# Pre-Analysis Plans for Survey Experiments: A Brief Primer

Joshua Kalla, Yale University

It is becoming increasingly common to file a pre-analysis plan before conducting experiments in the social sciences.[1] For example, the *Journal of Politics* announced that beginning in 2022, all original experimental work must have been pre-registered in order to be considered for publication at the *JOP*.

This document is intended to describe an easy process for submitting a pre-analysis plan. Pre-analysis plans come in many varieties and flavors. There is not one "right" way to draft a pre-analysis plan. Instead, this document provides a heavily opinionated process for filing a pre-analysis plan "as code."[2]

The motivations for this are three-fold:
- You will already need to write the code to clean and analyze your survey experiment. By doing this before conducting the experiment, you are able to draft a pre-analysis plan without adding any additional time to your research project.
- This allows you to very precisely specify your analytic approach. The more precise a pre-analysis plan, the better, since this means fewer researcher degrees of freedom. By specifying all of the cleaning and analysis in code, rather than in words, there is less room for post-hoc researcher interpretation of the pre-analysis plan.
- Once you have the analysis code, you can also go beyond your pre-analysis plan to assess the quality of the design, such as its statistical power. For example, you can use DeclareDesign for this: https://declaredesign.org/declare.pdf. Rather than arbitrarily creating fake data in Qualtrics, you would specify distributions of outcomes and treatment effects.

How should one file a pre-analysis plan for a survey experiment?[3] At a high-level:
1. Design the survey experiment in Qualtrics.
2. In Qualtrics, generate fake data.
3. Draft your cleaning and analysis code based on this fake data.
4. Upload your pre-analysis plan code to a research registry. You're now ready to launch!
5. Include your pre-analysis plan in any manuscripts you submit to a journal.

The following pages review each of these steps in greater detail.

As a running example, we will imagine we are conducting a question wording survey experiment to determine how public support varies depending on how we refer to the Affordable Care Act. We will randomly assign respondents to read, "Do you have a favorable or unfavorable opinion of X?" where X will either be: (1) the 2010 health reform law; (2) the Affordable Care Act; or (3) Obamacare.

---

[1] Pre-analysis plans are not just limited to experiments. One could file a pre-analysis plan for observational or qualitative data. While there are some cases of these types of pre-analysis plans, the norms for observational and qualitative research today are different from experimental research. In the interest of brevity, this document is limited only to survey experiments. Much of what is written here could be applied to lab or field experiments. However, given how common survey experiments are, that is the focus of this document.

[2] This is inspired by the "infrastructure as code" software movement (https://hackernoon.com/everything-as-code-explained-0ibg32a3).

[3] This is based on a thread from David Broockman: https://twitter.com/dbroockman/status/1350897805603205123

# 1. Design the survey experiment in Qualtrics.

First, set up your question.

