

Central Lancashire Online Knowledge (CLOK)

Title	Network meta-analyses in IBD: pitfalls and promise for clinicians
Type	Article
URL	https://knowledge.lancashire.ac.uk/id/eprint/57688/
DOI	https://doi.org/10.1177/17562848251408758
Date	2026
Citation	Gordon, Morris (2026) Network meta-analyses in IBD: pitfalls and promise for clinicians. <i>Therapeutic Advances in Gastroenterology</i> .
Creators	Gordon, Morris

It is advisable to refer to the publisher's version if you intend to cite from the work.
<https://doi.org/10.1177/17562848251408758>

For information about Research at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clok.uclan.ac.uk/policies/>

Network meta-analyses in IBD: pitfalls and promise for clinicians

Morris Gordon 

Ther Adv Gastroenterol

2026, Vol. 19: 1–6

DOI: 10.1177/
17562848251408758© The Author(s), 2026.
Article reuse guidelines:
sagepub.com/journals-permissions

Abstract: Network meta-analysis (NMA) has seen an exponential rise in use over the past two decades, with publications in gastroenterology increasingly using this approach to inform guidelines, such as for colorectal cancer surveillance and inflammatory bowel disease management. However, alongside this growth comes a parallel risk: that methodological complexity and statistical misinterpretation may lead to misuse of the findings. As this paper outlines, NMAs can unintentionally mislead when clinical certainty is obscured by statistical hierarchy. To address these limitations, we present the promise of NMAs and a novel way of presenting NMA results, the GORDON Plot (Grade Of Results Diagram Of Network meta-analysis).

Plain language summary

Network meta-analyses IBD: pitfalls and flaws

- Network Meta-analysis allows many different treatments to be compared, even though they haven't been studied together in real life studies.
- This allows them to fill gaps in research.
- As a complex technique, Network meta-analysis is at risk of being performed poorly.
- This article highlights the risks of ignoring the quality or certainty of the findings and that this can lead to drawing incorrect conclusions.
- An interesting visual way to present these results is presented—the GORDON plot.

Keywords: evidence-based medicine, inflammatory bowel disease, meta-analysis, network meta-analysis

Received: 25 July 2025; revised manuscript accepted: 3 December 2025.

Introduction

In the era of evidence-based medicine, synthesizing high-quality data is paramount to shaping best practice. Traditional pairwise meta-analyses remain central to this goal, but their limitations become evident when clinicians must choose between several treatment options not directly compared in trials. Network meta-analysis (NMA) emerged to fill this gap, allowing the combination of direct and indirect evidence to compare multiple interventions within a unified analytic framework.^{1–3} Using similar statistical

techniques to meta-regressions, the factors of study are not patient or study characteristics, but the treatments used within randomized controlled trials.

NMA has seen an exponential rise in use over the past two decades, with publications in gastroenterology increasingly using this approach to inform guidelines, such as for colorectal cancer surveillance and inflammatory bowel disease (IBD) management. However, alongside this growth comes a parallel risk: that methodological

Correspondence to:
Morris Gordon
School of Medicine,
University of Central
Lancashire, Harrington
Building HA132, Preston
PR1 2HE, UK
mgordon@uclan.ac.uk



complexity and statistical misinterpretation may lead to misuse of the findings. As this paper outlines, NMAs can unintentionally mislead when clinical certainty is obscured by statistical hierarchy.

To address these limitations, we present the promise of NMAs and a novel way of presenting NMA results, the GORDON Plot (Grade Of Results Diagram Of Network meta-analysis).⁴

The promise and problem of NMA

NMA has transformed systematic reviewing by allowing simultaneous comparison of multiple treatments across a web of studies, even when head-to-head trials are missing. This facilitates broader inference and can increase precision by pooling more data.

