Journal of Clinical Epidemiology 194 (2026) 112226

GRADE RELATED PAPER

An exploratory descriptive survey on the use of GRADE and CINeMA: time-consuming, process transparency and subjectivity vs high-speed, practical challenges, and poor understanding

Pier Carlo Battain[a,*], Giulia Lamiani[b,c], Cinzia Del Giovane[d,e,1], Silvia Minozzi[f,1]

[a]*Smart Homes, Assistive Technologies, and Occupational Therapy (DAT) Research Unit, IRCCS "S. Maria Nascente" - Fondazione Don Carlo Gnocchi, Via Capecelatro 66, Milan 20148, Italy*

[b]*Department of Health Science, University of Milan, Via Antonio Rudinì 8, Milan 20142, Italy*

[c]*Unit of Clinical Psychology, ASST Santi Paolo e Carlo, Via Antonio Rudinì 8, Milan 20142, Italy*

[d]*Department of Medical and Surgical Sciences for Children and Adults, University-Hospital of Modena and Reggio Emilia, via del Pozzo 71, Modena 41124, Italy*

[e]*Institute of Primary Health Care (BIHAM), University of Bern, Mittelstrasse 43, Bern 3012, Switzerland*

[f]*Laboratory of Methodology of Systematic Reviews and Guidelines production, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via Mario Negri 2, Milan 20156, Italy*

Accepted 10 March 2026; Published online 18 March 2026

Abstract

Objectives: To explore assessors' experience regarding the use of two approaches that evaluate the certainty of evidence from Network Meta-Analysis (NMA): Grading of Recommendations Assessment, Development and Evaluation (GRADE) and Confidence In Network Meta-Analysis (CINeMA).

Study Design and Setting: Thirteen assessors trained in NMA methodology and evaluation methods were randomly assigned to four groups to apply GRADE and CINeMA to two networks each to compare the approaches' concordance, inter-rater reliability, and application time (Minozzi et al, 2025). Assessors' experiences were collected through an online questionnaire including open-ended questions. The questionnaire explored usability, strengths, weaknesses, and perceived accuracy of each approach. Each assessor was surveyed twice: once after GRADE use and once after CINeMA use. Thematic analysis was used to analyze their responses.

Results: We identified seven themes: instruments' challenges, knowledge need, process execution, positive aspects, interpretative uncertainty, implementation, and level of confidence. GRADE assessment was considered long because of the extensive material to manage and at high risk of subjectivity but with a transparent and clear process. CINeMA automatization allowed a quick and easy use but impacted negatively on the process methodological understanding. Technical challenges related to website and files management were identified. Overall, assessors highlighted the need for further training on tool usage.

Conclusion: GRADE was considered time-consuming and at high risk of subjectivity but excelled in transparency and clarity. CINeMA was considered high-speed but showed some practical and output-interpretation challenges, particularly where specialized knowledge of NMA methods was required. Additional training courses were suggested for both tools.

Keywords: Network meta-analysis; GRADE; CINeMA; Survey; Expertise; Methodology

1. Introduction

Network Meta-Analysis (NMA) has gained prominence as a valuable tool for simultaneously evaluating the comparative effectiveness and safety of multiple interventions by combining direct and indirect evidence [1]. Currently, two approaches are available to assess the certainty of NMA results: the Grading of Recommendations Assessment, Development and Evaluation (GRADE) and

[1] Silvia Minozzi and Cinzia Del Giovane share the last authorship, as they equally contributed to the supervision and overall direction of the study.

* Corresponding author. IRCCS "S. Maria Nascente", Fondazione Don Carlo Gnocchi, Via Capecelatro 66, Milan 20148, Italy.

