Journal of Clinical Epidemiology 169 (2024) 111276

Journal of Clinical Epidemiology

OTHER GRADE PAPERS

The GRADE Working Group and CINeMA approaches provided inconsistent certainty of evidence ratings for a network meta-analysis of opioids for chronic noncancer pain

Atefeh Noori[a,b,c], Behnam Sadeghirad[a], Lehana Thabane[a], Mohit Bhandari[a,d], Gordon H. Guyatt[a], Jason W. Busse[a,e,f,*]

[a]*Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada*
[b]*The Michael G. DeGroote National Pain Center, McMaster University, Hamilton, Ontario, Canada*
[c]*Hand Program, Division of Plastic, Reconstructive and Aesthetic Surgery, University Health Network, Toronto Western Hospital, Affiliated with University of Toronto*
[d]*Division of Orthopaedic Surgery, Department of Surgery, McMaster University, Hamilton, Ontario, Canada*
[e]*Department of Anesthesia, McMaster University, Hamilton, Ontario, Canada*
[f]*The Michael G. DeGroote National Pain Centre, McMaster University, Hamilton, Ontario, Canada*

Accepted 4 February 2024; Published online 8 February 2024

Nantharat Apiwantanakul

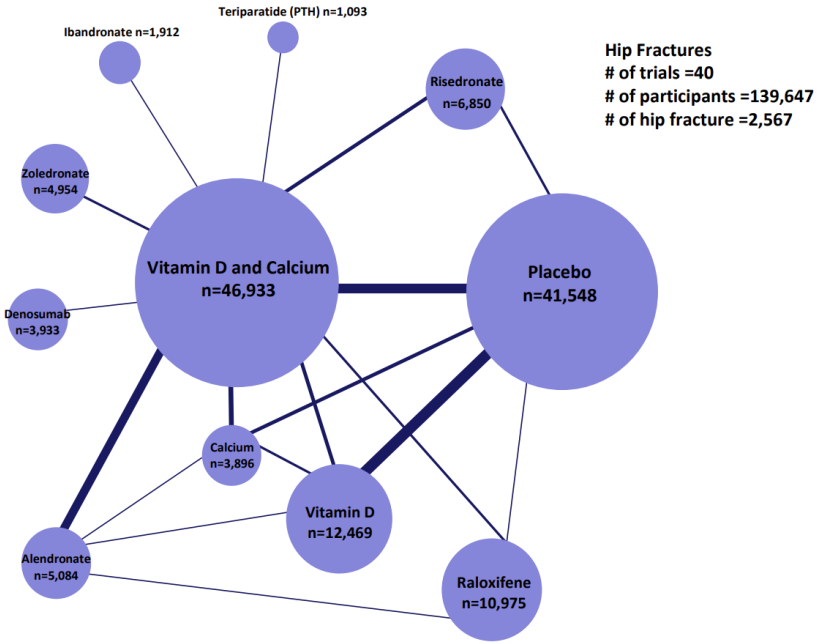
Clinical Epidemiology

06 June 2025

Comparative Effectiveness of Drug Treatments to Prevent Fragility Fractures: A Systematic Review and Network Meta-Analysis

Mohammad Hassan Murad, Matthew T. Drake, Rebecca J. Mullan, Karen F. Mauck, Louise M. Stuart, Melanie A. Lane, Nisrin O. Abu Elnour, Patricia J. Erwin, Ahmad Hazem, Milo A. Puhan, Tianjing Li, and Victor M. Montori

P: Postmenopausal women at risk of developing fragility fractures
I: Drug treatments
O: Hip fracture
S: RCT



Treatment	*P* (best)
Placebo	0.00
Teriparatide[a]	0.42
Denosumab[a]	0.13
Raloxifene[b]	0.00
Zoledronate[a]	0.05
Risedronate[b]	0.06
Ibandronate[a]	0.21
Alendronate[b]	0.14
Vitamin D[c]	0.00
Vitamin D + calcium[c]	0.00
Calcium[c]	0.00

Highest probability of being ranked first in hip fracture reduction

↓

To infer for clinical decision-making
(How likely they are to be true)

↓

Assessment of the certainty of evidence

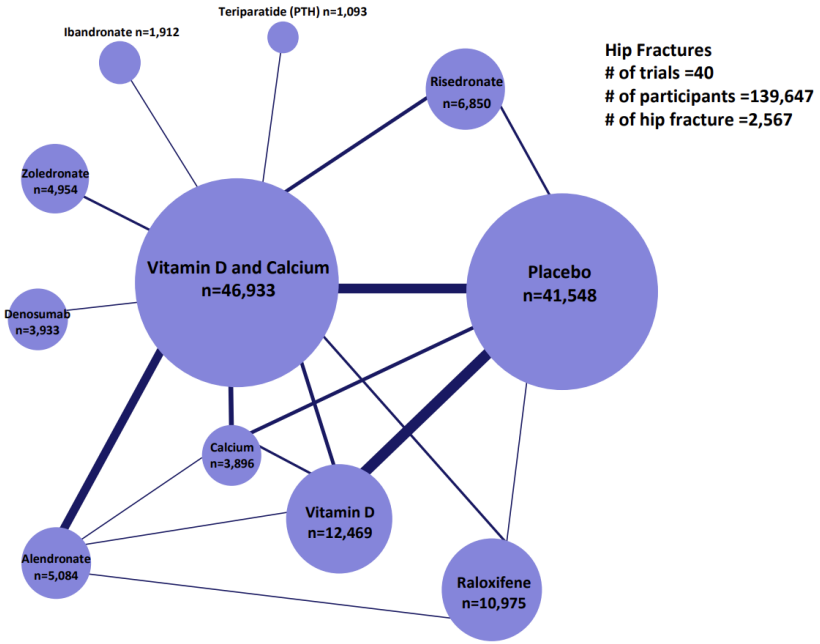
Murad MH, Drake MT, Mullan RJ, Mauck KF, Stuart LM, Lane MA, et al. Clinical review. Comparative effectiveness of drug treatments to prevent fragility fractures: a systematic review and network meta-analysis. J Clin Endocrinol Metab 2012;97:1871–80.

