Target Populations, Sampling Frames, and Site Evaluation

Presenter - Michael Dumelle (he/him/his; Dumelle.Michael@epa.gov)

Authors - Michael Dumelle and Tony Olsen

2022 National NARS Meeting: April 07, 2022

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.

Section 1

Introduction

NARS Survey Design and Analysis Steps

1. **Define target population**
2. **Develop sampling frame**
3. Select a sample (survey design)
4. **Evaluate sites**
5. Adjust weights
   - For design as implemented
   - For nonresponse
6. Estimate population parameters

Target Populations and Sampling Frames

- The target population (i.e., population of interest) is the population (i.e., set of sites) for which information is wanted and parameter estimates are required
- The sampling frame identifies the set of sites available to be selected by the survey design
- The sampling frame can:
  - Perfectly cover the target population
  - Overcover the target population
  - Undercover the target population
  - Overcover and undercover the target population

4

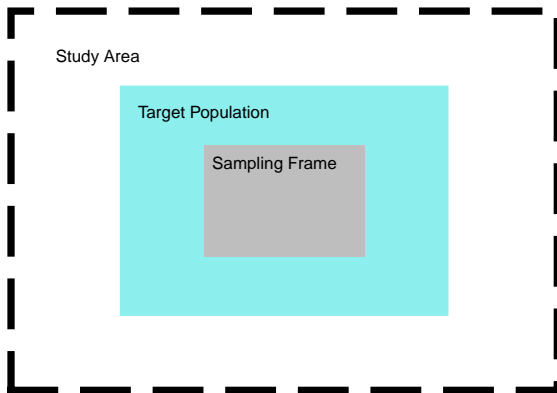## Target Population and Sampling Frames

Figure 1: Visual representation of study area, target population, and sampling frame for future use. The study area contains all area within the black, dashed line. The target population contains all area within the blue rectangle. The sampling frame contains all area shaded gray.

Perfect Coverage

- Perfect alignment of target population and sampling frame
- Target Population: Lakes with area greater than 10 hectares
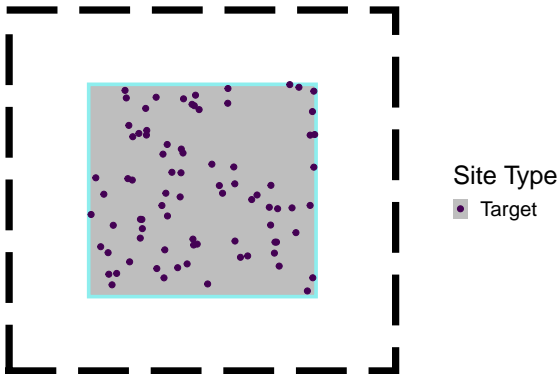- Sampling Frame: Lakes with area greater than 10 hectares

Perfect Coverage

Site Type
● Target

Figure 2: Perfect coverage. The target population and sampling frame are perfectly aligned.

Overcoverage

- The sampling frame contains sites outside of the target population
  - This implies that some sites evaluated are not target
- Target Population: Lakes with area greater than 10 hectares
- Sampling Frame: Lakes with area greater than 1 hectare
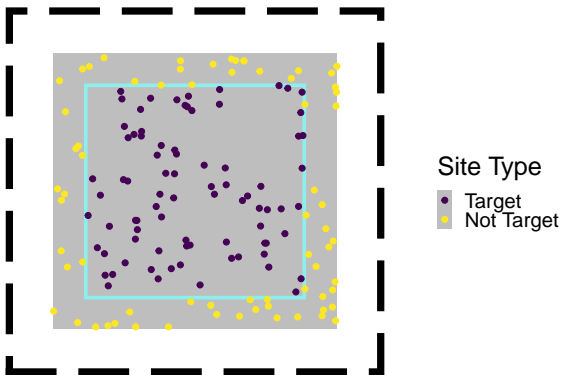- Overcoverage: Lakes with area between 1 and 10 hectares

Overcoverage

Figure 3: Overcoverage. The sampling frame contains sites outside of the target population.

