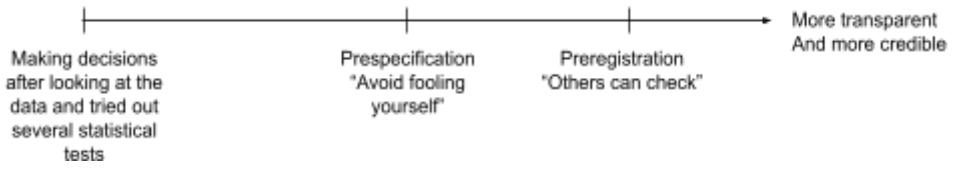
	Name	Email address
Current contact person	Steve Haroz	steve.haroz@gmail.com
Initial draft author(s)	Steve Haroz & Chat Wacharamanotham	steve.haroz@gmail.com chat@ifi.uzh.ch
Next actions: [ ] rearrange to focus and start with preregistration - break into prespecification and registration [ ] Chat add draft of this table [ ] revise the overall structure to What, Why, How [ ] overall structure ready for inner-circle review [ ] selected sections are READY-FOR-REVIEW (note it in the document when ready) [ ] ready to port to RMarkdown [ ] ported to RMarkdown at …(URL)… and ready for the reviewing process		

Preregistration: prespecifying and registering research plans

Placeholder for the "Experiment and analysis planning" FAQ which will be ported to RMarkdown here

Authors: Steve Haroz <steve.haroz@gmail.com>, Pierre Dragicevic <pierre.dragice@gmail.com>,

, Matthew Kay <matthew.kay@gmail.com> <please list anyone who has contributed.>
* Emails are used only for collaborated writing (comments/TODO). They will be removed in the final version of the guideline



[General comments]

TODO: restructure this chapter to “What” “Why” “How”

TODO: more citations for each statement

What is prespecification?

A prespecification is a description of a study and/or analysis prior to seeing the data. An aspect of a study is prespecified (or planned[1]) when it is decided ahead of time, typically before looking at the data or before running the study [cumming2013]. Because researchers are susceptible to memory slips and cognitive biases (Nosek2018 The preregistration revolution), an oral discussion during a meeting or an undocumented personal decision does not count as a prespecification. A prespecification is at minimum recorded on a document that the research team can refer to during the course of their research.

Recently, multiple sites have made it possible and to immutably and permanently host a time-stamped prespecification on freely accessible sites. According to the Transparent Statistics principles of Process Transparency and of Material Availability, there is now no reason to prespecify without preregistering. Without preregistration, a prespecificified study is no more reliable than a study that is not prespecified.

What is preregistration?

Preregistration is a form of prespecification where … (Steve?)

- Verifiable timestamp
- In a public registry
- Todo: copy/reword from COS

Preregister or Pre-register? Hyphen or no?

No-hyphen is less ambiguous, but use whatever floats your boat. Just be consistent.

Why prespecify?

> *The best way to avoid all of the biases Simmons et al. identified is to specify and commit to full details of a study in advance. Research falls on a spectrum, from such fully prespecified studies, which provide the most convincing results and must be reported, to free exploration of data, results of which might be intriguing but must—if reported at all—be identified as speculation, possibly cherry-picked.*
>
> *Confirmatory and exploratory are terms that are widely used to refer to research at the ends of this spectrum. Confirmatory, however, might imply that a dichotomous yes/no answer is expected and suggest, wrongly, that a fully planned study cannot simply ask a question (e.g., How effective is the new procedure?). I therefore prefer the terms* ***prespecified*** *and* ***exploratory****. An alternative is question answering and question formulating. [Cumming2013]*

[1] Also used in statistical “planned contrasts”

Conducting and reporting a study or analysis requires many decisions. For readers and reviewers to accurately interpret how those decisions impact results and conclusions, it should be clear whether the decisions were made before, during, or after inspecting the data. Because different modes of research assume either a data-led, inductive process or a data-independent, deductive hypothesis test, reported research must reflect this implied distinction to be accurately interpreted. If one mode of research veers into another without acknowledgement, it could diminish the credibility of reported results and conclusions, and it could cause the applicability of those conclusions to be misinterpreted. A preregistration helps mitigate these potential problems by differentiating what decisions were made before and after inspecting data.