Introduction

- Certainty of the evidence (CoE)
  - ~ Confidence in the estimates of effects
  - ~ Quality of the evidence
- Provide ratings for the confidence in the results from NMA
  - Can be high, moderate, low or very low

Comparative Effectiveness of Drug Treatments to Prevent Fragility Fractures: A Systematic Review and Network Meta-Analysis

Mohammad Hassan Murad, Matthew T. Drake, Rebecca J. Mullan, Karen F. Mauck, Louise M. Stuart, Melanie A. Lane, Nisrin O. Abu Elnour, Patricia J. Erwin, Ahmad Hazem, Milo A. Puhan, Tianjing Li, and Victor M. Montori



	Network meta-analysis	
Comparison	Odds ratio (95% credible interval)	Quality of evidence
Teriparatide *v* placebo	0.42 (0.10 to 1.82)	Very low
Denosumab *v* placebo	0.50 (0.27 to 0.86)	High
Raloxifene *v* placebo	0.87 (0.63 to 1.22)	Moderate
Zoledronate *v* placebo	0.50 (0.34 to 0.73)	High
Risedronate *v* placebo	0.48 (0.31 to 0.66)	Low
Ibandronate *v* placebo	0.49 (0.21 to 1.20)	Very low
Alendronate *v* placebo	0.45 (0.27 to 0.68)	Moderate
Vitamin D *v* placebo	1.13 (0.94 to 1.34)	Low
Vitamin D+calcium *v* placebo	0.81 (0.68 to 0.96)	Moderate
Calcium *v* placebo	1.14 (0.82 to 1.59)	Moderate
Denosumab *v* teriparatide	1.17 (0.24 to 5.54)	Low
Raloxifene *v* teriparatide	2.05 (0.47 to 9.47)	Very low
Zoledronate *v* teriparatide	1.18 (0.26 to 5.30)	Very low
Risedronate *v* teriparatide	1.12 (0.25 to 4.98)	Very low
Ibandronate *v* teriparatide	1.11 (0.22 to 6.42)	Very low
Alendronate *v* teriparatide	1.02 (0.24 to 4.82)	Very low
Vitamin D *v* teriparatide	2.67 (0.63 to 11.97)	Very low
Vitamin D+calcium *v* teriparatide	1.92 (0.45 to 8.42)	Low
Calcium *v* teriparatide	2.69 (0.63 to 12.23)	Very low

SUCRA calculates the probability of being the best treatments, however, this approach does not consider the certainty of evidence

Evidence supporting teriparatide is much lower quality than evidence supporting other lower ranked treatment.

↓

The strength of inferences for clinical decision-making

Puhan MA, Schünemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, Kessels AG, Guyatt GH; GRADE Working Group. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ. 2014 Sep 24;349:g5630.

Introduction

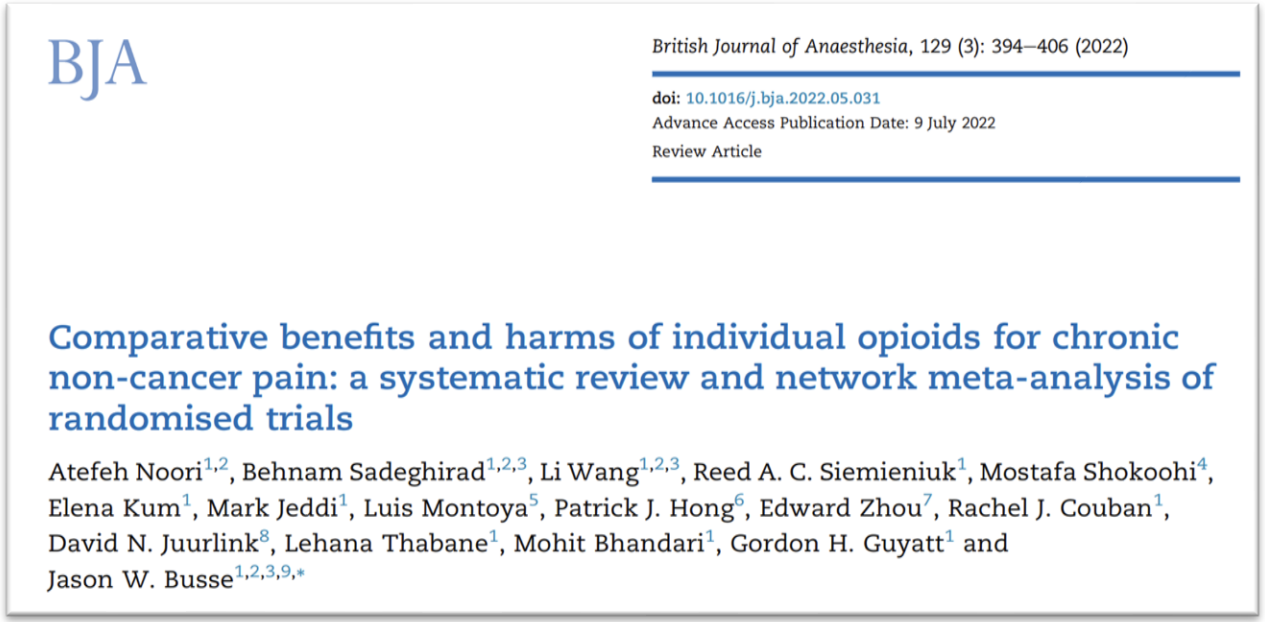
- Reports of NMAs often describe the risk of bias of included trials
- Risk of bias is *only one determinant* of certainty of the evidence.
  - E.g., the confidence in effect estimates will decrease
    - If there are large differences in results from study to study
    - If results are imprecise
- Assess the CoE of treatment effects informed by network meta-analysis
  - GRADE Working Group (GWG) approach
  - Confidence in Network Meta-Analysis (CINeMA) framework

GRADE working group

CINeMA
Confidence in Network Meta-Analysis

Objective

- Compare the concordance of CoEs between the GWG and CINeMA approaches for evaluating the certainty of network estimates
- NMA of opioids for chronic noncancer pain



