TITLE: Concise guidelines and complementary checklists for improving research reliability and reproducibility

Jason Pither*[1,2,3] & Mathew Vis-Dunbar[4,5]

[1] Okanagan Institute for Biodiversity, Resilience, and Ecosystem Services, University of British Columbia, Okanagan campus, Kelowna, British Columbia, Canada

[2] Department of Biology, University of British Columbia, Okanagan campus, Kelowna, British Columbia, Canada

[3] https://orcid.org/0000-0002-7490-6839

[4] Library, University of British Columbia, Okanagan campus, Kelowna, British Columbia, Canada

[5] https://orcid.org/0000-0001-6541-9660

* Correspondence: jason.pither@ubc.ca

Standfirst

Checklists to address reproducibility shortfalls have proliferated but tend to be discipline- or research stage-specific and lack context or learning opportunities. We offer an alternative: one page, every stage, with a reason for each recommendation.

Keywords: checklist, reproducibility, transparency, research reliability, credibility, replicability, guidelines, R package

Introduction

Building literacy and skills to enhance research reliability and reproducibility is often hindered by lack of institutional support or training and otherwise takes second stage to maximizing the quantity and "novelty" of research outputs. The consequences of this have been repeatedly highlighted in the literature, across many, if not most scientific disciplines [1–4]: a) insufficient documentation (studies lacking data, scripts, context, i.e. complex analytical processing decisions, etc.), resulting in an inability to identify questionable research practices and thus measure bias; b) inadequate study design, including insufficient replication or inappropriate statistical analyses limiting the reliability of conclusions; c) an overreliance on binary criteria (i.e. $P < 0.05$), combined with inadequate reporting of effect sizes, confidence intervals, and uncertainty around estimates; d) poor differentiation across the nexus of exploratory (hypothesis generating) and confirmatory (hypothesis testing) research, resulting in unreliable statistical interpretations; d) a body of literature that is severely biased and misrepresents the full scope of research being conducted [5].

The gaps

Among the myriad resources available to researchers to strengthen or evaluate the reliability and reproducibility of outputs, checklists have emerged as popular tools. Checklists offer an accessible means to standardize reporting and evaluation, asking binary or scaled questions about procedures within the research lifecycle. Checklists have been developed across a range of applications [e.g. 6–9], generally tailored to specific disciplines [e.g. 10] and study designs [e.g. clinical trials; 11,12], or research life cycle phases such as preregistration (https://www.cos.io/initiatives/prereg) or archiving data [13] and code [14]. A few are more comprehensive and include interactive dashboards [e.g. 9], but remain limited in scope and

prescriptive in nature. Specifically, they aren't designed to help the user understand the problem being addressed by the checkbox. Thus, we note three shortfalls in the current landscape of checklists: a) to our knowledge, none are sufficiently discipline-agnostic (within the sciences); (b) none address the full research life cycle; and c) while they contribute to evaluating reporting, they fall short on supporting critical interpretation.

The solution

To address these shortfalls, we present a resource, entitled "Concise Guidelines for Producing Reliable, Reproducible Research", that attempts to fill the gap between checklists and the well-developed literature that underpins what these checklists are attempting to evaluate. We anticipate it being especially useful for graduate students and early career researchers as it synthesizes common issues across the sciences and links out to core papers and resources through a shared public library (https://www.zotero.org/groups/6487727/concise_guidelines_for_reliable_research/library).

The need for such a resource became clear to us through the work we have been engaged in to enhance Open Science practices at the University of British Columbia's Okanagan campus [15] and beyond [16], and reflects discussions with graduate students and early career researchers through teaching, supervision, and project collaboration.

In conjunction with a companion R package (called "rrchecklist" for "reliable research" checklist [17]) that provides fillable checklists in HTML or PDF formats, we provide a proof-of-concept that merges the contextual resource with a checklist. Combined with the Concise Guidelines document, this resource enhances the ability of researchers to apply a critical lens to their interpretation of research outputs (their own and others) while directing users to a concise library of foundational resources for deeper learning.

