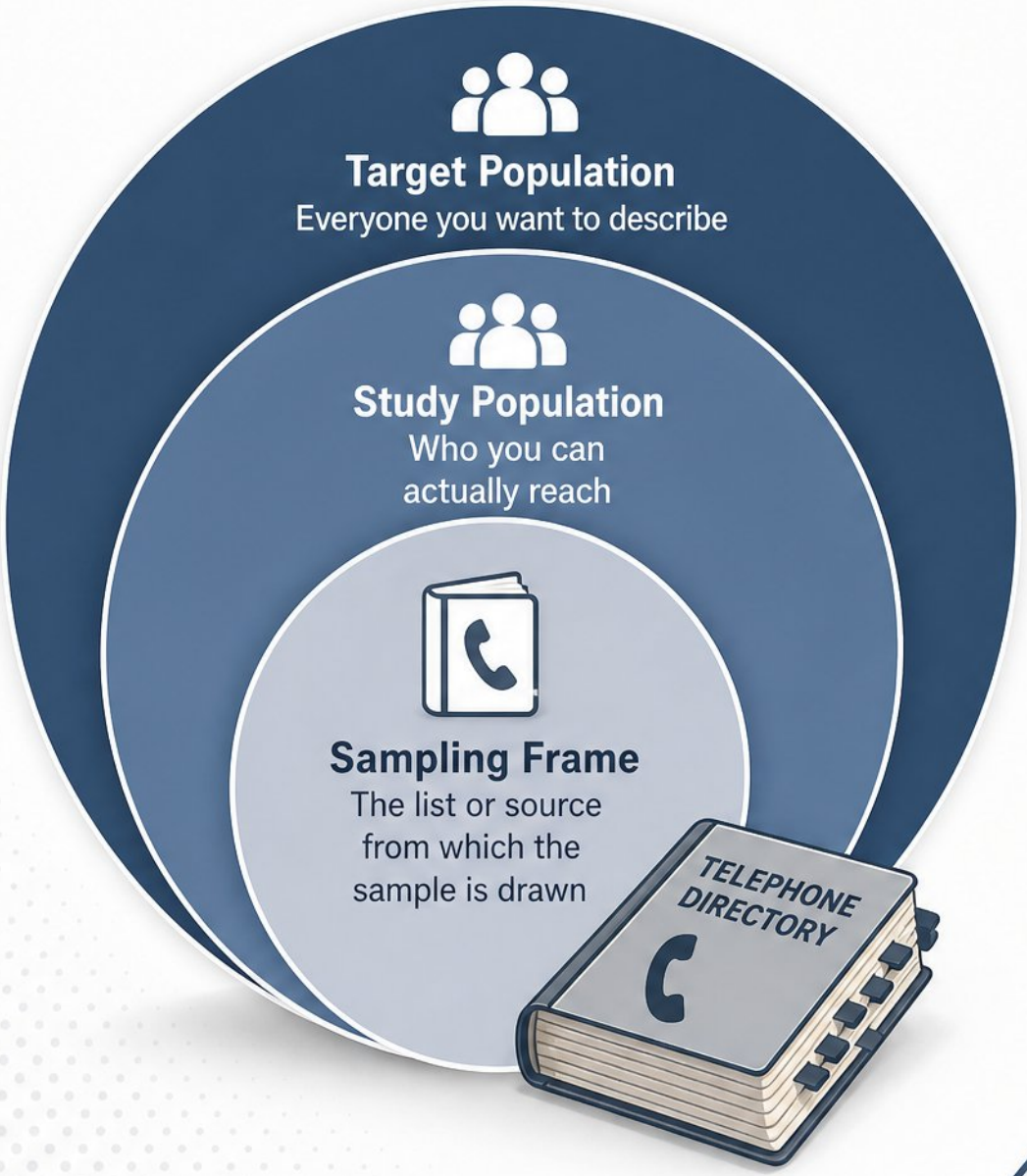
Introduction to Sampling and Bias

- Population = the whole pot of soup; sample = the spoonful you taste.
- We sample because measuring everyone is impossible: cost, time, logistics.
- Bias is systematic error in selection or measurement, not random chance.
- Non-representative samples mislead: wrong proportions = wrong conclusions.
- Real-world cost: female-focused startups lost 45% growth from biased beta tests; polls missed 3–4 points due to non-response bias.

