# Workshop Topics: Weights

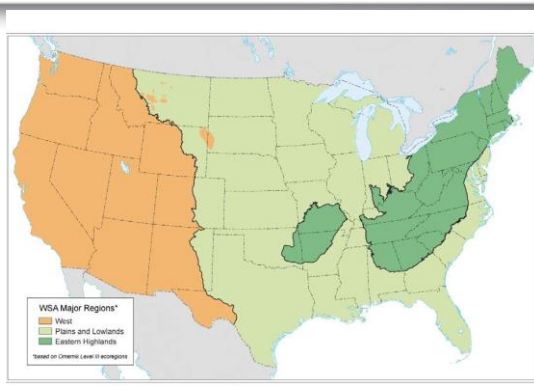
- What do people reading reports based on survey designs need to know about weights?
- How are weights used when survey data is used to produce results for monitoring reports, journal articles or special studies?
- How are weights calculated and what influences how they are calculated?
- Is the weight calculation process the same for NRSA, NLA, NCCA and NWCA?

# Hypothetical News Release

**A study by the US Environmental Protection Agency found that 60% of the lakes are in good biological condition. These results are based on a robust multi-metric index (MMI) that combines several measures relevant to the condition of aquatic benthic macroinvertebrates (e.g., aquatic insects and other creatures such as crayfish). The benthic macroinvertebrates are the most common biological indicator of water quality used by states. In collaboration with states, 100 lakes were sampled in each of three large geographic regions: western; plains & lowlands; and eastern highlands.**

# Are Study Results Correct? Maybe

- What we know
  - Sampled 300 lakes with 100 lakes in each of the three regions
  - 180 lakes were in good condition
- What we need to know
  - # lakes by region: 1000 lakes in West; 12,000 lakes in Plains and Lowlands; 2000 lakes in Eastern Highlands
  - # sample lakes in good condition by region




- Results Case A: 60 of 100 sample lakes in good condition in each region

$$\frac{180}{300} = \frac{1000*60/100 + 12000*60/100 + 2000*60/100}{15000} = \frac{\frac{1000}{100}*60 + \frac{12000}{100}*60 + \frac{2000}{100}*60}{15000} = 0.60$$

- Results Case B: 80 West, 40 Plains, 60 Eastern lakes in good condition
  - Unweighted estimate 60% in good condition: 180/300 = 0.60 WRONG
  - Weighted Estimate: 45% in good condition
  - $\frac{1000*80/100 + 12000*40/100 + 2000*60/100}{15000} = \frac{\frac{1000}{100}*80 + \frac{12000}{100}*40 + \frac{2000}{100}*60}{15000} = 0.45$
  - Weights: West 1000/100=10; Plains 12000/100=120; Highlands 2000/100=20 so that each lake in sample represents 10, 120 or 20 lakes depending on region
  - Weighted estimate calculated as (10*80 + 120*40 + 20*60)/15000 = .45

