	-17.19 (-24.65; -9.76)	0.46 (0.23; 0.73)	LOW
	S. boulardii + Zinc + LCF (L)	-16.74 (-36.05; 2.72)	0.42 (0.08; 0.88)	LOW
	S. boulardii (L)	-16.48 (-23.33; -9.69)	0.42 (0.19; 0.69)	LOW
	Yogurt (VL)	-16.43 (-30.49; -2.05)	0.42 (0.11; 0.85)	VERY LOW
	Yogurt + Probiotics + Zinc (VL)	-15.63 (-56.82; 26.63)	0.38 (0.00; 1.00)	VERY LOW
	Prebiotics (M)	-15.62 (-42.42; 11.28)	0.38 (0.00; 0.96)	VERY LOW
	LCF + Probiotics (VL)	-13.27 (-35.96; 9.19)	0.31 (0.00; 0.88)	VERY LOW
	LCF (VL)	-12.50 (-19.04; -5.99)	0.31 (0.15; 0.54)	VERY LOW
	S. boulardii + LCF (VL)	-12.32 (-30.01; 5.98)	0.27 (0.04; 0.81)	VERY LOW
Small beneficial effect	Vitamin A (VL)	-5.95 (-21.43; 9.32)	0.19 (0.00; 0.61)	VERY LOW
	Kaolin-Pectin (VL)	-5.32 (-33.76; 22.83)	0.15 (0.00; 0.89)	VERY LOW
Trivial to no effect	Micronutrients (L)	-0.68 (-33.29; 32.79)	0.08 (0.00; 0.85)	LOW
Small harmful effect	Diluted milk (VL)	3.02 (-14.32; 8.41)	0.04 (0.00; 0.23)	VERY LOW

Conclusions

- When considering all the interventions, S. boulardi+ Zinc, Smectite + Zinc, and Symbiotics result in a large reduction of diarrhea duration
- When considering all the interventions, LGG+ Smectite, Symbiotics + LCF, and Zinc + Probiotics may result in a large reduction of diarrhea duration
- When considering all the interventions, Zinc+ LCF, Zinc, Loperamide, and Zinc+ Micronutrients result in a moderate reduction of diarrhea duration

GRADE
Network Meta-Analysis

Final considerations

- Each framework presented in a separate paper
- Main change based on feedback: same example in both papers
- What do these frameworks add
  - Guiding principles
  - Process based on the degree of contextualization; consistency with EtD work
- What do these frameworks not create
  - Contextualization
  - How to interpret evidence

GRADE
Network Meta-Analysis

Conclusions

GRADE approach to rating certainty in NMA estimates

Summary of Findings Tables for NMA

Interpretation of results – key issues – four steps for consideration