However, this power introduces complexity. Interpreting indirect comparisons, navigating assumptions of transitivity and coherence, and accounting for multiple effect modifiers are demanding tasks. Even well-conducted NMAs risk being undermined if their findings are not conveyed clearly and accurately to clinicians and decision-makers. Furthermore, methodological errors—particularly in certainty assessment—are common, yet often overlooked.

Pitfall 1: Complexity obscures clinical interpretation

While statistical sophistication is a strength of NMA, it is also its Achilles heel. Clinicians accustomed to conventional forest plots or pairwise comparisons may struggle to understand league tables, rankograms, or network diagrams. These visualizations, though powerful, often lack intuitive clarity.

This opacity reduces the practical value of NMAs. If key stakeholders—such as guideline developers, clinicians, or patients—cannot readily interpret findings, the utility of the analysis is lost. Worse still, misunderstanding may lead to the adoption of interventions with weak evidence or limited applicability. This disconnect between statistical methods and clinical translation must be addressed.

Pitfall 2: Statistical rankings can mislead

The promise of treatment hierarchies is seductive. Ranking tables suggest simplicity—implying that the “best” treatment can be identified through probability estimates. However, this is rarely the case.

Rankings may reflect small, clinically irrelevant differences. For example, two therapies may have similar effectiveness, but a minor numerical difference may elevate one above the other in the hierarchy. If rankings are interpreted without considering effect size, precision, or relevance to patient priorities, the resulting conclusions can be misleading or even harmful.

Moreover, rankings do not account for the quality of underlying evidence. A top-ranked intervention based on low-certainty data should not be considered superior. Without robust contextualization, numerical rankings can foster false confidence. A recent example is a thorough and complete NMA in inducing remission in Crohn’s disease,⁵ but ranks therapies through their SUCRA and talks in superlatives of the “highest estimate.” There are treatments that, in other reviews^{4,6} are noted to be less relevant due to other method issues, but with a focus on ranking alone, this potential can leave readers with an incomplete message.

Pitfall 3: Certainty of evidence is often overlooked

GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) offers a structured, transparent method for assessing the certainty of evidence across multiple domains—including risk of bias, inconsistency, indirectness, imprecision, and publication bias. It is widely accepted as the gold standard in evidence synthesis.

Yet, GRADE is applied inconsistently in NMAs. In a systematic survey, less than one-third of NMAs included any form of certainty rating, and fewer than one in five applied GRADE.⁷ This omission is not trivial. Without a clear sense of how certain we are in the reported effects, statistical results become clinically hollow.

Moreover, applying GRADE to NMA introduces further complexity, as indirect and network estimates must also be graded—requiring methodological transparency and expertise that many teams lack. Nevertheless, failure to include GRADE devalues the output and weakens its utility for decision-making.

Using our recent examples, it can be seen that using GRADE⁶ has a major impact when compared to not using GRADE.⁵ This lessens the focus on ranking alone and instead only considers the network estimates in the context of how certain they are. This links to the previous point, not removing ranking, but contextualizing in a manner that is core to correct and evidence-based interpretation.

Pitfall 4: The misuse of GRADE—Neglecting outcome balance and decision-making context

When interpreting the findings of an NMA, it is tempting to treat certainty of evidence (as assessed by GRADE) as a series of distinct, discrete judgments tied to each individual outcome. However, GRADE was not only developed to rate certainty in isolation—it was also designed to support clinical decision-making. In fact, the strength of a GRADE-based recommendation relies on balancing multiple outcomes, not just rating them separately.

In practice, any treatment will typically have multiple relevant outcomes. It is not unusual for an NMA to include four or more primary outcomes—often including one or more safety endpoints. Focusing exclusively on the “most favorable” result risks privileging outcomes that are statistically significant but weak in certainty or clinical importance. Conversely, outcomes that are neutral or equivocal may have higher certainty and provide more clinically actionable insights, such as ruling out the value of a treatment.

