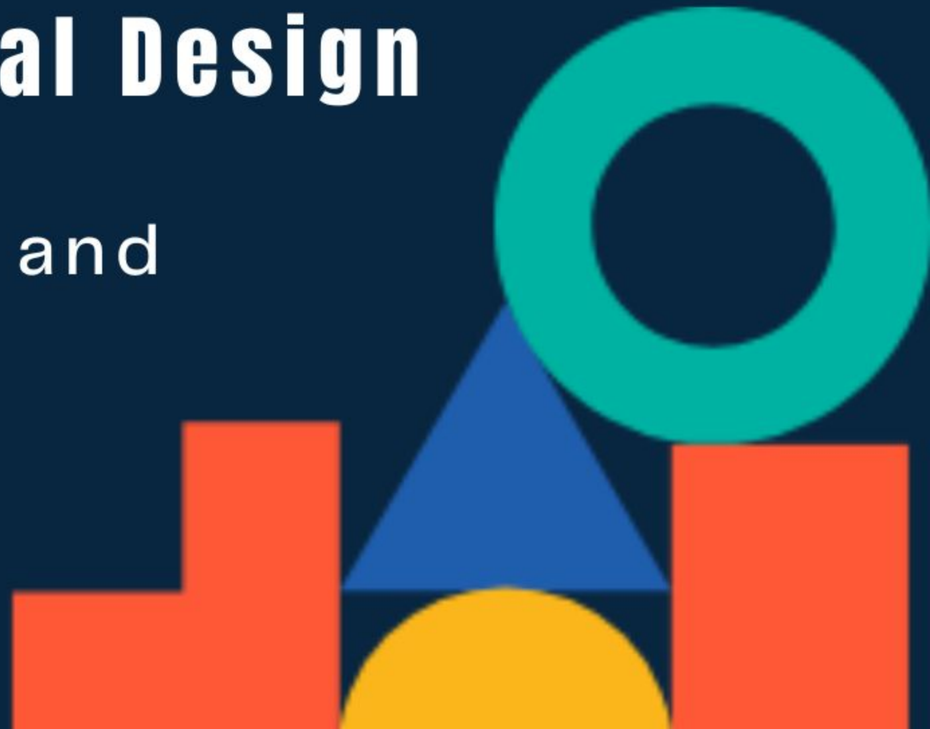
IGL

BEHIND THE SCENES

IGL PhD Community – Workshop Series

Pre-registration and Trial Design

with David McKenzie and
Amisha Miller



16 September 2025

Pre-registration and trial design

IGL Behind the Scenes Seminar

David McKenzie, *World Bank*

Tools for pre-commitment/pre-planning

- **Pre-registration of a study** – say what you are going to do before you see the data
- **Pre-analysis plan** – specify very precisely how you are going to the analysis
- **Registered report** – also get it peer-reviewed, write it up like a paper

Pre-registration

- For randomized experiments, main registry is the AEA RCT registry.
  - Mandatory now for submission to some journals (e.g. AEA journals:

> Research involving field experiments must be recorded in the AEA registry for randomized controlled trials (RCTs). (Laboratory experiments do not need to be registered at this time.)

  - Just requires basic information in mandatory fields:
    - Dates
    - Intervention details
    - Main outcomes – can be specified with more or less precision, up to authors
  - Over 10,000 registrations with locations in 168+ countries

INTERVENTIONS

Intervention(s)

Firms will be offered a mixture of information, training and assistance, and take-up incentives to encourage their take-up and usage of the new technology of digital marketing.

Experimental Design

Firms will be in one of four arms:

(1) Digital Marketing Training + Advertising Subsidy + Business Support
(2) Pay for Performance (Digital Marketing Presence and Advertising)
(3) Information Only
(4) Control

EXPERIMENT CHARACTERISTICS

Sample size: planned number of clusters

2550 firms

Sample size: planned number of observations

2550 small firms

Sample size (or number of clusters) by treatment arms

850 in training & subsidy, 850 in pay for performance, 300 in information, 550 in pure control

PRIMARY OUTCOMES

Primary Outcomes (end points)

We are interested in primary outcomes in two key domains:
Domain 1: Take-up and Usage of Digital Marketing Technology.
This will be measured by:
1. Percentage of firms with a Facebook page for the firm
2. Percentage of firms that have done paid advertising on Facebook in the last three months.

Domain 2: Firm performance
This will be measured by:
1. Sales made to new customers who have heard of firm through digital marketing efforts
2. Sales
3. Profits

Primary Outcomes (explanation)

These primary outcomes will be refined prior to collection of our first follow-up data, and will be explained in the pre-analysis plan

SECONDARY OUTCOMES

Secondary Outcomes (end points)

To be defined in pre-analysis plan

Specifying primary outcomes

- We often have lots of things we'd like to look at
  - Many different outcomes
  - Lots of different theories about heterogeneity
- I like to give myself and my co-authors what I call the Science/AER Insights/Policy brief test
  - Ask yourself – if at the end of this study, I was only allowed to show one short table or figure, what would be in it?
  - If you ask policymakers what outcomes matter most for them in making decisions based on this study, what would they say (and are you measuring these)?
- After all, you are not tying your hands very much if you say I intend to measure treatment effects on these 148 outcomes, and don't say whether some matter more than others.

