國立臺灣大學公共衛生學院公共衛生學系

學士班學生論文

Department of Public Health

College of Public Health

National Taiwan University

Bachelor's Thesis

證據偏誤對網絡統合分析結果影響之視覺化評估

Visual Assessment of the Impact of Evidence Bias on the Results of Network Meta-Analysis

梁嫚芳

Man-Fang Liang

指導教授：杜裕康 博士

Advisor: Yu-Kang Tu, DDS, Ph.D.

中華民國 112 年 4 月

April 2023

國立臺灣大學學士班學生論文
口試委員會審定書

證據偏誤對網絡統合分析結果影響之視覺化評估

Visual Assessment of the Impact of Evidence Bias
on the Results of Network Meta-Analysis

本論文係 梁嫚芳 君（B07801003）在國立臺灣大學公共衛生學系完成之學士班學生論文，於民國 112 年 4 月 7 日承下列考試委員審查通過及口試及格，特此證明

口試委員 ：　　（簽名）

（指導教授）

系主任　　：　　（簽名）

中文摘要

網絡統合分析 (Network meta-analysis, NMA) 透過單一統計模型同時比較多種療法，突破過去傳統統合分析的限制，但最終估計結果的品質優劣影響臨床決策的正確性，其中尤以所納入的研究的偏倚影響甚鉅，因此基於個別研究偏倚風險的證據品質評估對於臨床決策至關重要。

本研究以 NMA 所納入的個別研究的比例貢獻 (proportion contribution) 與偏倚風險 (Risk of Bias, RoB) 為基礎，自證據流網絡 (evidence flow network) 結構切入，參考二端串並聯電路 (Two-Terminal Series-Parallel Graphs, TTSPG) 概念，建立穩健性指標 (robustness index)，以量化評估證據偏誤對 NMA 結果的影響。此外，本研究開發互動式 Shiny app 介面，以新的視覺化方法呈現證據流網絡結構與穩健性評估結果。穩健性指標助於正確評估證據品質，Shiny app 使臨床從業人員能夠快速解讀 NMA 估計結果的品質，其中，網絡圖幫助識別低穩健性之療法對比，並能進一步檢視流網絡之結構與穩健性，長條圖助於查看證據流中各研究證據偏倚之分布情形。如此便使臨床從業人員能避免採用高偏倚風險之 NMA 估計結果，最終得以於進行決策時據此選擇適切的治療方式，改善患者預後情況。

關鍵字：網絡統合分析、比例貢獻、證據流網絡、穩健性、視覺化

Abstract

Network meta-analysis is a technique for comparing multiple treatments, overcoming the limitations of traditional meta-analysis. However, the quality of evidence of NMA affects the accuracy of clinical decision-making, especially the bias in the studies included. Therefore, it is essential to evaluate the quality of evidence based on the risk of bias in individual studies.

Based on the proportional contribution and Risk of Bias (RoB) of each trial, the robustness index for quantitatively assessing the quality of evidence in NMA was derived by considering the evidence flow networks that can be compared to two-terminal series-parallel graphs. To present the results of the assessment of evidence quality, the study developed a Shiny application including new interactive visualization techniques based on network plots and bar charts.

The robustness index improves the accuracy of the assessment of NMA quality by considering the structure of evidence flow networks. The Shiny application facilitates the interpretation of the quality of evidence in NMA. In this application, the network plots enable users to identify treatment contrasts with low robustness and to further examine the structure and robustness of flow networks. The bar chart visualizes the distribution of the risk of bias in studies of each stream. These results prevent clinical practitioners from placing confidence in the NMA estimation with a high risk of bias and inform clinical decision-making about interventions; hence improving the prognosis of patients.

Keywords: Network meta-analysis, Proportion contributions, Flow networks, Robustness, Visualization

Contents

List of Figures

List of Tables

1. Introduction

With the increasing number of clinical articles, comprehensively reviewing all the relevant literature has been a considerable challenge to researchers. Besides, the large variation among the results of each literature makes it difficult to provide a certain conclusion for researchers. Meta-analysis, a statistical method to calculate a pooled estimate of treatment effect based on a systematic review of multiple studies, has been developed to address these problems. This method provides evidence of the efficacy of interventions in healthcare by summarizing and evaluating research findings [1]. However, traditional pairwise meta-analysis (TMA) can only compare two treatments at a time. When multiple treatments are available for a particular disease, TMA can only provide pairwise comparisons and has limitations in conducting multiple comparisons, such as inconsistency of results and lack of evidence for direct comparison between some treatments. The outcomes of TMA cannot adequately inform clinical decision-making since the whole picture of treatment effects cannot be captured [2].

Network meta-analysis (NMA) compares multiple treatments simultaneously by synthesizing all available evidence into a single statistical model and generating a coherent network of evidence, thereby increasing the certainty of all outcome estimates. NMA also provides indirect evidence by comparing treatments to common comparators in the absence of direct evidence. NMA provides a comprehensive assessment of treatment effects by combining direct evidence and indirect evidence estimated by the connection of the network, which surpasses the limitations of TMA [3-5].

We wish to compare three treatments, A, B, and C, as shown in **Figure 1**, but the TMA method can only provide pairwise comparisons: direct evidence “A versus C” and “B versus C”. NMA connects the two direct comparisons with common comparator C and estimates the treatment effect size for the "A versus B" comparison. For example, if A is superior to C by 3 units and C is superior to B by 2 units, it can be inferred that A is superior to B by 5 units. The "A versus B" comparison is called an indirect comparison as they are not directly connected in the network graph, and have not been directly compared in previous studies. This cycle "A-B-C" in the network graph is called a triangular loop.

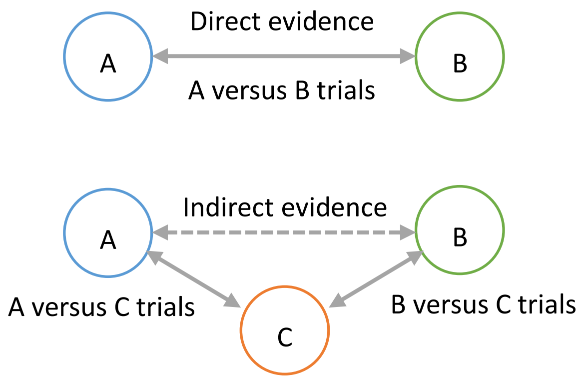


Figure 1. Visual representation of direct and indirect evidence toward the comparison of A versus B.

The appropriateness of interpreting the clinical effects of interventions is influenced by the quality of evidence obtained from NMA, which usually can be impaired by the within-study bias of the studies included in the evidence network. To evaluate the quality of evidence, the risk of bias (RoB) of each study and its contribution to the NMA estimates must be considered. Past research has established various approaches for estimating the proportion contribution of studies and guidelines for evaluating the RoB. Since an NMA contrast estimate can be seen as a linear combination of direct estimates, the evidence stream network should be considered when assessing the quality of NMA. Disregarding the flow network and only taking the weighted average by contribution and RoB led to neglecting the impact of the distribution of RoB in the network. To determine the potential bias due to conflicting directions of bias from the studies, the assessment of NMA quality must start with contributions of individual studies, instead of pooled direct evidence. Previous studies have not developed a quantitative index that considers the proportion contribution of individual studies and the structure of the evidence flow network for evaluating the quality of NMA evidence.

Therefore, the purpose of this study was to develop a robustness index to quantitatively assess the quality of NMA evidence based on the proportion contribution and RoB of studies as well as evidence stream network structure. Moreover, the other objective of the study was to develop new techniques for visualizing the quality of evidence based on network plots and bar charts. These enable clinical practitioners to easily interpret the quality of NMA estimates, increasing the accuracy of decision-making about interventions.

2. Literature Review

2.1. Quality of Evidence from Network Meta-Analysis

It is imperative to evaluate the quality of evidence from an NMA before incorporating it into clinical decision-making. The quality of evidence refers to the reliability or certainty of treatment effect estimates. Puhan et al. [6] and Salanti et al. [7] proposed methods of assessing the quality of evidence for NMA, based on the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) working group approach. In the standard GRADE guidelines [6, 8], five criteria will be used to downgrade the quality of evidence of randomized controlled trials (RCTs) based on the results of a systematic review to determine whether there is (1) risk of bias (RoB); (2) inconsistency; (3) indirectness; (4) imprecision; and (5) publication bias among the studies included in NMA. The quality of evidence is then classified as high, moderate, low, or very low according to the GRADE guidelines.

Risk of bias (RoB), one of the criteria in GRADE, refers to flaws in the design or implementation of a study, which result in an estimated treatment effect that systematically differs from the actual effect [9]. This kind of within-study bias is crucial for the evaluation of evidence quality in both TMA and NMA, as biased studies impair the validity of the estimates and, consequently, the appropriateness of clinical decisions. In TMA, RoB assessment is straightforward; the Cochrane RoB tool can be used to evaluate the included studies and generate a visual RoB table shown in **Figure 2**, to further assess the impact on the results [10]. However, RoB assessments in NMA pose a challenge since the contribution of direct evidence from individual studies to different treatment contrast estimates varies and cannot be directly obtained from the analysis results [11].

Figure 2. **Hypothetical presentation of risk of bias assessments for studies.**

The approach for rating the quality of NMA evidence in each direct and indirect estimate proposed by Puhan et al. is qualitative and lacks quantitative assessment. To increase efficiency for evaluating the indirect estimate, this approach focused on the assessment of the indirect evidence on the dominant first-order loop, the triangular loop as mentioned in Introduction. For example, a comparison between four interventions was presented as an evidence network shown in **Figure 3**, where the absence of lines between nodes indicates no trials comparing two treatments. For the indirect comparison "C versus D", the red solid line represents the first-order loop, which includes only one additional treatment "A" apart from the treatment being compared. The second-order loop involves two additional single treatments "A" and "B", as shown by the green dashed lines. Higher-order loops involve even more additional treatments. When both direct and indirect evidence are available for a specific comparison, the higher quality of the two estimates is used to determine the evidence quality. In other words, this approach does not consider the contribution of the remaining studies. Focusing on the first-order loop may lead, however, to ignoring an important proportion of the evidence and to inappropriate ratings of the certainty of estimates [12].

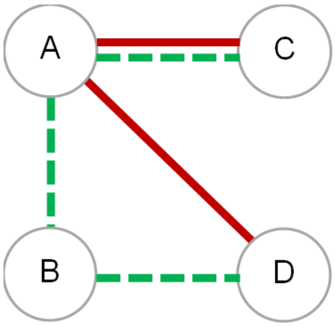


Figure 3. Hypothetical evidence network plot comparing the effects of four drugs.

Therefore, Puhan suggested researching on evaluating the contribution of individual studies, i.e., weight, to assess the quality of NMA [6]. On the other hand, Salanti et al. more extensively used weights in their approach and provided a more quantitative evaluation of the quality of evidence, a weighted average of "downgrading levels" that downgrades for the comparisons that are mostly contributed by studies with high risk of bias [7]. The weighted average, however, is based on the contribution of pooled direct evidence instead of individual studies. To understand the impact of RoB on the treatment contrast estimates and the entire network estimates, the weight of direct evidence from individual studies must be fully considered when assessing evidence quality [13].

2.2. Proportion Contributions and Flow Networks

Salanti's approach estimated the contribution matrix for each pooled direct estimate, instead of each direct evidence from individual studies, as shown in **Figure 4** [7]. The approach, however, cannot be used to determine the potential bias in contrast estimates due to conflicting directions of bias from the studies after pooling the same treatment comparisons from several studies into a single direct estimate. Thus, it is essential to estimate the amount of information that individual studies contributed to the estimation to identify positions and directions susceptible to bias in a network, providing an evaluation of the robustness of the evidence.

		Direct evidence				
		AB	AC	BC	BD	CD
NMA estimates	Mixed estimates					
	AB	15	37	10	33	5
	AC	12	70	4	7	7
	BC	2	2	67	10	19
	BD	5	5	20	52	18
	CD	4	4	33	32	27
	Indirect estimates					
	AD	9	28	23	23	17

Figure 4. Hypothetical contributions matrix: percentage contribution of each direct estimate to the network meta-analysis estimates.

Salanti et al. suggested considering the relative contribution of different studies to the overall NMA estimates to assess the impact of bias. To convert each entry in the absolute contribution matrix to the proportion contribution, they suggested standardizing the absolute values of each row and interpreting them as proportions [7]. However, each row of the contribution matrix can be interpreted as an evidence flow network [14, 15], including evidence streams with various paths and flows; the weight of each path is assigned to direct evidence involved. Thus, taking standardized absolute values would overestimate the contribution of direct evidence involved in long paths and underestimate the weights of the shortest paths by ignoring the multiple occurrences of identical values.

Nikolakopoulou and Papakonstantinou addressed this issue and proposed a proper method of converting contribution matrix entries into proportions based on the observation that the contribution matrix represents the flow of evidence [16]. They modified the H matrix from König et al.'s study [14] to proportions by decomposing flows into proportion, where each row sums to 100%. The study presented a new iterative algorithm that obtains the proportion contribution of pooled direct estimates and further calculates the contribution of individual studies to a single contrast estimate. In addition, the study visualized the proportion contributions of each study to each treatment contrast estimate with bar charts using Confidence In Network Meta-Analysis (CINeMA) [17], which were color-coded based on low (green), moderate (yellow), and high (red) risk of bias, as shown in **Figure 5** [16].