This is an underappreciated and often overlooked contribution to the evidence base. Unfortunately, reviewers and readers may fail to integrate outcomes into a cohesive synthesis. This synthesis is often left to the guideline developer or practicing clinician. Rarely, it is presented clearly by the review authors themselves. This disconnect between synthesis and interpretation is a critical methodological blind spot and represents one of

the most common and problematic pitfalls in NMA reporting.

Pitfall 5: Effect size and clinical meaning get lost

NMAs frequently focus on statistical ranking, which privileges relative statistical differences between treatments. Yet, a top-ranked therapy may only offer a marginal benefit—for example, 2% improvement over placebo—raising important questions about whether such a difference is clinically meaningful.

Clinical magnitude of effect is not just a footnote to statistical output; it should be central to interpretation. This issue was addressed directly in the British Society of Gastroenterology Guidelines for IBD.⁸ In that initiative, NMAs were used alongside pairwise comparisons, but not before a formal thresholding exercise was conducted. Over 100 expert members of the guideline development group, along with international collaborators, defined the minimum clinically important differences required for outcomes to be considered meaningful.⁹

This process allowed the guideline authors to assess precision and certainty not in abstract statistical terms but in relation to real clinical practice. For example, a wide confidence interval spanning from a large to a trivial effect would be downgraded for imprecision, while an equally wide interval from “high to high” would be interpreted more favorably. More importantly, the magnitude of effect was then aligned with GRADE certainty, allowing the team to report not just “high certainty” but “high certainty of a given magnitude of effect.”

This nuanced approach exemplifies the level of interpretation required to make NMAs truly decision-supportive. Without it, rankings remain superficial and easily misused.

Principles for responsible NMA use

To support more effective and trustworthy interpretation of NMAs in IBD and other specialties, we recommend five core principles (summarized in Figure 1):

Transparent reporting—Full disclosure of assumptions, methods, and network geometry

Network Meta-analyses in IBD: Pitfalls and Promise

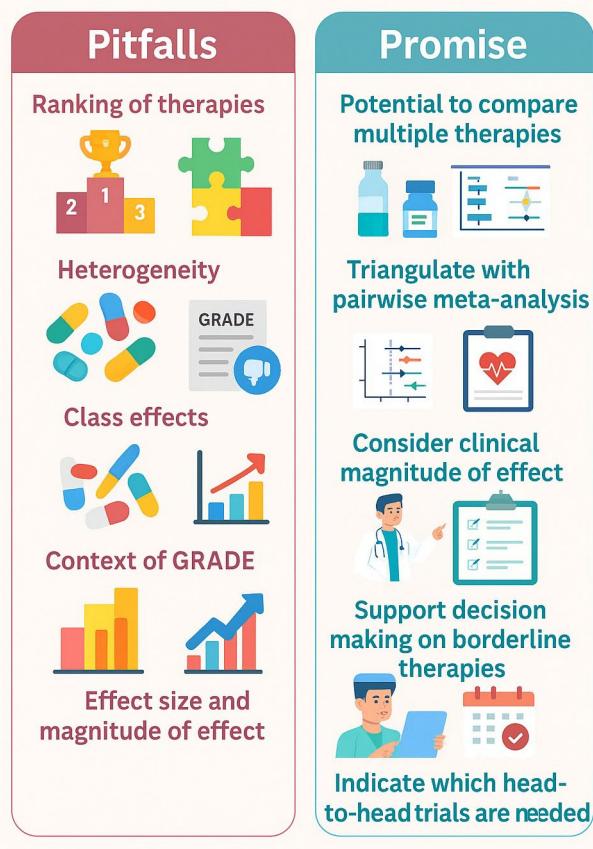


Figure 1. Summary of pitfalls and promise of NMA.
NMA, network meta-analysis.

should be standard. Readers must be able to trace the reasoning behind indirect comparisons and understand the scope of included evidence. The reader may not fully understand how or why such judgments were made, but it should be clear and have face validity to the reader.

