PP03

MetaSuite: Empowering Network Meta-Analysis in Clinical Research Through Interactive Visualizations and User-Friendly Evidence Synthesis Tools

Hirak Sen Roy, Princeps Technologies, Hyderabad, India

Abstract

Systematic reviews and meta-analyses are central to evidence-based medicine, yet conducting them remains surprisingly fragmented. Researchers typically assess risk of bias in spreadsheets, run statistical analyses in R or specialized software, and create figures in separate graphics tools. This workflow creates multiple opportunities for error, makes reproducibility difficult, and complicates the peer review process.

We developed **MetaSuite**, an open-source R Shiny platform that integrates the complete evidence synthesis pipeline—from risk-of-bias assessment through network meta-analysis—in a single environment. The platform supports five risk-of-bias frameworks (ROB 2, ROBINS-I, ROBINS-E, QUADAS-2, QUIPS), handles multiple outcome types and effect measures, and provides comprehensive diagnostics including heterogeneity assessment, publication bias detection, and network inconsistency evaluation. Unlike reactive analytical tools that silently recompute results, MetaSuite requires explicit user action for all analyses, maintaining clear audit trails between inputs and outputs.

Built on established R packages (metafor, netmeta, rankinma) with a Bootstrap 5 interface, MetaSuite separates statistical computation from visualization customization. This architecture enhances transparency for regulatory review while producing publication-ready outputs at each stage. The open-source implementation is available at https://hirak-sen-roy.shinyapps.io/metasuite/ with complete documentation.

Introduction

Meta-analysis has become essential for estimating treatment effects precisely, exploring heterogeneity across studies, and comparing multiple interventions simultaneously (Higgins et al., 2023). Network meta-analysis (NMA) extends traditional pairwise methods by combining direct head-to-head evidence with indirect comparisons through common comparators. This capability is par-

ticularly valuable when direct evidence is sparse or when comprehensive treatment rankings are needed to inform clinical guidelines and regulatory decisions.

Despite decades of methodological development, evidence synthesis workflows remain frustratingly disjointed. A typical systematic review requires juggling multiple tools: risk-of-bias assessments in spreadsheets, statistical analyses in R or Stata, and visualization in graphics software. Manual data transfers between these platforms create opportunities for transcription errors and undocumented changes. Separating statistical estimation from visualization obscures methodological choices and makes peer review more difficult. Many analytical environments use reactive computation, where changing one parameter silently triggers recalculation of all downstream results—a feature that can undermine reproducibility when users are unaware these recomputations have occurred.

MetaSuite addresses these challenges by integrating risk-of-bias assessment, pairwise meta-analysis, and network meta-analysis in a single reviewable environment. We enforce explicit execution control—no computation occurs without user action—and strictly separate statistical estimation from presentation choices. By building on established R packages while providing structured workflows, the platform balances methodological flexibility with the reproducibility requirements essential for regulatory and clinical decision-making.

Background and Motivation

Modern risk-of-bias frameworks evaluate specific threats to study validity without producing algorithmic quality scores. ROB 2 assesses randomized trials across five domains, while ROBINS-I evaluates non-randomized intervention studies through seven domains covering confounding, selection bias, and other concerns. ROBINS-E extends this approach to exposure studies, QUADAS-2 addresses diagnostic accuracy research, and QUIPS focuses on prognosis studies. These assessments help contextualize synthesis results and inform sensitivity analyses.

Pairwise meta-analysis pools effect estimates across studies, yielding more precise estimates than individual studies while quantifying between-study heterogeneity. Researchers can choose among several variance estimators—restricted maximum likelihood (REML), DerSimonian-Laird (DL), Hunter-Schmidt (HS), and Paule-Mandel (PM)—each with distinct statistical properties. Heterogeneity is typically assessed through I^2 and τ^2 statistics, which can be further explored through subgroup analyses. Publication bias assessment using funnel plots and influence diagnostics reveal studies with disproportionate impact on pooled estimates.