What it does

The resource adopts a full research lifecycle approach, starting with the evaluation of existing published research, enabling identification of the best available evidence to inform a new research project. It then builds on this, mapping the concepts used in this evaluation to one's own study design considerations, through to one's own reporting of outcomes. We see this as an important feature; academia trains aspiring researchers well to be critical of others' works, but is less effective at encouraging the application of a similarly critical lens to one's own work.

Another key feature of the resource is being prescriptive in what to consider, but not in how to consider it, i.e. one should look for hallmarks of poor study design and selective reporting, one should be able to distinguish between exploratory and confirmatory studies and appropriate reporting in each, etc. As such, and when partnered with a checklist that identifies specific criteria to look for in evaluating study design, the resource provides valuable informational context. It encourages the user to recall their own foundational training around study design, interpretation, and reporting – thereby resurrecting considerations that are otherwise often abandoned in favour of convention (e.g. not considering statistical power, emphasizing statistical significance over real world significance).

In its treatment of study design and statistical concepts, the resource implies a frequentist rather than Bayesian framework. We consider this more broadly applicable for most researchers – especially those at early career stages. We would welcome the development of a Bayesian version of the guidelines.

The resource also addresses ethical community engagement and encourages collaboration with those potentially impacted by the research [18], highlighting the importance of adhering to community-specific protocols. Additionally, it addresses key aspects of research integrity and

scholarly communication, ensuring due credit and maximizing discoverability, accessibility, and reach.

What it does not do

While the resource provides links to a few key references that serve as foundations for the user to build off, it does not provide a comprehensive rationale or an extended list of references for any of its recommendations or statements.

The resource does not provide introductory or background material on the reproducibility crisis, study design, or statistics. Rather, it assumes the user has sufficient introductory knowledge of these topics such that most, if not all, of the terminology and statements are familiar to them.

Conclusion

Both the guidelines document and the checklist package are meant to be adapted and enhanced; in spite of an attempt to create a discipline-agnostic resource for the sciences, we recognize limitations resulting from our own experiences and anticipate that we have used language and references that will not resonate with all potential audiences. Releasing the guidelines document in multiple formats under CC BY-NC, we hope it will generate discussion and modification, and be adopted and improved upon by many research labs and graduate programs. Further, the R ‘’rrchecklist’ package is licensed GPLv2 to encourage similar customized adoption.

References

1. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JPA, et al. Biomedical research: increasing value, reducing waste. The Lancet. 2014;383: 101–104. doi:10.1016/S0140-6736(13)62329-6

2. Purgar M, Klanjscek T, Culina A. Quantifying research waste in ecology. Nat Ecol Evol. 2022;6: 1390–1397. doi:10.1038/s41559-022-01820-0

3. Nosek BA, Hardwicke TE, Moshontz H, Allard A, Corker KS, Dreber A, et al. Replicability, Robustness, and Reproducibility in Psychological Science. Annu Rev Psychol. 2022;73: 719–748. doi:10.1146/annurev-psych-020821-114157

4. Errington TM, Mathur M, Soderberg CK, Denis A, Perfito N, Iorns E, et al. Investigating the replicability of preclinical cancer biology. eLife. 2021;10: e71601. doi:10.7554/eLife.71601

5. Ioannidis JPA. Transparency, bias, and reproducibility across science: a meta-research view. Journal of Clinical Investigation. 2024;134: e181923. doi:10.1172/JCI181923

6. Parker TH, Griffith SC, Bronstein JL, Fidler F, Foster S, Fraser H, et al. Empowering peer reviewers with a checklist to improve transparency. Nature Ecology & Evolution. 2018;2: 929–935. doi:10.1038/s41559-018-0545-z

7. Grant S, Corker KS, Mellor DT, Stewart SLK, Cashin AG, Lagisz M, et al. TOP 2025: An Update to the Transparency and Openness Promotion Guidelines. MetaArXiv; 2025. doi:10.31222/osf.io/nmfs6_v2

8. Wicherts JM, Veldkamp CLS, Augusteijn HEM, Bakker M, van Aert RCM, van Assen MALM. Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. Front Psychol. 2016;7. doi:10.3389/fpsyg.2016.01832