Contextualized rankings—Hierarchies in NMA should never be presented without effect size, confidence intervals, and an explanation of clinical relevance. Rankings must inform, not dictate, decision-making, and the presentation of them in isolation is reductionist at best and misleading to the point of dangerous at worst.

GRADE integration—Certainty must be evaluated rigorously for all estimates—direct, indirect, and network-derived. GRADE should not be optional or tokenistic, but central to

NMA interpretation. This cannot be presented as a distinct element or add-on, but as central to the results as the confidence intervals.

Outcome balance—Efficacy variables must be clear, decided a priori, clinically relevant, and all be weighed alongside safety, tolerability, and patient values. No single outcome should dominate the analytic landscape in isolation. NMA should have clear protocols or plans to clarify the primary or critical outcomes and the rationale for these. Then, the results for these should all be presented to allow the reader to consider the balance among efficacy outcomes and, more importantly, balance this against safety risks.

Clear visualization—Many NMAs give ranking tables, direct and indirect data, network plots, and similar. However, these are not intuitive or clear to the reader, and they do not naturally

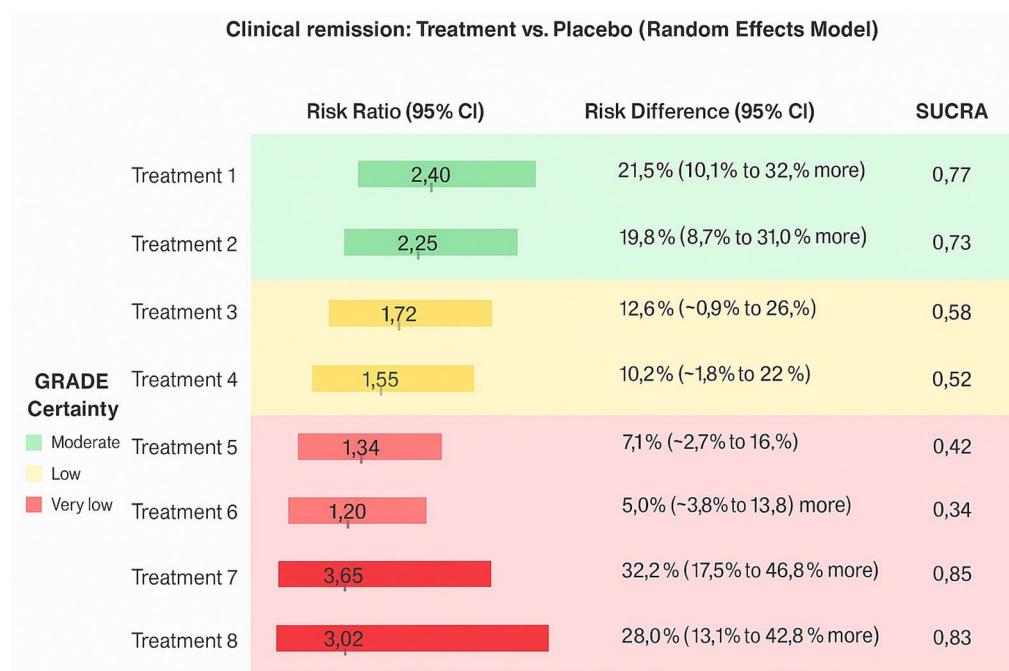


Figure 2. Dummy/example GORDON plot.
GORDON, Grade of Results Diagram of Network meta-analysis.

address the principles above. There is a need to use innovative tools to provide clarity by integrating ranking and certainty, enabling clinicians and stakeholders to allow understand both the strengths and limitations of the findings.