# Weights & Inference

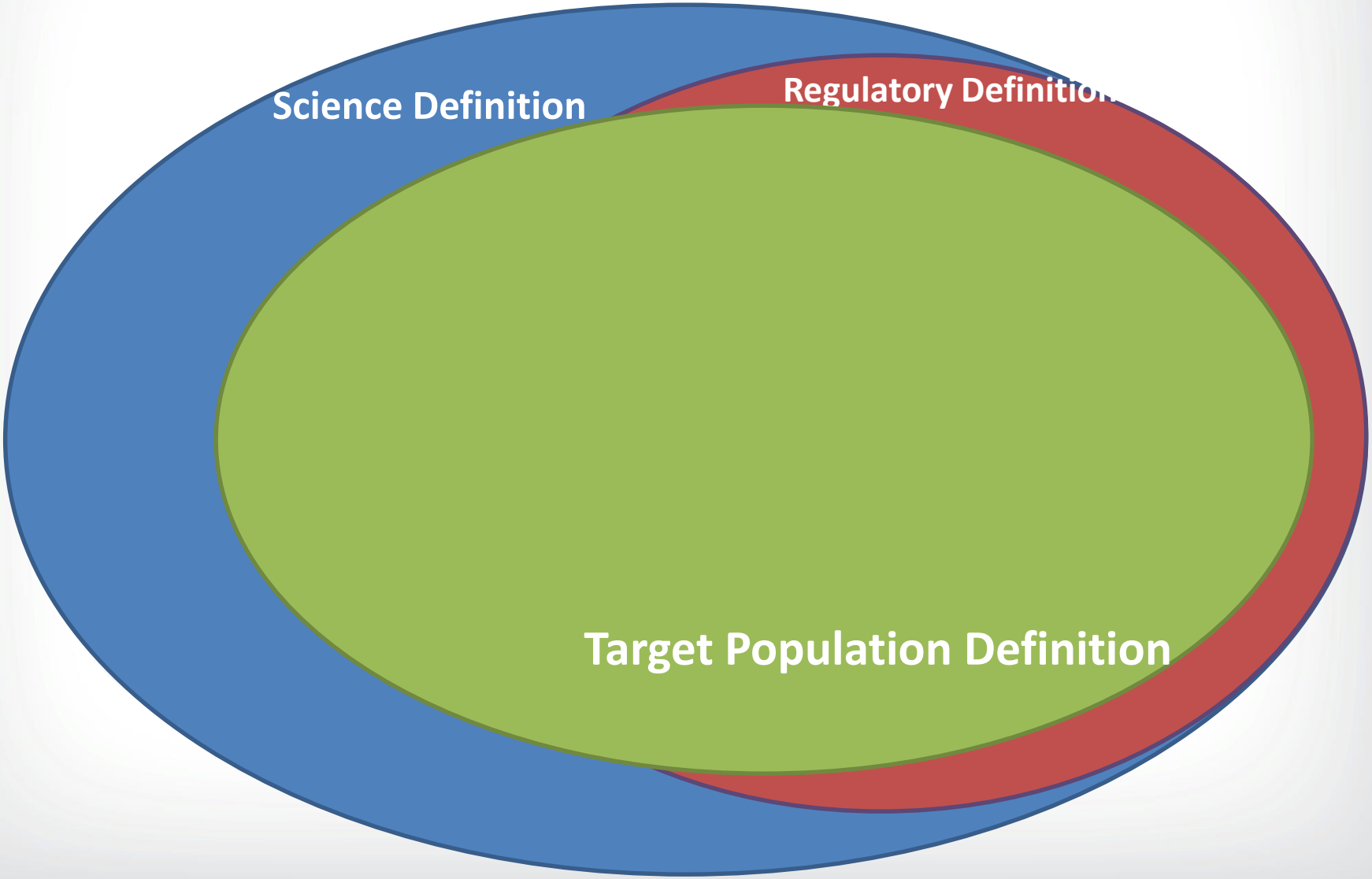
**Use of weights in survey analyses:**

**A scientifically defensible way to generalize from the sites sampled to the entire aquatic resource**

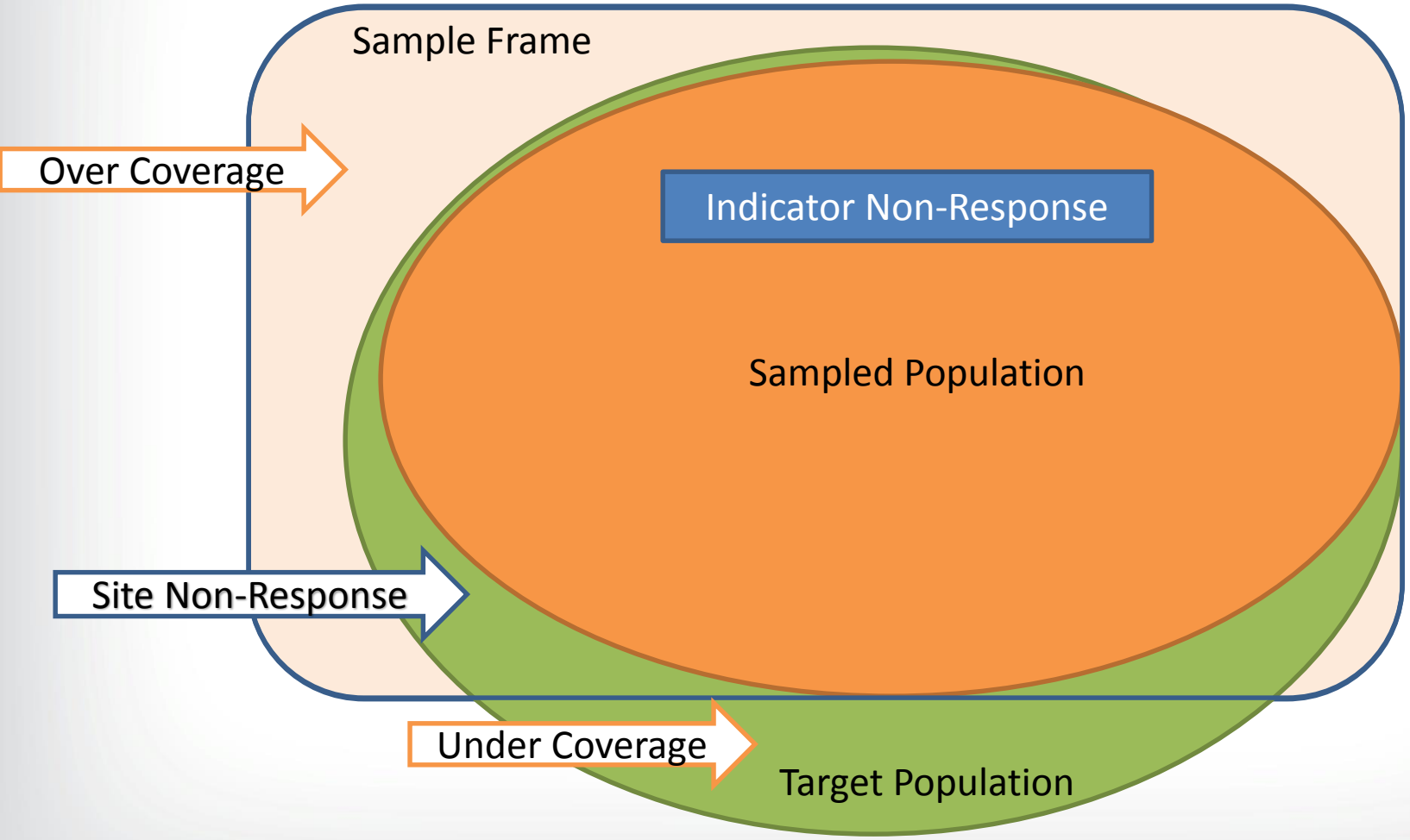
# When & How to use Weights

- Extent of aquatic resource estimation
  - National assessments
  - State assessments
- Condition of aquatic resource estimation
  - National assessments
  - State assessments
- Change between two time periods estimation
- Use in other statistical analyses