Network meta-analysis extends pairwise methods to compare multiple treatments simultaneously by combining direct and indirect evidence. When treatments A and B have both been compared to C, we can derive indirect evidence about A versus B through this common comparator. Key considerations include ensuring network connectivity, evaluating consistency between direct and indirect evidence, and appropriately handling multi-arm trials. Treatment rankings through SUCRA

values and probabilistic displays support decision-making, though these rankings should always be interpreted alongside effect magnitudes and uncertainty estimates.

Current evidence synthesis workflows face several persistent challenges. First, fragmentation across multiple platforms creates manual transfer steps that reduce reproducibility and introduce errors. Second, coupling analysis with visualization often obscures methodological choices, while reactive computation may silently recompute results without user awareness. Third, the absence of integrated validation workflows makes it difficult to verify that outputs correctly reflect analytical decisions. MetaSuite addresses these limitations through explicit modularity, user-controlled execution, and strict separation between statistical computation and presentation.

System Architecture

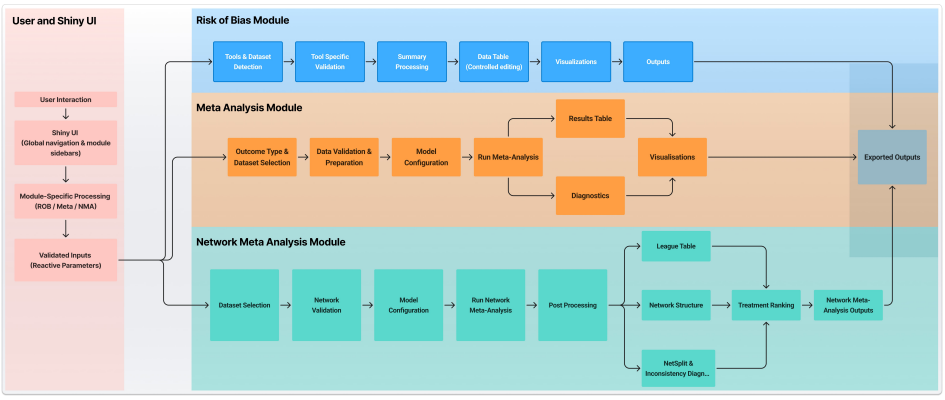


Figure 1: System architecture showing three core modules (Risk of Bias, Meta-Analysis, Network Meta-Analysis) with standardized processing pipelines enforcing explicit execution control and separation between statistical computation and visualization.

MetaSuite organizes functionality into three core modules—Risk of Bias Assessment, Pairwise Meta-Analysis, and Network Meta-Analysis—each following a standardized pipeline: input validation, data preparation, explicit execution, post-processing, diagnostics, and visualization. The architecture is built around five design principles that prioritize reproducibility and transparency.

Modularity ensures that each analytical module operates independently with well-defined inputs and outputs. Modules do not share computational state, which allows us to test, validate, and update methods without cascading effects across the platform. This isolation also means that adding new statistical methods or diagnostic tools does not require refactoring existing functionality.

Explicit execution control is central to our design. Users must trigger all computationally intensive operations through dedicated action buttons—no computation occurs automatically when data is uploaded or parameters change. This prevents the silent recomputation common in reactive environments and maintains clear audit trails linking inputs to outputs. While reactive interfaces offer convenience, they can undermine reproducibility when users are unaware that results have been recalculated.

Separation of concerns strictly isolates statistical estimation from presentation. Model configuration and execution occur in dedicated analytical stages, while visualization customization happens downstream and affects only display attributes, never statistical results. This means users can adjust colors, fonts, and plot dimensions without any risk of inadvertently altering effect estimates or confidence intervals.

Traceability maintains explicit linkage between input data, model parameters, statistical results, and final outputs. Diagnostic summaries and intermediate results remain accessible for review at any workflow stage, supporting verification workflows essential for regulatory submission and peer review.

Presentation safety restricts visualization customization to aesthetic attributes (colors, fonts, dimensions) that cannot affect statistical interpretation. Users cannot, for example, selectively hide studies from forest plots or adjust axis scales in ways that would misrepresent effect magnitudes. This constraint prevents inadvertent misrepresentation of analytical results through presentation choices.

