

Frontier Questions, Personnel Data, and Student Research Designs

Organizations, Personnel, and Internal Labor Markets
Week 10

Sai Zhang

A study inside one organization can answer a broader question

Central question

What makes a study within an organization informative about a broader economic mechanism?

- Begin with an organizational decision and two competing explanations.
- Connect records to a comparison, then connect that comparison to a claim.
- Explain which features of the setting matter for another organization's decision.
- End with a feasible next evidence product and an explicit remaining limit.

Capstone route: decision → mechanisms → records → comparison → interpretation → portability → feasibility. The final product is a research memo with an access/validation appendix.

The capstone separates identification, measurement, and portability

- Define the initial population, treatment regime, outcome, horizon, and comparison.
- Distinguish a causal effect, a mechanism, a model primitive, and a target prediction.
- Build a realistic data request, validation plan, and design-based inference strategy.
- Defend a narrow research contribution with worker and firm incidence made explicit.

Core points

A database is not a design. A coefficient is not a mechanism. Reweighting cannot create a missing context or treatment version. A runnable synthetic example does not establish an empirical finding.

Students leave with a defensible proposal; the course does not require a completed publishable paper or promise publication.

Earlier weeks supply the objects needed to define a project

Decision	Indispensable record	Contribution to the memo
Allocation	Worker, job, manager, task, and time	A decision-maker and a labor outcome
Recruiting	Candidate/vacancy risk sets and dispositions	Initial eligibility and selection stages
Measurement	Production, quality, ratings, information dates	Independent validation and metric versions
Incentives	Contract, assigned eligibility, earned payout	A treatment separate from receipt
Authority	Reporting lines and decision-right histories	Assignment threats and institutional scope

Course synthesis, Weeks 1–5. Use the institutional detail that changes the comparison; a familiar estimator cannot substitute for these records.

Careers, teams, and work design determine the claim's boundaries

Decision	Indispensable record	Contribution to the memo
Careers	Openings, eligibility, applications, offers, starts, exits	Risk sets and competing events
Teams	Rosters, exposure, joint output, helper costs	Spillover boundary and assignment unit
Pay and voice	Pay components, options, access/use/remedy	Incidence and missing worker outcomes
Work design	Components, versions, implementation, horizon	Bundle contrast and transition costs
Research design	Source/target frames, permissions, validation	Feasible memo and limits to the claim

Course synthesis, Weeks 6–10. These are research objects, not a prescription to fit the same model in every setting.

Freeze the claim ledger before choosing an estimator

Field	What must be specified
Decision-maker	Who can change the rule, and what choice is feasible?
Initial population	Who was eligible before assignment or adoption?
Regimes	Assigned treatment, control, actual access, receipt, versions
Outcomes	Economic meaning, units, denominator, recording process
Horizon	Announcement, implementation, follow-up, later crossover
Comparison	Lottery, timing rule, threshold, or other actual variation
Claim and limit	Strongest supported sentence and tempting unsupported sentence

Distinguish worker, application, vacancy, team, establishment, firm, and decision episode. Randomization, implementation, receipt, and observation can occur at different units.

An estimand fixes the population, exposure, horizon, and weights

$$\tau_S(h, e) = \mathbb{E}[Y_{ifh}(1, e) - Y_{ifh}(0, e) \mid S_f = 1, i \in \mathcal{E}_{f0}].$$

- i : worker; f : organization; h : horizon; e : peer environment.
- $S_f = 1$: participating organization; $\mathcal{E}_{f0}$: initial eligibility set.
- Specify whether regimes describe assignment or receipt. Leavers remain in initial eligibility.

Effects (1, 9), initial employment (900, 100) : $\tau_{\text{site}} = 5, \quad \tau_{\text{worker}} = 1.8.$

Original weighting example: $(1 + 9)/2 = 5$ versus $(900 \cdot 1 + 100 \cdot 9)/1000 = 1.8$. Post-reform employment weights can change the estimand through selection. The displayed source parameter is not automatically identified in every design.

Each stronger claim needs additional evidence

Claim	Additional requirement
A recorded variable changed	Valid measure and a defined denominator
An assigned regime caused a change	Credible comparison, exposure, and inference
A mechanism generated the effect	Distinguishing predictions or mechanism-specific variation
A model primitive is recovered	Restrictions linking observables to the primitive
A target organization benefits	Target, overlap, comparable regimes, effect stability
Welfare improves	Resources, transfers, worker conditions, distribution, horizon

This is an evidence ledger, not a publication hierarchy. A mediator regression or a familiar mechanism name does not independently identify the mechanism.

A recorded gain can be production or a changed measurement rule

$$R_{it} = m_t(Q_{it}, K_{it}, \text{task}_{it}) + \nu_{it}.$$

- R : recorded metric; Q : actual production; K : quality; m_t : recording rule.
- ν : measurement noise. The reform can affect production, task mix, or the rule itself.

