

Ulysses' Pact or Ulysses' Raft: Using Pre-Analysis Plans in Experimental and Non-Experimental Research

Jeffrey Michler, University of Arizona
Sarah Janzen, University of Illinois, Urbana-Champaign

Research Transparency and Reproducibility Training (RT2) for African Scholars
21 May 2026

Ulysses' Pact



They bound me hand and foot in the tight ship—
erect at the mast-block, lashed by ropes to the mast.

— Homer, *The Odyssey*, Book XII

Ulysses' Raft



And now the master mariner steered his craft,
sleep never closing his eyes, forever scanning
the stars.

— Homer, *The Odyssey*, Book V

The Pact or the Raft?

- ▶ Ulysses Pact - pre-commitment to a specific analysis plan
 - ▶ Common in law and medicine
 - ▶ Minimize the probability of false discoveries
 - ▶ Mirrors the way “proper” science is done
- ▶ Ulysses Raft - explore the data for discoveries only possible through searching
 - ▶ Way economics has always been done
 - ▶ Specification search may be the best way for new discoveries
 - ▶ Anything else is “impossible” for non-experimental data

Outline

- ▶ A brief history of pre-analysis plans
- ▶ Components of a pre-analysis plan
- ▶ Using a pre-analysis plan
- ▶ A third way?

Pre-Analysis Plans in Medicine

- ▶ Plans have been required by law in the United States for all drug trials since 1997
 - ▶ Stated goal was to counter strong incentives of drug companies to find positive results
- ▶ In 2004 the twelve journals of International Committee of Medical Journal Editors announced public trial registration would be pre-requisite for publication
 - ▶ Stated goal was to counter the “file drawer problem,” where only significant results get published
- ▶ ClinicalTrials.gov is where most are registered
 - ▶ Pre-2004, registry averaged 1,879 a year
 - ▶ Post-2004, registry averaged 22,088 a year

Pre-Analysis Plans in Social Science

- ▶ Pre-analysis plans predate registry
- ▶ Most commonly cited first studies:
 - ▶ Casey, Glennerster, and Miguel (2012). “Reshaping institutions: Evidence on aid impacts using a preanalysis plan.” *QJE*.
 - ▶ Finkelstein et al. (2012). “The Oregon health insurance experiment: Evidence from the first year.” *QJE*.
- ▶ In 2012 AEA established the AEA RCT Registry
 - ▶ Allows for registration of lab and field experiments
- ▶ Additional social science registries founded since
 - ▶ 3ie’s Registry for International Development Impact Evaluations (RIDIE) - impact evaluations in low income countries
 - ▶ Open Science Foundation (OSF) registry - any study

Pre-Analysis Plans v. Registrations

- ▶ In medical science, there has been a clear distinction between:
 - ▶ What pre-analysis plans prevent - *p*-hacking/HARKing/Specification Search
 - ▶ What public registration solve - file drawer problem
- ▶ In economics, these two have been muddled
 - ▶ See Olken (2015); Coffman and Niederle (2015); Burlig (2018); Duflo et al. (2020)
- ▶ Making pre-analysis plans work means understanding the difference between what a plan accomplishes and what a public registry accomplishes

Basic Pre-Analysis Plan

- ▶ **Data description:** Describe the process by which data was generated or collected
- ▶ **Hypotheses:** Clearly state the research questions and hypotheses to be tested
- ▶ **Outcome variables:** List the primary outcome variables as well as any secondary outcomes to be tested.
- ▶ **Variable construction:** Precisely define how each outcome and control variable is to be constructed
- ▶ **Model specification:** Describe the statistical method of analysis, including covariates and the approach to inference

Comprehensive Pre-Analysis Plans

- ▶ **Motivation:** Provide motivation and context for the study. A review of the literature can be helpful
- ▶ **Theoretical model or theory of change** The model to be tested or causal diagram that justifies the selected hypotheses
- ▶ **Multiple inference adjustments:** Address how multiple inference will be dealt with. Either Familywise Error Rate (FWER) or False Discovery Rate (FDR).
- ▶ **Data cleaning:** Outline basic operating principles for how data cleaning will be conducted, such as windsorizing outliers, correcting for attrition, and adjusting for contamination.

Comprehensive Pre-Analysis Plans

- ▶ **Subgroup analysis:** Describe any subgroup or heterogeneity analysis.
- ▶ **Power calculations:** If data allows, include information on sample size and power calculations.
- ▶ **Exploratory analysis:** What questions/methods/techniques are you unsure about but might want to explore.

