

One page, written before the data arrives

Every design in this course was reconstructed after the fact, and each reconstruction had a choice in it that the result could have influenced. This lesson writes the choices down first — the design, the outcome, the threshold and the analysis — on one page a programme will actually produce.

Pierre Robentz Cassion

Lesson 7 of 8

Impact Evaluation Methods — What the evaluation claims

What this lesson covers

- Six choices the result could have made for you
- The page
- Why each section is there
- Deviating from the plan
- Pre-specification does not mean no exploration
- When there is no plan and the data has arrived
- Report it whole
- What comes next

Six choices the result could have made for you

Lesson	The choice	What it could have been chosen to produce
1	Before-after, cross-section or difference-in-differences	+6.97, +2.12 or -0.96 points
2	Which covariates count as balance	A table that looks acceptable
3	Clustered or naive standard errors	An interval that excludes zero
4	Match or do not match	-0.71 or -0.96, and which schools are in
5	Which bandwidth	Whichever one is significant
6	Which ICC	An MDE below the observed effect

Six choices the result could have made for you

- Six binary choices produce sixty-four analyses — None of them is dishonest and any of them can be defended in. . .
- This is the multiple-comparisons problem from the statistics course wearing a different coat — There it was twenty-four. . .

Evaluation plan: school feeding and learning outcomes

Written 2024-01-15, before any endline data exists.

1. QUESTION

Does the school feeding programme raise literacy scores?

2. DESIGN

Difference-in-differences. 15 programme schools, 9 comparison schools, baseline and endline literacy assessment on the same children.

Assignment was not randomised; the balance table will be reported.

3. OUTCOME

Percent correct on the literacy assessment. Baseline is out of 40 items and endline out of 50, so raw scores are not comparable and percent correct is the primary outcome. Numeracy is secondary.

4. THRESHOLD OF INTEREST

3 percentage points. Below that the programme would not be expanded on these grounds, so an effect smaller than 3 points is reported as "no programme-relevant effect" regardless of its p-value.

5. ANALYSIS

OLS of the individual gain on a programme indicator, standard errors clustered on school. Adjusted for baseline score and district. Reported with a 95% confidence interval, not a p-value alone.

6. POWER

ICC assumed 0.065 from the attendance data. Minimum detectable effect 4.9 points, which is above the 3-point threshold. The evaluation is therefore underpowered for the effect size that matters, and this is stated in the report whatever the result.

7. WHAT WOULD CHANGE THE CONCLUSION

Attrition above 20%, or differential attrition between arms above 5 points, would make the panel unrepresentative and the primary analysis would be reported alongside a bounds analysis.

8. WHAT WE WILL NOT CLAIM

Causality beyond difference-in-differences. Effects on attendance, enrolment or nutrition, which are not measured here.

- Point 6 is the one that makes this document worth writing — It is knowable in January, it says the study cannot answer. . .

Why each section is there

- The question, in one sentence — If it takes more than one, it is more than one evaluation
- The design, named — "We will compare programme and non-programme schools" is not a design; it is three designs that. . .
- The outcome, defined to the item — The statistics course's exercise found a proposal comparing a 40-item paper to a. . .
- The threshold of interest, in programme units — This is the section nobody writes and it is the one that stops a. . .
- The analysis, specified — Including the standard errors, because the choice between naive and clustered moved the. . .
- The power, computed — Before, not after

Why each section is there (cont.)

- What would change the conclusion — Naming the failure modes in advance means the report does not have to argue about. . .
- What you will not claim — The shortest section and the one that protects the evaluation's credibility when someone else. . .

Deviating from the plan

- Deviations are allowed and must be listed — A short table in the report — what was planned, what was done, why — costs. . .

Deviating from the plan — Example

Deviations from the evaluation plan

Planned: adjust for baseline score and district.

Done: as planned.

Planned: primary analysis on all enrolled students.

Done: restricted to the 585 students with both assessment rounds.

Why: 156 students have no endline. Attrition analysis added; the students who left scored 39.3% at baseline against 56.8% for those who stayed, so the panel is not representative and this is reported as a limitation.

Deviating from the plan

- A listed deviation is a strength — An unlisted one, discovered by a reviewer, is the end of the evaluation's. . .

Pre-specification does not mean no exploration

- Two sections, clearly labelled — The pre-specified analysis answers the question that was asked

Pre-specification does not mean no exploration — In Python

```
PRIMARY = "gain ~ feeding_programme + baseline + district"    # pre-specified
EXPLORATORY = [                                                # labelled as such
    "gain ~ feeding_programme * sex",
    "gain ~ feeding_programme * baseline_quartile",
    "gain ~ feeding_programme + C(school_id)",
]
print(f"pre-specified: 1 model.  exploratory: {len(EXPLORATORY)} models.")
```

Pre-specification does not mean no exploration — In R

Two objects, two headings in the report. That is the whole discipline.

Pre-specification does not mean no exploration

- Subgroup findings are exploratory unless the subgroup was named in advance —
The statistics course found two. . .

When there is no plan and the data has arrived

- Write the plan anyway, dated, before you run the analysis — You have seen the data; you have not yet seen the result
- Pre-specify the primary analysis and report everything else as exploratory — The distinction still means something even. . .
- State how many analyses you ran — The report block in the statistics course ends with "comparisons run in total", and. . .

Report it whole — Example (cont.)

Analysis plan and deviations

The evaluation plan was written on 2024-01-15, before endline data collection, and is reproduced in Annex A.

Primary analysis as pre-specified: difference-in-differences on percent correct, clustered on school, adjusted for baseline and district.

Deviations: one, listed in Annex A. The panel was restricted to students with both rounds; an attrition analysis was added.

Exploratory analyses: 3, reported in Annex B and labelled as exploratory. None is presented as a finding.

Threshold of interest: 3 percentage points, agreed with the programme team in January. The observed estimate is -0.96 points (95% CI -3.5 to +1.6)

Report it whole — Example (cont.)

and is below the threshold in magnitude.

Report it whole

- The date in the first line is the whole document's load-bearing element — Everything else is a claim about intent; the...

What comes next

- The plan says what the evaluation will claim.

Read the full lesson, with runnable code

[https://data-analysis.cassion.dev/courses/impact-evaluation-methods/
ie-07-decide-the-design-first](https://data-analysis.cassion.dev/courses/impact-evaluation-methods/ie-07-decide-the-design-first)