For statistical computation, we leverage established R packages: **metafor** (Viechtbauer, 2010) for pairwise meta-analysis with diverse effect measures and heterogeneity estimators; **netmeta** (Balduzzi et al., 2023) for frequentist network meta-analysis with consistency evaluation; **rankinma** (Lin et al., 2023) for treatment ranking and probabilistic analyses. The frontend uses Shiny for the reactive web framework and Bootstrap 5 for modern UI elements. This technology stack balances statistical rigor with accessibility, using mature implementations while providing structured workflow orchestration.

User Interface and Workflow

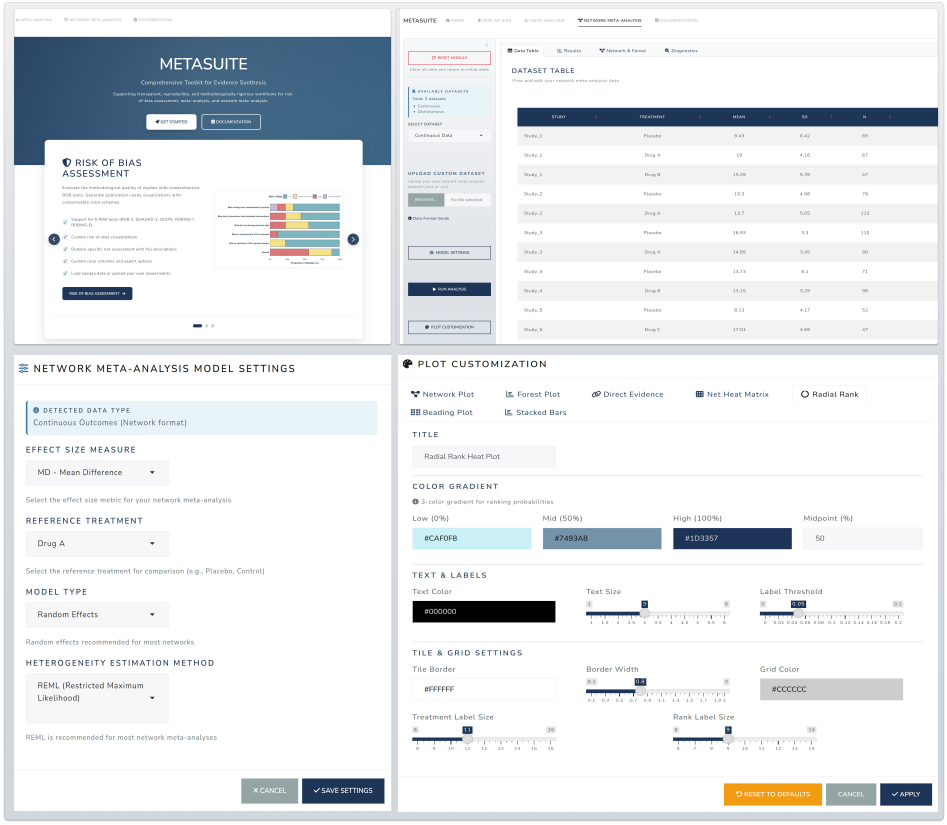


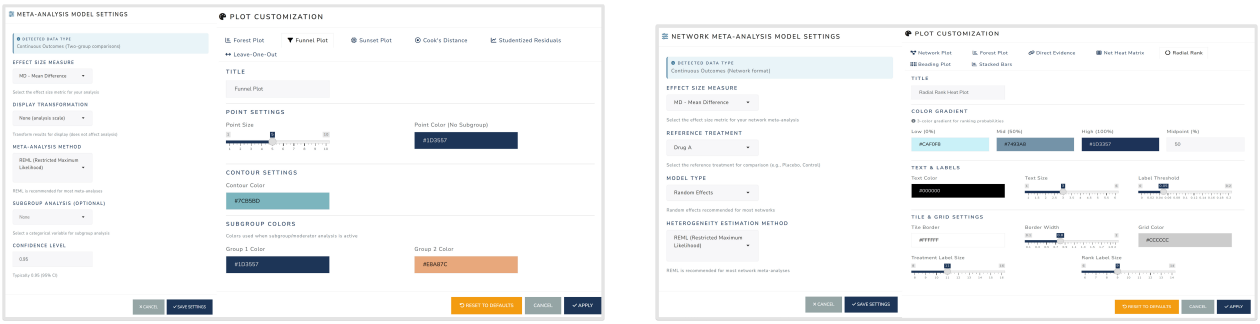
Figure 2: User interface displaying module tabs (Risk of Bias, Meta-Analysis, Network Meta-Analysis). The left sidebar contains sequential workflow controls including dataset selection, file upload, model configuration, execution triggers, and plot customization options. The main panel shows the data table with workflow stage tabs.