# Aquatic Resource Populations



# Survey Design & Sample Frame



# NWCA 2011 Extent Estimates

- **Purpose: Illustrate use of weights and site evaluation information from the NWCA 2011 survey to understand**
  - **relationship between sample frame and target population (TNT extent estimates)**
  - **relationship between target population and sampled population (Target extent estimates)**
  - **reasons why the sample frame includes elements that are non-target (NonTarget extent estimates)**
- **All data are in single csv file: NWCAdata_forPopEst_RelRisk.csv**

# What Information is Required?

- All site evaluation information for all sites evaluated
- Sites required
  - Include or exclude site
- Design information for each site
  - Location: projected coordinates (e.g. X_Albers, Y_Albers)
  - Stratum
  - Weight
- Subpopulations for which estimate is required
  - Subpopulation type #1
  - Subpopulation type #2
  - etc
- Data
  - Indicator #1
  - Indicator #2
  - etc

# R spsurvey code: Extent

```
# Sites to use in the analysis
sites_TNT <- data.frame(siteID=NWCA11$SITE_ID, Use=rep(TRUE, nr))
# Subpopulations requested
subpop <- data.frame(siteID=NWCA11$SITE_ID,
                     National=rep("National", nr),
                     SandT_Class=NWCA11$CLASS_FIELD_FWSST,
                     NWCA_Wetland_Group=NWCA11$NWCA_WET_GRP,
                     NWCA_Ecoregion=NWCA11$NWCA_ECO4,
                     NWCA_Reporting_Units=NWCA11$ECO_X_WETGRP)
# Design information
dsgn <- data.frame(siteID=NWCA11$SITE_ID,
                     wgt=NWCA11$WGT_11,
                     xcoord=NWCA11$XCOORD,
                     ycoord=NWCA11$YCOORD)
# site evaluation data
data.cat_TNT <- data.frame(siteID=NWCA11$SITE_ID,
                                 Target_NonTarget=NWCA11$TNT)
# calculate estimates
TNT_Extent_Estimates <- cat.analysis(sites_TNT, subpop, dsgn, data.cat_TNT)
```

# Target-NonTarget Estimates

| Type | Subpopulation | Indicator | Category | NResp | Estimate.P | StdError.P | LCB95Pct.P | UCB95Pct.P | Estimate.U | StdError.U | LCB95Pct.U | UCB95Pct.U |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| National | National | Target_NonTarget | Target | 1644 | 82.4 | 1.1 | 80.3 | 84.5 | 94,913,706 | 3,162,376 | 88,715,564 | 101,111,848 |
| National | National | Target_NonTarget | NonTarget | 669 | 17.6 | 1.1 | 15.5 | 19.7 | 20,268,466 | 1,219,212 | 17,878,855 | 22,658,078 |
| National | National | Target_NonTarget | Total | 2313 | 100.0 | 0.0 | 100.0 | 100.0 | 115,182,172 | 3,141,865 | 109,024,230 | 121,340,114 |
| SandT_Class | E2EM | Target_NonTarget | Target | 308 | 98.1 | 0.6 | 97.0 | 99.2 | 5,600,398 | 529,939 | 4,561,738 | 6,639,059 |
| SandT_Class | E2EM | Target_NonTarget | NonTarget | 16 | 1.9 | 0.6 | 0.8 | 3.0 | 106,842 | 30,303 | 47,448 | 166,235 |
| SandT_Class | E2EM | Target_NonTarget | Total | 324 | 100.0 | 0.0 | 100.0 | 100.0 | 5,707,240 | 530,120 | 4,668,225 | 6,746,255 |
| SandT_Class | E2SS | Target_NonTarget | Target | 159 | 99.8 | 0.1 | 99.7 | 100.0 | 989,572 | 153,459 | 688,798 | 1,290,346 |
| SandT_Class | E2SS | Target_NonTarget | NonTarget | 5 | 0.2 | 0.1 | 0.0 | 0.3 | 1,578 | 768 | 73 | 3,084 |
| SandT_Class | E2SS | Target_NonTarget | Total | 164 | 100.0 | 0.0 | 100.0 | 100.0 | 991,150 | 153,457 | 690,380 | 1,291,921 |
| SandT_Class | PEM | Target_NonTarget | Target | 432 | 88.2 | 1.5 | 85.3 | 91.1 | 21,151,755 | 842,345 | 19,500,789 | 22,802,722 |
| SandT_Class | PEM | Target_NonTarget | NonTarget | 87 | 11.8 | 1.5 | 8.9 | 14.7 | 2,834,178 | 364,641 | 2,119,495 | 3,548,862 |
| SandT_Class | PEM | Target_NonTarget | Total | 519 | 100.0 | 0.0 | 100.0 | 100.0 | 23,985,934 | 842,006 | 22,335,632 | 25,636,236 |
| SandT_Class | PF | Target_NonTarget | Target | 87 | 19.5 | 4.6 | 10.6 | 28.4 | 2,000,354 | 527,022 | 967,410 | 3,033,299 |
| SandT_Class | PF | Target_NonTarget | NonTarget | 225 | 80.5 | 4.6 | 71.6 | 89.4 | 8,268,651 | 542,050 | 7,206,253 | 9,331,050 |
| SandT_Class | PF | Target_NonTarget | Total | 312 | 100.0 | 0.0 | 100.0 | 100.0 | 10,269,006 | 629,207 | 9,035,783 | 11,502,229 |
| SandT_Class | PFO | Target_NonTarget | Target | 325 | 92.0 | 1.2 | 89.7 | 94.4 | 46,473,841 | 2,241,845 | 42,079,906 | 50,867,775 |
| SandT_Class | PFO | Target_NonTarget | NonTarget | 55 | 8.0 | 1.2 | 5.6 | 10.3 | 4,032,538 | 576,466 | 2,902,686 | 5,162,390 |
| SandT_Class | PFO | Target_NonTarget | Total | 380 | 100.0 | 0.0 | 100.0 | 100.0 | 50,506,379 | 2,167,257 | 46,258,632 | 54,754,125 |
| SandT_Class | PSS | Target_NonTarget | Target | 218 | 91.9 | 1.3 | 89.4 | 94.4 | 16,130,361 | 1,238,947 | 13,702,069 | 18,558,652 |
| SandT_Class | PSS | Target_NonTarget | NonTarget | 56 | 8.1 | 1.3 | 5.6 | 10.6 | 1,419,638 | 201,902 | 1,023,917 | 1,815,358 |
| SandT_Class | PSS | Target_NonTarget | Total | 274 | 100.0 | 0.0 | 100.0 | 100.0 | 17,549,998 | 1,223,994 | 15,151,014 | 19,948,982 |
| SandT_Class | PUBPAB | Target_NonTarget | Target | 115 | 41.6 | 6.9 | 28.1 | 55.1 | 2,567,424 | 620,289 | 1,351,680 | 3,783,169 |
| SandT_Class | PUBPAB | Target_NonTarget | NonTarget | 225 | 58.4 | 6.9 | 44.9 | 71.9 | 3,605,041 | 449,129 | 2,724,763 | 4,485,318 |
| SandT_Class | PUBPAB | Target_NonTarget | Total | 340 | 100.0 | 0.0 | 100.0 | 100.0 | 6,172,465 | 722,440 | 4,756,509 | 7,588,421 |