Table 1. Impact of training programs on business survival, profitability, and sales. Data are from four rounds of surveys and show the average impacts over the 2.5 years after training. All regressions include randomization strata and survey wave dummies. Huber-White robust standard errors (in parentheses) are clustered at the firm level. Business survival is a binary indicator that takes the value 1 if the business survives. Sales are winsorized (capped) at the 99th percentile and profits at the 1st and 99th percentiles, reducing the influence of outliers. Sales and profits are expressed in terms of real CFA francs. The profits and sales index is the mean of the standardized *z*-scores of our various profits and sales measures. An *F* test was used to test equality of the impacts of the two training programs. *$P < 0.1$; **$P < 0.05$; ***$P < 0.01$.

	Business survival	Monthly sales	Monthly profits	Weekly profits	Profits and sales index
Traditional business training	−0.005	38,077	10,746	3086	0.029
	(0.008)	(57,812)	(6802)	(2050)	(0.030)
Personal initiative training	−0.003	114,733*	28,709***	6685***	0.100***
	(0.008)	(58,619)	(7110)	(1979)	(0.031)
Number of observations	5792	5642	5642	5633	5643
Number of firms	1499	1492	1492	1492	1492
P value from test of equality of treatments	0.813	0.171	0.014	0.091	0.025
Control group mean	0.960	680,807	96,089	30,417	0.000

Table 2. Mechanisms through which training operates. Huber-White robust standard errors (in parentheses) are clustered at the firm level. *$P < 0.1$; **$P < 0.05$; ***$P < 0.01$.

	Business practices	Personal initiative	Capital and labor inputs	Innovation index	Diversified product line	Access to finance index
Traditional business training	0.060*** (0.008)	0.065*** (0.015)	0.032* (0.020)	0.117*** (0.050)	0.044** (0.018)	0.070** (0.033)
Personal initiative training	0.054*** (0.007)	0.124*** (0.015)	0.078*** (0.020)	0.309*** (0.070)	0.092*** (0.018)	0.147*** (0.040)
Number of observations	5646	5538	5655	5639	5632	4207
Number of firms	1492	1484	1492	1492	1492	1473
P value from test of equality of treatments	0.458	0.000	0.024	0.011	0.010	0.043
Control group mean	0.618	4.32	0.000	0.000	0.335	0.000

What about sub-groups/heterogeneity?

- Useful to pre-specify if one or two dimensions that theory/policy strongly focused on.
- But typically lots of potential dimensions of heterogeneity that might apply
  - Which are of most interest depends on whether main purpose is policy targeting or understanding economic mechanism
    - E.g. heterogeneity by geographic region, age group, gender vs
      heterogeneity by risk aversion, level of credit constraints, baseline level of capital
- Most experiments have low power for looking at heterogeneity
  - Consider as exploratory only
  - Use machine learning methods to look at heterogeneity over many variables, rather than testing lots of bivariate hypotheses.

What about non-experimental studies?

- Only makes sense if you can credibly commit to not having seen the data when you specify analysis.
- Examples:
  - Policy change (new program for firms in certain industries) coming in, and specify how will use government survey data to be collected after the change.
  - Prospective impact evaluation:
    - E.g. regression-discontinuity design of some policy, and now planning follow-up survey to see what impacts are
    - E.g. prospective difference-in-differences design or synthetic control design – might pre-specify now how comparison groups be formed, and follow-up data not yet collected.
- Main place to register: Open science foundation (osf) (now includes EGAP)
- Alternative: Aspredicted.org

Pre-analysis plans

- A **pre-analysis plan** is a step-by-step plan setting out how a researcher will analyze data which is written in advance of them seeing this data (and ideally before collecting it in cases where the researcher is collecting the data).
- Typically more detailed than AEA registration:
  - Specify how data will be cleaned, exact regression specifications, exactly how variables will be constructed, etc.
- Some debate in the literature as to when/whether to do one and how extensive it should be:
  - Early pre-analysis plans often were 30+ pages, tried to pre-specify all eventualities
  - Olken (2005) suggests simpler plans may have most of advantages
  - Coffman and Niederle (2005) – may not be needed if easy to replicate – e.g. lab experiments.
  - Duflo et al. (2020) call for moderation

Pros of pre-analysis plans/when to use

- Most useful for field experiments that may be expensive and difficult to replicate
- As well as credibility, several other uses:
  - Helps focus policymakers on what key outcomes they most care about, and get agreement on this in advance rather than after seeing results.
  - Really helpful if you are designing questionnaires – helps make sure you are measuring everything you need to, and also on what can be cut from questionnaire if necessary (lots of “nice to ask” questions never get used).
  - Records upfront a lot of design and intervention details that may be harder to remember/reconstruct 2-3 years later.
  - Upfront investment in thinking through analysis can make it much faster/easier when data arrives to get headline results

What should go into a pre-analysis plan?