Imagine you want to know if a coin is evenly weighted. So you have 50 people each flip a coin 10 times. Out of those 50, one person gets heads 9/10 times. You then only report the result of that one person without mentioning that you only decided to report it after seeing the results of 49 other attempts. If anyone tries to replicate the experiment, they'd likely get a different result from your statistical fluke. If, however, you preregister

What is hidden flexibility?

Hidden flexibility (also called researcher's degree of freedom, ….) refers to ….

Table 1 in Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking provides an excellent overview of the flexibility (or degrees of freedom) that can occur in planning, running, and reporting a study.

What kind of papers can include a preregistration?

Preregistration does not just apply to human-subjects experiments. Preregistration can be used for many research methods, though it is perhaps best suited for inferential studies where a sample is being used to draw a conclusion to a wider population. Any research where data is collected and analyzed as evidence for a reported result can be preregistered. Examples include scraping, categorizing, or coding data. Generally, any result (such as “A performs better than B”, “A helps people to…”, or “A is preferred over B”) is more transparently communicated if it explicitly states whether the evidence collection that supports it was preregistered, and it is more credible if it was indeed preregistered.

What are the differences between preregistration and the document for the institutional review board (IRB)?

Many institutions that conduct research with human participants has a committee that review and approve study procedures. Such committee is usually called an institutional review board (IRB), ethics committee, or human subjects committee. Researchers are expected to submit a document about the study for approval before they can conduct the study.

Preregistration focuses on elements that affect how can the results be interpreted. (For a full list of components, see <link to what-are-needed for pre-specify data collection and data analysis>.)
A preregistration does not need to include statements on ethical issues as long as they are not influencing how the results could be interpreted. Example of issues discussed in IRB submissions but usually out of scope of preregistrations are:

- how you will obtain informed consent from study participants,
- how you plan to compensate the study participants, or
- risks and benefits to the participants.

While submission that pass IRB approvals could be adequately precise enough about the measurement procedure, they can be imprecise about how the dependent variables are derived from the raw measurements. Therefore, IRB submissions are usually not adequate for preregistration.
Studies that are exempted from the IRB still benefit increased credibility if they are preregistered.

How can you prespecify your plans in a reliable way?

A reliable preregistration is:

- Publicly accessible - Any reviewer or reader should be able to follow the URL without paying for or signing up for any service.
- Time-stamped - If you measure new data, try to timestamp the data too, so it is clear that the preregistration happened first.
- Immutable - Documents are either unchangeable or there is a clear indication of any changes.
- On a registry that has a long-term plan for permanence - A preregistration should be an archival document, lasting as long as the published paper.

Personal websites, university websites, Github, and Google docs do not meet these criteria.

Examples of repositories provide that meet these criteria are:

- Preregistration on osf.io
- aspredicted.org

Example preregistrations in HCI and related topics:

- Fernandes, Walls, Munson, Hullman, & Kay. CHI. 2018
- Yuan, Haroz, & Franconeri. Psychonomic Bulletin & Review. 2018
- Kale, Nguyen, Kay, & Hullman. Transactions on Visualization and Computer Graphics. 2019

Beyond preregistration, some journals have a “Registered Reports” (RR) category of submission. The submissions are essentially incomplete papers where the introduction, methods description, and analysis description are complete, but the data and results are not yet known. The work is reviewed on the basis of the merits of the research question and methodological approach, not the results. If accepted, the authors may collect the data and fill in the results and conclusions. This approach avoid a bias of only publishing certain results.

Will preregistration increase the risk of my research being scooped?

- Pre-registrations can be embargoed or hidden from searches for a specified duration of months or years
- Only people with the URL can access it until either the embargo ends or the authors make it public

What should I do if I need to change something during/after collecting the data?