BJA

British Journal of Anaesthesia, 129 (3): 394–406 (2022)

doi: 10.1016/j.bja.2022.05.031
Advance Access Publication Date: 9 July 2022
Review Article

Comparative benefits and harms of individual opioids for chronic non-cancer pain: a systematic review and network meta-analysis of randomised trials

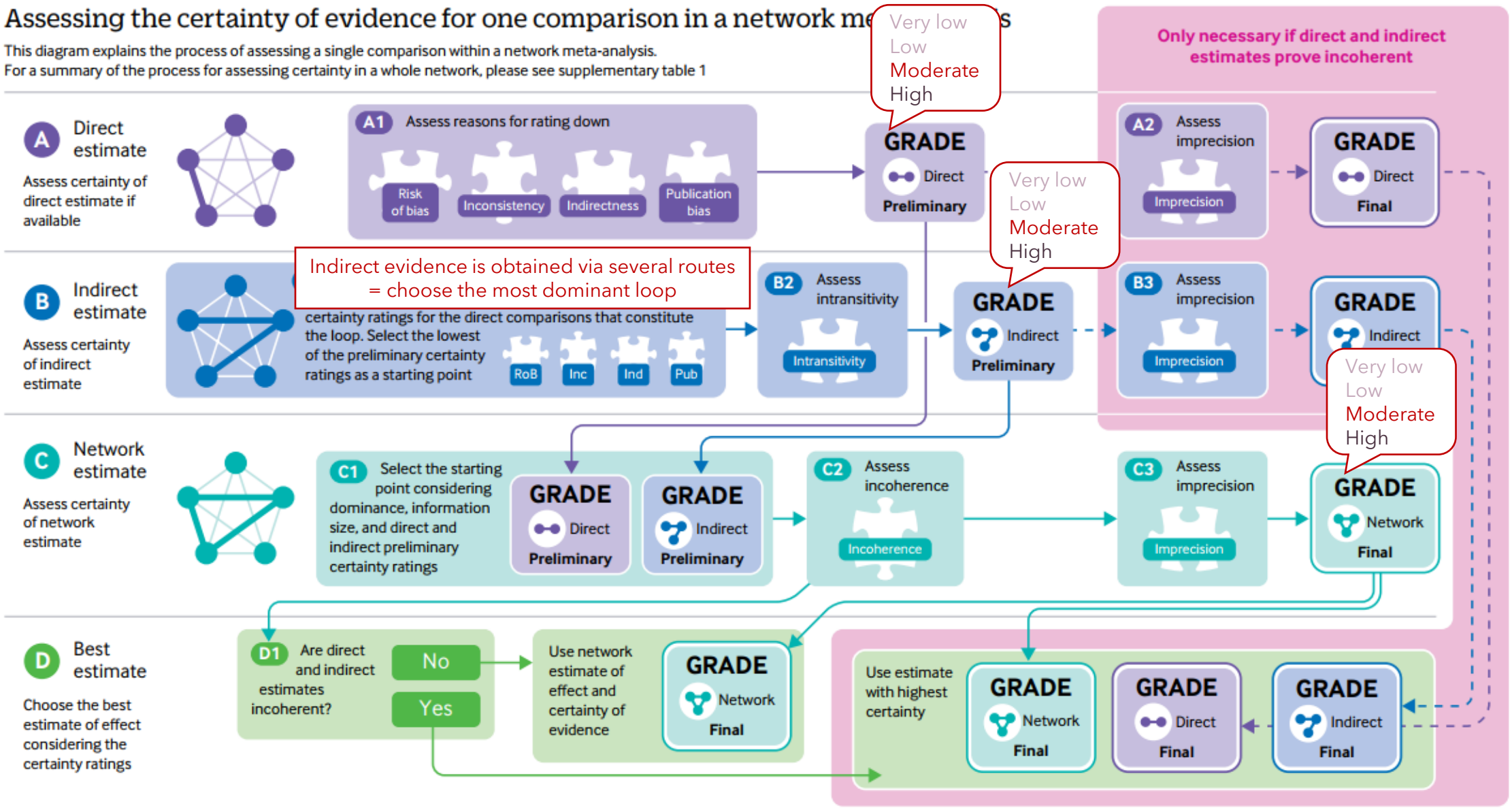
Atefeh Noori[1,2], Behnam Sadeghirad[1,2,3], Li Wang[1,2,3], Reed A. C. Siemieniuk[1], Mostafa Shokoohi[4], Elena Kum[1], Mark Jeddi[1], Luis Montoya[5], Patrick J. Hong[6], Edward Zhou[7], Rachel J. Couban[1], David N. Juurlink[8], Lehana Thabane[1], Mohit Bhandari[1], Gordon H. Guyatt[1] and Jason W. Busse[1,2,3,9,*]