Populations, Frames, and the Ideal Sample



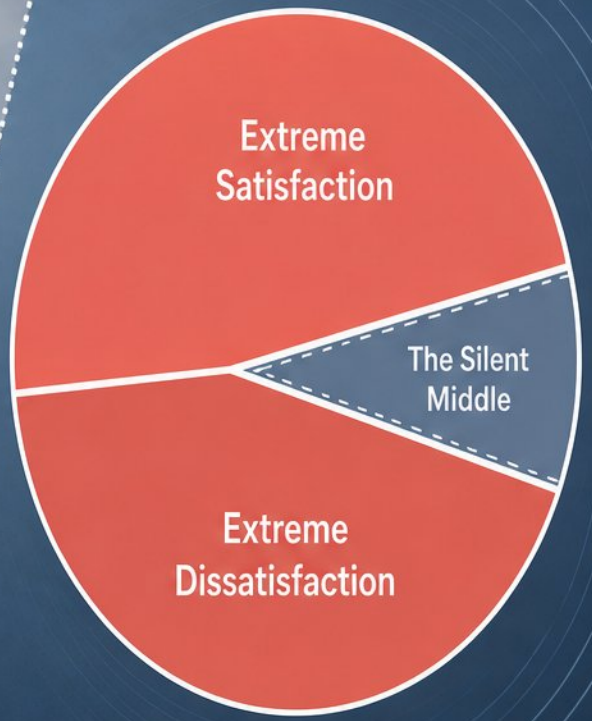
- **Target population:** everyone you want to describe
- **Study population & frame:** who you can actually reach
- **Coverage error:** the frame misses groups (e.g., no-landline households)
- **Simple random sampling** as the gold standard for representativeness
- **Non-response bias:** even a perfect frame fails when people don’t respond

Selection Bias:
How Sample Composition Goes Wrong

- **Systematic distortion** in sample composition vs. the target population
- **Undercoverage bias:** missing groups like offline rural/older populations
- **Self-selection bias:** only extreme experiences shown (e.g., opt-in reviews)
- **Survivorship bias:** ignoring cases filtered out (WWII plane armor example)
- **Check:** Who is missing? How were participants recruited?

Response and Non-Response Bias

- Non-response distorts data when those who don't answer differ from those who do
- 2024 polls with ~1.4% response rates show total error of **±49%** without knowledge of non-responders
- Voluntary response overrepresents extreme opinions; the silent middle is invisible
- 2020 and 2024 elections underestimated Trump by **~3 points** due to differential non-response
- Mitigation: callback designs, weighting adjustments, early vs. late responder comparison