Prevent p -Hacking

- ▶ Forming a Ulysses Pact is a good way to prevent p -hacking
- ▶ It binds the researcher to limit “research degree of freedoms”
- ▶ Important when there are strong incentives to cheat (drug trials)
- ▶ Or, when there is leeway for what researchers can “discover”
 - ▶ Coffman and Niederle (2015) - economic theory limits what a research might “find”
 - ▶ Brodeur et al (2016) - little evidence of p -hacking in experiments

Example 1: Casey et al. (2012)

“We show how misleading an undisciplined interpretation of treatment effects can be in the absence of a PAP by constructing two opposing and equally erroneous narratives based on our data.”

- ▶ Without a pre-analysis plan, selective emphasis can make the same evidence support conflicting stories.
- ▶ Casey, Glennerster, and Miguel use their own results to illustrate the risk of ex post cherry-picking.

Example 1: Casey et al. (2012) – Cherry Picking

ERRONEOUS INTERPRETATIONS UNDER “CHERRY PICKING”

Outcome variable	(1) Mean for controls	(2) Treatment effect	(3) Robust std. err.	(4) N	(5) Hypo
Panel A: GoBifo “weakened” institutions					
Attended meeting to decide what to do with the tarp	0.81	-0.04 ⁺	(0.02)	236	H5
Everybody had equal say in deciding how to use the tarp	0.51	-0.11 ⁺	(0.06)	232	H5
Community used the tarp (verified by physical assessment)	0.90	-0.08 ⁺	(0.04)	233	H4
Community can show research team the tarp	0.84	-0.12 [*]	(0.05)	232	H5
Respondent would like to be a member of the VDC	0.36	-0.04 [*]	(0.02)	236	H10
Respondent voted in the local government election (2008)	0.85	-0.04 [*]	(0.02)	236	H10
Panel B: GoBifo “strengthened” institutions					
Community teachers have been trained	0.47	0.12 ⁺	(0.07)	173	H4
Respondent is a member of a women’s group	0.24	0.06 ^{**}	(0.02)	236	H8
Someone took minutes at the most recent community meeting	0.30	0.14 [*]	(0.06)	227	H5
Building materials stored in a public place when not in use	0.13	0.25 [*]	(0.10)	84	H5
Chieftom official did not have the most influence over tarp use	0.54	0.06 [*]	(0.03)	236	H6
Respondent agrees with “Responsible young people can be good leaders” and not “Only older people are mature enough to be leaders”	0.76	0.04 [*]	(0.02)	236	H6, H12
Correctly able to name the year of the next general elections	0.19	0.04 [*]	(0.02)	236	H9

Notes: Significance levels (per comparison p -value) indicated by ⁺ $p < .10$, ^{*} $p < .05$, ^{**} $p < .01$. Treatment effects estimated on follow-up data. Includes fixed effects for the district council wards (the unit of stratification) and the two balancing variables from the randomization (total households and distance to road) as controls.

Example 1: Casey et al. (2012)

“These two plausible, opposite, and equally erroneous interpretations illustrate the risks of allowing researchers complete discretion to choose the subset of outcomes to highlight ex post and the potential value of employing a PAP.”

- ▶ The pre-analysis plan disciplines which outcomes are primary and how the evidence will be interpreted.
- ▶ The goal is not to eliminate judgment, but to separate pre-specified tests from exploratory interpretation.

Solve the File Drawer Problem

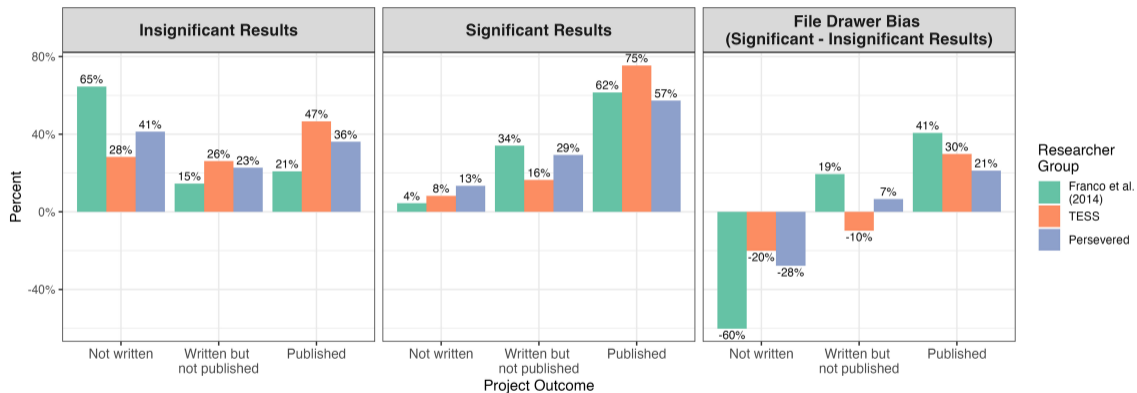
- ▶ Ulysses Pacts don't solve the file drawer problem but public registries do
- ▶ Regardless of the empirical methods a “hypotheses registry” creates a database on the universe of research questions asked
- ▶ Pre-analysis plans and public registries work together but are not the same thing
 - ▶ Registered reports at *Journal of Development Economics*, the *Journal of Accounting Research*, *The Review of Financial Studies*, and *PLOS One*

