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

Abstract

Checklists to improve research reliability and reproducibility have proliferated, but most remain discipline-specific, narrow in scope, or lack context and learning opportunities. Here we present "Concise guidelines for improving research reliability and reproducibility": a one-page resource that addresses all stages of the research lifecycle, is discipline-agnostic, and provides context for each item, supported via hyperlinks by a public library of open-access resources. Licensed CC BY-NC, the resource is designed for adaptation across disciplines and career stages, but is geared towards researchers-in-training. We additionally provide a companion R package ('rrchecklist') that generates fillable checklists, also customizable under a GPLv2 license.

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 a lack of institutional support or training, or 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]: 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.

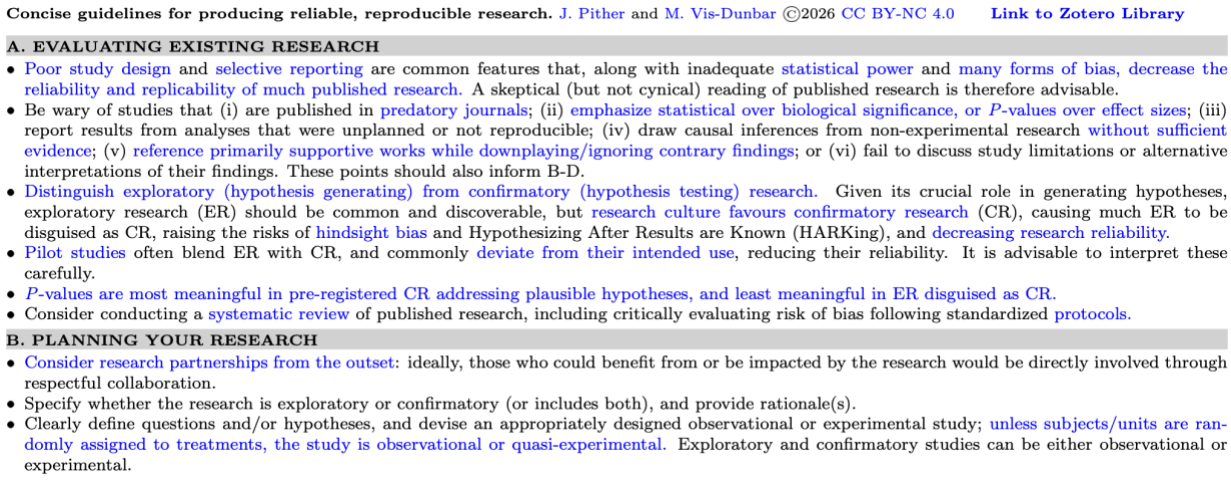
The gaps

Among the myriad resources available to researchers to strengthen or evaluate the reliability and reproducibility of outputs, checklists have emerged as popular tools. Checklists have been developed across a range of applications [2–4], generally tailored to specific disciplines [5] and study designs [6,7], or research life cycle phases such as preregistration (https://www.cos.io/initiatives/prereg) or archiving data [8] and code [9]. A few are more comprehensive and include interactive dashboards [e.g. ref. 3]. However, these checklists remain limited in scope, prescriptive in nature, and oriented towards researchers who are either already well-versed in the reproducibility crisis and its causes, or are motivated to apply a checklist

regardless of their level of understanding of the underlying issues. These characteristics make checklists less directly useful to researchers-in-training (including undergraduate and graduate students), who, in our experience, are keen to learn about the proximate and ultimate causes of the replicability crisis, and gain motivation to apply best practices when they see the extent of the problem. 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 (including the step of evaluating existing research); and c) while they contribute to evaluating reporting, they fall short on supporting learning and critical interpretation.

The solution

To address these shortfalls, we present a resource, entitled "Concise Guidelines for Producing Reliable, Reproducible Research" (Figure 1), that attempts to fill the gap between checklists and the well-developed literature that underpins what these checklists are attempting to evaluate. We anticipate the resource being especially useful for graduate students and early career researchers (ECRs) more generally 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).