- Documenting preregistration planning and deviation https://twitter.com/annaveer/status/1133097905898889216

What are valid reasons to keep a preregistration embargoed? For how long is the embargo appropriate?

Is planned the same as confirmatory?

Not exactly. Planned does not imply confirmatory: a planned study/analysis can just ask questions, and does not need to state hypotheses to be confirmed. Confirmatory does not imply planned: few studies/analyses that are presented as confirmatory are actually planned.

Is preregistration possible with qualitative methods or exploratory research?

Yes. Hypothesis-generating and exploratory research methods are compatible with preregistration. Preregistration only aims to limit **hidden** flexibility in experiments, not enforce rigidity. Stating what kind of information will be recorded/measured, participant exclusion criteria, and that an open-coding or hypothesis-generating approach will be used is a valid form of preregistration.

- Preregistering Qualitative Research discusses how preregistration can apply to various types of qualitative approaches, including grounded theory. Table 1 is a template.
- Exploring Pre-registration and Pre-analysis Plans for Qualitative Inference has a similar discussion and describes a template for a qualitative preregistration
- The Preregistration Revolution: “Preregistration does not favor prediction over postdiction; its purpose is to make clear which is which.”

Example preregistrations for qualitative studies:

- https://royalsocietypublishing.org/doi/10.1098/rsos.171474 (published as registered report)
- https://osf.io/wyu78/
- https://osf.io/yj97r/
- https://osf.io/yz25d/
- https://osf.io/egqjp/
- https://osf.io/7qegs/

Is preregistration compatible with double-blind submission?

Yes!

- For OSF, you can share an anonymized version of registration (instructions).
- AsPredicted provides a link to an anonymized PDF of the preregistration with author names removed (example).

What should reviewers do when a submission is preregistered

Make sure that at least a reviewer checks if the preregistration exists, is accessible, and reliable (see [Link to Reliability question]. Aim for long-term scrutinizability, meaning that future readers can check for discrepancies in the preregistration and the paper.

If a reviewer check the preregistration and found some deviation from the preregistration undocumented in the paper, the reviewer should ask the authors to discuss about this in the paper.

DATA COLLECTION

Discuss flexibility in data collection and reporting.

What are needed to credibly prespecify data collection?

In the course of a research project, many decisions that influences the results and the interpretations [Wicherts2016]. A prespecification declares these decisions upfront, hence increase the credibility of your interpretation of the results. The more you prespecify, the more credibility you establish. The table below shows how answering each of the common questions in prespecification can improve the credibility. (The list of questions is modified slightly from aspredicted.org. We reorder them based on their dependency.):

At least, a prespecification should describe how the data will be collected. A credible specification give clear and unambiguous answer to the following questions (slightly modified from aspredicted.org):

- Have any data been collected for this study already?
- How many and which conditions will participants be assigned to?
- What are the key dependent variable(s) and how they will be measured?
- How many observations will be collected or what will determine sample size (e.g., stopping criteria)?
- Exclusion criteria
- ??? include hypothesis - it's nice to have, but I'm not sure if it's necessary
    - Either include

In addition to prespecifying data collection, you can increase the credibility by prespecifying data analysis as well. See section [link: prespecify data analysis] for details.

	Prespecified			
Have any data been collected for this study already?				
How many and which conditions will participants be assigned to?				
What are the key dependent variable(s) and how they will be measured?				

How many observations will be collected or what will determine sample size (e.g., stopping criteria)?				

What are credible ways to prespecify the sample size or stopping criteria?

- Power analysis from a pilot study or previous studies looking at similar effects
- Base the sample size on a resource limits such as the number of available subjects, funding limits, or time
- An arbitrary number. While not ideal, it is still avoids temptation to “Data Peek”
- Precision analysis
- Krushke’s rule
- Sequential experimental design

Old text: It is unrealistic to require all authors to conduct a power analysis, but it is recommended that they at least explain how they decided on the number of participants and whether this was done ahead of time [simmons2011false]. Also, power analysis focuses on null hypothesis significance testing and may be irrelevant if other inferential methods are used [refs]. For example, alternative methods exist for planning the precision of estimation analyses [refs].

