



Reproducibility and Registration in Sexuality Research

Tierney K. Lorenz¹

Published online: 6 February 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Most readers of the *Archives of Sexual Behavior* will, by now, be familiar with an upsurge of interest in best practices for reproducibility and replicability across modern scientific disciplines (Ioannidis, 2005; National Academies of Sciences Engineering and Medicine, 2019; Simmons, Nelson, & Simonsohn, 2011). Concerns about the basic validity of findings that cannot be replicated—and the widespread practices and incentive systems that lead to unverifiable results—are particularly important for social and behavioral scientists, whose fields of study are relatively younger and whose work is often interpreted as more subjective than other sciences (Lilienfeld, 2012).

As sexuality research continues to mature as an established field of study with its own methods, theories, and bodies of literature, we too must address growing concerns regarding our commitment to scientific rigor. The “replication crisis” has been covered at length in other settings; in this Guest Editorial, I will outline some specific concerns with regards to improving open science practices in sexuality research. Additionally, I will describe some best practices recommended to researchers looking to submit to *Archives*, and for reviewers at *Archives* to consider when recommending manuscripts for publication. As *Archives* receives many submissions using qualitative methods, I have made special efforts to include consideration of how reproducibility and replicability apply to these scholars as well. My intention is to start a broader discussion about methodical standards across the *Archives*’ readers, authors, and reviewers; as such, I welcome critiques and commentary.

Reproducibility and Replicability

Although often used together, the terms “reproducibility” and “replicability” refer to different constructs. Reproducibility occurs at the study level: A study is reproducible when its data can be verified to produce the same results when using the same analytic method. For example, the same statistical output may be reproduced using the same code shared across two laboratory groups. In sex research, issues with reproducibility most typically arise when there is poor access to the original data or lack of transparency in reporting on the analytic method (rather than being a function of complex processes introducing non-systematic error, as in experimental physics). While the need for reproducibility applies primarily to quantitative research with computational analytic methods, many of the efforts to improve reproducibility may benefit qualitative researchers as well: Creating open-source materials and clearly documenting the rationales for analytic decisions is equally important for all types of scholarly effort.

Replicability occurs at the finding level: A finding is replicable when multiple independent research efforts investigating the same scientific question all arrive at the same pattern of results. This is typically interpreted to mean finding the same patterns of results using the same methods in a new sample (National Academies of Sciences Engineering and Medicine, 2019). However, this narrow interpretation somewhat misses *why* scientists care about replicability: namely, that a replicable finding is more likely to reflect a fundamental truth that is robust to observer, sample, *and* method. As such, the means of establishing replicability should include not only reports of duplications of previous methods, but also meta-analysis and meta-synthesis (Slavin, 1995), Bayesian or cumulative analytic frameworks (Braver, Thoemmes, & Rosenthal, 2014), and multi-level or multimodal analyses. Again, although qualitative researchers are often left out of these sorts of discussions, their contributions are critical as a variety of methods are needed to establishing true replicability.

✉ Tierney K. Lorenz
Tierney.lorenz@unl.edu

¹ Department of Psychology, and Center for Brain, Biology and Behavior, University of Nebraska–Lincoln, Lincoln, NE 68588, USA

Confirmatory versus Exploratory Research and Preregistration

One of the fundamental issues in the replication crisis is the confusion—often unintentional, occasionally deliberate—between confirmatory and exploratory research. Both types of research are vital to the scientific enterprise, but standards and practices for one can differ in ways that would critically undermine the validity of the other. Exploratory research aims to methodically observe a phenomenon, to contribute to model building, and speculate about possible causal relationships between variables; as such, exploratory work can (and should) follow a wandering path that leads researchers through a variety of models. In contrast, confirmatory research aims to test pre-specified hypotheses in order to make causal claims about the mechanisms of an effect (Nosek, Ebersole, DeHaven, & Mellor, 2018).

In sexuality research, which is a maturing but still developing field, both forms of analysis are valuable (and should be considered equally appropriate for publication in top journals). *Archives* does not share the derogatory view that exploratory research is intrinsically less systematic (and thus less “science-y”) than confirmatory work, or that only the very best work “deserves” a confirmatory label (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012)—these are distinctions not of quality but aim. Nevertheless, researchers *must* be transparent about which analyses are confirmatory and which exploratory, and interpret their findings accordingly.

