The Journal of Open Source Software

Risk of Bias: Explainable, human-in-the-loop software for general risk-of-bias assessment

Robert Luke [1] and **Jessica Biesiekierski** [2]

1 Macquarie University, Macquarie University Hearing & Department of Linguistics, Australian Hearing Hub, Sydney, New South Wales, Australia **2** The University of Melbourne, School of Agriculture, Food and Ecosystem Sciences, Human Nutrition Group, Victoria, Australia

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Gabriela Alessio Robles

Reviewers:

- @tellarin
- @RISbO001

Submitted: 12 June 2025
Published: unpublished

Summary

Assessing risk of bias (RoB) is a fundamental component of evidence synthesis, directly affecting the credibility and interpretability of systematic reviews and meta-analyses. RoB assessment clarifies to what extent findings from primary studies can be trusted, guiding both research conclusions and downstream policy or clinical recommendations (Higgins et al., 2019; Page et al., 2021; P. Whiting et al., 2016). Despite its importance, RoB assessment remains time-intensive and demands specialized expertise. The *risk-of-bias* Python package provides a general, framework-agnostic software assistant for risk-of-bias assessment, combining explainable AI with open, programmable infrastructure. The tool is designed to support any domain-based risk-of-bias instrument, with the widely adopted Cochrane RoB 2 tool for randomized trials (Sterne et al., 2019) implemented as a framework exemplar. By storing explicit evidence and reasoning for every answer, and providing both command line and web interfaces, *risk-of-bias* enables explainable, auditable, and efficient assessment, supporting reproducible research workflows.

Statement of Need

Manual risk-of-bias assessment is a critical bottleneck in evidence synthesis, with typical reviews requiring 10–60 minutes of expert time per study, repeated across dozens or hundreds of manuscripts (Savovic et al., 2014). The gold standard remains two independent human reviewers with a third for adjudication—an approach that is resource intensive and often unattainable in time- or resource-constrained projects. Commercial platforms (e.g., Covidence (Kellermeyer et al., 2018), DistillerSR) offer user-friendly interfaces, while AI-driven tools like RobotReviewer (Marshall et al., 2016) provide partial automation for specific tasks, but few options combine openness, programmability, explainability, and affordability. The *risk-of-bias* package addresses this gap by delivering a fully open-source, scriptable tool that provides structured, explainable output, and supports integration with both human and AI-driven workflows. Because the package is open source and developed with modern software best practices (including continuous integration), it can be rapidly updated to support the latest AI models as they emerge. Unlike commercial software, where underlying models may be opaque and lag behind the state of the art, this approach puts the technology directly in the hands of researchers—empowering them to select, use, or even contribute the most powerful and up-to-date AI models for their needs.

Software Overview & Architecture

Risk-of-bias is built around a generic, hierarchical assessment structure:

Framework → Domain → Question → Response

This mirrors all major RoB instruments, enabling use beyond the Cochrane RoB 2 tool. The package offers:

- A modular core, where any framework is defined as a JSON schema capturing its domains, questions, and allowed responses.
- Data classes (via Pydantic) that explicitly store, for each answer: the response, supporting evidence (verbatim text from the manuscript), and a natural language reasoning/explanation.
- Multiple user interfaces: a command-line interface (CLI) for batch assessment and workflow integration, and a web interface for interactive analysis and report download. This enables both technical and non-technical users to use the tool effectively.
- An engine that systematically applies the framework to imported manuscripts, walking through the assessment questions and storing structured outputs.
- Export functions for RobVis-compatible CSV summaries, facilitating high-quality visualizations using the *robvis* R package or web app (McGuinness & Higgins, 2021).

For batch analysis, the CLI allows processing of entire directories of manuscripts, automatically generating summary CSVs for cross-study visualization or meta-analysis.

Explainable AI & Evidence-linked Reasoning

Unlike "black-box" AI tools, *risk-of-bias* stores and surfaces both the **evidence** (exact textual excerpts) and the **reasoning** (explanation of how evidence informs the answer) for every question in every domain. This design meets the auditability requirements of leading journals and systematic review standards, supporting both transparent reporting and dispute resolution in collaborative review teams. When used alongside independent human reviewers, the software's explicit justifications make it easy to compare and resolve discrepancies, and to understand why the software or a reviewer made a particular assessment. This explainable approach is particularly valuable as LLM-based and hybrid systems become more common in evidence synthesis, ensuring assessments remain interpretable and verifiable.