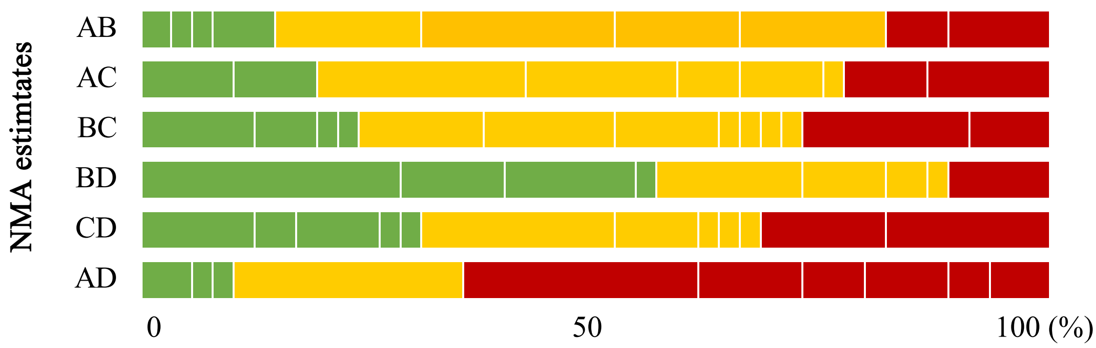


Figure 5. Hypothetical bar chart displaying the study proportion contributions to treatment contrast estimates with low (green), moderate (yellow), and high (red) RoB.

However, the visualization of the contributions of direct evidence from different studies and their RoB has limited applicability to clinical practice. A contrast estimate can be decomposed into multiple evidence streams with respective paths and flows, each of which is composed of evidence with varied amounts of contributions and qualities, affecting the robustness of the NMA estimate. Take a contrast estimate comprised of a high proportion of high RoB evidence for example, which would be like the contrast bar "A versus D" consisting large proportion of red color in **Figure 5**. There are three streams "A-B-D", "A-C-D", and "A-B-C-D" for the treatment contrast estimate "A versus D", as shown in **Figure 6**. The width of the lines represents the flow of evidence streams. The high RoB studies would either mostly distribute in a single stream of the contrast estimate as shown in **Figure 6**(a) or evenly disperse among several streams as shown in **Figure 6**(b). In the former case, *stream 1* in **Figure 6**(a) would be a high RoB stream, requiring cautious evaluation of the contrast estimate's quality if the stream also had a high flow.

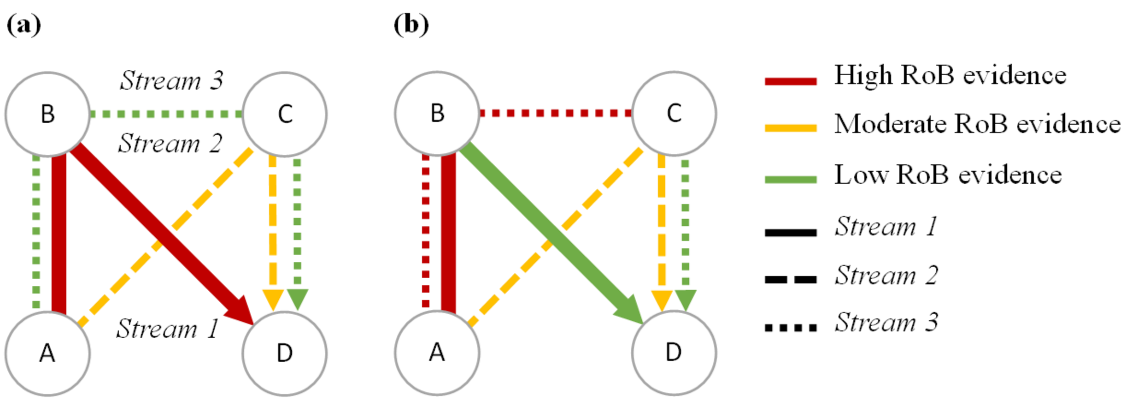


Figure 6. Concentration (a) and dispersion (b) of high RoB evidence among streams in a treatment contrast estimate.

Moreover, for the composition of a single evidence stream; the concentration and dispersion of high RoB evidence among edges also affect the robustness of estimates. This study compares the stream networks to two-terminal series-parallel graphs (TTSPG), where two-terminal (source and sink) multi-graphs built from edges use two operations: series and parallel compositions [18]. The networks can also be regarded as electrical circuits, where high RoB is equivalent to the high risk of open circuits, and the robustness of an evidence stream is equivalent to the probability of current flow through the stream. Once high RoB evidence (i.e., high risk of an open circuit) is concentrated in one edge in the stream as shown in **Figure 7(a)**, the current is less likely to flow in the circuit, resulting in lower robustness than when high RoB evidence is dispersed across different edges as shown in **Figure 7(b)**. An assumption underlined the idea that the concentration of high RoB evidence on one edge has lower robustness is that the evidence comparing the same treatment comparison is likely to have the same direction of bias. This within-edge correlation leads to an accumulation of bias with the same direction in an edge. In contrast, the evidence comparing different treatment comparisons have different directions of bias, which can be canceled out. If the quality of evidence was evaluated merely based on proportion contributions and RoB, the outcomes would be the same regardless of the distribution of high RoB evidence.

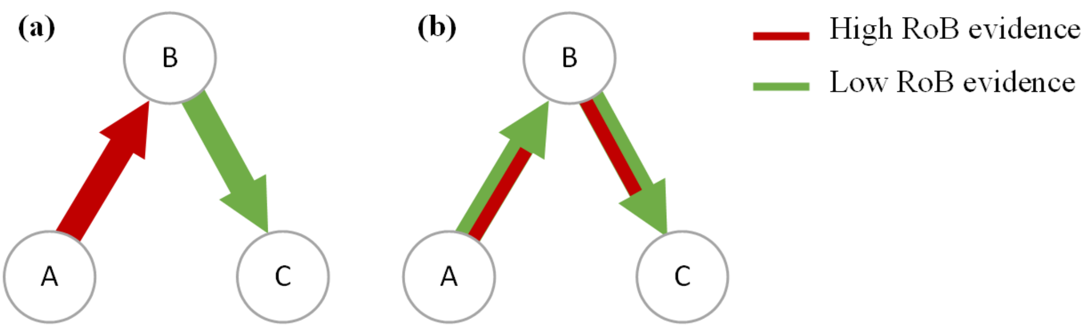


Figure 7. Concentration (a) and dispersion (b) of high RoB evidence among edges in a stream.

When assessing the quality of NMA evidence in clinical research, we should also investigate the evidence flow network, including flow, proportion contribution distribution, and RoB, to further identify the distribution of studies susceptible to bias in the evidence network structure. This can provide a more nuanced understanding of the quality of evidence from NMA contrast estimates; and further, indicates whether to remove high RoB research evidence.

2.3. Visualization of the Quality of Evidence

The network structure of NMA can be visualized using a network plot to display the relationships between different interventions, as shown in **Figure 8**. Lines that connect nodes representing different treatments indicate comparisons made in RCTs, while the absence of such lines indicates no studies (i.e., no direct evidence) comparing these two treatments. Typically, the width of lines is proportional to the number of trials comparing the treatments, and the size of nodes is proportional to the number of participants (sample size) with random assignment. Various software, such as Stata and R, can be used to generate network plots. Two articles published in 2017 and 2021 reviewed the differences in features of packages for R network visualization [19, 20]. Based on these reviews, this study estimated and visualized the robustness of NMA with the network plot and bar chart; and further developed an interactive user interface using the Shiny package.

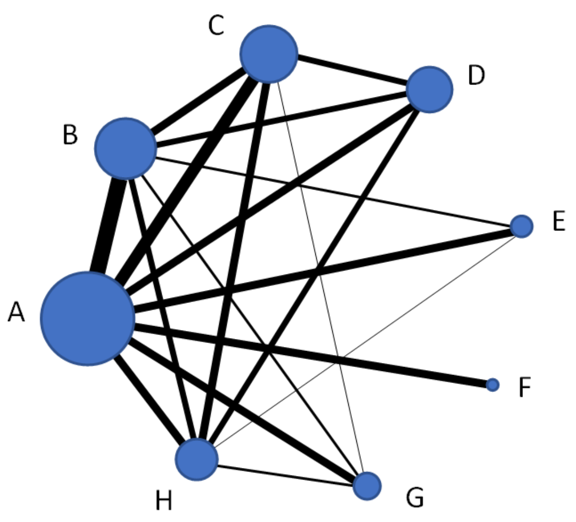


Figure 8. Hypothetical network plot.

2.4. Summary

Direct evidence from individual studies in NMA can be rated as low, moderate, or high RoB according to the GRADE guidelines. The approach for assessing the quality of NMA using RoB proposed by Puhan et al. [6] was qualitative and ignored the impact of higher-order loops. Hence, Puhan et al. suggested considering the weights of studies for evaluation. Although the quantitative approach developed by Salanti et al. [7] included contributions of studies for evaluating the quality of evidence, the weighted average they proposed was for “downgrading levels” of quality, instead of a general quantitative index for assessing the impact of RoB of individual studies. In addition, Salanti et al.’s approach only considered the contribution of pooled direct evidence rather than individual studies, thus the potential bias in contrast estimates due to conflicting directions of bias from the studies cannot be determined.

Methods for decomposing the flows in a network and estimating the proportion contribution have been developed [16]. Each NMA treatment contrast estimate can be decomposed into an evidence flow network. Without considering the flow network structure and directly calculating the weighted average based on proportion contribution for assessing the quality of NMA would result in a neglect of the influence of RoB distributions in the network. Flow networks can be compared to two-terminal series-parallel graphs (TTSPG).

2.5. Aims of This Study

This study aims to quantify the impact of evidence bias on NMA estimates. We compared RoB to the probability of open circuits and combined it with proportion contribution for constructing a robustness index that can be understood as the probability of functioning circuits. Furthermore, to enhance the interpretability of the quality of direct evidence in NMA, we developed an interactive Shiny app to visualize the robustness estimation results, supporting clinical decision-making.

3. Methods

3.1. Theoretical Foundation

Based on the method proposed by Krahn and König [14, 21], the H matrix, also known as the contribution matrix, can be estimated. The entries in the matrix represent the generalization of the study weights between -1 and 1, depending on the precision of the studies, heterogeneity between studies, and network structure. For the fixed-effects models in NMA, the network can be expressed as the following generalized linear model (GLM):

$$\hat{\theta}^{D} = X\theta^{net} + \epsilon, \qquad \epsilon \sim MVN(0, V)$$

The model represents a network containing T treatment nodes, with a total number of D treatment effects of direct comparisons from the studies included. The vector $\hat{\theta}^{D}$ of length D contains all the direct effect estimates. X is a design matrix with dimensions $D \times (T-1)$, representing the linear relationship between the direct effects and the basic parameters. The vector θ^{net} of length $T-1$ represents the basic parameters, corresponding to $T-1$ treatment effects of all comparisons to the reference treatment. The vector ϵ represents the measurement error associated with each estimate, assumed to be independent across studies and distributed normally with a known covariance structure $cov(\epsilon) = V$. If the network does not include multi-armed studies, V will be a covariance matrix that has a diagonal form.

The vector of the basic parameters θ^{net} can be estimated in a classical frequentist manner by generalized least squares as follows:

$$\hat{\theta}^{net} = (X'V^{-1}X)^{-1}X'V^{-1}\hat{\theta}^{D}$$

$$\hat{\theta}^{NMA} = Y\hat{\theta}^{net} = Y(X'V^{-1}X)^{-1}X'V^{-1}\hat{\theta}^{D} = H\hat{\theta}^{D}$$

The vector $\hat{\theta}^{NMA}$ of length C_2^T contains all the contrast estimates in NMA, including direct, indirect, or mixed evidence. Y is a design matrix with dimensions $C_2^T \times (T-1)$, representing the linear relationship between the contrast estimates and the basic parameters estimate. When direct evidence exists for all treatment comparisons, design matrix X and Y are equivalent. The H matrix can be obtained from the above equation:

$$H := Y(X'V^{-1}X)^{-1}X'V^{-1}$$

The dimension of the H matrix is $D \times C_2^T$. The rows of the H matrix contain all the NMA effects estimates, and each column corresponds to a direct comparison. Each entry indicates the weight of a direct comparison to a contrast estimate. These weights, also known as contributions, may be positive or negative depending on the direction of comparison.

Nikolakopoulou and Papakonstantinou translated the entries of the H matrix proposed by König et al. [14] to proportion contributions, where the sum of each row is 100% [16]. The study developed a novel iterative algorithm for calculating the proportion contribution of each direct evidence to a contrast estimate; and further, the proportion contribution of each study. To demonstrate the concepts, the study used a network of topical antibiotics for the treatment of chronic otitis media with ear discharge in patients with eardrum perforations.

The network compared four treatments x, y, u, and v, as shown in **Figure 9** [16]. The treatment effect was measured using the odds ratio (OR). Direct evidence existed for all comparisons except u versus x. The H matrix showing the contribution of each direct evidence to a contrast estimate for the network was demonstrated in **Table 1** [16], where columns correspond to direct evidence and rows correspond to NMA treatment contrast estimates. Positive values favor the first treatment. The matrix P derived by Papakonstantinou et al. showed the proportion contribution (%) of direct evidence indicated in the columns to the NMA treatment contrast estimates indicated in the rows, as shown in **Table 2** [16].

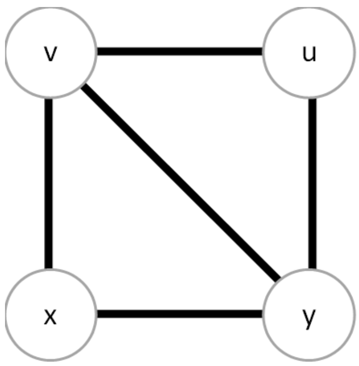


Figure 9. Hypothetical network plot with four interventions x, y, u, and v.

