Experimental Analysis of Public Policies

Session 3

Randomised Experiments: Design and Power

Etienne Dagorn - Universite de Lille

Session Objectives

By the end of this session, students should be able to:

- → explain why random assignment eliminates selection bias in expectation;
- → choose between individual, stratified, and cluster randomisation;
- → read a balance table without confusing diagnostics and proof;
- → identify spillovers, attrition, non-compliance, and measurement threats;
- → compute and interpret an MDE in words;
- → explain why clustered designs have lower effective sample size;
- → write a short experimental protocol aligned with the policy decision.

Roadmap

Why Randomise?

Choosing the Assignment Design

Implementation Threats

Power and MDE

Protocol and Bridge

Roadmap

Why Randomise?

Choosing the Assignment Design

Implementation Threats

Power and MDE

Protocol and Bridge

From Selection to Randomisation

→ Session 2 asked whether observed covariates can justify comparability.

→ RCTs create comparability through the assignment mechanism.

→ The main task shifts from controlling selection to protecting implementation.

Central question

How do we design an experiment whose comparison group is credible and informative for policy?

Example: Education Trial

- → Eligible schools apply for a tutoring programme.
- → Randomise schools within districts to reduce spillovers.
- → Primary outcome: standardised maths score after one year.
- → Main threats: school attrition, heterogeneous implementation, political pressure.

Class discussion: if politicians insist that the weakest schools must be treated first, the design moves from RCT towards phased rollout or DiD.

Why Randomise?

- → Random assignment breaks the link between treatment status and potential outcomes.
- → It makes selection bias zero in expectation.
- → It gives a transparent basis for inference.
- → It does not automatically solve spillovers, attrition, non-compliance, or external validity.

In plain words: randomisation gives us comparable groups before treatment; the rest of the lecture is about keeping them comparable after implementation begins.

The Design Question

Before randomising

Define the policy lever, eligible population, randomisation unit, treatment arm, outcome, and timing.

→ Randomising a vague intervention creates a vague estimand.

→ Randomising access and randomising actual receipt are different.

→ A good design anticipates implementation frictions.

Example: "being offered tutoring" can be randomised; "learning from tutoring" also depends on attendance, teacher quality, and effort.

Assignment Mechanism

$$Z_i \perp \{Y_i(1), Y_i(0)\}$$

- Z_i is the randomised assignment, offer, or access rule.
- D_i is treatment received; under full compliance only, $D_i = Z_i$.
- Randomisation can be simple, stratified, clustered, or phased in.
- The assignment mechanism should be documented clearly.

Design audit: who generated the random numbers, when, with which eligible list, and who could alter assignment?

Estimand in a Simple RCT

$$ATE = E[Y_i(1) - Y_i(0)]$$

$$\widehat{\tau} = \bar{Y}_{Z=1} - \bar{Y}_{Z=0}$$

→ The sample contrast compares assigned arms, not self-selected recipients.

→ With full compliance, $D_i = Z_i$, so it estimates the treatment effect for the experimental sample.

→ With imperfect compliance, the assigned-arm contrast estimates the ITT: the effect of assignment or offer.

Roadmap

Why Randomise?

Choosing the Assignment Design

Implementation Threats

Power and MDE

Protocol and Bridge

RCT Design Workshop: Protocol

Policy: schools can receive a new maths-tutoring package, but resources cover only half of eligible schools.

Group task

Draft a five-line protocol: eligibility rule, randomisation unit, treatment arm, primary outcome, and follow-up date.

Deliverable: a protocol is good if another team could implement the same assignment rule without asking you what you meant.

Design Choices: First Tradeoffs

Choice	Tradeoff to discuss
School randomisation	fewer clusters, fewer spillovers
Pupil randomisation	more units, more contamination risk
Stratification	more balance, more logistical complexity

Simple Random Assignment

- → Each eligible unit has the same probability of assignment.
- → Easy to explain and implement.
- → Can produce chance imbalances in small samples.
- → Works best when the sample is large and treatment is individual-level.

Teaching example: toss a coin for each pupil; randomisation is transparent, but school-level logistics may become messy.

Stratification and Blocking

- → Randomise within groups defined by important baseline variables.
- → Common strata: school, region, gender, baseline outcome level.
- → Improves balance on variables chosen ex ante.
- → Can improve precision when strata predict outcomes.
- → Analyse consistently with the design, for example with stratum indicators.

Example

Randomise pupils within schools so treated and control pupils face the same school environment.