- See checklist: https://blogs.worldbank.org/impactevaluations/a-pre-analysis-plan-checklist

1) **Describe how sample is selected, expected sample size, how randomization is done**
2) **Key data sources and timing of data collection**
3) **Estimating equation:** e.g. what controls will be used in regression, how will standard errors be calculated, what adjustments will be used for multiple testing, etc.
4) **How will attrition be handled?**

What should go into a pre-analysis plan?

5) Hypotheses, families and outcomes:

- I see this as the most important.
- Separate outcomes into separate families/domains, and ideally also into primary and secondary outcomes, or main effects, mechanisms and heterogeneity.
- For outcomes, be really specific – so RA could create from these instructions:
  - E.g. don't just say "Wage earnings in last month", but make clear
    - Log, Levels, or some other transform? How will 0s be handled?
    - Any winsorizing to deal with outliers?
    - If you have multiple questions on this (e.g. earnings in main job, earnings in past week, etc) how will overall measure be constructed.
  - If forming summary index measures, define components of this and how will be constructed.

Example 1:

- YouWin business plan

FAMILY D: CHANGES IN BUSINESS SALES AND PROFITABILITY

HYPOTHESIS D1: Treatment leads to greater sales and profits in the medium term, but likely has no discernible impact in the first follow-up survey.

This will be measured as the following set of outcomes:

1. **Number of customers in a typical week** (B12). This will be top-coded at the 99th percentile of the overall distribution to account for outliers.
2. **Total sales in the last month with no truncation:** BF5. For businesses not answering the exact answer, but answer the range question, the midpoint of the range will be used. For firms in the top range, a value equal to the median of firms with sales in this top range will be used.
3. **Total sales in the last month truncated at the 99th percentile.** As in 2, except truncated at the top 99th percentile.
4. **Total sales in 2012 to date, truncated at the 99th percentile**. BF6 – measured as per 3.
5. **Sales are higher than one year ago.** BF7=3
6. **Total profits in the last month with no truncation:** BF9. For businesses not answering the exact answer, but answer the range question, the midpoint of the range will be used. For firms in the top range, a value equal to the median of firms with sales in this top range will be used.
7. **Total profits in the last month truncated at the 99th percentile.** As in 6, except truncated at the top 99th percentile.
8. **Total profits in the best month of the year, truncated at the 99th percentile.** BF10, measured as per 7.
9. **The inverse hyperbolic sine transformation** of total business profits in the past month $\log(y+(y^2+1)^{1/2})$ – which is similar to the log transformation, but can deal with zero profits. BF9. For businesses not answering the exact answer, but answer the range question, the midpoint of the range will be used. For firms in the top range, a value equal to the median of firms with sales in this top range will be used.
10. **Sales of main product in past month**. BF12b*BF12d
11. **Mark-up profit on main product in past month:** (BF12b-BF12c)*BF12d
12. A standardized **profits and sales** impact will be obtained by aggregating these different effects as described below in our methods section as a standardized z-score.

Example 2:

Business training program

Family A: Business Practices

These questions were asked at both the 1-2 month and 6-8 month follow-ups. However, questions indicated by * were only asked at the longer-term follow-up. For the short-term follow-up measure, the outcome will omit these questions.

- Index of Marketing Practices: the proportion of the following 11 marketing practices currently used by the business:
  - Monitored prices of a competitor's business (*bus4_1_1_6m*=1 *or bus4_1_2_6m=1*)
  - Monitored products of a competitor's business (*bus4_2_1_6m=1 or bus4_2_2_6m=1*)
  - Asked customers if there are other products they would like business to sell (*bus4_3*=1)
  - Spoke with an ex-client to find out why they had stopped purchasing (*bus4_4=1)*
  - Asked a vendor which products sell best (*bus4_5=1)*
  - Used a special offer to attract customers (*bus4_6=1)*
  - Did some form of publicity (*bus4_7=1)*
  - Compared the prices and quality offered by other vendors (*bus4_10=1)*
  - * Performed customer segmentation to help determine marketing strategy (*bus4_11=1).*
  - * Has a logo for their brand (*bus5_17=1)*
  - * Has a registered trademark (*bus5_18=1 or 3 (in progress))*
- Index of accounting and financial practices: the proportion of the following 11 accounting and finance practices used by the business:
  - Keeps written records (*bus5_1=1).* If *bus5* is missing, answer to *bus3* will be used (how do you keep accounts), with answers*<=4* taken to mean written records are kept.
  - Records every purchase and sale (*bus5_2=1)*
  - Uses records to find out how much money business has (*bus5_3=1)*
  - Uses records to know if sales of a product go up or down from one month to another (*bus5_4=1)*
  - Calculates how much each of the major products or services it sells costs the business (*bus5_5=1)*
  - Knows which products/services are the most profitable (*bus5_6=1)*
  - * Pays themselves a salary as an employee of the business (*bus5_9=1)*
  - Have records showing business would have enough money to pay off a loan (*bus5_11=1 bus5_8=1 in short-term follow-up)*
  - Has document detailing annual profits and loss of company (*bus5_14=1, bus5_11=1 in short-term follow-up)*
  - Tracks cash income annually (*bus5_15=1, bus5_12=1 in short-term follow-up)*
  - Separates household and personal finances (*bus5_16=1, bus5_15=1 in short-term follow-up)*

