

The preregistration process

In a project, the researcher (user) preregisters their study design before data collection and analysis. After completing the preregistration, the user executes the project and produces a final paper. Both the paper and the preregistration are available to the reviewers, editors and other readers, so that they can *check* if the protocol has been followed.

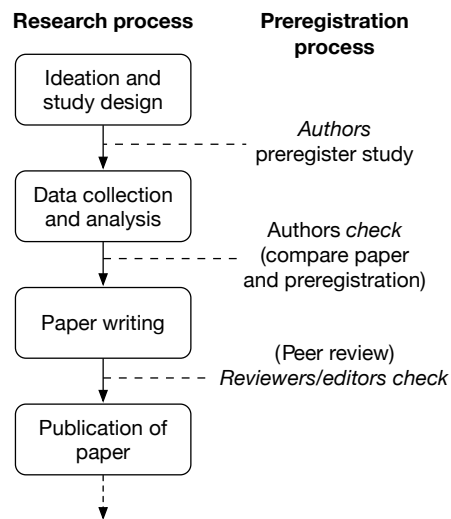


Figure 1: Preregistration steps adapted from Center for Open Science at cos.io/rr.

Designing for Preregistration: A User-Centered Perspective

Xiaoying Pu

Computer Science and Engineering,
University of Michigan
Ann Arbor, MI, USA
xpu@umich.edu

Matthew Kay

School of Information, University of Michigan
Ann Arbor, MI, US
mjskay@umich.edu

Licheng Zhu

School of Information, University of Michigan
Ann Arbor, MI, US
zlicheng@umich.edu

Frederick Conrad

Institute for Social Research, University of Michigan
Ann Arbor, MI, US
fconrad@umich.edu

ABSTRACT

The replication crisis—a failure to replicate foundational studies—has sparked a conversation in psychology, HCI, and beyond about scientific reliability. To address the crisis, researchers increasingly adopt preregistration: the practice of documenting research plans before conducting a study. Done properly, preregistration should reduce bias from taking exploratory findings as confirmatory. It is crucial to treat preregistration, often an online form/template, as a user-centered design problem to ensure preregistration achieves its intended goal. To understand preregistration in practice, we conducted 14 semi-structured interviews with preregistration users (researchers) who ranged in seniority and experience. We identified two main purposes researchers have for using preregistration,

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5971-9/19/05.

<https://doi.org/10.1145/3290607.3312862>

Preregistration templates and registries

Studies are typically preregistered using web templates submitted to online registries, which support archiving and anonymous links for peer review. Two common templates are hosted at the Open Science Foundation (OSF; osf.io) and AsPredicted (aspredicted.org). These templates vary in length and content. Besides the general-purpose sites, different research fields have their own preregistration sites, such as ClinicalTrials.gov for medical research.

Table 1: Interviewee demographics.

Title	
PhD student	4
Assistant professor	3
Associate professor	3
Professor	4
Discipline	
Psychology	11
HCI	2
Behavioral Econ	1
Gender	
Female	7
Male	7
Have written a preregistration?	
Yes	13
No	1
Have reviewed/edited a preregistration?	
Yes	7
No	6
Unclear	1

in addition to different user roles and adoption barriers. With the ultimate goal of improving the reliability of scientific findings, we suggest opportunities to explicitly support the different aspects of preregistration use based on our findings.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; *Interaction design*.

KEYWORDS

Preregistration; User-centered Design; Open Science.

ACM Reference Format:

Xiaoying Pu, Licheng Zhu, Matthew Kay, and Frederick Conrad. 2019. Designing for Preregistration: A User-Centered Perspective. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI'19 Extended Abstracts)*, May 4–9, 2019, Glasgow, Scotland Uk. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3290607.3312862>

INTRODUCTION

A “replication crisis” has shaken fields such as psychological science, where many foundational studies have failed to be independently replicated [7]. Preregistration has been proposed as a *tool* to address the replication crisis and to improve the scientific process [11]. One definition of preregistration is to specify “the research questions and analysis plan before observing the research outcomes” [5]. Overall, preregistration distinguishes preregistered, confirmatory findings from exploratory ones, constraining the researchers’ flexibility in data collection and analysis, the very flexibility that contributed to the replication crisis. The Sidebars on page 1 and 2 show the preregistration process and common web-based templates. Preregistration adoption has increased rapidly in recent years. The number of preregistrations in the OSF (Open Science Foundation) registry alone grew from 38 in 2012 to 18,000 in 2018 [6].

