Impact Evaluations - Practice Sessions

Summer School 2025 at the University of Ghana, Accra

Claire Stein
`c.l.stein@rug.nl`

Faculty of Economics and Business
University of Groningen

July 26-31, 2025

Overview

1. Pre-Analysis Plans (PAP)
2. Randomization
3. Balance tests
4. Analyses techniques – ITT, ANCOVA
5. Non-compliance – LATE and Lee Bounds
6. Presenting PAPs
7. Power calculations
8. Presenting impact analyses

Brief introduction

A few facts about me:

- PhD candidate at the University of Groningen working on a RCT project in Vietnam
- Worked with different institutions – e.g., HelpAge Myanmar, Population Services International (PSI), French Research Institute for Development (IRD), AgriCord
- On different topics – e.g., women's empowerment, public health, rural livelihoods
- Mostly focused on South East Asia, brief experience in Uganda relatively new to the African context! :)

What about you?

- Name
- Background (research interest)
- Experience/knowledge with RCTs

Overall objectives and practice material

Objective: After learning about different components of impact evaluations–specifically randomized controlled trials (RCTs)–we will practice key steps, from pre-analysis plans (PAPs) to randomization and impact estimation, using a simulated dataset. The session will emphasize real-world applications and data analysis in Stata.

Material you'll need for the class:

- Datasets and do-files can be downloaded from GitHub:
- `https://github.com/ClaireStn/Impact_evaluation_practice`
- Please download the relevant dataset and do-file for each session, and we'll practice together, step-by-step!

Pre-Analysis Plans (PAP)

Claire Stein, FEB, University of Groningen

Section layout

Saturday 26 July, 11:30 - 13:00 (1.5 hours)

Objective: We will see an example of a Pre-Analysis Plan (PAP) and we will then think of our own PAP.

- Refresher - What goes in a PAP?
- PAP example - Psychosocial intervention in Vietnam
- Thinking of our own PAP (groups of 5)

Refresher: PAP checklist (1/2)

Based on McKenzie's PAP checklist:

1. **Sample Description**: Describe sampling strategy, expected sample size, randomization procedure, balance and attrition tests.
2. **Key Data Sources**: Identify surveys, administrative records, or monitoring tools.
3. **Hypotheses and Outcomes**: Specify primary and secondary outcomes, causal chain steps (ToC), subgroups. Link outcomes to precise measures (e.g., survey questions).
4. **Variable Construction**: Define transformations, coding, outlier handling, and missing data strategies.
5. **Treatment Effect Estimation**: Specify model (ANCOVA, DID, etc.), controls, standard errors.

Refresher: PAP checklist (2/2)

6. **Multiple Hypothesis Testing**: Plan to address multiplicity (e.g., indices, sharpened q-values).
7. **Survey Attrition**: State how you'll test and correct for differential attrition.
8. **Low-Variation Outcomes**: How to deal with low var outcomes (e.g., exclude variables with 95% obs with same values for C group ex-post).
9. **Model Specification (Optional)**: Include theory/model if testing specific behavioral or economic mechanisms.
10. **Archiving**: Register PAP with time-stamped registries (e.g., AEA, EGAP, JPAL).

PAP example - Intervention in Vietnam (1/3)

Link to our PAP on the AEA RCT Registry

1. **Introduction**: In collaboration with a MicroFinance Institution (MFI) in Vietnam, we develop and evaluate a *psychosocial training* (material) for female members, framed independently (T1) or interdependently (T2).
   - **Research question**: What is the impact of the intervention on psychological agency and economic outcomes?
   - **ToC**: (*Input*) Psychosocial training developed → (*Output*) 610 women trained → (*Outcomes*) SMART goals and enhanced agency → *(*Impact*) Economic well-being (increase in income, savings, etc.)

TRAINING OVERVIEW

This training is specially designed for the women members of TYM with the aim to helping them enhance their skill set to achieve their life goal.

This training includes 3 main topics with 4 sessions:

- Topic 1: Goal setting
- Topic 2: Confidence building
- Topic 3: Communication
- Recap session

Session 1: Goal setting

In the first session you will take a moment to reflect on your life goals. You will reflect on how smaller saving goals can help you in achieving your bigger life goals. Lastly, you will learn how to make your saving goals **SMART**.

Session 2: Confidence

In this session, you will realize how confidence can help you achieve your goals. You will learn the **3C model** for confidence, in which you will identify your own strengths and think of how they can help you in overcoming challenges.

Session 3: Communication

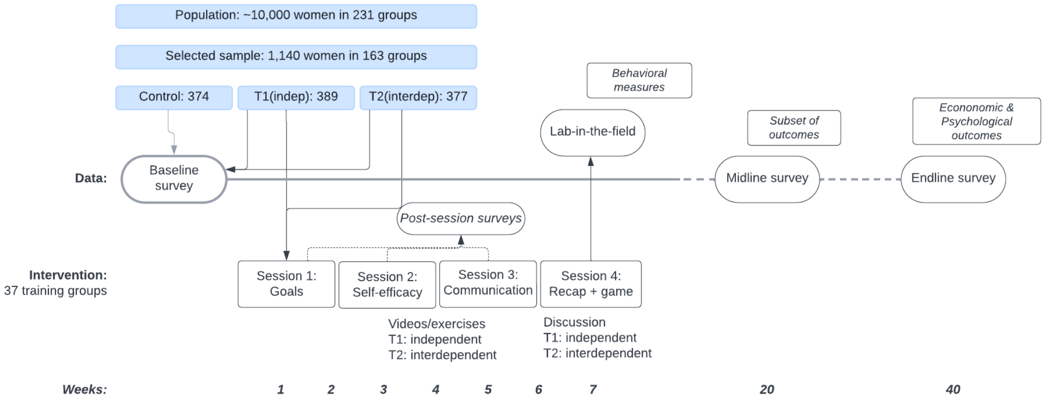
You will learn the importance of communicating your goals to other around you. You will learn about the 3W1H model in which you will learn effective communication skills that can help you in communicating your goals well.

Session 4: Recap

This is the occasion to reflect on what you have learned and discuss how you can further develop and apply your skills after this training.

PAP example – Intervention in Vietnam (2/3)

2. **Study design**: Women member of the MFI randomized at the MFI group level
    - **Sample**: Women below 52, member of the MFI (N = 1140)
    - **Intervention**: A psychosocial training aimed at helping women in progressing towards their goals
    - **Randomization**: MFI group level
        - T1: Independent-framed training (55 MFI groups N = 389)
        - T2: Interdependent-framed training (55 MFI groups N = 377)
        - C: Control group, no training (53 MFI groups N = 374)
3. **Data strategy**:
    - **Data collection**: We collected baseline (pre-treatment) and endline data in-person, online post-session surveys, and midline data by phone
    - **Variables**: We are interested in *intermediate outcomes* (goals and agency outcomes), and *primary outcomes* (economic well-being)


Population: ~10,000 women in 231 groups
Selected sample: 1,140 women in 163 groups
Control: 374
T1(indep): 389
T2(interdep): 377
Behavioral measures
Lab-in-the-field
Subset of outcomes
Econonomic & Psychological outcomes
Data:
Baseline survey
Midline survey
Endline survey
Post-session surveys
Intervention:
37 training groups
Session 1: Goals
Session 2: Self-efficacy
Session 3: Communication
Session 4: Recap + game
Videos/exercises
T1: independent
T2: interdependent
Discussion
T1: independent
T2: interdependent
Weeks:
1
2
3
4
5
6
7
20
40

PAP example - Intervention in Vietnam (3/3)

4. **Estimation strategy**: Women member of the microfinance randomized at the MFI group level
   - **Estimations**: We estimate Intention to Treat (ITT) using baseline outcomes when available (ANCOVA).
   - **MHT**: We create indices for different sets of outcomes, and correct for Multiple Hypothesis Testing (MHT).
   - **Attrition**: We use Leebounds if and when we observe differential attrition by treatment.
5. **Ethical issues**: We obtained ethical clearance from the University's IRB. We ensure that no sensitive issues arise, and engage in co-creation process with the local partners,

5.1.3. Main econometric specifications

To study impact of the "progress with your goals" curriculum on savings, we will estimate the following ANCOVA specifications (McKenzie, 2012):

$$y_{ijst} = \beta_0 + \beta_1 Training_{js} + \beta_2 y_{ijs0} + \theta_s + \varepsilon_{ijst} \tag{1}$$

where y_{ijst} is the (primary, secondary or intermediate) outcome variable from Sections 4.1.1-4.1.3 for individual i in group j from randomization stratum s at midline or endline, $Training_{js}$ is a dummy that equals 1 for either of the treatments and 0 for control, y_{ijs0} is the baseline value of the outcome variable and is included whenever available, θ_s represents stratum fixed effects, and standard errors will be clustered at the group level.

We will test $H_0: \beta_1 = 0$ vs. $H_a: \beta_1 \neq 0$.

