



FROM THE RESEARCH BENCH

What is selection bias? The sample can change the answer

Selection bias begins with the inclusion mechanism. Name the target, eligibility rule, exclusions, and conditioning event before comparing outcomes.

Sinan Isoglu, MBA (Quantic)

Commercial growth leader, lecturer and doctoral researcher

PUBLISHED

6 September 2026

READING TIME

5 min

WORDS

1,041

<https://isoglu.com/insights/journal/what-is-selection-bias/>

Management summary	2
What does selection bias mean?	3
Is selection bias the same as confounding?	3
What does targeted exposure teach us?	4
What does an eligibility ledger look like?	4
How should a team review a selected sample?	5
References	7

MANAGEMENT SUMMARY

Selection bias is a distortion created when the mechanism that puts units into an observed or analyzed sample is related to variables relevant to the target question. It is not the same as confounding, missingness, nonresponse, measurement error, or attrition, although those mechanisms can interact. Lewis and Rao provide a bounded advertising example in which targeted exposure creates a central selection concern and randomized trials add unbiased information under their design. This article builds a synthetic eligibility ledger with target population, inclusion rule, observed sample, selection variable, possible direction, and design response. The ledger and causal examples are author synthesis. They do not estimate bias for a current sample or prescribe adjustment without a design and identification argument.

Keywords: Selection Bias · Sample Selection · Eligibility Rule · Collider Bias · Causal Inference · Target Population

A conversion rate can look excellent because only the units that reached a late stage are counted. A survey can look representative because the people who answered were described carefully. A campaign comparison can look balanced because the exposed and unexposed groups have the same label. In each case, the question is earlier: who had a chance to enter the sample, and who did not?

Selection bias begins with an inclusion mechanism that changes which units are observed or analyzed relative to the target question. The sample can change the answer before the model is fitted.

The evidence-over-anecdote article owns the wider standard for claims that survive scrutiny. This page owns the population boundary: which units entered, which were excluded, and what that choice means for the comparison.

What does selection bias mean?

Name four objects before calculating a rate:

Table 1 What does selection bias mean?

Object	Question	Example
Target population	To whom or what should the conclusion speak?	All eligible accounts in a defined market and period
Eligibility rule	Which units are allowed to enter the intended population?	Accounts meeting region, product, and date criteria
Observed sample	Which units were actually measured or analyzed?	Accounts with a response, event, or usable record
Conditioning event	What later state or variable restricted the comparison?	Reached proposal, answered survey, or accepted treatment

Table from this essay. Sources and interpretation are given in the article.

Selection bias is not the fact that a sample is smaller than a target. It is the risk that the path into the sample depends on variables that matter for the outcome or exposure comparison. A small random sample can be informative. A large selected sample can still answer a different question.

Is selection bias the same as confounding?

No. Confounding concerns a common cause of exposure and outcome that distorts a comparison. Selection bias concerns the mechanism that determines which units are observed, retained, or analyzed. Measurement error concerns how a target is observed. Nonresponse is a missing response from a selected unit. Attrition is loss after entry. The mechanisms can coexist, but the repair question differs.

Collider bias is one important selection pathway. Suppose prior buyer intent and sales attention both increase the chance that a lead becomes “sales accepted.” If analysis conditions on accepted leads, that shared consequence can create an association between the two causes even if none existed in the target population. The label “accepted” does not make the selected comparison neutral.

Do not infer collider bias from every post-treatment variable. Draw the inclusion path, name the variables, and state which comparison the conditioning event changes.

What does targeted exposure teach us?

Lewis and Rao analyze large digital advertising experiments and emphasize two bounded points. Randomized trials add unbiased information under the assignment design. Observational advertising methods face a serious selection concern because exposure is targeted rather than random (Lewis & Rao, 2015). Their advertising setting is not a general conversion benchmark. It is a clear example of why a selected comparison needs a design argument.

The same logic applies to a commercial sample only as a question, not as a transferred estimate:

- Were units invited, routed, exposed, or retained by a rule related to the outcome?
- Did the treatment or commercial team influence who received a measurement?
- Did the analysis condition on a later stage, response, or acceptance event?
- Are excluded or missing units described well enough to test the target boundary?

What does an eligibility ledger look like?

The six rows are synthetic. They do not represent participants, customers, employees, or a current company sample.

Figure 1 The synthetic selection-bias eligibility ledger

ID	Target and intended unit	Inclusion or conditioning event	Selected sample	Selection variable to inspect	Possible direction	Design response
S-01	All eligible trial accounts	Account completed setup	Setup completers only	Baseline capability and motivation	Unknown	Compare entry population; retain non-completers
S-02	All invited buyers	Survey response	Respondents	Interest and response burden	Unknown	Track non-response; compare frame variables
S-03	All routed leads	Sales acceptance	Accepted leads	Qualification and seller capacity	Could change both ways	Preserve rejected leads; model routing path
S-04	All campaign targets	Ad exposure	Exposed and hold-out users	Targeting score and prior behavior	Likely selection risk	Random assignment or validated design
S-05	All open opportunities	Proposal reached	Proposal-stage records	Deal maturity and seller choice	Unknown	Report stage entry; do not call late-stage rate funnel-wide
S-06	All retained accounts	Remained observable through day 60	Day-60 responders	Early value and survival	Unknown	Treat attrition as a separate outcome and sensitivity

Author's synthetic eligibility ledger grounded in Lewis and Rao (2015). Targets, selection paths, possible directions, tests, and dispositions are illustrative.

S-03 is not automatically biased in one direction. If seller capacity favors high-value leads, the selected sample may look better than the target. If difficult leads are routed for extra attention, the selected sample may look worse. The direction belongs to the mechanism and the estimand, not to the word “selection.”

How should a team review a selected sample?

- 1 State the target population, unit, time window, treatment or exposure, and outcome.
- 2 Draw the path from target to invited, observed, retained, and analyzed units.
- 3 Record eligibility, invitation, take-up, response, retention, and exclusion rules separately.

- 4 Identify variables that may affect both inclusion and the target comparison.
- 5 Name whether the response is a design, weighting, stratification, sensitivity, or new-data question.
- 6 Report the excluded and missing population beside the selected result.

If the inclusion mechanism is unknown, narrow the conclusion to the observed sample or stop. Do not repair selection bias by changing the sample label, adding an unmotivated covariate, or comparing two selected groups without an identification argument.

Selection bias is a population and process problem before it is a modelling problem. The first question is not “what adjustment should we run?” It is “which units did the data-generating path leave out?”

The external-validity article takes up the next question: whether an identified comparison can travel to a target population.

REFERENCES

Lewis, R. A., & Rao, J. M. (2015). The unfavorable economics of measuring the returns to advertising. *The Quarterly Journal of Economics*, 130(4), 1941-1973. <https://doi.org/10.1093/qje/qjv023>

END OF ESSAY

HOW TO CITE THIS ESSAY

Isoglu, S. (2026, September 6). What is selection bias? The sample can change the answer.
Sinan Isoglu. <https://isoglu.com/insights/journal/what-is-selection-bias/>

KEEP READING

From the research bench

What is external validity? A result needs a transfer boundary

<https://isoglu.com/insights/journal/what-is-external-validity/>

From the research bench

Staggered difference-in-differences needs a cohort-time estimand

<https://isoglu.com/insights/journal/staggered-difference-in-differences-needs-a-cohort-time-estimand/>

From the research bench

What are parallel trends? The assumption behind difference-in-differences

<https://isoglu.com/insights/journal/what-are-parallel-trends/>



Sinan Isoglu, MBA (Quantic)

Commercial growth leader, lecturer and doctoral researcher