Visual Explanation of the Assessment of Certainty of Evidence for Systematic Review and Meta-analysis

Syed Arsalan Ahmed Naqvi*
Department of Oncology
Mayo Clinic

Kunwer Sufyan Faisal
Ziauddin University

Manal Imran
Dow University of Health Sciences

Muhammad Ali Khan
Department of Oncology
Mayo Clinic

Kaneez Zahra Rubab Khakwani
The University of Arizona

M. Hassan Murad
Mayo Clinic Evidence Based Practice Center, Mayo Clinic

Huan He
Department of Biomedical Informatics and Data Science
Yale University

Irbaz Bin Riaz†
Department of Oncology
Mayo Clinic

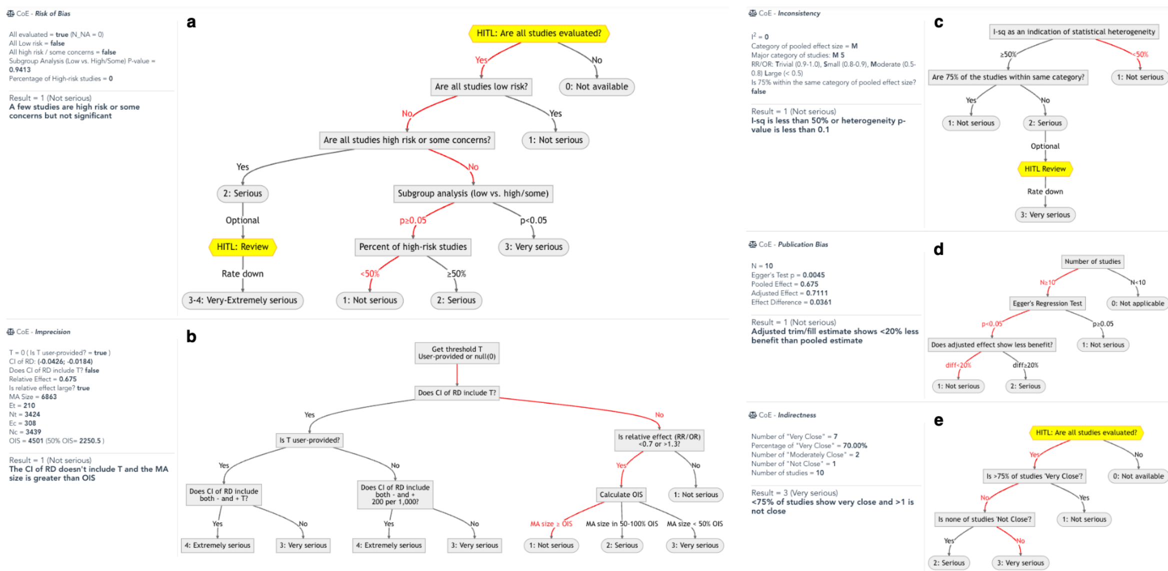

Figure 1: The visual explanations of the assessment of certainty of evidence for pairwise meta-analysis, including (a) risk of bias, (b) imprecision, (c) inconsistency, (d) publication bias, and (e) indirectness. In each sub-figure, the yellow boxes represent Human-in-the-loop process or parameters, which are provided by users. The rectangle boxes represent those steps in the certainty of evidence assessment. The round-rectangle boxes represent the decisions. The conditions between each step and decision are visualized as lines and their values are displayed as labels on the line. The red lines indicate the how the decision is made in the assessment process.

Abstract

Practitioners of evidence-based medicine need to know the level of certainty in the evidence they are applying to patient care. We use a living interactive evidence synthesis framework to create and maintain living, interactive systematic reviews (LISRs). With each new update, it is critical to report any changes to the confidence level or certainty of synthesized evidence (CoE) for patient important endpoints. Ascertaining CoE is a complex task and thus challenging in the setting of LISRs. Therefore, we propose a hybrid approach, which leverages an interactive web-based visualization techniques to accelerate the CoE evaluation.

Index terms: data visualization, systematic review, meta-analysis, certainty of evidence.