One means of verifying the conditions necessary for confirmatory research is through study preregistration. Preregistration is a process by which researchers clearly describe the details of their research plan, including methods and analysis decisions, *prior to conducting the research*. A key element of preregistration is the publication or cataloging of the (preliminary) research plan in a publicly accessible format that provides time-stamped evidence of when the plan was established. Of note, both quantitative and qualitative work can be preregistered (see <https://osf.io/j7ghv/> for guidelines for preregistering qualitative research). A variety of databases have been established to make preregistration accessible (and free); a few commonly used by sex researchers are ClinicalTrials.gov (particularly for biomedical or intervention research), the Open Science Framework (osf.io), and aspredicted.org. Each allows research teams to specify study design and analysis/interpretation plans and clarify which components of the study are intended as exploratory or confirmatory.

There are many benefits of preregistration, even aside from the obvious benefits to replicability. The process of pre-specifying research design can clarify one’s own thought process and highlight likely decision points that

may benefit from forethought rather than happening on the fly as they arise. If one chooses not to embargo a research plan (more on this below), one can get feedback from other researchers on design at an earlier juncture—when one might actually be able to make good on that feedback, rather than during the journal review process when there is nothing one can do. Creating a date and time-stamped proof of one’s hypotheses not only provides evidence necessary for establishing the conditions for confirmatory research, it also gives one leverage to convince Editors of originality if one is scooped. At a broader level, preregistration helps to reduce the false positive rate, which is all the more important for researchers working with small samples (as is often unavoidable when studying sexual or gender minority groups). And finally, given sex research is often treated as a “niche” area (at best), the respect afforded to preregistered studies doesn’t hurt.

There are costs to preregistration worth considering. One concern that we can quickly lay to rest is the preconception that preregistering irrevocably locks one into a design or analysis plan: If the original plan does not fit the final data, it is acceptable to describe whatever modifications were necessary and the rationale behind them (DeHaven, 2017). For example, if the actual data collected have a different distribution than anticipated that make them inappropriate to model using the originally planned methods, it is perfectly fine to apply whatever corrections are needed to more appropriately model those data—so long as one acknowledges this is a (needed) deviation from the original plan.

Another concern, which is harder to address, is that preregistration does take time and there is a sharp learning curve. This time investment can be somewhat softened by having students assist (and gain valuable experience in research design). It is time well spent, as the process of registering one’s research design can help hone one’s own understanding of the project—and potentially highlight issues that can be addressed before they become problems. For example, in a recent preregistration, because I was forced to think about how I will handle missing data for a survey design with randomized question blocks, I realized that I needed to change how the survey software would mark questions the participant saw but chose not to answer (i.e., missing not at random) versus did not see at all (missing at random).

An issue that is unique to sex research is the possibility that preregistration could open our research efforts up to unwanted scrutiny, which could interfere with our ability to conduct the work. Many of us work in cultural settings that do not view sex research kindly or study topics that are considered taboo. All of us have at some point been the subject of raised eyebrows; some of us are the subject of targeted campaigns of stalking and harassment. Although (thankfully) rare, there are periodically efforts to shut down our ongoing research. I know of several such cases in fellow

sex researchers: people who called in false complaints to the IRB, groups that posted fake ads on social media to confuse potential participants, and even protesters who sat outside laboratories to discourage participants from entering. Pre-registering a study could make this harassment all the easier by clearly and publicly describing one's plans. Luckily, this concern can be circumvented by putting an embargo on the preregistration until the primary data collection is completed (see here for a guide on managing embargoes on OSF preregistrations: <https://bit.ly/2WFCsqf>).