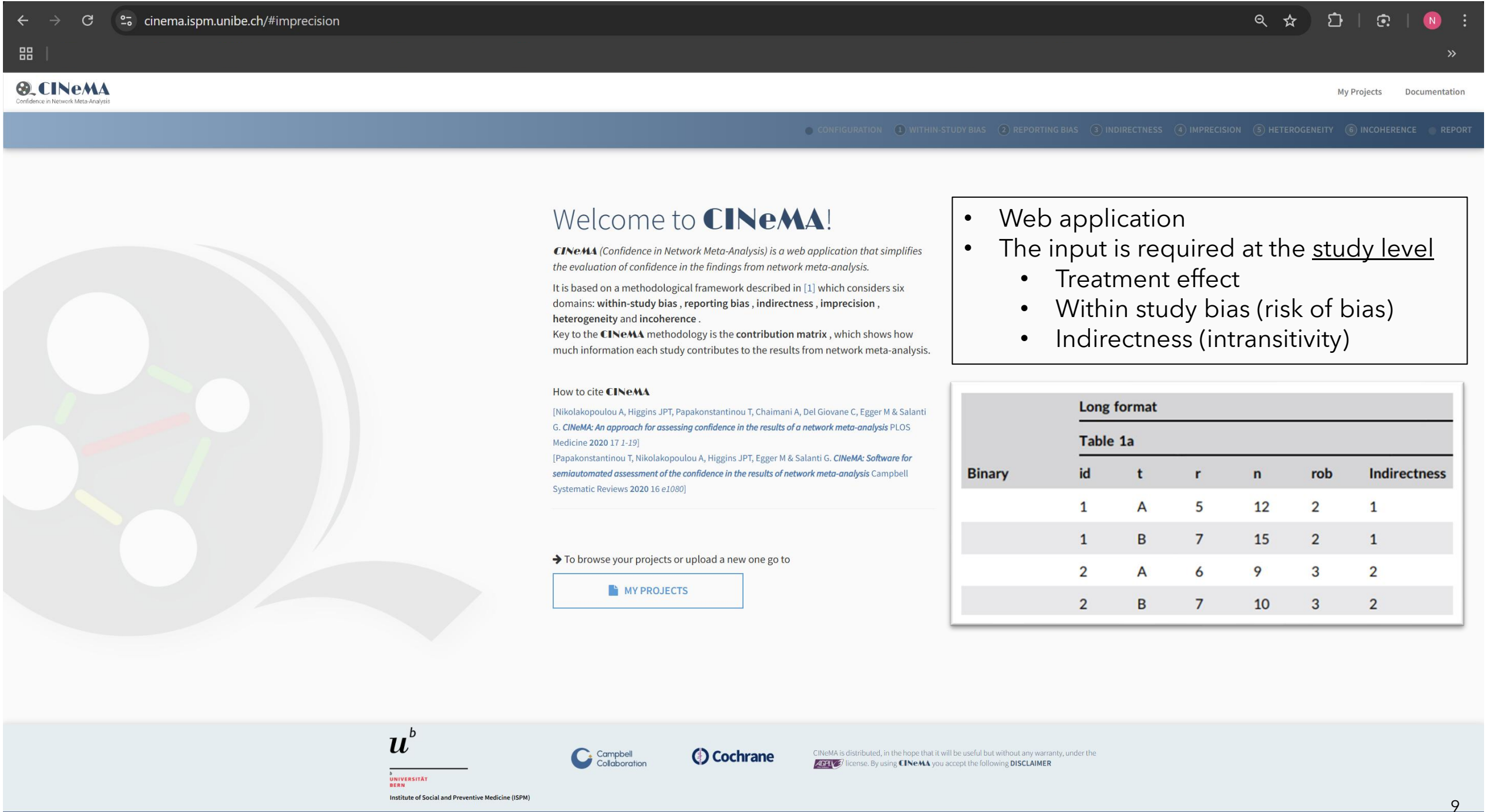
P	Patients with chronic noncancer pain
I	Opioids
C	Placebo
O	Pain relief, Physical functioning
S	RCT

Methods

- Assess the CoE for treatment effects using the GWG approach and CINeMA framework
  - For pain relief outcome
  - For physical functioning outcome
- Both approaches evaluate the CoE as high, moderate, low or very low.
- Quantified the number of discrepant CoE ratings between approaches and the magnitude of the difference (ie, one level, two levels, or three levels)
- GWG approach
  - Two reviewers, independently, resolved any discrepancies by discussion
- CINeMA framework
  - One reviewer, the senior author confirmed the final assessment.

GRADE working group

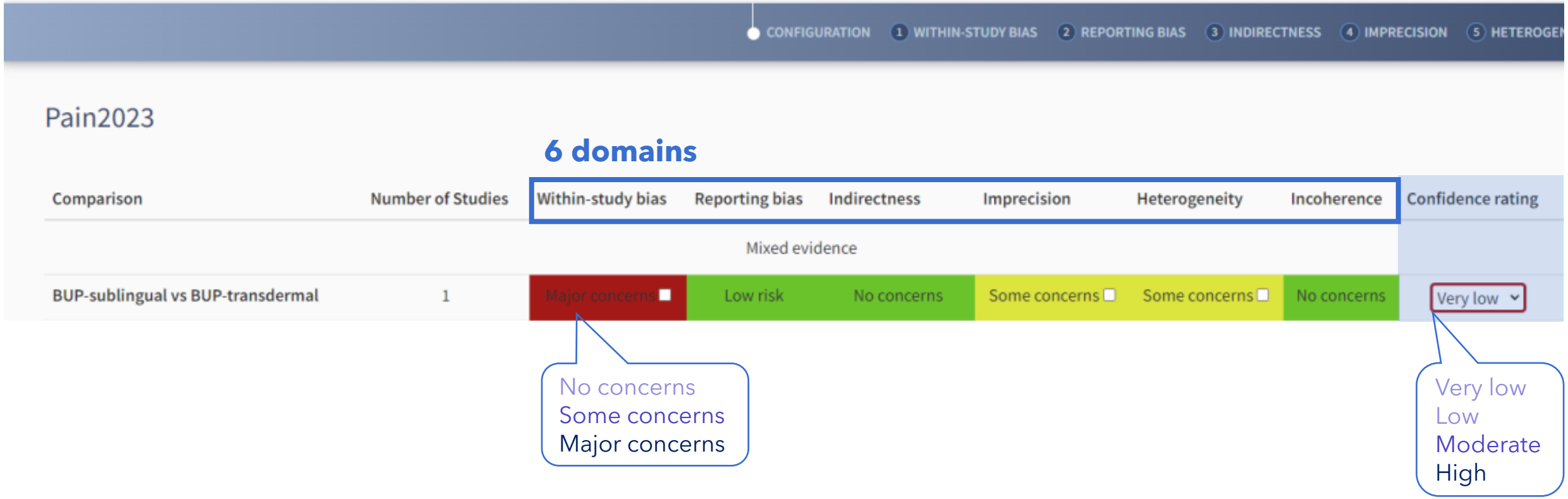




- Web application
- The input is required at the <u>study level</u>
  - Treatment effect
  - Within study bias (risk of bias)
  - Indirectness (intransitivity)