Table 1. H matrix for the network of four interventions x, y, u, and v.

	xy	**xv**	**yu**	**yv**	**uv**
xy	0.64	0.37	–0.11	–0.25	–0.11
xu	0.60	0.40	0.63	–0.03	–0.37
xv	0.55	0.46	0.17	0.38	0.17
yu	–0.03	0.03	0.75	0.22	–0.26
yv	–0.09	0.09	0.28	0.63	0.28
uv	–0.06	0.06	–0.46	0.40	0.54

Table 2. Matrix P of proportion contribution for the network of four interventions x, y, u, and v.

%	xy	xv	yu	yv	uv
xy	63.5	16.4	3.8	12.6	3.8
xu	30.1	19.4	31.1	1	18.4
xv	24.4	45.5	5.7	18.8	5.7
yu	1.1	1.1	74.5	11.1	12.2
yv	4.5	4.5	14.2	62.7	14.2
uv	1.9	1.9	22.1	20.2	53.8

To illustrate the algorithm of deriving the matrix P shown in **Table 2**, take one row of the H matrix (i.e., a single contrast estimate) for example, the NMA effect estimate xy is denoted as h^{xy}. König et al. [14] stated that h^{xy} can be interpreted as a flow network from source x to sink y and visualized with a directed acyclic graph (DAG), as shown in **Figure 10**(a) [16]. The contribution of direct evidence uv to the NMA estimate xy is denoted by h_{uv}^{xy}. The flow of direct evidence uv to the NMA estimate xy is denoted by f_{uv}^{xy},

$$f_{uv}^{xy} = \left|h_{uv}^{xy}\right|,$$

corresponding to the absolute value of each cell (i.e., contribution) in the H matrix. Since the example focused on the decomposition of the NMA estimate xy, the superscript "xy" will be neglected in the following expressions.

The number of iterations I in this algorithm corresponds to the number of evidence streams from source x to sink y as shown in **Figure 10**(b) [16]. A stream is obtained by searching for the shortest path from vertex x to y in each iteration, denoted by S_i, where $i = 1, \dots, I$. A stream S_i is defined as the composition of a path and its associated flow, $S_i = (\pi_i,\ \varphi_i)$. The flow of a stream φ_i is defined as the minimum value among flow of direct evidence f_{uv} along the path π_i of the stream. For instance, the stream S_2 of path π_2 in **Figure 10**(b) is composed of edge (i.e., direct evidence) "x to v" and "v to y" with the flow of $f_{xv} = 0.365$ and $f_{vy} = 0.251$, respectively. The corresponding flow φ_2 is determined by the minimum value among f_{xv} and f_{vy}, which is 0.251. The proportion contribution of direct evidence xv to the NMA contrast estimate xy, denoted by p_{xv}^{xy} (or p_{xv}), is the cell corresponding to the column xv in the row xy in the matrix P. The calculation of p_{xv} first requires identifying the relevant paths.

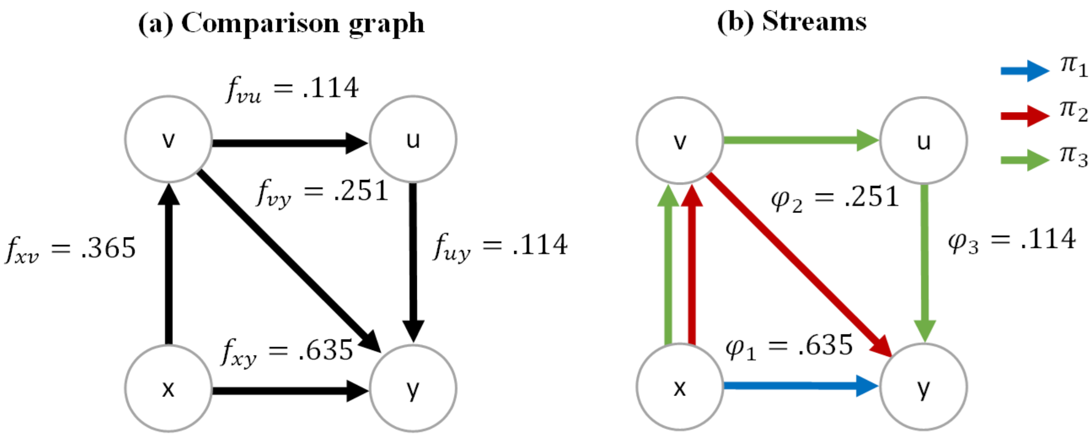


Figure 10. Decomposition of streams from the hypothetical network plot.

In this case, both path π_2 and π_3 with lengths $|\pi_i|$ of 2 and 3, flow through the direct evidence xv. In other words, this direct evidence contributes to both streams S_2 and S_3. Thus, the formula of p_{xv} can be expressed as follows:

$$p_{xv} = \frac{\varphi_2}{|\pi_2|} + \frac{\varphi_3}{|\pi_3|} = \frac{0.251}{2} + \frac{0.114}{3} = 0.164 = 16.4\%$$

The calculations of reamaining direct evidence are shown in **Figure 11** [16].

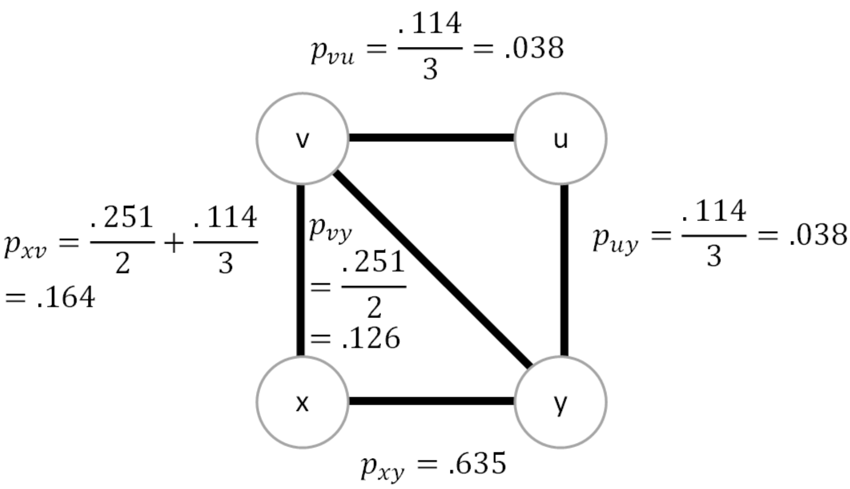


Figure 11. Proportion contributions of direct evidence.

The matrix P in **Table 2** shows the proportion contribution of each direct evidence to each NMA contrast estimate. The study further calculated the matrix P* consisting of the proportion contribution of direct evidence from individual trials to each NMA contrast estimate, as shown in **Table 3**. For simplification, the study considered that there are no multi-arm studies. A direct estimate from TMA is a weighted average effect size combining the results of multiple studies, where the weights refer to the amount of information they contribute, usually the inverse variance of each effect-size estimate [22]. Accordingly, the proportion contribution of direct evidence to a contrast estimate can be allocated to the studies contributing to the evidence based on their respective weight, which can be seen as the proportion contribution of direct evidence from the individual trial to a contrast estimate, denoted by p_i^*. The equation is expressed as follows:

$$p_i^* = \frac{w_i}{\sum_i^I w_i} \cdot p$$

$i = 1, \dots, I$ where I is the number of studies included in each direct evidence; w_i is the weight of individual study within the direct evidence, obtained from TMA; p is the proportion contribution of the direct evidence to a given contrast estimate.

For example, the proportion contribution of direct evidence xy to the contrast estimate xy, denoted by p_{xy}^{xy} (or p_{xy}), is 63.5%. The direct evidence xy is contributed by two studies with weights of 0.69 and 1.54 from TMA, respectively. Thus, the proportion contributions of the two studies to contrast estimate xy are 19.6% and 43.8%, respectively.

$$p_1^* = \frac{0.69}{0.69 + 1.54} \cdot 63.5\% = 19.6\%$$

$$p_2^* = \frac{1.54}{0.69 + 1.54} \cdot 63.5\% = 43.8\%$$

Table 3. Study proportion contribution matrix P* for the network of four interventions x, y, u, and v.

Study	1	2	3	4	5	6	7	8	9	10	11	12	13
xy	19.9	63.5	2.7	0.4	0.7	0.5	0.5	0.6	2.8	1.2	0.7	2.4	4.1
xu	9.5	40.5	8.7	4.3	5.4	3.6	5	5.3	7	6.4	3.8	0.3	0.2
xv	7.8	67.2	4.1	0.7	1.1	0.7	0.8	0.9	4.2	1.9	1.1	3.5	6.1
yu	0.5	4.73	13.5	10.4	12.8	8.5	12.1	12.8	12.3	4.2	2.5	2.1	3.6
yv	1.6	23.6	12	1.9	2.5	1.7	2.2	2.3	12.3	4.9	2.9	11.4	20.6
uv	0.8	8.5	18.9	3	3.9	2.6	3.5	3.7	14.5	19	11.4	3.7	6.6

3.2. Estimation of Network Robustness

This study modified and called several functions within the R package "flow contribution" published by Papakonstantinou et al. [23], including "getHatMatrix", "comparisonStreams", "getStudyContribution", and "getComparisonContribution". The variables, function arguments, and return values are described in **Appendix A**.

To obtain a more nuanced evaluation and visualization of the quality of NMA estimates, this study constructed a function of estimating the proportion contribution of direct evidence from each study in each evidence stream to the interested treatment contrast estimate, involved the further calculation of the matrix P and matrix P* (see **Appendix B**).

A set of evidence streams of a single contrast estimate, also known as a flow network, can be linked to electric circuits converted into a two-terminal series-parallel graph (TTSPG) with a source and a sink. In this study, a method has been developed to estimate the robustness of a network, where robustness is defined as the probability of current flow through the whole stream. In contrast, the risk of open circuits is determined by the RoB of the studies included in NMA.

Streams, edges, and direct evidence from an individual study can be compared to two-terminal components in electrical networks, which can be connected in series or parallel. An evidence stream network is composed of streams connected in parallel; a stream is composed of edges connected in series; an edge is composed of direct evidence from individual studies connected in parallel. Since NMA estimates are linear combinations of direct evidence, a biased study will impair the quality of the entire evidence flow, similar to how an open circuit can cause the entire circuit to fail. A high RoB study exists in an evidence stream makes the entire stream susceptible to bias. The influence of biased studies on the stream network varies depending on their proportion contribution to the network.

Therefore, the structure of an evidence streams network (flow network), the proportion contribution of studies, and their RoB were considered for the robustness index. The robustness index is defined as the probability of evidence successfully flowing through the stream without the occurring of open circuits. The probability that an open circuit occurs can be set to RoB levels of different risk values (low with 0.1, moderate with 0.5, and high with 0.9 in the following demonstration), representing the probability that an NMA contrast estimate is unlikely to be biased. Once the proportion contributions of studies were obtained through a further calculation of matrix P, a function for estimating the robustness of the flow network can be developed (**Appendix C**).

To illustrate the idea of the robustness index, firstly take two simple examples of computing the probability that the current passes through a series and parallel circuit. The example circuits were demonstrated in **Figure 12**, with the probability of failing to an open circuit for each component (denoted by p_i) assumed to be independent.

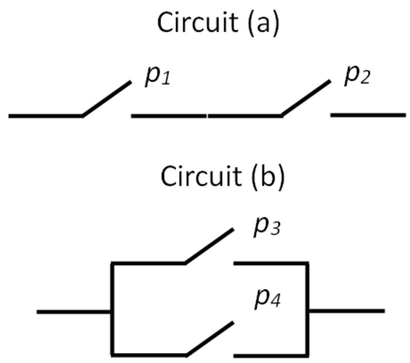


Figure 12. Series circuit (a), and parallel circuit (b).

For circuit (a), the probability that the current passes through can be expressed as $(1-p_1)(1-p_2)$ since only one opened component would fail the entire circuit. For circuit (b), the probability that the current passes through can be expressed as $1-p_3p_4$ since only one working component is required to support the entire circuit. Since the whole stream network can be seen as a series and parallel circuit, the robustness can be evaluated based on the probability that the current passes through.

It is noteworthy that this study, however, did not employ the above calculation when computing the probability that a parallel circuit works since it would overestimate the robustness by ignoring the high RoB evidence as long as any low RoB evidence exists regardless of how low its contribution might be. For parallel circuits, instead, the study computed the probability by summing up the product of contribution and RoB of each evidence, similar to the pooling of effects in TMA.

The algorithm to calculate the robustness index starts with the robustness of edges. Since an edge consists of direct evidence from individual studies, the equation of the robustness of an edge is expressed as follows:

$$R^e = \sum_{i=1}^{I} c_i(1-r_i)$$

$$c_i = \frac{p_i'}{\frac{\varphi}{|\pi|}}$$

In an edge, $i = 1, \dots, I$ where I is the number of direct evidence from individual studies; c_i is the proportion contribution of direct evidence from an individual study, with a value between 0 and 1; p_i' is the proportion contribution of direct evidence from an individual study in a given stream to the contrast estimate, derived from the matrix P and matrix P* (see **Appendix B**); φ is the flow of the stream; $|\pi|$ is the length (i.e., number of edges) of the stream; r_i is the RoB of the study, converted to the risk value between 0 and 1.