Overcoverage

- Use oversample sites to reach desired target population sample size
- Critical to record site evaluation
    - Are sites target or not target? What made the site not target?
    - Are target sites sampled or not? What made the site unable to be sampled?
- The cost of overcoverage is consumption of finite resources

Undercoverage

- The sampling frame is a subset of the target population
- Target Population: Lakes with area greater than 10 hectares
- Sampling Frame: Lakes with area greater than 15 hectares
- Undercoverage: Lakes with area between 10 and 15 hectares
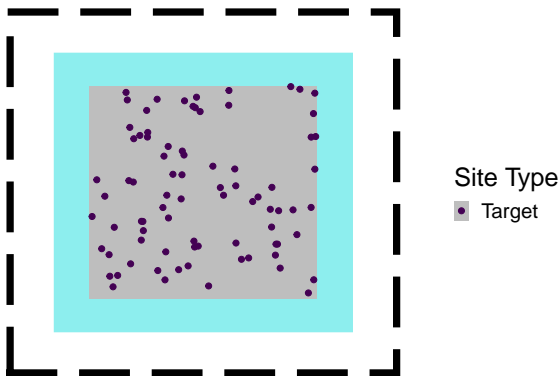
Undercoverage

Figure 4: Undercoverage. The sampling frame is a subset of the target population.

Undercoverage

- The sample gives us no information about sites outside the sampling frame but inside the target population
- To make statements about the entire target population, it must be assumed that the sites outside the sampling frame are missing completely at random
    - A missing completely at random assumption implies the cause of the missingness is unrelated to the response
    - There is no characteristic of the undercovered sites that makes the indicators behave fundamentally differently than covered sites
- The cost of undercoverage is assumptions

Overcoverage and Undercoverage

- Part of the sampling frame contains sites outside of the target population (overcoverage) and part of the sampling frame is a subset of the target population (undercoverage)
- Target Population: Lakes with area between 10 and 20 hectares
- Sampling Frame: Lakes with area between 5 and 15 hectares
- Overcoverage: Lakes with area between 5 and 10 hectares
- Undercoverage: Lakes with area between 15 and 20 hectares
- Overcoverage adjustments and undercoverage assumptions must simultaneously occur in order to make statements about the target population
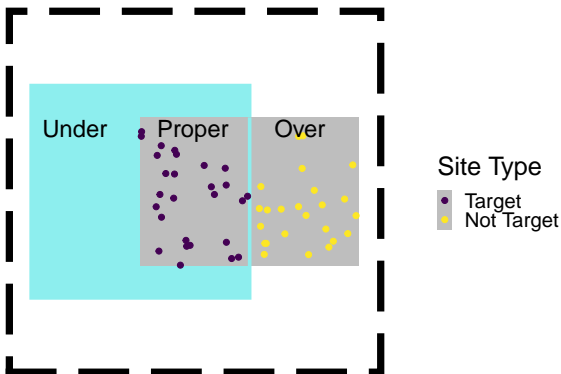
14

Overcoverage and Undercoverage

Figure 5: Overcoverage and undercoverage. "Under" represents undercovered area. "Proper" indicates properly covered area. "Over" represents overcovered area.

Non-response

- Non-response for an indicator occurs when a target population site has no data regarding that indicator
- Non-response is dominated by lack of landowner permission or lack of access (e.g., the site is physically inaccessible)
- Other sources of non-response: equipment failure, accidental lack of recording, etc.
- Critical to record whether sites were sampled and if they were not sampled, the reason why

Analysis

- Sites have a response value (e.g., NITR_COND) and a design weight (e.g., WGT_DSGN)
- Design weight indicates how many units the site represents in the sampling frame
- Adjust weights for the design as implemented
  - Sample size is the number of sites evaluated
- Adjust weights for non-response (if non-response is present)
- Make assumptions regarding undercoverage

Analysis

- Combine response values and adjusted design weights to estimate population parameters for target population, sampled population, and subpopulations
- Knowing the design is important in analysis
  - For example, knowing whether the design is stratified or not can impact variance estimation!