E-mail address: piercarlo.battain@gmail.com (P.C. Battain).

https://doi.org/10.1016/j.jclinepi.2026.112226

Plain Language Summary

Researchers use NMA to compare multiple medical treatments. To rate how trustworthy the results are, they can choose between two tools: Grading of Recommendations Assessment, Development and Evaluation (GRADE) and Confidence In Network Meta-Analysis (CINeMA). We asked thirteen researchers who had used both tools to describe their experiences in a survey. Their feedback highlighted a key trade-off. GRADE was praised for being transparent and educational, as its manual process helped assessors understanding each step. However, it was consistently described as very time-consuming and somewhat subjective. CINeMA was appreciated for its speed and ease of use, thanks to its automated, web-based system. Yet, this automation made the process feel like a ‘‘black box,’’ reducing assessors’ understanding on how conclusions were reached. Practical issues, like file upload problems, were also a hurdle. A common theme was that both tools require solid training and background knowledge to be used correctly. In short, GRADE offers depth and clarity at the cost of time, whereas CINeMA offers efficiency at the cost of transparency. These insights can help guide future training and improve these important research tools.

the Confidence in Network Meta-Analysis (CINeMA). The GRADE approach was developed by the GRADE Working Group [2], and the CINeMA approach was inspired by the GRADE methodology [3,4]. Despite different terminological nuances, both approaches share a common foundation and consider similar domains that can impact the certainty of evidence. Both approaches adopt the same final levels of certainty of evidence, but their operationalization differs. Within GRADE direct estimates are evaluated first and, only under certain conditions, indirect evidence from the first-order loop are also rated, whereas, within CINeMA, direct and indirect evidence are considered simultaneously. GRADE requires a step-by-step manual judgment, whereas CINeMA provides semiautomated evaluation via a dedicated web application.

A comparison of the two approaches in terms of concordance of results, inter-rater agreement, and application time of the instruments was reported in a recent publication [5]. The tools were applied by thirteen assessors to four networks with different structures and sizes. The authors found that the concordance of results between the two methods was good for small networks with no or few indirect comparisons and decreased as the number of comparisons and indirect evidence increased. Furthermore, the inter-rater agreement was moderate to substantial with both approaches (range from 0.49 with GRADE to 0.73 with CINeMA), except for CINeMA when applied to a large and complex network showing a slight agreement. Finally, the application time was long with both methods, especially with GRADE for large networks.

Despite this quantitative analysis, the assessors’ experiences with both tools remain unexplored. Gathering assessors’ feedback is fundamental, as it may contribute to increasing the usability of the instruments, identifying technical or practical challenges, and reducing possible interpretative doubts. As an extension of the study from Minozzi et al [5], we conducted this exploratory descriptive study that aimed to explore assessors’ experience regarding the usability, strengths, weaknesses, and perceived accuracy of the two approaches.

2. Materials and methods

2.1. Methodology

In Minozzi et al [5], thirteen assessors were randomly assigned to four groups. The assessors were selected from those who participated in a course focused on NMA methodology, which took place in 2023 at the University of Milan. We intended to recruit participants who were interested in the topic and had a good knowledge of the two approaches. We used a computer-generated randomization to assign assessors to each group. Three groups were formed of three assessors and one group of four. Two groups first applied GRADE and then CINeMA to two different networks, whereas the other two groups first applied CINeMA and then GRADE to two other networks. The networks were selected from four NMAs of randomized controlled trials [6−9]. All assessors worked independently.

We developed a questionnaire to investigate the experiences of those assessors (Supplementary S1). The questionnaire was composed of two parts. The first part included closed-ended questions to collect data on assessors’ clinical experience, methodological knowledge, and previous experience using the instrument. The second part consisted of open-ended questions to collect qualitative feedback on technical and practical challenges encountered when using the instrument, challenges in understanding theoretical concepts, and the need for additional skills to correctly apply the instrument. Assessors were also asked about the positive and negative aspects of using the instrument, as well as their opinion on its usability by clinicians involved in research processes. The questionnaire ended with a self-assessment of the assessors’ validity and accuracy. The questionnaire was administered online via Microsoft Forms (Microsoft Corporation, Redmond, USA).

Participants were informed of the study objectives and requirements. An informed consent was obtained and participation was anonymous.

The timing of questionnaire administration followed the timeline of the study conducted by Minozzi et al [5]. The

What is new?