# Target Extent: Sampled Population

| Type | Subpopulation | Indicator | Category | NResp | Estimate.P | StdError.P | LCB95Pct.P | UCB95Pct.P | Estimate.U | StdError.U | LCB95Pct.U | UCB95Pct.U |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| National | National | Evaluation_Status | Target_AA | 87 | 2.3 | 0.5 | 1.3 | 3.2 | 2,156,058 | 469,270 | 1,236,306 | 3,075,809 |
| National | National | Evaluation_Status | Target_Access_Denied | 429 | 24.7 | 1.8 | 21.2 | 28.2 | 23,462,233 | 1,895,029 | 19,748,044 | 27,176,422 |
| National | National | Evaluation_Status | Target_Inaccessible | 126 | 6.8 | 1.0 | 4.7 | 8.9 | 6,448,142 | 1,005,751 | 4,476,906 | 8,419,378 |
| National | National | Evaluation_Status | Target_Other | 35 | 0.7 | 0.2 | 0.3 | 1.2 | 691,074 | 207,600 | 284,186 | 1,097,963 |
| National | National | Evaluation_Status | Target_Sampled | 967 | 65.5 | 2.0 | 61.6 | 69.4 | 62,156,199 | 2,692,262 | 56,879,463 | 67,432,935 |
| National | National | Evaluation_Status | Total | 1644 | 100.0 | 0.0 | 100.0 | 100.0 | 94,913,706 | 2,918,088 | 89,194,358 | 100,633,054 |
| SandT_Class | E2EM | Evaluation_Status | Target_AA | 2 | 0.2 | 0.1 | 0.0 | 0.4 | 9,291 | 7,872 | 0 | 24,719 |
| SandT_Class | E2EM | Evaluation_Status | Target_Access_Denied | 22 | 5.7 | 1.7 | 2.4 | 9.1 | 321,616 | 94,133 | 137,119 | 506,114 |
| SandT_Class | E2EM | Evaluation_Status | Target_Inaccessible | 25 | 4.9 | 1.0 | 2.8 | 6.9 | 272,786 | 53,577 | 167,777 | 377,794 |
| SandT_Class | E2EM | Evaluation_Status | Target_Other | 1 | 0.2 | 0.1 | 0.0 | 0.4 | 8,881 | 8,070 | 0 | 24,698 |
| SandT_Class | E2EM | Evaluation_Status | Target_Sampled | 258 | 89.1 | 2.1 | 85.0 | 93.1 | 4,987,824 | 528,104 | 3,952,758 | 6,022,890 |
| SandT_Class | E2EM | Evaluation_Status | Total | 308 | 100.0 | 0.0 | 100.0 | 100.0 | 5,600,398 | 530,019 | 4,561,580 | 6,639,217 |
| SandT_Class | E2SS | Evaluation_Status | Target_AA | 9 | 1.7 | 1.4 | 0.0 | 4.5 | 16,708 | 13,906 | 0 | 43,963 |
| SandT_Class | E2SS | Evaluation_Status | Target_Access_Denied | 48 | 7.6 | 2.9 | 2.0 | 13.3 | 75,669 | 26,260 | 24,200 | 127,139 |
| SandT_Class | E2SS | Evaluation_Status | Target_Inaccessible | 29 | 40.3 | 7.3 | 26.0 | 54.7 | 398,974 | 35,754 | 328,898 | 469,050 |
| SandT_Class | E2SS | Evaluation_Status | Target_Other | 4 | 0.0 | 0.0 | 0.0 | 0.1 | 399 | 209 | 0 | 809 |
| SandT_Class | E2SS | Evaluation_Status | Target_Sampled | 69 | 50.3 | 8.6 | 33.5 | 67.1 | 497,821 | 158,300 | 187,559 | 808,084 |
| SandT_Class | E2SS | Evaluation_Status | Total | 159 | 100.0 | 0.0 | 100.0 | 100.0 | 989,572 | 153,455 | 688,806 | 1,290,338 |
| SandT_Class | PEM | Evaluation_Status | Target_AA | 11 | 2.1 | 1.0 | 0.1 | 4.1 | 448,240 | 213,340 | 30,101 | 866,380 |
| SandT_Class | PEM | Evaluation_Status | Target_Access_Denied | 135 | 33.9 | 2.5 | 28.9 | 38.8 | 7,167,684 | 568,768 | 6,052,920 | 8,282,448 |
| SandT_Class | PEM | Evaluation_Status | Target_Inaccessible | 19 | 4.9 | 1.0 | 3.0 | 6.8 | 1,034,536 | 201,108 | 640,371 | 1,428,701 |
| SandT_Class | PEM | Evaluation_Status | Target_Other | 5 | 1.0 | 0.4 | 0.3 | 1.8 | 218,234 | 83,089 | 55,382 | 381,086 |
| SandT_Class | PEM | Evaluation_Status | Target_Sampled | 262 | 58.1 | 2.8 | 52.6 | 63.6 | 12,283,062 | 838,118 | 10,640,380 | 13,925,743 |
| SandT_Class | PEM | Evaluation_Status | Total | 432 | 100.0 | 0.0 | 100.0 | 100.0 | 21,151,755 | 825,972 | 19,532,881 | 22,770,630 |