Concise guidelines for producing reliable, reproducible research. J. Pither and M. Vis-Dunbar ©2026 CC BY-NC 4.0 **Link to Zotero Library**

A. EVALUATING EXISTING 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. A skeptical (but not cynical) reading of published research is therefore advisable.
- 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 in generating hypotheses, 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. It is advisable to interpret these 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.
- Specify whether the research is exploratory or confirmatory (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. Exploratory and confirmatory studies can be either observational or experimental.

Figure 1: A screenshot showing the beginning portion of the 1-page guidelines document rendered to PDF from LaTeX. In the upper right is the link to the public Zotero library that hosts all the citations that are hyperlinked in the document (blue text). The document is also provided in Microsoft Word format, LaTeX, and Markdown, and the CC BY-NC license enables users to tailor and reformat as desired. The companion 'rrchecklist' R package, licensed GPLv2, provides users the option of creating more conventional checklists of action items based on these guidelines; one checklist for evaluating existing research and one for when conducting one's own research.

The need for such a resource became clear to us through the work we have been engaged in over the past decade to enhance Open Science practices at the University of British Columbia's Okanagan campus [10] and beyond [11], and reflects discussions with graduate students and ECRs through teaching, supervision, and project collaboration. For instance, the guidelines directly address common points of confusion for many researchers, especially graduate students, such as the distinction between exploratory and confirmatory research, and the fact that both types of research can employ either observational or experimental study designs. We also sought and received valuable feedback from colleagues from an array of disciplines (economics, social sciences, statistics, ecology, neuroscience, metaresearch).

What it does

The resource adopts a full research lifecycle approach, starting with the evaluation of existing published research, thereby 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 that it is prescriptive in what to consider, but not in how to consider it. For example, one should look for hallmarks of poor study design and selective reporting, and one should be able to distinguish between exploratory and confirmatory studies and appropriate reporting in each. But it is left to the researcher to gauge the contextual impact of the potential sources of bias, the quality or thoroughness of reporting, and the potential misapplication of methods or interpretations of the data. When partnered with the complementary checklist (from the 'rrchecklist' package) 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, such as not considering statistical power, and emphasizing statistical significance over real world significance.

In contrast with a recently published checklist [4], our guidelines (i) deliberately emphasize the important role that that exploratory research plays, (ii) underscore the distinction between it and confirmatory research at the outset, and (iii) revisit this distinction in the multiple stages of the research lifecycle where it has implications (e.g. for the interpretation of *P*-values). As mentioned above, the exploratory-confirmatory distinction is a common source of confusion for researchers-in-training and ECRs, likely due in part to the emphasis that conventional training and instruction often places on hypothesis-driven (confirmatory) research (at the

111 expense of instruction around exploratory research), and an unfounded favouring of confirmatory
112 over exploratory research more generally (cf. [14]).

113 Another example of a distinction that carries implications through multiple stages of the
114 research lifecycle, and that is often a source of confusion, is that of observational versus
115 experimental study designs. In the "Planning your research" section we describe what
116 distinguishes experimental from observational research, then, in the "Reporting, interpreting, and
117 communicating your research findings" section we highlight how experimental versus
118 observational studies warrant different terminology when describing outcomes (e.g. "affect"
119 versus "is associated with"). Highlighting such connections across stages of the research
120 lifecycle provides an example of how our guidelines offer learning opportunities that other
121 checklists tend to overlook.

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

126 The resource also addresses ethical community engagement and encourages collaboration
127 with those potentially impacted by the research [12], highlighting the importance of adhering to
128 community-specific protocols. Additionally, it addresses key aspects of research integrity and
129 scholarly communication, ensuring due credit and maximizing discoverability, accessibility, and
130 reach.