9. Aczel B, Szaszi B, Sarafoglou A, Kekecs Z, Kucharský Š, Benjamin D, et al. A consensus-based transparency checklist. Nat Hum Behav. 2019;4: 4–6. doi:10.1038/s41562-019-0772-6

10. Ekhtiari H, Zare-Bidoky M, Sangchooli A, Valyan A, Abi-Dargham A, Cannon DM, et al. Reporting checklists in neuroimaging: promoting transparency, replicability, and reproducibility. Neuropsychopharmacol. 2025;50: 67–84. doi:10.1038/s41386-024-01973-5

11. Hopewell S, Chan A-W, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 statement: updated guideline for reporting randomized trials. Nat Med. 2025;31: 1776–1783. doi:10.1038/s41591-025-03635-5

12. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019; l4898. doi:10.1136/bmj.l4898

13. Jones S, Grootveld M. How Fair Are Your Data? Zenodo; 2017. doi:10.5281/ZENODO.1065991

14. Australian Research Data Commons. FAIR software checklist. In: FAIR software checklist [Internet]. 2025. Available: https://ardc.edu.au/resource/fair-software-checklist/

15. Hanna S, Pither J, Vis-Dunbar M. Implementation of an Open Science Instruction Program for Undergraduates. Data Intelligence. 2021;3: 150–161. doi:10.1162/dint_a_00086

16. Srivastava D, Hunt D, Binning S, Emry S, Bledsoe E, Reemeyer J, et al. Rescuing the Past to Prepare for the Future: Environmental Data Rescue as a Key Activity of Open Science. BISS. 2025;9: e180327. doi:10.3897/biss.9.180327

17. Pither J. rrchecklist: an R package providing markdown templates for checklists to support research assessment and the production of reliable research. 2026. Available: https://github.com/pitherj/rrchecklist

18. Hoekstra F, Mrklas KJ, Khan M, McKay RC, Vis-Dunbar M, Sibley KM, et al. A review of reviews on principles, strategies, outcomes and impacts of research partnerships approaches: a first step in synthesising the research partnership literature. Health Res Policy Sys. 2020;18: 51. doi:10.1186/s12961-020-0544-9

A. Evaluating Existing Research

• Be skeptical but not cynical when reading published research. Poor study design and selective reporting are common features that, along with inadequate statistical power and many forms of bias, decrease the reliability and replicability of much published research. • Be wary of studies that (i) are published in predatory journals; (ii) emphasize statistical over biological significance, or *P*-values over effect sizes; (iii) report results from analyses that were unplanned or not reproducible; (iv) draw causal inferences from non-experimental research without sufficient evidence; (v) reference primarily supportive works while downplaying/ignoring contrary findings; or (vi) fail to discuss study limitations or alternative interpretations of their findings. These points should also inform **B-D.** • Distinguish exploratory (hypothesis generating) from confirmatory (hypothesis testing) research. Given its crucial role, exploratory research (ER) should be common and discoverable, but research culture favours confirmatory research (CR), causing much ER to be disguised as CR, raising the risks of hindsight bias and Hypothesizing After Results are Known (HARKing), and decreasing research reliability. • Pilot studies often blend ER with CR, and commonly deviate from their intended use, reducing their reliability. Interpret them carefully. • P-values are most meaningful in pre-registered CR addressing plausible hypotheses, and least meaningful in ER disguised as CR. • Consider conducting a systematic review of published research, including critically evaluating risk of bias following standardized protocols.