To study how the impact of the "progress with your goals" curriculum depends on cultural framing, we limit the sample to those that were directly exposed to framing, i.e. those that are assigned to the training and participated in at least the second or third session.[24] We estimate the following ANCOVA specifications:

$$y_{ijst} = \beta_0 + \beta_1 Independent_{js} + \beta_2 y_{ijs0} + \theta_s + \varepsilon_{ijst} \tag{2}$$

Making your own PAP – Assignment (1/2)

Now you will make your own PAP in groups of 5. Below is an outline for your PAP, and in the next slides, there are a few examples from which you can choose from.

Outline for your PAP:

1. **General information**: Think of your research question, hypotheses, and general outcomes (#3 in the checklist)
   - Main research question
   - Simple Theory of Change (ToC)
2. **Study design**: Think of your sample, the intervention you are setting up, and how you will randomize (#1 in the checklist)
   - Sample selection
   - Brief intervention description
   - Randomization strategy and variables for balance

Making your own PAP – Assignment (2/2)

3. **Data**: Think of your data collection plan (when, how, and what data will you collect)
   - Data collection plan
   - Key data source(s) (#2 of the checklist)
   - Variables (#4 of the checklist)
4. **Estimation strategy**: Think of what estimation strategy you will use (#5 in the checklist), and of how you will address potential attrition (#7 in the checklist).
5. **Ethics**: Briefly think of potential ethical issues and how you could address these.

As we have not covered everything, you will start working on parts **1 and 2**. You can keep working on the rest the next days. We will present PAPs Monday after lunch.

Making your own PAP – Examples

Vocational Training in Nigeria

Does short-term training improves youth employment?

Cash Transfer and Child Nutrition in Malawi

Do unconditional cash transfers improve child nutrition and food security?

Community health workers in Uganda

Do monthly visits from community health workers reduce child mortality?

Randomization

Claire Stein, FEB, University of Groningen

Session layout

Monday 28 July, 11:45–12:30 (45 min)

Objective: We will practice stratified randomization with a simulated dataset based on a specific case study.

- Setting the stage for our case study
- Setting up in Stata
- Doing a stratified randomization

Files for the session:

- `GhanaVocationalFinance_schoollist.dta`
- `Randomization.do`

Practical case

We will build on this practical case study throughout the sessions. It may be helpful to keep these details in mind or go back to it while doing the exercises.

Context: The World Bank wants to implement and evaluate a financial literacy and entrepreneurship training for youth in Ghana. They will work in **40 vocational schools**, and plan to deliver the training to up to **50 eligible students per school**.

We will be using a simulated dataset for this section and the followings.

What will you do?

Let's walk through how we can simulate a randomized controlled trial with one treatment and one control group using **stratified randomization**.

- **Population**: 2,000 youth (50 students × 40 vocational schools), aged 17–24
- **Randomization level**: School level
- **Stratified by**:
  - % Female students in the school (above/below median)
  - School size (above/below median)

Practice

You now receive a dataset and are asked to randomize schools to T or C.

Steps in a stratified random assignment

- Divide eligible units (here it's the 40 schools) into strata based on observables (% of females, school size)
- Do simple random assignment for each stratum by:
  - order sublist randomly
  - allocate units to randomization cells from this sublist
- Randomly pick which unit is treatment and which is comparison
- Why are we doing this?
  - To ensure balance.
  - To increase power.
  - To facilitate sub-group analysis.

Setting up in Stata

```
clear
set more off
cd "/Users/clairestein/Documents/PhD_material/Impact_evaluation_summerschool"

cap log close
set logtype text
log using RandomizationPractice.txt, replace
```

- Prepare a clean Stata session
- Prevent annoying screen pauses
- Set your working directory
  - For Windows users you will need to add C: in front of your file directory
- Start a log file to record your session output

Guided practice: Randomization (1/3)

Step 1. Load the school list dataset

- `GhanaVocationalFinance_schoollist.dta`
- 40 rows (1 per school)
- Variables:
  - Proportion of female students (`prop_female`)
  - School size (`school_size`)

Step 2a. Explore the data

- `describe`, `summarize`, `tab schoolid`
- Understand the number of schools and key characteristics

```
. /*** STEP 1: Load school list (one row per school) ***/
. /*
> We start with a dataset of 40 vocational schools
> Variables: % female (prop_female), school size (school_size)
> */
.
. use Dataset/GhanaVocationalFinance_schoollist.dta, clear

.
. /*** STEP 2a: Explore the dataset ***/
. /*
> Quick look at the dataset:
>  - Variable names and labels
>  - Summary stats for key variables
>  - How many schools
> */
.
. describe

Contains data from Dataset/GhanaVocationalFinance_schoollist.dta
 Observations:            40
    Variables:             3                  25 Jun 2025 09:48
--------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
--------------------------------------------------------------------------------
schoolid        float   %9.0g                 School ID
prop_female     float   %9.0g                 (mean) prop_female
school_size     float   %9.0g                 (mean) school_size
--------------------------------------------------------------------------------
Sorted by: schoolid

. summarize

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
    schoolid |         40        20.5    11.69045          1         40
 prop_female |         40    .5245357    .1138585   .3132428   .6959848
 school_size |         40      82.125    14.56935         52        113
```

Guided practice: Randomization (2/3)

Step 2b. Construct strata

- Median split:
  - High vs Low % female students
  - Large vs Small school size
- 4 strata in total (2x2)
- Command: `egen strata = group(high_female large_school)`

```
. /*** STEP 2b: Construct strata: % female x school size ***/
. /*
> Four strata: High female + large school, etc.
> Why? To improve balance across treatment/control on key school characteristics
> */
.
. sum prop_female, detail

                         (mean) prop_female
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .3132428       .3132428
 5%     .3562749       .3502552
10%     .3769126       .3622946       Obs                  40
25%     .4184239       .3758824       Sum of wgt.          40

50%     .5483128                      Mean           .5245357
                        Largest       Std. dev.      .1138585
75%     .6356594       .6710637
90%     .6697772       .6766917       Variance       .0129637
95%     .6805419        .684392       Skewness      -.1449739
99%     .6959848       .6959848       Kurtosis       1.661819

. gen high_female = prop_female > r(p50)

.
. sum school_size, detail

                         (mean) school_size
-------------------------------------------------------------
      Percentiles      Smallest
 1%           52             52
 5%         55.5             54
10%         63.5             57       Obs                  40
25%           71             62       Sum of wgt.          40

50%         82.5                      Mean             82.125
                        Largest       Std. dev.      14.56935
75%         92.5            104
90%        101.5            105       Variance        212.266
95%          105            105       Skewness      -.051367
99%          113            113       Kurtosis       2.466424

. gen large_school = school_size > r(p50)
```

```
.
. tab high_female large_school    // 2x2 cross-tab

high_femal |     large_school
         e |         0          1 |     Total
-----------+----------------------+----------
         0 |        13          7 |        20
         1 |         7         13 |        20
-----------+----------------------+----------
     Total |        20         20 |        40

.
. egen strata = group(high_female large_school), label

. label variable strata "Strata number: % female x school size"

.
. tab strata

     Strata |
  number: % |
   female x |
school size |      Freq.     Percent        Cum.
------------+-----------------------------------
        0 0 |         13       32.50       32.50
        0 1 |          7       17.50       50.00
        1 0 |          7       17.50       67.50
        1 1 |         13       32.50      100.00
------------+-----------------------------------
      Total |         40      100.00

.
```

Guided practice: Randomization (3/3)

Step 3. Randomize schools

- `set seed 314159` → To make the randomization *replicable*
- `gen random = runiform()` → Draw random number for each school
- `bysort strata (random): gen strata_index = _n` → Rank schools randomly within each stratum
- `bysort strata: gen strata_size = _N` → Count the number of schools in each stratum
- `gen treat = strata_index <= strata_size/2` → Assign *50% of schools to Treatment* in each stratum
- `gen treatment = "Control"; replace treatment = "Treatment" if treat == 1` → Create Treatment variable

```
. /*** STEP 3: Stratified randomization ***/
. /*
> Randomization unit = school
> 50% of schools assigned to Treatment within each stratum
> Ensures T/C balance across: % female and school size
> */
.
. set seed 314159                                  // Set seed for reproducibility

. gen random = runiform()                          // Random number per school
.
. bysort strata (random): gen strata_index = _n    // Rank within stratum

. bysort strata: gen strata_size = _N              // Stratum size

. gen treat = strata_index <= strata_size/2        // 50% Treatment
.
. gen treatment = "Control"

. replace treatment = "Treatment" if treat == 1
variable treatment was str7 now str9
(18 real changes made)

. label variable treatment "Treatment group"

.
. tab treatment strata                             // Show balance across strata

 Treatment |    Strata number: % female x school size
     group |       0 0        0 1        1 0        1 1 |     Total
-----------+--------------------------------------------+----------
   Control |         7          4          4          7 |        22
 Treatment |         6          3          3          6 |        18
-----------+--------------------------------------------+----------
     Total |        13          7          7         13 |        40

.
```

Congratulations - You completed the randomization!

