

# Aquatic Resources Monitoring: Design, Analysis and Reporting Using EPA Tools

Garrett Stillings

Contributions from:  
Karen Blocksom and Michael Dumelle

2024 Southeastern Water Pollution Biologist Association

The views expressed in this presentation are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

## 1 Monitoring

## 2 Survey Design

## 3 Sampling

## 4 Adjusting Weights

## 5 Population Estimates

## 6 Communication

## 7 Resources

# A Quick Note

- The `spsurvey` R package, originally created in 2004, has been used extensively by NARS to carry out survey designs and analyze data
- The survey design and population estimates R Shiny apps discussed today are user-friendly, non-R interfaces that actually use `spsurvey` “under the hood”
- `spsurvey` had a major update released in October, 2021 (version 5.0)

## Section 1

Monitoring

# Clean Water Act Goals

- What is the overall quality of waters in the State?
  - **Section 305(b)**
- To what extent is water quality changing over time?
  - **Section 305(b), Section 314(a)(1)(F)**
- What are the problem areas and areas needing protection?
  - **Section 303(d)**
- What level of protection is needed?
  - **TMDLS, BPMs, Use attainability analyses, water quality standards**
- How effective are clean water projects and programs?
  - **Section 319, Section 314, Section 303(d), Section 402....**

# Survey Sampling Types

- ① **Census:** Sample every unit in the population.
- ② **Targeted:** Sample a selected area(s) of interest.
- ③ **Probability:** Randomly sample a subset of population and use the sample to draw inferences regarding the target population.

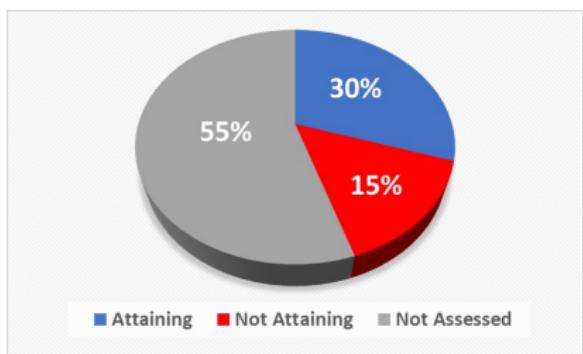


Figure 1: Targeted Monitoring

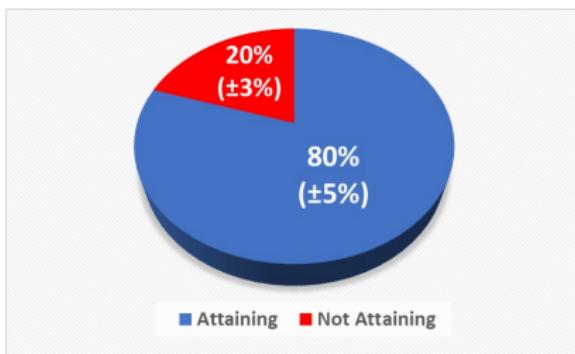


Figure 2: Probability Survey

# Generalized Random Tessellation Stratified (GRTS) Design

- Spatial balance disperses the 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.
  - More precise estimates of population parameters.