Key findings

- Grading of Recommendations Assessment, Development and Evaluation process is transparent but time-consuming
- Confidence In Network Meta-Analysis is fast but limits insight due to automated processes

What this adds to what is known?

- Highlights assessors' real-world experiences with both tools
- Reveals training needs and usability issues not previously explored

What is the implications and what should change now?

- More training resources are needed to support a proper use of the tools

assessors completed the questionnaire twice: once at the end of the first assessment phase using one instrument and once at the end of the second assessment phase using the other instrument, approximately 1 month after the first completion. Each assessor reported their feedback for both GRADE and CINeMA.

2.2. Analysis

We conducted a thematic analysis of open-ended responses to identify recurring themes within data. Data analysis followed the six-phase method for reflexive thematic analysis as outlined by Braun and Clarke [10], an iterative process consisting of the following phases: (1) familiarization with the data through repeated reading of the transcripts; (2) generating initial codes: each researcher independently identified and labeled meaningful features in the data; these initial categories were subsequently grouped into broader, more inclusive candidate themes; (3) searching for themes, where codes were collaboratively collated into broader patterns of meaning (candidate themes); (4) reviewing themes, ensuring they worked in relation to both the coded extracts and the full dataset; (5) defining and naming themes to capture the essence and scope of each; and (6) producing the report, selecting vivid extracts and crafting the final analytical narrative.

The reporting for this study was guided by the Standards for Reporting Qualitative Research checklist [11], the completed version of which is provided in the Supplementary S2. Microsoft Excel 16.0 (Microsoft Corporation, Redmond, USA) was used to facilitate data management and coding.

3. Results

3.1. Participants and settings

Thirteen questionnaires were collected. Around 70% of participants were clinicians. They had slightly above average experience in clinical and methodological fields. Six of them (46%) had previous experience with GRADE, whereas only two (15%) with CINeMA (Table 1).

3.2. Qualitative analysis

Seven themes emerged: (1) difficulties related to the instrument, (2) need for knowledge, (3) process execution, (4) positive aspects, (5) uncertainty of interpretation, (6) implementation, and (7) level of confidence.

The results are presented by themes, including the most relevant excerpts from the questionnaires (Table 2). A detailed articulation of themes, categories (24 overall), subcategories, citation frequencies, and all quotes is available in the Supplementary S3.

3.3. Theme 1: difficulties related to the instrument

For CINeMA, assessors reported general technical and informatic challenges related to website access or file upload. "I experienced some difficulties accessing the CINeMA website (offline for a few days) and the risk of bias due to missing evidence in network meta-analysis (ROB-MEN) subsection (online but not accessible due to server errors). Once I had access to both, there were no problems." "I experienced initial practical difficulties in uploading the file, in .csv format, to the CINeMA application."

For GRADE, assessors highlighted challenges in collecting and organizing a large amount of information that should be downloaded from a statistical package (eg, pairwise meta-analyses, contribution matrix, test of heterogeneity, network estimates, and incoherence results). "The most

Table 1. Assessors' characteristics

Characteristics	Assessors (n = 13)
Time to complete the first interview, mean in minutes (range)	12 (1–58)
Time to complete the second interview, mean in minutes (range)	8 (2–23)
Clinicians, n (%)	9 (69)
Experience as a clinician[a], mean (SD)	7.6 (1.9)
Experience in methodological field[a], mean (SD)	6.3 (2)
Experience in the tool used	
CINeMA, n (%)	2 (15)
GRADE, n (%)	6 (46)

GRADE, Grading of Recommendations Assessment, Development and Evaluation; CINeMA, Confidence in Network Meta-Analysis.

[a] Likert scale from 1 (no experience) to 10 (a lot of experience).