However, it is unclear how well preregistration is working in practice. Several cases have emerged in psychology where prominent researchers have investigated published, preregistered papers and found that the papers did not follow their preregistrations or did not report significant deviations from protocol [8, 10]. Given the importance and rapid adoption of preregistration, we approach preregistration as a *user-centered design problem*: does the current implementation of preregistration in practice—the design of preregistration forms, how preregistration integrates into research practice and peer review, etc.—effectively support its intended goals?

To our knowledge, we are the first to tackle the user-centered design problem around preregistration. Therefore, we ground our understanding with qualitative interviews of 14 preregistration users—primarily academic psychologists, whose field is at the forefront of preregistration adoption. The

interviews revealed 1) two perceived purposes for preregistration, 2) different user roles, and 3) adoption barriers. Our findings can help explain the current preregistration tool design and use, the inadequacy of which might undermine the overarching goal of preregistration: improving the reliability of scientific findings. We conclude by making design suggestions based on our findings, and aim to initiate conversations about user-centered designs for preregistration.

Terms in interview results.

Flexibility: Researcher’s freedom in selecting sample size, covariates, and experimental conditions, *etc.*, potentially leading to desirable conclusions [11].

Transparency: Includes production transparency (open access data or data collection procedures) and analytic transparency (“a full account of how they draw their analytic conclusions from the data”) [3].

Power analysis: The calculation of the sample size needed to detect an effect of a certain size with a given probability, under the significance testing framework.

Exploratory studies: Studies where hypotheses are found in the data by post-hoc theorizing [11].

Confirmatory studies: The study components are preregistered such that there is no room for exploration in data collection and data analyses, if applicable [11].

Delimiting flexibility: P10 on excluding power analysis

“So I feel like people are confusing disclosure, which often you want more the better, but I think the preregistration should be reserved for things that it’s really important to know ahead of time... power analysis is not relevant for preregistration”

INTERVIEW STUDY PROTOCOLS

To understand the use of preregistration in research, writing, and the peer-review process, we conducted fourteen ($N = 14$) semi-structured interviews. Participants all had self-reported experience with preregistration and were from eight universities in the United States, recruited through word-of-mouth and at open science workshops in psychology. Table 1 shows more details about participant demographics. We primarily recruited in psychology because these researchers have increasingly adopted, advocated for, and extensively discussed preregistration [5, 11]. This user group should thus inform us about how preregistration is used in practice.

The interview sessions lasted 43 minutes on average ($min = 23, max = 57, sd = 11.2$). The interview questions targeted different roles in which a participant might encounter preregistration; e.g., as an author of a preregistration or as an editor/reviewer of a preregistered manuscript. We also asked more general questions about preregistration, such as motivations to adopt preregistration and opinions about preregistration formats. Two independent coders inductively developed 485 codes (affinity notes) from the interview transcripts and summarized themes collaboratively with affinity diagrams, following the protocol described in [2].

INTERVIEW FINDINGS

We list some of the terms that appeared in interviews in the Sidebar on page 3.

Purposes of preregistration

Two purposes of preregistration, *delimiting flexibility* and *increasing transparency*, emerged from our interviews. It is important to note that the two purposes are not mutually exclusive: explicitly *delimiting* (or stating) *flexibility* is in a way being transparent. That said, the preference of one purpose over another can have tangible effects on the design of preregistration, see Discussion section.

1. *Delimiting Flexibility:* We found that preregistration can be viewed to delimit, mark, or contain the choices a researcher can make throughout a study. Endorsing this purpose, P9 said that preregistration was “a way to manage flexibility in analyses”. P10 called flexibility “decision points” and emphasized that preregistration should only contain time-sensitive decisions, see Sidebar. To state the sample size is to delimit flexibility because the sample size is needed to *decide when to stop* data collection in a study.