Measurement Bias: When Instruments and Questions Mislead

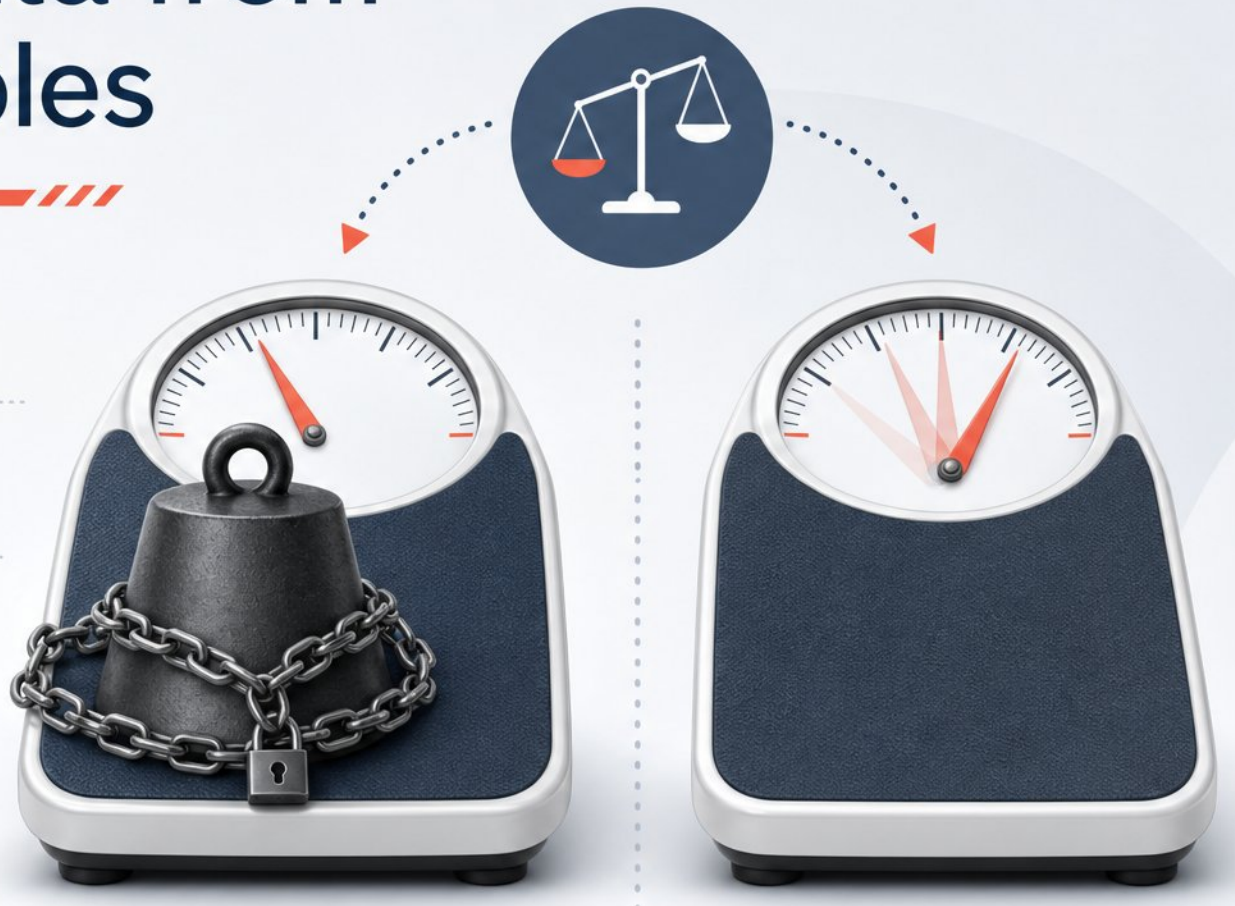
- Distortion from how data are collected, not just who is sampled
- Wording shapes answers: leading questions push responses
- Social desirability: underreporting bad habits, overreporting virtue
- Recall errors, interviewer effects, faulty sensors introduce bias
- Thumbs-down buttons capture frustration but miss "good enough" users

Real-World Consequences of Biased Samples

- Female-focused products saw 45% less growth due to early-user gender skew
- 6-10% of A/B tests have invalid splits, silently shipping artifacts
- Enumerator manipulation underreported marginalized populations, biasing fertility estimates
- Feedback buttons over-represent extreme users; the "good enough" majority stays silent

Analyzing Data from Biased Samples

- Larger N shrinks random variance but never fixes systematic bias
- Bias = consistently off-target scale; Variance = fluctuating scale
- Weighting adjusts under/over-represented groups using known benchmarks
- Non-ignorable non-response: response likelihood tied to what you measure
- Transparency about limitations is ethical; overclaiming harms decisions



SYSTEMATIC BIAS
Consistently off-target scale

RANDOM VARIANCE
Fluctuating scale

Spotting Bias in Reports, Polls, and Dashboards

- Verify sponsorship, population, mode, dates, sample size, margin of error
- Missing methodology, small samples, selective crosstabs are red flags
- Likely vs. registered voter screen can create 2-point systematic swing
- Media favors dramatic outliers over averages, horse-race over substance
- A 3-point lead with ±3% MOE is a statistical tie