Next, the robustness of streams will be calculated. Since a stream consists of edges connected in series with the same contribution $\frac{\varphi}{|\pi|}$, the expression of the robustness of a stream is shown as follows:

$$R^s = \prod_{j}^{J} R_j^e$$

In a stream, $j = 1, \dots, J$ where J is the number of edges; R_j^e is the robustness of an edge, with a value between 0.1 and 0.9.

Then, the robustness of treatment contrasts will be calculated. Since a contrast can be decomposed to streams connected in parallel with the contribution φ_k, the expression of the robustness of a treatment contrast is shown as follows:

$$R^c = \sum_{k=1}^{K} \varphi_k \, R_k^s$$

In a contrast, $k = 1, \dots, K$ where K is the number of streams; φ_k is the proportion contribution of each stream; R_k^s is the robustness of a stream, with a value between 0 and 0.9.

Finally, the robustness of the entire network estimates R^n can be calculated, with a value between 0 and 1. Since a complete network consists of all NMA treatment contrasts estimates, the expression of the robustness of a network is shown as follows:

$$R^n = \frac{\sum_{l=1}^{L} R_l^c}{L}$$

In a network, $l = 1, \ldots, L$ where L is the number of treatment contrasts estimates; R_l^c is the robustness of a contrast estimate, with a value between 0 and 0.9.

3.3. Visualization

To present the above results for better supporting clinical decisions, several functions for graphing were developed to build an interactive user interface with the Shiny package. The modules of visualization include a robustness network plot of all NMA estimates, a robustness network plot containing streams of a single contrast estimate, and a bar chart of the proportion contributions of each study to a single contrast estimate. These modules were integrated into a Shiny app and deployed to Shinyapps.io (https://maliang.Shinyapps.io/NMA-Robustness-Visualization/) [24].

One dataset for the demonstration was based on a systematic review of the effect of different classes of antihypertensive drugs on incident diabetes mellitus up to 2006 by Elliot et al., which included 22 clinical trials (143,153 participants). There were six different drug therapy used in the trials [25], which is a well-connected network as shown **Figure 13**. The other dataset was based on a systematic review and NMA of the effectiveness of Calcimimetic agents for secondary Hyperparathyroidism in adults, with 36 trials (11,247 participants) included. The study compared three Calcimimetic agents and their combinations [26], which is a sparse network as shown in **Figure 14**. The R code of the Shiny application and demo datasets are available on GitHub (https://github.com/ShibaNekoL/NMA-Robustness-Visualization.git) [27].

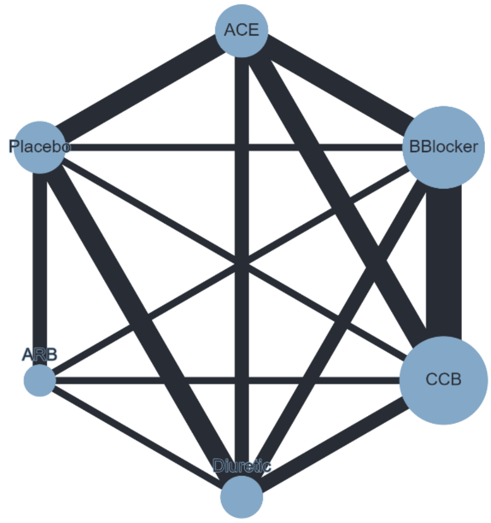


Figure 13. Network plot of the antihypertensive drugs on incident diabetes mellitus network.

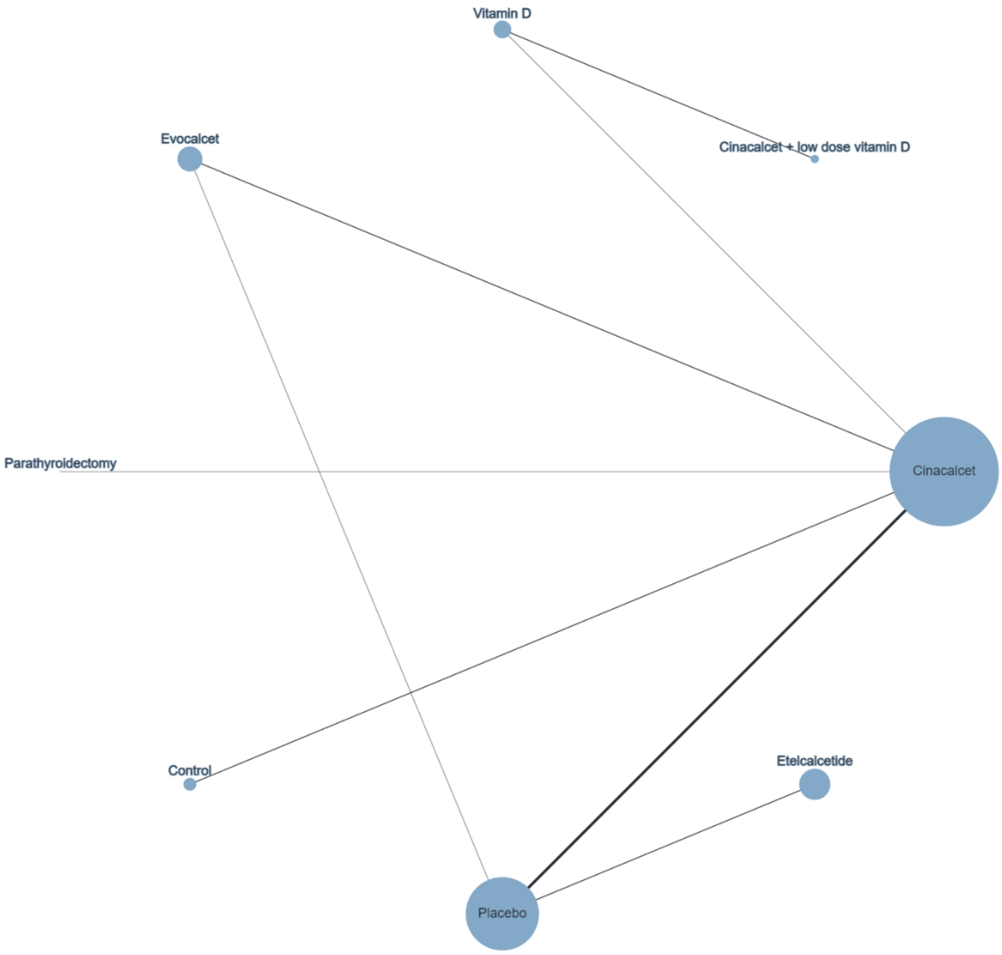


Figure 14. Network plot of Calcimimetic agents for secondary Hyperparathyroidism in adults.

4. Results

4.1. Robustness index

Demonstrated with the dataset of antihypertensive drugs on incident diabetes mellitus, the study estimated the robustness of each stream and overall robustness for the treatment contrasts “ACE versus placebo” and “ARB versus Diuretic”. The study compared the results of evaluating the quality of evidence by taking weighted averages of RoB and proportion contributions of studies with the robustness index considering flow networks.

The RoB was set to the same risk values between the two conditions, where high as 0.1, moderate as 0.5, and low as 0.9. The overall quality of the contrast estimate “ACE versus Placebo” evaluated by the robustness index was 0.76, whereas the estimate without considering the flow network was 0.85, as shown in **Table 4**. This indicates that evaluating the quality of evidence based merely on RoB and proportion contributions leads to an overestimation compared to the robustness index that takes into account the evidence flow network. Moreover, the more the length of an evidence stream (i.e., the number of edges) increases, the larger difference between the two evaluations becomes. The robustness index of a longer stream is significantly lower due to the property of calculating the probability of a serial circuit being open. Since the evidence stream in NMA is a linear combination of various direct evidence, if single direct evidence (edge) is biased, the quality of the entire stream will be impaired. Therefore, longer evidence streams including more edges are more susceptible to bias.

The quality of the contrast estimate “ACE versus Diuretic” evaluated by the robustness index was 0.48, whereas the estimate without considering the flow network was 0.74, as shown in **Table 5**. The disparity of estimates between the two methods was larger than that of the contrast estimate “ACE versus Placebo”. In “ARB versus Diuretic”, the streams with higher flow had lower robustness, thus the overall robustness was a lot lower than the overall weighted average. The low robustness of a stream was due to the high RoB direct evidence connected inside, which would be visualized in section **4.2.2**.

Demonstrated with the other dataset of Calcimimetic agents for secondary Hyperparathyroidism in adults, the study estimated the quality of contrast estimate “Cinacalcet versus Control” and compared it with the weighted average of RoB and Proportion Contributions, as shown in **Table 6**. The contrast estimate consists of one stream with the length of one direct evidence, hence the evaluation of the weighted average and the robustness index were the same.

Table 4. Comparison between evaluations of quality of contrast estimate “ACE versus Placebo” by weighted average and the robustness index.

Path of evidence streams of ACE versus Placebo contrast estimate	Flow	Weighted Average	Robustness Index
ACE:Placebo	0.59	0.90	0.90
ACE:Diuretic:Placebo	0.14	0.86	0.75
ACE:CCB:Placebo	0.10	0.82	0.66
ACE:BBlocker:Placebo	0.04	0.74	0.51
ACE:BBlocker:CCB:Diuretic:Placebo	0.01	0.86	0.47
ACE:BBlocker:ARB:Placebo	0.06	0.73	0.37
ACE:BBlocker:CCB:Diuretic:ARB:Placebo	0.00	0.84	0.37
ACE:CCB:ARB:Placebo	0.03	0.65	0.26
ACE:BBlocker:CCB:ARB:Placebo	0.03	0.66	0.17
ACE:BBlocker:Diuretic:Placebo	0.01	0.58	0.17
Overall evaluation	1.00	0.85	0.76

Table 5. **Comparison between evaluations of the quality of contrast estimate "ARB versus Diuretic" by weighted average and the robustness index.**

Path of evidence streams of ARB versus Diuretic contrast estimate	**Flow**	**Weighted Average**	**Robustness Index**
ARB:Diuretic	0.01	0.90	0.90
ARB:Placebo:Diuretic	0.25	0.78	0.60
ARB:Placebo:ACE:Diuretic	0.13	0.84	0.59
ARB:BBlocker:CCB:ACE:Diuretic	0.02	0.85	0.51
ARB:BBlocker:ACE:Diuretic	0.08	0.79	0.46
ARB:CCB:Diuretic	0.31	0.70	0.45
ARB:Placebo:CCB:ACE:Diuretic	0.01	0.82	0.43
ARB:Placebo:BBlocker:CCB:ACE:Diuretic	0.00	0.82	0.37
ARB:CCB:ACE:Diuretic	0.02	0.71	0.33
ARB:BBlocker:Diuretic	0.17	0.63	0.32
Overall evaluation	1.00	0.74	0.48

Table 6. Comparison between evaluations of the quality of contrast estimate "Cinacalcet versus Control" by weighted average and the robustness index.

Path of evidence streams of Cinacalcet versus Control contrast estimate	**Flow**	**Weighted Average**	**Robustness Index**
Cinacalcet:Control	1	0.23	0.23
Overall evaluation	1	0.23	0.23

4.2. Shiny App, an Interactive User Interface

The modules for visualizing the results of network robustness were integrated into a Shiny app and deployed to Shinyapps.io (https://maliang.Shinyapps.io/NMA-Robustness-Visualization/) [24].

4.2.1. Visualization of Network Robustness

The interactive user interface (UI) of the robustness network with antihypertensive drugs on incident diabetes mellitus was shown in **Figure 16**. The nodes represent interventions; the solid lines represent direct evidence or mixed evidence, while the dotted edges represent indirect evidence. The width of the edges is proportional to the robustness index, and the color also depends on the robustness index, from red, yellow to green. The values of robustness are displayed in the table in the lower part of the UI. The UI allows for the adjustment of parameters including color display, node size, edge width, and filter. Besides, it displays a prompt window when hovering over any element on the network to show detailed information, as shown in **Figure 18**. When hovering over a contrast (edge), the prompt window shows the robustness index, mean length, and the number of streams. When hovering over a treatment (node), the prompt window shows the sample size and the number of studies. Multiple contrasts can be selected for further analysis by either clicking on any rows in the table or the edges of the network plot, as shown in **Figure 19**.

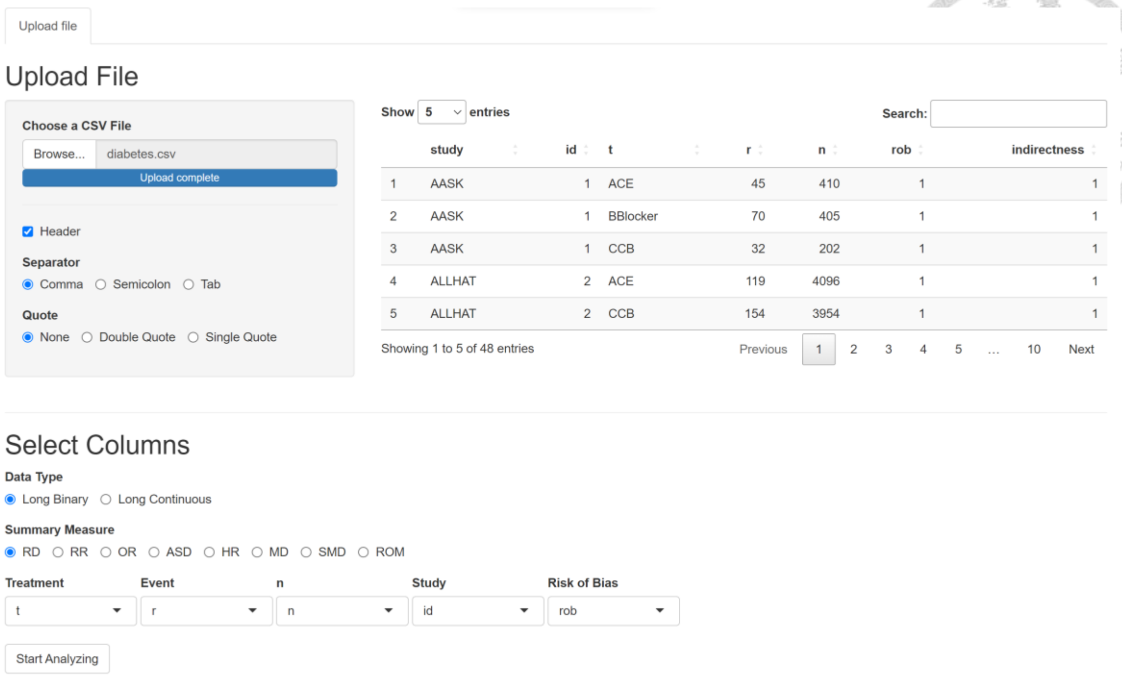


Figure 15. The first tab "Upload file" of the Shiny app "Network Robustness Visualization".

