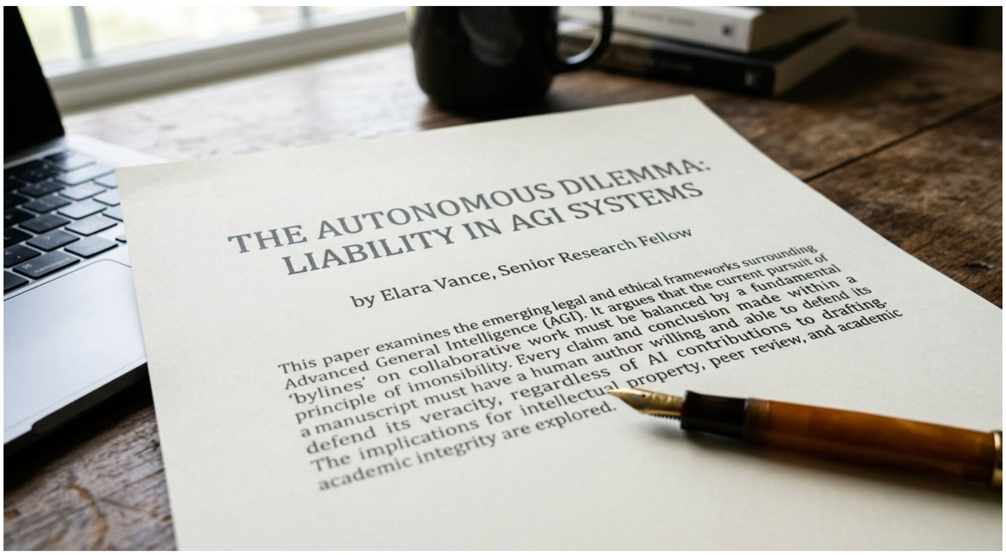
Review Methodology for AI-Assisted Research Paper Drafting

Ronald G. Thomas

2026-08-21



A manuscript outlined by a human and drafted by an assistant still needs one author who can defend every claim: this is the checklist for becoming that author.

Introduction

A drafting pattern increasingly common in academic and scientific writing pairs a human researcher, who designs the study, runs or directs the analysis, and sets the argument, with an AI assistant that drafts prose, summarizes literature, or helps structure results. The arrangement compresses drafting time considerably, but it collides with a settled point of publishing policy. Every major journal body that has published a position on the question agrees that an AI tool cannot be an author, because authorship requires accountability, and accountability cannot attach to a tool. The named human authors carry the full obligation to stand behind every claim, citation, and figure, regardless of what fraction of the prose an assistant drafted first.

This post presents a methodological framework for closing the gap between "an assistant helped draft this" and "I can defend every claim in this manuscript." It is a transposition of a companion framework applied elsewhere in this series to R package code, blog prose, and textbook chapters:

Code Review for AI-Assisted R Package Development, Review Methodology for AI-Assisted Blog Post Drafting, and Review Methodology for AI-Assisted Textbook Drafting. Unlike those three siblings, this post also draws on an external constraint the others do not have: published authorship and disclosure policy from ICMJE, COPE, WAME, and a set of major publishers, which set a floor this framework's ownership test builds on rather than invents from scratch.

Motivations

- **The responsibility gap in a research paper carries professional and ethical weight the other media in this series do not.** A hallucinated citation, a misreported statistic, or a claim the data do not support becomes part of the scientific record, cited by other researchers who trust the byline.
- **AI-drafted manuscript prose has characteristic failure modes.** Fabricated or mismatched citations, overstated conclusions relative to the actual results, and confident literature-review claims that do not check out against the cited source recur often enough to warrant a dedicated review pass.
- **Journal and publisher policy already forecloses one path entirely.** No major venue accepts an AI tool as an author, and most now require a specific, located disclosure statement when AI assistance was used; a review process that ignores this treats a compliance requirement as optional when it is not.
- **Ownership needs a definition, not a feeling.** "The assistant drafted a clean paragraph" and "I can defend this paragraph's claim against a reviewer's specific challenge, unaided" are different claims, and only the second one licenses submitting the manuscript under your name.