Implementation note: define strata before randomisation and report them in the protocol.

Cluster Randomisation

- → Assign treatment to groups: schools, villages, firms, regions.
- → Useful when the intervention is delivered at group level.
- → Reduces contamination between treated and untreated units.
- → Requires more clusters for the same statistical power.

Intuition: 30 pupils in one treated classroom provide less independent information than 30 treated classrooms.

Choosing the Randomisation Unit

Unit	When useful
Individual	treatment is private and spillovers are small
Classroom	teacher or peer interactions matter
School	implementation is school-wide
Region	policy is administered territorially

Ethics and Feasibility

- → Randomisation must be acceptable to participants and institutions.
- → Scarce resources can make lotteries defensible.
- → Control groups may receive delayed treatment.
- → Ethical constraints shape the estimand and design.

Discussion cue: ethical feasibility is not separate from econometrics; it determines which counterfactual can actually be observed.

Roadmap

Why Randomise?

Choosing the Assignment Design

Implementation Threats

Power and MDE

Protocol and Bridge

Balance Tables

- → Show whether treatment and control groups look similar at baseline.
- → Use pre-treatment variables only.
- → Focus on substantively important differences, not just p-values.
- → Large imbalance can occur by chance but still matter for precision and credibility.

Do Not Overread Balance Tests

Subtle point

Randomisation validity comes from the assignment process, not from obtaining insignificant balance tests.

- → In small samples, important differences may be statistically insignificant.
- → In large samples, tiny differences may be significant.
- → Balance tables are diagnostics and communication tools.

Spillovers

- → Treatment of one unit may affect another unit's outcome.
- → Tutored pupils may help untreated classmates.
- → Job-search assistance may change competition for untreated workers.
- → Spillovers can make the control group partly treated.

Design response

Change the randomisation unit, measure networks, or define spillover estimands explicitly.

Key distinction: spillovers are not always a nuisance; sometimes they are part of the policy effect policymakers care about.

Attrition

- → Attrition occurs when outcome data are missing for some units.
- → It threatens validity if missingness differs by treatment and potential outcome.
- → Report attrition rates by treatment arm.
- → Use tracking protocols and sensitivity checks.

Example: if low-performing treated pupils are more likely to miss the final test, the observed treatment mean may be too optimistic.

Non-compliance Preview

- → Assignment may differ from actual treatment received.
- → Some assigned to treatment do not participate.
- → Some assigned to control find treatment elsewhere.
- → Session 4 studies ITT, first stage, IV, and LATE.

Preview question: do we want the effect of being offered the programme, or the effect for pupils whose participation was changed by the offer?

Outcome Measurement

- → Outcomes should map onto the policy objective.
- → Measurement must be comparable across groups.
- → Multiple outcomes create interpretation and multiple-testing issues.
- → Baseline outcomes can improve precision and diagnose comparability.

Design habit: name one primary outcome before looking at results, then classify secondary outcomes as mechanisms or complements.

Implementation Threats

Threat	Design response
Contamination	cluster randomisation or tracking
Attrition	follow-up protocol and bounds
Non-compliance	report ITT and first stage
Measurement changes	common instruments and timing

Roadmap

Why Randomise?

Choosing the Assignment Design

Implementation Threats

Power and MDE

Protocol and Bridge

Power: Intuition

Question

If the policy has a meaningful effect, is the study likely to detect it?

→ Power depends on sample size, outcome variance, treatment share, clustering, and effect size.

→ Low power can turn a useful experiment into an inconclusive one.

→ Precision should be considered before data collection.

Power: Numerical Interpretation

Suppose:

- → school-level outcome standard deviation is 20 test-score points;
- → 40 schools are randomised, 20 assigned $Z = 1$ and 20 assigned $Z = 0$;
- → two-sided 5% test and 80% power.

$$SE(\widehat{\tau}) = \sqrt{\frac{20^2}{20} + \frac{20^2}{20}} \approx 6.3, \qquad MDE \approx 2.8 \times 6.3 \approx 18.$$

Interpretation

This 40-school design cannot reliably detect a 6-point effect under these assumptions.

Minimum Detectable Effect

$$MDE \approx \left(z_{1-\alpha/2} + z_{1-\kappa}\right) SE(\widehat{\tau})$$

- → $1 - \kappa$ is desired power; κ is the Type II error rate.
- → MDE is the smallest effect the design can detect with that chosen power.
- → It should be compared to a policy-relevant effect size.
- → $SE(\widehat{\tau})$ must reflect clustering, blocking, treatment shares, and planned covariate adjustment.