Some practical tips with PAPs

- Don’t overcomplicate and try and specify every eventuality
  - E.g. 1: Some outcomes/mechanisms only make sense/are of interest if you see impact on a key outcome first.
    - For example, a primary outcome of a management improvement program might be the number of workers in the firm.
      - IF you find the number of workers has increased, you might then want to look at whether they are trying new ways of hiring, the wages they pay workers and whether they pay for performance, whether they are hiring young women, etc.
      - IF you find employment falls, you might want to look at whether they are firing workers more versus just reducing hiring, at what types of workers they got rid of, at whether they are now using more capital instead of workers, etc.
      - IF you find no change in employment, then much less interesting to look at all these channels.
    - One approach would be to try and pre-specify complicated IF/ELSE plans, but hard to anticipate everything
    - Instead just focus on main outcomes, and then acknowledge that analysis understanding mechanisms/channels is exploratory.

Some practical tips with PAPs

- *Timing:*
  - Need more information than for AEA registry – so register project "early" (once getting underway) on the registry, and then add PAP after baseline/once understand intervention better (but before any follow-up data). Pilots can help.
  - Can update over time – e.g. add new version before second follow-up.
- *Write-up:* Too mechanical an adherence to the PAP makes for boring papers that include irrelevant information and exclude relevant information
  - Also reason for not specifying too long a PAP – no one wants to read a zillion appendix tables saying we pre-specified we would look at these extra 50 outcomes and 10 types of heterogeneity.
  - I want the authors to bring in descriptive data, qualitative information, new thoughts, and their own exploratory hypotheses to help explain the results they got - just so long as they indicate that this is exploratory and post-hoc
  - Duflo et al. (2020) suggested a separate "populated PAP" that can compliment the paper

Registered reports

- Even more work than a PAP – write up as much of the paper as you can without seeing results, and get it peer-reviewed
- Journal of Development Economics

> Registered Reports: The JDE offers authors the opportunity to have their prospective empirical projects reviewed and approved for publication before the results are known (referred to as 'Registered Reports'). This pre-results review track may be particularly suitable for authors working on research projects for which they have not yet collected or accessed data. Submissions in this track will follow existing policies outlined in the Author Information Pack, including the Mandatory Replication Policy, but specific information is available in the JDE Registered Reports Author Guidelines. A website including the Guidelines and information on Phase 1 acceptances to data is available here ↗ . To submit a Registered Report, select "Registered Report Stage I: Proposal" as the article type in the submission portal. "Registered Report Stage II: Full Article" should only be used for articles derived from accepted Stage I submissions.

JPE Micro Registered Reports

- https://www.journals.uchicago.edu/journals/jpemi/data-generation-guidelines

To address gaps in information about data generation, the editors of *JPE Micro* have developed a set of guidelines that will be put in place in early 2024. Authors of studies that generate their own data will be expected to register their project *before* data collection in one or more of the various trial registries *and* write a pre-analysis plan (PAP) *before* data collection. If a study is not pre-registered *and* has no PAP, then it should not be submitted to *JPE Micro* . **Studies submitted without the required pre-registration and PAP will be sent back to the author without a review.** [2]

Pre-registration is a publicly documented statement summarizing the study and the hypotheses the researcher plans to investigate before data are collected and analyzed. A subset of pre-registrations are **PAPs** , which offer greater detail about the researcher's plans. [3] If used appropriately, registries and PAPs can help to tackle key issues in the credibility crisis such as p-hacking and the file drawer problem.

Pre-Registration
The following Registration Checklist should be used in creating your pre-registration. More information may be included at the author's discretion:

- Experiment Name
- Country, State, or City
- Short description (< 280 characters)
- Intervention Start Date
- Intervention End Date
- Outcomes (Surrogates and End Points)
- Experimental Design, including treatment arms
- Planned Number of Observations
- Sampled Population
- Was IRB approval obtained (or exempt)? If approved:
  - IRB Name
  - IRB Approval Date
  - IRB Approval Number
- If exempt from IRB approval, please provide a brief explanation.

The AEA Registry is our preferred registry. However, social scientists have created a variety of specialized trial registries, summarized in Exhibit 1.

Allowing RR submissions places a particular trust in authors not to use the editors and referees of *JPE Micro* for their comments and then shop the finished product to other journals. In this spirit, our policy is strict: authors of accepted RRs can submit their study to the *Journal of Political Economy* as a *de novo* submission or they may send it back to *JPE Micro* . Submission to any other journal is strongly discouraged. We reserve the right to create a public list of authors who violate this policy.