The specifics of what needs to be entered into a preregistration plan will depend on the study question and methods: An intrinsically iterative design (such as scale development) may require more flexibility and thus fewer pre-defined parameters. But at minimum, a good preregistration will include: (1) succinct, clear definitions of the independent and dependent variables, including their operationalization (see <http://datacolada.org/64> for excellent guidelines here); (2) clear description of the sample to be collected (including recruitment strategies, sample size, and stopping rules); and (3) for any confirmatory analyses, directional a priori hypotheses with the tests/models planned to address those hypotheses. Description of all statistical tests to be performed should include pre-specified decision points for when one cares about the results of those tests (e.g., effect sizes over threshold X , a p value under threshold Y , degree of variance explained, etc.). As a reviewer of *Archives*, I have seen many adirectional statements of group-wise differences presented as hypotheses, so let's be very clear: "We expected gender/sex differences in variable X " is not a hypothesis—it is a topic sentence and a senseless one at that. There can be a sex/gender difference if the mean is higher in men, in women, or in gender non-binary people; if one group has a bimodal distribution and another unimodal; if one group has a larger variance than others; if there are more outliers in one group than others; and on and on. Don't present weak "hypotheses" when you mean you *explored* group differences—again, exploration is just as important as confirmation.

Issues with Data Analysis and Data Sharing

Preregistration is not necessary in all research designs (e.g., for exploratory work), but transparency in one's research decisions and open sharing of data and materials will always be critical for good science.

At an interdisciplinary journal like *Archives*, reviewers can be asked to review manuscripts that are within their topical, but not methodological, expertise and thus may not always know what details are critical for future replication efforts. This makes it all the more critical that as part of our review process we ensure that authors make their data and materials publicly available, in a freely accessible format with good

metadata and syntax (see here for an excellent guide to writing metadata: <https://bit.ly/2KMbohW>). There are many free repositories that support a wide variety of data formats, such as OSF (which can be linked to your preregistration!), Dataverse.org, and the ICPSR (<https://www.icpsr.umich.edu/>). Not only is this good scientific practice, it benefits the researcher directly as others use (and cite) their data. Along the same lines, *Archives'* interdisciplinarity necessitates our authors to be proactive in transparency of our "researcher degrees of freedom" (Simmons et al., 2011): disclosing *all* measures collected and tests performed, being consistent with one's rationales for analytic decisions (e.g., keeping the same covariates across models), and presenting evidence of robustness of the finding when analytic decisions are manipulated (e.g., when an outlier is included vs. excluded).

An issue that is increasingly common in sex research is the use of datasets that cannot be openly shared—either because the data are extremely sensitive (such as sexual networks of individuals with HIV/AIDS) or the data are proprietary (such as from collaborations with corporations). In these cases, best practices include: engaging in strong de-identification processes (such as those available through <https://amnesia.openaire.eu>); making data available but under limited circumstances (e.g., through a data sharing agreement); making summary data available (e.g., at the group level, rather than individual level); statistically altering the original data in ways that reduce risk of identifying individuals, but do not change their analysis (e.g., standardizing raw data); and being as detailed as possible in describing data collection from proprietary sources, particularly if those sources change over time (e.g., Web sites whose algorithms are constantly updated). For this last point, this includes sharing both the details of data collection (e.g., the version of the data used and/or the methods for scraping public data) and analysis scripts (e.g., code used to construct relevant variables).

One final minor note on transparency issues specific to sex research: I urge those of us who regularly use psychophysiological assessments of sexual response to either use published methods for data cleaning (e.g., Prause, Williams, & Bosworth, 2010; Pulverman, Meston, & Hixon, 2018) or to fully document the many, many decisions that occur when cleaning and condensing psychophysiology data and presenting these as an "Appendix". I get it—I am also guilty of not reporting everything, because it's tedious and only a few reviewers are ever going to raise a stink about it—but our subfield will not move forward unless we have methodological consensus.

Interpretations and Biases

Sometimes lost among discussion of registration plans and open-source materials is the need to address interpretations and biases in our research. Here too, sexuality research must

strive to establish best practices in transparency and rigor in appropriately contextualizing our findings and considering how our research practices and biases contribute to our results.

The National Academy of Sciences, Engineers, and Medicine's (2019) *Report on Reproducibility and Replicability in Science* highlights the importance of qualified interpretation of individual study results, noting that no one study is definitive and that the strongest claims should be reserved for the strongest evidence. This is particularly true for those areas of research pertaining to sensitive topics (i.e., practically all of sex research) or for which new findings have important implications for policy decisions (again, much of sex research) (van Anders et al., 2017). This is an area where we sometimes—perhaps often—fail. By virtue of being sex researchers, we can become accustomed to thinking and communicating about very sensitive sexual matters; thus, we may fail to recognize when a claim is particularly bold and requires extraordinary evidence. Or, more rarely, we deliberately over-interpret the evidence to bring public attention to an issue. But even if that issue legitimately deserves greater attention, no one is well served by undermining the public perception of objectivity in sex research. Instead, we should consider the potential audiences for our work—including the lay public and policymakers—and make appropriately cautious interpretations of any individual finding. As sexuality research continues to grow, there is greater potential for meta-analyses, systematic reviews, and other forms of research synthesis for which stronger claims can be justified.