Table 2. Themes, categories, citation frequency for each tool and the most representative quotes

Themes	Categories	Response frequency CINeMA vs GRADE	Quotes
Difficulties related to the instrument	Technical difficulties	18:3	*''Initially when uploading the .csv files due to problems with the extension (I had to manually reformat the file into comma-separated values because I could not save the one received so that it would be accepted by the software).'' [CINeMA]* *''Initially when uploading .csv files due to problems with the extension (I had to manually reformat the file into comma-separated values because I could not save the one, I received so that it would be accepted by the software).'' [CINeMA]*
	Practical difficulties	*5:9*	*''The most challenging phase, in my opinion, concerns the organization of the working material. In fact, the main criticality of the tool lies in the correct organization of the material to allow quick access to the required information.'' [GRADE].*
	Theoretical difficulties	*13:15*	*''The biggest conceptual difficulty concerns the ROB-MEN tool, which, having never used and being unfamiliar with the concepts, I found difficult to integrate into the rest of the CINeMA procedure.'' [CINeMA].* *''The excel file is not entirely intuitive especially in the part of the explanation of what to put in the different cells to get the OIS and how to calculate/get that information.'' [GRADE]*
Needed knowledge	Methodological knowledge	*10:24*	*''Personally, I would need more methodological knowledge.'' [GRADE].* *''Having more methodological and statistical knowledge could speed up the evaluation timelines.'' [GRADE]* *''I think that, in general, it is necessary to have a high level of expertise in methodology and statistics in order to fully understand each step and to apply any corrections in evaluations.'' [CINeMA]*
	Statistical knowledge	*4:5*	*''Having more methodological and statistical knowledge could speed up the evaluation time.'' [GRADE]* *''Yes, but pure clinicians need to have a solid methodological and statistical basis on NMA.''[CINeMA].*
	Clinical knowledge	*1:3*	*''[…] I think a solid clinical knowledge of the topic is indispensable […].'' [GRADE]*
	Knowledge of the instrument	*5:3*	*''You need a very clear and thorough knowledge of all the concepts underlying the individual domains and the rules implemented by the software to evaluate them.'' [CINeMA]* *''Being unfamiliar with the evaluation of certainty of evidence, the difficulties I encountered generally concerned the conceptual steps from one step to the next with which I had to become familiar.'' [GRADE]*
	Preliminary training needed	*3:1*	*''It is definitely necessary to attend a further training course or at least a specific in-depth study of the tool, spending time on its operation.'' [CINeMA]* *''The concepts of inconsistency and heterogeneity are not always immediately understandable, specific training is needed.'' [GRADE]*
Process execution	Process automatization	*10:0*	*''[…] the automatism associated with assessments, if not properly managed and monitored by the assessor, can be a limitation.''[CINeMA]*
	Length process and iteration	*0:14*	*''The difficulty of figuring out which steps to go through to make the work smoother and simpler.'' [GRADE]* *''The application of the tool is conceptually complex and extremely 'time-consuming' at least in the initial stages.'' [GRADE]*
Positive aspects	Speed of use	*6:0*	*''Speed of the instrument (if familiar with its use).'' [CINeMA]*

(*Continued*)