* e-mail: naqvi.syed@mayo.edu
† e-mail: riaz.irbaz@mayo.edu

1 Introduction

Certainty of evidence (CoE) is widely used in systematic reviews and meta-analyses (SRMAs) and is an essential component of appraising evidence for clinical practice guidelines and policy making [1]. It generally requires a detailed assessment of results in the context of risk of bias, inconsistency, indirectness, imprecision, and publication bias and reflects the level of confidence in the results across different patient important endpoints. This evaluation usually follows the GRADE (Grading of Recommendation, Assessment, Development, and Evaluation) approach [2], which is a guideline for CoE assessment. The rapid evolution of medical knowledge and advances in research methods make it imperative to update systematic reviews and meta-analyses regularly to ensure that the most current evidence is being used to inform decision-making.

However, there remains a lack of efficient means to assess and interpret CoE as it requires repetitive evaluations. First, the manual efforts required to assess CoE necessitate considerable time and resources which becomes even more cumbersome with periodic

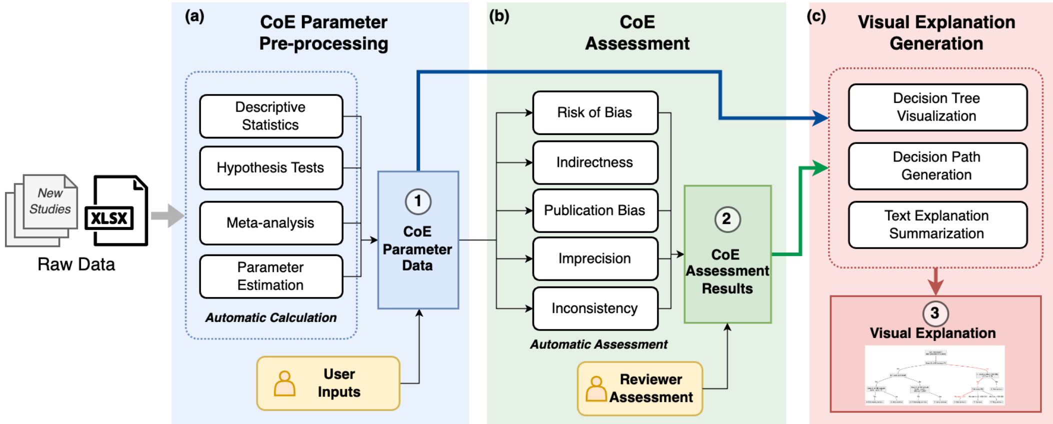

Figure 2: The architecture of our proposed system, which consists of four components: (a) CoE parameter pre-processing that calculates essential parameters for CoE assessment; (b) CoE assessment that evaluates different CoE domains based on the GRADE guideline, and (c) Visual Explanation generation that creates both visual and text explanations.

updates. Secondly, the CoE assessment results can be complex and difficult to interpret, especially for non-experts. The presentation of the results should, therefore, be clear and concise and should use appropriate terminology and visual aids to aid understanding.

To address the above challenges, we propose an interactive system to improve the interpretability of CoE assessment results by leveraging data visualization techniques. We have released an online demo to present the visual explanations of the CoE assessment results with a sample dataset at our website: https://lisr.org/coe-vis/.

2 Related Works

The GRADE approach defines the quality of a body of evidence as the extent to which one can be confident that an estimate of effect or association is close to the quantity of specific interest [1]. The construct of the GRADE approach provides criteria for rating the certainty in evidence across five relevant domains including risk of bias [3], inconsistency [4], indirectness [5], imprecision [6], and publication bias [7] for systematic reviews of randomized controlled trials. The GRADE approach defines four levels of confidence including high, moderate, low, or very low certainty in evidence depending on assessment for each domain at the level of each outcome summarized using evidence from the included trials.

While certain tools such as GRADEpro GDT [8] and MAGIC App [9] are available to facilitate the process of adjudicating CoE, the results are usually presented in static text or a few colors (e.g., red, yellow, and green indicate high, moderate, and low respectively). It's not transparent for researchers to understand how the results are generated and the decision process. Our proposed system provides a visual explanation of the assessment results using automatically generated decision trees and text summarizations, which aims to improve the efficiency and user experience.

3 System Requirements

We worked with domain experts and clinical methodologists to shape the requirements for our framework.