Step 4. Save randomization file

- One row per school with: `schoolid`, `treatment`, `strata`, etc.

	schoolid	prop_female	school_size	high_female	large_school	strata	random	strata_index	strata_size	treat	treatment
1	7	.4817641	81	0	0	0 0	.1478543	1	13	1	Treatment
2	11	.4288765	78	0	0	0 0	.1692771	2	13	1	Treatment
3	12	.4085802	71	0	0	0 0	.2227877	3	13	1	Treatment
4	19	.4768413	65	0	0	0 0	.4900562	4	13	1	Treatment
5	24	.4900551	76	0	0	0 0	.4963467	5	13	1	Treatment
6	40	.4759239	72	0	0	0 0	.6331453	6	13	1	Treatment
7	32	.422444	76	0	0	0 0	.684197	7	13	0	Control
8	15	.3758824	70	0	0	0 0	.6942776	8	13	0	Control
9	5	.494808	71	0	0	0 0	.7572148	9	13	0	Control
10	31	.4328734	52	0	0	0 0	.8035679	10	13	0	Control
11	36	.3622946	70	0	0	0 0	.8392463	11	13	0	Control
12	8	.3779428	79	0	0	0 0	.8686824	12	13	0	Control
13	14	.3851774	77	0	0	0 0	.9812849	13	13	0	Control
14	3	.3132428	97	0	1	0 1	.25898	1	7	1	Treatment
15	33	.3502552	91	0	1	0 1	.3022594	2	7	1	Treatment

Balance tests

Claire Stein, FEB, University of Groningen

Section layout

Monday 28 July, 14:00-14:30 (30 min)

Objective: We will have a quick refresher on balance tests and practice in Stata.

- Basic idea and example)
- Practicing in Stata two methods:
  - Balance tests at the school level
  - Balance tests at the individual level

Files for the session:

- `GhanaVocationalFinance_randomized_baseline.dta`
- `Balance.do`

Refresher: Why do balance tests?

- After randomization, we want to check:
  - Are Treatment and Control groups **similar at baseline**?
  - Was randomization successful?
- We do this by testing if baseline covariates are balanced
- We **do NOT expect many significant differences – 5% rule of thumb**
- If there are large imbalances → investigate!

Reminder

Balance tests should be done at the **level of randomization**.
We will practice the 2 methods discussed earlier.

But first... a practical example

Dupas, P. (2011). Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya. American Economic Journal: Applied Economics, 3(1), 1-34.

- **Main idea**: RCT in Kenya to examine how different types of HIV information affect teenage sexual behavior. *Does providing relative risk info (e.g., older partners have higher HIV prevalence) change behavior more than abstinence-only edu?*
- **Experimental Design**:
  - Sample: 328 Kenyan primary schools
  - Two main arms: Abstinence-only curriculum (govt standard), and Relative risk information (RR): facts on age-specific HIV prevalence
- **Key Finding**: The RR information reduced teen pregnancy by 28%, suggesting that students substituted away from older, riskier partners.

TABLE 2—VERIFYING BALANCE BETWEEN GROUPS IN TERMS OF SCHOOL CHARACTERISTICS AND OUTCOMES FOR PRE-PROGRAM COHORT

	RR information			TT on HIV/AIDS curriculum		
	Comparison group (C) (1)	Treatment group (T) (2)	Difference T–C (3)	Comparison group (C) (4)	Treatment group (T) (5)	Difference T–C (6)
Panel A. School characteristics at baseline						
Class size	38.2	34.4	−3.8	37.4	37.3	−0.06
	[15.9]	[17.4]	(1.540)**	[16.9]	[15.7]	(1.281)
Pupils' sex ratio (girls/boys)	1.07	1.12	0.049	1.06	1.10	0.040
	[0.489]	[0.668]	(0.072)	[0.476]	[0.586]	(0.059)
Teacher-pupil ratio	0.026	0.026	0.000	0.025	0.027	0.003
	[0.026]	[0.022]	(0.003)	[0.021]	[0.028]	(0.003)
Teachers' sex ratio (females/males)	1.033	0.921	−0.112	1.003	1.014	0.011
	[0.914]	[0.777]	(0.119)	[0.92]	[0.852]	(0.099)
KCPE results (2003)	251.0	249.4	−1.6	252.2	249.0	−3.2
	[29.0]	[27.4]	(3.9)	[28.6]	[28.5]	(3.2)
Sampled for TT on HIV/AIDS curriculum	0.50	0.49	−0.003	0.00	1.00	
			(0.067)			
Sampled for RR information	0.00	1.00		0.22	0.22	−0.002
						(0.046)

Notes: These are school averages. Panel A: school characteristics collected through school visits in 2004. Panels B and C: students' outcomes collected in 2004 for the cohort of students enrolled in grade 8 in 2003, which is the cohort just one year older than the cohort involved in the RR experiment. Data collected by asking whereabouts of students at their 2003 primary school. Standard deviations in brackets. Columns 3 and 6: standard errors are in parenthesis. Five schools did not have an eighth grade class in 2003 and therefore are excluded from the table.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Questions

1. Can you **interpret the table?**

2. Which additional **individual baseline characteristics** do you think the authors should check for balance between treatment and control groups?

Panel B. Girls enrolled in eighth grade in 2003 (control cohort for RR): outcomes at the end of 2004						
Percent repeating class 8	0.246	0.209	−0.04 (0.021)*	0.236	0.238	0.002 (0.018)
Percent in secondary school	0.449	0.458	0.008 (0.026)	0.472	0.430	−0.043 (0.021)**
Percent in professional training	0.037	0.036	−0.001 (0.008)	0.037	0.036	−0.001 (0.007)
Percent out of school	0.259	0.289	0.029 (0.022)	0.246	0.286	0.040 (0.018)**
Percent married	0.077	0.083	0.006 (0.012)	0.071	0.085	0.014 (0.010)
Percent who had begun childbearing	0.144	0.139	−0.004 (0.018)	0.134	0.152	0.018 (0.015)
Observations	4,783	1,212	5,995	3,016	2,979	5,995
Panel C. Boys enrolled in eighth grade in 2003 (control cohort for RR): outcomes at the end of 2004						
Percent repeating class 8	0.226	0.220	−0.006 (0.022)	0.227	0.223	−0.004 (0.018)
Percent in secondary school	0.521	0.509	−0.012 (0.027)	0.528	0.508	−0.020 (0.023)
Percent in professional training	0.015	0.009	−0.006 (0.004)	0.017	0.010	−0.007 (0.004)*
Percent out of school	0.229	0.254	0.025 (0.023)	0.217	0.252	0.036 (0.019)*
Percent married	0.009	0.007	−0.002 (0.005)	0.010	0.007	−0.003 (0.004)
Observations	4,845	1,229	6,074	3,079	2,995	6,074
Number of schools	252	71	323	163	160	323

Notes: These are school averages. Panel A: school characteristics collected through school visits in 2004. Panels B and C: students' outcomes collected in 2004 for the cohort of students enrolled in grade 8 in 2003, which is the cohort just one year older than the cohort involved in the RR experiment. Data collected by asking whereabouts of students at their 2003 primary school. Standard deviations in brackets. Columns 3 and 6: standard errors are in parenthesis. Five schools did not have an eighth grade class in 2003 and therefore are excluded from the table.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Our case: Ghana Vocational Finance RCT

- Randomization level: **school level** (40 schools)
- Data: individual-level data on 2000 students
- Dataset: `GhanaVocationalFinance_randomized_baseline.dta`
- Variables: savings, financial literacy, confidence, budgeting, mobile money use, etc.

Practice

How do we conduct a balance test when the data is at **individual level**?

School and individual level

- **School level: Collapse to school-level**
  - Calculate average of covariates per school
  - Run t-tests or regressions at school level
- **Individual level: Regressions with clustered SEs**
  - Use individual data
  - Regress covariate on treatment
  - Use `vce(cluster schoolid)` to correct SEs

Reminder

SEs must account for randomization at school level!