Table 2. Continued

Themes	Categories	Response frequency CINeMA vs GRADE	Quotes
	Usability	*13:5*	*''The tool also allows for use by those who do not have in-depth knowledge and skills in the methodological, statistical or clinical fields, enabling greater usability even among those without much experience.'' [CINeMA]* *''Grading of evidence in a fairly useable manner by clinical colleagues not overly familiar with EBP principles […].'' [GRADE]*
	Reproducibility	*2:2*	*''[…] reproducibility.'' [GRADE]* *''The automatism associated with certain evaluations that facilitates the standardization of the procedure.'' [CINeMA]*
	Evaluation transparency	*3:5*	*''It also allows for quality scaling by excluding the most problematic studies allowing for an assessment of whether or not this changes the final evaluations […].'' [CINeMA]* *''[…] the various steps of the tool can help you actually understand what you are doing, thus allowing you to correct procedural errors as you go along.'' [GRADE]*
	Intuitive user interface	*10:0*	*''The very simple and intuitive graphic interface of the application, from the initial screen of which all the instructions for the correct compilation and nomenclature of files, demonstration datasets and links for access to in-depth methodological documentation can be found; the possibility of automatic compilation of fields, and a wide availability of graphs and tables, extremely useful in the preparation of the final work.'' [CINeMA]*
Interpretative uncertainty	Length of process	*0:4*	*''It is a useable tool, but it takes time to do the evaluation.'' [GRADE]*
	Subjectivity of evaluation	*1:3*	*''In some respects, it is more time-consuming and complicated to apply and may leave more freedom to assess things more 'subjectively'.'' [GRADE]* *''[…] there always remains an element of subjectivity in the evaluation.'' [GRADE]*
	Interpretability of results	*2:1*	*''When the automatization of the results takes place, there are interpretative doubts, especially for those who do not have an established experience in this area.'' [CINeMA]* *''I also think that they might have difficulties in interpreting the results.'' [GRADE]*
Implementation	Automatization of steps	*0:4*	*''Automatizing the assessment of inconsistency and heterogeneity.'' [GRADE]*
	User interface	*5:0*	*''I would probably include a more in-depth tutorial in the navigation interface.'' [CINeMA]*
	GRADE software	*0:3*	*''The possibility of having dedicated software.'' [GRADE]*
	OIS excel file configuration	*0:1*	*''If it were possible I would improve the excel for calculating the OIS.'' [GRADE]*
Level of confidence	Good confidence	*5:5*	*''[…] Regarding the ability of the indicated procedure to arrive at reliable and accurate results, there is no uncertainty.''[GRADE]* *''[…]I have tried to be as accurate as possible, I have carefully reread the methodological papers provided several times, I have listened again and again to the teaching material provided during the advanced course, I have carefully followed the instructions provided for the formulation of qualitative judgments and I have preliminarily reproduced the results of the examples shown during the course in order to have an indirect confirmation of the accuracy of the logical reasoning followed.'' [CINeMA]*

(*Continued*)

Table 2. Continued

Themes	Categories	Response frequency CINeMA vs GRADE	Quotes
	Poor confidence	*4:12*	*''The concept of 'validity' and 'accuracy' must however always be contextualized with respect to the use of a tool that has a strong subjective and qualitative dimension in reaching judgments […].''[GRADE]* *''I think they might come close to being valid and/or accurate because they seem to me to be consistent, but some aspects remain less 'objective' than others.'' [GRADE]*

GRADE, Grading of Recommendations Assessment, Development and Evaluation; CINeMA, Confidence in Network Meta-Analysis; NMA, Network Meta-Analysis; ROB-MEN, risk of bias due to missing evidence in network meta-analysis; CSV, comma-separated values; OIS, optimal information size.

This table shows the themes, relative categories, citation frequency for each tool and the quotes that we considered representative of each category. Italics indicate verbatim participant quotes.

challenging phase […] concerns the organization of the working material. In fact, the main criticality of the tool lies in the correct organization of the material to allow quick access to the required information.''

On the theoretical side, assessors encountered difficulties related to the assessment of the reporting bias with the ROB-MEN tool within CINeMA and of imprecision with GRADE. They distinctly found challenging the integration of the results obtained with the ROB-MEN tool into the CINeMA procedure and understanding the fundamentals for making the reporting bias assessment. Regarding GRADE, the evaluation of imprecision using an absolute measure anchored to a predefined Minimally Important Difference was considered conceptually complex and methodologically challenging.

3.4. Theme 2: need for knowledge

For both instruments, the assessors highlighted their need for more in-depth knowledge regarding methodology, statistics, clinical aspects, and how to use the instrument: ''a deep knowledge of many theoretical contents is needed, which is not always possessed by clinicians active in research [GRADE].'' Although the CINeMA tool is easy to use, ''there must be a minimum knowledge of basic methodological concepts.'' Furthermore, ''having methodological and statistical knowledge could fasten the evaluation time [GRADE].'' ''Especially in case of very complex networks, I think that […] methodological and statistical skills are required […] [CINeMA].''

Regarding the necessity of solid clinical knowledge, participants stated: ''especially in the case of very complex networks, solid clinical knowledge of the subject matter is necessary [CINeMA].'' ''[…] I think a solid clinical knowledge of the topic is indispensable […] [GRADE].''