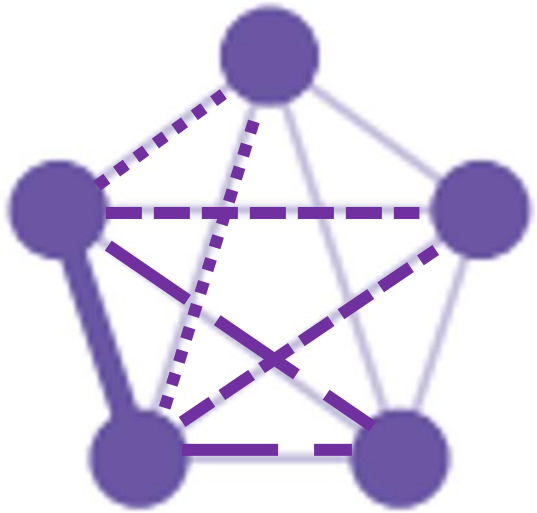
	Long format					
	Table 1a					
Binary	id	t	r	n	rob	Indirectness
	1	A	5	12	2	1
	1	B	7	15	2	1
	2	A	6	9	3	2
	2	B	7	10	3	2

CINeMA
Confidence in Network Meta-Analysis



CINeMA
Confidence in Network Meta-Analysis

- Network estimates resulted from direct and indirect evidences
- There are multiple comparisons informing indirect evidence
- Evaluate the impact of every study included in the network using percentage contribution matrix
  - The percentage contribution from studies judged to be at low, moderate, and high risk of bias are presented for each comparison in bar charts.



Both approaches incorporate risk of bias, publication bias, indirectness, imprecision, heterogeneity and incoherence; however, there are several conceptual differences between them.

Table 1. Comparison in approaches between GRADE Working Group (GWG) and CINeMA approaches to derive the overall certainty of evidence

	Direct evidence		Indirect evidence		Network evidence	
Domain assessment	**GWG**	**CINeMA framework**	**GWG**	**CINeMA framework**	**GWG**	**CINeMA framework**
Study limitations	Yes	Yes	No	Yes	-	-
Indirectness	Yes	Yes	No	Yes	-	-
Heterogeneity	Yes	Yes	No	Yes	-	-
Publication bias	Yes	Yes	No	Yes	-	-
Intransitivity[a]	No	No	Yes	No	-	-
Imprecision	-	-	-	-	Yes	Yes
Incoherence	-	-	-	-	Yes	Yes
Automatic compute	-	-	-	-	No	Yes
Overall rating across domains	Yes	No	Yes	No	Yes	Yes

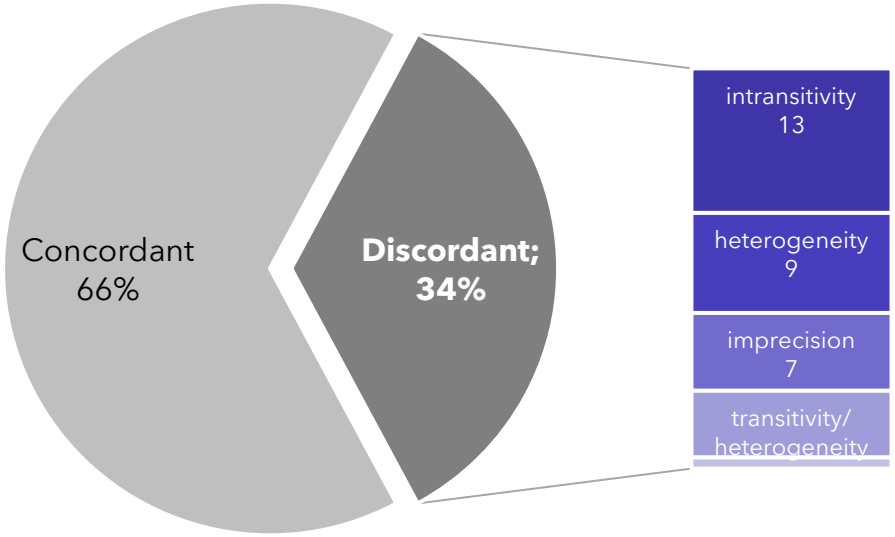
Abbreviations: CINeMA, Confidence in Network Meta-Analysis; GRADE, Grading of Recommendations Assessment, Development and Evaluation; GWG, GRADE Working Group.

[a] Intransitivity is assessed as a part of the consideration of indirectness with the CINeMA framework.

Results - discrepancies

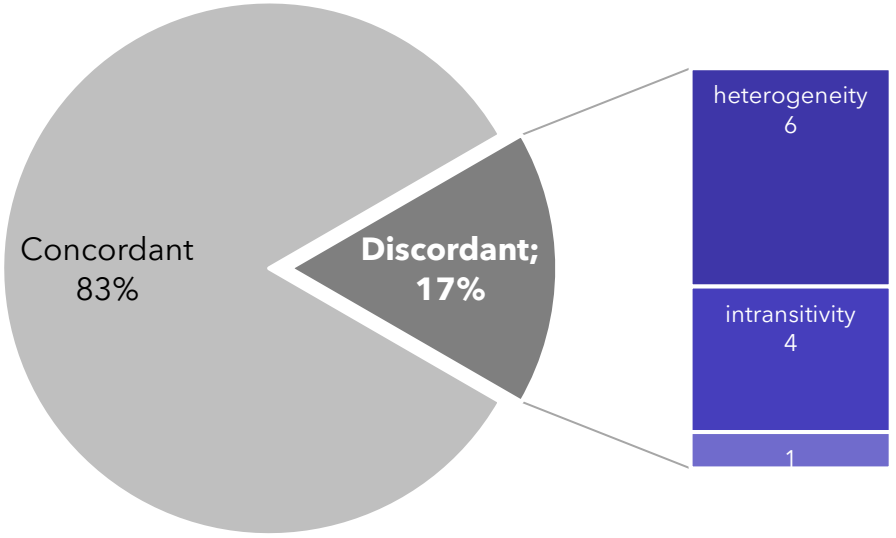
Pain relief

- GWG and CINeMA approaches provided different CoE ratings in 34% of cases (36 of 105)



Physical functioning

- GWG and CINeMA approaches provided different CoE ratings in 17% of cases (11 of 66)