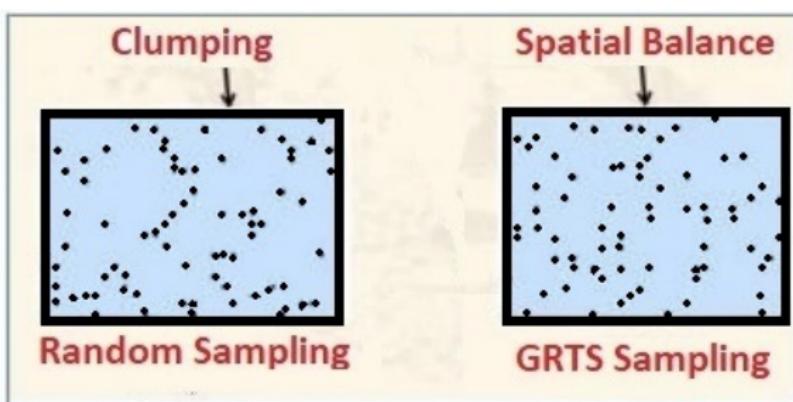


Figure 3: Spatial Balance

Monitoring  
oooo

Survey Design  
●ooooo

Sampling  
oooooooo

Adjusting Weights  
ooo

Population Estimates  
oooooo

Communication  
ooo

Resources  
oo

## Section 2

### Survey Design

# Survey Design Tool

The screenshot shows the Survey Design Tool interface. At the top, there's a navigation bar with the EPA logo, a search bar, and links for Environmental Topics, Laws & Regulations, Report a Violation, and About EPA. Below the navigation is a contact link labeled "CONTACT US". The main title "Survey Design Tool (v. 2.0.0)" is displayed, followed by a horizontal navigation bar with four tabs: Step 1: Instructions for Use (selected), Step 2: Prepare Survey Design, Step 3: Survey Design Results, and Step 4: Adjust Survey Weights.

**Step 1: Instructions for Use**

**Select the Survey Sample Frame**

Choose all files of the Sample Frame  
Required: (.shp, .dbf, .prj, .ahx)

Browse... 4 files Upload complete

Transform CRS to NAD83 / Conus Albers

**Design Attributes**

Choose Design Type  GRTS  IRS

Select Attribute Which Contains Strata None

Select Attribute Which Contains Categories None

Optional Design Attributes

**Calculate Survey Design**

**Stratum: None**

Stratum 1 None 50

Base Sites 10

Replacement Sites 10

Sample Frame Summary

**Sample Frame Summary**

GROUP	RESOURCE_units
None	776

Figure 4: <https://rconnect-public.epa.gov/survey-design-tool/>

# Define Survey Objectives

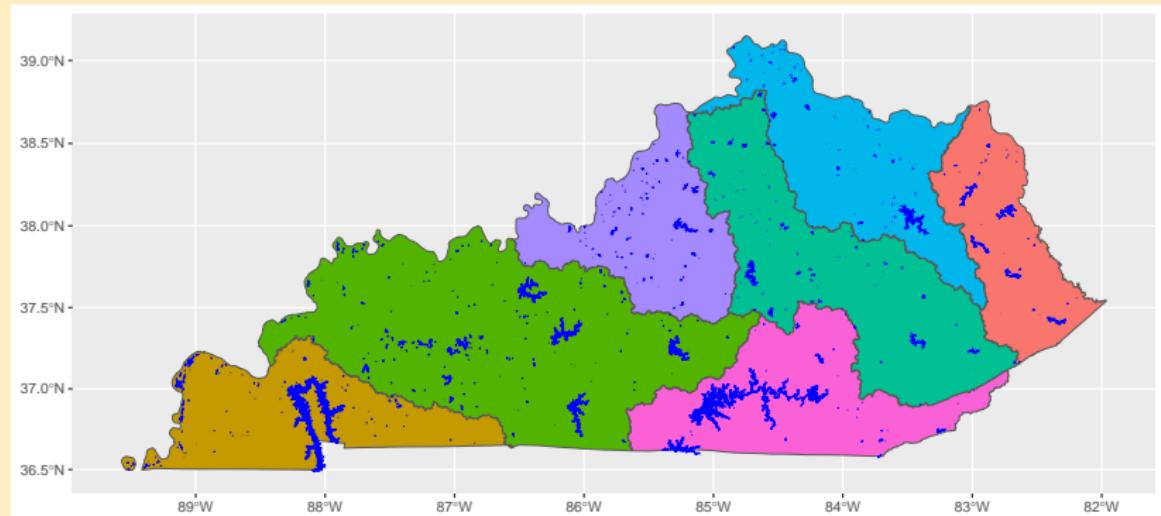
- Objectives should be linked to management decisions and reporting requirements.
- Ensure the level of effort will sufficiently answer the questions.
- Define the **target population** of interest.

## Target Population

- The target population defines the complete collection of units that will be monitored.
- Must define what elements make up the target population (e.g., Lakes >1m depth).
- The target population should align with your organization's assessment goals and monitoring strategy.