How do I know if my/their sample size was sufficient?

It is important to keep in mind that without context, sample size in itself tells nothing. In particular, a reviewer is not justified in rejecting a study based on sample size alone, as there is no magic sample size that works for all studies.

Is there such a thing as too large of a sample size?

No. Often in HCI we are interested in effects that we know, a priori, are non-zero: for example, one would not expect, given two different user interfaces, for people's performance on the two interfaces to be *exactly* the same. Therefore, with a large sample size we will almost always "find" an effect (e.g., a p value less than some threshold). Some therefore argue that there is such a thing as too large a sample size, which will generate "spurious" effects. Yet a large sample size will give a more precise estimate of the effect size. Thus, if we inspect the effect size we will find that it is present but small, and so we can judge the practical significance of the result. The idea that samlpe sizes can be "too large" is only true if one mindlessly subscribes to the so-called null ritual (Gigerenzer).

DATA ANALYSIS

Discuss flexibility in analyses

Can I preregister without an analysis?

Yes, but there is less flexibility and therefore more credibility with an analysis plan.

Should I replace excluded subjects?

This potential source of flexibility can be mitigated by clearly prespecifying whether new subjects will be recruited to replace excluded subjects.
Not replacing excluded subjects will reduce the statistical power but may be necessary due to subject availability or time and budget limits.

What are credible approaches to exclude data from analysis?

Sometimes dropping subjects or individual data points is necessary for valid analysis. Prespecifying can improve credibility. as long as you prespecify.

- Failed catch trials
- at floor for easy condition, at ceiling for hard condition

How do I know if my analysis plan is detailed enough?

Minimize (or eliminate!) ambiguity in:

- Statistical analyses
- Exclusion criteria
- IVs
- DVs
- Subset grouping (e.g. including subject ID in the model)

Check other resources

What if I use visual interpretation?

It's subjective and not reproducible, which makes conclusions based on visual inspection less credible than those based on quantitative criteria. But if it you do use visual inspection, prespecifying that you'll do so avoids perceived flexibility of trying numerical reporting and then changing your mind of the results don't look impressive.

Potential source of hidden flexibility: what you emphasize in plots (by using marks and channels)

Is prespecification the same as confirmatory research?

Not exactly. Planned does not imply confirmatory: a planned study/analysis can just ask questions, and does not need to state hypotheses to be confirmed. Confirmatory does not imply planned: few studies/analyses that are presented as confirmatory are actually planned.

After collecting and analyzing data

Can I preregister after either looking at the data or discussing the data with someone who has looked at it?

No. If you do, state that the results are exploratory, purely descriptive, or will be used to generate new hypotheses for testing.

Do I have to report a preregistration? Can I pretend it never happened?

If a study has a preregistration, failing to point to the preregistration in any write-up of the study makes it harder for the research community to accurately understand the work conducted. The primary goal of a preregistration is to increase the transparency of flexibility in a study or analysis. So failed predictions or changed plans should be reported without shame of being “wrong”.

What if plans change after preregistration?

A preregistration should be thought of as “a plan, not a prison”. Psychic powers are not expected from researchers. After inspecting data, it may become clear that assumptions about the data are not true, experiment subjects did not follow directions, or the quality of the data is lower than expected. In those cases, transparently report any differences between the preregistration and the analysis described in the paper. If you think of an interesting new analysis that had not been considered a priori, you can report it in the paper, provided it is described as an exploratory analysis. This addition could change how the results are interpreted, or whether the conclusions can be used for inferences to a wider population. That is OK, as preregistration enhances the utility and value of these methods by making it more acceptable to report them in fields that traditionally rely on deductive, confirmatory statistical approaches.