The MetaSuite interface emphasizes workflow clarity and reproducibility. Users access the three core analytical modules through dedicated tabs on the landing page. This tab-based design provides immediate visual access to all capabilities while maintaining clear functional separation.

When users select a module, the sidebar dynamically presents workflow controls in a standardized sequence: dataset selection (with sample data access), file upload, model settings, explicit analysis execution buttons, and plot customization options. This vertical arrangement enforces logical analytical progression—data loading precedes model configuration, which precedes execution, which precedes visualization. The sidebar design ensures that only methodologically valid options appear at each stage, preventing invalid analytical configurations.

The main content area uses secondary tabbed panels for distinct workflow stages: Data Input for file upload and format validation, Data Table for interactive review and editing, Model Settings for detailed configuration, Results for statistical outputs and diagnostics, and Plots for visualization generation and export. This dual-level structure maintains clear separation between what analysis is being performed and where in the analytical pipeline the user is currently working.

All computationally intensive operations require explicit user action through “Run Analysis” buttons in the sidebar. No statistical computation occurs automatically when data is uploaded or parameters are modified. This execution model, combined with visual feedback during computation, maintains clear audit trails and prevents the silent recomputation issues that plague reactive analytical environments. Once analysis completes, the interface provides immediate access to results and automatically activates the plots tab, while preserving full navigation to review inputs, modify settings, or export data at any stage.



(a) Meta-analysis configuration **(b)** Network meta-analysis configuration

Figure 3: Model settings and plot customization interfaces. The left panel provides analytical configuration including outcome type detection, effect measure selection, and model specification. The right panel controls presentation attributes without altering statistical computations.

The pairwise meta-analysis module automatically detects outcome types (continuous, dichotomous, or correlation) from the uploaded data structure, guiding appropriate effect measure selection. Users configure effect size metrics, heterogeneity estimation methods (REML, ML, DL, HS, PM), and optional subgroup moderators. Network meta-analysis requires additional specifications: reference treatment selection, model type (fixed or random effects), and consistency evaluation approaches. The interface validates network connectivity before permitting execution.

Plot customization happens strictly after statistical estimation. Once analysis executes, users access dedicated panels controlling visual attributes: plot types, titles and labels, point sizes and colors, contour fills, and color schemes. All customization affects presentation only; re-customizing plots never triggers statistical recomputation.

