

Sampling: Ethics

```
$ echo "Data Science Institute"
```

Key Texts

- Peter Bacchetti, Leslie E. Wolf, Mark R. Segal, Charles E. McCulloch, Ethics and Sample Size, *American Journal of Epidemiology* , Volume 161, Issue 2, 15 January 2005, Pages 105–110, <https://doi.org/10.1093/aje/kwi014>
- Patrick Dattalo, Ethical Dilemmas in Sampling, *Journal of Social Work Values and Ethics* , Volume 7, Issue 1, 2010, <https://jswve.org/download/2010-1/2dattalo-Ethical-dilemmas-in-sampling.pdf>
- Bornstein, M. H., Jager, J., & Putnick, D. L. (2013). Sampling in Developmental Science: Situations, Shortcomings, Solutions, and Standards. *Developmental review: DR* , 33(4), 357–370. <https://doi.org/10.1016/j.dr.2013.08.003>

Respondent Burden

Respondent Burden

- **Respondent burden** refers to any risk, inconvenience, or discomfort respondents endure as a result of their participation in a study or survey
 - Low respondent burden: filling out a survey that takes 15 minutes of a participant's free time
 - High respondent burden: taking an experimental medication with potential negative side effects
- **Net burden** occurs when respondent burden outweighs the personal benefit received for a given respondent

Total participant burden = Net burden per participant · Number of participants

Impacts of High Respondent Burden on Data Quality

- Unit or item non-response
 - Respondent may avoid parts of a survey that they feel are most involved, or may avoid responding entirely
- Measurement error
 - Respondents providing rushed or inaccurate answers as a result of feeling overly burdened
- Sustainability
 - Inability to recruit new or repeat respondents once potential participants become aware of the burden involved

Study Value

- **Study value** refers to the projected societal or clinical benefits of the results of a given study
 - Low study value: a survey measuring university students' opinions on local coffee shops
 - High study value: a clinical trial of a new cancer treatment
- Value can also include benefit to individual participants in the study (i.e. compensation or insight/knowledge)
- Value is sometimes measured in terms of the probability of generating statistically significant results (also known as **statistical power**)

Respondent Burden and Sample Size

- If there is no net burden for each participant, ethical considerations do not need to constrain sample size
- If there is a net burden for each participant, the total burden increases as sample size increases
- However, a larger sample may increase the probability of achieving statistically significant or reliable results, which would increase the study value

A sample size could be considered ethically acceptable if the projected value of the study is greater than the total participant burden generated by the sample size



External Validity

Validity

- **Internal validity** refers to the validity of results *within* a particular study
 - Were all relevant variables measured? Were measurements accurate? Is the analysis appropriate for the study design?
- **External validity** describes the extent to which the results of study can be validly applied *outside* of the context of that specific study
 - Are the study participants representative of the target population? Do the conditions of the study mimic real-world conditions?
 - A study must be internally valid in order to be externally valid

Sampling and External Validity

- How can we make sure samples maximize external validity?
 - Reduce selection, coverage, and non-response bias
 - Ensure sample size is sufficiently large
 - Use **probability samples**
 - SRS, stratified samples, cluster samples, etc.
 - Each sampling unit has a known selection probability
 - Representative and minimally biased

Issues with Probability Sampling

- Resource intensive, expensive, and time consuming
- Sample size may be too small to be representative
- Incomplete randomization
 - Sampling units may be randomized, but still all drawn from one location or over a short time frame (i.e. time and place are not random, which impacts selection prob)
- Presence of other errors
 - Measurement error
 - Coverage error (incomplete sampling frame)
 - Low response rate or high non-response bias

Non-probability Sampling

- Becoming more common in social and scientific research
- Generally **easier** and **cheaper** than probability sampling
- **More difficult to analyze**, but more robust analysis and estimation techniques are being developed
- Examples
 - **Convenience sampling**: selecting individuals from accessible populations or allowing individuals to volunteer without explicit selection
 - **Deliberate or purposive sampling**: choosing individuals with specific characteristics (either for diversity or homogeneity)

External Validity and Ethical Concerns

- **Representation**
 - Convenience sampling: Who is the sample convenient for? Is the sample systematically biased towards some populations and away from others?
 - Deliberate sampling: How were individuals with the desired characteristics identified and selected?
- **Estimation and generalizations from non-representative samples**
 - Demographic variation and covariation may go undetected
 - Potential inaccurate results for population-level estimates
 - Potential negative outcomes on populations that were not represented in the sample

Working with Non-Representative Samples

- Consider impacts of bias
 - What are the potential consequences if inference from a non-representative sample is inaccurate?
- Restrict estimates to a subpopulation
 - If your sample only contains people under 30, do not generalize results to people of all age groups
- Weighting methods
 - Post-stratification
 - Raking
- Modelling
 - Multi-level models to extract information from underrepresented sub-populations

Next

Estimation and survey quality