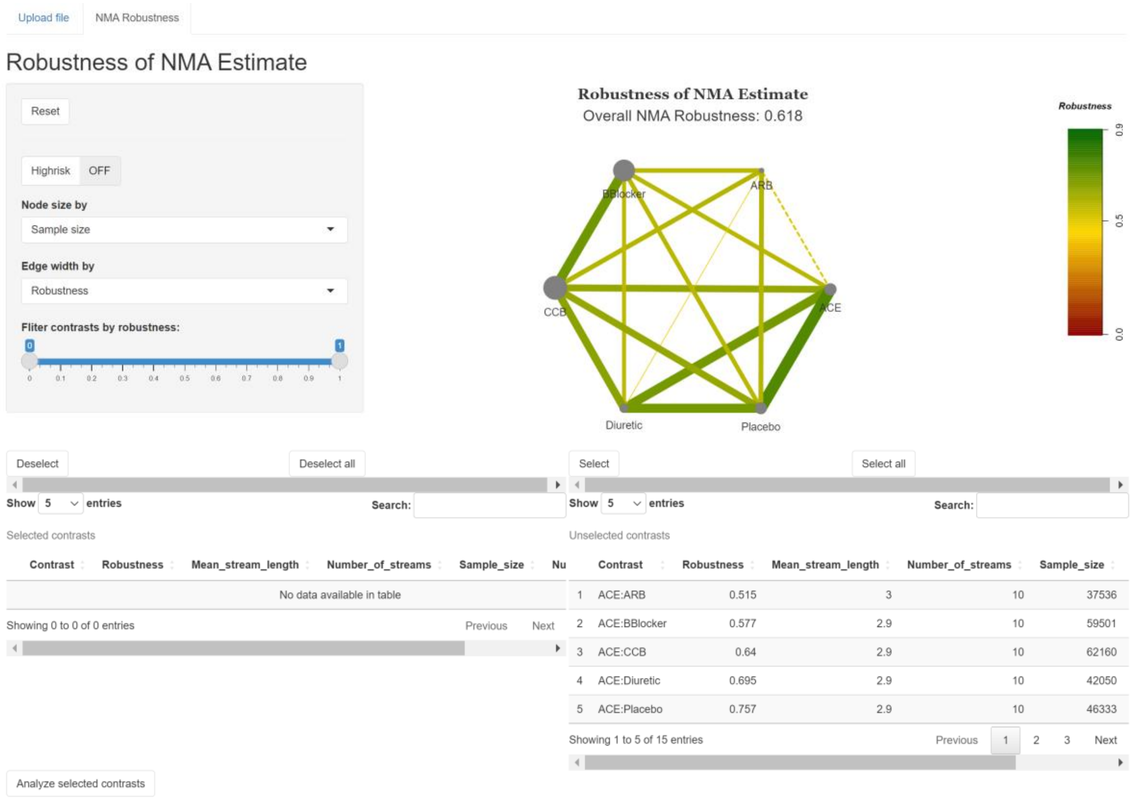


Figure 16. Module network robustness in the second tab "NMA Robustness" of the Shiny app "Network Robustness Visualization".

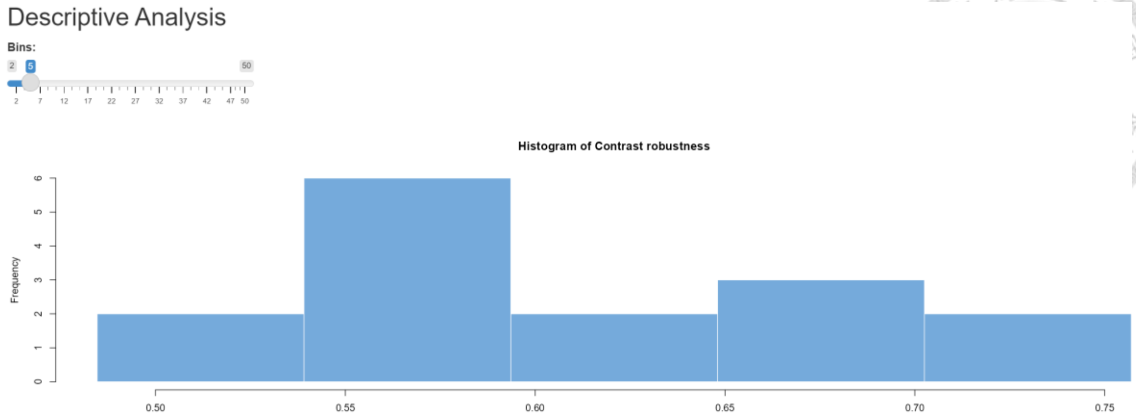


Figure 17. Histogram of contrast robustness in the second tab “NMA Robustness” of the Shiny app “Network Robustness Visualization”.

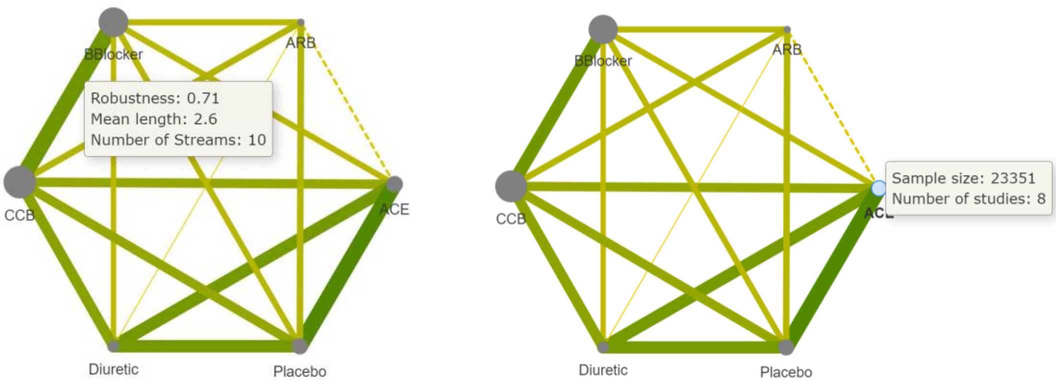
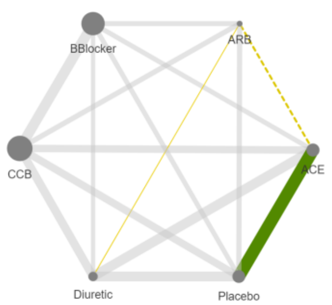


Figure 18. Demonstration of hovering over the elements in network robustness plot.



Selected contrasts

	Contrast	Robustness	Mean_stream_length	Number_of_streams	Sample_size
1	ACE:ARB	0.515	3	10	37536
5	ACE:Placebo	0.757	2.9	10	46333
8	ARB:Diuretic	0.485	2.9	10	32884

Figure 19. Demonstration of selecting multiple contrasts in network robustness plot.

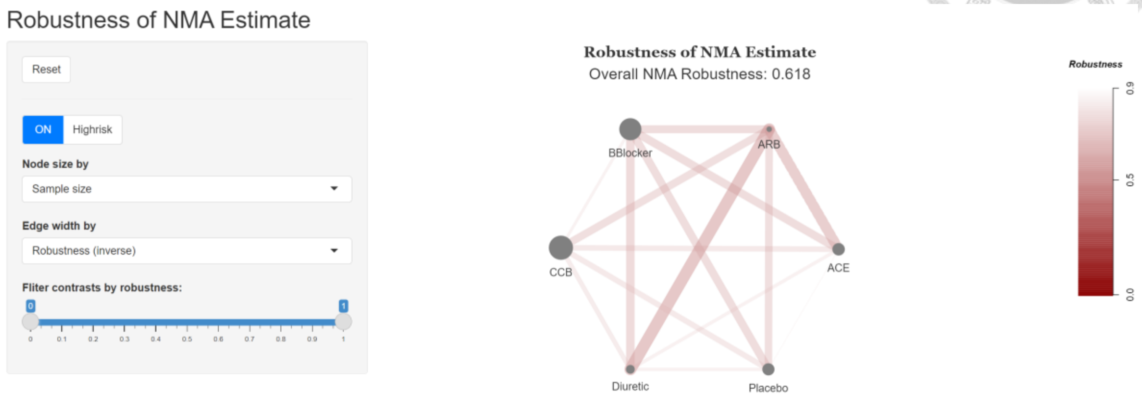


Figure 20. Network robustness plot with low robustness contrasts highlighted.

This UI allows users to comprehensively evaluate the quality of evidence in NMA. In the demonstration shown in **Figure 16**, the overall robustness of the network was 0.618, which indicated that the quality of evidence was acceptable since there was no contrast estimate (edge) with extremely low robustness (in red). There was still low-quality evidence such as “ARB versus Diuretic” and “ARB versus ACE”, though. The latter comparison consisted of only indirect evidence and was therefore presented as a dotted line. It is important to cautiously interpret the results of these comparisons in clinical decision-making. The histogram of the robustness of comparisons shown in **Figure 17** indicated that attention should be paid to the peak of the number of evidence with lower robustness on the left. The network robustness plot can also be switched to the display that highlights the contrasts with low robustness, indicating which evidence has a high risk to be biased, as shown in **Figure 20**. For example, the contrast estimate “ARB versus Diuretic” with darken color was the one required to be noticed.

The other demonstration used the dataset of the effects of Calcimimetic agents for secondary Hyperparathyroidism. The sparse network shown in **Figure 21** with an overall robustness of 0.054 consisted of several dotted lines in red, representing the low robustness of indirect evidence. In addition, the distribution of contrasts robustness was right-skewed, as shown in **Figure 22**; hence, the quality of the NMA estimate of this dataset was inferior.

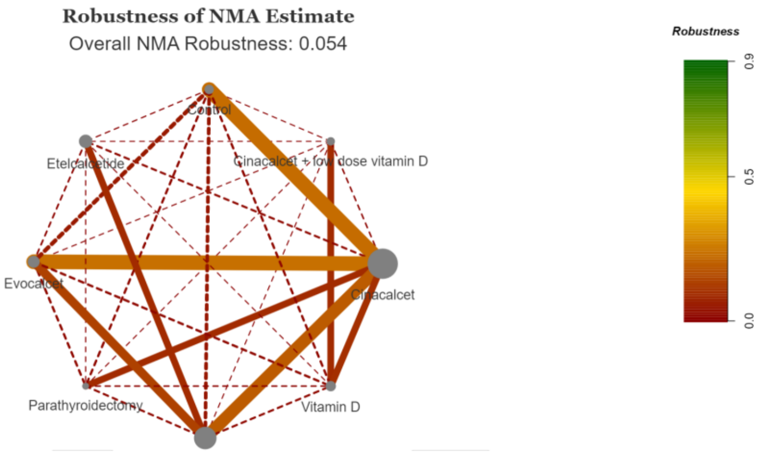


Figure 21. Network robustness with the dataset of the effects of Calcimimetic agents for secondary Hyperparathyroidism.

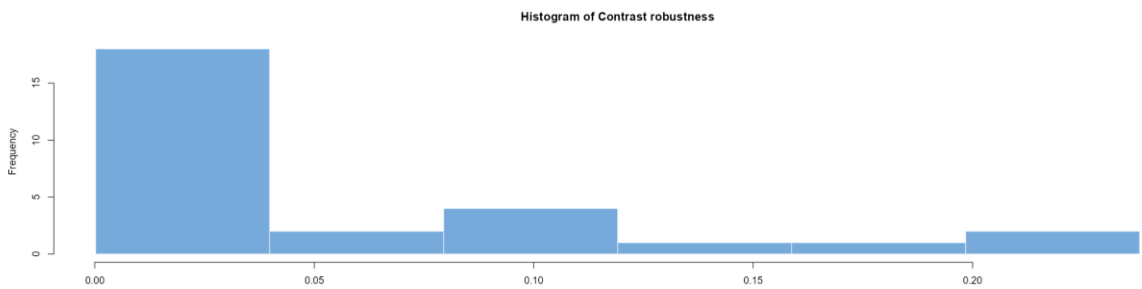


Figure 22. Histogram of contrast robustness with the dataset of the effects of Calcimimetic agents for secondary Hyperparathyroidism.

This module provides a comprehensive evaluation of the overall quality of NMA and the identification of low-quality contrast estimates. To examine the composition of the quality of evidence within a single contrast, it is essential to visualize the flow network shown in the following demonstration.