Invented explanation	Recorded gain	Independent audited gain
Real production improvement	10	10
Recording-only change	10	0

Require a validation measure with a different failure mode. A second field from the same logging pipeline may share its error. Randomizing the intervention and measuring R does not turn R into Q .

Rising survivor averages need not indicate anyone's improvement

Invented workforce	Initially	After one exit
Observed individual outputs	(80, 40)	(80)
Observed-worker mean	60	80
Total employer output	120	80

- The mean rises by 20; total employer output falls by 40; nobody improves.
- Preserve the initial cohort, exits, transfers, and follow-up dispositions.
- Zero output for this employer does not imply zero earnings or welfare at the destination.

Original composition diagnostic. A stayer-only comparison is selected after treatment; it does not establish a gain for the initial workforce.

The hybrid experiment is a claims audit with multiple endpoints

- **Setting:** Bloom–Han–Liang, eligible Trip.com employees.
- **Variation:** random assignment to a hybrid option on specified days, compared with the original office regime.
- **Outcomes:** retention, satisfaction, performance grades, promotions, and code measures; their samples and meanings differ.
- **Horizon:** distinguish the trial from later periods after controls receive access.
- **Limit:** the assigned option is neither universal remote work nor identical attendance by every assigned worker.

Bloom, Han, and Liang (2024), Nature 630:920–925. Required reading: audit eligibility, assignment, outcomes, inference, timing, and the availability statement.

Different endpoints and work arrangements support different claims

Comparison	What to check	What does not follow
Hybrid performance tests	Endpoint-specific equivalence margins and samples	Every nonsignificant result proves no harm
Hybrid promotion tests	Ordinary comparisons and uncertainty	Nonsignificance alone establishes equivalence
Later control access	Earlier versus later access histories	A permanent treated/never-treated contrast
Remote call-center evidence	Selection, office closure, distinct workers and regime	A contradiction with the hybrid option experiment

Bloom–Han–Liang (2024); Emanuel–Harrington (2024), AEJ: Applied Economics 16(4):528–559. Explain different populations, treatments, and outcomes before trying to reconcile coefficients.

A one-firm contribution does not require a universal effect

- **Hoffman–Stanton:** a synthesis of personnel practices, productivity, and generalizability; it is not a new experiment on every practice.
- A selected setting can reveal a decision process, separate explanations, identify a local effect, or recover a primitive under explicit restrictions.
- Insider knowledge can reveal assignment rules and measurement failures; access does not create exogeneity.

Write a scope condition

The effect concerns eligible workers facing A rather than B over H. The mechanism may operate where C and D hold; its magnitude could change with E, which this study does not identify.

Required: Hoffman–Stanton (2025), selected Handbook synthesis/external-validity sections. Companion: Ichniowski–Shaw (2009), specified NBER 15618. Count versions of a work once.

Partnerships, reported pilots, and worker follow-up select differently

Relevant firms → reachable → willing partners → eligible sites → observed workers.

- **Site selection:** which organizations enter the experiment? Keep a permitted recruitment-stage log.
- **Evidence selection:** which pilots and results are retained or reported? Preserve a full pilot registry.
- **Follow-up selection:** which workers have later outcomes? Retain response and exit records.
- A within-partner lottery randomizes none of the earlier partnership decisions.

A pilot registry does not recover missing employee outcomes or effects at unrepresented firms. Reasons for nonparticipation may be sensitive; publish only permitted information.

Allcott separates within-site efficacy from predictions for later sites

- **Setting:** energy-conservation programs at utility sites.
- **Variation:** experimental comparisons within sites combined with a selected sequence of program sites.
- **Target:** whether effects learned at evaluated sites predict effects at later sites.
- **Limit:** site participation and willingness to partner are not randomly assigned.

Personnel-data translation

Record how participating employers differ in baseline support, constraints, implementation, and recruitment timing; do not assume every difference is observed or causal.

Allcott (2015), Site Selection Bias in Program Evaluation, QJE 130(3):1117–1165. Required methodological comparator; this is not a personnel experiment.

The evidence base can select the most favorable pilot

- **DellaVigna–Linos:** comprehensive trials at two nudge units compared with published academic evidence.
- This comparison studies which results enter an evidence base; it does not randomize small versus large implementation.

Invented pilots $(-4, -1, 0, 1, 4, 12)$: $\bar{\tau} = 2$, $\max \hat{\tau} = 12$.

- A separately stipulated holdout effect of 2 illustrates how a selected winner can mislead.
- Neither this toy maximum nor the paper's decomposition is a universal correction for personnel studies.

DellaVigna–Linos (2022), *Econometrica* 90(1):81–116, published version. The six values and holdout are invented; this is not an estimator of publication bias.