Section 2

NRSA

Target Population Definition (NRSA)

- All streams and rivers with flowing water during the summer index period above head-of-salt

  - Excludes run-of-river reservoirs with greater than seven-day residence time

- The target population definition is the same for all 2008-2009, 2013-2014, 2018-2019, and 2023-2024 design cycles

- Partner intensification designs have restricted the target population

  - Wadeable streams
  - Subregions of a state
  - Non-headwater streams (Strahler order greater than one)

Sampling Frame (NRSA)

- Source: National Hydrography Dataset
  - 2008-2009 and 2013-2014: NHDPlus Version 1 (1:100K resolution)
  - 2018-2019: NHDPlus Version 2 (1:100K resolution)
  - 2023-2024: NHDPlus Hi-Res (at least 1:24K resolution)
- Sampling frame: 2008-2009, 2013-2014, and 2018-2019
  - Changed every design cycle based on FCODE
  - Relied on "perennial" and "intermittent" coding, which sometimes was not a reliable indicator of streams with flowing water
  - Used Strahler order to define small streams, large streams, and rivers
- Sampling frame: 2023-2024
  - Uses FCODE to exclude reaches that are not expected to be in the target population
  - Uses mean annual flow to define small streams (1 - 5 cfs), large streams (5 - 100 cfs), and rivers (greater than 100 cfs)

Introduction
○○○○○○○○○○○○○○○○○○○○

NRSA
○○○

NLA
●○○

NWCA
○○○

NCCA
○○○

NGLA
○○○

NARS Coverage
○○○○

Conclusions
○○○

Section 3

NLA

Target Population Definition (NLA)

- All natural and man-made freshwater lakes, ponds, and reservoirs greater than one hectare (approximately 2.5 acres) in the conterminous U.S., excluding the Great Lakes

- The target population definition is the same for the 2012, 2017, and 2022 design cycles

- The target population for 2007 was lakes greater than four hectares

- Partner intensification designs have restricted definition to

  - Different lake area minimums
  - Subregion of a state
  - Public lakes

Sampling Frame (NLA)

- Source: National Hydrography Dataset
  - 2007 and 2012: NHDPlus Version 1 (1:100K resolution)
  - 2017: NHDPlus Version 2 (1:100K resolution) supplemented with NHDPlus Hi-Res (at least 1:24K) for lakes less than 5 hectares
  - 2022: NHDPlus Hi-Res (at least 1:24K resolution)

Section 4

NWCA

Target Population Definition (NWCA)

- All tidal and nontidal wetlands of the conterminous US, including certain farmed wetlands not currently in crop production. The wetlands have rooted vegetation and (when present) open water less than one meter deep.

- The target population definition is the same for the 2011, 2016, and 2021 design cycles

- Partner intensification designs have restricted definition to

    - Subregions of a state

Sampling Frame (NWCA)

- Source:
  - 2011: Fish and Wildlife Service (FWS) Status and Trend 4 square mile plots ($\sim$ 5000 plots)
  - 2016: FWS Status and Trend plots supplemented with the National Wetland Inventory (NWI) digitized map of wetlands (from a sample of 4 square mile FWS plots)
  - 2021: NWI digitized map of wetlands
- Sampling frames included a subset of wetland polygons within their respective source expected to meet the target population definition

Section 5

NCCA

Target Population Definition (NCCA)

- All coastal waters of the conterminous United states from the head-of-salt to confluence with the ocean
  - Includes inland waterways, tidal rivers and creeks, lagoons, fjords, bays, and major embayments
  - Seaward boundary extends to the point at which a straight line intersecting two land features fully encloses a body of coastal water
  - Salinities at least as high as 0.5 ppt are defined as estuarine (regardless of depth)
- The target population definition is the same for the 2010, 2015, and 2020 design cycles

Sampling Frame (NCCA)

- Source:
    - 2010: Developed by EMAP and ORD/GED
    - 2015: Updated 2010 sampling frame with information from NCA 1998-2001, NCA 2004, and NCA 2005-2006 sampling frames
    - 2020: Minor update of 2015 sampling frame for small and large estuaries
    - 2025: New sampling frame under construction based on NOAA continuously updated shoreline and an outer boundary matching the NCCA outer boundary