To bridge the gap between complexity and clinical utility, we propose the use of the GORDON Plot—an innovative visualization method that aligns two critical dimensions: Ranking Probability (derived from the statistical model of the NMA) and Certainty of Evidence (evaluated using GRADE for each treatment estimate). This technique has been used in several recent IBD publications and offers great potential to address many of the pitfalls^{4,6,10} which show GORDON plots. However, included in Figure 2 is an example which demonstrates the clear focus on GRADE and how ranking must be considered in the context of GRADE, with some higher-ranked treatments being very uncertain. There is currently no alternative to combine ranking, absolute effects, and GRADE judgments in a single visual output, and so this is currently a unique approach.

Conclusion

NMA offers a powerful means of comparing complex interventions across IBD treatment landscapes. However, when misapplied or misunderstood, it can mislead clinicians and erode trust in evidence-based practice.

By embedding certainty into ranking and grounding findings in clear, balanced visual communication, the NMAs fulfill their intended role: not to confuse with complexity, but to clarify with confidence.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Author contributions

Morris Gordon: Conceptualization; Formal analysis; Investigation; Writing – original draft; Writing – review & editing.

Acknowledgements

None.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Competing interests

The author declares that there is no conflict of interest.

AI assistance

This manuscript was drafted by the named author, using content and concepts developed solely by the author. The language and structure were refined using ChatGPT (OpenAI), which provided editorial support only. No original content, references, or novel ideas were generated by the tool, and no external sources were accessed. All material reflects the author's original work and intellectual contribution.

Availability of data and materials

No original data, so not applicable.

ORCID iD

Morris Gordon  <https://orcid.org/0000-0002-1216-5158>

References

1. Caldwell DM, Dias S and Welton NJ. Extending treatment networks in health technology assessment: how far should we go? *Value Health* 2015; 18: 673–681.
2. Cooper NJ, Peters J, Lai MCW, et al. How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? *Value Health* 2011; 14(2): 371–380.
3. Chaimani A, Caldwell DM, Li T, et al. Chapter 11: Undertaking network meta-analyses (last updated October 2019). In: Higgins JPT, Thomas J, Chandler J, et al. (eds) *Cochrane handbook for systematic reviews of interventions* version 6.5. London: Cochrane, 2024.
4. Gordon M, Sinipoulou V and Moran G. Immunomodulators and advanced therapies for induction of remission in Crohn's disease: a systematic review and network meta-analysis. *Inflamm Bowel Dis*. Epub a head of print September 2025. DOI: 10.1093/ibd/izaf191.
5. Versteegh M, Idema D, Huygens S, et al. Comparative efficacy of all available pharmaceutical therapies for moderate to severe Crohn's disease: a systematic review and network meta-analysis. *Gastro Hep Adv* 2025; 5: 100764.
6. Sinopoulou V, Nigam GB, Gordon M, et al.; British Society of Gastroenterology Colorectal IBD Surveillance Guideline Development Group. Comparative efficacy and safety of endoscopic modalities for colorectal cancer screening in inflammatory bowel disease: a systematic review and network meta-analysis. *Clin Gastroenterol Hepatol* 2025; 23(12): 2128–2143.
7. Lovsletten PO, Wang X, Pitre T, et al. A systematic survey of 200 systematic reviews with network meta-analysis (published 2020–2021) reveals that few reviews report structured evidence summaries. *J Clin Epidemiol* 2024; 173: 111445.
8. Moran GW, Gordon M, Sinopoulou V, et al. British Society of Gastroenterology guidelines on inflammatory bowel disease in adults: 2025. *Gut* 2025; 74(Suppl. 2): s1–s101.
9. Gordon M, Shaban N, Sinopoulou V, et al. Developing IBD outcome effect size thresholds to inform research, guidelines, and clinical decisions. *Inflamm Bowel Dis* 2025; 31(10): 2798–2804.
10. Ananthakrishnan AN, Murad MH, Scott FI, et al. Comparative efficacy of advanced therapies for management of moderate-to-severe ulcerative colitis: 2024 American Gastroenterological Association Evidence Synthesis. *Gastroenterology* 2024; 167(7): 1460–1482.