Example 2: Moniz et al. (2025)

“We find evidence of the problem, stemming from scholars opting to not write up insignificant results.”

- ▶ Moniz, Druckman, and Freese revisit the file drawer problem in social science survey experiments after the rise of open science norms.
- ▶ They follow TESS applicant-projects from 2012–2018, comparing projects with significant and insignificant findings.
- ▶ The key mechanism is not only journal rejection: many null results are never written up or submitted.

Example 2: Moniz et al. (2025) – File Drawer Bias



Example 2: Moniz et al. (2025)

“The file drawer problem we document appears to stem from researchers’ choices to not write up or submit null results.”

- ▶ Significant results are more likely to appear in print: 30 percentage points among TESS researchers and 21 percentage points among persevering researchers.
- ▶ The bias is smaller than in Franco et al. (2014), suggesting progress, but not disappearance.
- ▶ Public registries help because they preserve the universe of research questions, even when results are null or never published.

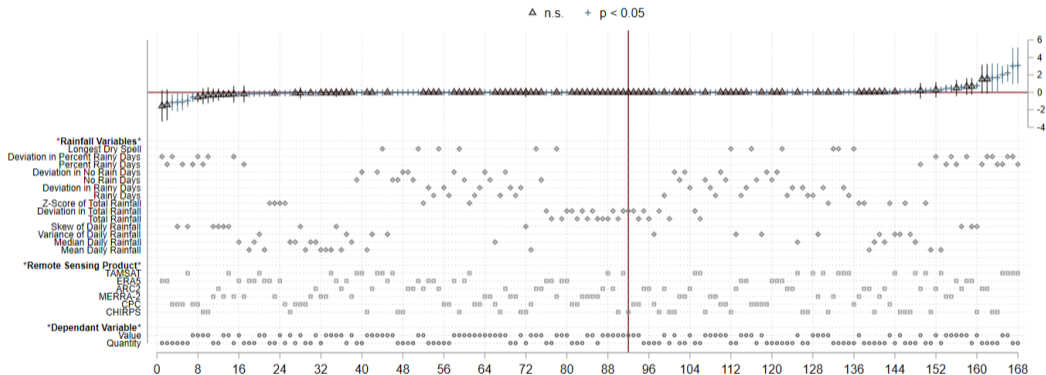
Pre-Analysis Plans for Observational Studies

- ▶ Typical history of pre-analysis plans in econ start with two RCTs (Casey et al. and Finkelstein et al.) both in *QJE*
- ▶ First published paper with a pre-analysis plan in econ was Neumark (2001) in *Industrial Relations*
 - ▶ Neumark tests impact of increase of minimum wage on employment
 - ▶ Developed pre-analysis plan prior to use of new data
 - ▶ Submitted plan to editors ahead of data release
 - ▶ Neumark (2020) uses same process to test effects of CA minimum wage law
- ▶ Econ has muddled the point, it has focused on pre-analysis plans for experiments only
- ▶ No reason why pre-analysis plans & registries can't be used in observational studies - especially to solve the file drawer problem

Example 3: Weather and Agriculture – Why a PAP?

- ▶ How much does weather-measurement error matter when estimating the impact of weather on agricultural productivity?
- ▶ 7 LSMS-ISA countries, 19 survey waves, ~70,000 observations; 9 remote-sensing products with daily weather series over 34 years.
- ▶ 10 GPS obfuscation methods, 27 weather metrics, and 6 econometric specifications implied roughly 200,000 potential regressions.
- ▶ The PAP converted an enormous model-selection problem into a documented design, filed at osf.io/z3snh.

Example 3: Weather and Agriculture – The Specification Space



Each point represents a plausible modeling choice. The PAP made those choices visible before the results could guide which weather product, metric, or specification to emphasize.