Objectives

1. Define the five review phases (structural, section-by-section, verification of results and citations, cross-section integration, research-integrity and disclosure safety) and what each phase is for.
2. Catalog the AI-generated manuscript patterns that deserve specific scrutiny during the section-level pass.
3. Summarize what published journal, publisher, and funder policy actually requires for AI-assisted manuscripts, since this is a compliance question, not only a quality one.
4. Establish a verifiable test for authorship: what a researcher must be able to do, unaided, before submitting a paper as their own.

What Is This Framework For?

The scope is deliberately narrow. It addresses review of a research manuscript where a human researcher designed the study, performed or directed the analysis, and defined the argument, an AI assistant drafted some portion of the prose, literature summary, or exposition from that material, and the researcher now seeks complete understanding and defensible ownership of the manuscript before submission. It is not a statistics or study-design review guide, and it does not replace a journal's own editorial and peer-review process, an institution's research-integrity office, or a statistical co-author's sign-off; it sits alongside all three.

A thorough review in this context serves five purposes:

1. **Comprehension.** The author can explain every claim and can trace every result back to its source.
2. **Correctness.** The reported results match what the analysis actually produced.
3. **Argumentative soundness.** The discussion follows from the results rather than overstating them.
4. **Consistency.** Terminology, notation, and reported figures agree across the abstract, body, and tables.
5. **Compliance.** The manuscript meets the target journal's specific disclosure requirements for AI-assisted work, not a generic assumption about what disclosure means.

Prerequisites

Before starting, assemble the analysis code and its output exactly as it produced the numbers reported in the manuscript, the target journal's author guidelines and AI-disclosure policy specifically (these vary by publisher, sometimes by journal within a publisher), the conversation log or prompts used during drafting, and every source cited in the literature review, not just the ones the drafting assistant flagged as uncertain. Then set up a working environment: a rendered preview of the manuscript, the analysis re-run from the underlying data in a clean session, and a citation-verification pass that checks each reference against the actual source rather than against the assistant's summary of it.

The Five Review Phases

Phase 1: Structural Review

Begin at the manuscript level, before reading any individual section closely.

Structure fidelity. Compare the drafted section structure against both the target journal's required format (IMRaD or its variant) and the researcher's own intended argument. Sections present that were not part of the intended structure, and sections the argument needs but the draft omits, are structural defects worth flagging before section-level review starts.

Author-contribution accuracy. Confirm the author list and any CRediT or contribution statement accurately reflects who did what, independent of the AI-assistance question; the two are related but distinct compliance obligations.

Disclosure placement. Confirm the manuscript includes an AI-assistance disclosure in the location the target journal actually requires. This is not uniform: some venues want it in the Methods section, some in Acknowledgments, and at least one major publisher requires a dedicated statement immediately above the reference list rather than folded into either. Get this from the current author guidelines, not from memory of a previous submission.

Phase 2: Section-by-Section Review

This is the core of the process. For every section, work through five questions in order:

Correctness. Does every reported number, claim, and citation match its source? Does the section state a conclusion with more certainty than the underlying result supports?

Argumentative alignment. Does the section build the argument the researcher actually intends, or has the assistant introduced a framing, emphasis, or transition that drifts from the intended claim while still reading smoothly?

Result-to-claim traceability. Can every reported result be traced back to a specific line of analysis output? A number that reads plausibly but does not match the actual output is the manuscript equivalent of a hallucinated function call.

Terminology and notation consistency. Does the section use variable names, statistical terms, and abbreviations as defined earlier in the manuscript, rather than introducing a synonymous but locally inconsistent term?

Clarity. Can you explain, in your own words, what the section claims and why the evidence in hand supports it?