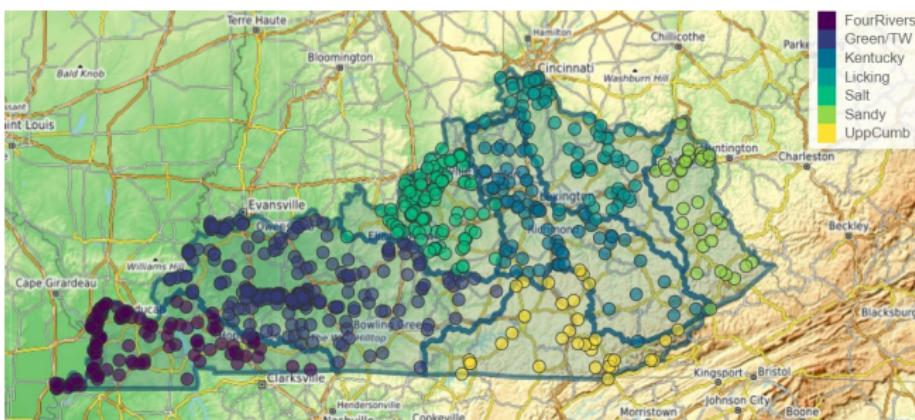
# Target Population

Example - In this example, we chose to target all lakes in Kentucky which are publicly accessible. This consists of 776 lakes across seven river basins.



# Sampling Frame

- A **sampling frame** is a list that contains all sites available to be sampled and intended to be in the target population.
- Sampling frames are often stored as GIS representations (e.g., ESRI shapefile) and contain additional information about the site that can be used in the design process:
  - Stratification variables (Stratified sampling)
  - Categorical/Continuous variables (Unequal or Proportional sampling)



# Sampling Frame Types

- ① **Discrete Objects:** Represented by **points** which serve as sampling units for selection (e.g. ponds and lakes treated as a whole).
- ② **Linear Features:** Represented by **polylines** in which a sampling unit is selected along the networks length (e.g. river and stream length).
- ③ **Areas:** Represented by **polygons** where sampling units are selected within these regions (e.g. lake, wetland and coastline area).

## Section 3

### Sampling

# Inclusion Probability Methods

- ➊ **Equal Probability:** Selection where all members of the sample frame have equal probabilities of being selected.
- ➋ **Unequal Probability:** Selection where the chance of being included is calculated relative to the proportion of a **Categorical Variable** (e.g., size class) across the population (user samples sizes are not guaranteed). Can give smaller populations a greater chance of being selected.



Figure 6: Equal Probability.



Figure 7: Unequal Probability.

# Sample Selection

- 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** (e.g., river basin) from which independent random samples are drawn.



Figure 8: Unstratified Sampling (Equal Probability).



Figure 9: Stratified Sampling (Equal Probability).

# Inclusion Probability Methods

- ③ **Proportional Probability Sampling:** Selection where the chance of being included is proportional to the values of a positive **Auxiliary Variable** (e.g., numeric size). Larger values of the auxiliary variable result in higher inclusion probabilities.



Figure 10: Proportional Sampling.