Analytical Modules and Outputs

Risk of Bias Assessment

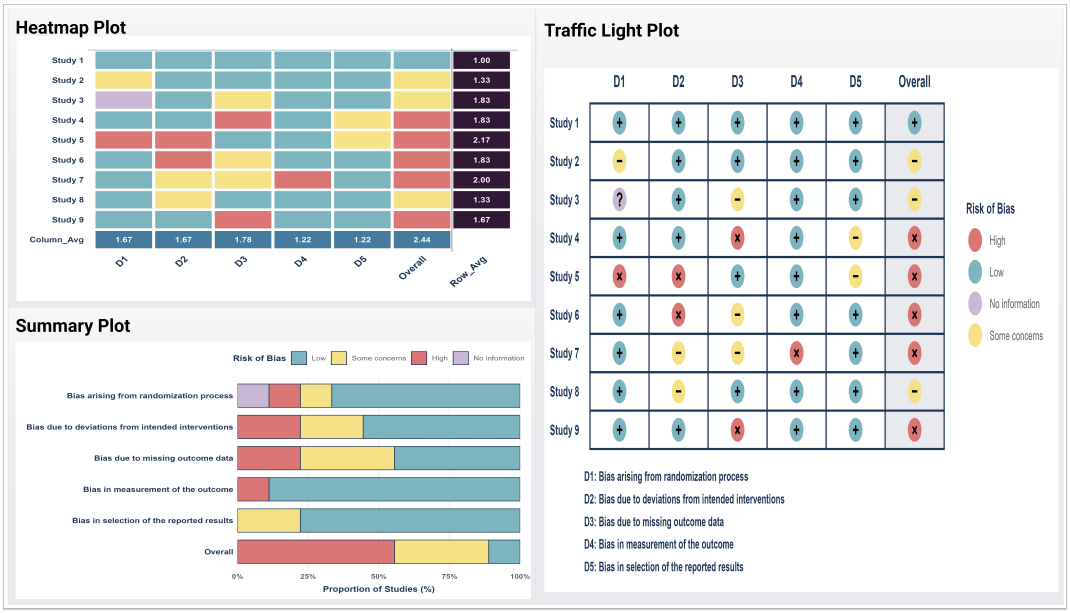


Figure 4: Risk of Bias module visualization outputs showing heatmap plot with row and column averages (left), weighted bar plot displaying proportional distribution of risk judgments across domains (middle), and traffic light plot presenting individual study-by-domain risk assessments (right). Color coding indicates risk levels with domain abbreviations corresponding to tool-specific bias domains.

MetaSuite supports five risk-of-bias frameworks, each designed for specific study types. ROB 2 assesses randomized controlled trials across five domains plus an overall judgment. ROBINS-I evaluates non-randomized intervention studies through seven domains addressing confounding, selection, classification, deviations, missing data, measurement, and reporting biases. ROBINS-E extends this to exposure studies, while QUADAS-2 assesses diagnostic accuracy studies and QUIPS evaluates prognosis studies.

The workflow enforces tool-specific validation. Data must contain a Study identifier column, domain columns (D1, D2, …, Dn) matching the selected tool's structure, and an Overall risk judgment. Risk categories must exactly match tool specifications—for example, "Low", "Some Concerns", "High" for ROB 2. This validation prevents category mismatches that would produce invalid visualizations.

Three complementary visualizations (McGuinness and Higgins, 2021) serve different analytical needs. **Traffic-light plots** display individual study-by-domain assessments in a matrix format, allowing rapid identification of problematic studies or consistently high-risk domains. **Weighted bar plots** (summary plots) show the proportional distribution of risk judgments across domains, highlighting areas of methodological concern. **Heatmap plots** use numeric encoding with row and column averages to support pattern detection and identification of outlier studies. All visualizations export at 300 DPI in PNG or JPG formats, suitable for publication and presentation.

