

## Designing a Survey

Garrett Stillings ([Stillings.Garrett@epa.gov](mailto:Stillings.Garrett@epa.gov))

Monitoring and Analysis Branch-NARS

NARS National Workshop: April 5 & 6, 2022

- 1 Design Planning
- 2 Sample Frame Creation
- 3 Survey Design
- 4 Survey Sample Size
- 5 Survey Weights
- 6 References

## Section 1

### Design Planning

# Probability Sampling VS a Census

- **Census:** Examination of every unit in a population of interest.
- **Probability Sampling:** Allows for the extrapolation of information from a subset of the population (sample).

## Probability Sampling Benefits

- ① Less time consuming.
- ② Less costly.
- ③ Estimates can exhibit higher accuracy because staff can perform field operations more thoroughly, reducing measurement error.

# Define Survey Objectives

- Objectives should be linked to management decisions and reporting requirements.
- Ensure the level of effort will sufficiently answer the questions.
- Define the target population of interest.

## Target Population

- The target population defines the extent of units that will be monitored.
- Must define what elements make up the target population (i.e. Subpopulations).
- The target population should align with your organizations monitoring strategy and objectives.

# Target Population

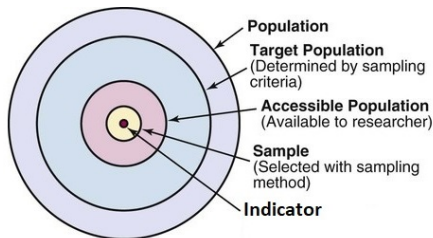


Figure 1: Survey Plan

## Target Population Example

If the designer only has an interest in assessing the condition of perennial waters in a state, the target population is defined as perennial waters and intermittent and ephemeral waters are defined as **Non-Target populations** and should be omitted from the selection process.

## Section 2

### Sample Frame Creation

# Sample Frame

- A **Sample Frame** is a GIS representation (e.g. shapefile) of the target population which is used to select the sample sites.
- The corresponding attribute file should contain covariates representing design features such as Strata and Categories.
- The CRS should be an area-preserving projection so that spatial distances are equivalent for all directions.

## Types of Sample Frames

- 1 **Discrete Objects:** Represented by point resources which serve as sampling units for selection (i.e. Lakes).
- 2 **Linear Features:** Represented by a linear resource in which a sampling unit is selected along the networks length (i.e. Streams and Rivers).
- 3 **Areas:** Represented by polygon resources where sampling units are selected within these regions (i.e. Wetlands and Coastlines).



## Section 3

### Survey Design

# Design Preperation

## Generalized Random Tessellation Stratified (GRTS) Design

- Spatial balance disperses sampling effort across the extent of the resource so that samples achieve a similar spatial distribution as the population (Olsen et al. 2012).
  - More representative of the population.
  - Give more precise estimates.

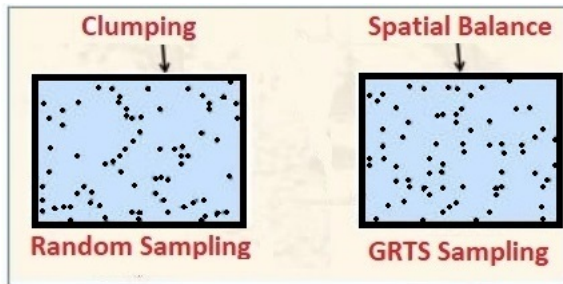


Figure 2: Spatial Balance

# Sample Selection

## Sampling Methods

- a. **Unstratified Sampling:** Random samples are drawn from the sample frame as a whole.
- b. **Stratified Sampling:** The sample frame is divided into non-overlapping **Strata**, or subpopulations, from which independent random samples are drawn.

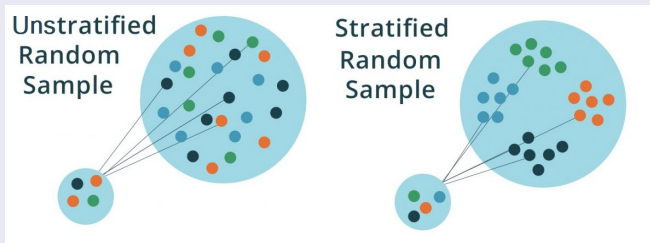


Figure 3: Stratified Sampling

# Sample Selection (cont.)

## Sampling Methods

- a. **Equal Probability Sampling:** Selection where all members of the sample frame have equal probabilities of being selected.
- b. **Unequal Probability Sampling:** Selection where the chance of being included is calculated relative to the proportion of a **Categorical Variable** across the population. Can give smaller populations a greater chance of being selected.