Advantages of registered reports

- Get feedback on design and useful/constructive comments at a time where you can do something about it.
- Builds on PAP to get a lot of the write-up done in advance on paper and help you think through carefully what study is doing
- Acceptance without seeing results helps guard against difficulty of publishing null results, and can give you something earlier on to show for study that is going to take years
- May provide some commitment/protection against changes in policy partners objectives/openness to reporting

- *But is a lot of work – like registration/PAPs, makes more sense for longer-term, riskier, hard to replicate projects.*

Some tips on registered reports

- Introduction – should be written like a regular paper, making clear what this paper does, why it is interesting, and how it contributes to what we know from the existing literature
- Reports often lack key specifics:
  - Provide more context and outline the status quo/problem that the intervention is trying to solve: tell us the context and details of the sample, and whether there is a market failure or problem that needs to be solved
    - E.g. if you are doing an experiment on helping youth find jobs- tell us what the status quo process is like, what the background of the youth are, etc.
    - Very common for people to just put in table of summary statistics and not describe the sample at all or offer more details.
  - Unpack the black box – describe in detail the interventions
  - Outline a clear theory of change linked to hypotheses
    - This should also help justify choices of outcomes, and of timing of follow-ups – e.g. is it reasonable to expect impacts in 6 months?
  - Outcomes need to be defined precisely

Some tips on registered reports

- Power calculations often particularly problematic:
  - Discuss assumptions about take-up rates, control means and standard deviations, etc.
  - Discuss effect sizes in meaningful units where possible – e.g. what is the percentage point change in employment you can detect?
  - Relate MDEs to the existing literature to discuss reasonableness – e.g. don't just tell me your MDE is a 8% change in employment – but given the existing literature (and economic theory/cost effectiveness), is it reasonable to think the program will have this effect?

Conclusions

- Very rare in field experiment to just have single treatment and single outcome of interest

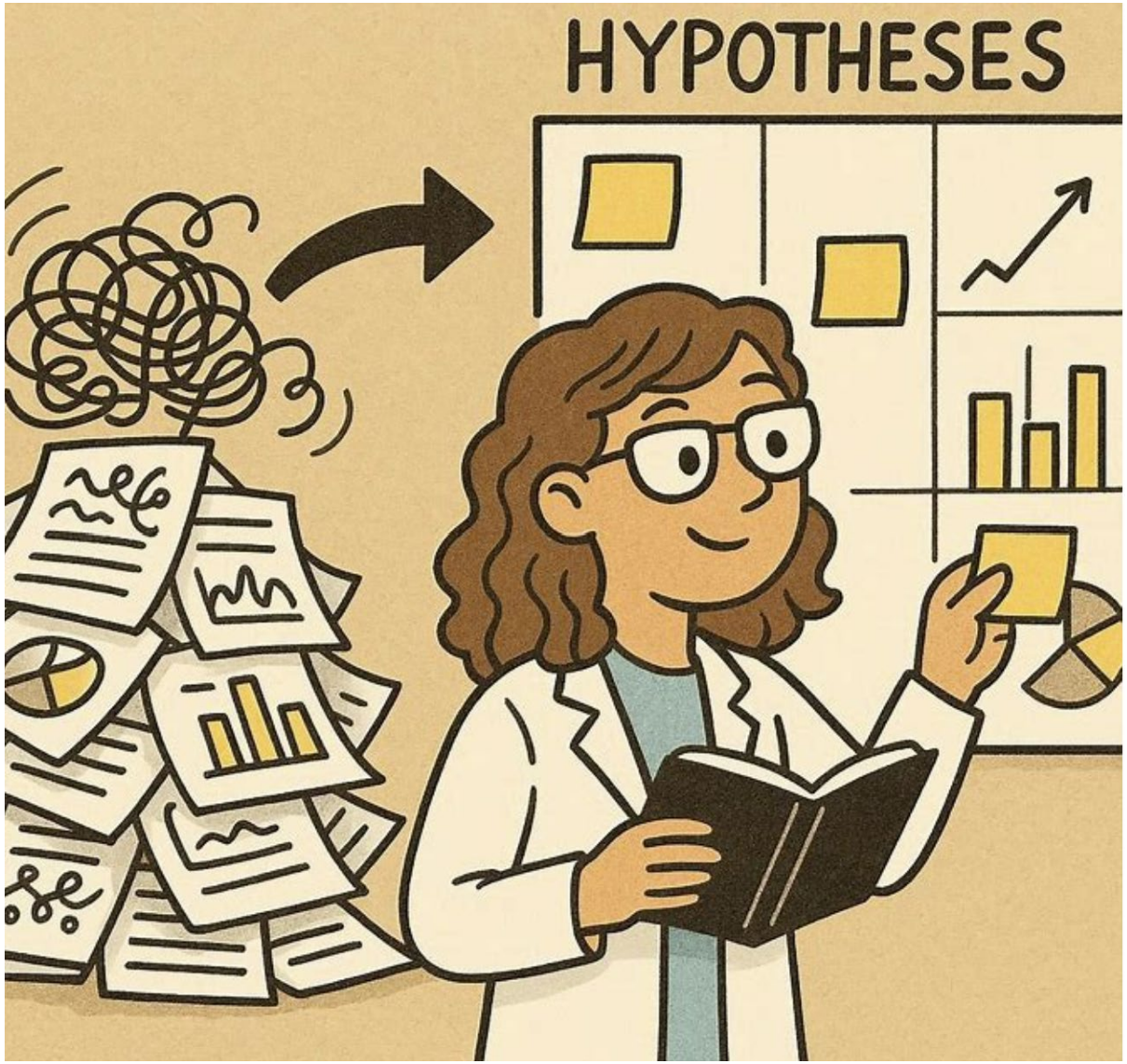
(contrast A/B testing in online experiments)