For each section, read the underlying analysis output before reading the drafted prose, independently restate what the data show before comparing against the drafted claim, verify that every citation in the section actually supports the point it is attached to, and deliberately question any transition that moves from "the data show" to "this demonstrates" without an intermediate step, since AI-drafted exposition tends to compress that step smoothly.

Patterns specific to AI-generated manuscript prose

A handful of patterns recur often enough in assistant-drafted manuscripts to warrant a dedicated pass:

- **Hallucinated or mismatched citations.** A reference that does not exist, or one that exists but does not say what the manuscript attributes to it.
- **Statistical overstatement.** A result reported as significant, or a trend reported as an effect, beyond what the actual test output supports.
- **Literature-review claims that do not check out.** A confident summary of prior work that reads plausibly but paraphrases the cited source's finding incorrectly, or cites it for a claim the source does not actually make.
- **Discussion drift beyond the results.** A discussion section that argues a broader or more general conclusion than the reported results, on their own, license.
- **Notation drift between Methods and Results.** A variable, parameter, or model term defined one way in Methods and reported under a different name or symbol in Results.
- **Pattern mimicry in the literature review.** A review paragraph that follows the structure of a familiar review sentence convincingly while citing a source that does not actually support the specific claim attached to it.

Phase 3: Verification of Results and Citations

A manuscript that reads cleanly is a starting point, not a conclusion. For every reported number, re-derive it from the analysis output rather than trusting the drafted prose's restatement of it. For every figure and table, confirm it was generated from the same data and code version the manuscript's Methods section describes. For every citation, open the actual source and confirm it says what the manuscript claims it says, since a citation that merely looks correctly formatted is the manuscript equivalent of a test that asserts nothing.

Then check correspondence in both directions: every claim in the Discussion should trace to a specific result reported in Results, and every result reported in Results should be addressed somewhere in the Discussion. A discussion claim with no supporting result is as much a defect as a result the discussion never engages with.

Phase 4: Cross-Section Integration Review

Trace consistency across section boundaries rather than within a single section. Does the Abstract accurately summarize what the body actually reports, rather than a rounder or more favorable version of it? Do the Methods described match the analysis that actually produced the reported Results? Does terminology introduced in the Introduction stay consistent through Methods, Results, and Discussion, rather than drifting toward whatever phrasing the assistant found convenient in each section independently? Verify that every in-text reference to a figure or table number resolves to the correct figure or table, and that the supplementary material, if any, is consistent with the main text's claims about it.

Phase 5: Research-Integrity and Disclosure Safety Review

Examine the manuscript for the exposure classes specific to publication under a researcher's name. Confirm no AI tool is listed as an author or co-author. Every major journal body that has published a position on this, including ICMJE, COPE, WAME, and the policies of Nature/Springer Nature, Science/AAAS, JAMA Network, Elsevier, and PLOS, agrees on this point without exception, on the grounds that authorship requires accountability that an AI tool cannot hold. Confirm the AI-disclosure statement is present, correctly located per the target journal's specific policy, and accurately describes what the tool was used for. Several venues, Science and JAMA Network among them, go further and require the actual prompts used to be disclosed when AI contributed to the manuscript's substantive content, not only its prose. Confirm no data were fabricated or embellished by an AI tool at any stage, since this is treated as a bright line by every policy reviewed for this framework, not a matter of degree. Finally, if the manuscript is tied to a grant application or a funder disclosure, confirm compliance with the funder's own AI policy; NIH, for instance, has stated that applications substantially developed by AI will not be considered the applicant's original work, which is a stricter standard than most journal disclosure requirements.