The instrument knowledge category reflects the need to understand how the tools work. ''Since I am not familiar with the evaluation of certainty of evidence, the difficulties I encountered concerned the conceptual steps, which I had to become familiar with [GRADE].'' This points to a need for conceptual orientation. For CINeMA, the emphasis is on mastering both the concepts and the software logic: ''you need to have a very clear and thorough knowledge of all the concepts underlying the individual domains and the rules implemented by the software to evaluate them.''

Sometimes they reported the need for prior training: ''it is definitely necessary to attend a further training course or at least a specific in-depth study of the tool, spending time on its operation/usage [CINeMA]'' ''The concepts of inconsistency and heterogeneity are not always immediately understandable; specific training is needed [GRADE].''

3.5. Theme 3: process execution

For GRADE, the ''process automatization'' was not an issue but rather the ''process length and iteration'': ''the length and repetitiveness of the procedure, which easily exposes one to errors, especially in the evaluation of the network plot […].'' ''The application of the tool is conceptually complex and extremely 'time-consuming,' at least in the initial stages.''

Also, assessors found a ''greater understanding of methodological steps'' compared to CINeMA: ''the difficulty of figuring out which steps to go through to make the work smoother and simpler [GRADE].'' However, once understood, ''the 'manual' procedure is intuitive and easy to follow, providing a hands-on reading of the quality of the NMA results [GRADE].''

The category ''process automatization'' and the subcategory ''less understanding of methodological steps'' refer exclusively to CINeMA. Assessors found that the automatization process makes the rationale and methodological procedure behind it difficult to be fully understood. Although the theoretical concepts do not seem complicated, ''it is sometimes very difficult to understand some of the results obtained with CINEMA because they are the result of an elaboration of the system itself.'' It is challenging to determine whether the processed results are correct and plausible because the system processes the data automatically.

3.6. Theme 4: positive aspects

GRADE was positively valued for its transparency and support for critical reflection: ''the various steps of the

tool can help you actually understand what you are doing, thus allowing you to correct procedural errors as you go along.''

CINeMA was appreciated for its usability by people with little experience of the methodology and for its speed: ''the tool also allows for use by those who do not have in-depth knowledge and skills in the methodological, statistical, or clinical fields, enabling greater usability even among those without much experience.'' ''Speed of the instrument (if familiar with its use).'' Its automatization was also seen as a facilitator of reproducibility: ''the automatism associated with certain evaluations that facilitates the standardization of the procedure.''

3.7. Theme 5: uncertainty of interpretation

For GRADE, assessors noted a degree of subjectivity: ''there always remains an element of subjectivity in the evaluation.''

For CINeMA, uncertainty was linked to the automatization in producing the assessments, which can hinder interpretability: ''when the automatization of the results takes place, there are interpretative doubts, especially for those who do not have an established experience in this area.''

3.8. Theme 6: implementation

For GRADE, the need for a dedicated software and automatization of specific steps was highlighted: ''automatizing the assessment of inconsistency and heterogeneity.'' ''The possibility of having a dedicated software.''

For CINeMA, suggestions for improvement included a better guidance within the web-interface: ''I would include a more in-depth tutorial in the interface.''

3.9. Theme 7: level of confidence

For GRADE, some assessors expressed strong trust in the method: ''regarding the ability of the indicated procedure to arrive at reliable and accurate results, there is no uncertainty.'' However, others emphasized the subjective nature of the judgments: ''the concept of 'validity' and 'accuracy' must however always be contextualized with respect to the use of a tool that has a strong subjective and qualitative dimension in reaching judgments […].''

In CINeMA, the level of confidence was often linked to the effort invested in understanding the methodology: ''I have carefully followed the instructions provided for the formulation of qualitative judgments and I have preliminarily reproduced the results of the examples shown during the course in order to have an indirect confirmation of the accuracy of the logical reasoning followed.''

4. Discussion

This exploratory survey collected the experiences of thirteen assessors in using both the GRADE and CINeMA tools in evaluating the certainty of NMA results.

