# Pre-Registration

EDRS Open Science Pre-Conference Workshop

Shirley B. Wang

September 17, 2021

# Today's Topics

Provide a brief introduction to **preregistration.** What is it, how and when should we do it, and why should we care?

Overview of common preregistration **templates and websites** (e.g., Open Science Framework, AsPredicted), including templates for secondary data analysis.

Address common **concerns and questions** about preregistration (e.g., "what if someone scoops my ideas?").

Provide you with enough knowledge to start **pre-registering your own studies**!

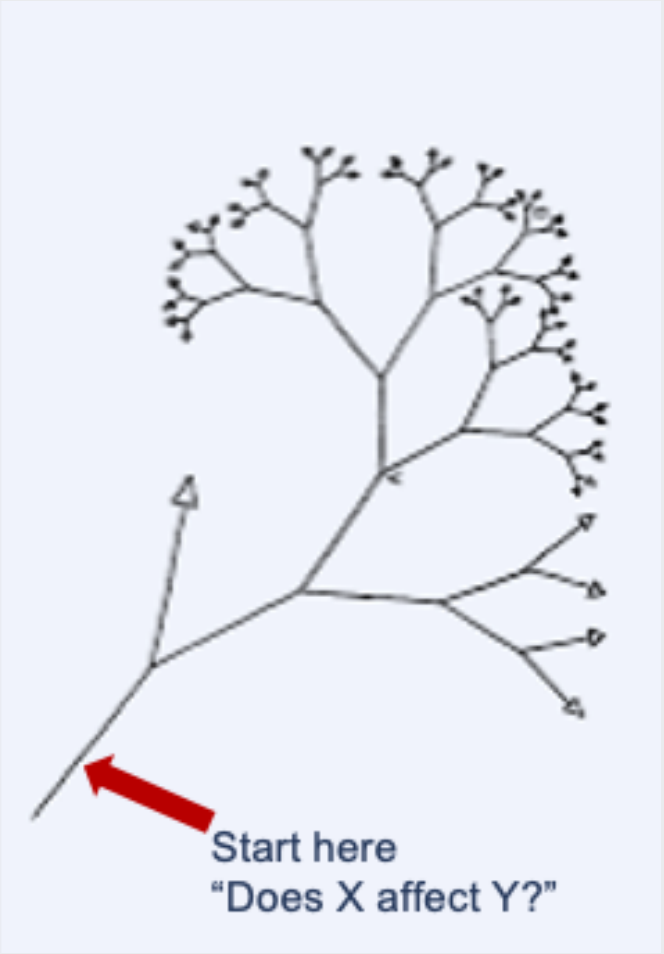
# What is Pre-Registration?

# Why Pre-Register?

Start with a question (and a hypothesis).

But then, do you...

- Control for demographics?
- Exclude outliers?
- Use the median or mean?
- Use the full scale or subscales?
- Which timepoints?
- Listwise deletion, pairwise deletion, or imputation?