Transport needs stable effects as well as source identification

$$\mathbb{E}[Y(1) - Y(0) \mid X = x, T] = \tau_S(x), \quad \tau_T = \mathbb{E}_T[\tau_S(X)].$$

- S : source; T : defined target; X : measured pretreatment effect modifiers.
- Identify source effects, then require compatible treatment versions, outcome definitions, and horizons.
- Require conditional mean-effect stability and source support covering the target.
- Equal untreated outcomes do not establish equal treatment responses; these are distinct assumptions.

Effect stability is an assumption about unobserved target responses. Balance, job titles, a precise source estimate, or a familiar mechanism do not prove it. Fix exposure regimes as well as own treatment.

Changing the context mix changes the population average

Implementation resources	Conditional effect	Source share	Target share
High support	8	2/3	1/4
Low support	2	1/3	3/4

$$\tau_S = \frac{2}{3}(8) + \frac{1}{3}(2) = 6, \quad \tau_T = \frac{1}{4}(8) + \frac{3}{4}(2) = 3.5.$$

$$r(x) = \frac{f_T(x)}{f_S(x)}, \quad r_H = 0.375, \quad r_L = 2.25.$$

Original hypothetical effects and populations. “High support” describes implementation resources; both high and low contexts are in statistical support. The target interpretation requires the assumptions on the preceding slide.

Weight concentration does not change the randomization unit

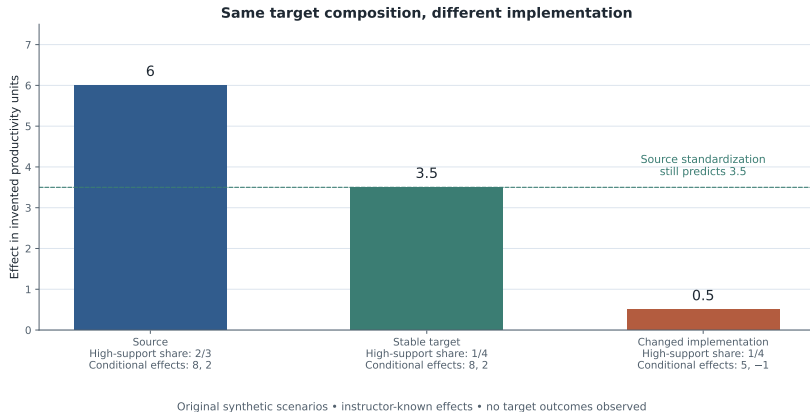
$$8(0.375) + 4(2.25) = 12, \quad 8(0.375)^2 + 4(2.25)^2 = 21.375.$$

$$n_{\text{eff}} = \frac{(\sum_f r_f)^2}{\sum_f r_f^2} = \frac{144}{21.375} \approx 6.737.$$

- The example has 8 high-support and 4 low-support source establishments.
- The target gives greater weight to the less common source context.
- ESS diagnoses concentration; it is not 6.737 independent clusters or degrees of freedom.
- The experiment still has six independent matched-pair allocations.

Original density-ratio calculation. Source randomization does not reveal unobserved target outcomes; reweighting changes which source effects receive weight.

Balanced observed contexts can conceal changed implementation



Original teaching worlds: source effects (8, 2) imply 6 in the source and 3.5 in the stable target. If target effects are (5, -1), its effect is $0.25(5) + 0.75(-1) = 0.5$. Instructor-known scenarios are distinct from realized estimates; recorded balance cannot discover this hidden shift.

Absent contexts require a narrower target or an explicit range

$$\tau_T = 0.8\tau_{\text{supported}} + 0.2\tau_{\text{absent}}.$$

$$\tau_{\text{supported}} = 3.5, \quad \tau_{\text{absent}} \in [-20, 5] \implies \tau_T \in [-1.2, 3.8].$$

- Twenty percent of this invented target has no comparable source context.
- The endpoints depend on the assumed effect bounds for that absent context.
- This identification range omits sampling uncertainty in the supported effect.
- More precise source estimates do not identify the absent-context effect.

Original assumption-indexed example, not a confidence interval. Trimming the target can be the honest substantive answer; silently extrapolating is a new assumption.

Population adjustment and treatment comparability are separate

Comparator	Capstone use	Boundary
Hotz–Imbens–Mortimer	Training-program comparisons across populations and locations	Similar untreated outcomes do not imply identical treatment components or responses
Vivalt	Heterogeneity across impact evaluations and study characteristics	Observed site modifiers are not randomly assigned; associations need not be causal
Next organization	Select a setting with a proposed constraint that differs	A deliberately contrasting site does not randomize its institutions

Hotz–Imbens–Mortimer (2005), *Journal of Econometrics* 125:241–270; Vivalt (2020), *JEEA* 18(6):3045–3089. A precise pooled mean need not be a precise prediction for a new organization.