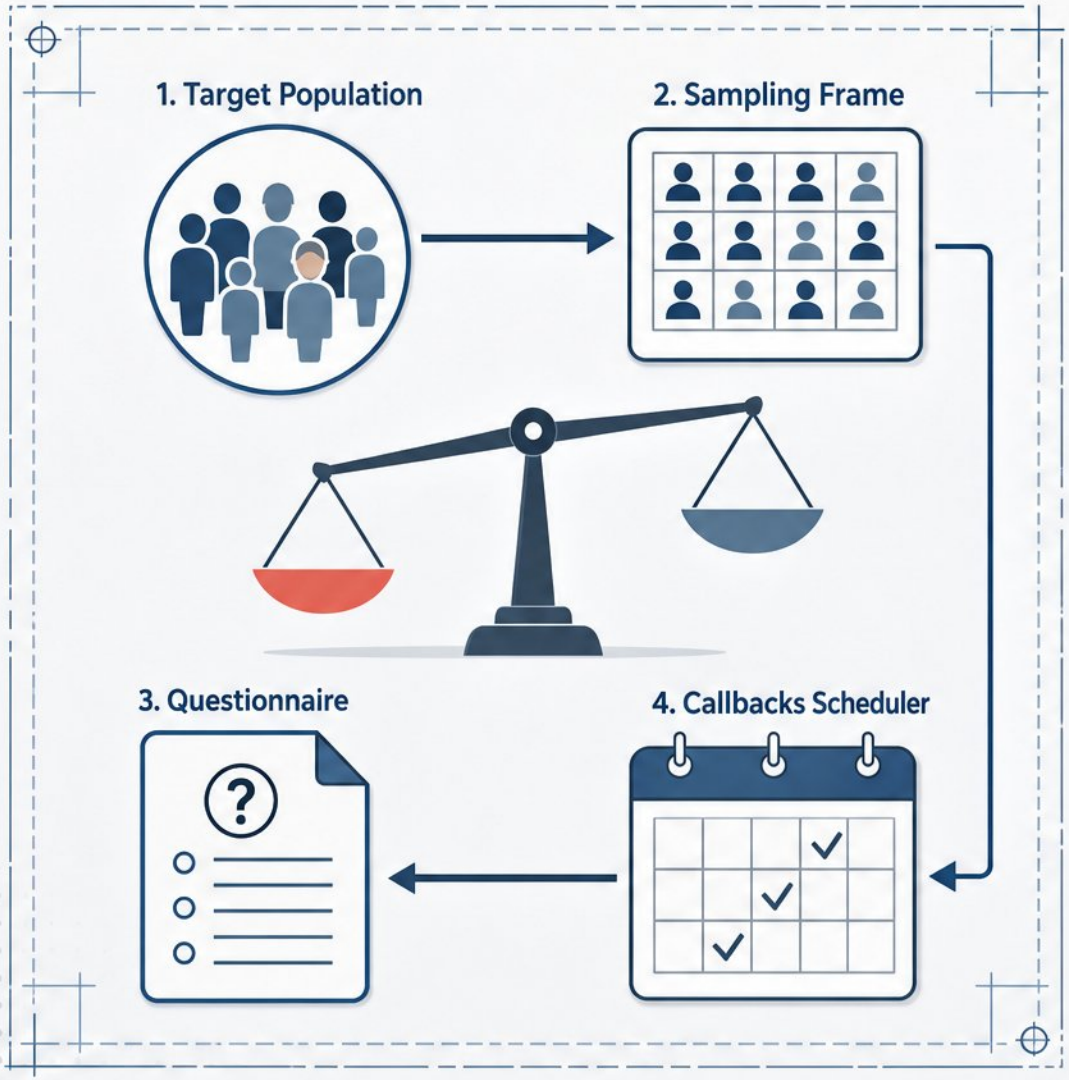
Case Walkthrough: The 3am Eval Set

- 3am sampling over-represents batch retries and APAC traffic
- Totally missed North American business-hours advisory queries
- New model looked better on eval but spiked support tickets
- Every sampling choice encodes geography, intent, and workload shape
- Cohort coverage should be a published, first-class eval statistic

Case Walkthrough: The Thumbs-Up Button That Poisoned the Eval Set

- Feedback-based eval dominated by thumbs-up prompts creates survivorship bias
- Casual queries score high; demanding enterprise prompts that drive churn vanish
- Eval, NPS, and churn cohorts produce three different distributions
- Every eval row needs a chain of custody: source pool and represented cohorts
- A signal can be real yet measure the wrong distribution when sampling is ungoverned

