The reference sheet, field by field

One page per indicator, thirteen fields, and a test that tells you when it is finished — hand it to a second analyst and see whether they get your number without asking a question.

Pierre Robentz Cassion

Lesson 3 of 8

What this lesson covers

- The test the sheet has to pass
- The fields
- Worked: penta3 coverage
- The field everyone omits
- Keep it beside the code
- Write the exclusions as code, not as prose
- Version it
- The two-analyst test, in practice
- What comes next

The test the sheet has to pass

- An indicator reference sheet is finished when a competent analyst who has never met you can compute your number from it, from the raw data, without asking you a question.

The fields

Field	What it holds
Name	Including the unit of measure
Definition	One sentence a non-analyst can read
Numerator	The exact condition, in words
Denominator	The exact population, and where it comes from
Exclusions	What is deliberately left out, and why
Unit of measure	Percent, count, per 1,000, litres per person per day
Direction	Higher is better, lower is better, or neither
Disaggregation	The cuts that will be reported, and the minimum cell size

. . .

Worked: penta3 coverage — Example (cont.)

Name	penta3_coverage_percent_monthly
Definition	The share of the target infant population that received a third dose of pentavalent vaccine in the month.
Numerator	Third doses of pentavalent vaccine administered, as reported by facilities that submitted a monthly report for that month.
Denominator	Surviving infants in the catchment, from the MoH population estimates 2024, projected from the 2015 census at 2.4% annual growth, summed across facilities that reported.
Exclusions	Facility-months with report_submitted = false are excluded from both numerator and denominator. Doses given to children outside the catchment are not separable and remain in the numerator.
Unit	Percent
Direction	Higher is better
Disaggregation	Month, facility type, district. Minimum cell 30 in denominator.
Frequency	Monthly, reported 6 weeks after month end
Source	DHIS2 data element PENTA3_DOSES, org unit level 4, monthly

Worked: penta3 coverage — Example (cont.)

Computation	`100 * sum(doses) / sum(target_population)`, over reporting facility-months only. Not the mean of facility-level rates.
Decision	Whether a district is prioritised for outreach in the next quarterly microplan.
Limitations	Reporting completeness was 76.5% in 2024 and fell to 29% in August; the figure is a lower bound in any month where completeness is below 90%. Catchments overlap, so facility-level values are unreliable; use district level or above.

Worked: penta3 coverage

- "Over reporting facility-months only. Not the mean of facility-level rates." — Those two sentences settle a difference...

The field everyone omits

- **It exposes indicators that inform nothing.** Some of those are still required — a donor asks for them — and that is a...
- **It exposes indicators that inform two decisions.** Coverage used both to prioritise outreach *and* to trigger a...
- **It sets the precision you need.** An indicator that prioritises a district needs to be right about the ranking. One...

Keep it beside the code — In Python

```
import json
from pathlib import Path

sheet = json.loads(Path("indicators/penta3_coverage.json").read_text())

assert sheet["denominator_source"], "denominator source is required"
assert sheet["decision_informed"], "an indicator with no decision needs saying so"

numerator = vax.loc[vax["reported"] & (vax["antigen"] == "penta3"),
                    "doses_administered"].sum()
denominator = vax.loc[vax["reported"] & (vax["antigen"] == "penta3"),
                      "target_population"].sum()

print(f"{sheet['name']}: {100 * numerator / denominator:.1f}")
```

Keep it beside the code — In R

```r
sheet <- jsonlite::read_json(here::here("indicators", "penta3_coverage.json"),
                             simplifyVector = TRUE)

stopifnot(nzchar(sheet$denominator_source), nzchar(sheet$decision_informed))

vax |>
  filter(report_submitted, antigen == "penta3") |>
  summarise(value = 100 * sum(doses_administered) / sum(target_population))
```

Write the exclusions as code, not as prose — In Python

```
EXCLUSIONS = [
    ("non-reporting facility-months", lambda d: ~d["reported"]),
    ("other antigens", lambda d: d["antigen"] != "penta3"),
]

remaining = vax.copy()
for label, rule in EXCLUSIONS:
    dropped = rule(remaining).sum()
    remaining = remaining[~rule(remaining)]
    print(f"excluded {dropped:>5} rows: {label}")
print(f"{len(remaining)} rows in the indicator")
```

Write the exclusions as code, not as prose — In R

```r
EXCLUSIONS <- list(
  "non-reporting facility-months" = function(d) !d$report_submitted,
  "other antigens"                = function(d) d$antigen != "penta3"
)

remaining <- vax
for (label in names(EXCLUSIONS)) {
  rule <- EXCLUSIONS[[label]]
  cat(sprintf("excluded %5d rows: %s\n", sum(rule(remaining)), label))
  remaining <- remaining[!rule(remaining), ]
}
```

Version it — Example

```
version   2.1
changed   2026-07-28
change    Denominator restricted to reporting facilities. Previously all
          facilities, with non-reporters counted at their target population,
          which understated coverage by about 23% in August 2024.
effect    Series revised from 2024-01. Values before v2.1 are not comparable.
```

Version it

- Never silently improve a definition — A coverage figure that rises eight points because the definition changed,...

The two-analyst test, in practice

- The colleague used the calendar month; you used the reporting month.
- The colleague computed a mean of facility rates; you computed a ratio of sums.
- The colleague included the district hospital; your extract excluded it because of an org-unit level filter nobody...

What comes next

- The field that generates more disagreement than the other twelve combined is the denominator.

Read the full lesson, with runnable code

https://data-analysis.cassion.dev/courses/indicator-design-logframe/indicators-03-the-reference-sheet