Deep personnel records and broad surveys answer different questions

Data source	Potential contribution	Persistent blind spot
One employer	Detailed decisions, risk sets, task processes	Selected setting and missing destinations
WMS / MOPS	Broader management measurement and institutional variation	Comparability, response, selected universes
Linked employer–worker data	Conditions, earnings, moves, administrative follow-up	Linkage selection; missing internal alternatives
Commercial histories	Reported transitions across organizations	Incomplete workforce and decision processes

Bloom–Van Reenen (2007); Bloom et al. (2019), What Drives Differences in Management Practices?, AER 109(5):1648–1683. Dispersion or explained productivity variation is not the causal return to installing a practice.

An access map separates documentation from usable records

Route	Available starting point	Verify before claiming access
WMS	Public-data and methodology pointers	Coverage, terms, and separate special access
Census MOPS	Program documentation and statistics	Confidential establishment-data requirements
IAB LPP / LPP-ADIAB	Survey/linkage catalogue	Version, linkage population, research route
Campus File	Software/data-handling practice	Anonymization makes it unsuitable for substantive analysis or publication
Article archives	Named data/code or supporting-file links	Actual files, permissions, object, executability

Official resource routes in the Week 10 data-access map, checked in the source pack on September 22, 2026. A public listing does not establish that all required microdata are accessible or that a result has been reproduced.

Build a minimal data request backward from the comparison

- Request a schema or synthetic example first: row unit, stable keys, creation timestamps, coverage, and metric revisions.
- For each required object, record its owner, economic meaning, sensitivity, missingness, and fallback.
- Label status as **available and checked**, **permission pending**, or **not observed**.
- Check joins, duplicate IDs, interval overlaps, initial risk sets, outcome versions, and missing dispositions.

Access is a design milestone

Agree permitted uses, secure processing, null-result publication, disclosure review, and retention arrangements through the appropriate review process.

A journal policy is not data-use permission. Do not export confidential text or linkage information to an external service without authorization for that processing. The memo template itself authorizes no contact or data transfer.

A discriminating design makes a favored explanation vulnerable

Competing mechanisms	Useful additional record	Claim to avoid
Information versus effort	Comprehension plus independent production audit	A usage regression identifies mediation
Opportunity versus ability	Eligibility, applications, slots, offers, later skill	Promotees reveal all eligible workers' gains
Manager skill versus sorting	Reassignments, independent skills, comparable tasks	Noise correction makes assignment random
Capacity versus incentives	Queues, review capacity, costs, component regimes	A bundle effect identifies every complement

Weidmann et al. (2026): manager–team randomization, selection experiment, and field validation are separate layers.
Kline–Saggio–Sølvsten (2020) correct specific variance-component noise problems, not general assignment endogeneity.

Inference should reproduce the actual source of variation

- Preserve randomized blocks and exposure boundaries; analyze assignment before endogenous use.
- For rollout, specify cohorts, treatment histories, controls, anticipation, and supported horizons.
- Modern DiD does not automatically handle endogenous timing, arbitrary reversals, or interference.
- Cluster corrections address dependence; they do not add independent assignments or identify effects.
- Simulate power using the actual assignment, serial dependence, and planned estimator.

Cameron–Miller (2015); Callaway–Sant’Anna (2021); Roth (2022). Nonrejection of a pretrend test does not prove parallel trends. Transactions and worker-days do not create additional randomized firms.

Pre-specification and sensitivity make the analysis history legible

- Fix primary outcomes, adjustment rules, heterogeneity families, weighting targets, and the uncertainty method before outcome-driven choices.
- Date amendments and separate confirmatory results from exploratory discoveries.
- Ask which model moments and restrictions drive a structural counterfactual.
- Sensitivity to a moment is not an independent identification test of that restriction.

A sufficient contribution can be reduced form

Use a structural calculation when the question needs its additional restrictions; do not add one merely to make the proposal look more sophisticated.

Olken (2015), Promises and Perils of Pre-Analysis Plans; Andrews–Gentzkow–Shapiro (2017), Measuring the Sensitivity of Parameter Estimates to Estimation Moments.

Missing outcomes can leave even the sign unidentified

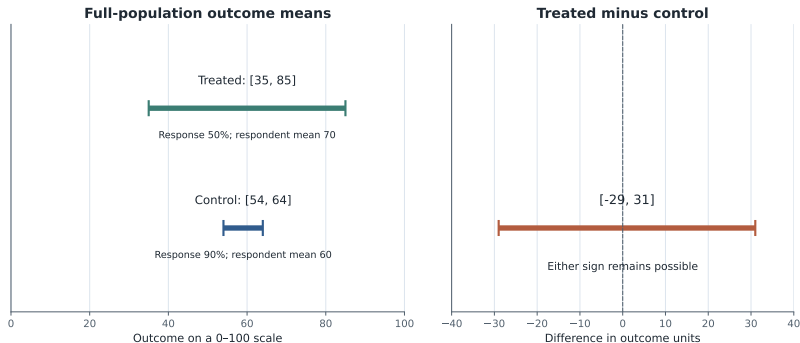
$$\mu_z \in [p_z \bar{Y}_{z,\text{obs}} + (1 - p_z)L, \quad p_z \bar{Y}_{z,\text{obs}} + (1 - p_z)U].$$