# NonTarget: Sample Frame Issues

| Type | Subpopulation | Indicator | Category | NResp | Estimate.P | StdError.P | LCB95Pct.P | UCB95Pct.P | Estimate.U | StdError.U | LCB95Pct.U | UCB95Pct.U |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| National | National | Evaluation_Status | NonTarget_Active_Crop | 204 | 48.0 | 3.2 | 41.7 | 54.3 | 9,727,984 | 861,224 | 8,040,016 | 11,415,952 |
| National | National | Evaluation_Status | NonTarget_Ag_Aqua | 100 | 6.2 | 1.0 | 4.3 | 8.1 | 1,254,370 | 191,242 | 879,543 | 1,629,196 |
| National | National | Evaluation_Status | NonTarget_Inundated | 95 | 12.6 | 2.1 | 8.6 | 16.6 | 2,550,110 | 450,698 | 1,666,758 | 3,433,461 |
| National | National | Evaluation_Status | NonTarget_NotWetland | 183 | 17.5 | 1.9 | 13.8 | 21.2 | 3,549,857 | 399,742 | 2,766,377 | 4,333,336 |
| National | National | Evaluation_Status | NotTarget_WetlandType | 87 | 15.7 | 2.5 | 10.8 | 20.6 | 3,186,146 | 504,727 | 2,196,899 | 4,175,393 |
| National | National | Evaluation_Status | Total | 669 | 100.0 | 0.0 | 100.0 | 100.0 | 20,268,466 | 948,851 | 18,408,753 | 22,128,179 |
| SandT_Class | E2EM | Evaluation_Status | NonTarget_Ag_Aqua | 1 | 8.7 | 7.9 | 0.0 | 24.1 | 9,261 | 7,859 | 0 | 24,664 |
| SandT_Class | E2EM | Evaluation_Status | NonTarget_Inundated | 3 | 22.7 | 10.8 | 1.5 | 43.8 | 24,245 | 11,114 | 2,463 | 46,027 |
| SandT_Class | E2EM | Evaluation_Status | NonTarget_NotWetland | 3 | 9.4 | 5.6 | 0.0 | 20.3 | 10,027 | 5,405 | 0 | 20,621 |
| SandT_Class | E2EM | Evaluation_Status | NotTarget_WetlandType | 9 | 59.3 | 14.0 | 31.8 | 86.7 | 63,309 | 22,630 | 18,954 | 107,663 |
| SandT_Class | E2EM | Evaluation_Status | Total | 16 | 100.0 | 0.0 | 100.0 | 100.0 | 106,842 | 16,672 | 74,166 | 139,517 |
| SandT_Class | E2SS | Evaluation_Status | NonTarget_Inundated | 2 | 54.7 | 28.4 | 0.0 | 100.0 | 863 | 527 | 0 | 1,896 |
| SandT_Class | E2SS | Evaluation_Status | NonTarget_NotWetland | 2 | 5.9 | 5.4 | 0.0 | 16.5 | 93 | 59 | 0 | 209 |
| SandT_Class | E2SS | Evaluation_Status | NotTarget_WetlandType | 1 | 39.5 | 28.8 | 0.0 | 95.9 | 623 | 554 | 0 | 1,708 |
| SandT_Class | E2SS | Evaluation_Status | Total | 5 | 100.0 | 0.0 | 100.0 | 100.0 | 1,578 | 573 | 455 | 2,702 |
| SandT_Class | PEM | Evaluation_Status | NonTarget_Active_Crop | 18 | 24.7 | 5.9 | 13.1 | 36.2 | 698,804 | 158,514 | 388,122 | 1,009,486 |
| SandT_Class | PEM | Evaluation_Status | NonTarget_Ag_Aqua | 5 | 9.4 | 4.5 | 0.6 | 18.2 | 266,889 | 123,738 | 24,367 | 509,411 |
| SandT_Class | PEM | Evaluation_Status | NonTarget_Inundated | 10 | 19.9 | 7.5 | 5.2 | 34.5 | 563,511 | 236,303 | 100,367 | 1,026,655 |
| SandT_Class | PEM | Evaluation_Status | NonTarget_NotWetland | 45 | 31.0 | 6.1 | 19.0 | 43.0 | 878,839 | 180,264 | 525,529 | 1,232,150 |
| SandT_Class | PEM | Evaluation_Status | NotTarget_WetlandType | 9 | 15.0 | 5.6 | 4.0 | 26.0 | 426,135 | 165,441 | 101,876 | 750,394 |
| SandT_Class | PEM | Evaluation_Status | Total | 87 | 100.0 | 0.0 | 100.0 | 100.0 | 2,834,178 | 235,721 | 2,372,174 | 3,296,183 |