First, we reviewed literatures to understand the workflow to form initial design ideas. Secondly, we invited domain experts with extensive experience in SRMA and interviewed them to assess our initial ideas. Then, we iterated through weekly design and development cycles to discuss the latest progress and gather feedback from our domain experts. Each cycle begins with a demonstration of the latest visual designs, and then the comments are collected to form new requirements to guide the design refinement and prototype development.

After about 3-4 months of iterative design and development cycles, the designed framework was implemented into a prototype system with the following core features to meet the functional requirements:

R.1 Human-in-the-loop assessment: The system should involve a human-in-the-loop review and refinement of the preliminary CoE assessment results. A team of experts in the cancer-related field should be able to review the preliminary CoE assessment and provide feedback to improve the accuracy and clarity of the results. This feedback would be incorporated into the system, and the CoE assessment results would be iteratively refined until the experts are satisfied with the quality and clarity of the results.

R.2 Interactive visual summary: The system should present the CoE assessment results in a clear and understandable way. For example, users could use interactive graphs and charts to visualize the CoE assessment results for different outcomes, subgroups, or study designs.

R.3 Natural language summary: In addition to the visual summary, the system should also provide a concise and structured text summary of the evidence and its implications, such as the main findings, the magnitude of the effect, and the quality of the evidence for the main outcomes.

4 System Design

To address the abovementioned requirements, we proposed a human-in-the-loop system to facilitate the CoE assessment and interpretation by leveraging automated tools and data visualization techniques. As shown in Figure 2, the proposed system consists of three components, including (1) CoE parameter pre-processing (Fig. 2a), (2) semi-automatic CoE assessment (Fig. 2b), and (3) visual explanation generation (Fig. 2c).

4.1 CoE parameter pre-processing

According to the GRADE guideline, the final CoE adjudication is decided based on a series of answers to assessment questions. However, the uploaded raw data from users cannot be processed by the CoE algorithms due to various factors, we thus designed this component to ensure the quality of the input parameters for CoE algorithms with supports from human experts (R.1) by the following three steps:

1) Data cleaning: As the raw outcome data provided in the raw data may contain various errors (e.g., missing, or null values, wrong format, etc.), the first step in this component is to clean the data to ensure that it is accurate and consistent.

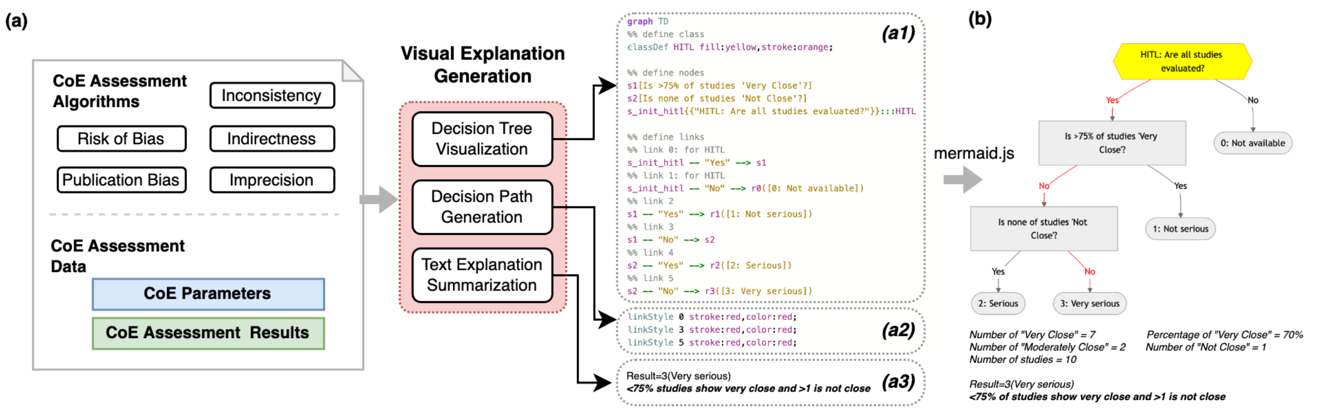

Figure 3: Creating tree diagrams based on the mermaid.js graph description language. (a) the workflow of generating the tree diagram description for visualizing the decision tree of indirectness, including (a1) the graph definition of node and links, (a2) the decision path highlighting showing the assessment process, and (a3) the text explanation related to the assessment process. (b) The tree diagram rendered by mermaid.js showing the nodes, links, decision path, and text summarizations of parameters and explanations.