- A variety of different approaches for dealing with multiple testing (lecture 3).
  - “horses-for-courses” – different methods useful for different issues
- Thinking about how you will do this in advance is helpful both for your research, and for credibility
  - Pre-registration, PAP, and RR all helpful here in setting out key outcomes, thinking through what matters most.

Pre-Registration and Trial Design

September 16, 2025

Amisha Miller



Hi, I'm Amisha Miller

Research: How organizations evaluate potential and the ramifications for inequality

- Longitudinal field research, field experiments

Build from mixed methods and experience in the field:

- **Education:** Warwick (B.A. History), LSE (MSc. Population & Development), Boston University (PhD. Strategy & Innovation)
- **Professional experience:** Entrepreneur – UK | Entrepreneurship research & policy – UK, Brazil, US
- **Partnerships:** World Bank/IFC, investment firms, accelerators, entrepreneurs

→ I run field experiments to **link x and y causally**, based on organizational theories. BUT the constructs and mechanisms are rarely nailed down 100% before running the experiment

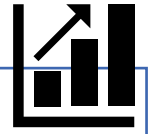
NYU

Management scholars' priorities are diverse

Psychology

- Experiments unpack **process theories** through careful design, main effects analysis, **moderation and mediation**
- Field experiments show the relevance of theory in the **real world** when many factors are at play, and for **specific populations**. Can be combined with lab experiments
- ***Significance and direction*** *of effects is important to uncover theoretical links. Incentives can often be aligned in the lab*

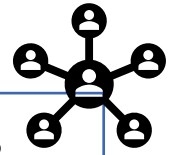
(Falk & Heckman 2009, Mitchell 2012)



Economics

- "Gold standard of evidence"
- Establish a **causal link** between and intervention and outcome
- Demonstrate how **effective interventions** are / whether a theory holds under **real-world** conditions
- *Replicate* ***incentives*** *and show the* ***size of effects*** *for policy implications*

(Banerjee & Duflo 2017, Hsu et al. 2017, Czibor et al. 2019)