Section 6

NGLA

Target Population Definition (NGLA)

- All shallow water nearshore zones of Lake Superior, Lake Michigan, Lake Huron, Lake Erie, and Lake Ontario (Great Lakes) within the United States
  - Shallow water is defined as within 5km of shoreline and up to 30m depth, whichever is reached first (does not include connecting channels between the Great Lakes and St. Laurence River)

- The target population definition changes based on target embayments as 2010, 2015, and 2020 target populations are all different

- Intensification designs have further studied:
  - Lake Erie, Illinois Lake Michigan near shore, connecting channel studies, National Parks Service studies, island studies, Green Bay study

Sampling Frame (NGLA)

- Source:
  - 2010: Developed by Duluth Ecology Lab
  - 2015: Updated 2010 sampling frame using a new shoreline to define nearshore and using a new embayment criteria
  - 2020: New sampling frame developed by Duluth Ecology Lab. In the new sampling frame, the nearshore uniquely "coastal" land-water interface zone includes open and semi-enclosed bays and embayments with greater than a 200 m wide connection to open water and more open waters adjacent to shorelines (5 km from shoreline and less than 30m depth).
  - 2025: No decision yet

Section 7

NARS Coverage
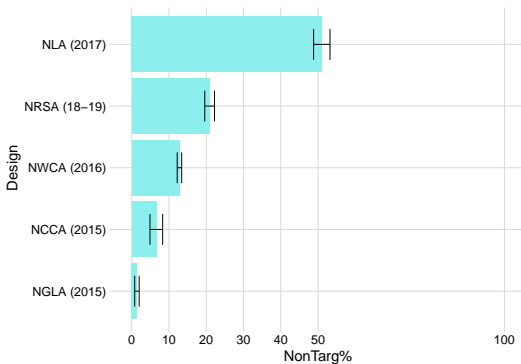
NARS Overcoverage

Figure 6: Estimated non-target rates in each survey's most recent design cycle alongside error bars.

- NRSA, NLA, and NWCA have high rates of overcoverage; NCCA and NGLA do not
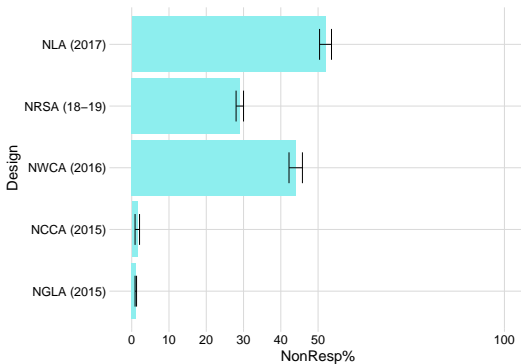
NARS Non-response

Figure 7: Estimated non-response rates in each survey's most recent design cycle alongside error bars.

- NRSA, NLA, and NWCA have high rates of non-response; NCCA and NGLA do not

Topics for Discussion

- Can the sampling frame be improved? In this improvement regional? Is it national? Is it both? Something else?
- Can non-response be reduced?
- Most effective ways to integrate NARS and partner surveys?

Section 8

Conclusions

Conclusions

- The target population (i.e., population of interest) is the population (i.e., set of sites) for which information is wanted and parameter estimates are required
- The sampling frame identifies the set of sites available to be selected by the survey design
- Incorporating overcoverage, undercoverage, and non-response is important when analyzing data to make population estimates
- NRSA, NLA, NWCA have high rates of overcoverage and non-response
- NCCA and NGLA have low rates of overcoverage and non-response

Acknowledgements

- Thanks to Tony Olsen, Steven Paulsen, and Karen Blocksom for helping with the presentation
- Thanks to Elizabeth Smith for facilitating
- Thanks to CTIC, Sarah Lehmann, and the rest of the organizing committee for helping to host this wonderful workshop
- Thanks to you for attending!
- Questions?