Pairwise Meta-Analysis

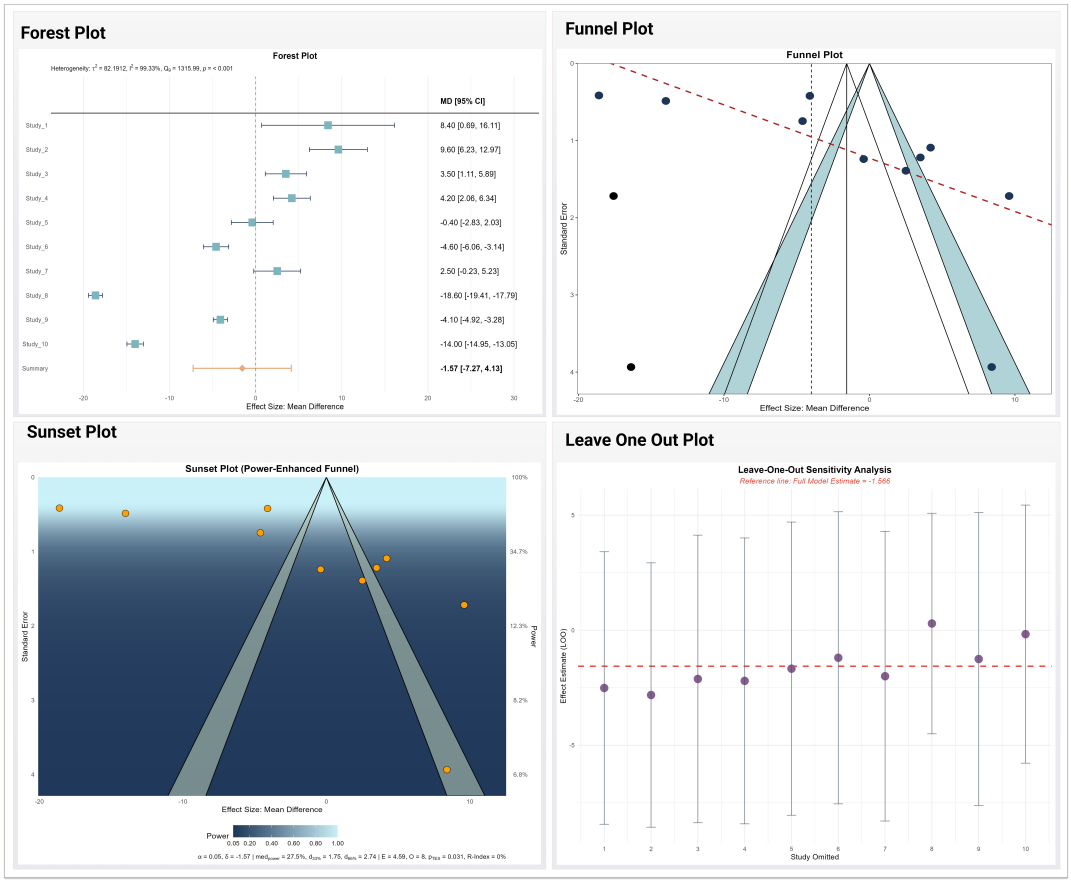


Figure 5: Pairwise meta-analysis visualization outputs. Forest plot (top left) displays individual study effects and pooled estimate with heterogeneity statistics. Funnel plot (top right) evaluates publication bias. Sunset plot (bottom left) provides power-enhanced funnel visualization with significance contours. Leave-one-out plot (bottom right) presents influence diagnostics.

The pairwise meta-analysis module handles three outcome types, each with distinct effect measures. **Continuous outcomes** (requiring study, treatment, mean, sd, n columns) support mean difference (MD) and standardized mean difference (SMD) calculation. **Dichotomous outcomes** (study, treatment, events, n) enable odds ratio (OR), risk ratio (RR), or risk difference (RD) estimation. **Correlation outcomes** (study, treatment, correlation, n) use Fisher's r-to-z transformation. The interface automatically detects outcome type from data structure and restricts effect measure selection accordingly.

Five heterogeneity estimators provide methodological flexibility. **REML** (Restricted Maximum Likelihood) is the recommended default, providing unbiased variance estimation. **ML** (Maximum Likelihood) enables likelihood-based inference. **DL** (DerSimonian-Laird) offers computational simplicity but has known bias in small samples. **HS** (Hunter-Schmidt) adjusts for sampling error, while **PM** (Paule-Mandel) uses iterative variance estimation. Users select estimators based on sample size, heterogeneity magnitude, and inferential goals.

Forest plots present the core synthesis results. Individual study effect estimates appear as data points with horizontal lines representing 95% confidence intervals. Studies are listed vertically with associated sample sizes and effect magnitudes. The pooled estimate appears as a diamond at the bottom, with diamond width indicating the 95% confidence interval. Heterogeneity statistics (I^2, τ^2, Q-statistic with p-value) are annotated directly on the plot.

Funnel plots assess publication bias by plotting effect size against standard error. Larger studies appear at the top (smaller standard error) with smaller studies scattered below. In the absence of bias, studies distribute symmetrically within a triangular 95% confidence region. Asymmetry—particularly empty regions in lower corners—suggests potential missing studies. The plot overlays Egger's regression line to detect small-study effects and displays trim-and-fill imputed studies as open circles.

Sunset plots (Kossmeier et al., 2020) enhance funnel plot interpretation by adding statistical power contours at 50%, 80%, and 95% levels through graduated color gradients. Studies in upper contour regions demonstrate adequate statistical power, while those in lower zones indicate underpowered investigations. The pooled effect estimate appears as a vertical line, allowing immediate assessment of whether the meta-analytic conclusion is driven primarily by well-powered studies.