## ACA Example Survey Experiment

▼ DV

☐ Q1    Do you have a favorable or unfavorable opinion of ${e://Field/wording}?
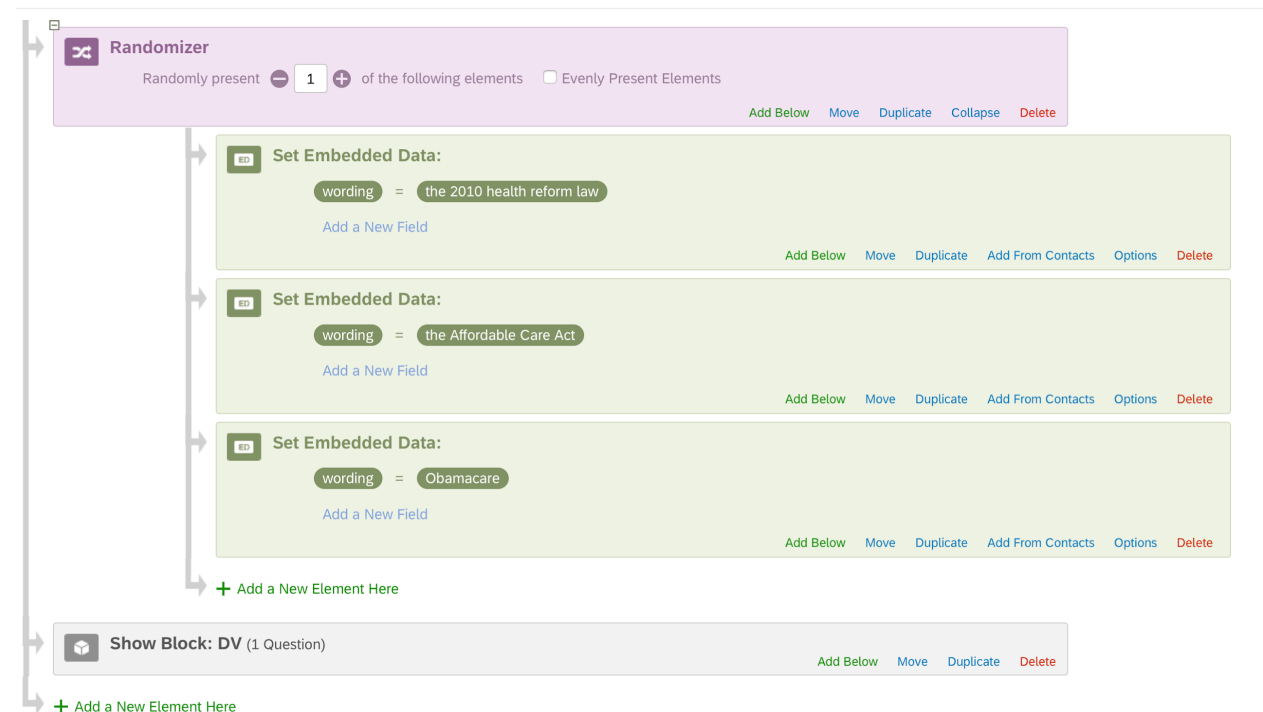
⚙️
    ○ Very favorable

    ○ Somewhat favorable

    ○ Neither favorable nor unfavorable
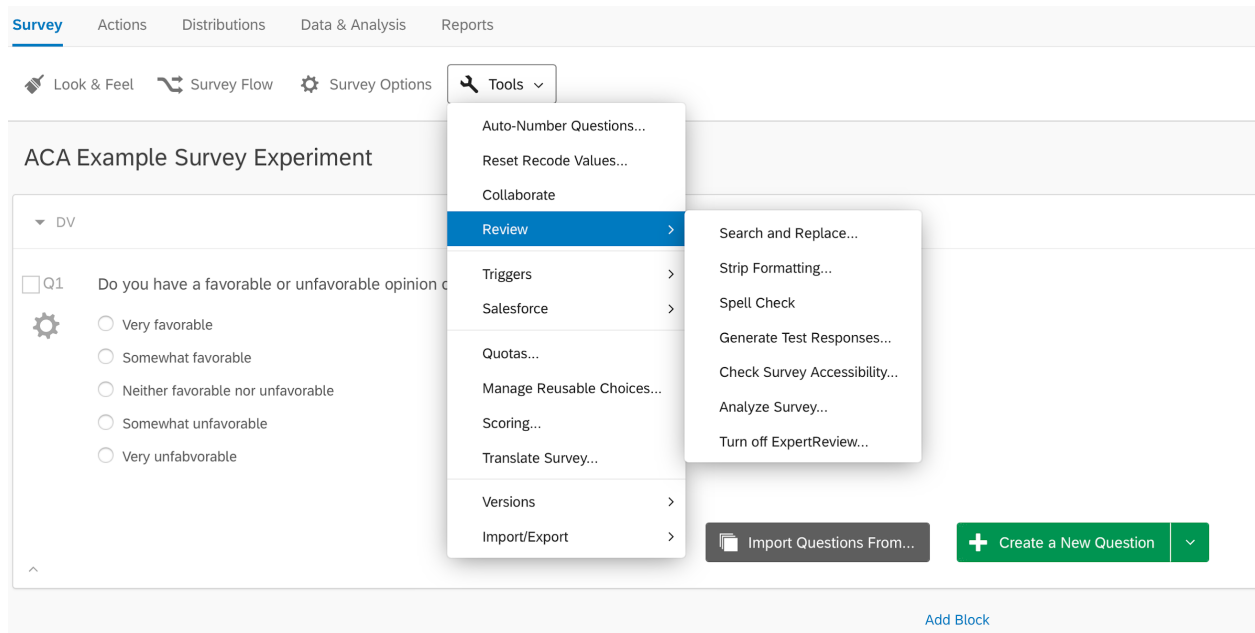
    ○ Somewhat unfavorable