Human-in-the-Loop Augmentation

The *risk-of-bias* package is designed to augment, not replace expert human judgment. Its intended workflow is to augment the established approach of two independent human reviewers, with a third human reviewer adjudicating discrepancies. AI can meaningfully enhance this process: for example, the software can serve as an additional reviewer alongside human experts, providing a systematically derived perspective while leaving final adjudication to a human. Incorporating an AI perspective can help reveal potential biases in both directions—including those arising from the AI itself—and offer a complementary lens for evaluating studies. In situations where resource constraints make the gold standard unachievable, AI tools can support more consistent and thorough assessments, helping raise the overall quality of risk-of-bias evaluations as the field moves toward best practice.

Current Framework Support & Extensibility

While RoB 2 for randomized trials is implemented end-to-end (with both CLI and web UI, and export to RobVis/CSV (McGuinness & Higgins, 2021)), the architecture is framework-agnostic by design. Additional frameworks—such as ROBINS-I (Sterne et al., 2016), ROBINS-D (Higgins et al., 2024), QUADAS-2 (P. F. Whiting et al., 2011), and PROBAST (Wolff et al., 2019)—can be registered as JSON schemas immediately, leveraging the same hierarchical logic (framework → domain → question → response), and is in the roadmap for future explicit inclusion in the software. This software is already being utilised to support study design and systematic literarure reviews.

Acknowledgements

We acknowledge the foundational work of the *robvis* package (McGuinness & Higgins, 2021), the authors of RoB 2 (Sterne et al., 2019), and the wider open-source and evidence synthesis community whose contributions inform both the methodology and the software ecosystem.

References

Higgins, J. P., Morgan, R. L., Rooney, A. A., Taylor, K. W., Thayer, K. A., Silva, R. A., Lemeris, C., Akl, E. A., Bateson, T. F., Berkman, N. D., & others. (2024). A tool to assess risk of bias in non-randomized follow-up studies of exposure effects (ROBINS-e). *Environment International*, *186*, 108602. https://doi.org/10.1016/j.envint.2024.108602

Higgins, J. P., Savovic, J., Page, M. J., Elbers, R. G., & Sterne, J. A. (2019). Assessing risk of bias in a randomized trial. *Cochrane Handbook for Systematic Reviews of Interventions*, 205–228. https://doi.org/10.1002/9781119536604.ch8

Kellermeyer, L., Harnke, B., & Knight, S. (2018). Covidence and rayyan. *Journal of the Medical Library Association: JMLA*, *106*(4), 580. https://doi.org/10.5195/jmla.2018.513

Marshall, I. J., Kuiper, J., & Wallace, B. C. (2016). RobotReviewer: Evaluation of a system for automatically assessing bias in clinical trials. *Journal of the American Medical Informatics Association*, *23*(1), 193–201. https://doi.org/10.1093/jamia/ocv044

McGuinness, L. A., & Higgins, J. P. (2021). Risk-of-bias VISualization (robvis): An r package and shiny web app for visualizing risk-of-bias assessments. *Research Synthesis Methods*, *12*(1), 55–61. https://doi.org/10.1002/jrsm.1411

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., & others. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Bmj*, *372*. https://doi.org/10.1136/bmj.n71

Savovic, J., Weeks, L., Sterne, J. A., Turner, L., Altman, D. G., Moher, D., & Higgins, J. P. (2014). Evaluation of the cochrane collaboration's tool for assessing the risk of bias in randomized trials: Focus groups, online survey, proposed recommendations and their implementation. *Systematic Reviews*, *3*, 1–12. https://doi.org/10.1186/2046-4053-3-37

Sterne, J. A., Hernan, M. A., Reeves, B. C., Savovic, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., & others. (2016). ROBINS-i: A tool for assessing risk of bias in non-randomised studies of interventions. *Bmj*, *355*. https://doi.org/10.1136/bmj.i4919

Sterne, J. A., Savovic, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., & others. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *Bmj*, *366*. https://doi.org/10.1136/bmj.l4898

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M., Sterne, J. A., Bossuyt, P. M., & Group*, Q. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, *155*(8), 529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

Whiting, P., Savovic, J., Higgins, J. P., Caldwell, D. M., Reeves, B. C., Shea, B., Davies, P., Kleijnen, J., Churchill, R., & others. (2016). ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology*, *69*, 225–234.

https://doi.org/10.1016/j.jclinepi.2015.06.005

Wolff, R. F., Moons, K. G., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., Mallett, S., & Group†, P. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine*, *170*(1), 51–58. https://doi.org/10.7326/M18-1376