Guided practice: Balance tests

Load and check the data first

- `use GhanaVocationalFinance_randomized_baseline.dta, clear`
- Check variables: `describe`, `summarize`

```
. describe

Contains data from Dataset/GhanaVocationalFinance_randomized_baseline.dta
 Observations:         2,000
    Variables:            18                  25 Jun 2025 16:10
------------------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
------------------------------------------------------------------------------------------
schoolid        float   %9.0g                 School ID
school_size     float   %9.0g                 School size (students)
prop_female     float   %9.0g                 Proportion female in school
studentid       float   %9.0g
female          float   %9.0g                 Female (1=yes)
age             float   %9.0g                 Age
household_inc~e float   %9.0g                 Household income (GH₵)
caregiver_edu   float   %9.0g                 Caregiver education level (0–5)
rural           float   %9.0g                 Rural residence
savings         float   %9.0g                 Amount saved last month (GH₵)
knows_budgeting float   %9.0g                 Knows how to budget
mobile_money_~e float   %9.0g                 Uses mobile money
confidence      float   %9.0g                 Financial confidence (1–5)
financial_lit~e float   %9.0g                 Financial literacy score (0–100)
id              float   %9.0g
strata          float   %9.0g      strata     Strata number: % female x school size
treatment_group str9    %9s                   Treatment group
treatment       float   %9.0g                 Treatment (1=Treatment, 0=Control)
------------------------------------------------------------------------------------------
Sorted by:
```

Guided practice: Balance tests

Let's first do the tests at the school level

- `collapse (mean) savings financial_literacy_score confidence ..., by(schoolid treatment strata)`
- `ttest savings, by(treatment)`
- `reg savings treatment`

```
. collapse (mean) savings financial_literacy_score confidence                 ///
>                                knows_budgeting mobile_money_use              ///
>                                household_income caregiver_edu age female rural ///
>                                , by(schoolid treatment strata)

.
. /* T-tests at school level */
. ttest savings                                   , by(treatment)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
       0 |      22    587.3528    8.278888    38.83143    570.1359    604.5697
       1 |      18    622.7118    13.96381    59.24342    593.2507    652.1728
---------+--------------------------------------------------------------------
Combined |      40    603.2643    8.153548    51.56756    586.7722    619.7564
---------+--------------------------------------------------------------------
    diff |           -35.35899    15.58122               -66.90153   -3.816451
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -2.2693
H0: diff = 0                                     Degrees of freedom =       38

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0145         Pr(|T| > |t|) = 0.0290          Pr(T > t) = 0.9855
```

```
. /* Optional: regressions at school level */
. reg savings                                     treatment

      Source |       SS           df       MS      Number of obs   =        40
-------------+----------------------------------   F(1, 38)        =      5.15
       Model |  12377.5549         1  12377.5549   Prob > F        =    0.0290
    Residual |  91331.7725        38  2403.4677    R-squared       =    0.1193
-------------+----------------------------------   Adj R-squared   =    0.0962
       Total |  103709.327        39  2659.21352   Root MSE        =    49.025

------------------------------------------------------------------------------
     savings | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   treatment |   35.35899   15.58122     2.27   0.029     3.816451    66.90153
       _cons |   587.3528    10.4522    56.19   0.000     566.1934    608.5122
------------------------------------------------------------------------------

. reg financial_literacy_score    treatment

      Source |       SS           df       MS      Number of obs   =        40
-------------+----------------------------------   F(1, 38)        =      0.05
       Model |  .102919297         1  .102919297   Prob > F        =    0.8214
    Residual |  75.6624674        38  1.99111756   R-squared       =    0.0014
-------------+----------------------------------   Adj R-squared   =   -0.0249
       Total |  75.7653867        39  1.94270222   Root MSE        =    1.4111

------------------------------------------------------------------------------
financial_~e | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   treatment |   .1019602   .4484674     0.23   0.821    -.8059145    1.009835
       _cons |   40.01514   .3008411   133.01   0.000     39.40612    40.62416
------------------------------------------------------------------------------
```

What if we see imbalances?

- Small differences can happen by chance – even after randomization
- Always ask:
  - Is this imbalance **large in magnitude**?
  - Is it **statistically significant**?
  - Could it affect interpretation of results?
- If yes → consider adjusting in analysis:
  - Include imbalanced covariates as controls
  - Use ANCOVA specification (control for baseline outcomes)

Question

If you see differences in baseline savings – what might this imply?

Guided practice: Balance tests (individual-level)

- Load data again: `GhanaVocationalFinance_randomized_baseline.dta`
- Regress baseline covariate on treatment, using clustered SEs:
  - `reg savings treatment i.strata, vce(cluster schoolid)`
  - `reg financial_literacy_score treatment i.strata, vce(cluster schoolid)`
- Compare results with Method 1

Why do we add `i.strata`?

Because randomization was stratified → Controlling for strata improves precision.

Why do we use `vce(cluster schoolid)`?

Because randomization was at school level, so SEs must be clustered at this level → students in same school are not independent!

```
.
. reg savings                      treatment       i.strata, vce(cluster schoolid)

Linear regression                               Number of obs     =      2,000
                                                F(4, 39)          =       4.14
                                                Prob > F          =     0.0068
                                                R-squared         =     0.0081
                                                Root MSE          =     298.74

                                (Std. err. adjusted for 40 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
     savings | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   treatment |   36.23698    14.3073     2.53   0.015      7.29773    65.17623
             |
      strata |
        0 1  |   20.75804   20.97694     0.99   0.328    -21.67183    63.18792
        1 0  |      60.33   19.75926     3.05   0.004     20.36311    100.2969
        1 1  |   23.14008   18.27146     1.27   0.213    -13.81743    60.09759
             |
       _cons |   565.2467   15.12816    37.36   0.000     534.6472    595.8463
------------------------------------------------------------------------------

. reg financial_literacy_score    treatment       i.strata, vce(cluster schoolid)

Linear regression                               Number of obs     =      2,000
                                                F(4, 39)          =       2.75
                                                Prob > F          =     0.0414
                                                R-squared         =     0.0040
                                                Root MSE          =     9.9315

                                (Std. err. adjusted for 40 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
financial_~e | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   treatment |   .0839877   .4015849     0.21   0.835    -.7282943    .8962698
             |
      strata |
        0 1  |   .0375688   .5250565     0.07   0.943    -1.024458    1.099596
        1 0  |   .1666723   .6617242     0.25   0.802    -1.171791    1.505136
        1 1  |   1.390428   .4671654     2.98   0.005     .4454967    2.335359
             |
       _cons |    39.5356   .3791107   104.29   0.000     38.76877    40.30242
------------------------------------------------------------------------------
```

Analyses techniques – ITT, ANCOVA

Claire Stein, FEB, University of Groningen

Section layout

Monday 28 July, 15:30–17:00 (1.5h)

Objective: We will practice (simple) analyses techniques, namely Intention to Treat (ITT), ANCOVA, and Difference-in-Difference.

- Refresher: Estimations
- Practicing in Stata
- Interpret results

Files for the session:

- `GhanaVocationalFinance_randomized_endline.dta`
- `ITT_ANCOVA_DD.do`

Refresher: Intention-to-Treat (ITT)

Basic logic: If randomization worked, then a simple comparison of treatment and control outcomes gives us unbiased treatment effect estimates.

$$Y_i = \alpha + \beta T_i + \varepsilon_i$$

Where:

- Y_i: Endline outcome (e.g., savings, financial literacy)
- T_i: Treatment dummy (1=treated, 0=control)

Add controls (X): To increase precision, or if some imbalance remains.

$$Y_i = \alpha + \beta T_i + \gamma X_{iB} + \varepsilon_i$$

Refresher: ANCOVA

ANCOVA = ”Analysis of Covariance” model

$$Y_{i1} = \alpha T_i + \beta X_{iB} + \rho Y_{iB} + \varepsilon_i$$

Where:

- Y_{iE}: Endline outcome
- Y_{iB}: Baseline value of the outcome
- X_{iB}: Baseline covariates

Why ANCOVA?

It increases precision by adjusting for baseline outcomes.

Refresher: Double-Differences (DD)

DD logic: Estimate the difference in change over time for treated vs. control

$$Y_{it} = \alpha X_{it} + \beta t_t + \theta T_i + \delta(T_i \times t_t) + \varepsilon_{it}$$

Where:

- t_t: Time dummy (1 = endline, 0 = baseline)
- T_i: Treatment dummy
- X_{it}: Baseline covariates
- δ: DD estimator (of interest)

Why DD?

Similar to FE with time dummy – useful if some imbalance is suspected.

Recap

Which method to use?

- If randomization was clean and baseline outcomes are available: **ANCOVA** often preferred for precision.
- If there's concern about imbalance or selection: **DD** is more robust.
- Always check balance first (and attrition).

Let's practice now!

We will first **practice** these estimation strategies in Stata using the same dataset and compare results - we will use a few variables from the dataset.
Then you will have **15 min to do the assignment in your groups of 5**.

Your assignment

After we practiced together, do you own estimations in groups of 5:

- Use variables we have not practiced with in the session
- Run some estimations (ITT, ANCOVA, DD)
- Save results (e.g., take a screenshot of the Stata output)
- Include it in a PPT, you will later present results (Wedn 30th)

But first... Let's understand the data

Before jumping into estimation, let's take a moment to understand the dataset:

- `GhanaVocationalFinance_randomized_endline.dta`
- 2,000 students (50 students $\times$ 40 schools)
- Treatment assigned at school level
- Variables include:
  - Treatment status (0/1)
  - School ID, strata
  - Endline outcomes: `end_savings`, `end_finlit_score`, `end_confidence`, etc.
  - Baseline outcomes: `savings`, `financial_literacy_score`, etc.

Practice

Open the dataset (`br`), explore the variables (`describe`, `summarize`). What outcomes do you expect to be most affected?