What Published Policy Actually Requires

Four requirements recur across every source reviewed for this framework, independent of field or publisher, and they are worth stating plainly because they are compliance obligations, not suggestions:

1. **AI is categorically excluded from authorship.** ICMJE, COPE, and WAME each ground this in the same reasoning: authorship requires the capacity to be accountable for the work, to consent to being listed, and to give final approval of the version published, none of which an AI tool can do.
2. **Disclosure is mandatory and location-specific.** Where the disclosure goes is not standardized. Nature/Springer Nature and COPE point to the Methods section; JAMA Network accepts Acknowledgments unless the AI use was part of the formal research design, in which case Methods is required; Elsevier requires a dedicated statement placed immediately above the reference list. Check the target journal's current guidelines rather than reusing language from a previous submission to a different venue.
3. **Human accountability is total, not proportional to the fraction drafted.** Every source reviewed treats AI-assisted content as the named authors' own work for purposes of accuracy, plagiarism, and bias, with no discount for the portion an assistant drafted first.
4. **Verification is an explicit obligation, not an implicit expectation.** ICMJE, WAME, and Science's policy each state directly that authors must review AI-generated content for correctness, bias, and plagiarism before submission; this is the published-policy analogue of Phase 3 above, not a separate requirement invented for this framework.

Basic AI-assisted copy editing (grammar, spelling, reference formatting) is exempted from disclosure by several of these policies, but the exemption is narrow; confirm what the target journal specifically excludes before assuming a given use qualifies.

Annotation and Severity

Maintain structured notes during the review rather than relying on memory. A minimal per-section template:

```
## Section: section_heading

### Status: [Reviewed | Needs Revision | Approved]

### Understanding
[Summary of what the section claims, in your own words]

### Concerns
- [Issue 1]
- [Issue 2]

### Questions
- [Question for further investigation]

### Changes Required
- [ ] Change 1
- [ ] Change 2
```

Classify every finding by severity: **Critical** (a fabricated or mismatched citation, a reported result that does not match the analysis output, a data-fabrication risk, or a missing or misplaced AI-disclosure statement), **Major** (a discussion claim that overstates the supporting result, notation inconsistent across sections, or a broken figure/table cross-reference), **Minor** (an awkward transition, an inconsistent abbreviation, or a formatting slip), or **Enhancement** (a suggestion beyond what the intended argument called for). The severity tier determines the order of remediation, not whether an issue gets fixed at all; a missing disclosure statement is a submission blocker regardless of how minor it feels to fix.

Remediation and Final Verification

Work through issues in severity order: critical issues affecting correctness, integrity, or compliance first, then major issues affecting argumentative soundness or cross-section consistency, then minor issues and enhancements as time allows. For each fix, trace the error to its root (a citation pulled from the wrong source, a number transcribed from a stale analysis run) before rewriting the affected passage, rewrite it yourself rather than asking the assistant to patch it without independently re-verifying the patch, and update any cross-section reference the fix affects.

After remediation, re-read the entire manuscript start to finish as a peer reviewer would, re-verify every citation and reported number one final time against source and output, and review the diff against the last reviewed draft before considering the review closed. Before calling the manuscript submission-ready, run a full consistency pass: confirm the abstract still matches the body after all fixes, confirm every in-text reference resolves, and confirm the AI-disclosure statement is present, accurate, and in the location the target journal's current guidelines specify.

Per-Paper Review Checklist

The five phases above are the reasoning behind this checklist; the checklist itself is what to actually run through for a given manuscript. Copy it per paper rather than trying to hold it in memory across multiple manuscripts in progress at once.

Phase 1: Structural

- ☐ Section structure matches the target journal's required format
- ☐ Author list and contribution statement are accurate, independent of the AI-assistance question
- ☐ AI-disclosure statement is present and located where the target journal's *current* guidelines specify

Phase 2: Section-by-section

- ☐ Every claim in every section has an identifiable evidentiary basis
- ☐ No section states a conclusion with more certainty than its supporting result
- ☐ Terminology, variable names, and abbreviations are consistent with their first definition
- ☐ Every AI-drafted-prose pattern in this post's list (citation mismatch, statistical overstatement, unchecked literature claims, discussion drift, notation drift, pattern mimicry) has been checked for, not just read past