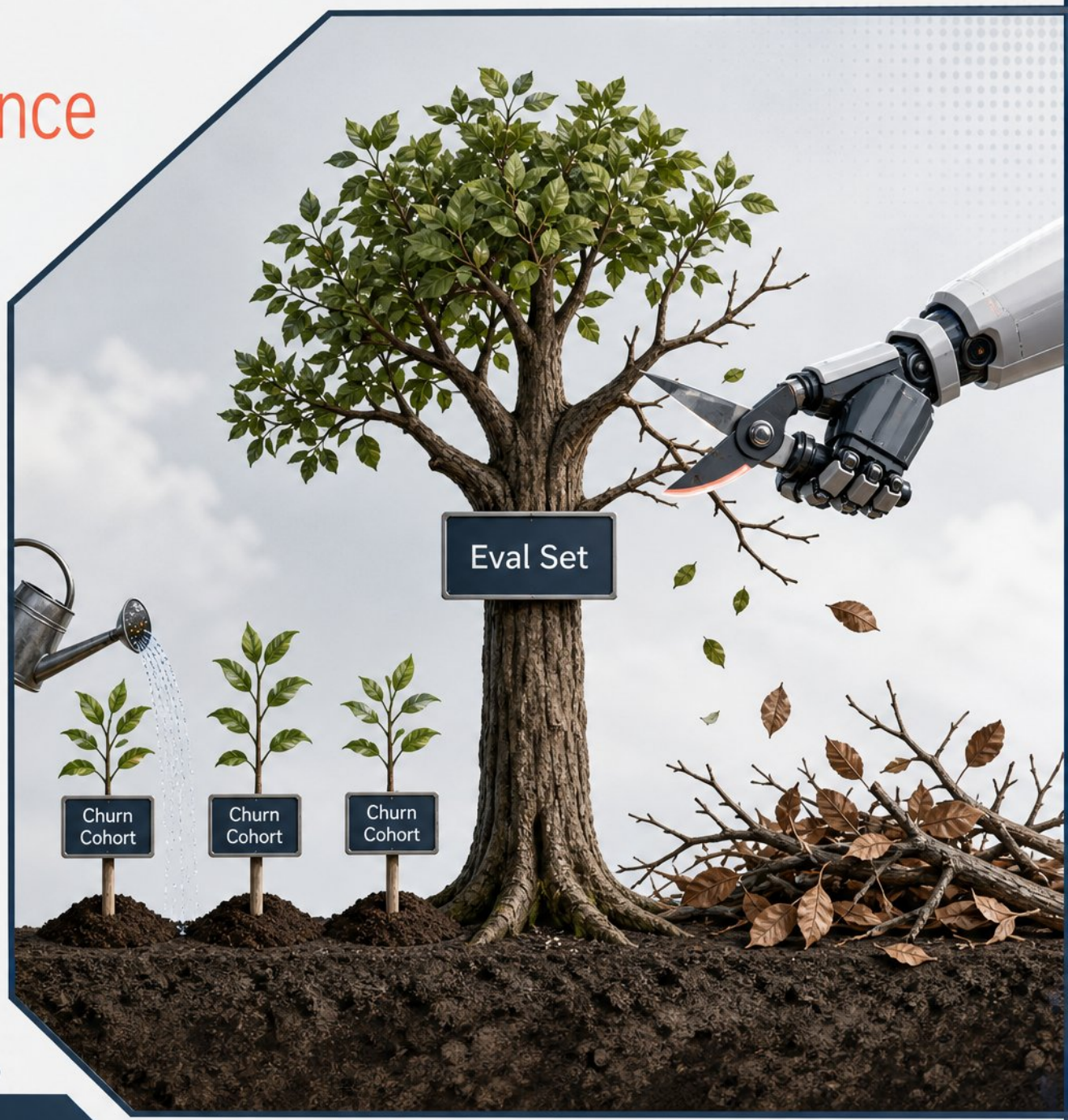
Reducing Bias During Study and Survey Design



- Define target population and build a frame covering all sub-groups
- Use probability sampling; acknowledge non-probability design limits
- Write neutral questions and pre-test via cognitive interviewing
- Monitor response rates by demographic in real time
- Plan for non-response: callbacks, weighting variables, pattern modeling

Governance and Continuous Vigilance

- Treat datasets as living artifacts that drift over time
- Quarterly reviews of cohort coverage and source mix
- Separate training signals from audited evaluation signals
- Apply time-based decay and recalibrate thresholds
- Track implicit signals (abandonment, edits) that buttons miss



Communication and Ethical Responsibility

- Be honest about what data can and cannot say
- Misrepresented sampling harms marginalized groups, biasing policy
- When citizens can't judge survey quality, trust in evidence erodes
- Use clear disclosure: "This reflects respondents' views, which may differ from the broader population because..."
- Rigorous transparency is the foundation of credible, ethical data work

PersonWise.

PersonWise • AI digital-human course platform

Scan the code or click anywhere to watch this course live, with voice Q&A

personwise.ai/t/bb3a2cc0/public