Based on the assessors' experience, both GRADE and CINeMA have shown strengths and challenges. For GRADE, the main positive aspect that emerged was related to the way the assessors had to proceed for the evaluation of the confidence. Judging each domain ''manually'' gave them the opportunity to understand what they were doing and therefore to have a strong trust in the method, although at high risk of subjectivity. However, such manual evaluation involved considering and managing a large volume of material, which caused practical and organizational issues and made the evaluation process very time-consuming, especially for large and complex networks. The domain that was particularly challenging was imprecision. Although a lot of materials was provided to the assessors for its assessment [12], its methodological understanding and how it was supposed to be correctly evaluated was considered very challenging. This was one of the areas for which a further deep knowledge of some theoretical content was considered needed. Compared to CINeMA, GRADE offered researchers a greater level of awareness and understanding throughout the evaluation process. Each stage in GRADE requires careful analysis of specific aspects, such as the weight of studies, risk of bias, and interpretation of the forest plot, which are not automatically calculated by the software, as they are in CINeMA. This encourages a more critical and reflective assessment. While CINeMA simplifies the process, GRADE demands specific competencies, ultimately allowing for a deeper and more informed evaluation.

The CINeMA framework has been implemented as a web application, which allowed people to evaluate the certainty of the evidence easily and quickly, especially for complex and large networks. However, the assessor often lacked the skills to understand the logical steps, potentially leading to cognitive bias [13]. Without expertise, interpreting CINeMA results may be challenging. Indeed, to reduce this limitation and have a better understanding of the procedure, the need for more in-depth methodological knowledge was expressed. In addition, technical issues emerged when the server or the web hosting were offline. The web application seemed to perform well on some browsers but poorly on others. Using a web app facilitates the evaluation process, though it could be at risk of information technology problems. However, such problems may have been limited to the period of our study by chance. Assessors had also reported concrete difficulties, including problems with comma-separated values (CSV) files uploading, and importing ROB-MEN results within CINeMA. These issues can cause frustration to the assessors and organizational challenges. A further support from the developers of the

tool and further explanation of how to create the CSV files for both CINeMA and ROB-MEN may help the assessors in using the instrument for its purpose. Assessors described CINeMA as ''automated''. In accordance with Minozzi's study [5], CINeMA's automatization effectively reduces evaluation time, especially for large networks, but it represents a limitation because it compromises understanding. GRADE, whereas less manageable and more error-prone, offered a complete and transparent overview. However, automatizing the GRADE process has been suggested by some assessors as an aspect for improvement of the tool.

The provision of additional educational resources and training materials would improve the effective application of both GRADE and CINeMA, given the methodological complexity and specific competencies required by each tool.

This study represents the first systematic exploration of assessors' experience, contributing novel insights into the comparative use of GRADE and CINeMA. The tools were applied by different assessors, and this introduces variability that may reflect individual reasoning styles rather than differences inherent to the tools themselves. However, the fact that they were used on the same clinical question provided a unique opportunity to examine how each framework is interpreted and operationalized in practice. This approach allowed for the identification of overlapping challenges, divergent reasoning paths, and specific knowledge requirements associated with each tool.

The study employed a qualitative methodology, which enabled a rich and nuanced analysis of participants' reflections. Through written open-ended responses, it was possible to explore not only the technical aspects of the tools but also the cognitive processes, uncertainties, and learning needs that emerged during the evaluation. This methodological choice proved particularly effective in capturing the complexity of judgment involved in evidence assessment and in highlighting areas where further guidance or training may be beneficial.

The use of an open-ended survey approach may not have captured the depth of reflections and experiences that could have emerged from using interviews or group discussions but allowed us to efficiently collect feedback from a structured group of assessors, balancing time constraints with the need for descriptive insights. Future research could employ in-depth qualitative methods to explore these experiences more fully.

5. Conclusion

GRADE was considered slow and time-consuming, but it excelled in transparency and clarity. CINeMA was considered quick to use but showed some practical and output-interpretation challenges, particularly where specialized knowledge of NMA methods was required. Specific training courses for effective use were suggested for both tools. Adequate methodological preparation is essential for a good understanding of the evaluation process. Future studies involving assessors with different experience and knowledge may provide further insights.