Invented group	Response p_z	Observed mean	Full mean range
Treated	0.5	70	[35, 85]
Control	0.9	60	[54, 64]

$$L = 0, \quad U = 100 : \quad \mu_1 - \mu_0 \in [35 - 64, 85 - 54] = [-29, 31].$$

Arbitrary missingness on a known 0–100 scale: the observed difference is 10, but its population sign is not identified. These are identification ranges, not sampling confidence intervals or Lee bounds.

An assumption range is not an uncertainty interval



Arbitrary missingness + bounded outcomes • identification ranges, not confidence intervals or Lee bounds

Original synthetic example • respondent difference of 10 does not determine the population sign

Original bounded-outcome example. Sampling uncertainty is separate. Lee-style trimming requires additional selection assumptions and concerns a principal stratum; it does not automatically recover welfare for the initial workforce (Lee, 2009).

A transfer changes incidence without adding joint resources

Invented monetary ledger	Firm	Workers	Combined
Verified output value	12	0	12
Setup and operating resources	-4 - 2	0	-6
Additional pay	-2	2	0
Burden and amenity	0	-3 + 1	-2
Total	4	0	4

- Define valuation, horizon, and accounting boundary; this is not a welfare estimate.
- Leavers are a subset of the initial workforce, not an additional workforce to sum again.

Original monetary illustration assumes the stated valuations and omits taxes and external effects. Productivity units in the main lab have no automatic monetary conversion. Unobserved destination welfare is not zero.

Training and outside-option evidence require distinct outcome claims

Anchor	Capstone use	Interpretation limit
Soft-skills training	Randomized workplace training; productivity, promotion, wage, and retention outcomes	Published productivity/promotion gains do not imply wage or retention gains
Outside-options model	Summarize opportunities under a matching model	The index is not an inventory of actual available offers
New proposal	Combine employer outcomes with confidential worker follow-up	Pay and retention alone do not reveal worker welfare

Adhvaryu–Kala–Nyshadham (2023), published JPE version; Caldwell–Danieli (2024), REStud. State what could reveal worker gains after exit and what remains unobserved.

The pilot effect can differ from the effect of full rollout

$$Y(z, e) = 100 + 4z - 3e, \quad z \in \{0, 1\}, \quad e \in [0, 1].$$

$$Y(1, e) - Y(0, e) = 4, \quad Y(1, 1) - Y(0, 0) = 1.$$

- z : own assignment; e : specified peer-exposure share.
- Holding exposure fixed gives a direct effect of four; changing own and peer regimes gives a rollout effect of one.
- Shared review capacity, helpers, and competitors can make these different policy objects.
- A pilot-to-rollout prediction needs a defined exposure regime, not only a larger sample.

Original spillover example. In actual designs, ask whether both required exposure states have support and whether an exposure mapping adequately represents interference.

Two observed regimes leave the mechanism decomposition unresolved

$$\underbrace{0}_{\text{tool}} + \underbrace{0}_{\text{redesign}} + \underbrace{12}_{\text{interaction}} = 12 = \underbrace{-4}_{\text{tool}} + \underbrace{3}_{\text{redesign}} + \underbrace{13}_{\text{interaction}} .$$

- Both original decompositions fit a no-change versus bundle contrast of 12.
- A missing mixed regime cannot be reconstructed from a positive bundle effect alone.
- Design feasible component comparisons or measure predictions that distinguish the proposed constraint.

Original observational-equivalence illustration. Bloom et al. (2013) assigns management assistance rather than each practice; Atkin et al. (2017) uses sequential technology and later worker-payment interventions rather than a full factorial of all components.

Frontier A: observe opportunity before an internal application

- **Question:** does changing information or release approval expand access, improve allocation, or simply redistribute fixed slots?
- **Proposed design:** change an information/approval rule for an initial eligibility set, with opening supply and competition included in the estimand.
- **Records:** who sees openings, eligibility, permission, applications, offers, starts, independent skills, and displaced candidates.
- **Limit:** applications are not skill gains; promotees are selected; slot competition creates spillovers.

Haegle (2026), talent hoarding in one internal labor market; Benson–Li–Shue (2019), old- versus new-job allocation across 131 firms. Their samples and identified margins are not interchangeable. This slide proposes a next study.

Frontier B: separate a tool from a reorganized job

- **Question:** does access speed a task, or does coordinated review/routing redesign increase verified throughput?
- **Proposed design:** feasible tool-access and workflow regimes with independent assignment where possible.
- **Records:** assigned access, versions, nonuse, task mix, queues, independent defects, labor time, and peer exposure.
- **Limit:** app time is not output; individual access does not independently vary organizational complements.