Increasing transparency: P6 on including power analysis

“I think [statistical] power is one of the biggest mistakes that are made in psychology, so not asking people to actually prove that they have the power of the test is a mistake...”

¹Explanation at <https://sometimesimwrong.typepad.com/wrong/2015/06/why-p-048-should-be-rare-and-why-this-feels-counterintuitive.html>

P6 on the OSF template: “I can’t go into the OSF and register a correlation study. We should be able to pre-register anything. And the fact that things are set up really for experiments makes it difficult because you’re like, “Well, that doesn’t really fit what I’m doing.” And it can turn you off so then you just stop, and so it’s turning off people from using it.”

However, P10 considered power analysis irrelevant as it justifies the sample size, a time-insensitive justification that should stay valid whether data have been collected or not.

2. *Increasing Transparency:* Lupia and Elman define *transparency* as open-access data or data collection procedures and “a full account of” analyses details [3]. Different from participants describing *delimiting flexibility* as a purpose, P6 thought power analysis (sample size rationale) should be part of preregistration (see Sidebar). This may reflect a desire to use preregistration to force researchers to “prove” they are following good practices by making deviations from good practices more *transparent*.

Checking preregistration

We roughly define *checking* as comparing a manuscript to its preregistration and verifying that the data collection and analyses are conducted as the preregistration has specified. Authors, reviewers, and editors may all engage in checking. When discussing the paper *authoring* process, participants described their checking as *linear and thorough*, probably because they would take full responsibility for their papers. P7 said that while writing a paper, they “went back to [the preregistration] multiple times”. P14 compared their paper and the preregistration “point-by-point”, likening the checking process to doing journal paper revisions. When acting as *reviewers and editors*, on the other hand, participants did not describe an exhaustive checking process. Instead, they reported heuristics to *query* a preregistration. One way to query the preregistration was to check “when something looks out of place” (P6), e.g., an unlikely “ p -value = 0.48”¹. When asked why they did not check thoroughly, P8 attributed it to trust: “if somebody said that [they disclosed all flexibilities] in their paper, I would trust that, and I wouldn’t necessarily look to the preregistration.”

Barriers to adopting and using preregistration

Some researchers, especially those who conduct non-experimental studies, met barriers to preregistration. For example, P5 primarily conducted secondary data analysis and had concerns that their idea might be scooped from a preregistration, because anyone can replicate the secondary analysis if the data are publicly available. P6, who did not work in the experimental, hypothesis testing framework, complained about how they could not use the OSF template readily, see Sidebar. Understanding preregistration or open science concepts (see Sidebar on page 3) is also a conceptual barrier to understanding what preregistration is for and how to use it.

DISCUSSION

Designing for multiple purposes

We speculate that the current preregistration templates reflect what their designers consider to be the primary *purpose* of preregistration. For example, if the AsPredict template (see Sidebar on page 2)

is primarily for *delimiting flexibility*, it would explain why its designers required *Sample Size* but not *Sample Size Rationale*. By contrast, the OSF template prompts for this rationale, perhaps because its designers wanted to maximize *transparency*. To support these purposes, a preregistration template could be dynamic, allowing authors or reviewers to adjust the template view (hide and show questions) to their purpose of choice. In addition, an explicit statement of purpose could be especially useful to new users. Without such guidance (as now), new users experience confusion about which questions are important and “what the right level of nuance is”, as P13 put it.

Designing for different user roles explicitly

Since an author and a reviewer/editor use preregistration differently, we propose designing preregistration tools targeting these user roles separately.

To aid *authors’* checking “point-by-point” and to improve report quality, we propose to better integrate preregistration into their existing workflow. Interactive tools for experimental design and/or analysis (for example, TouchStone [4], which aids in experimental design, or StatSplorer [9], which aids in analysis), could have support for preregistration explicitly designed into them. Such a tool could aid the user throughout the research process while keeping track of any preregistration protocol deviations. It can even give suggested texts for writing up results with clear indications of what is exploratory/confirmatory and where protocol deviations occurred.