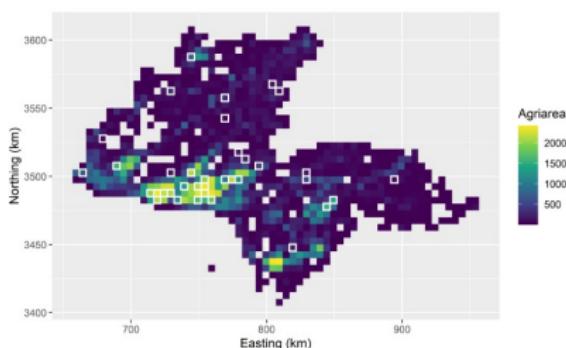


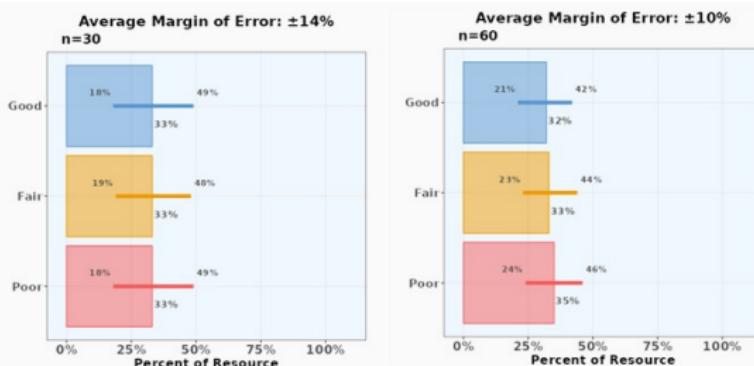
Figure 11: Image from: Dick Brus (2023). Spatial sampling with R.

# Additional Sampling Methods

- ① **Replacement Sampling:** Provides additional sites to replace sites in the base sample for which data cannot be collected.  
**Sites should be replaced with the next site in Site ID order within the same stratum.**
  - **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.
- ② **Legacy Sampling:** Sites that were selected in a previous sampling design and should be included in the current sample.
- ③ **Minimum Distance Sampling:** Enforces a minimum distance between sites.

# 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.



# Sample Size Considerations

- Compare the **Spatial Balance** of surveys as estimates from spatially balanced surveys are more precise.
- Set a reasonable sample size of replacement sites for increased spatial balance and lower variance estimates.
- 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. Use the Survey Design tools **Population Estimate Simulation** to test.



# Design Weights

- Design weights (i.e., sampling weights) are assigned to each site and **quantify the number of units** in the target population this site represents.
- Combined with the observed data to construct population estimates that represent the target population.

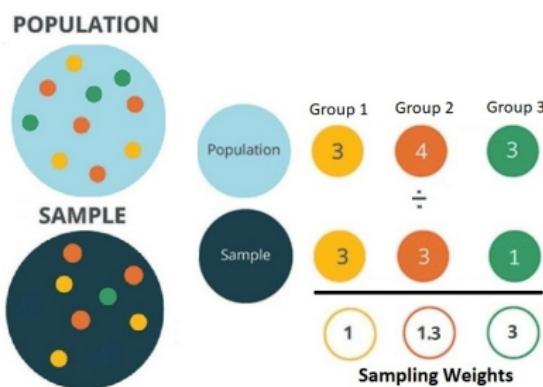


Figure 14: Design Weights.

## Section 4

### Adjusting Weights

# Site Evaluation

- Often the design does not go exactly as planned and the original design weights must be adjusted so that they continue to accurately represent the target population.
- A sampling frame has **overcoverage** if it contains some sampling units that are missing from the target population and/or **undercoverage** if it is missing some sampling units present in the target population.

## Site Evaluation Designations

- 1 Target Site** - A site which was in the target population and was sampled.
- 2 Non-Target Site** - A site which was **not** part of the target population and was **not** sampled.
- 3 Non-Response Site** - A site which was in the target population and was **not** sampled (inaccessible site, landowner denial)

# Design Weight Adjustments

- **Weight Category:** A weight adjustment category represents if a Stratum and/or a category was used in the design. Separate weight adjustments are applied to each initial weight so the final weights sum to a desired frame size.
- **Frame Size:** Size of the target population of the sampling frame. **Note** The Survey Design Tools Sample Frame Summary can assist.