Phase 3: Verification of results and citations

- ☐ Every reported number re-derived from analysis output, not trusted from the drafted prose
- ☐ Every figure and table confirmed to come from the data and code version Methods actually describes
- ☐ Every citation opened at the source and confirmed to say what the manuscript attributes to it
- ☐ Every Discussion claim traced to a specific Results finding, and every Results finding addressed somewhere in Discussion

Phase 4: Cross-section integration

- ☐ Abstract accurately summarizes the body, not a rounder version of it
- ☐ Methods as described match the analysis that produced the reported Results
- ☐ Every in-text figure/table reference resolves to the correct figure or table
- ☐ Supplementary material, if any, is consistent with the main text's claims about it

Phase 5: Research-integrity and disclosure safety

- ☐ No AI tool listed as an author or co-author
- ☐ AI-disclosure statement content matches what was actually used, not a boilerplate copied from a prior submission
- ☐ Prompts disclosed if the target journal requires it (Science and JAMA Network do; most others do not, as of this writing)
- ☐ No data fabricated or embellished by an AI tool at any stage
- ☐ Funder-specific AI policy checked and satisfied, if the work is tied to a grant

Before submission: ownership sign-off

- ☐ Can explain every claim's evidentiary basis without the draft in front of you
- ☐ Can trace every reported number back to the analysis that produced it
- ☐ Can identify which citation supports which specific claim
- ☐ Can predict how a peer reviewer's specific challenge to any result or citation would be answered
- ☐ Every co-author, not only the one who drafted with AI assistance, can stand behind the manuscript's claims

A "no" anywhere in this checklist is a blocker, not a note for later; see Establishing Ownership below for what each unchecked box actually costs if it ships unresolved.

Things to Watch Out For

- **A well-written paragraph that is quietly unsupported.** Clean prose is not evidence of correctness; trace the specific claim back to a specific result.
- **Delegating the fix back to the assistant.** Asking the assistant to correct a flagged citation or number and accepting the correction unread repeats the original problem one level down.
- **Treating a mismatched citation as a rare edge case.** It surfaces often enough in assistant-drafted literature reviews that checking every citation against its source is worth the time, not an optional spot check.
- **Assuming disclosure requirements are the same across journals.** They are not; a disclosure statement copied from a prior submission to a different venue can be placed in the wrong section or omit something the new venue specifically requires.
- **Skipping the "why," not just the "what."** A sentence can be factually accurate and still misaligned with the argument the researcher actually intends to make.
- **Conflating "read it" with "can defend it."** The annotation template's "Understanding" field is not decorative; if you cannot restate the section's claim and its evidentiary basis from memory, it has not actually been reviewed.

Lessons Learned

Conceptual understanding:

- Authorship of a submitted manuscript is a testable and a compliance-bound claim, not a feeling of familiarity with the topic.
- Correctness and argumentative soundness are different axes; a section can pass one while failing the other, most often when a discussion claim outruns its supporting result.
- AI-drafted manuscript prose has a distinct failure signature (citation mismatch, statistical overstatement, literature-review claims that do not check out) that ordinary proofreading does not catch.

Technical:

- Re-deriving every reported number from analysis output, rather than spot-checking a sample, is the only reliable way to confirm the Results section is accurate.

- Opening the actual cited source, rather than trusting a summary of it, is the only reliable way to catch a citation that is correctly formatted but substantively wrong.
- A cross-section terminology log catches notation drift between Methods and Results that no single section's review can, because the inconsistency only exists in the comparison.

Gotchas:

- A reported result can be numerically correct while the surrounding claim describes it with more certainty than the underlying test supports; verify both the number and the claim's hedging.
- A manuscript's disclosure statement can be present and still noncompliant, if it is placed in the wrong section for the target journal or omits a detail (such as the prompts used) that venue specifically requires.
- A citation can exist and be real while still not supporting the specific point it is attached to; verify the claim-to-source match, not just the source's existence.