For *reviewers/editors*, their checking by *query* might miss discrepancies between the paper and preregistration, leading to publication with flawed preregistration. Existing preregistration formats are designed mainly for authors to input study information quickly, verging on being write-only media. Thus, we recommend exploring designs that make relevant preregistration content easier to *query*. For example, Hardwicke’s SMART pre-registration [1] proposes adding tooltips containing pre-registration content into the relevant sections of the online version of a published paper. Such paper-to-preregistration lookup (and vice versa) could improve preregistration checking.

Designing for different user disciplines to overcome adoption barriers

One barrier to preregistration adoption is the lack of templates that support other non-experimental ways of knowing. This barrier can be addressed by community-based approaches, letting researchers design for themselves. As an example, at the Society for the Improvement of Psychological Science conference in 2018, a group of psychologists collaboratively drafted a preregistration template² for secondary data analysis, a kind of research previously not compatible with existing templates.

Implications for the CHI community

Though our main contribution is in improving our understanding of users to inform preregistration design, we do acknowledge the interests in replicability and rigor in general at CHI. Thus, we make

²Available at <https://osf.io/bpuw3/>

recommendations to preregistration design in CHI given our diverse ways of knowing: 1) since experimental HCI draws upon methods from psychology, it thus encounters the same statistical issues that motivate preregistration, so our findings are likely applicable to experimental HCI; 2) for other non-experimental studies, such as qualitative, design, or engineering work, we wish to initiate discussions at CHI about whether adapting/adopting preregistration can help improve research rigor.

CONCLUSION

In this work, we investigated preregistration practices as a user-centered design problem. We uncovered two main purposes—increasing transparency and delimiting flexibility, which may explain the differences in preregistration template design. We suggest that future preregistration tool design make these purposes explicit, and explore ways to support preregistration continually across the research pipeline (from experimental design, to analysis, to write-up, to paper review). We also invite discussions about designing preregistration tools for different user roles (authors and reviewers/editors) and user backgrounds (non-experimental paradigms).

REFERENCES

- [1] Tom Hardwicke. 2018. SMART Preregistration. Accessed on Jan 2, 2019. https://twitter.com/Tom_Hardwicke/status/1010325387488866310.
- [2] Andrés Lucero. 2015. Using affinity diagrams to evaluate interactive prototypes. In *Human-Computer Interaction*. Springer, 231–248.
- [3] Arthur Lupia and Colin Elman. 2014. Openness in political science: Data access and research transparency. *PS - Political Science and Politics* 47, 1 (2014), 19–42. <https://doi.org/10.1017/S1049096513001716>
- [4] Wendy E Mackay, Caroline Appert, Michel Beaudouin-Lafon, Olivier Chapuis, Yangzhou Du, Jean-Daniel Fekete, and Yves Guiard. 2007. Touchstone: exploratory design of experiments. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1425–1434.
- [5] Brian A. Nosek, Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. 2018. The preregistration revolution. *Proceedings of the National Academy of Sciences* 2017 (2018), 201708274. <https://doi.org/10.1073/pnas.1708274114>
- [6] Brian A Nosek and D Stephen Lindsay. 2018. Preregistration becoming the norm in psychological science. *APS Observer* 31, 30/3 (2018).
- [7] Harold Pashler and Eric Jan Wagenmakers. 2012. Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? *Perspectives on Psychological Science* 7, 6 (2012), 528–530. <https://doi.org/10.1177/1745691612465253>
- [8] Dan Simons. 2015. HI-BAR: A gold standard brain training study? <http://blog.dansimons.com/2015/11/hi-bar-gold-standard-brain-training.html>
- [9] Chat Wacharamanotham, Krishna Subramanian, Sarah Theres Völkel, and Jan Borchers. 2015. Statsplor: Guiding novices in statistical analysis. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2693–2702.
- [10] Eric-Jan Wagenmakers and Quentin Gronau. 2018. Musings on Preregistration: The Case of the Facial Feedback Effect.
- [11] Eric-Jan Wagenmakers, Ruud Wetzels, Denny Borsboom, Han L. J. van der Maas, and Rogier A. Kievit. 2012. An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science* 7, 6 (2012), 632–638.