Results - discrepancies

Pain relief

- All discrepancies were separated by one level
  - (26/36) the GWG resulted in higher certainty ratings than the CINeMA
    - 25; Low vs Very low
    - 1; Moderate vs Low
  - (10/36) the GWG resulted in lower certainty ratings than the CINeMA
    - 10; Very Low vs Low

Physical functioning

- All discrepancies were separated by one level
  - (11/11) the GWG resulted in higher certainty ratings than the CINeMA
    - Low vs Very low

Transitivity assessment

GWG approach

- A credible subgroup effect in a previous meta-analysis of opioids based on the duration of follow-up (<3 months vs. ≥3 months)
- When there was a large imbalance (50% or more difference) in the length of follow-up between the two direct comparisons
  - Rated down the CoE of indirect evidence one level for intransitivity

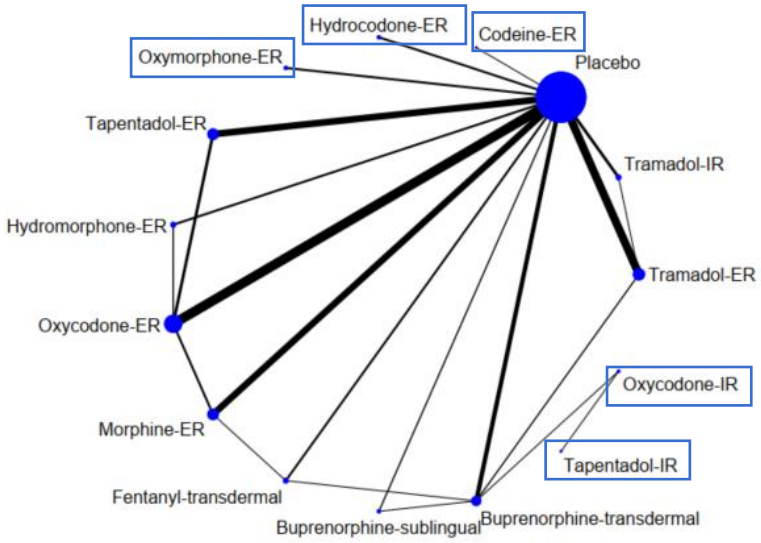
CINeMA framework

- Evaluating each included study according to its relevance to the research question in terms of effect modifying variables.
- The effect modifier distribution can only be evaluated in networks with enough studies available for a specific comparison.
  - Downgraded the CoE for opioids that were assessed in a single study or only compared vs placebo without direct comparisons with other opioids.

Results - intransitivity

Pain relief

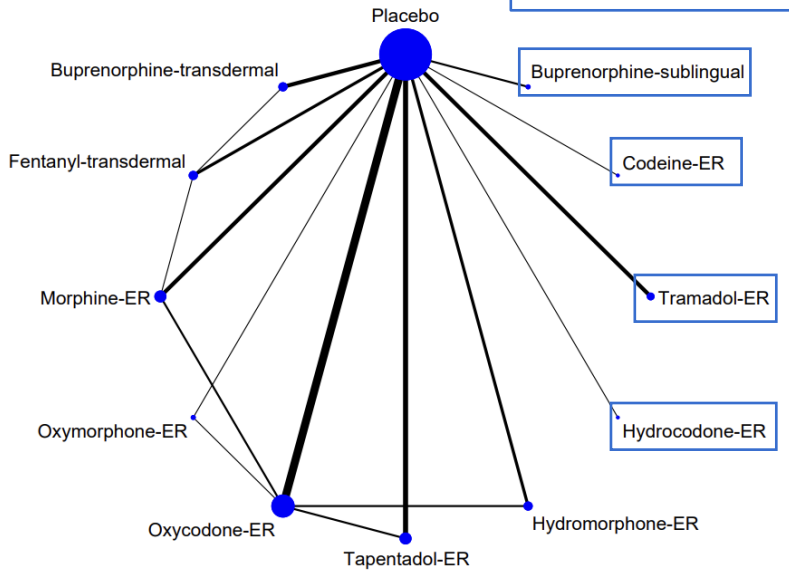
In the CINeMA framework, a total of 58 comparisons were rated down



In the GWG approach, 36% (21 out of 58) of them were rated down for violating the transitivity assumption

Physical functioning

In the CINeMA framework, a total of 41 comparisons were rated down



Using the GWG approach, (78%) 32 out of 41 of them were rated down due to intransitivity.

Heterogeneity assessment

GWG approach

- The GWG evaluates heterogeneity for direct evidence based on the I^2 statistic
- *(only comparisons for which at least two studies included the comparison)*

CINeMA framework

- CINeMA considers heterogeneity for each network estimate according to the agreement between the 95% CI and the prediction interval, which is a range of values between which the true effect of a new study is likely to lie, in relation to the range of equivalence

Results – heterogeneity

GWG approach

- GWG only evaluated 15 out of the 105 comparisons for heterogeneity.

CINeMA framework

- CINeMA rated down 55 out of 105 comparisons (52%) for heterogeneity

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
GWG	no	no	no	no	no	no	no	**concern**	**concern**	no	no	no	**concern**	**concern**	**concern**
CINeMA	**concern**	**concern**	**concern**	**concern**	**concern**	**concern**	**concern**	no	no	no	no	no	**concern**	**concern**	**concern**

Imprecision

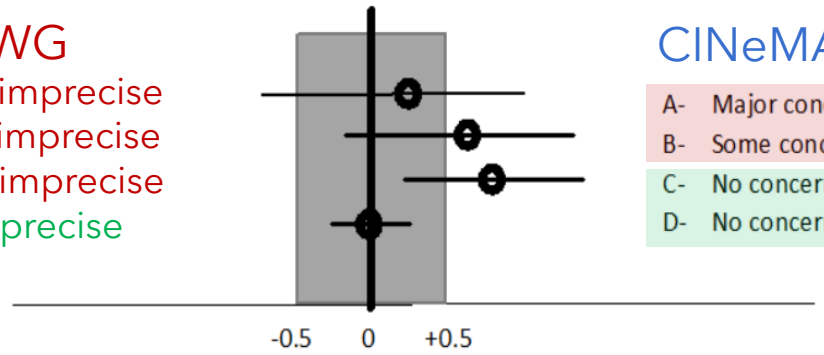
Clinically important effect was set at **half of MID**