Limitations

- The framework is scoped to manuscript prose and its relationship to an underlying analysis; it does not cover study design review, statistical methodology review, or peer review itself, each of which is its own discipline.
- It assumes the human author already has the domain and statistical fluency to evaluate correctness and argumentative soundness; it is not a substitute for a qualified co-author or statistical reviewer's expertise.
- Published journal and publisher AI-disclosure policy is an actively moving target; the specific requirements summarized in this post reflect the sources current as of this writing and should be re-checked against the target journal's current guidelines before submission, not assumed to be stable.
- The ownership test in the final section is self-administered and is original synthesis, not a restatement of any single published authorship policy; it provides no external verification, such as a co-author's independent sign-off, that the answers given are accurate.

Opportunities for Improvement

1. A companion checklist scaled down for short communications or correspondence, where the full five-phase process is disproportionate to the manuscript's length.
2. A per-journal disclosure-requirement reference table, since the location and content of the required AI-disclosure statement varies by publisher and this framework can only summarize the pattern, not substitute for checking current guidelines.
3. Guidance on reviewing an AI-assisted *revision* to an already-submitted or already-published manuscript, as distinct from reviewing a manuscript drafted from scratch.
4. Worked severity-classification examples specific to manuscripts, since "Major" versus "Critical" boundary cases (a discussion claim that overstates a result versus a citation that does not exist) are where reviewers disagree most.

5. An extension of the research-integrity phase for multi-author manuscripts, where AI assistance may vary by section and by co-author, and the disclosure obligation applies to the manuscript as a whole.

Establishing Ownership

Before submitting a manuscript, verify that you can, without referring to the draft: explain every claim's evidentiary basis, trace every reported number back to the analysis that produced it, identify which citation supports which specific claim, and predict how a peer reviewer's specific challenge to any result or citation would be answered. Separately, confirm compliance readiness: does the manuscript's AI-disclosure statement match what the target journal's current guidelines actually require, is the author list and any contribution statement accurate independent of the AI-assistance question, and can every co-author, not only the one who drafted with AI assistance, stand behind the manuscript's claims?

This test is original synthesis, built for this framework rather than drawn from any single published source; no journal, publisher, or research-integrity body reviewed for this post specifies a procedural test of this kind. What those sources do establish, and what this test is built to satisfy, is the underlying obligation: ICMJE, COPE, and WAME each require that a named author be able to take accountability for the entire work, and accountability that has not been tested against a reviewer's likely challenge is accountability claimed, not demonstrated.

Consider also documenting the drafting process itself beyond the required disclosure statement: what review process was undertaken, which sections received AI assistance, and a version-control or document-revision history showing the verification work. This exceeds what any policy reviewed for this framework requires, but it gives a co-author, a future collaborator on a follow-up study, or a research-integrity inquiry the context this review process assumed from the start.

Review does not end at submission. Reviewer comments that request revised or additional analysis should go through the same verification before the revised manuscript is resubmitted, and a manuscript that sits in revision for an extended period is worth re-verifying against its own prior claims before resubmission, since an author's understanding of the work, and occasionally the underlying data, can shift in the interim. Knowledge decays without use: keeping review notes accessible and re-reading the full manuscript before responding to reviewers are what keep the authorship claim true through the review cycle, not just true at the moment of first submission.

Wrapping Up

What Did We Learn?

Thorough review of an AI-assisted research manuscript requires the same rigor a careful peer reviewer would bring to any submission, plus a specific sensitivity to the failure modes assistant-drafted academic prose tends to produce, and a compliance obligation neither of the other posts in this series carries in the same form. The process demands more than a read-through; it requires achieving genuine, defensible understanding of every claim, citation, and reported number before accepting responsibility for it under the author's name. The five-phase structure (structural, section-by- section, verification of results and citations, cross-section integration, research-integrity and disclosure safety) gives that process a checklist rather than leaving it to instinct, and the ownership test at the end gives it a stopping condition: review is complete when the answers to the ownership questions are honestly yes and the disclosure statement matches the target journal's actual requirements, not when the researcher is tired of reading the draft.