Finally, we must consider how our biases and assumptions may impact the replicability of our findings by changing how the reporting on methods is interpreted across scholars. In an interdisciplinary journal such as *Archives*, which has an international audience, these sorts of errors are sadly typical. As a reviewer, I have often noted assumptions of national or cultural context, such as describing a sampling pool in “a large Midwestern university” without naming the country where that university is found. And many authors use discipline-specific terminology (or worse, terms that have different meanings across disciplines): For example, I have seen the term “panel data” referring to either cross-sectional data collected by a large nationally representative panel and to longitudinal data with multiple measurements of the same respondent. At a minimum, I advise authors to scan their manuscripts for such assumptions; having a colleague from another discipline, or a student who is new to the field, read through a draft can be helpful in this effort.

At a broader level, we should be mindful of how more complex biases that influence the conduct of sex research overall: How we operationalize our definitions, what variables we choose to consider or not, and how we design

inclusion/exclusion criteria may influence the replicability of research across contexts. Here, we may look to the fields of gender/sexuality studies as leaders in critical analysis of sexological methods (see, e.g., Barker, 2016). In 2014, neuroscientists at McGill University made headlines by showing that measures of stress and pain in rodent models were systematically skewed by the sex of the experimenter handling the rodents (Sorge et al., 2014). Prior to that report, no one would have even thought to record the sex of the experimenter, let alone account for its effects on replicability of findings across studies. When we encounter a similar “failure to replicate,” we should take it as a call to interrogate our assumptions about what factors might influence our findings (National Academies of Sciences Engineering and Medicine, 2019). If one has documented the hell out of one's methods and been open with one's data, a truly unexpected finding—including failure to replicate—should spark not shame, but excitement: that is the heart of discovery.

Summary

If we are to continue to advance as a field, sexuality researchers must address concerns about reproducibility and replicability. As such, I strongly recommend researchers wishing to publish at the *Archives of Sexual Behavior* consider preregistration of their research designs. Regardless of whether or not formal preregistration is feasible (or necessary given the research aim), authors should be as transparent and open with all aspects of the research enterprise as possible, including (1) highlighting which are exploratory versus confirmatory analyses, (2) making available all (de-identified) materials and data in public repositories with metadata, (3) clearly describing decisions regarding data cleaning and statistical testing, (4) appropriately qualifying interpretations of results from individual studies, and (5) proactively seeking out and addressing one's own biases in the conduct of sexuality research. I welcome commentary from the authors and readership of the *Archives* regarding these issues in the form of a Letter to the Editor.

Appendix

The following are intended to assist reviewers with determining how well a manuscript meets standards for data transparency, reproducibility, and replicability. None of the criteria below should be treated as absolute requirements (as some will only apply to certain kinds of studies or papers) but the answers to these questions should guide the reviewers' ultimate decisions

and recommendations for authors. Reviewers are encouraged to familiarize themselves with the current best practices in reproducibility and replicability (e.g., as outlined in recent reports from the National Academies of Sciences, Engineering and Medicine [<http://sites.nationalacademies.org/sites/reproducibility-in-science/index.htm>]).

Does the manuscript describe results from a pre-registered research plan?

We encourage authors to pre-register research plans wherever possible and reasonable for the research question. Most empirical research designs that collect quantitative data to test hypotheses can be pre-registered. Some kinds of qualitative research may also be appropriate for pre-registration (e.g., as described here [<https://osf.io/j7ghv/>]). However, certain kinds of manuscripts do not require pre-registration (e.g., purely exploratory analyses; narrative research reviews). Any manuscript that reports on confirmatory hypothesis testing must present either pre-registration, or some other evidence that all hypotheses and plans for data collection and statistical testing were, in fact, determined prior to the research being conducted.