OLD stuff before 26.06.19 15:12

Note: some of the content has been taken from the Reviewer FAQ draft which is now deprecated. See if there are more questions that could fit here or in another chapter.

In a factorial design, do I counterbalance (or use Latin square) for each variable or for all of them? How to take the replications (repeats of the same conditions) into account?

- Positional effect vs. “order effect” [Alferes2012]
- Counterbalance, manipulated variable

How do I increase a likelihood that my study won’t go into a file drawer (for a lack of better words: “nothing is significant”)?

- pilot study
- “severe test” [Cairns 2019, chapter 1]
- plan the experiment such that you gain knowledge regardless of the directions of the results

References


```
@article{simmons2011false,
    title={False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant},
  author={Simmons, Joseph P and Nelson, Leif D and Simonsohn, Uri},
  journal={Psychological science},
  volume={22},
  number={11},
  pages={1359--1366},
  year={2011},
```

  publisher={Sage Publications Sage CA: Los Angeles, CA}
}

@article{maxwell2004,
    title={The persistence of underpowered studies in psychological research: causes, consequences, and remedies.},
  author={Maxwell, Scott E},
  journal={Psychological methods},
  volume={9},
  number={2},
  pages={147},
  year={2004},
  publisher={American Psychological Association}
}
@article{Gelman2017,
  title={Ethics and Statistics: Honesty and Transparency Are Not Enough},
  author={Gelman, Andrew},
  journal={Chance},
  volume={30},
  number={1},
  pages={37--39},
  year={2017},
  publisher={Taylor \& Francis}
}
@article{Bakker2016,
  title={Researchers' intuitions about power in psychological research},
   author={Bakker, Marjan and Hartgerink, Chris HJ and Wicherts, Jelte M and van der Maas, Han LJ},
  journal={Psychological science},
  volume={27},
  number={8},
  pages={1069--1077},
  year={2016},
  publisher={Sage Publications Sage CA: Los Angeles, CA}
}
@book{Alferes2012,
  title={Methods of randomization in experimental design},
  author={Alferes, Valentim R},
  volume={171},
  year={2012},
  publisher={Sage Publications}
}
@article {Nosek2018,
      author = {Nosek, Brian A. and Ebersole, Charles R. and DeHaven, Alexander C. and Mellor, David T.},
      title = {The preregistration revolution},

```
      year = {2018},
      doi = {10.1073/pnas.1708274114},
      publisher = {National Academy of Sciences},
      issn = {0027-8424},
      URL = {http://www.pnas.org/content/early/2018/03/08/1708274114},
      eprint                                                    =
{http://www.pnas.org/content/early/2018/03/08/1708274114.full.pdf},
      journal = {Proceedings of the National Academy of Sciences}
}

@article{nosek2018preregistration,
author = {Nosek, Brian A and Lindsay, D Stephen},
journal = {APS Observer},
mendeley-groups = {Pre-reg/Why preregister},
number = {30/3},
title = {{Preregistration becoming the norm in psychological science}},
volume = {31},
year = {2018}
}


@article{cumming2014,
  title={The new statistics: Why and how},
  author={Cumming, Geoff},
  journal={Psychological science},
  volume={25},
  number={1},
  pages={7--29},
  year={2014},
  publisher={Sage Publications Sage CA: Los Angeles, CA}
}
```

Resources to be incorporated into the FAQ

- Maxwell2004: lack of awareness about different “power” in ANOVA (with worked examples)
    - A. power to have at least one effect significant
    - B. power for a particular effect to be significant
    - C. power for all effects to be significant

- Gelman2017: “honest reporting is not enough to save a poor design”
- Bakker2016: discrepancy about researchers’ estimate of statistical power
    - Psychologists overestimated power when the expected effect sizes are small (N = 291 research psychologists)
    - Survey (N = 214 respondents), 95% underestimated sample size required to obtain 0.80 power for small effect (Cohen’s d = 0.20)
- Nosek2018: List of several challenges in preregistration observed in practice
-