4.2.2. Visualization of Flow Network and Robustness

This UI module provides users with a further understanding of the structure of a stream network, flow, and robustness of a single NMA contrast estimate. The demonstration visualized the flow network and robustness of the contrast “ACE versus Placebo”, as shown in **Figure 23**. The nodes represent treatments, and the lines in the same color represent an evidence stream. The robustness of streams is displayed as transparency, where lower transparency (darker color) indicates greater robustness. The details of streams are shown in the table below the network plot. The UI allows users to adjust parameters including the size of nodes, the width of edges, filters, and colors. It also provides a hover-over prompt window and a select-to-highlight feature, as shown in **Figure 24**.

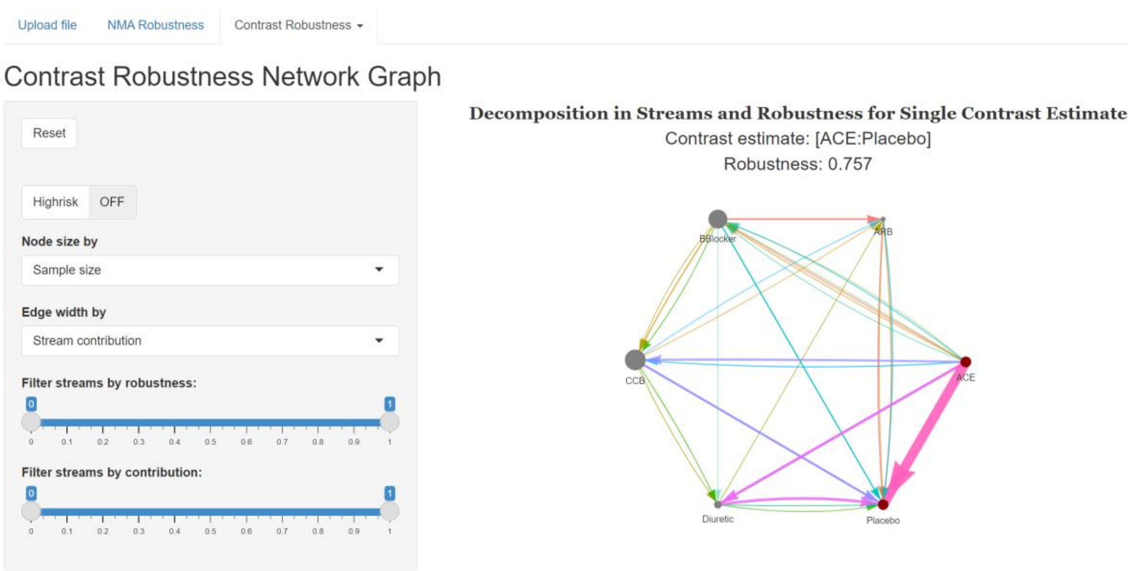


Figure 23. Module stream network and robustness plot in The third tab "Contrast Robustness" of the Shiny app "Network Robustness Visualization".

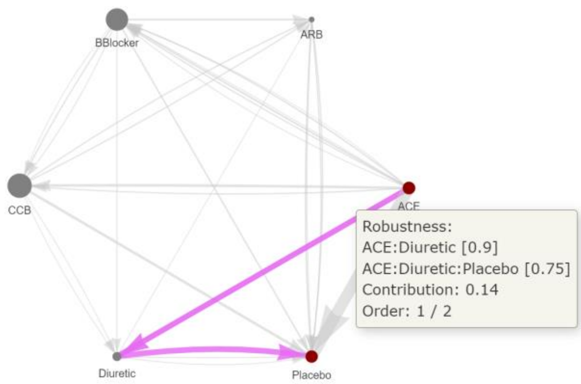


Figure 24. Demonstration of hovering over and selecting a stream in the flow network plot.

It is important to note that streams with thicker edges and lighter colors, indicating high flow but low robustness, can have an adverse impact on the overall evidence quality. In the example shown in **Figure 23**, the robustness of the comparison "ACE versus Placebo" is 0.757, indicating decent robustness since evidence streams with high flow, such as "ACE-Placebo", did not have low robustness. **Figure 25** illustrates different colors to represent robustness, whereas darker colors indicate lower robustness. This enables easier identification of high-risk and susceptible streams in the flow network.

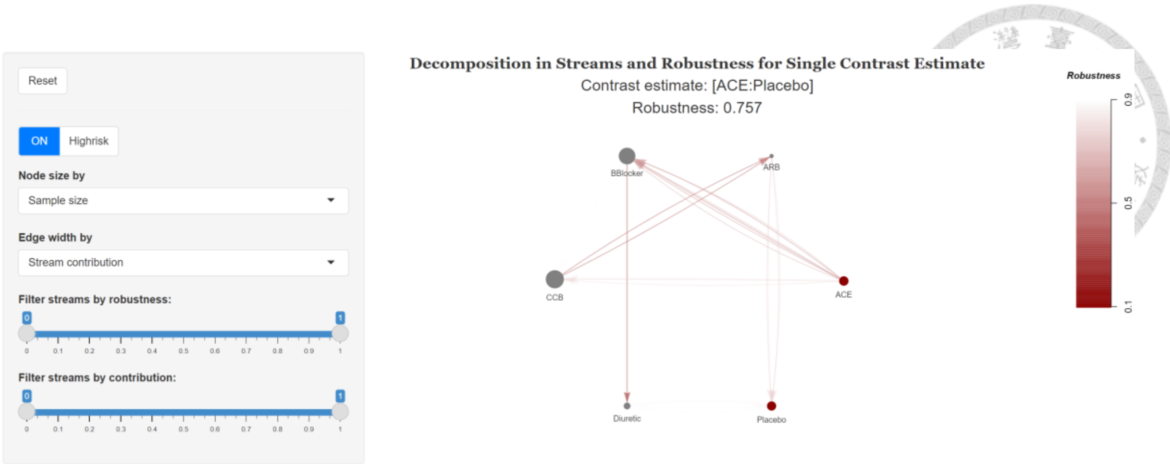


Figure 25. Flow network plot with low robustness streams highlighted.

The other demonstration was the stream network of the contrast estimate "ARB versus Diuretic" which had the lowest robustness of 0.485 in the network of antihypertensive drugs, as shown in **Figure 26**. The streams with the higher flow were "ARB-CCB-Diuretic", "ARB-Placebo-Diuretic", and "ARB-BBlocker-Diuretic", as shown in the left panel. The direct evidence "ARB-CCB", "ARB-Placebo", and "BBlocker-Diuretic", the lower robustness edges as shown in the right panel, were in the higher flow streams. Any biased edge would cause the entire stream to be biased, thus these streams with higher flow had low robustness.

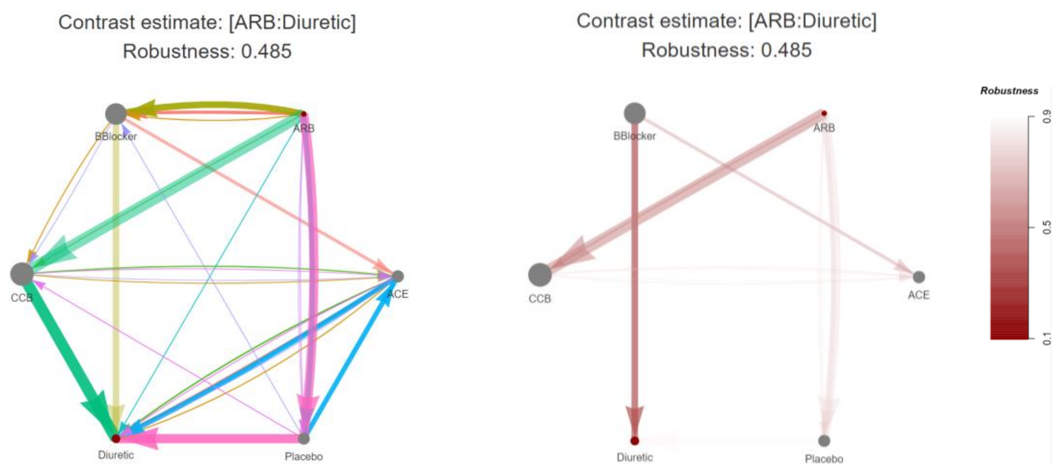


Figure 26. Streams network of the contrast estimate "ARB versus Diuretic".

The other example using the dataset of the effects of Calcimimetic agents for secondary Hyperparathyroidism showed the streams network of contrast estimate "Evocalcet versus Placebo" in **Figure 27**. This comparison was only contributed by two streams. In the left panel, the thicker stream has a lighter color, indicating that it had a high flow but low robustness. Streams with high flow and low robustness have a negative impact on NMA contrast estimates, hence they are essential to be identified. To gain a nuanced understanding of why a particular evidence stream has low robustness, it is essential to visualize the proportion contribution and RoB of each evidence from individual studies, which will be shown in the following demonstration.

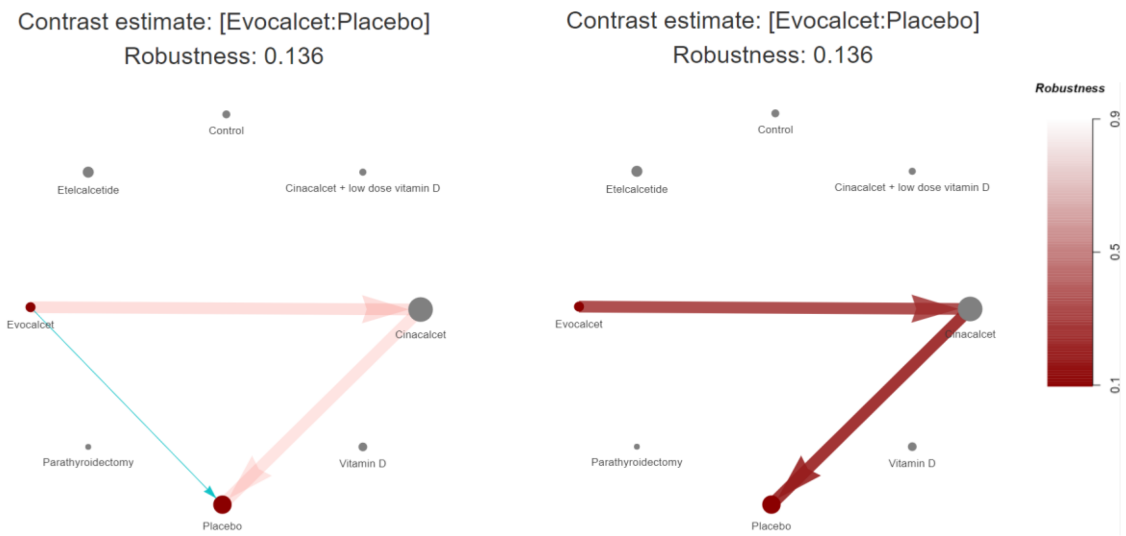


Figure 27. Flow network plot with the dataset of the effects of Calcimimetic agents for secondary Hyperparathyroidism.

4.2.3. Visualization of Streams RoB and Proportion Contributions of Studies

The bar chart consisting of the proportion contributions of each study and their RoB to the contrast estimate “ACE versus Placebo” is shown in **Figure 28**. The vertical axis represents each evidence stream; the horizontal axis represents the proportion contributions of each study. The black diamonds represent treatments (i.e., nodes in the network plot); the white lines separate the different studies. The RoB of each study is presented as a high, moderate, or low risk in red, yellow, or green, respectively. The UI allows for zooming, and hovering over a bar displays information including study ID and proportion contrition, as shown in **Figure 29**. Clicking on a bar highlights the corresponding study, and multi-arms studies may occur multiple times.

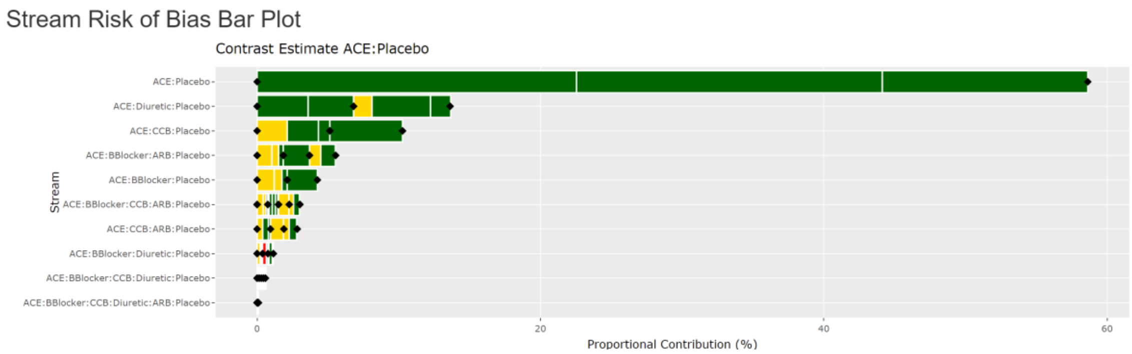


Figure 28. Module stream RoB bar chart in the third tab “Contrast Robustness” of the Shiny app “Network Robustness Visualization”.