# Weight Calculation: Types

- **Basic design weight calculation**
  - Initial design
  - As-implemented design
    - Prior survey sites
    - New survey sites
    - Combined Prior and New survey sites
- **Adjusting weights to address sample frame imperfections and inability to sample target sites**
  - Unknown eligibility weight adjustment
  - Non-response weight adjustment

# Oregon Stream Example

- **Sampling frame stream length (km) by ecoregion, stream size and land ownership**

| Ecoregion | | Land ownership | | | | | |
|---|---|---|---|---|---|---|---|
| | Stream size | Non-federal | USFS | BLM | Other federal | Tribal | Total |
| **Mountain** | | | | | | | |
| | Small | 26413 | 21842 | 5163 | 353 | 505 | 54276 |
| | Large | 10983 | 4220 | 1404 | 81 | 344 | 17033 |
| | Total | 37396 | 26063 | 6566 | 435 | 849 | 71309 |
| **Lowland** | | | | | | | |
| | Small | 7472 | 7 | 1517 | 414 | 9 | 9420 |
| | Large | 5628 | 3 | 1444 | 251 | 60 | 7386 |
| | Total | 13100 | 10 | 2961 | 666 | 69 | 16806 |
| **Total** | | | | | | | |
| | Small | 33885 | 21849 | 6680 | 768 | 514 | 63696 |
| | Large | 16611 | 4223 | 2847 | 333 | 405 | 24419 |
| | Total | 50496 | 26073 | 9527 | 1101 | 919 | 88115 |

# Design Weights

- **300 site equal probability design using GRTS**
  - **Weight is $w_i$ = L/n = 88115/300 = 440.5735 km where L is stream length and n is sample size**
- **300 site design stratified by ecoregion and stream size with 75 sites in each stratum**
  - **Weights are** $w_{hi} = M_h / n_h$

| Ecoregion | Stratum extent | | | Stratum weight | |
|---|---|---|---|---|---|
| | Stream size | | | Stream size | |
| | Small | Large | Total | Small | Large |
| Mountain | 54,276 | 17,033 | 71,309 | 723.6822 | 227.1034 |
| Lowland | 9,420 | 7,386 | 16,806 | 125.5954 | 98.4817 |
| Total | 63,696 | 24,419 | 88,115 | | |

# Design Weights Unequal Prob

- **300 site design with unequal probability by ecoregion and stream size with 75 sites in each of the four categories**
  - **Weights are** $u_{ki} = 1/\pi_{ki} = M_k/75$

| Ecoregion | Class Extent | | | Class Weight | | |
|---|---|---|---|---|---|---|
| | Stream Size | | | Stream size | | |
| | Small | Large | Total | Small | Large | |
| Mountain | 54,276 | 17,033 | 71,309 | 723.6822 | 227.1034 | |
| Lowland | 9,420 | 7,386 | 16,806 | 125.5954 | 98.4817 | |
| Total | 63,696 | 24,419 | 88,115 | | | |