Reading reflex: an underpowered null result should be read as "inconclusive", not necessarily "no effect".

What Drives Precision?

- → More observations usually reduce standard errors.
- → Balanced treatment shares improve precision.
- → Baseline covariates can reduce residual variance.
- → Blocking can improve precision when the analysis reflects the blocked design.
- → Clustered assignment reduces effective sample size.

Practical reading: adding 200 pupils in the same few schools may help much less than adding more schools.

Sample Size Tradeoffs

- → More units can improve precision but cost more.
- → More clusters can matter more than more individuals per cluster.
- → Measuring baseline outcomes may be cheaper than expanding the sample.
- → The design should be feasible for the implementing institution.

Policy connection: the best sample size is not the largest possible one; it is the one that can answer the decision question with acceptable uncertainty.

Cluster Design Effect

$$DE = 1 + (m - 1)\rho$$

- m: average cluster size.
- ρ: intra-cluster correlation.
- When outcomes are highly correlated within clusters, effective sample size falls.
- This is why clustered RCTs need many clusters.

Numerical feel: if $m = 30$ and $\rho = 0.10$, the design effect is $1 + 29 \times .10 = 3.9$, so precision is much lower than individual counts suggest.

Roadmap

Why Randomise?

Choosing the Assignment Design

Implementation Threats

Power and MDE

Protocol and Bridge

Design Protocol

- → Record eligibility, assignment, treatment arms, outcomes, and timing.
- → Specify the primary outcome and main estimating equation.
- → Record planned subgroup or robustness analyses.
- → Keep the protocol short enough to be usable during implementation.

Code Mini-Lab

- → Simulate random assignment.
- → Compare balance under simple and stratified randomisation.
- → Compute a difference in means.
- → Change sample size and observe the standard error.

Student deliverable: one balance table, one treatment estimate, and one sentence explaining how sample size changes precision.

Exercise Bridge

- → Theory sheet: `exercises/theory/session_03_theory.md`.
- → QCM: `assessments/qcm/session_03_qcm.md`.
- → Applied exercise source: RCT design, power, and compliance scenario.
- → Exam skill: diagnose whether an RCT estimate answers the policy question.

Continuity: the exercise begins with design choices and ends with interpretation, mirroring the lecture rhythm.

Design Checklist

1. What is randomised?
2. At what level?
3. What is the estimand?
4. Is $D_i = Z_i$, or is receipt different from assignment?
5. What outcomes are measured and when?
6. What threatens implementation?
7. Is the study powered for a meaningful effect?

Key Takeaways

- → Randomisation creates comparability through the assignment process.
- → The unit of randomisation should match treatment delivery and spillover risks.
- → Balance tables support transparency but do not prove randomisation.
- → Power is about detecting meaningful effects, not just significant ones.
- → Implementation problems motivate compliance analysis in Session 4.

One-sentence memory aid: a randomised design is credible when assignment, implementation, measurement, and inference all point to the same estimand.

Transition to Session 4

Next question

What if people do not comply with their assigned treatment?

- → A clean lottery first identifies the effect of assignment or offer.
- → Treatment received D_i may differ from assignment Z_i.
- → Session 4 separates ITT, take-up, and the local treatment effect for compliers.

Bridge: when assignment is credible but receipt is endogenous, the source of identifying variation is the part of receipt moved by Z_i.

Appendix: Estimating Equation

$$Y_i = \alpha + \tau Z_i + \varepsilon_i$$

$$Y_i = \alpha + \tau Z_i + \gamma X_i + \lambda_s + \varepsilon_i$$

- → These equations estimate assigned-arm contrasts.
- → Pre-treatment covariates and stratum indicators can improve precision.
- → Randomisation, not the equation alone, justifies causality.
- → Under full compliance, $Z_i = D_i$; otherwise the coefficient is an ITT.

References and Further Reading

- → Duflo, Glennerster, and Kremer (2007), "Using Randomization in Development Economics Research: A Toolkit".
- → Gerber and Green (2012), *Field Experiments: Design, Analysis, and Interpretation*.
- → Glennerster and Takavarasha (2013), *Running Randomized Evaluations: A Practical Guide*.
- → Muralidharan (2017), "Field Experiments in Education in Developing Countries".

Reading reflex

When reading an RCT, identify the assignment unit, treatment delivered, first-stage take-up, attrition rate, and primary outcome before reading the coefficient table.