## Survey Design Tool Weight Adjustment Outputs

- WGT\_TP\_EXT: Weights based on all target and non-target sites (design as implemented).
- WGT\_TP\_CORE: Weights based on the target population.  
**Use WGT\_TP\_CORE for population estimates**

## Section 5

### Population Estimates

# Population Estimate Tool

The screenshot shows the NARS Population Estimate Calculation Tool (v. 2.2.1) on the EPA website. The tool is designed for calculating population estimates from various data sources. Key features include:

- Input Options:** Allows users to input data from a URL or a local directory, or select a delimited file for analysis. A browse button points to "DataFor\_PopEstimates\_exam".
- Site Variable:** Set to "SITE\_ID".
- Weight Variable:** Set to "WEIGHT\_TP\_CORE".
- Response Variables:** A dropdown menu lists "AQUATIC\_LIFE\_USE PTL\_COND NTL\_COND".
- Analysis Options:** Includes checkboxes for performing change analysis by year or cycle, calculating overall estimates, and calculating estimates for subpopulations.
- Geographic Variables:** X coordinate variable is set to "XCOORD", and Y coordinate variable is set to "YCOORD".
- Stratification:** A dropdown menu for selecting a categorical stratum variable is set to "None".
- Display Options:** Set to "Head".
- Subset Data:** A checkbox for "Subset data using a single categorical variable" is checked.

Figure 15:

[https://rconnect-public.epa.gov/NARS\\_Population\\_Estimates\\_Tool/](https://rconnect-public.epa.gov/NARS_Population_Estimates_Tool/)

# Analyses available in app

- ① Categorical variable analysis (Proportions and Totals)
- ② Continuous variable analysis (CDFs, Percentiles)
- ③ Change analysis (Categorical and Continuous variables)

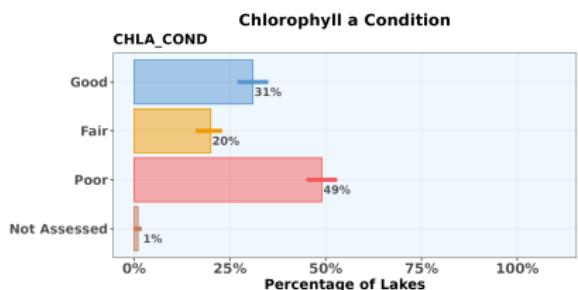


Figure 16: Proportion Estimates.

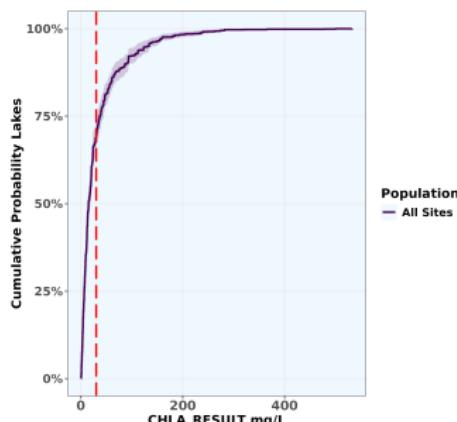


Figure 17: Cumulative Distribution Function.

# Analyses available using R code

## ④ Risk analysis

- Relative risk, attributable risk, risk difference

## ⑤ Trend analysis

- Categorical and continuous variables
- e.g., indicator trend over 3+ survey cycles

### U.S. EPA National Lakes Assessment 2022

#### Estimated Risk to Biota Associated with Stressors

In Relation to: Benthic Macroinvertebrates | National | All Lakes ( $\geq 1$  hectare)



Figure 18: Risk Analysis

# Considerations for Analysis

- ① All data must be in a single file
  - One row per site and year/cycle. No Visit 2's.
  - For Change Analysis, if legacy sites are included, site variable needs to be the same name.
- ② Weight Column
- ③ Response Variable Column
  - Categorical or Continuous variable(s)
- ④ Subpopulation(s) Column (optional)
  - Variable for different subsets of the population which estimates will be calculated for.
- ⑤ Variance Estimator
  - Local neighborhood variance is preferable and decreases variance estimates
- ⑥ X/Y Coordinates Columns
  - Should be an area-preserving projection CRS instead of geographic (XCOORD/YCOORD).