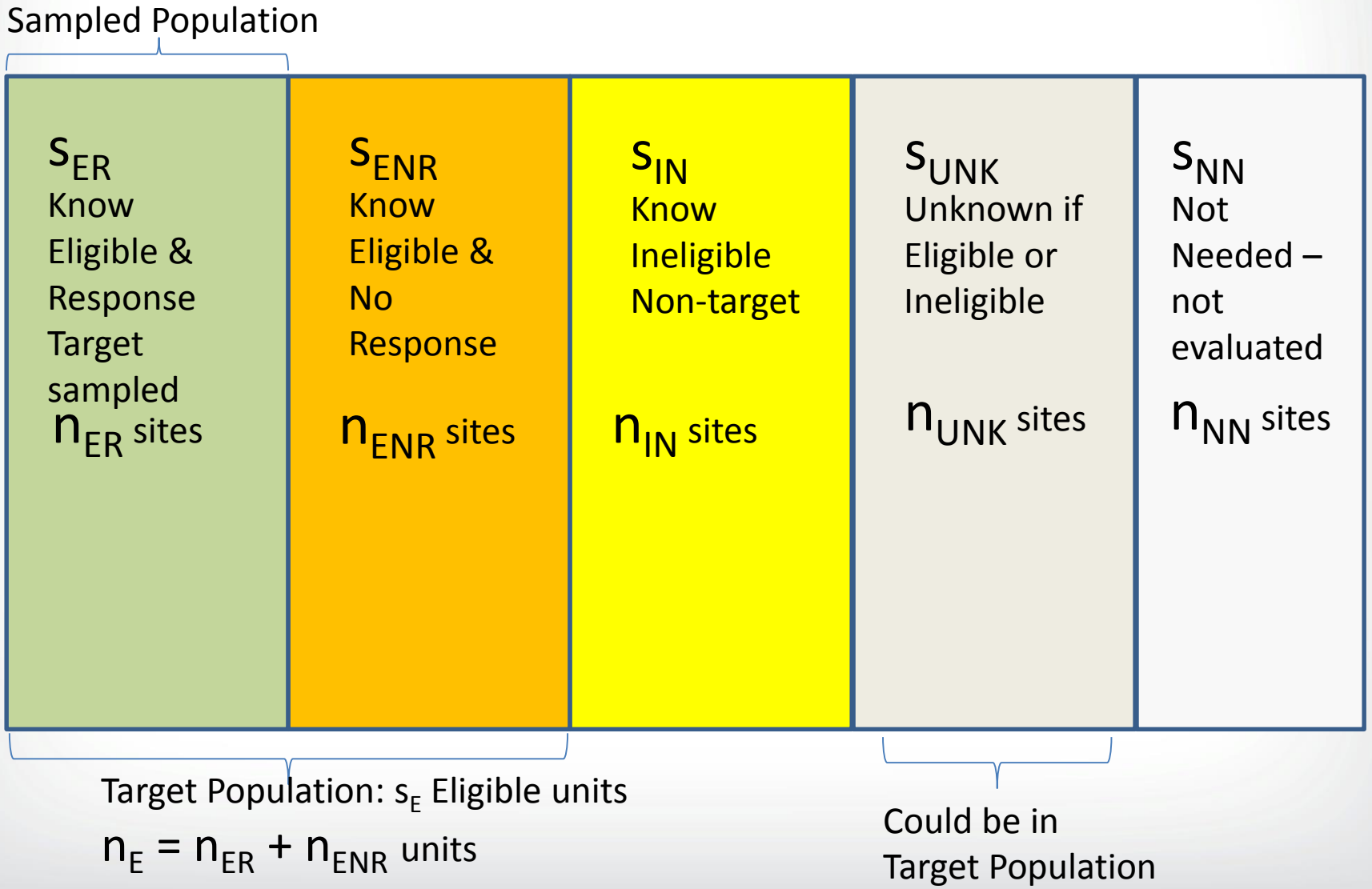
| Ecoregion | Class realized sample size | | | Sum of design weights | | |
|---|---|---|---|---|---|---|
| | Stream Size | | | Stream size | | |
| | Small | Large | Total | Small | Large | Total |
| Mountain | 70 | 86 | 156 | 50,658 | 19,531 | 70,189 |
| Lowland | 80 | 64 | 144 | 10,048 | 6,303 | 16,351 |
| Total | 150 | 150 | 300 | 60,706 | 25,834 | 86,540 |

- **selection process guarantees a fixed 300 sample size but it does not guarantee 75 in each category**
- **Weight may not sum to the extent of the sampling frame**

# As-Implemented Design Weights

- **Designs implemented as planned DO NOT require any adjustment of design weights**
  - **Assumes all sites included in the "Base" design are evaluated**
  - **No "over sample" sites are evaluated**
  - **Number of sites sampled will depend on**
    - **Gaining landowner permission to sample**
    - **Gaining physical access to site**
- **Aquatic resource designs are rarely implemented as planned**
  - **Sample units are found to be ineligible**
  - **Eligible sample units cannot be sampled due to physical access, landowner permission not obtained, permits not approved, and other reasons**
  - **Study planned to sample the planned number of "Base" sites so "over sample" sites are evaluated and sampled to meet that planned number**
- **Example As-Implemented design**
  - **Plan to sample 150 sites, design is based on 150 sites with an over sample of 150**
  - **Evaluated 204 sites to achieve 150 sampled sites**
  - **Weights must be adjusted since design and initial weights were based on 150 sites**