- 0.5 cm for pain on a 0-10 cm visual analog scale
- 2.5 points for physical function on the 0-100 points short form-36 physical component summary score

GWG approach

- Rated the network estimate as ***imprecise*** if the associated 95% CI included half of the MID:

GWG
A- imprecise
B- imprecise
C- imprecise
D- precise



CINeMA
A- Major concern
B- Some concern
C- No concern
D- No concern

CINeMA framework

- CINeMA uses two strategies to rate imprecision of the network estimate:
- (1) if the point estimate is less than the threshold then the estimate is considered ***imprecise*** if the associated 95% CI includes half of the MID
- (2) if the point estimate is greater than the threshold then the estimate is considered ***imprecise*** if the associated 95% CI includes the null effect

Results – imprecision

GWG approach

- rated down the network estimates for transdermal buprenorphine and fentanyl, ER hydromorphone, ER hydrocodone, ER oxycodone, and IR oxycodone vs placebo for imprecision, as the associated 95% CI included half the MID.
- rated down the CoE (ER codeine vs placebo) one level due to imprecision, as the direct evidence was informed by less than 300 observations

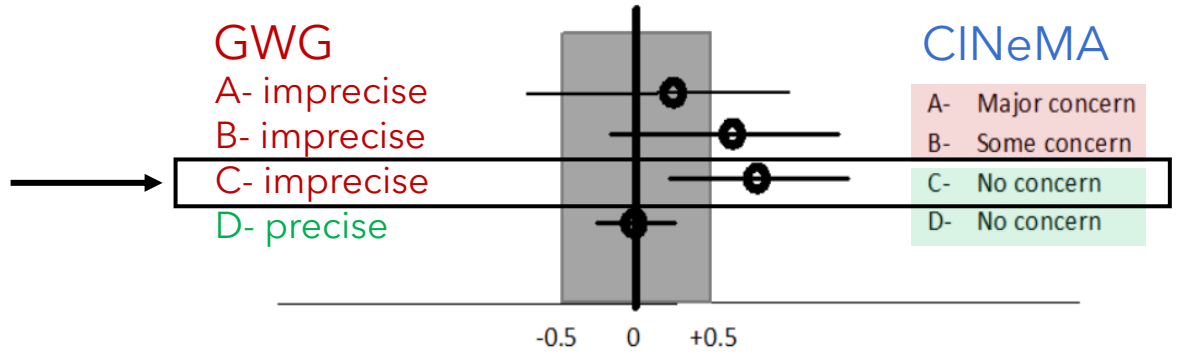
CINeMA framework

- These comparisons were not rated down using the CINeMA framework.

Sensitivity analysis

- Using the **same approach for adjudicating imprecision** for both the GWG and CINeMA approaches
- Incorporating the imprecision assessment approach used by CINeMA into the GWG approach to evaluate the robustness of our results

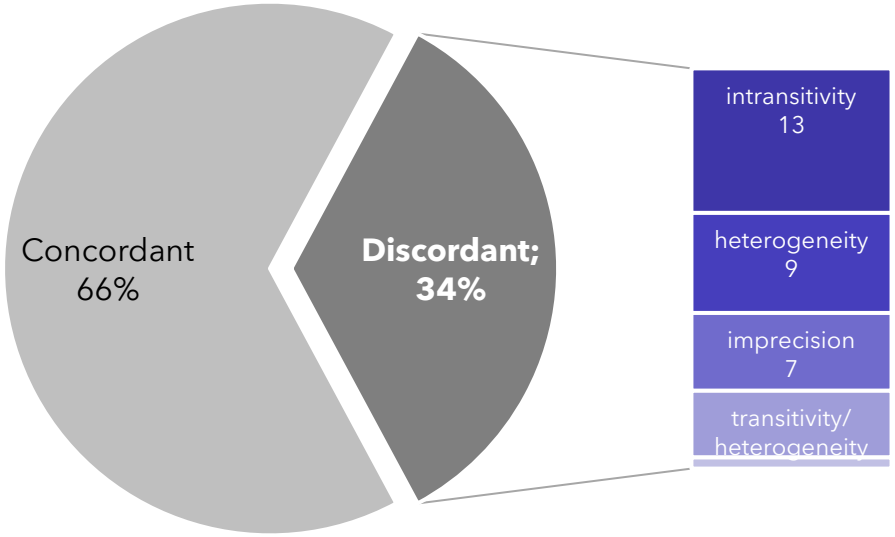
Reported no concerns with imprecision when the point estimate exceeded the threshold of half of the MID, and the CI included half of the MID but not the null effect



Results – sensitivity analysis (pain)

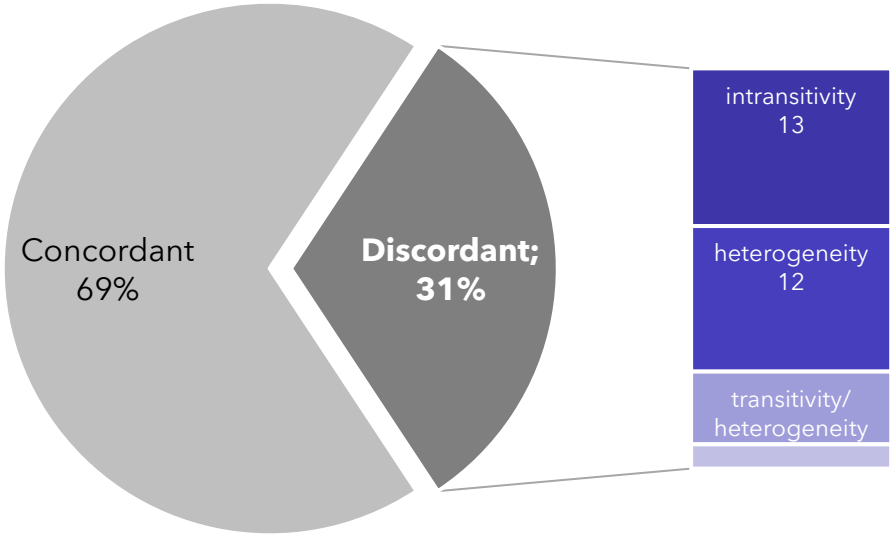
Main analysis

- GWG and CINeMA approaches provided different CoE ratings in 34% of cases (36 of 105)