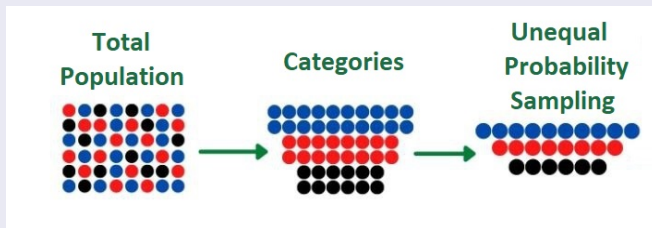


Figure 4: Unequal Sampling

# Sample Selection (cont.)

## Sampling Methods

- **Proportional Probability Sampling:** Selection where the chance of being included is proportional to the values of a positive **Auxiliary Variable** (Continuous variable).

Total Population	Auxiliary Variable	Proportional Sampling
21313121	1111111111	
32212312	1111111111	1111111111
12121213	22222222	22222222
21213121	22222222	333333
32312332	333333	
13121211	333333	

Figure 5: Proportional Sampling

# Sample Selection (cont.)

## Additional Sampling Methods Available

- ① **Legacy Sampling:** Sites that were selected in a previous sampling design and should be included in the current sample.
- ② **Minimum Distance Sampling:** Enforces a minimum distance between sites.
- ③ **Replacement Sampling:** Provides sites available to replace sites in the base sample for which data cannot be collected.
  - **Reverse Hierarchical Ordering (Default):** Sites are first selected using the GRTS algorithm then determined as base sites or replacement sites in a way that preserves as much spatial balance as possible.
  - **Nearest Neighbor:** Closest site measured by Euclidean distance to the base site.

## Section 4

### Survey Sample Size

# Determining the Sample Size

- Setting an appropriate sample size and considering how they should be allocated across a sample frame is a fundamental step in designing a successful survey.
- Many surveys are limited by budgetary and logistical constraints. The designer must determine a sample size which can overcome these constraints while ensuring the survey estimates the parameter(s) of interest with a low margin of error.



# Determining the Sample Size

## Things To Consider

- Compare the **Spatial Balance** of surveys. Typically, estimates from spatially balanced surveys are more precise (vary less) than estimates from non-spatially balanced surveys.
- Consider if the parameter(s) of interest will result in low variation across the survey. A smaller sample size can yield an estimate with a low margin of error in this case. Use the tools **Population Estimate Simulation** to test this.
- Allocate additional sampling time to survey extra sites if needed. When designing the survey, be sure to generate **Replacement Sites** to use for oversampling.

## Section 5

### Survey Weights

# Weights

- Allows for a dataset to be re-balanced so that results more accurately represent the population.
- Sample sites must be proportioned according to the target population.

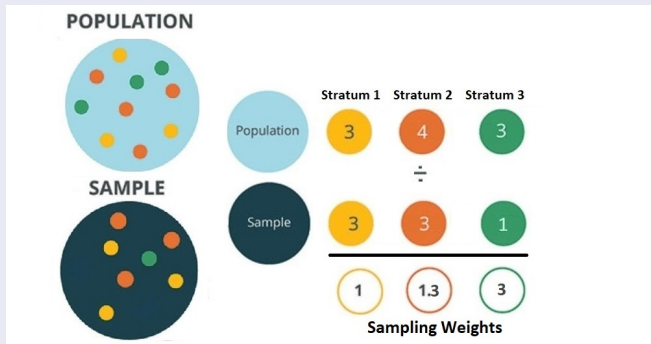


Figure 6: Sampling Weights

# Adjusting Design Weights

- Unfortunately, most sample frames are imperfect and will include non-target samples in the frame (**over coverage**) and include parts of the target population that cannot be sampled (e.g. access denials, barriers) (**under coverage**).
- To correct for this bias, the designer must adjust sampling weights to the new sample frame size prior to data analysis.

## When to Adjust Weights

- When **Non-Target** sites are found (remove from target population).
- Samples that are smaller or larger than planned (sites should be weighted more or less).

## Section 6

### References

# References

- [Survey Design Tool Link](#)
- [Survey Design Tool GitHub](#)

## References

- ① [Stevens and Olsen \(2004\) Spatially Balanced Sampling of Natural Resources](#)
- ② [Spatially Balanced Sampling Vignette](#)
- ③ [A GRTS User's Manual for the SDrawNPS Package](#)
- ④ [NARS Website](#)
- ⑤ Olsen, A., Kincaid, T., & Payton, Q. (2012). Spatially balanced survey designs for natural resources. In R. Gitzen, J. Millspaugh, A. Cooper, & D. Licht (Eds.), *Design and Analysis of Long-term Ecological Monitoring Studies* (pp. 126-150). Cambridge: Cambridge University Press.