Main takeaways:

- Clean prose is a floor, not a target; verify every citation and reported number independently, do not just read past them.
- Published journal and publisher policy is unanimous that AI cannot be an author and that disclosure is mandatory; treat this as a compliance requirement to check against the current target-journal guidelines, not a stable fact to memorize once.
- Authorship is demonstrated by the ability to defend every claim against a reviewer's specific challenge, not claimed by completing a checklist.

If you are trying this yourself: start with the structural review and the target journal's current AI-disclosure guidelines before reading any section closely, keep a running citation-verification log open alongside the annotation template, and do not consider a manuscript submission-ready until you can honestly answer every question in the Establishing Ownership section, including tracing every reported number back to its source.

See Also

This post extends a framework first written for a different medium: Code Review for AI-Assisted R Package Development applies the same five-phase structure and ownership test to R package code, Review Methodology for AI-Assisted Blog Post Drafting applies it to blog-length prose, and Review Methodology for AI-Assisted Textbook Drafting applies it to book-length, pedagogically structured material. This post adds the phase and requirements specific to academic manuscripts: verification against published results and citations, and compliance with journal, publisher, and funder AI-disclosure policy.

Key resources:

- ICMJE Recommendations, updated Jan. 2024 — International Committee of Medical Journal Editors
- COPE Position Statement on Authorship and AI Tools — Committee on Publication Ethics
- WAME Recommendations on Chatbots and Generative AI — World Association of Medical Editors
- Nature Portfolio AI editorial policy
- Springer Nature AI guidance for researchers
- Science/AAAS policy on generative AI and large language models
- JAMA Network AI authorship policy — AMA Style Insider
- Elsevier Generative AI Policies for Publishing
- NIH Notice NOT-OD-23-149 on generative AI in peer review

Reproducibility

Source document: adapted from `analysis/report/index.qmd` in the sibling posts `rp-code-review-methodology`, `rp-blog-review-methodology`, and `rp-book-review-methodology` (`~/prj/rgtlab/posts/`). This post transposes and extends that framework's five-phase structure and ownership test to academic research manuscripts, and grounds the research-integrity phase and the "What Published Policy Actually Requires" section in published journal, publisher, and funder policy current as of this writing; those policies should be re-verified against the target journal's current guidelines before relying on them for a live submission.

Session information:

```
R version 4.6.1 (2026-06-24)
Platform: aarch64-apple-darwin25.4.0
Running under: macOS Tahoe 26.6.2

Matrix products: default
BLAS:   /opt/homebrew/Cellar/openblas/0.3.34/lib/libopenblasp-r0.3.34.dylib
LAPACK: /opt/homebrew/Cellar/r/4.6.1/lib/R/lib/libRlapack.dylib;  LAPACK version 3.12.1

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

time zone: America/Los_Angeles
tzcode source: internal

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

loaded via a namespace (and not attached):
 [1] compiler_4.6.1  fastmap_1.2.0   cli_3.6.6       tools_4.6.1
 [5] htmltools_0.5.9 parallel_4.6.1  otel_0.2.0      yaml_2.3.12
 [9] rmarkdown_2.31  knitr_1.51      jsonlite_2.0.0  xfun_0.58
[13] digest_0.6.39   rlang_1.3.0     evaluate_1.0.5
```

Rendered on 2026-08-21 at 12:00 PDT. Source: ~/prj/rgtlab/posts/rp-paper-review-methodology/analysis/report/in

---

Let's Connect

Questions, corrections, or a different view on where this framework is too strict or too loose, or on whether a cited policy has since changed, are welcome in the comment thread below.

- **GitHub**: rgt47
- **Email**: Contact form