Declaration of generative AI and AI-assisted technologies in the writing process

We did not use any generative AI and AI-assisted technologies in the writing process.

CRediT authorship contribution statement

Pier Carlo Battain: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Giulia Lamiani:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis. **Cinzia Del Giovane:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Silvia Minozzi:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization.

Declaration of competing interest

SM is a member of the GRADE working group. CDG is a co-author of the publications about CINeMA. There are no competing interests for any other author.

Acknowledgments

This project was developed during a postgraduate academic course focused on NMA methodology, which took place in 2023 at the University of Milan. The authors would like to thank all the teachers and students who participated in the project. In particular, the authors would like to thank Michela Cinquini for her external support and her comments.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2026.112226.

Data availability

All data collected are available in the supplementary material.

References

[1] Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. Ann Intern Med 2013;159(2):130–7. https://doi.org/10.7326/0003-4819-159-2-201307160-00008.

[2] Puhan MA, Schünemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ 2014;349:g5630. https://doi.org/10.1136/bmj.g5630.

[3] Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JPT. Evaluating the quality of evidence from a network meta-analysis. PLoS One 2014;9:e99682. https://doi.org/10.1371/journal.pone.0099682.

[4] Nikolakopoulou A, Higgins JPT, Papakonstantinou T, Chaimani A, Del Giovane C, Egger M, et al. CINeMA: an approach for assessing confidence in the results of a network meta-analysis. PLoS Med 2020; 17(4):e1003082. https://doi.org/10.1371/journal.pmed.1003082.

[5] Minozzi S, Cinquini M, Arienti C, Battain PC, Brigadoi G, Del Vicario M, et al. Grading of Recommendations Assessment, Development, and Evaluation and Confidence in Network Meta-Analysis showed moderate to substantial concordance in the evaluation of certainty of the evidence. J Clin Epidemiol 2025;184:111811. https://doi.org/10.1016/j.jclinepi.2025.111811.

[6] Lattanzi S, Trinka E, Zaccara G, Striano P, Russo E, Del Giovane C, et al. Third-generation antiseizure medications for adjunctive treatment of focal-onset seizures in adults: a systematic review and network meta-analysis. Drugs 2022;82(2):199−218. https://doi.org/10.1007/s40265-021-01661-4.

[7] Cortese S, Adamo N, Del Giovane C, Mohr-Jensen C, Hayes AJ, Carucci S, et al. Comparative efficacy and tolerability of medications for attention-deficit hyperactivity disorder in children, adolescents, and adults: a systematic review and network meta-analysis. Lancet Psychiatry 2018;5(9):727−38. https://doi.org/10.1016/S2215-0366(18)30269-4.

[8] Cinquini M, Annibale B, Attanasio R, Busetto L, Chianelli M, Tralongo A, et al. What's the best intervention for overweight and obese adult patients affected by metabolic comorbidities? PROSPERO. 2022. Available at: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022351409. Accessed January 17, 2024.

[9] Del Giovane C, Boncoraglio GB, Bertù L, Banzi R, Tramacere I. Antiplatelet drugs for secondary prevention in patients with ischemic stroke or transient ischemic attack: a systematic review and network meta-analysis. BMC Neurol 2021;21(1):319. https://doi.org/10.1186/s12883-021-02341-2.

[10] Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77−101. https://doi.org/10.1191/1478088706qp063oa.

[11] O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014;89(9):1245−51. https://doi.org/10.1097/ACM.0000000000000388.

[12] Brignardello-Petersen R, Guyatt GH, Mustafa RA, Chu DK, Hultcrantz M, Schünemann HJ, et al. GRADE guidelines 33: addressing imprecision in a network meta-analysis. J Clin Epidemiol 2021; 139:49−56. https://doi.org/10.1016/j.jclinepi.2021.07.011.

[13] Korteling JEH, Paradies GL, Sassen-van Meer JP. Cognitive bias and how to improve sustainable decision making. Front Psychol 2023;14: 1129835. https://doi.org/10.3389/fpsyg.2023.1129835.