B. Planning Your Research

• Consider research partnerships from the outset Ideally, those who could benefit from or be impacted by the research would be directly involved through respectful collaboration. • Complete ethics approvals and initiate a data management plan (DMP) prior to any data collection, ensuring ongoing respect for and adherence to community-specific engagement protocols, data sovereignty policies, and data management procedures, as per FAIR and CARE guidelines. • Consider using version-controlled reporting systems. • Keep track of all contributors and their contributions from the outset; consider using Dragon Kill Points, or the MeRIT or CRediT frameworks for reporting contributions to project outputs. Encourage all contributors to acquire an ORCID to facilitate discoverability and accurate attribution. • Specify whether the research is ER or CR (or includes both), and provide rationale(s). • Clearly define questions and/or hypotheses, and devise an appropriately designed observational or experimental study; unless subjects/units are randomly assigned to treatments, the study is observational or quasi-experimental. • Ethical and/or logistical constraints often necessitate observational or quasi-experimental designs, in which case causal inference is more challenging (compared to experiments) but possible, requiring careful pre-planning, and meeting myriad criteria and assumptions. • For experimental studies, ensure proper controls and randomization, and implement blinding where possible. • For all studies, define the target population, scope of inference, dependent and independent variables, units of observation, and replication (sample sizes). • Choose statistical methods best suited to addressing the question(s) / hypotheses; new methods are rarely required. • Multiple alternative data pre-processing and/or analysis methods may legitimately be suitable, so consider evaluating the robustness of the findings to these varied protocols (i.e. a form of "multiverse" or sensitivity analysis). • Plan to accommodate non-Gaussian error distributions, multivariate responses and/or hierarchical/non-independent sampling designs where appropriate. • Specify how model assumptions will be checked and violations dealt with. • For CR, clearly define what would constitute a meaningful effect size, and specify what will constitute evidence (i) consistent with and (ii) contrary to each hypothesis.• For CR, consider prospective power analyses that employ the planned statistical methods and plausible effect sizes. • Especially when power is limited, acknowledge uncertainty and focus on sound study design and data useability. • For CR, plan to adjust *P*-values for multiple testing when appropriate. • For ER, justify sample sizes based on feasibility, resource constraints, or desired precision. • Especially for CR, strongly consider detailing your study design and analysis plan in a pre-registration (PR), which can help improve research reliability. PRs shift effort appropriately towards the planning stage, where it has the greatest benefit. Deviations from PRs are absolutely okay if transparently reported (see section C). • Have your research plan reviewed by both subject matter experts and others (possibly as a registered report). • Foster a peer community that values providing and receiving constructive criticism.

C. Conducting Your Research

• Careful planning (part B) helps make the "conducting research" stage smoother. • Document procedures transparently (ideally with versioning), detailing all deviations from planned protocols. Data quality control and updates to data management plans are ongoing responsibilities.

D. Reporting, Interpreting, and Communicating Your Research Findings

• Transparently report all evaluations of model assumptions, and remedies to violations, as per analysis plans. • Prioritize reporting and interpreting effect sizes, their uncertainty (e.g. confidence intervals), and their real-world relevance. • When appropriate, report effect sizes alongside *P*-values. • Absence of evidence (i.e. a statistically non-significant result) is not necessarily evidence of absence. • Statistical significance does not necessarily imply biological relevance. • Do not imply causation with words like "cause" or "affect" unless appropriately justified (e.g. from experimental results); consider whether "is associated with" is more suitable. • Use effective visualizations: reveal, don't conceal data. • Clearly report sample sizes for all treatments, each analysis, and in each relevant figure and table. • Maximize the accessibility and inclusivity of outputs. • There should be no surprise analyses: each analysis should be linked to a pre-specified question/hypothesis. • Adding unplanned analyses is sometimes warranted (e.g. as a follow-up to planned), but these tend to yield inflated false positive rates (e.g. due to researcher degrees of freedom), so label them clearly as unplanned; only planned analyses retain the intended statistical relevance. • Interpret findings objectively and discuss alternative interpretations where appropriate. • Avoid using persuasive language, overselling, or over-generalizing your findings. • Do not extrapolate findings beyond the target population without sufficient justification. • Speculation can be valuable when labeled clearly as such. • Evaluate your draft manuscript against points in section "**A**" above.

E. Enhancing the Replicability, Discoverability, Accessibility, and Reach of Your Research

• Ensure data and executable analysis code are as open as possible and as closed as necessary according to FAIR and CARE guidelines, archived with appropriate metadata, and shared to facilitate independent replication. • Publish Open Access and/or a preprint, and follow best practices to maximize discoverability when writing your title, abstract, and keywords.

These are guidelines not requirements: every step taken towards more reliable, reproducible research is valuable and should be celebrated.