    ○ Very unfabvorable

Second, set up the randomization in the survey flow.

### Survey Flow  ACA Example Survey Experiment

**Randomizer**

Randomly present ⊖ 1 ⊕ of the following elements  ☐ Evenly Present Elements

Add Below    Move    Duplicate    Collapse    Delete

**Set Embedded Data:**

wording  =  the 2010 health reform law

Add a New Field

Add Below    Move    Duplicate    Add From Contacts    Options    Delete

**Set Embedded Data:**

wording  =  the Affordable Care Act

Add a New Field

Add Below    Move    Duplicate    Add From Contacts    Options    Delete

**Set Embedded Data:**

wording  =  Obamacare

Add a New Field

Add Below    Move    Duplicate    Add From Contacts    Options    Delete

＋ Add a New Element Here

**Show Block: DV** (1 Question)

Add Below    Move    Duplicate    Delete

＋ Add a New Element Here

# 2. Generate fake data.

In Qualtrics, go to Tools -> Review -> Generate Test Responses.



Next, generate several hundred fake responses. This may take a few minutes.



Then, navigate to Data & Analysis to export your data. You may need to be patient and refresh this page.

You will want to export as choice text.



## Download a data table

Use the legacy exporter

| CSV | TSV | Excel | XML | SPSS | Google Drive | User-submitted files | Tableau |

**Comma separated values**

This is a .csv file that can be imported into other programs. Each value in the response is separated by a comma and each response is separated by a newline character. If your responses contain special characters and you will open this export in Microsoft Excel we recommend using the TSV export. Qualtrics CSV exports use UTF-8 encoding, which Excel will not open correctly by default.

Learn more

☑ Download all fields

○ Use numeric values
⦿ Use choice text

More options          Close          ↓ Download

You now have your fake data. You should now delete this from Qualtrics, since you don't want to include the fake data in your real analysis.

# 3. Draft your pre-analysis plan as code.

Using your preferred statistical software, draft your code to clean and analyze the survey data. You should make this code easy for reviewers (or yourself in a few months) to understand. That means to use variable names that are sensible -- avoid names like var1 or obscure acronyms that only you may understand. Also, write very clear comments: highlight for a reviewer what your main tests are.