Influence diagnostics identify studies with disproportionate impact. Cook's distance quantifies each study's influence on the overall pooled estimate. Studentized residuals identify statistical outliers by plotting standardized residuals with Bonferroni-adjusted confidence limits. Leave-one-out analysis provides comprehensive sensitivity assessment by sequentially removing each study and recalculating the pooled estimate. Results appear as a forest plot showing the revised pooled estimate and confidence interval for each omission.

Comprehensive heterogeneity evaluation includes I^2 (percentage of variability attributable to between-study differences rather than sampling error), τ^2 (estimated variance of true effects across studies), Q-statistic with associated p-value, and prediction intervals indicating the expected range of effects in future studies. Optional subgroup analysis partitions studies by categorical moderators, providing subgroup-specific pooled estimates with formal tests of between-group differences.

Network Meta-Analysis

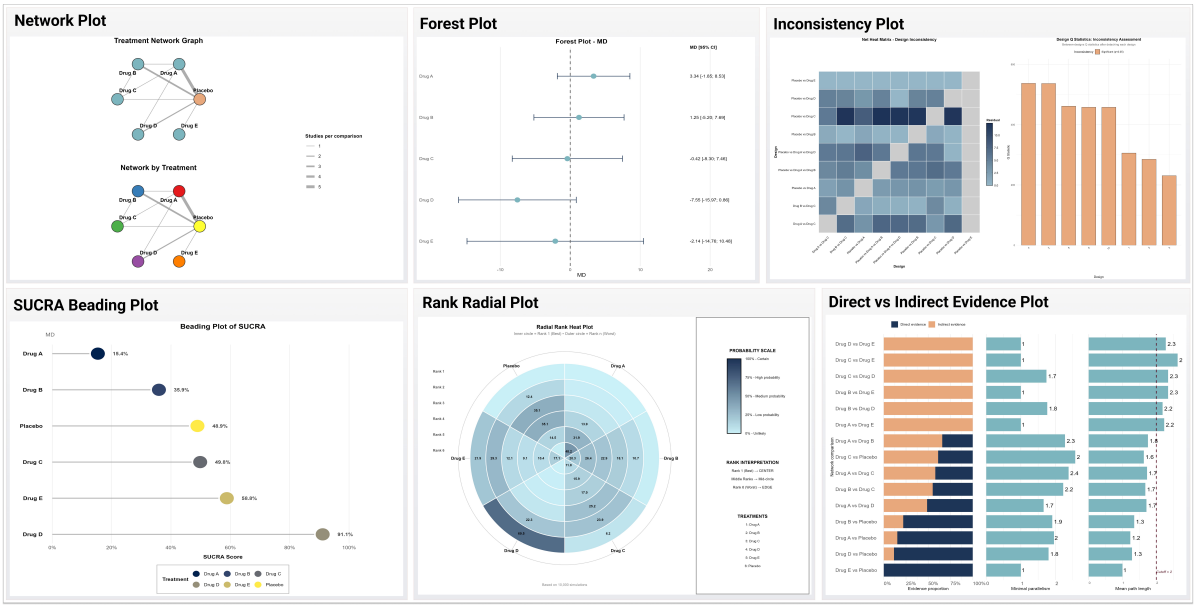


Figure 6: Network meta-analysis outputs. Network plot (top left) displays treatment connectivity structure. Forest plot (top center) shows all pairwise comparison estimates. Inconsistency plot (top right) presents design-based inconsistency assessment. SUCRA beading plot (bottom left) visualizes treatment ranking probabilities. Rank radial plot (bottom center) displays probabilistic ranking across treatments. Direct vs. indirect evidence plot (bottom right) compares effect estimates from different evidence sources.

The network meta-analysis module extends pairwise methods to simultaneous multi-treatment comparison. Data must be in network format where each row represents a treatment arm within a study, including study identifier, treatment name, outcome measures (mean/sd/n for continuous; events/n for dichotomous), and treatment_class for grouped visualization. The module validates network connectivity, verifying that all treatments link through common comparators.