This can involve correcting errors and standardizing the data format. In addition, the missing value is a common issue in SRMA studies. To avoid calculation errors, we removed those records with missing or non-processable values (e.g., non-standardized values and data encoding issues).

2) Standardization: The input raw data format is usually designed for data storage efficiency, which cannot be directly used in statistics and parameter calculation. Therefore, we need to convert the cleaned data those formats required by SRMAs.

For example, as the pairwise meta-analysis library requires a tabular-like data frame, we convert the database records from a JSON (JavaScript Object Notation) dictionary format to vectors and ensure the data type of each value matches the library requirements (e.g., some libraries require input values are represented in the logarithmic format while others require differently).

3) Calculation: Once the data is cleaned and standardized, this component will perform meta-analyses provided by multiple R libraries (e.g., meta and metafor) and other user inputs to generate the parameters for CoE assessment.

A pre-processed sample dataset is available for reference at our website: https://lisr.org/coe-vis/pwma_sample.xlsx.

4.2 Semi-automatic CoE Assessment

Once all the essential CoE parameters are calculated, this component can assess the five CoE subdomains in a semi-automatic approach, including the risk of bias, indirectness, inconsistency, publication bias, and imprecision (R.1).

We followed the GRADE guidelines to implement the algorithms of five subdomains with a human-in-the-loop design. Reviewers can also review the results and adjudicate them. The workflow of each subdomain is as follows:

Risk of bias (Fig. 1a): First, the risk of bias of each individual study included in the meta-analysis is evaluated manually by reviewers. Then, the overall risk of bias of each outcome is calculated using R packages.

Imprecision (Fig. 1b): both the user-input thresholds and the meta-analysis results will be used to assess the imprecision.

Inconsistency (Fig. 1c): the pre-processed heterogeneity test results and meta-analysis results are used to assess the inconsistency. Reviewers may also be involved to review the final results.

Publication bias (Fig. 1d): the pre-defined thresholds and subjective evaluation of domain-specific questions are used to assess the publication bias.

Indirectness (Fig. 1e): like the evaluation of the risk of bias, the indirectness of each study included in the meta-analysis needs to be assessed by reviewers first. If the indirectness of all studies is evaluated, the overall indirectness will then be calculated by using both meta-analysis results and reviewers' results.

4.3 Visual Explanation Generation

To improve the interpretability of the CoE assessment result, we proposed a visual explanation generation method by combining both visual and text information to present the decision-making process based on dynamic tree diagrams (R.2).

First, as the CoE assessment process can naturally be depicted as a decision tree, we adopted the tree diagram to visualize the CoE assessment process. As shown in Fig. 3a, the tree diagram can be defined using three sections of mermaid.js chart syntax [10], including a diagram definition that describes the figure settings, nodes that describe the tree leaves for CoE assessment, and links describing the tree branches (Fig. 3(a1)).

Secondly, the decision path on the tree diagram can be highlighted by converting CoE parameter data and CoE results into visual styles (Fig. 3(a2)), which can help users to identify how the decision is made through the tree diagram. Each line in Fig. 3(a2) represents a tree branch that needs to be highlighted in a tree diagram. By combing and rendering the tree diagram (Fig. 3(a1)) and the highlighted path (Fig. 3(a2)), we can get a tree diagram that shows the CoE assessment process with the visual guidance of how the decision is made step by step (Fig. 3b).

Thirdly, to help users understand the CoE assessment result, we developed a rule-based algorithm to generate a text summarization as a brief explanation (Fig. 3(a3)) for the CoE assessment (R.3). The summarization algorithm takes CoE parameters and results as input, follows the same algorithm of each CoE domain, and outputs a text explanation, which includes the text description of the related CoE parameters and a text summary of reason.

Lastly, the tree diagram of the CoE assessment process is created from the diagram definition by mermaid.js (Fig. 3b), and the text explanation is added along with the tree diagram so that users can understand the assessment process through both visual and textual information.