Sensitivity analysis

- GWG and CINeMA approaches provided different CoE ratings in 31% of cases (33 of 105)



Results – sensitivity analysis (pain)

Main analysis

- GWG and CINeMA approaches provided different CoE ratings in 34% of cases (36 of 105)

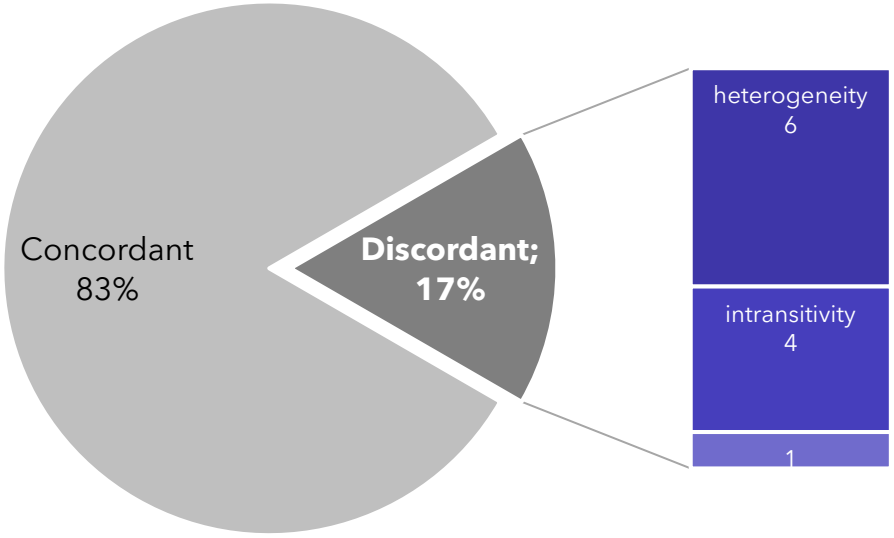
Sensitivity analysis

- GWG and CINeMA approaches provided different CoE ratings in 31% of cases (33 of 105)
  - (30/33) the GWG approach assigned a higher rating than CINeMA
  - (3/33) the GWG approach assigned a lower rating than CINeMA

Results – sensitivity analysis (physical functioning)

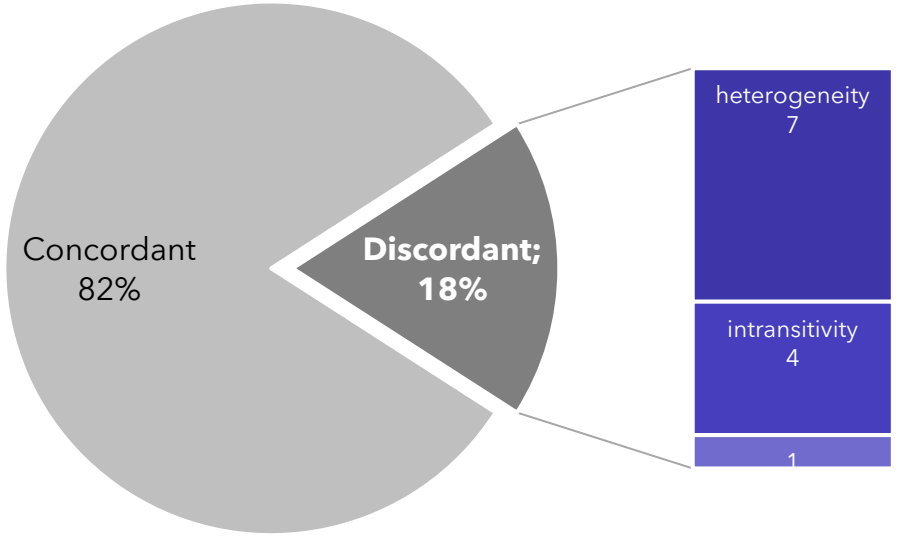
Main analysis

- GWG and CINeMA approaches provided different CoE ratings in 17% of cases (11 of 66)



Sensitivity analysis

- GWG and CINeMA approaches provided different CoE ratings in 18% of cases (12 of 66)



Results – sensitivity analysis (physical functioning)

Main analysis

- GWG and CINeMA approaches provided different CoE ratings in 17% of cases (11 of 66)

Sensitivity analysis

- GWG and CINeMA approaches provided different CoE ratings in 18% of cases (12 of 66)
  - (12/12) the GWG approach rate the CoE higher than CINeMA

Discussion

- After removing differences in the assessment of imprecision through our sensitivity analysis, there remained substantial discordance as a result of different approaches to the assessment of transitivity and heterogeneity.
- The starting level of confidence in the network estimate (before intransitivity, heterogeneity, and imprecision are assessed) was not a source of inconsistency between rating systems.
  - Likely because of very limited variability
  - However, GWG and CINeMA do use different approaches.

Limitations

- The findings are based on a single case study and the generalizability to other contexts is uncertain.
- Most of the evidence in NMA of opioids for chronic pain was rated as low or very-low certainty
  - This limited variability may have attenuated differences between the GWG and CINeMA approaches.
  - Despite this, discrepancies in CoE ratings were common

Conclusion

- Differences between CoE ratings provided by the GWG and CINeMA approaches when applied to network meta-analyses are common
- Disagreements were separated by one level of magnitude in CoE ratings
- Disagreements were predominantly due to considerations around intransitivity or heterogeneity.
- When discordant, CINeMA tended to provide lower CoE ratings than the GWG approach.
- For imprecision, the GWG approach rated down the CoE more often than CINeMA.
- Authors should be aware of differences in CINeMA and GWG approached when evaluating CoE and justify the selected approach.