Guided practice: ITT and ANCOVA

Step-by-step in Stata:

1. Open: `GhanaVocationalFinance_randomized_endline.dta`
2. Pick outcome: e.g. `end_savings`
3. Run ITT:
   - `reg end_savings treatment i.stratum, vce(cluster schoolid)`
4. Add controls:
   - `reg end_savings treatment female age i.stratum, vce(cluster schoolid)`
5. Run ANCOVA:
   - `reg end_savings treatment savings i.stratum, vce(cluster schoolid)`

Interpretation questions

How big is the estimated effect? Is it statistically significant? Does it change with controls?

```
. reg end_savings treatment            i.strata, vce(cluster schoolid)

Linear regression                               Number of obs     =      1,834
                                                F(4, 39)          =      20.95
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0420
                                                Root MSE          =     314.09

                              (Std. err. adjusted for 40 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
 end_savings | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   treatment |    128.733   14.81445     8.69   0.000     98.76791     158.698
             |
      strata |
        0 1  |   15.95006     20.256     0.79   0.436    -25.02157     56.9217
        1 0  |   44.50038   20.17414     2.21   0.033     3.694336    85.30643
        1 1  |   16.51201   19.23444     0.86   0.396    -22.39331    55.41734
             |
       _cons |   631.2152   14.50168    43.53   0.000     601.8828    660.5477
------------------------------------------------------------------------------

.
. /*Add controls (age, gender)*/
. reg end_savings treatment age female i.strata, vce(cluster schoolid)

Linear regression                               Number of obs     =      1,834
                                                F(6, 39)          =      14.02
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0439
                                                Root MSE          =     313.96

                              (Std. err. adjusted for 40 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
 end_savings | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   treatment |   128.5213   14.86166     8.65   0.000     98.46071    158.5818
         age |  -6.430506   4.251858    -1.51   0.138     -15.0307    2.169688
      female |  -14.29067   13.87337    -1.03   0.309    -42.35221    13.77086
             |
      strata |
        0 1  |   16.04221   20.32117     0.79   0.435    -25.06123    57.14565
        1 0  |   47.33272   20.36087     2.32   0.025     6.148973    88.51646
        1 1  |     19.721   18.77372     1.05   0.300    -18.25243    57.69444
             |
       _cons |   766.5833   88.12459     8.70   0.000     588.3345    944.8321
------------------------------------------------------------------------------
```

```
.
. /*** STEP 3: ANCOVA Estimation ***/
. /*
> ANCOVA = Adjust for baseline outcome. Higher power than DD when baseline is predictive.
> */
. reg end_savings   treatment savings        i.strata, vce(cluster schoolid)

Linear regression                                  Number of obs     =      1,834
                                                   F(5, 39)          =    3147.47
                                                   Prob > F          =     0.0000
                                                   R-squared         =     0.9036
                                                   Root MSE          =      99.65

                              (Std. err. adjusted for 40 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
 end_savings | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   treatment |   95.10366   4.449578    21.37   0.000     86.10354    104.1038
     savings |   .9941231    .008127   122.32   0.000     .9776847    1.010562
             |
      strata |
        0 1  |  -3.676124   8.579742    -0.43   0.671    -21.03029    13.67804
        1 0  |  -6.918866    5.50784    -1.26   0.217    -18.05952    4.221791
        1 1  |  -5.730463   4.934736    -1.16   0.253    -15.71191    4.250983
             |
       _cons |   62.36888   7.014158     8.89   0.000      48.1814    76.55635
------------------------------------------------------------------------------
```

Guided practice: Running a DD regression (1/2)

Step 1: Instead of reshaping, we use an easier approach:

- Stack baseline and endline datasets (append)
- Create time dummy: 0 = baseline, 1 = endline
- Construct outcome variable that equals:
  - Baseline value if time = 0
  - Endline value if time = 1

```
gen time = 1
tempfile endline
save `endline'

use Dataset/GhanaVocationalFinance_randomized_baseline.dta, clear
gen time = 0

append using `endline'

gen     Y_savings        = savings
replace Y_savings        = end_savings   if time == 1

gen     post_T = treatment * time

/*Let's check what happened in the dataset after these commands*/
sort unique_ID
br unique_ID treatment time post_T Y_savings
```

Let's check what happens in the dataset...

```
br unique_ID treatment time post_T Y_savings
```

	unique_ID	treatment	time	Y_savings	post_T
95	48	0	1	556.257	0
96	48	0	0	484.2915	0
97	49	0	1	760.6503	0
98	49	0	0	690.8914	0
99	50	0	0	620.8835	0
100	50	0	1	579.2139	0
101	51	1	1	.	1
102	51	1	0	759.8786	0
103	52	1	1	667.7073	1
104	52	1	0	809.4647	0
105	53	1	1	900.408	1
106	53	1	0	827.9123	0
107	54	1	0	620.2534	0
108	54	1	1	921.2762	1
109	55	1	0	1124.388	0
110	55	1	1	1437.51	1

Guided practice: Running a DD regression (2/2)

Step 2: Estimate DD

- You can use either command:
  - `reg Y_savings treatment time post_T i.strata, vce(cluster schoolid)`
  - `reg Y_savings treatment##time i.strata, vce(cluster schoolid)`

Interpretation

The interaction term `post_T` (or `treatment##time`) is your DD estimate!

```
.
. reg  Y_savings treatment time  post_T   i.strata, vce(cluster schoolid)

Linear regression                               Number of obs     =      3,834
                                                F(6, 39)          =     371.08
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0521
                                                Root MSE          =     306.07

                               (Std. err. adjusted for 40 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
   Y_savings | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   treatment |   36.13559    14.3616     2.52   0.016     7.086524    65.18466
        time |   60.11643   4.459379    13.48   0.000     51.09649    69.13637
      post_T |   92.68769   5.980576    15.50   0.000     80.59084    104.7846
             |
      strata |
        0 1  |   18.46149   20.17474     0.92   0.366    -22.34578    59.26875
        1 0  |   52.76723   19.67564     2.68   0.011     12.96948    92.56497
        1 1  |   19.97234   18.36474     1.09   0.283    -17.17385    57.11853
             |
       _cons |   568.0473    14.6693    38.72   0.000     538.3758    597.7187
------------------------------------------------------------------------------
```

Summary: ITT vs. ANCOVA vs. DD

Estimator	Coefficient (SE)	Interpretation	Strengths	Weaknesses
ITT (Intention-to-Treat)	128.73 (14.81)	Simple difference in endline outcomes	Easy to compute: policy-relevant.	Can be biased if baseline imbalance.
ANCOVA	95.10 (4.45)	Controls for baseline outcome: more precise	High precision, accounts for baseline outcome.	Requires baseline data.
Difference-in-Differences (DD)	92.69 (5.98)	Estimate change over time, accounts for trends	Controls for time-invariant unobservables.	Less precision than ANCOVA, more complex setup.

TABLE 1. ITT EFFECTS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Amount saved last month (GH₵)	**Business knowledge (0-10)**	**Entrepreneurship index (0-1)**	**Financial confidence (1–5)**	**Financial literacy score (0–100)**	**Knows how to budget**	**Uses mobile money**
treatment	128.733***	2.057***	0.186***	0.492***	10.540***	0.089***	0.110***
	(14.814)	(0.090)	(0.012)	(0.049)	(0.589)	(0.016)	(0.021)
Constant	631.215***	5.054***	0.482***	3.423***	44.982***	0.715***	0.596***
	(14.502)	(0.102)	(0.012)	(0.055)	(0.661)	(0.019)	(0.020)
# Schools	40	40	40	40	40	40	40
Observations	1834	1834	1834	1834	1834	1834	1834
Model	ITT	ITT	ITT	ITT	ITT	ITT	ITT

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

TABLE 2. ANCOVA ESTIMATIONS

	(1)	(2)	(3)	(4)	(5)
	Amount saved last month (GH₵)	**Knows how to budget**	**Financial confidence (1–5)**	**Financial literacy score (0–100)**	**Uses mobile money**
treatment	95.104***	0.081***	0.452***	10.370***	0.079***
	(4.450)	(0.007)	(0.021)	(0.429)	(0.008)
Baseline outcome	0.994***	0.873***	0.922***	0.992***	0.910***
	(0.008)	(0.024)	(0.013)	(0.022)	(0.018)
Constant	62.369***	0.085***	0.247***	5.797***	0.053***
	(7.014)	(0.018)	(0.051)	(1.013)	(0.012)
# Schools	40	40	40	40	40
Observations	1834	1834	1834	1834	1834
Model	ANCOVA	ANCOVA	ANCOVA	ANCOVA	ANCOVA

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

TABLE 3. DID ESTIMATIONS

	(1)	(2)	(3)	(4)	(5)
	Amount saved last month (GH₵)	**Knows how to budget**	**Financial confidence (1–5)**	**Financial literacy score (0–100)**	**Uses mobile money**
post_T	92.688***	0.077***	0.457***	10.464***	0.071***
	(5.981)	(0.010)	(0.025)	(0.495)	(0.011)
treatment	36.136**	0.013	0.035	0.081	0.040*
	(14.362)	(0.017)	(0.048)	(0.410)	(0.022)
time	60.116***	0.006	-0.030	4.890***	0.007
	(4.459)	(0.005)	(0.020)	(0.329)	(0.006)
Constant	568.047***	0.710***	3.454***	39.801***	0.589***
	(14.669)	(0.020)	(0.042)	(0.408)	(0.019)
# Schools	40	40	40	40	40
Observations	3834	3834	3834	3834	3834
Model	DiD	DiD	DiD	DiD	DiD

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Non-compliance - LATE and Lee Bounds

Claire Stein, FEB, University of Groningen

Session Layout

Tuesday 29 July, 11:00-12:30 (1.5h)

Objective: Estimate treatment effects when there is attrition or non-compliance.