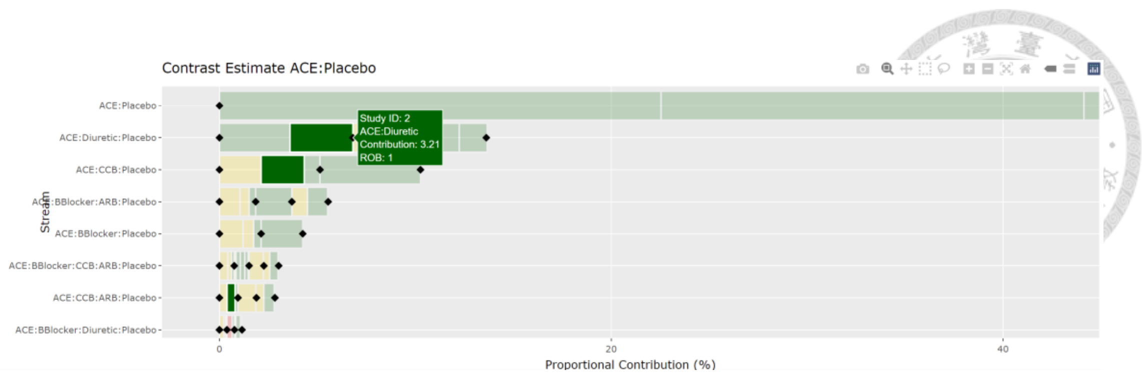


Figure 29. Demonstration of selecting a study in the stream RoB bar chart.

Another example revealed the streams and proportion contributions of each study and their RoB for the contrast estimate "ARB versus Diuretic" which had the lowest robustness in the network of antihypertensive drugs, as shown in **Figure 30**. The top three high-flow streams "ARB-CCB-Diuretic", "ARB-Placebo-Diuretic", and "ARB-BBlocker-Diuretic" contained a large proportion of moderate and high RoB studies. The large red bar on the second edge of the stream "ARB-BBlocker-Diuretic" indicated the low robustness of the stream.

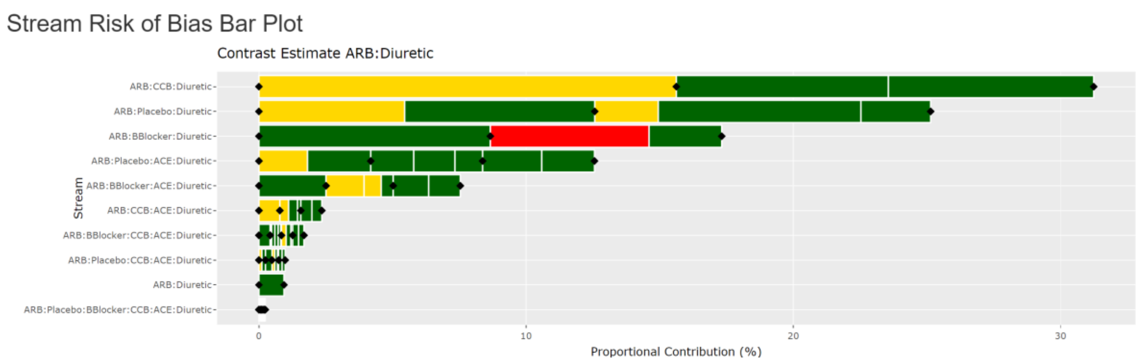


Figure 30. Stream RoB bar chart of contrast "ARB versus Diuretic"

The other demonstration for the contrast "Evocalet versus Placebo" with the dataset of the effects of Calcimimetic agents for secondary Hyperparathyroidism showed two streams, where the one with a larger flow consists mostly of studies with high RoB and contribution, as shown in **Figure 31**. This caused low overall robustness of the contrast estimate.

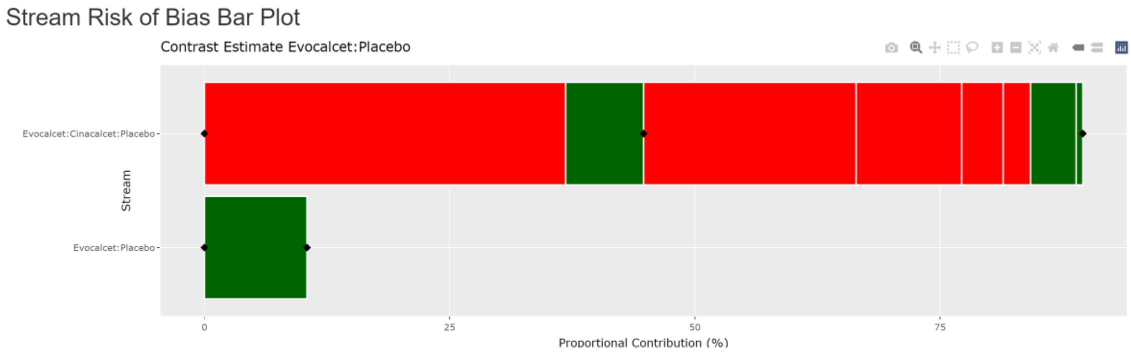


Figure 31. Stream RoB bar chart with the dataset of the effects of Calcimimetic agents for secondary Hyperparathyroidism.

This visualization module presents not only the number, length, and flow of evidence streams but also the distribution of the proportion contribution and RoB of studies within each stream. The distribution of RoB supports clinical decision-making at the level of individual study. For example, both contrast estimate consists of multiple streams with even contributions or a few with high and the concentration or dispersion of high RoB evidence at the edges within a stream affect the assessment of robustness. This information can be used as a reference for removing high RoB studies.

5. Discussions

In this study, we derive a robustness index for quantitatively assessing the quality of evidence in NMA, based on the proportion contribution and RoB of individual studies. Instead of taking a weighted average, this study considered the ideas of evidence flow networks that can be compared to two-terminal series-parallel graphs.

An approach for rating the quality of NMA treatment effect estimates has been developed by Puhan et al. based on GRADE [6]. This approach considers within-study bias in the most dominant one-step loop for each contrast estimate, thus neglecting substantial amounts of information, which hinders its practical application in large networks. Although the other approach developed by Salanti et al. [7] considered the weights of studies, the weighted average of RoB was not a quantitative index for assessing the impact of bias in individual studies. Besides, the weighted average did not consider the evidence flow network structure, disregarding the impact of the distribution of bias in studies in the network. In contrast to these approaches, the robustness index relies on the proportion contribution and takes the impact of every study included in the network into account, increasing the accuracy in evaluating the quality of evidence.

To derive the robustness index, this study compared the evidence flow networks to two-terminal series-parallel graph (TTSPG) circuits, the RoB of an individual study to the risk of an open circuit, and the robustness index to the probability of a functioning circuit. A functioning circuit can be decomposed into compositions connected either in series or parallel. The probability of a functioning series composition is the multiplication of the probability of each component working, where the component represents the direct evidence from an individual study. The characteristic of this multiplication method of a stream is that it is more conservative for assessing the quality when the true size of bias is unknown. Since the effect estimate of a stream is a linear combination of the effect estimate of each edge connected in series, the size of bias of the effect estimate of a stream can be accumulated by each edge bias. In other words, the existing within-study bias can flow through the entire stream. The impact of a single edge bias on a stream remains the same regardless of the length of a stream, which cannot be obtained by the weighted average of RoB and Proportion contributions.

For example, three streams were decomposed from the contrast "x versus y" in a hypothetical network comparing x, y, u, and v, as shown in **Figure 32**. The effect estimates of each direct evidence were shown along the edges. The true effect size of the contrast "x versus y" was 2, while the effect estimates of stream φ_2 and φ_3 are biased to 4 due to the biased direct evidence xy. Thus φ_2 and φ_3 are high RoB streams for the bias of 100% higher than the true value. The biased effect estimate of 4 of the direct evidence xy was also 100% higher than the true effect size of 2, so the direct evidence was also high RoB. The RoB was set to different risk values (low with 0.1, moderate with 0.5, and high with 0.9) in the following estimation. Since the direct evidence xy was high RoB, the quality of it was set to 1-0.9-0.1, where 0.9 corresponds to the risk value set for high RoB. The desired quality of the stream φ_2 and φ_3 would also be 1-0.9=0.1 since they were also high RoB.

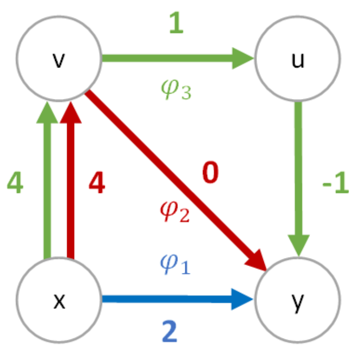


Figure 32. A hypothetical streams network for comparing the weighted average method and the robustness index.

The weighted average method for assessing the quality of streams is expressed as follows:

$$Quality\ of\ \varphi_2 = 0.1 \cdot \frac{1}{2} + 0.9 \cdot \frac{1}{2} = 0.5$$

$$Quality\ of\ \varphi_3 = 0.1 \cdot \frac{1}{3} + 0.9 \cdot \frac{1}{3} + 0.9 \cdot \frac{1}{3} = 0.6\overline{3}$$

The robustness index for streams is expressed as follows:

$$Robustness\ of\ \varphi_2 = 0.1 \cdot 0.9 = 0.09$$

$$Robustness\ of\ \varphi_3 = 0.1 \cdot 0.9 \cdot 0.9 = 0.081$$

Instead of being overestimated by the weighted average method, the quality estimated by the robustness index was close to the desired value of 0.1. Besides, when the length of streams increased, the impact of the bias of the direct evidence xy should not be decreased. The weighted average method, however, weakened this impact with a raising of quality estimation from 0.5 to 0.63.

The actual impact of a single edge bias on a stream depends on the effect size of the other edges, thus this impact might be weakened by a large effect size from the other edges. Since the true effect size of each edge is unknown, the study used the multiplication method for evaluating the quality of a stream to obtain a more conservative assessment.

One underlying assumption for the multiplication method is that the RoB of direct evidence from individual studies is independent, disregarding the positive correlation between different trials induced by multi-arms studies. This may lead to an underestimation of the robustness.

Assuming that the RoB level was an interval scale, the study converted RoB to the risk value of 0.1, 0.5, and 0.9 in the previous demonstration. The different risk value settings were compared for the stream network of the contrast estimate "ACE versus Placebo", as shown in **Figure 33**. The patterns of the robustness index among different risk values were the same, while the ranges of different settings were different. There were three reasons for the setting of 0.1, 0.5, and 0.9 in the demonstration. First, the study aimed to differentiate the quality of evidence, thus the settings with larger variation were preferred. Second, if high RoB was set to 0, the robustness of a stream circuit would be 0 when any edge with a robustness of 0 existed, which was unrealistic. The study set high RoB to 0.1 to avoid the situation. Finally, the study set low RoB to 0.9 instead of 1 based on the concept that the long path will be more vulnerable.

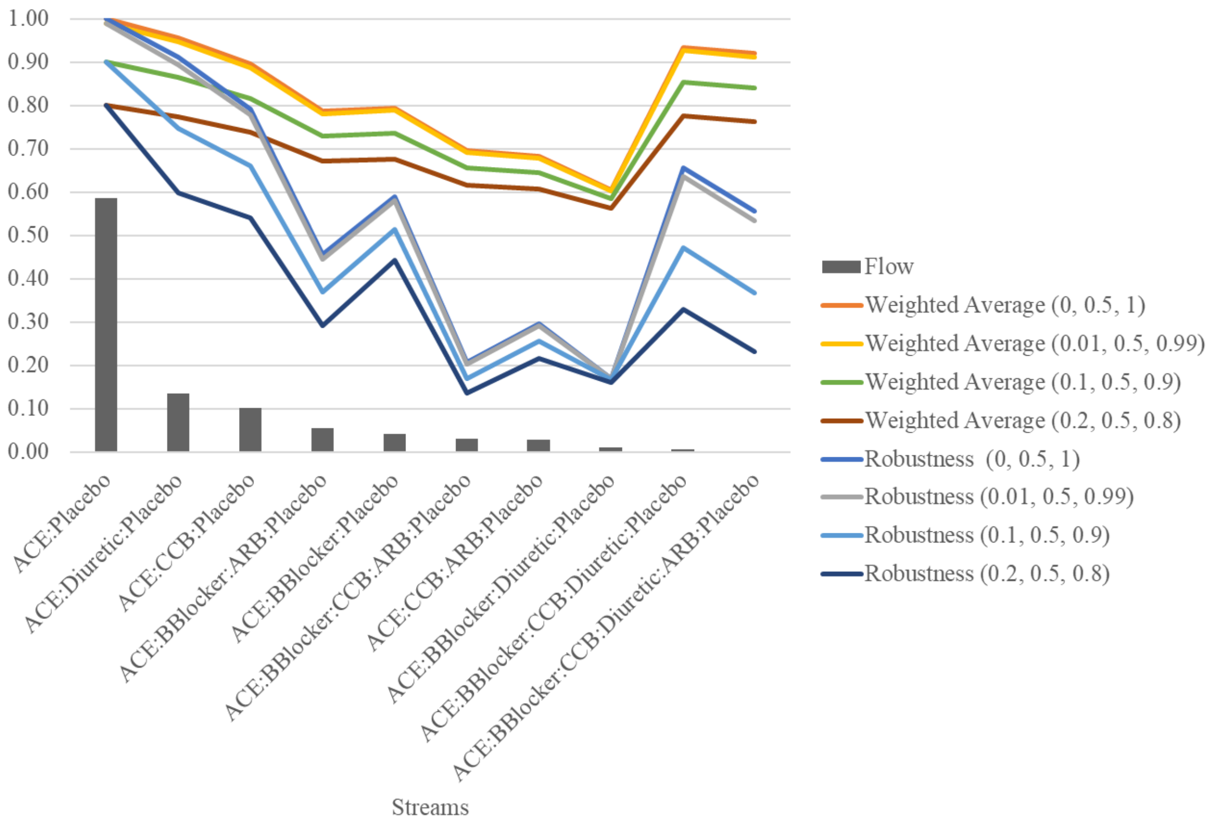


Figure 33. Comparison of different risk values of RoB.