Brynjolfsson–Li–Raymond (2025): customer-support rollout evidence, not an RCT. Dillon et al.: multi-firm randomized individual access, specified November 2025 NBER/arXiv v4 revision. Proposed complementary redesign is not a finding from either design.

Frontier C: measure learners and the coworkers who help them

- **Question:** who gains skills and who bears the production/time cost of mentoring?
- **Proposed design:** vary mentoring schedules or credit rules, accounting for shared helpers and task assignment.
- **Records:** both sides' time, feedback, task-quality audits, completed work, and later independent skill measures.
- **Limit:** selection into helping remains a threat; messages, learning, and helper costs are distinct outcomes.

Emanuel–Harrington–Pallais (2026), published QJE 141(3):1825–1870. Initial closure/feedback evidence differs from later return-to-office code-quality analysis; those quality outcomes are not available around the initial closures.

Frontier D: distinguish safe reporting from underlying harm

- **Question:** does a communication channel change access, use, response, resolution, or the problems themselves?
- **Proposed design:** an access intervention with independent problem audits and confidential initial-worker follow-up.
- **Records:** availability, use, response times, remedies, underlying conditions, survey response, and exits.
- **Limit:** more complaints can mean safer reporting; low use does not establish no benefit to nonusers.

Adhvaryu, Gade, Molina, and Nyshadham (2026), *Sotto Voce*, EJ 136(674):798–812. Access is randomized, use is not; option value is a mechanism interpretation. A new project needs an appropriate ethics/confidentiality plan.

Frontier E: test why a common reform travels differently

- **Question:** which measured implementation constraint changes the response to a specified reform?
- **Proposed design:** recruit deliberately contrasting organizations and randomize the reform within supported contexts.
- **Records:** recruitment stages, pretreatment constraints, common intervention stages, independent outcomes, versions, and target context.
- **Limit:** reform randomization does not randomize organizational context; cross-site heterogeneity needs a separate causal argument.

Allcott (2015); Bloom et al. (2019). A multi-site study contributes when it makes a modifier observable and preserves comparability, not merely when it adds partner names. These five frontiers are proposed directions, not claims that the questions are unstudied.

The laboratory separates a source experiment from a target frame

12 source sites, 6 pairs, 1 treated per pair, $2^6 = 64$.

- Four pairs have high implementation support; two have low support.
- All potential outcomes are fixed before a separate assignment draw; all 64 allocations are equally likely.
- Twenty target establishments: five high support and fifteen low support, with no target outcomes.
- Complete source follow-up, stable measurement, and no cross-establishment interference are explicit toy assumptions.

Original synthetic reconstruction motivated by Allcott's design distinction. It does not reproduce any paper's data, protocol, outcomes, or estimates. Firms are not randomly sampled into the source frame.

Observed pair contrasts estimate two weighted source objects

$$\hat{d}_p = (Y_{p,T,\text{post}} - Y_{p,T,\text{pre}}) - (Y_{p,C,\text{post}} - Y_{p,C,\text{pre}}).$$

$$\hat{\tau}_S = \frac{2}{3}\hat{\tau}_H + \frac{1}{3}\hat{\tau}_L, \quad \hat{\tau}_{T,\text{std}} = \frac{1}{4}\hat{\tau}_H + \frac{3}{4}\hat{\tau}_L.$$

- p : matched pair; T, C : assigned treated/control site; H, L : context means.
- Both formulas estimate predetermined weighted effects for source establishments.
- Interpreting the second as an unobserved target effect additionally requires stability and compatibility.
- The change-score adjustment rule is fixed before assignment; randomization supplies identification.

Student inputs: `data/teaching/source_sites.csv` and `data/teaching/target_frame.csv`. Use observed changes only; instructor potential outcomes never enter the student estimator.

Target weighting concentrates inference on two low-support pairs

$$\hat{\tau}(a) = \sum_{c \in \{H, L\}} a_c \hat{\tau}_c, \quad \hat{V}(a) = \sum_{c \in \{H, L\}} a_c^2 \frac{s_c^2}{P_c}.$$

- a_c : fixed source or target context mass; s_c^2 : sample variance of pair contrasts in context c .
- There are $P_H = 4$ and $P_L = 2$ independent pair allocations in the two contexts.
- Heterogeneous pair-average effects make this variance estimator conservative in expectation here.
- A realized variance estimate need not exceed true design variance; six pairs do not justify automatic normal bands.

Design-aware uncertainty concerns the weighted source estimator. Its standard error does not account for failure of transport assumptions or supply sampling uncertainty for an unobserved target.