- Test for attrition (by treatment)
- Instrumental Variables: Local Average Treatment Effect (LATE)
- Lee Bounds

Files for the session:

- `GhanaVocationalFinance_randomized_endline.dta`
- `LATE_Leebounds.do`

Context: Participation in training and survey

Our endline dataset contains important variables that help distinguish between different groups of students:

- `treatment` – *Assignment* to the treatment group
- `took_training` – Actual *participation* in the training (take-up)
  - Some students assigned to the treatmedid not take up the training
- `endline_observed` – *Surveyed* at endline (attrition)
  - Some students (in both treatment and control) did not respond at endline

We can now study:

- Effects among compliers (LATE)
- Effects accounting for differential attrition (Lee Bounds)

```
.
. /*** STEP 2: Descriptive table of compliance and attrition ***/
. tab took_training endline_observed if treatment == 1

   Actually |
    took up |
   training |  Observed at endline
    (1=yes, |  (1=yes, 0=attrited)
      0=no) |         0          1 |     Total
-----------+----------------------+----------
          0 |         4        134 |       138
          1 |        37        725 |       762
-----------+----------------------+----------
      Total |        41        859 |       900

. /*
> Interpretation:
> - 4 students neither took up training nor participated in the survey
> - 134 students did not take up training but responded to the survey
> - 37 students took training but did not respond to the survey
> - 725 students took training and responded to the survey
> */
.
```

LATE: Imperfect compliance

Local Average Treatment Effect (LATE) logic: What is the causal effect of treatment among compliers – those who actually participated (take-up) in the training?

Why LATE?

- Treatment was randomized, but not all assigned received it (non-compliance).
- IV regression uses assignment as an instrument for actual take-up.
- Focuses on effect for compliers (Local Average Treatment Effect).

2-Stage Least Squares (2SLS):

1. First stage: $\text{Takeup}_i = \alpha + \beta T_i + \varepsilon_i$
2. Second stage: $Y_i = \alpha + \gamma \widehat{\text{Takeup}}_i + \varepsilon_i$

Guided practice: Estimating LATE

Step 1: First stage (IV):

- `reg took_training treatment i.strata, vce(cluster schoolid)`

Step 2: Second stage (2SLS)

- `ivregress 2sls end_savings (took_training = treatment) i.strata, vce(cluster schoolid)`

```
.
. /*** STEP 3: LATE Estimation ***/
. /*
> Estimate TOT (LATE) using IV: treatment instrumenting actual take-up (took_training)
> Only among observed students (compliers)
> */
. reg took_training treatment i.strata, vce(cluster schoolid)

Linear regression                               Number of obs     =      2,000
                                                F(4, 39)          =    2592.46
                                                Prob > F          =     0.0000
                                                R-squared         =     0.7527
                                                Root MSE          =      .2418

                              (Std. err. adjusted for 40 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
took_train~g | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   treatment |       .847   .0094295    89.82   0.000     .8279271    .8660729
             |
      strata |
        0 1  |   .0167143   .0113379     1.47   0.148    -.0062188    .0396474
        1 0  |   .0252857   .0170295     1.48   0.146    -.0091598    .0597312
        1 1  |        .02   .0090445     2.21   0.033     .0017058    .0382942
             |
       _cons |      -.014   .0069165    -2.02   0.050      -.02799    -.00001
------------------------------------------------------------------------------
```

```
.
. ivregress 2sls end_savings (took_training = treatment) i.strata, vce(cluster schoolid)

Instrumental variables 2SLS regression            Number of obs   =      1,834
                                                  Wald chi2(4)    =      89.24
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.0290
                                                  Root MSE        =      315.8

                               (Std. err. adjusted for 40 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
 end_savings | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
took_training |   152.4713   17.33859     8.79   0.000     118.4883    186.4543
             |
      strata |
        0 1  |   13.32058   19.75129     0.67   0.500    -25.39123    52.03239
        1 0  |   40.58592   19.49451     2.08   0.037     2.377371    78.79446
        1 1  |   13.30208   19.35968     0.69   0.492     -24.6422    51.24635
             |
       _cons |   633.4263   14.37813    44.05   0.000     605.2456    661.6069
------------------------------------------------------------------------------
Instrumented: took_training
 Instruments: 2.strata 3.strata 4.strata treatment
```

Lee Bounds: Correcting for attrition bias

When treatment changes attrition, we may be estimating treatment effects on **different sub-populations**

Lee Bounds provide a range within which the true treatment effect likely lies

How it works:

- Trims the distribution of outcomes in the group with **less attrition**
- Keeps only units most comparable to the other group
- Gives **lower and upper bounds** on the treatment effect

Let's first check attrition

Before conducting Lee Bounds analysis, we first need to examine attrition patterns.

Key questions:

- Are there differences in attrition between treatment and control groups?
- Are specific types of students more likely to attrit (e.g., gender)?

Reminder

Attrition that differs systematically across groups can bias treatment effect estimates – especially if related to outcomes.

Guided practice: Checking attrition

- `reg endline_observed treatment i.strata, vce(cluster schoolid)`

```
.
. /*** STEP 4: Attrition test ***/
. reg endline_observed treatment i.strata, vce(cluster schoolid)

Linear regression                               Number of obs     =      2,000
                                                F(4, 39)          =       7.82
                                                Prob > F          =     0.0001
                                                R-squared         =     0.0152
                                                Root MSE          =     .27412

                               (Std. err. adjusted for 40 clusters in schoolid)
--------------------------------------------------------------------------------
               |               Robust
endline_ob~d   | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
---------------+----------------------------------------------------------------
     treatment |   .0682222   .0136038     5.01   0.000     .0407059    .0957386
               |
        strata |
          0 1  |   .0097216   .0140854     0.69   0.494    -.0187688     .038212
          1 0  |   .0011502   .0305932     0.04   0.970    -.0607303    .0630307
          1 1  |   .0015385   .0148866     0.10   0.918    -.0285726    .0316495
               |
         _cons |   .8838974   .0132756    66.58   0.000     .8570449    .9107499
--------------------------------------------------------------------------------

. * Check if attrition differs across treatment and control
.
```

Guided practice: Lee Bounds

Install the command (if needed):

- `ssc install leebounds, replace`

Estimate Lee Bounds:

- `leebounds end_savings treatment, select(endline_observed)`

With tightening:

- `leebounds end_savings treatment, select(endline_observed) tight(strata)`

What is ”tightening”?

Tightening uses baseline covariates (like strata, age, or gender) to form cells. Bounds are computed within each cell and then averaged – this leads to narrower bounds when attrition patterns differ across types of individuals.

```
.
. leebounds end_savings treatment, select(endline_observed)

Lee (2009) treatment effect bounds

Number of obs.                          =    2000
Number of selected obs.                 =    1834
Trimming porportion                     =    0.0713

------------------------------------------------------------------------------
 end_savings | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
treatment    |
       lower |   82.43525   16.39844     5.03   0.000     50.29489    114.5756
       upper |   174.5305   16.45936    10.60   0.000     142.2708    206.7903
------------------------------------------------------------------------------
```

Interpretation:

- Are the bounds narrow or wide?
- Do they include zero?

Let's do Lee bounds manually (1/2)

The `leebounds` command does not allow for clustering of SEs, FEs, etc.
We can also estimate Lee bounds manually...