5 Case Study and Discussion

As our proposed framework is still in the development stage, we didn't conduct a formal usability evaluation. We presented the prototype system to our domain experts and SRMA researchers during the development and got very positive feedback from them. The screenshot of our prototype system is shown in Fig. 4a. In

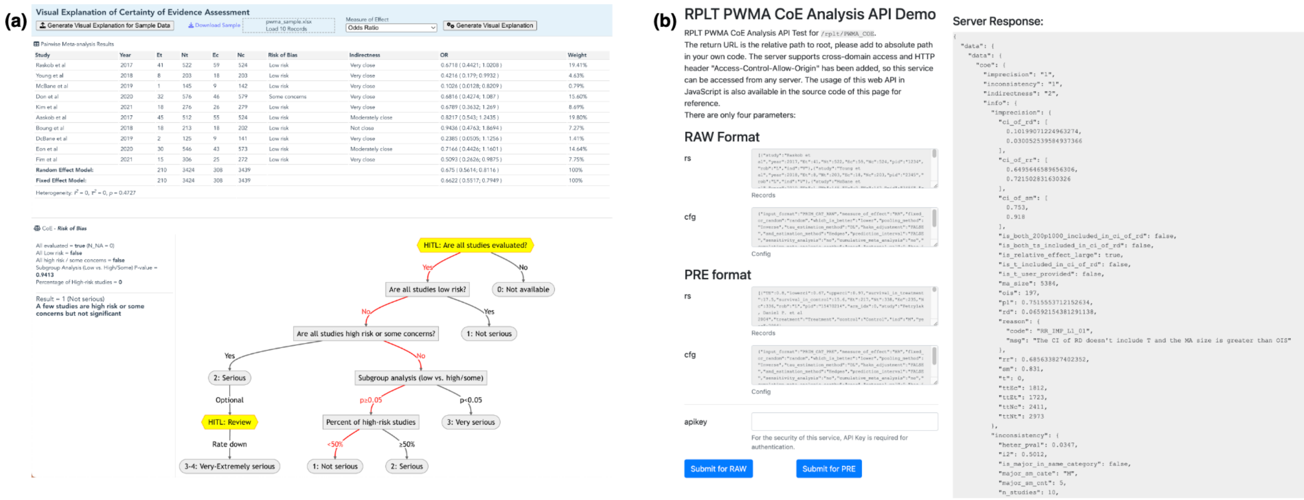

Figure 4: (a) The screenshot of the online demonstration that allow users to upload an Excel file to generate visual explanations of the CoE assessment results. (b) The screenshot of the web API that allow users to send data directly to our web service for CoE assessment and visual explanation generation, which enables the integration to other external systems.

addition, we also built a web API to provide the visual explanations of the CoE assessment results per users' input (Fig. 4b).

First, domain experts and researchers commented that the visual explanations could significantly improve the understanding of the assessment process. By reading the tree diagram and colored path, it's intuitive to interpret the CoE assessment results and make more informed decisions based on the available evidence. Moreover, they commented that those visual aids, both the tree diagram and text information can help to improve communication and facilitate discussion among stakeholders. By presenting intuitive diagrams, stakeholders from different backgrounds and areas of expertise can collaborate more effectively.

Secondly, domain experts were impressed by the quick response of the system, which is a significant benefit in terms of resource-saving and time-saving. The system can help to reduce resource requirements and save time that would otherwise be spent on repetitive manual efforts and interpreting complex statistical results. They suggested that our approach could potentially decrease the workload of CoE assessment.

Thirdly, domain experts indicated that the CoE assessment of the certainty of evidence could be influenced by personal biases or preconceptions. Researchers may interpret the evidence in a way that supports their own beliefs or research agenda, leading to a biased presentation of the results. It is important to be transparent about potential biases and to present the evidence objectively.

Therefore, the feedback from domain experts indicate that our proposed visualization system could help to increase transparency in the CoE assessment process for researchers.

5.1 Limitations

While visual explanations can be helpful in improving the interpretability of CoE assessment results, our domain experts also commented that there are several potential limitations that should be considered:

The complexity of visual representations: as the number of branches in the tree diagram increases, the visual representations can be complex and difficult to interpret, particularly for non-experts. Especially when the assessment process involves multiple conditions and steps, the complex tree branches and nodes may lead to misunderstandings and incorrect conclusions.

Limited standardization: There is currently limited standardization in the use of visual representations for presenting CoE assessment results. The visual design and terminology may not be different from other SRMA projects, which can make it difficult to compare and interpret results across different projects.