Effect measures parallel pairwise options but are interpreted relative to a user-specified reference treatment (typically placebo or standard-of-care). The interface generates league tables showing all pairwise comparisons in matrix format, providing comprehensive comparative effectiveness summaries. Forest plots display effect estimates relative to the reference treatment with confidence intervals.

Consistency evaluation assesses agreement between direct and indirect evidence—a fundamental assumption of network meta-analysis. We implement design-based inconsistency assessment, comparing design-specific direct estimates against network predictions. Significant inconsistency may indicate effect modification, differential bias patterns, or violation of transitivity assumptions. Inconsistency heatmaps and bar charts identify specific treatment comparisons or study designs contributing to inconsistency.

Treatment ranking supports decision-making through multiple approaches. SUCRA (Surface Under the Cumulative Ranking curve) values range from 0 (worst treatment) to 100 (best treatment),

providing single-number summaries. However, rankings depend on the chosen outcome and may differ across efficacy versus safety endpoints. Probabilistic ranking displays show the probability that each treatment occupies each rank position, revealing ranking uncertainty. Rank radial plots (Veroniki et al., 2016) visualize the full probability distribution through color gradients, immediately highlighting treatments with consistent ranking versus uncertain positioning.

Network plots offer structure visualization showing connectivity patterns, with node sizes representing study numbers and edge widths indicating direct evidence quantity. The direct versus indirect evidence plot evaluates consistency by comparing direct head-to-head estimates against indirect estimates derived through common comparators. Agreement within confidence intervals supports consistency; systematic disagreement suggests potential violations requiring investigation.

All network outputs export at publication quality with comprehensive legends, statistical annotations, and methodological details. The implementation uses the **netmeta** package (Balduzzi et al., 2023) for frequentist network meta-analysis and **rankinma** (Lin et al., 2023) for advanced treatment ranking methods.

Conclusion

MetaSuite addresses critical limitations in evidence synthesis workflows by integrating risk-of-bias assessment, pairwise meta-analysis, and network meta-analysis in a unified, transparent, and reproducible platform. Through modular architecture, explicit execution control, and strict separation of statistical computation from presentation, the platform enhances reproducibility, supports regulatory review, and maintains methodological integrity.

Built on established R packages and aligned with contemporary reporting standards (Higgins et al., 2023), MetaSuite produces publication-ready outputs while enforcing structured analytical workflows. The open-source implementation enables independent verification, methodological extension, and integration into institutional evidence synthesis processes.

For clinical researchers, health technology assessment bodies, and regulatory evaluators, MetaSuite offers a reviewable alternative to fragmented workflows, supporting transparent and reproducible evidence synthesis from risk-of-bias assessment through treatment ranking and comparative effectiveness evaluation.

Software Availability

Live Demo: https://hirak-sen-roy.shinyapps.io/metasuite/

GitHub Repository: https://github.com/hirak8981/metaAnalysis

The repository includes complete source code, sample datasets for each module, comprehensive documentation, and installation instructions for local deployment.

References

Balduzzi, S., Rücker, G., and Schwarzer, G. (2023). netmeta: An R package for network meta-analysis using frequentist methods. *Journal of Statistical Software*, 106(2):1–40.

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., and Welch, V. A. (2023). *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane, version 6.4 edition.

Kossmeier, M., Tran, U., and Voracek, M. (2020). *metaviz: Forest plots, funnel plots, and visual funnel plot inference for meta-analysis*. R package version 0.3.1.

Lin, L., Zhang, J., Hodges, J. S., and Chu, H. (2023). *rankinma: Ranking in Network Meta-Analysis*. R package version 0.2.3.

McGuinness, L. A. and Higgins, J. P. T. (2021). Risk-of-bias visualization (robvis): An R package and web app for visualizing risk-of-bias assessments. *Research Synthesis Methods*, 12(1):55–61.

Veroniki, A. A., Straus, S. E., Fyraridis, A., and Tricco, A. C. (2016). The rank-heat plot is a novel way to present the results from a network meta-analysis including multiple outcomes. *Journal of Clinical Epidemiology*, 76:193–199.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3):1–48.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Hirak Sen Roy

Company: Princeps Technologies Inc.

Address: Hyderabad, India

Email: hirak@princepstech.com

Website: https://www.princepstech.com/