You should also make sure you don't include any identifying information in this code. You want the pre-analysis plan to remain blinded for peer review.

```
                                                                              sample_pap
1    clear all
2
3    // NOTE: Don't include any identifying information, such as in a working directory.
4
5    /* Sample Information:
6    Provider: Lucid
7    Sample: 18+ US general population
8    Expected sample size: 1,000 completed surveys
9    Expected field dates: February 2021
10   Exclusion: Must provide informed consent and pass pre-treatment attention checks
11   */
12
13   *** IMPORT DATA ***
14   import delimited using "~/Downloads/ACA+Example+Survey+Experiment_February+4%2C+2021_11.20.csv", ///
15       clear rowrange(4:) varnames(2) bindquotes(strict) delim(",")
16
17   *** EXCLUSIONS ***
18   drop if consent != "Yes" // Drop if no consent
19   keep if understandac == "I understand" // Drop if failed attention check
20
21   *** CREATE A NUMERIC TREATMENT VARIABLE ***
22   encode wording, gen(wording_enc)
23
24   *** CLEAN THE DEPENDENT VARIABLE ***
25   rename doyouhaveafavorableorunfavorable health_support
26   replace health_support = "5" if health_support == "Very favorable"
27   replace health_support = "4" if health_support == "Somewhat favorable"
28   replace health_support = "3" if health_support == "Neither favorable nor unfavorable"
29   replace health_support = "2" if health_support == "Somewhat unfavorable"
30   replace health_support = "1" if health_support == "Very unfabvorable"
31   destring health_support, replace
32
33   *** MAIN TEST: Does support vary by question wording? ****
34   reg health_support i.wording_enc
35   margins wording_enc
36
37   /*
38   PRIMARY FIGURE:
39   We expect to plot mean favorability across condition using 83% confidence intervals.
40   See Austin and Hux (2002) "A brief note on overlapping confidence intervals" for
41   a discussion of these CIs: https://core.ac.uk/download/pdf/82702323.pdf.
42   */
43   marginsplot, scheme(s1color) ytitle("Favorability (1-5") xtitle("") ///
44       title("Effect of Question Wording on Favorability") ///
45       l(83) note("Error bars denote 83% confidence intervals")
46
47   *** SECONDARY TESTS ***
48   /*
49   We could also pre-specify any secondary analyses. For example, we might be interested
50   in heterogenous treatment effects by a respondents partisanship. We could draft
51   the code for that here.
52   */
53
54   /*
55   For anything not included in this pre-analysis plan, I will default to using the
56   standard operating procedures (version 1.05) from Don Green's lab at Columbia/
57   See https://alexandercoppock.com/Green-Lab-SOP/Green_Lab_SOP.html
58   */
59
```

In your pre-analysis plan as code, I also suggest you specify any sample restrictions or screenings. Sometimes this will be done in code (e.g., dropping a subject if they fail an attention check question or do not provide informed consent), in which case you should clearly specify this in the code (see Lines 17-19 above). But other times, this might occur at an earlier stage in the process (e.g., you tell a sample provider to only include subjects of a particular demographic). While this won't require code, you should still make a note of this at the top of your code file. Include something like the comments in Lines 5-11 above.