131 Our guidelines come with a companion R package (called "rrchecklist" for "reliable
132 research" checklist [13]) that provides fillable checklists in HTML or PDF formats. Together,
133 these resources can enhance the ability of researchers to apply a critical lens to their

interpretation of research outputs (their own and others), and to build literacy by linking to a concise library of foundational resources for deeper learning.

What it does not do

While the resource provides links to some key references that serve as foundations for the user to build on, it does not provide a comprehensive rationale or an extended list of references for any of its recommendations or statements. We anticipate the publicly-accessible Zotero library serving as a starting point for further learning.

The resource does not provide introductory or background material on the reproducibility crisis, study design, or statistics. Rather, it (i) 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, and (ii) provides links to external resources that provide relevant background material.

Conclusion

Both the guidelines document and the checklist R package are meant to be adapted and enhanced; despite our 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 (Microsoft Word, PDF, Markdown, LaTeX), 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.

Supplemental Materials

All supplemental materials including the guidelines document (in various formats) are publicly available at this OSF archive: https://doi.org/10.17605/OSF.IO/DQX8K, and the R package is available at https://github.com/pitherj/rrchecklist.

Acknowledgments

For helpful feedback on the guidelines and/or manuscript, we thank Diane Srivastava, Adam Ford, Guillaume Blanchet, Malgorzata Lagisz, Trudy Kavanagh, Ross Hickey, John Thompson, Phil Barker, Shirley Chau, and members of the Biodiversity and Landscape Ecology Research Facility. We also thank the University of British Columbia for financially supporting our Open Science efforts at both university campuses, through the following internal funding opportunities: Excellence Fund (JP & MV-D), Open Educational Resources (JP), ALT 2040 (JP, and Curricular Innovation Awards (JP).

Contributions

JP wrote the initial draft of the guidelines resource, and MV-D wrote the initial draft of the manuscript. Both authors helped with revisions on both documents. JP produced the 'rrchecklist' R package with assistance from Claude Code (Sonnet 4.6). JP used Claude Code (Sonnet 4.6) to generate the initial draft LaTeX and Markdown versions of the guidelines document.

References

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

2. 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

3. 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

4. Banzi R, Varga M, Gelsleichter YA, Vinatier C, Moher D, Naudet F, et al. An international consensus on core reproducibility items in research. PLoS Biol. 2026;24: e3003726. doi:10.1371/journal.pbio.3003726

5. 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

6. 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

7. 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

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

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

10. 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

11. 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

12. 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

13. 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

14. Fife DA, Rodgers JL. Understanding the exploratory/confirmatory data analysis continuum: Moving beyond the “replication crisis”. American Psychologist. 2022;77: 453–466. doi:10.1037/amp0000886

Concise guidelines for producing reliable, reproducible research. J. Pither and M. Vis-Dunbar **Link to Zotero Library**

A. EVALUATING EXISTING 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. A skeptical (but not cynical) reading of published research is therefore advisable.
- 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 in generating hypotheses, 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. It is advisable to interpret these 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.
- Specify whether the research is exploratory or confirmatory (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. Exploratory and confirmatory studies can be either observational or 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 & independent variables, units of observation, and replication (sample sizes).
- Describe study materials and procedures in detail.
- 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 preregistration, which can help improve research reliability. Preregistrations shift effort appropriately towards the planning stage, where it has the greatest benefit. Deviations from preregistrations 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.
- Complete ethics approvals and initiate a data management plan 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.
- Plan to use version-controlled reporting systems if possible.
- 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.

C. CONDUCTING YOUR RESEARCH

- Careful planning (part B) helps make the "conducting research" stage smoother.
- Document procedures transparently, 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 exploratory analyses to planned confirmatory analyses is sometimes warranted, but these tend to yield inflated false positive rates due to researcher degrees of freedom, so label them clearly as exploratory; only planned confirmatory analyses retain the intended statistical meaning.
- 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 or data access and sharing agreements, 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.