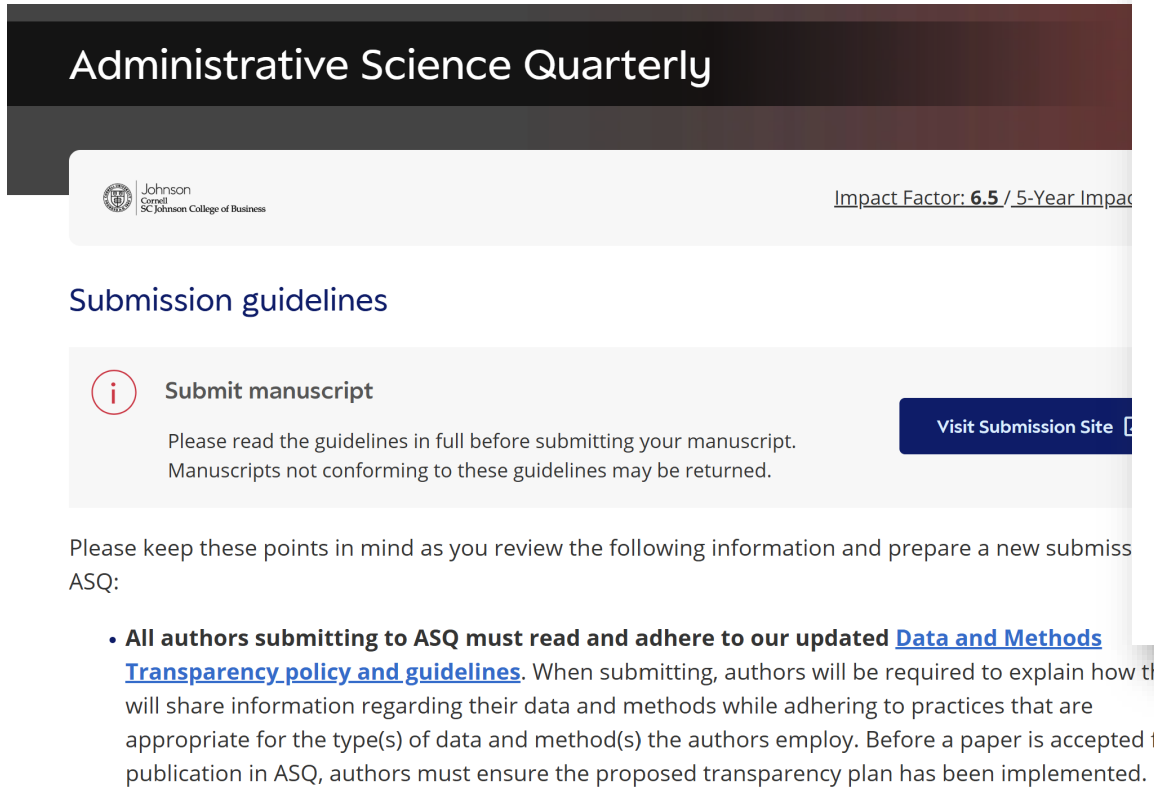
Organizational Theories

- Experiments can uncover **understudied topics** in "invisible domains", and can provide strong evidence for underlying processes
- Examining behavior in **controlled circumstances with robust measures** can cut through the complexity and ambiguity of organizations and identify causal relationships
- ***Openness to complexity*** *(e.g. factorial designs, and combining interviewing and observational techniques with experiments)*

(Levine Schilke Kacperczyk Zucker 2023)

NYU

Pre-registration for management:
Few norms



Guidelines for Pre-Registration

Our expectations about pre-registration take into account the diversity of methods and types of data used in *ASQ* articles. While we do not require pre-registration, we encourage it for experiments and original surveys, as pre-registration can strengthen other scholars' inferences about the research. For studies that test hypotheses using data that has already been collected, we value pre-registration of analytical plans before analyses are conducted.

Pre-registrations are useful for identifying what aspects of a study should be considered deductive and what parts are not. *ASQ* is open to manuscripts that test theory when designs permit but also engage in informed speculation based on unanticipated results. Pre-registration helps authors and readers distinguish these intellectual objectives. *ASQ*'s editorial team believes that we can advance knowledge by sharing positive, negative, and null results. We are, therefore, open to publishing studies based on pre-registered analyses that produce either supported or unsupported hypotheses. We ask authors to be transparent about which parts of a manuscript are consistent with their pre-registered research design and analyses and which parts are not. At the time of submission, you may share pre-registration by including a link in the manuscript to anonymous materials housed on a site such as OSF or aspredicted.org.

When authors want or need to deviate from pre-registrations, they should disclose this in the submission, keeping in mind that deviations do not make the study inadequate or necessarily invalidate the pre-registration. An editor or reviewers may request follow-up action based on such information, such as asking to see results from both pre-registered and exploratory analyses.

NYU

Pre-registration for management:
Issues with construct validity

- There are few field experiments to build on in some organizational or macro management theories
  - You might need to **register a novel intervention** – and you'll need to defend construct validity somehow
- Field sites work on different time scales to us
  - You **may not have time** to develop a perfectly theorized and validated construct to pre-register before you run the experiment
- Theories develop over time
  - Especially if you or other scholars **learn** between the field experiment and publication

Things I have learned…

From reviews (both sides), feedback on pre-registrations, making lots of mistakes

Define ONE PRIMARY: 1) outcome; 2) intervention; 3) empirical model

4) Add sample guesses; and 5) register (ltd.) supplementary analysis

NYU

Define ONE outcome per pre-reg

Don't do what I did

I pre-registered **two outcomes** – one for the bundled treatment and one for the first stage

- I had to explain that the bundled treatment outcome is the outcome, the first-stage outcome is an interim outcome (and post hoc, I now see it as a mechanism)
- I did not gain clarity on this before pre-registering

*DO BETTER: "This is the **primary outcome** (investment decision). I will also analyze a **supplementary outcome** (investor assessment) in the first stage, to re-examine theories from previous research"*

NYU

Define ONE outcome per pre-reg

Don't do what I did

I pre-registered **two outcomes** – one for the bundled treatment and one for the first stage

- I had to explain that the bundled treatment outcome is the outcome, the first-stage outcome is an interim outcome (and post hoc, I now see it as a mechanism)
- I did not gain clarity on this before pre-registering

*DO BETTER: "This is the **primary outcome** (investment decision). I will also analyze a **supplementary outcome** (investor assessment) in the first stage, to re-examine theories from previous research"*

If you need multiple outcomes

Include an adjustment / correction to analyze significance conservatively

- David's blog is really good at explaining this and pointing to research explaining your choices
- E.g., Bonferroni adjustment (e.g., Katz et al. 2007)

NYU

Define ONE intervention

Try to design **ONE strong intervention** (w/ multiple parts, arms, or levels of intervention)

→ maximize chances of finding an effect

→ minimize chances of confusing reviewers

Theorize WHY this intervention should work, and demonstrate **construct validity by registering a manipulation check**

- E.g., I theorize consistency would be important to fostering objective evaluation. Here is my consistency intervention in stage 1. Here is how I will check if investors were more consistent in their assessment stage 1. Here is how I will check if investors were more consistent in stage 2…