Some other features of the above sample code that should likely always be included in a pre-analysis plan:
- Specify how you will recode your variables. Note how I am very clear what the numeric values 1-5 for health_support mean substantively.
- Specify how your code relates to your hypotheses. Note how in Line 33 I add a comment on what the main hypothesis is and how I will test it. Also, will you test this hypothesis using one- or two-tailed tests? The default in political science will be two-tailed. If you plan to do anything different, specify that and have a sentence or two justifying/explaining it.
  - **To belabor this point, be sure to clearly state the main hypotheses. This is one of the most important parts of a PAP.**
- If you plan to use pre-treatment covariates to improve precision, make sure you specify which covariates and how they are each recoded. Write down the full regression as code and make sure it runs with your fake data. You should also clearly specify in your code what, if anything, you will do with observations with missing data (e.g., will you impute the mean, delete observations with missing data, etc.?).
- Your pre-analysis plan should also have code testing for covariate imbalances and differential attrition. These are often serious problems in survey experiments (e.g., http://datacolada.org/89).
- Ideally, your code could even generate the key tables/figures that you intend to include in the paper.
- Throughout, it is typically good to include some general descriptions of the intention and motivation for the different choices you make. This is to restrict oneself in case the code one wrote turns out not to work, and changes are needed. Without a description of the underlying intentions and motivations, one might be tempted to do larger changes.
- It is often helpful to include a generic statement like: "For anything not included in this pre-analysis plan, I will default to using the standard operating procedures (version 1.05) from Don Green's lab at Columbia (see https://alexandercoppock.com/Green-Lab-SOP/Green_Lab_SOP.html)." In this case, if anything unforeseen happens, you have a fall-back pre-analysis plan. I do this in Lines 54-58. Remember, pre-analysis plans are learning experiences. You won't include everything you "should have" included when you draft your first one. Keep a note of what you perhaps should have included that way you will remember to include those features in future pre-analysis plans.
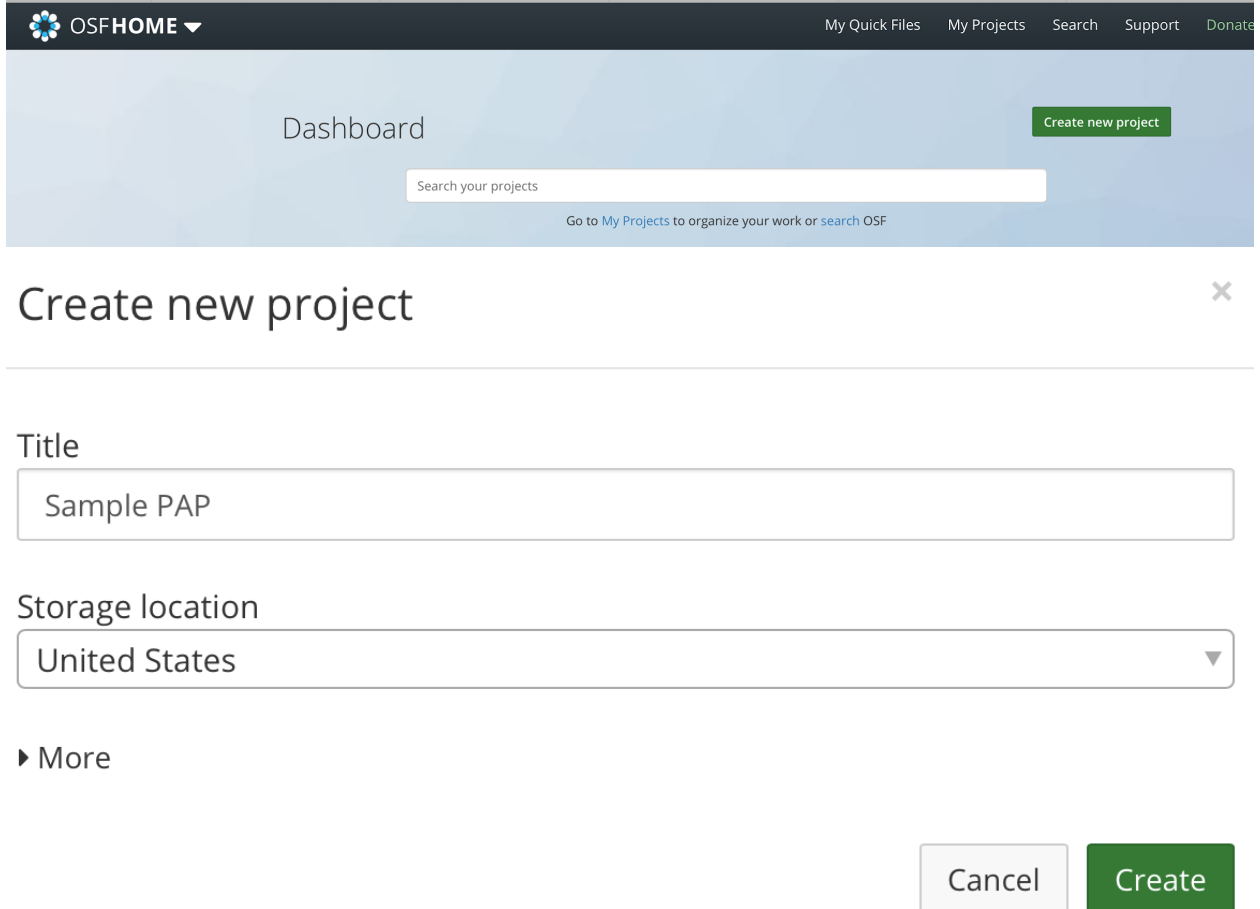
At this stage, you could also modify the fake data with an assumed treatment effect in order to conduct a power analysis. This does not need to be included in a pre-analysis plan, but this is often an incredibly helpful step for assessing the quality of your research design. For some examples on how to do this in R, see https://declaredesign.org/getting-started.html and https://egap.org/resource/script-power-analysis-simulations-in-r/. For an example in Stata, see