# Plotting Options

- Data can come from within the app or can upload a file of analysis results
- For categorical data, plot all categories for single subpopulation **OR** all subpopulations for single category
- For continuous data, only CDFs can be plotted
- For continuous data, confidence bounds and reference lines can be added
- Plots can be saved as .png files

## Section 6

Communication

# ATTAINS-Surveys Module

- Information about the population surveyed
  - Resource Type (Streams/Rivers, lakes, wetlands, estuaries)
  - Population (statewide, basins, individual waterbodies)
  - Extent and number of sites sampled
- Condition estimates for designated uses, conditions and stressors
  - Margin of error
  - Confidence level

The screenshot shows the ATTAINS Surveys Module interface. At the top, there is a navigation bar with icons for Home, Assessment Units, Assessments, Actions, Reports, EPA Reports, Priorities, and Surveys (which is highlighted with a red box). To the right of the navigation bar are two radio buttons: 'Use or Condition' (selected) and 'Stressor'. Below the navigation bar, the title 'State Statistical Survey - 2000' is displayed. A horizontal bar indicates the current step is 'Survey Water Type Groups'.

On the left, a sidebar titled 'Survey Water Groups' contains the following fields:

- Waterbody Type Group: STREAM/CREEK/RIVER
- Subpopulation: Statewide
- Target Population Size: 3506 Miles
- Number of Sites Sampled: 50

On the right, the survey parameters are listed:

- Use or Condition\*: RECREATION (selected)
- Category\*: Least Disturbed
- Statistic\*: Condition Estimate
- Metric Value\*: 50 %
- Margin of Error\*: ± 8
- Confidence Level\*: 95 %

A search bar at the bottom right contains the placeholder text 'Select Use(s) or Condition(s)'.

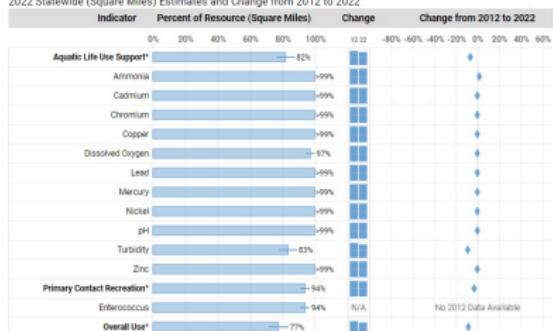
Figure 19: ATTAINS Surveys Module

# ARMADA

## Aquatic Resource Monitoring and Assessment Dashboard

### South Carolina | 2022 | Percent of Bay/Estuary Square Miles in Fully Supporting Category

2022 Statewide (Square Miles) Estimates and Change from 2012 to 2022



\*Represents that the Indicator is identified as a **Designated Use** or **Ecological Condition** by the State/Territory/Tribe.

### South Carolina | 2022 | Percent of Bay/Estuary Square Miles in Each Condition Category

Aquatic Life Use Support\* | 2022 Statewide (Square Miles) Estimates and Change from 2014 to 2022



\*Represents that the Indicator is identified as a **Designated Use** or **Ecological Condition** by the State/Territory/Tribe.

About the Data: No comments available from State/Territory/Tribe

About the Dashboard: This dashboard displays statistical survey results which provide an overall picture of water quality condition across a State/Territory/Tribe or subpopulation. From left to right, the graphs display the percentage of aquatic resources in different condition categories for the most recent survey year available for a population and a change comparison from the selected survey years, if available. Please note that the years shown are the years survey data was reported and not necessarily the collection year. Explore different resource types, subpopulations, condition categories and survey years by using the dropdowns on the right. Hover over a result to see more information and an expansion of the results. For national survey data and results, please visit [EPA's webpage for the National Aquatic Resource Survey](#).

Figure 20: <https://rconnect-public.epa.gov/armada/>

Monitoring  
oooo

Survey Design  
ooooooo

Sampling  
oooooooo

Adjusting Weights  
ooo

Population Estimates  
oooooo

Communication  
ooo

Resources  
●○

## Section 7

Resources

# Applications and Websites

- *spsurvey Website*
- *Survey Design Tool Link*
- *Survey Design Tool GitHub*
- *Population Estimate Shiny Tool Link*
- *Population Estimate Shiny Tool GitHub*
- *ARMADA*

## Additional Resources

- ① *spsurvey: Spatial Sampling Design and Analysis in R.*
- ② *Spatially Balanced Sampling of Natural Resources*
- 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.