If yes, do the authors clearly identify:

- a. Was the plan registered, with a link to the registration?
- b. Any deviations from this plan, with rationale for those deviations?
- c. Which analyses are intended to be confirmatory hypothesis tests, and which analyses were exploratory/unplanned?

If no, do the authors clearly identify:

1. Which analyses are intended to be confirmatory hypothesis tests, and which analyses were exploratory/unplanned? (See note above)
2. All decisions regarding data cleaning and statistical testing?

Some issues that may be relevant to answering this question:

- a. Are rules about including/excluding outliers defined and well-defended?
- b. Is there a *theory-driven* reason given for each covariate, and are covariates included consistently across models?
- c. Is there a description of *all* tests performed (which may or may not match the final models presented in the paper)?
- d. Is there a description of *all* data collected during this study (not just that which is reported in this specific report)?

Does the manuscript describe where research data, materials, syntax for analyses, codebooks, and other relevant materials necessary for replication can be accessed?

We encourage authors to make these materials accessible wherever possible. If there is some reason the data or materials cannot be made available, authors should be asked to describe this reason within the manuscript. Also, even in cases where the data cannot be made available for ethical or proprietary reasons, other materials such as survey batteries, stimuli, syntax and code, or qualitative codebooks could still be presented. Note that “accessible” means either open access (e.g., through a publicly accessible repository such as the ICPSR [<https://www.icpsr.umich.edu/icpsrweb/>] or OSF) or as supplementary materials attached to the manuscript—simply stating “data are available from the authors on request” is not considered to meet the standards for accessibility.

Do the authors consider data (or perspectives) from diverse populations, as relevant to that topic?

The best research includes data that are truly representative. Do the authors include participants (or perspectives) from different cultures, racial or ethnic backgrounds, developmental stages, socioeconomic statuses, ability levels, sexes, gender identities or sexual orientations, and so on? If not, do the authors acknowledge this as a significant limitation?

Do the authors consider issues of potential research bias, as relevant to that topic?

Are field-specific terms defined for an interdisciplinary readership? Are convenience samples appropriately identified as such (e.g., undergraduate participant pools) with appropriate qualification in the interpretation of results? Do the authors make a good-faith effort to consider their findings within alternative frameworks, and include contrasting points of view?

Are the interpretations of results appropriately qualified given the nature of the data?

Individual studies are rarely conclusive. Meta-synthesis of multiple studies across a range of research designs (and researchers) offers greater confidence and may allow stronger conclusions. But related to the point above, even meta-analytic results must be appropriately qualified if they do not include data from a diverse range of people (e.g., the only data available were collected from Western countries).

References

- Barker, M. J. (2016). *Queer: A graphic history*. London: Icon Books.
- Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science*, 9(3), 333–342.
- DeHaven, A. (2017). *Preregistration: A plan, not a prison*. Retrieved October 29, 2019 from <https://cos.io/blog/preregistration-plan-not-prison/>.
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>.
- Lilienfeld, S. O. (2012). Public skepticism of psychology: Why many people perceive the study of human behavior as unscientific. *American Psychologist*, 67(2), 111–129.
- National Academies of Sciences Engineering and Medicine. (2019). *Reproducibility and replicability in science*. Washington, DC: The National Academies Press.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606.
- Prause, N., Williams, K., & Bosworth, K. (2010). Wavelet denoising of vaginal pulse amplitude. *Psychophysiology*, 47(2), 393–401.
- Pulverman, C. S., Meston, C. M., & Hixon, J. G. (2018). Automated artifact-detection procedure for vaginal photoplethysmography. *Journal of Sex and Marital Therapy*, 44(6), 566–590.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366.
- Slavin, R. E. (1995). Best evidence synthesis: An intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology*, 48(1), 9–18.
- Sorge, R. E., Martin, L. J., Isbester, K. A., Sotocinal, S. G., Rosen, S., Tuttle, A. H., ... Leger, P. (2014). Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nature Methods*, 11, 629–632.
- van Anders, S. M., Schudson, Z. C., Abed, E. C., Beischel, W. J., Dibble, E. R., Gunther, O. D., ... Silver, E. R. (2017). Biological sex, gender, and public policy. *Policy Insights from the Behavioral and Brain Sciences*, 4(2), 194–201.
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7(6), 632–638.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.