Example 3: Weather and Agriculture – What the PAP Bought Us

- ▶ Household and remote-sensing data were public, but exact household GPS coordinates were private and required specialized data construction.
- ▶ The populated PAP became a 180-page public record!!!!
- ▶ Lots of null results
- ▶ Has produced 2 papers in *JDE* and several follow-on papers.

Pre-Analysis Plans for Replication

- ▶ There has been recent growth in pre-analysis plans for replications
- ▶ Restrict incentives to find positive results (p -hacking)
- ▶ Replications have limited scope for “exploration” (no loss of freedom from Ulysses raft)
- ▶ Protects against accusation of bias in undermining a famous study
- ▶ Andrew Chang (US Fed) has 6 replications studies in which he filed pre-analysis plans

Pre-Analysis Plans for Teaching

- ▶ Grad students frequently struggle to articulate their research question or method
- ▶ Pre-analysis plans (registered or not) can help focus grad student ideas
- ▶ Easy to make part of a prospectus or an assignment for a research methods course
- ▶ Extremely helpful in giving guidance to students that are part of a large project

Ulysses' Pact or Ulysses' Raft

- ▶ Pre-analysis plans are an Ulysses' pact
 - ▶ They bind the researcher to a plan
 - ▶ They prevent p -hacking but also the ability to explore
 - ▶ Registering them helps solve the file drawer problem
 - ▶ In econ, focus has been on RCTs
- ▶ Typical econ uses an Ulysses' raft
 - ▶ Allows researchers to explore messy social science data
 - ▶ Does not limit the questions we can ask to what we think we might have know ahead of time
 - ▶ Could register exploratory work to solve the file drawer problem
 - ▶ In econ, been excluded from pre-analysis and registry discussion

Ulysses' Packraft



Of immense value in a country where the want of wood renders it impossible to form any sort of raft.

— French explorer Joseph Renè Bellot

Value of the Packraft

- ▶ A packraft is a portable inflatable boat first designed in the 1840s for use in exploring the Canadian arctic
- ▶ The researcher develops a basic pre-analysis plan and files it with a public registry
- ▶ Unlike a binding pact, researchers are allowed to update the plan as new information arises as in ArXiv
- ▶ Transparency about deviations from the original plan allows for flexibility, while also addressing issues of p -hacking
- ▶ Encourages a disciplinary practice of telling readers at what point in the research process (and for what reason) a particular model emerged

Example 4: COVID-19 – A Packraft PAP

- ▶ Did livelihood diversification help households maintain welfare during the COVID-19 pandemic?
- ▶ 4 LSMS-ISA countries, 1 pre-COVID baseline round, and 12 COVID phone-survey rounds (~94,000 observations).
- ▶ The PAP could be filed after initial COVID rounds but before later rounds because data collection was ongoing and the sampling frame/questionnaire were known.
- ▶ Diversification, welfare, and heterogeneity could each be measured in many defensible ways; the plan narrowed the analysis before later outcomes were observed.
- ▶ Filed at: osf.io/nu593.

Example 4: COVID-19 – Designing for Messy Data

What the PAP organized

- ▶ Measures of income and livelihood diversification
- ▶ Welfare outcomes across multiple countries and survey rounds
- ▶ Heterogeneity analyses that were plausible ex ante
- ▶ A common framework for a large research team

What remained hard

- ▶ Too many proposed metrics
- ▶ Some variables could not be constructed
- ▶ Missing outcomes in some countries
- ▶ Weak alignment between pre- and post-COVID data
- ▶ No formal power calculations

Example 4: COVID-19 – Lessons for the Packraft

- ▶ The populated PAP was 80 pages long; the empirical results were mostly null and still required careful interpretation.
- ▶ The plan helped define complex concepts, coordinate and focus the team, and support student learning.
- ▶ Observational PAPs can easily become a “kitchen sink,” and when the data are not yours it is hard to guarantee that every planned variable will be feasible.
- ▶ The packraft model preserves flexibility, but requires transparency about when the analysis changed and why.

Promoting the Packraft

- ▶ AEA registry should allow for studies beyond RCTs
- ▶ Other association journals should mandate pre-registration for:
 - ▶ Studies that collect new data
 - ▶ Studies that claim to establish causal relationships
- ▶ Basic pre-analysis plans should be the norm for conducting good economic science
- ▶ Public registration of a study's objectives should be the norm in economics
- ▶ Registered plans should be allowed to deviate as long as deviations are documented

Ulysses' Journey Continues



Then go, taking with you a well-shaped oar,
until you come to people who know nothing of the sea.

— Homer, *The Odyssey*, Book XI