https://blog.stata.com/2019/08/13/calculating-power-using-monte-carlo-simulations-part-3-linear-and-logistic-regression/. If you run a power analysis, you should include that in your pre-analysis plan.

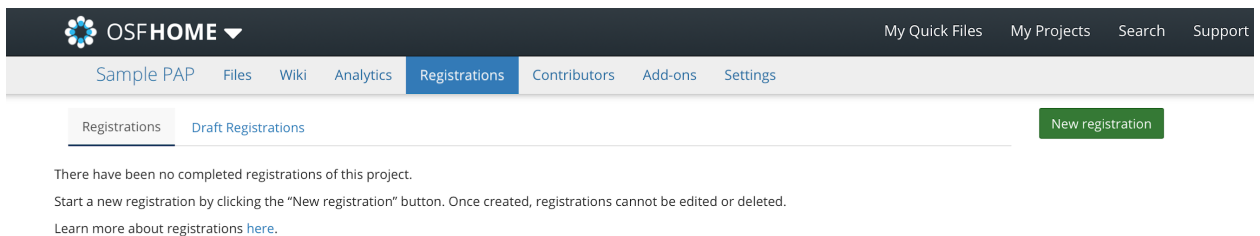# 4. Upload your pre-analysis plan code to a research registry.

https://osf.io is very easy to use for this.

First, after logging in, create and name a new project.



After navigating to the new project's home page, create a new registration.



It is easiest to create an open-ended registration.

# Register

Registration creates a frozen version of the project. Your original project remains editable and will have the registration linked. Things to know about registration:

- **Registrations cannot be edited or deleted.**
- Withdrawing a registration removes its contents, but leaves behind basic metadata: title, contributors, date registered, date withdrawn, and justification (if provided).
- Registrations can be public or embargoed for up to four years. Embargoed registrations will be made public automatically when the embargo expires.

Continue your registration by selecting a registration form:

- ○ **OSF Preregistration** ❶
- ● **Open-Ended Registration** ❶
- ○ **Qualitative Preregistration** ❶
- ○ **Registered Report Protocol Preregistration** ❶
- ○ **OSF-Standard Pre-Data Collection Registration** ❶
- ○ **Preregistration Template from AsPredicted.org** ❶
- ○ **Replication Recipe (Brandt et al., 2013): Post-Completion** ❶
- ○ **Replication Recipe (Brandt et al., 2013): Pre-Registration** ❶
- ○ **Pre-Registration in Social Psychology (van 't Veer & Giner-Sorolla, 2016): Pre-Registration** ❶

[ Cancel ]  [ Create draft ]

You should then enter metadata for the experiment. Here, you could choose to include other information, such as expected field dates, intended sample sizes, etc. Some pre-analysis plans will include this type of information.

O **Metadata**

● **Summary**

● **Review**

<

## Registration Metadata

This metadata applies only to the registration you are creating, and will not be applied to your project.

Next ➔

Auto-saved:
a few seconds ago

**Title** *

Sample PAP

**Description** *

Enter in a few sentences about the project.
This is a sample pre-analysis plan for a survey experiment testing whether question wording has an effect on support for the ACA.