You may need to **post-hoc rationalize** why the intervention worked differently than you expected, which should inform your **contribution to theory development**

Pre-specify ONE Analysis

I like to register **ONE simple empirical model for the main effect:**

Outcome = intervention + error

Outcome = intervention + population + interaction + error

(This helps a varied reviewer pool understand what my starting point was, which may not have been a commonly used empirical model in e.g., psychology)

- I define the model based on previous research, and include clustered standard errors at the level of the treatment
- If I need **controls**, I explain why (small, varied sample)
- *Note: I am always fighting for power due to small field samples, so I love simplification*

NYU

Define the sample

I have never got the exact sample I designed, BUT I either:

1. **Explain the recruitment strategy**. Explain that my population (investors) select in or out at any point, and I'll take as many as show up at an event
2. Define a **population** and a **theoretical stopping rule** for collecting data

NYU

Define the sample

I have never got the exact sample I designed, BUT I either:

1. **Explain the recruitment strategy**. Explain that my population (investors) select in or out at any point, and I'll take as many as show up at an event
2. Define a **population** and a **theoretical stopping rule** for collecting data

I always:

- Explain what happened in the field to the sample
  - (i.e. more investors showed up because they were bored during Covid, tech glitches on video calls / with avatars caused attrition, etc.)
- Run a balance check before and after attrition on treatment

You may have the power to conduct more analysis on attrition (ITT, etc)

NYU

Supplementary analysis (Hedging)

EVERYTHING IN THE PRE-REGISTRATION NEEDS TO GO IN THE PAPER

(I had an Appendix N.

Do not recommend.)

Robustness check measures if there are other commonly used, good options

- E.g., I measured "female founder" as the person that presented the venture (following micro / OB literature I used to developing the intervention). I added robustness check measures i.e., one of the founders on the venture overview (which links to macro/ strategy literature on gendered financing outcomes in entrepreneurship)

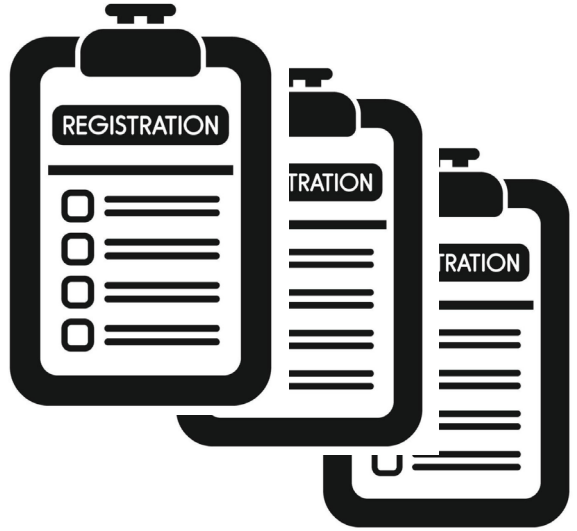
Heterogenous effects for populations if this is core to your theory. Pre-define a measurement for population

- E.g., High/low skill can drive heterogenous training effects, which I measure using SAT scores

Underlying mechanisms that you will explore/test using specific data

- Some reviewers love the cleanliness of field experiments and hate post-hoc analysis. Others call for more analysis of mechanisms to build generalizable theory
- Explain why the analysis is necessary (i.e. the theory doesn't have a clear mechanism). Explain your methods and cite other recently published field experiments

Creating multiple pre-registrations



I was taught to register every site separately BUT I heard from a friend that multiple pre-registrations looked suspicious to their reviewers

I create a new pre-reg if a **new setting is different in a way that might affect my experiment** (population, intervention, tech)

- E.g., I pre-registered an experiment in Mexico separately from the pilot in Brazil because: a) my intervention tech (avatars) did not work perfectly as planned in Brazil; b) I suspected the field site recruited investors in a slightly different way
  - In pre-reg 1, I mentioned that it was a pilot, and I would run further experiments
  - In pre-reg 2, I disclosed the other pre-reg, and registered pooled, exploratory analysis with both Mexico and Brazil, which included heterogeneity analysis on investor types

NYU

Updating pre-registrations

- **After baseline,** I often edit the pre-registration to further define sampling and the empirical model if needed

- You could add a PAP instead. I **would not** include this in the pre-registration document as I would be worried about how a diverse management reviewer pool might react / their expectations for reporting

NYU

Writing about pre-registrations

- Your pre-registration is not be 100% correct → we can develop theory in surprising ways!
- My reviewers have valued a story that links cleanly to the theory I end up developing. I use:
  - **Long introductions** to help reviewers understand where the paper is going from initial theory to results and theory development (no surprises)
  - **Methods** sections only include **main effects** and **my choice of supplementary/exploratory analysis**
  - I use pre-registrations, **footnotes and appendices** to make it clear what exploratory analysis I only reported in the appendix (and footnotes to explain why I did not prioritize this analysis in the paper)
- Changes can be difficult to explain to reviewers → good friends might help by checking footnotes and justifications!

Thank you!

Amisha Miller
amisha.miller@nyu.edu