Overreliance on visuals: the tree diagram is intuitive but also needs more space to display and browser. It needs interactivity designs to reduce the burden of reading many diagrams, such as making the diagram as foldable panels, which can be integrated into other systems for validating or exploring the CoE assessment results.

Technical limitations: While our online demonstration provides an example of how the visual explanation of CoE assessment results works, it still requires users to understand the data format and some parameters for meta-analysis to use it in their own system. In addition, some stakeholders may lack the technical expertise or resources needed to use or interpret the visualization provided by our web service API. It would be great to provide more tutorial and samples for users.

Overall, by providing a comprehensive visualization of the CoE assessment results and parameters, users can more easily assess the reliability and validity of the findings. Meanwhile, while those visual explanations can be a valuable tool for improving the interpretability of CoE assessment results, electronic medical records we need to recognize their potential limitations and use them alongside other forms of evidence to ensure the completeness and correctness.

6 Conclusion and Future Work

To facilitate the interpretability of the CoE assessment results, we developed a system to generate visual explanations by leveraging a semi-automatic approach and data visualization techniques. We implemented a prototype to validate the effectiveness of the system based on open-source software. Qualitative feedback from experienced domain experts is positive, while there are also some potential limitations. They suggest that our system provides a clear and intuitive representation of the CoE assessment results, which could enhance the interpretability and transparency of the CoE assessment process.

In the future, we plan to improve our system from the following aspects. First, we plan to improve the interactivity design of the tree diagram to make it more intuitive. Secondly, we plan to implement the CoE assessment on the network meta-analysis and integrate this system into other actual systematic review and meta-analysis projects to validate its performance. Thirdly, we plan to conduct a formal usability evaluation to improve system design.

References

[1] M. H. Murad, R. A. Mustafa, H. J. Schünemann, S. Sultan, and N. Santesso, "Rating the certainty in evidence in the absence of a single estimate of effect," *Evid Based Med*, vol. 22, no. 3, pp. 85–87, Jun. 2017, doi: 10.1136/ebmed-2017-110668.

[2] G. Guyatt *et al.*, "GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables," *J Clin Epidemiol*, vol. 64, no. 4, pp. 383–394, Apr. 2011, doi: 10.1016/j.jclinepi.2010.04.026

[3] G. H. Guyatt *et al.*, "GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias)," *J Clin Epidemiol*, vol. 64, no. 4, pp. 407–415, Apr. 2011, doi: 10.1016/j.jclinepi.2010.07.017.

[4] G. H. Guyatt *et al.*, "GRADE guidelines: 7. Rating the quality of evidence--inconsistency," *J Clin Epidemiol*, vol. 64, no. 12, pp. 1294–1302, Dec. 2011, doi: 10.1016/j.jclinepi.2011.03.017.

[5] G. H. Guyatt *et al.*, "GRADE guidelines: 8. Rating the quality of evidence--indirectness," *J Clin Epidemiol*, vol. 64, no. 12, pp. 1303–1310, Dec. 2011, doi: 10.1016/j.jclinepi.2011.04.014.

[6] G. H. Guyatt *et al.*, "GRADE guidelines 6. Rating the quality of evidence--imprecision," *J Clin Epidemiol*, vol. 64, no. 12, pp. 1283–1293, Dec. 2011, doi: 10.1016/j.jclinepi.2011.01.012.

[7] G. H. Guyatt *et al.*, "GRADE guidelines: 5. Rating the quality of evidence--publication bias," *J Clin Epidemiol*, vol. 64, no. 12, pp. 1277–1282, Dec. 2011, doi: 10.1016/j.jclinepi.2011.01.011.

[8] A. Basu, "A tutorial on how to use Gradepro GDT Tool for writing reviews," PeerJ Inc., e2520v1, Oct. 2016. doi: 10.7287/peerj.preprints.2520v1.

[9] "MAGICevidence website." Accessed: Sep. 12, 2024. [Online]. Available: https://www.magicevidence.org/

[10] K. Sveidqvist and Contributors to Mermaid, *Mermaid: Generate diagrams from markdown-like text*. (Dec. 2014). JavaScript. Accessed: Sep. 12, 2024. [Online]. Available: https://github.com/mermaid-js/mermaid