Steps:

1. Calculate the number of observations that need to be trimmed in the group with less attrition (here treatment) to get to the same attrition rate

2. Rank observations in ascending order (worst (1) to best (e.g., 1600) and descending order (opposite)

```
* i. Tabulate attrition by treatment status

tab endline_observed treatment
/*
- Control attrition rate = 125/1100 = 11.36%
- Treatment attrition rate = 41/900 = 4.56%
→ Differential attrition = 6.8% (higher in control)
→ We will trim 6.8% of the observed treatment group to match
→ 6.8% × 900 = ~61 observations to trim
Will lead to 102 observations missing in the treatment (=11.3% attrition)
*/


* ii. Generate ranked variables for trimming (within treatment group)
set seed 124        // for reproducibility

* Only rank non-missing endline observations
egen Asc_rank_end_savings = rank(end_savings) if treatment == 1 & end_savings < ., unique
egen Desc_rank_end_savings = rank(-end_savings) if treatment == 1 & end_savings < ., unique
```

Let’s do Lee bounds manually (2/2)

3. Prepare upper and lower bound variables by trimming the ”worst” and ”best” performers
    - Worst case scenario - the extra dropouts are the ”best” performers - **lower bound**
    - Best case scenario - the extra dropouts are the ”worst” performers - **upper bound**
4. Get lower and upper bound by regressing the variables on the treatment.

```
* iii. Create upper and lower bound outcome variables
gen     upper_end_savings = end_savings
replace upper_end_savings = . if treatment == 1 & Asc_rank_end_savings <= 61
// Trimming best outcomes -> conservative upper bound

gen     lower_end_savings = end_savings
replace lower_end_savings = . if treatment == 1 & Desc_rank_end_savings <= 61
// Trimming worst outcomes -> conservative lower bound

* iv. Estimate treatment effects using trimmed outcomes
reg lower_end_savings treatment i.strata, vce(cluster schoolid)

reg upper_end_savings treatment i.strata, vce(cluster schoolid)
```

	end_savings	Asc_rank_e~s	Desc_rank_~s	upper_end_~s	lower_end_~s
4	0	2	857	.	0
6	0	3	858	.	0
10	0	1	859	.	0
12	3.316856	4	856	.	3.316856
14	12.32289	5	855	.	12.32289
17	28.66033	6	854	.	28.66033
26	65.33304	7	853	.	65.33304
27	72.12195	8	852	.	72.12195
32	77.09892	9	851	.	77.09892
33	77.76219	10	850	.	77.76219
38	87.36093	11	849	.	87.36093
41	89.86476	12	848	.	89.86476
46	104.2899	13	847	.	104.2899
53	113.2368	14	846	.	113.2368
55	122.1922	15	845	.	122.1922
56	125.5838	16	844	.	125.5838
63	134.7546	17	843	.	134.7546
65	139.8108	18	842	.	139.8108
66	140.0837	19	841	.	140.0837
67	141.7988	20	840	.	141.7988
70	144.2497	21	839	.	144.2497
72	145.0282	22	838	.	145.0282
78	160.8831	23	837	.	160.8831
81	162.0157	24	836	.	162.0157
83	165.917	25	835	.	165.917
89	169.1493	26	834	.	169.1493

```
.
. * iv. Estimate treatment effects using trimmed outcomes
. reg lower_end_savings treatment i.strata, vce(cluster schoolid)

Linear regression                               Number of obs     =      1,773
                                                F(4, 39)          =      10.15
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0209
                                                Root MSE          =     297.61

                             (Std. err. adjusted for 40 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
lower_end_~s | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   treatment |   83.21802   13.98462     5.95   0.000     54.93146    111.5046
             |
      strata |
        0 1  |   20.73041   18.13583     1.14   0.260    -15.95276    57.41359
        1 0  |   40.59709   20.34838     2.00   0.053    -.5613972    81.75557
        1 1  |   15.73187   18.23518     0.86   0.394    -21.15226      52.616
             |
       _cons |   631.3042   14.23065    44.36   0.000     602.5199    660.0884
------------------------------------------------------------------------------

.
. reg upper_end_savings treatment i.strata, vce(cluster schoolid)

Linear regression                               Number of obs     =      1,773
                                                F(4, 39)          =      43.27
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0796
                                                Root MSE          =     297.72

                             (Std. err. adjusted for 40 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
upper_end_~s | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   treatment |   174.1891   13.72379    12.69   0.000     146.4301    201.9481
             |
      strata |
        0 1  |    18.6037   20.21572     0.92   0.363    -22.28645    59.49385
        1 0  |    34.7988   17.53282     1.98   0.054    -.664687    70.26228
        1 1  |   22.70651    18.1381     1.25   0.218    -13.98125    59.39427
             |
       _cons |   630.4826   14.27637    44.16   0.000     601.6059    659.3593
------------------------------------------------------------------------------

.
```

Recap: When to Use What?

ITT/ANCOVA/DD: Use if compliance is perfect and attrition is balanced
LATE (IV): Use if take-up is imperfect, but assignment is random
Lee Bounds: Use when there is differential attrition by treatment group

In Practice

- Check compliance to treatment and attrition first!
- If non-compliance: consider LATE
- If attrition differs by group: consider Lee bounds

TABLE 4. LATE ESTIMATIONS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Amount saved last month (GH¢)	**Business knowledge (0-10)**	**Entrepreneurship index (0-1)**	**Financial confidence (1–5)**	**Financial literacy score (0–100)**	**Knows how to budget**	**Uses mobile money**
Actually took up training	152.471***	2.437***	0.220***	0.583***	12.484***	0.106***	0.131***
	(17.339)	(0.111)	(0.013)	(0.057)	(0.687)	(0.019)	(0.025)
Constant	633.426***	5.089***	0.485***	3.431***	45.163***	0.717***	0.598***
	(14.378)	(0.103)	(0.011)	(0.054)	(0.653)	(0.019)	(0.020)
# Schools	40	40	40	40	40	40	40
Observations	1834	1834	1834	1834	1834	1834	1834
Model	LATE	LATE	LATE	LATE	LATE	LATE	LATE

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

TABLE 5.ITT WITH LEEBOUNDS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Amount saved last month (GH₵)	**Business knowledge (0-10)**	**Entrepreneurship index (0-1)**	**Financial confidence (1–5)**	**Financial literacy score (0–100)**	**Knows how to budget**	**Uses mobile money**
treatment	128.733***	2.057***	0.186***	0.492***	10.540***	0.089***	0.110***
	(14.814)	(0.090)	(0.012)	(0.049)	(0.589)	(0.016)	(0.021)
Constant	631.215***	5.054***	0.482***	3.423***	44.982***	0.715***	0.596***
	(14.502)	(0.102)	(0.012)	(0.055)	(0.661)	(0.019)	(0.020)
# Schools	40	40	40	40	40	40	40
Observations	1834	1834	1834	1834	1834	1834	1834
Model	ITT	ITT	ITT	ITT	ITT	ITT	ITT
Lee Bounds							
Lower	83.218***	1.763***	0.161***	0.410***	8.694***	0.073***	0.087***
	(13.985)	(0.083)	(0.011)	(0.047)	(0.585)	(0.017)	(0.022)
Upper	174.189***	2.358***	0.220***	0.650***	12.411***	0.150***	0.163***
	(13.724)	(0.091)	(0.012)	(0.045)	(0.579)	(0.016)	(0.021)
# Schools	40	40	40	40	40	40	40
Observations	1773	1773	1773	1773	1773	1773	1773

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Presenting PAPs

Claire Stein, FEB, University of Groningen

Session Layout

Tuesday 29 July, 13:30–15:00 (1.5h)

Objective: To present your Pre-Analyses Plans (PAP)

- Take 15 min to wrap up in your groups (decide who will present: ideally one person per section for equity!)
- Each group has 10-15 min to present their analysis
  - Present your PAP
  - Briefly discuss each section of the PAP

Power calculations

Claire Stein, FEB, University of Groningen

Session Layout

Wednesday 30 July, 14:00–15:00 (1h)

Objective: We will simulate a graph in Stata to visualize how power improves when we increase sample size

- Refresher: Why does power analysis matter?
- Basic formula
- Doing power calculations in stata

Files for the session:

- `power.do`

Power Analysis: Why It Matters

Statistical power is the probability of detecting a true effect when it exists.

Key concepts:

- **Power** (1 - κ): Probability of rejecting a false null hypothesis
- **Significance level** (α): Probability of false positive (Type I error)
- **Minimum Detectable Effect** (MDE): Smallest effect size you can detect with given power
- **Sample size** (N): Larger samples increase power

Power depends on: Effect size, standard deviation, sample size, and allocation ratio

Power Formula and MDE

Minimum Detectable Effect (MDE) is:

$$MDE = (t_{1-\kappa} + t_{\alpha}) \times SE_{\hat{\beta}}$$

With OLS and simple randomization:

$$SE_{\hat{\beta}} = \sqrt{\frac{\sigma^2}{P(1-P)N}}$$

Implications:

- Increasing sample size (N) decreases MDE
- Balanced treatment assignment (P=0.5) maximizes power

What Affects Power? Practical Considerations

- **Clustering**: Increases standard errors $\Rightarrow$ higher MDE
- **Covariates**: Can reduce residual variance if predictive $\Rightarrow$ lower MDE
- **Noncompliance**: Lowers effective treatment contrast $\Rightarrow$ higher MDE
- **Stratification**: Improves balance and efficiency

Rule of thumb: 80% power ($\kappa = 0.20$), 5% significance ($\alpha = 0.05$)

Comparing Estimators: Post, DiD, ANCOVA

Model	Uses Pre Data?	Variance	Power
Post-only	No	$\frac{2\sigma^2}{n}$	Lowest (reference)
DiD	Yes (as diff)	$\frac{4\sigma^2}{n}(1-\rho)$	Higher if $\rho > 0.5$
ANCOVA	Yes (as control)	$\frac{2\sigma^2}{n}(1-\rho^2)$	Always highest (if RCT)

Note: ANCOVA gives the most power when outcomes are correlated over time ($\rho > 0$) – especially in RCTs.