**Contributors**
Edit contributors on your project

| Name | Permission | Citation |
|------|-----------|----------|
| Joshua Kalla | Administrator | Yes |

**Category**

📦 Project                                                                    ▼

**Affiliated institutions**
You have no institutional affiliations

**License** *

A license tells others how they can use your work in the future and only applies to the information and files submitted with the registration. For more information, see this *article on licenses*.

No license                                                                    ▼

License FAQ

**Year** *

2021

**Copyright Holders** *

Authors

Show full text

**Subjects** *

Social and Behavioral Sciences ✖

Browse all subjects        Search subjects

Then, upload your pre-analysis plan code.

Sample PAP >
# New registration

## Summary

**Provide a narrative summary of what is contained in this registration or how it differs from prior registrations. If this project contains documents for a preregistration, please note that here. ***

This PAP is written as code. See attached .do file for the PAP.

**Add supplemental files or additional information**

You may attach up to 5 file(s) to this question. You may attach files that you already have in OSF Storage in this project or upload (drag and drop) a new file from your computer. Uploaded files will automatically be added to this project so that they can be registered. To attach files from other components or an add-on, first add them to this project.

sample_pap.do ✖

➕

| Name ∧∨ | Last modified ∧∨ |
|---|---|
| 🗋 sample_pap.do | 2021-02-04 11:40 AM |

**Review**

← Metadata

Auto-saved:
a few seconds ago

You can then choose to make this registry available immediately or impose an embargo. It is fairly common to impose an embargo.

## Almost done...

Please keep in mind that:

- Registrations cannot be modified or deleted once completed.
- The content and version history of Wiki and OSF Storage will be copied to the registration.
- This project contains links to other projects. These links will be copied into your registration, but the projects that they link to will not be registered. If you wish to register the linked projects, they must be registered separately. Learn more about links.

○ **Make registration public immediately**
○ **Enter registration into embargo**

[ Submit ]  [ Back ]

You will then receive an email from OSF asking you to approve the embargo. Be sure to click accept.

**You are now ready to run your experiment!**

# 5. Prepare your pre-analysis plan for peer review.

You will want to include an anonymized version of your pre-analysis plan for peer review.

To do this, navigate to the project's home page, click the 3 dots in the top right, and select "Create view-only link."



You can then create a fully anonymized version of your pre-analysis plan.

This will generate a link (e.g., https://osf.io/sdb5w/?view_only=bf115e8b67f44de4b49b2f348aed2804) that you can share with reviewers. There, reviewers can find a time-stamped version of your pre-analysis plan.

**Sample PAP**

1.6KB   Private   Make Public   ₽ 0   …

Contributors: Anonymous Contributors
Date created: 2021-02-04 11:34 AM | Last Updated: 2021-02-04 11:44 AM
Category: 🗂 Project
Description: Add a brief description to your project
License: Add a license

**Wiki**

Add important information, links, or images here to describe your project.

**Components**   Add Component   Link Projects

Add components to organize your project.

**Files**

Q Filter   i

| Name ∧ ∨ | Modified ∧ ∨ |
|---|---|
| 🗂 Sample PAP | |
| ⚙ OSF Storage (United States) | |
| 📄 sample_pap.do | 2021-02-04 11:40 AM |

**Tags**

Add a tag to enhance discoverability

**Recent Activity**

🖥 A user created a view-only link to a project
2021-02-04 11:44 AM

🖥 A user approved an embargoed registration of a project
2021-02-04 11:41 AM

🖥 A user initiated an embargoed registration of a project
2021-02-04 11:41 AM

🖥 A user added a file to OSF Storage in a project
2021-02-04 11:40 AM

🖥 A user created a project
2021-02-04 11:34 AM

After the paper is accepted for publication, you can end the embargo. You can also upload your replication files, data, and code to this OSF project.

That's it!