A Shiny app, an interactive user interface, was developed in this study to visualize the flow network and the results of the robustness assessment. With the new visualization techniques in this app, users can quickly assess the robustness of the entire network, spot low-robustness contrast estimates, and gain a deeper understanding of the contribution and robustness of streams in the evidence flow network as well as the distribution of RoB and contribution of direct evidence from individual studies. In contrast to the CINeMA approach for assessing and visualizing confidence of NMA [9], the app provides a more nuanced reveal of the quality of evidence by considering how evidence streams flow in a network, how direct evidence is connected in a stream, and what their proportion contribution and RoB are.

The results of this study provide an assessment of the quality of evidence from NMA, which informs inferences about the effect estimates of interventions. Future investigations are necessary to validate the practical interpretation of the size of the robustness index and the use of comparison between different networks.

6. Conclusions

This study proposed a robustness index for quantifying the quality of evidence in NMA by combining proportion contributions and RoB of individual studies, based on the concept of flow networks that can be compared to two-terminal series-parallel graph circuits. Moreover, a Shiny app with multiple new visualization techniques was developed to present the results of the robustness assessment. The robustness index improves the accuracy of the assessment of NMA quality by considering the structure of evidence flow networks. The Shiny app facilitates the interpretation of the quality of NMA evidence and informs clinical decision-making about medical interventions. Future research is needed to investigate the interpretation of the size of the robustness index with more datasets.

References

1. Uman, L. S., *Systematic reviews and meta-analyses.* Journal of the Canadian Academy of Child and Adolescent Psychiatry, 2011. **20**(1): p. 57.
2. Tu, Y.-K., *Use of generalized linear mixed models for network meta-analysis.* Medical Decision Making, 2014. **34**(7): p. 911-918.
3. White, I. R., *Network meta-analysis.* The Stata Journal, 2015. **15**(4): p. 951-985.
4. Greco, T., Biondi-Zoccai, G., Saleh, O., Pasin, L., Cabrini, L., Zangrillo, A., and Landoni, G., *The attractiveness of network meta-analysis: a comprehensive systematic and narrative review.* Heart, lung and vessels, 2015. **7**(2): p. 133–142.
5. Cipriani, A., Higgins, J. P., Geddes, J. R., and Salanti, G., *Conceptual and technical challenges in network meta-analysis.* Annals of internal medicine, 2013. **159**(2): p. 130-137.
6. Puhan, M. A., Schunemann, H. J., Murad, M. H., Li, T., Brignardello-Petersen, R., Singh, J. A., Kessels, A. G., and Guyatt, G. H., *A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis.* BMJ, 2014. **349**: p. g5630.
7. Salanti, G., Del Giovane, C., Chaimani, A., Caldwell, D. M., and Higgins, J. P., *Evaluating the quality of evidence from a network meta-analysis.* PloS one, 2014. **9**(7): p. e99682.
8. Brignardello-Petersen, R., Bonner, A., Alexander, P. E., Siemieniuk, R. A., Furukawa, T. A., Rochwerg, B., Hazlewood, G. S., Alhazzani, W., Mustafa, R. A., and Murad, M. H., *Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis.* Journal of clinical epidemiology, 2018. **93**: p. 36-44.

9. Nikolakopoulou, A., Higgins, J. P., Papakonstantinou, T., Chaimani, A., Del Giovane, C., Egger, M., and Salanti, G., *CINeMA: an approach for assessing confidence in the results of a network meta-analysis.* PLoS medicine, 2020. **17**(4): p. e1003082.

10. Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J., Schulz, K. F., Weeks, L., and Sterne, J. A. C., *The Cochrane Collaboration's tool for assessing risk of bias in randomised trials.* BMJ, 2011. **343**: p. d5928.

11. Furukawa, T. A., Miura, T., Chaimani, A., Leucht, S., Cipriani, A., Noma, H., Mitsuyasu, H., Kanba, S., and Salanti, G., *Using the contribution matrix to evaluate complex study limitations in a network meta-analysis: a case study of bipolar maintenance pharmacotherapy review.* BMC research notes, 2016. **9**(1): p. 1-13.

12. Yao, L., Guyatt, G., and Brignardello-Petersen, R., *Considering the weight of loops for assessing the certainty of indirect evidence of network meta-analysis.* Advances in Evidence Synthesis: special issue Cochrane Database of Systematic Reviews, 2020. **9**.

13. Chaimani, A., Higgins, J. P., Mavridis, D., Spyridonos, P., and Salanti, G., *Graphical tools for network meta-analysis in STATA.* PloS one, 2013. **8**(10): p. e76654.

14. König, J., Krahn, U., and Binder, H., *Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons.* Statistics in medicine, 2013. **32**(30): p. 5414-5429.

15. Rücker, G., *Network meta-analysis, electrical networks and graph theory.* Research synthesis methods, 2012. **3**(4): p. 312-324.

16. Papakonstantinou, T., Nikolakopoulou, A., Rücker, G., Chaimani, A., Schwarzer, G., Egger, M., and Salanti, G., *Estimating the contribution of studies in network meta-analysis: paths, flows and streams.* F1000Research, 2018. **7**: p. 610.
17. *CINeMA*. Available from: https://cinema.ispm.unibe.ch/.
18. Lomnicki, Z., *Two-terminal series-parallel networks.* Advances in Applied Probability, 1972. **4**(1): p. 109-150.
19. Sadler, J. *Introduction to Network Analysis with R*. 2017; Available from: https://www.jessesadler.com/post/network-analysis-with-r/.
20. Ognyanova, K. *Network visualization with R*. 2021; Available from: www.kateto.net/network-visualization.
21. Krahn, U., Binder, H., and König, J., *A graphical tool for locating inconsistency in network meta-analyses.* BMC medical research methodology, 2013. **13**(1): p. 1-18.
22. Sánchez-Meca, J. and Marin-Martinez, F., *Weighting by inverse variance or by sample size in meta-analysis: A simulation study.* Educational and Psychological Measurement, 1998. **58**(2): p. 211-220.
23. Papakonstantinou, T. *Flow contribution R package*. 2020; Available from: https://github.com/esm-ispm-unibe-ch/flow_contribution.
24. Liang, M. F. *Shinyapps of NMA Robustness Visualization*. 2023 [cited 2023; Available from: https://maliang.shinyapps.io/NMA-Robustness-Visualization/.
25. Elliott, W. J. and Meyer, P. M., *Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis.* The Lancet, 2007. **369**(9557): p. 201-207.
26. Palmer, S. C., Mavridis, D., Johnson, D. W., Tonelli, M., Ruospo, M., and Strippoli, G. F., *Comparative effectiveness of calcimimetic agents for secondary hyperparathyroidism in adults: a systematic review and network meta-analysis.* American Journal of Kidney Diseases, 2020. **76**(3): p. 321-330.

27. Liang, M. F. *NMA Robustness Visualization*. 2022 [cited 2023; Available from: https://github.com/ShibaNekoL/NMA-Robustness-Visualization.git.

Appendix A

R Function Documentation of the Revised “flow_contribution” Package

1. \`getHatMatrix\` (path: “flow_contribution/R/hatmatrix.R”)

   1.1. Arguments

   1.1.1. \`indata\`: A csv file, the dataset for NMA.

   1.1.2. \`type\`: A character string indicating the data type of \`indata\`, either “long_binary” or “long_continuous”.

   1.1.3. \`model\`: The type of NMA model, including “fixed” (default) for fixed effect model, “random” for random effect model.

   1.1.4. \`tau\`: Square-root of between-study variance (additive model). \`NA\` by default.

   1.1.5. \`sm\`: A character string indicating underlying summary measure, e.g., "RD", "RR", "OR", "ASD", "HR", "MD", "SMD", or "ROM".

   1.2. Returns: A list containing contribution matrix of direct evidence to each NMA treatment contrast estimate.

2. \`comparisonStreams\` (path: “flow_contribution/R/streamStatistics2.R”)

   2.1. Arguments

   2.1.1. \`hatmatrix\`: A list returned from \`getHatMatrix\` function.

   2.1.2. \`comparison\`: A character string indicating the treatment comparison user interested in, with format “treatment A:treatment B”.

2.2. Returns: Paths and flows of streams in a single NMA treatment contrast estimate.

3. getStudyContribution (path: “flow_contribution/R/studycontribution.R”)

3.1. Arguments:

3.1.1. `hatmatrix`: A list returned from `getHatMatrix` function.

3.1.2. `comparison`: A character string indicating the treatment comparison user interested in, with format “treatment A:treatment B”.

3.2. Returns: Proportion contributions of direct evidence from individual studies to a single NMA treatment contrast estimate.

4. getComparisonContribution (path: “flow_contribution/R/contributionrow.R”)

4.1. Arguments:

4.1.1. `hatmatrix`: A list returned from `getHatMatrix` function.

4.1.2. `comparison`: A character string indicating the treatment comparison user interested in, with format “treatment A:treatment B”.

4.2. Returns: Proportion contribution of each direct evidence to a single NMA treatment contrast estimate.

Appendix B

R Function Documentation: Estimating the Proportion Contributions of Direct Evidence from Individual Studies to Evidence Streams in a Single NMA Treatment Contrast Estimate.

1. `get.streams_df`

    1.1. Arguments

    1.1.1. `hatmatrix`: A list returned from `getHatMatrix` function.

    1.1.2. `comparison`: A character string indicating the treatment comparison user interested in, with format “treatment A:treatment B”.

    1.2. Returns: A dataframe containing paths, flows, length, proportion contributions of direct evidence in a single NMA contrast estimate.

    1.3. Algorithm: Call the `comparisonStreams` function in the revised `flow_contribution` package to obtain the data of evidence streams and construct to a dataframe.

2. `get.studies_df`

    2.1. Arguments

    2.1.1. `hatmatrix`: A list returned from `getHatMatrix` function.

    2.1.2. `comparison`: A character string indicating the treatment comparison user interested in, with format “treatment A:treatment B”.

    2.2. Returns: A list containing two dataframes `StudyContributionPerFlow` and `StudyReport`.

2.2.1. \`StudyContributionPerFlow\`: A dataframe containing paths, flows, lengths of each stream, proportion contributions of direct evidence, study ID, proportion contributions of direct evidence from individual studies to a contrast estimate, and proportion contributions of direct evidence from individual studies to a stream in a single NMA treatment contrast estimate.

2.2.2. StudyReport: A simplified dataframe containing paths, source and sink of each stream, study ID, and proportion contributions of direct evidence from individual studies to a stream in a single NMA treatment contrast estimate.

2.3. Algorithm

The function first merges two dataframe returned from \`get.streams_df\` and \`get.studies_df\` horizontally by the columns \`contrastinflow\` (edges in each stream) and \`comparison\` (direct evidence from individual studies).

Then, the function calculates the proportion contributions of direct evidence from individual studies in each stream to the contrast estimate (column \`study.cont\`) based on the matrix P and matrix P* since the proportion contributions of individual studies (p^*) may be distributed in multiple streams. For example, p_{xv} (or p_{xv}^{xy}) in **Figure 11** was distributed in two streams with paths "x→v→y" and "x→v→u→y", separately. The formula of the proportion contribution of an individual study in a given stream S_i, denoted by $p'_{xv,i}$, can be expressed as follows:

$$p'_{xv,i} = p^*_{xv} \times \frac{\frac{\varphi_i}{|\pi_i|}}{p_{xv}},$$

$i = 1, \ldots, I$ where I is the number of streams flow through the given edge xv; p_{xv}^{*} is the proportion contribution of direct evidence from individual studies, $\phi_{xv}/|\pi_{xv}|$ is the flow of each edge in the stream, p_{xv} is the proportion contribution of direct evidence to a treatment contrast estimate.

Appendix C

R Function Documentation: Developing the Robustness Index

1. `get.studies_rob`

    1.1. Arguments

        1.1.1. `indata`: A csv file, the dataset for NMA.

        1.1.2. `hatmatrix`: A list returned from `getHatMatrix` function.

        1.1.3. `comparison`: A character string indicating the treatment comparison user interested in, with format “treatment A:treatment B”.

    1.2. Returns: A dataframe merged from the dataframe `StudyContributionPerFlow` returned from `get.studies_df` function and the RoB of studies by study ID.

2. `get.robustness`

    2.1. Arguments

        2.1.1. `indata`: A csv file, the dataset for NMA.

        2.1.2. `hatmatrix`: A list returned from `getHatMatrix` function.

        2.1.3. `comparison`: A character string indicating the treatment comparison user interested in, with format “treatment A:treatment B”.

    2.2. Returns: A list containing three dataframes named as `allvariables`, `robustness_df`, `pathrobust` and an array named as `network_robustness.contrast`. The `edge_p` and `path_p` columns represent the robustness of edges of each stream and the robustness of streams in an NMA contrast estimate. The array `network_robustness.contrast` indicates the robustness of the stream network.

2.3. Algorithm of developing the robustness index (see Method 3.2)

3. `get.contrast_robust`

3.1. Arguments

3.1.1. `indata`: A csv file, the dataset for NMA.

3.1.2. `hatmatrix`: A list returned from `getHatMatrix` function.

3.2. Returns: A dataframe containing all NMA treatment contrast estimates, their robustness, average length of streams, numer of streams.

3.3. Algorithm: Collect data by looping `get.robustness` function over all treatment comparisons.