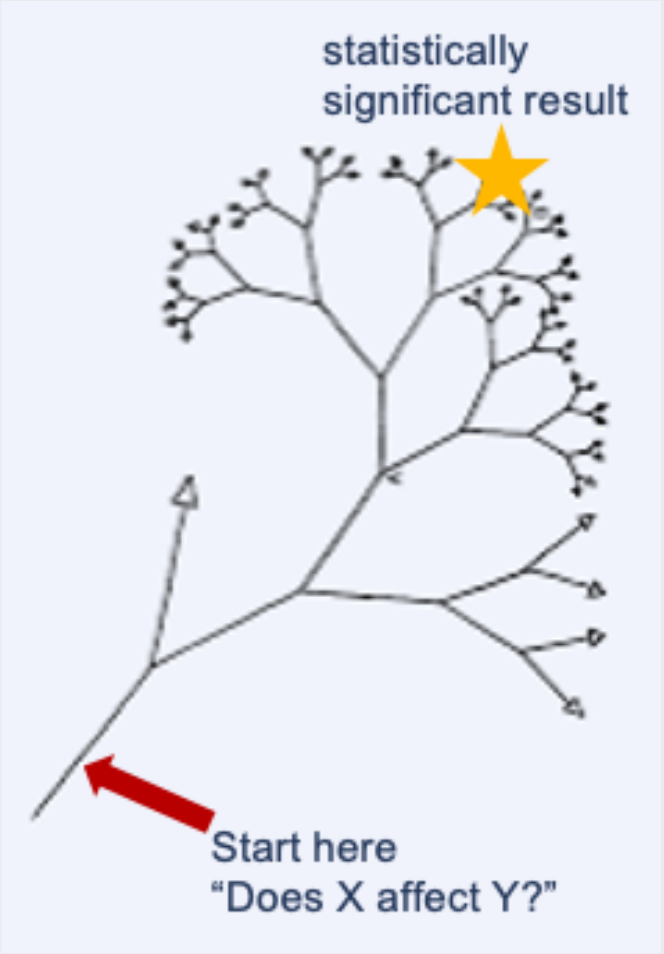


# Why Pre-Register?

Start with a question (and a hypothesis).

But then, do you...

- Control for demographics?
- Exclude outliers?
- Use the median or mean?
- Use the full scale or subscales?
- Which timepoints?
- Listwise deletion, pairwise deletion, or imputation?



# Questionable research practices are common

# Pre-Registration Basics


Pre-registrations are **time-stamped research plans** typically created **before** starting data collection[1].

The goal is to make a clear distinction between *exploratory* and *confirmatory* work.

There are many templates available on the Open Science Framework; they vary in specificity/length but typically include:

- Study information
- Specific hypotheses
- Methods
- Specific analysis plan
- Whether data collection has begun

[1] However, there are preregistration templates available for secondary data collection.

# Common Pre-Registration Templates

# Standard OSF Registration

Standard, comprehensive, general purpose preregistration form (template | tutorial)

Preregister:

- Study information (title, authors, brief description, hypotheses[1])
- Design plan (study type, blinding, study design, randomization)
- Sampling plan (existing data, data collection procedures, sample size)
- Variables (manipulated variables, measured variables, indices[2])
- Analysis plan (statistical models[1], transformations, inference criteria, data exclusion, missing data, exploratory analysis)

[1] Hypotheses should be specific, concise, and testable.
[2] Precisely specify if and how measurements will be combined into an index (e.g., sum score, subscales)
[3] As described on the template, this is perhaps the most important and complicated question. It is crucial to provide a specific recipe for analyzing the data, and avoid leaving any uncertainties/open questions.

# AsPredicted Registration

Shorter general-purpose preregistration (template | website)

Answer eight questions:

1. Have any data been collected for this study already?
2. What's the main question being asked or hypothesis being tested in this study?
3. Describe the key dependent variable(s) specifying how they will be measured.
4. How many and which conditions will participants be assigned to?
5. Specify exactly which analyses you will conduct to examine the main question/hypothesis.
6. Any secondary analyses?
7. How many observations will be collected or what will determine the sample size?
8. Anything else you would like to preregister?

# Secondary Data Registration

For preregistering research using an existing dataset (FAQ | Template | Paper).

This template includes many of the same questions as prior templates (e.g., clear description of measures, statistical analysis plans, etc.), with some important additions specific to working with existing data:

- Data description (datasets used, data availability, data access date)
- Knowledge of data[1]:
  - Prior publications/presentations you have worked on from the dataset
  - Prior knowledge that may be relevant for proposed analyses (e.g., from working with the data firsthand, reading previously published research, codebooks)
  - Transparently report any direct or indirect knowledge about hypothesized associations of your variables.

[1] Describe prior knowledge of data for each co-author separately

# Common Questions and Concerns

# Pre-registration sounds great, but...

- "What if someone scoops me"?
  - Time stamps verify your claims
  - pre-registrations can be embargoed

- "Does this mean I can't look at anything else in my data?"
  - No! You can still explore your data; pre-registrations help differentiate *confirmatory* vs. *exploratory* research.

- "What if I find a better way to analyze my data?"
  - That's great! You can indicate this in your paper and explain any deviations from pre-registration protocols.

- "Can't people cheat?"
  - Yes. Pre-registrations can be "hacked" - but this makes questionable research practices more intentional.
  - Pre-registrations help you stay accountable to yourself.

# Thank you!

Shirley Wang
PhD Candidate, Harvard University
Clinical Psychology
Computational Science & Engineering

Email: shirleywang@g.harvard.edu

Twitter: @ShirleyBWang

Website: shirleywang.rbind.io

All materials available on github. For more, see https://osf.io/pbv27/.