A target-weighted statistic still tests a source sharp null

$$H_0 : Y_f(1) = Y_f(0) \quad \text{for every source establishment } f.$$

- Hold observed changes and predetermined source/target masses fixed; relabel only within each original pair.
- Enumerate 64 allocations and count two-sided statistics at least as extreme, including ties and the realized allocation.
- Both source and target-weighted statistics test this global source sharp null.
- Neither tests transport stability or a standalone hypothesis about the unobserved target effect.

The separate instructor audit recomputes potential outcomes across allocations to check estimator expectations and true design variance. It is not the fixed-outcome Fisher reference; oracle variance is not an empirical standard error.

Reproduce the design, diagnose its limits, then transfer the logic

- **Reproduce:** observed sites and target frame $\rightarrow$ pair/context contrasts $\rightarrow$ weighted estimates and uncertainty.
- **Diagnose:** changed implementation, missing contexts, worker response, logging, composition, weights, spillovers, and incidence.
- **Verify:** all 67 supplied assertions; observed-CSV parity; two identical default runs; seed 88771; Python compilation; overwrite refusal with unchanged files.
- **Transfer:** use one frontier to specify two mechanisms, an access fallback, and the strongest and unsupported claims.

The original offline runner uses the Python standard library. Significance and realized equality to instructor truth are not acceptance criteria. A successful synthetic test verifies declared logic, not empirical transportability.

Reproducibility must describe what another researcher can rerun

Mode	Evidence product	Boundary
Exact or reduced replication	Named object from inspected original inputs and programs	Version, files, terms, and execution must be documented
Published-statistic reconstruction	Arithmetic from reported summaries	No recovery of unavailable microdata comparisons
Synthetic demonstration	Executable design and measurement logic	No empirical effect reproduced
Restricted-data workflow	Permitted outputs, versions, code, acquisition route	Public code does not imply public records

Week 10 uses original synthetic reconstruction. Original research microdata and estimation programs were not downloaded or executed for this week. AEA Data and Code Availability Policy guides documentation, not data acquisition permission.

Three required readings support three different capstone tasks

Required reading	Prepare for the workshop
Hoffman–Stanton (2025), selected synthesis/external-validity sections	State the broader labor contribution and a scope condition
Allcott (2015), Site Selection Bias in Program Evaluation	Separate within-site identification, selected sites, and later-site prediction
Bloom–Han–Liang (2024), hybrid-work study	Audit eligibility, assigned option, outcomes, horizon, inference, and data availability

- Each project adds one personnel anchor, one needed method, and one actual resource page.

Week 10's reference bank has 28 scholarly works and six official resources. Focused companions include Hotz–Imbens–Mortimer, DellaVigna–Linos, Emanuel–Harrington, and Bloom et al. (2019); the full bank is not compulsory reading.

The four-page memo makes the next research step concrete

Page	Main job	Required objects
1	Economic decision and contribution	Two mechanisms, literature position, strongest/unsupported claims
2	Population, timing, measurement	Eligibility, units, regimes, independent outcomes, missingness, event schema
3	Comparison and inference	Estimand, assumptions, independent variation, discriminating evidence, sensitivity
4	Portability and feasible next step	Source/target, modifiers, incidence, next-month product, fallback

Separate appendices: A, minimal data-access map; B, validation/reproducibility plan; C, source-to-target comparison. The workshop appendix gives the complete prompts. A feasible proposal is the deliverable, not a finished paper.

The defense ends with a supported claim and an explicit limit

Strongest feasible claim: proposed access experiment

With the stated assignment and follow-up, we could estimate the offer's effect on independently audited cases for initially eligible workers over the trial horizon.

Tempting claim the same design cannot support

The tool raises productivity and worker welfare at every employer through a separately identified learning mechanism.

- Review the largest remaining threat and the most valuable next observation.
- Give one justified revision and one unresolved limitation in the written response.

Suggested workshop: eight-minute presentation, five minutes of questions, one-page response. Verdicts: proceed with the scoped design; narrow the question/target; or obtain a missing comparison/measurement.

A defensible research plan is the course's final product

- Name what changes, for whom, relative to what, and over which horizon.
- Make competing mechanisms observable and validate the outcome independently.
- Match inference to the design; separate source identification from target prediction.
- Record who gains, who bears costs, what is missing, and what can be executed now.

The next step is your project

Organizational detail becomes a labor-economics contribution when it makes a question answerable and the limits of the answer visible.

End of the ten-week course. The following eight appendix frames support memo writing and peer review; they do not add another lecture or a publication guarantee.

Workshop appendix: memo page 1 defines the economic contribution

- **Working title and decision-maker:** what can the organization actually change?
- **Labor outcome and question:** begin with the choice and unresolved distinction.
- **Two competing mechanisms:** write a distinct observable implication for each.
- **Literature position:** one personnel application, one method, and one data/resource reference.
- **Strongest feasible claim:** one sentence the proposed design could support.
- **Unsupported claim:** one tempting sentence requiring evidence you do not have.