Power in Practice: Visualizing Effect Size and SD

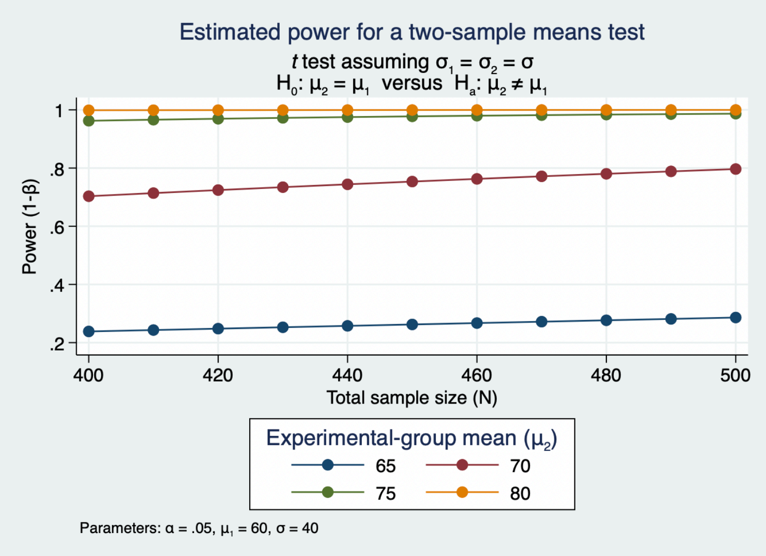
How big does our sample need to be to detect an increase in savings?

Parameters:

- Mean in control group: $60/month
- We expect a treatment effect of +$5 to +$20
- Standard deviation: 40
- Sample size range: 400 to 500 participants (200–250 per group)

STATA command:

```
power twomeans 60 (65(5)80), sd(40) n(400(10)500) graph
```



Estimated power for a two-sample means test
t test assuming $\sigma_1 = \sigma_2 = \sigma$
H_0: $\mu_2 = \mu_1$ versus H_a: $\mu_2 \neq \mu_1$
Power (1-β)
1
.8
.6
.4
.2
400
420
440
460
480
500
Total sample size (N)
Experimental-group mean (μ_2)
65
70
75
80
Parameters: $\alpha = .05$, $\mu_1 = 60$, $\sigma = 40$

Interpretation

- Shows how power increases with bigger effects or larger samples
  - We would need a *much much bigger* sample size to detect the smallest effect, but costly... may need to think of a stronger program...!
- Helps determine the minimum detectable effect (MDE) for a fixed sample
- Reflects two-group RCT design, where total sample is split

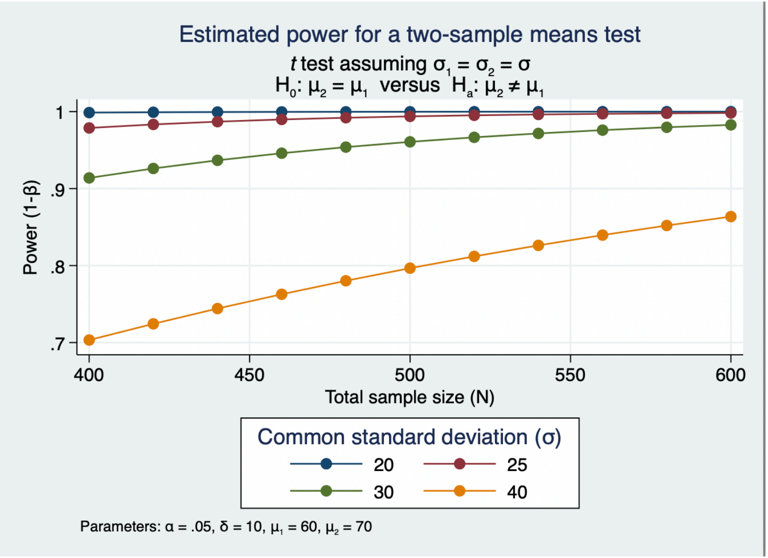
Impact of Standard Deviation on Power

What if savings vary more than expected?

Same savings example, fixed effect size (60 vs. 70), varying standard deviation.

STATA command:

```
power twomeans 60 70, sd(20 25 30 40) n(400(20)600) graph
```


Estimated power for a two-sample means test
t test assuming σ1 = σ2 = σ
H0: μ2 = μ1 versus Ha: μ2 ≠ μ1
Power (1-β)
1
.9
.8
.7
400
450
500
550
600
Total sample size (N)
Common standard deviation (σ)
20
25
30
40
Parameters: α = .05, δ = 10, μ1 = 60, μ2 = 70

Interpretation

- Larger outcome variance (SD) reduces our ability to detect effects
- Important to measure outcomes precisely or use covariates to reduce variance
- Power drops sharply when SD increases, especially with small samples

Power Calculation: Individual Randomization

What sample size is needed to detect a $5 increase in monthly savings?

Design:

- Mean in control group: $60
- Mean in treatment group: $65
- Standard deviation: 20 in both groups
- Desired power: 80% at significance level $\alpha = 0.05$

STATA command:

```
sampsi 60 65, sd1(20) sd2(20) alpha(0.05) power(0.8)
```

```
Estimated sample size for two-sample comparison of means

Test H0: m1 = m2, where m1 is the mean in population 1
                    and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.8000
            m1 =       60
            m2 =       65
           sd1 =       20
           sd2 =       20
         n2/n1 =     1.00

Estimated required sample sizes:

            n1 =      252
            n2 =      252
```

Interpretation: Individual Randomization

STATA tells us:

- Estimated sample size needed per arm: 252 individuals
- Total sample size = 252 (control) + 252 (treatment) = 504

Takeaway:

- With equal SDs and no clustering, detecting a 5-unit effect on savings requires about 250 people per group.
- Smaller effect sizes or higher SDs would require a larger sample.

Power Calculation: Clustered Randomization

What if we randomize at the group (cluster) level?

Design:

- Mean in control group = \$60; treatment = \$65
- SD = 15, ICC = 0.05
- Average cluster size = 30

STATA command:

```
clustersampsi samplesize means, mu1(60) mu2(65) sd1(15) sd2(15)
rho(0.05) m(30) alpha(0.05) beta(0.8)
```

```
For the user specified parameters:

mean 1:                                                     60.00
mean 2:                                                     65.00
standard deviation 1:                                       15.00
standard deviation 2:                                       15.00
significance level:                                         0.05
power:                                                      0.80
baseline measures adjustment (correlation):                 0.00
average cluster size:                                       30
intra cluster correlation (ICC):                            0.0500
coefficient of variation (of cluster sizes):                0.00


clustersampsi estimated parameters:


Firstly, assuming individual randomisation:
sample size per arm:                                        142

Then, allowing for cluster randomisation:
design effect:                                              2.45
sample size per arm:                                        390
number clusters per arm:                                    13

Note: sample size per arm required under cluster randomisation is rounded up
to a multiple of average cluster size and includes the addition
of one extra cluster per arm (to allow for t-distribution).
To understand sensitivity to these conservative allowances:
power with m clusters per arm:                              0.82
power with m-1 clusters per arm:                            0.78
```

Interpretation: Clustered Randomization

STATA tells us:

- Under individual randomization: 142 people per arm
- Under clustering (ICC = 0.05, 30 per cluster): 390 people per arm
- Requires 13 clusters per arm
  - **Note:** STATA internally adds 1 extra cluster during calculations to be conservative (for small sample correction)
  - *Why?* Small number of clusters → use of t-distribution → extra margin added
  - *Sensitivity:* If one cluster drops out, power can fall below 80% (e.g., from 82% to 78%)

Takeaway:

- Clustering reduces power due to intra-cluster correlation.
- Even with the same means and SDs, you need more people when using a clustered design.

If you want more practice...

In Stata:

1. Try adjusting n, SD, effect size, with a two-sample comparison of means

```
sampsi 600 660, sd1(300) sd2(300) n1(500) n2(500) alpha(0.05)
```

2. Visualize how power changes

```
power onemean 600 (620(10)660), sd(300) n(500) graph
```

Bonus Step. Try a clustered design, adjusting ICC (rho)

```
clustersampsi samplesize means, mu1(600) mu2(660) sd1(300) sd2(300) rho(0.05)
m(30)
```

Practice with a **JPAL** online tool

What happens to power when the N increases? When SD is higher? When ICC changes?

For more advanced Stata code on power: click here

If you want more practice...

Presenting impact analyses

Claire Stein, FEB, University of Groningen

Session Layout

Wednesday 30 July, 15:30–17:00 (1.5h)

Objective: To present your impact analyses

- Take 15 min to wrap up in your groups (decide who will present)
- Each group has 10–15 min to present their analysis
  - Present screenshots of your results
  - Briefly discuss estimations and provide interpretation

References

Pascaline Dupas (2011)
Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya
American Economic Journal: Applied Economics 3(1), 1 – 34.

Nina Hansen et al. (2024)
The impact of a "progress with your goals" curriculum and its cultural framing among female microfinance clients in Vietnam.
AEA RCT Registry

David McKenzie, WB blog (2012)
A pre-analysis plan checklist
World Bank blogs

university of groningen

Thank you for your attention

Claire Stein
c.l.stein@rug.nl

Faculty of Economics and Business
University of Groningen

July 26-31, 2025