Research-memo template, page 1. A new country, dataset, estimator, or prominent partner alone does not name the contribution. Explain what readers could learn beyond the partner's internal report.

Workshop appendix: memo page 2 makes observation auditable

- **Population and units:** initial eligibility, observation unit, and assignment unit.
- **Regimes:** treatment/control components, access, timing, versions, and possible crossover; assignment is separate from use.
- **Two outcome families:** primary and independent validation, with units, denominator, horizon, owner, creation time, and revisions.
- **Missingness:** unobserved candidates, workers, sites, exits, destinations, and response.
- **Event ledger:** mock schema with stable IDs and timestamps; use synthetic records unless real access is authorized.

Research-memo template, page 2. A useful schema includes initial eligibility, assignment/exposure history, outcome creation, and dispositions; a file of current employees alone may omit the relevant risk set.

Workshop appendix: memo page 3 defends the comparison

- **Variation:** which comparisons are permitted, and why would they reveal the counterfactual?
- **Estimand:** population, treatment/exposure regime, weighting rule, and horizon.
- **Assumptions:** separate assignment, measurement, interference, and attrition; name the largest remaining threat.
- **Inference:** independent units, strata/clusters, test null, uncertainty, and design-based power simulation.
- **Discriminating evidence:** which pattern weakens the favored mechanism; which proposed control is post-treatment?
- **Sensitivity:** justify alternatives before choosing among them by results.

Research-memo template, page 3. A list of fixed effects and standard-error options is not an explanation of identifying variation.

Workshop appendix: memo page 4 defines scope and feasibility

- **Source and target:** whose effect is learned, and whose decision is informed?
- **Modifiers and support:** pretreatment worker/organization attributes, absent contexts, and changed treatment versions.
- **Incidence:** firm, initially eligible workers, entrants, and external parties; keep leavers nested correctly.
- **Accounting:** resources versus transfers, valuations, horizon, and unobserved welfare.
- **Next month:** one bounded evidence product, not a promise of a completed paper.
- **Fallback:** what public documentation, an available archive, or synthetic testing allows; which substantive claim must wait.

Research-memo template, page 4. A narrower target or a first measurement study can be the best feasible contribution.

Workshop appendix A maps each required object to access

One row for each object	Complete these fields for every row
Eligibility / openings	Owner/source; required fields and linkage keys
Assignment or rollout	Timing, baseline/follow-up coverage, version histories
Production and audit	Sensitive content and permitted processing environment
Pay and conditions	Status: available and checked / permission pending / not observed
Exits and follow-up	Expected missingness, including missing destinations
Target context	A fallback and the claim that cannot yet be made

- The second column is a common field checklist, not a one-to-one restriction on each row.

Full template: `research-memo-template.md` in the Week 10 lab. Filling out the form does not authorize provider contact, access applications, worker recruitment, or data transfer.

Workshop appendix B validates records and the reproducible route

Validation area	Prespecify at least one concrete check
Identity and time	Joins/IDs; duplicate records; time ordering; overlapping spells
Population	Initial risk sets; exits; survey response; missing destinations
Measurement	Recording changes; independent labels/audits; denominator definitions
Permissions	Allowed uses, storage, linkage, disclosure review, retention
Reproducibility	Versioned inputs; source-to-output scripts; restricted-file documentation
Mode	Exact, reduced, published-statistic reconstruction, or synthetic

Each mode needs its own label. Preserve transformation code and permitted outputs. A synthetic companion tests workflow and logic; it does not recreate unavailable empirical evidence.

Workshop appendix C makes the transport assumption concrete

Compare source with target on	For each dimension, state
Workers and skills	Source and target definitions
Tasks and review capacity	Why this feature could modify effects
Pay and authority	Whether it is measured before treatment
Outside options	Whether both populations have relevant support
Implementation and version	What compatibility or stability must be assumed
Scale and peer exposure	The consequence for the supported claim
Horizon and outcome definition	The most useful missing observation

The right-hand checklist applies to every dimension. Do not infer effect stability from similar controls or observed balance.
An absent treatment version requires new evidence or a narrower claim.

Workshop review rewards evidence and a constructive next step

Review dimension	Evidence to look for
Contribution and estimand	Decision, two mechanisms, eligibility, regimes, horizon, weights
Identification and measurement	Actual variation, remaining threat, independent validation
Inference and portability	Independent units, test null, target, modifiers, support, version
Welfare and feasibility	Consistent incidence, minimal schema, access status, fallback
Reproducibility and care	Permissions, privacy, versions, documented restricted inputs

- Mark each dimension: **clear**, **needs clarification**, or **not yet supported**. Do not total a numeric score.
- Respond with one revision, one remaining limit, and the smallest useful new observation.

Full proposal-review-rubric.md: nine dimensions, grouped here. Judge the evidence rather than the partner's prestige, method, data size, or statistical significance.