# Site Evaluation Subsets



# Class-based Weight Adjustment

- Define $s_{EV}$ as the set of sample units in $s$ that have been evaluated for potential eligibility and sampling. The number of sample units in $s_{EV}$ is $n_{EV} = n_{IN} + n_{ER} + n_{ENR} + n_{UNK}$
- Create k=1, … , K classes based on information that is known for all sample units in the sampling frame. The classes are formed to reflect how sample units are chosen to be evaluated for potential use.
- Define $M_k$ as the extent in the sampling frame in the $k$th class and $M = \sum_{k=1}^{K} M_k$ is the total extent of the sampling frame
- Define $s_k$ as the set of sample units in class $k$ where information in the sampling frame is used to assign the sample unit to class $k$. All units must be assigned to one and only one class.
- Define $s_{k,EV}$ as the set of sample units in class $k$ that were evaluated for potential eligibility and sampling, $S_{k,EV} = S_k \cap S_{EV}$. Then the number of sample units in k that have been evaluated is $n_{k,EV}$ and $n_{EV} = \sum_{k=1}^{K} n_{k,EV}$
- Calculate the adjustment for sample units in class $k$ as $a_k = \frac{M_k}{\sum_{i \in S_{k,EV}} w_{Di}}$ where $w_{Di}$ is the design weight for the $i$th sample unit.
- Adjusted weight for sample unit $i$ in $S_{k,EV}$ is $w_{Vi} = w_{Di} a_k$
- Adjusted weight $w_{Vi}$ is zero for the sample units not needed in class $k$, i.e. in $S_{k,NN} = S_k \cap S_{NN}$

# Lake As-Implemented Example 1

- Plan to sample 150 lakes based on an unequal probability survey design
- Unequal probability selection using a combination of ecoregion and lake area categories
- Based on past experience sample size set at 300 with expected sample size of 75 in each combination
- Evaluated $n_{EV}$ = 213 lakes in "siteID order" to achieve 150 sampled lakes
- Weights must be adjusted since design and initial weights were based on 300 lakes
- Single adjustment class is used with M = 1831 sample units in the class
- Adjustment factor is $$a_1 = \frac{M}{\sum_{i \in s_{1,EV}} w_{Di}} = \frac{1831}{1321.787} = 1.385246$$
- Adjusted weight is $$w_{Vi} = w_{Di} * 1.385246$$
- weights are 16.1981 for Mountain lakes ≤ 10 ha, 9.4751for Lowland lakes ≤ 10 ha, 5.19001 for Mountain lakes >10 ha and 2.955192 for Lowland lakes >10 ha.

# Lake Example: 1 Class Adjustment

| Ecoregion | Class Extent | | | Number of Sample Units Evaluated | | |
|---|---|---|---|---|---|---|
| | Lake Area | | | Lake Area | | |
| | ≤ 10 ha | >10 ha | Total | ≤ 10 ha | >10 ha | Total |
| Mountain | 877 | 281 | 1158 | 58 | 51 | 109 |
| Lowland | 513 | 160 | 673 | 49 | 55 | 104 |
| Total | 1390 | 441 | 1831 | 107 | 106 | 213 |

| Ecoregion | Design Weights | | | Sum of Design Weights | | |
|---|---|---|---|---|---|---|
| | Lake Area | | | Lake Area | | |
| | ≤ 10 ha | >10 ha | | ≤ 10 ha | >10 ha | Total |
| Mountain | 11.6933 | 3.7467 | | 982.2 | 264.1 | 1246.3 |
| Lowland | 6.8400 | 2.1333 | | 441.2 | 172.8 | 614.0 |
| | | | | 1423.4 | 436.9 | 1860.3 |

| Ecoregion | Adjusted Weights | | | Sum of Adjusted Weights | | |
|---|---|---|---|---|---|---|
| | Lake Area | | | Lake Area | | |
| | ≤ 10 ha | >10 ha | | ≤ 10 ha | >10 ha | Total |
| Mountain | 16.1981 | 5.19001 | | 939.49 | 264.69 | 1204.19 |
| Lowland | 9.4751 | 2.955192 | | 464.28 | 162.54 | 626.81 |
| Total | | | | 1403.77 | 427.23 | 1831 |

# Lake As-Implemented Example 2

- Plan to sample 150 lakes based on an unequal probability survey design
- Unequal probability selection using a combination of ecoregion and lake area categories
- Based on past experience sample size set at 300 with expected sample size of 75 in each combination
- Evaluated total of $n_{EV}$ = 216 lakes in "siteID order" within each combination to achieve 38/37 sampled lakes in each combination
- Weights must be adjusted since design and initial weights were based on 300 lakes
- Four adjustment classes used since lakes replaced by lake within same unequal probability classs
- Adjustment factors are $$a_k = \frac{M_k}{\sum_{i \in s_{k,EV}} w_{Di}}$$
- Adjusted weights is $$w_{Vi} = w_{Di} a_k$$

# Lake Example: 4 Class Adjustment

| Ecoregion | Class Extent | | | Number of Sample Units Evaluated | | |
|---|---|---|---|---|---|---|
| | Lake Area | | | Lake Area | | |
| | ≤ 10 ha | >10 ha | Total | ≤ 10 ha | >10 ha | Total |
| Mountain | 877 | 281 | 1158 | 59 | 58 | 117 |
| Lowland | 513 | 160 | 673 | 64 | 51 | 115 |
| Total | 1390 | 441 | 1831 | 123 | 109 | 232 |

| Ecoregion | Design Weights | | | Sum of Design Weights | | |
|---|---|---|---|---|---|---|
| | Lake Area | | | Lake Area | | |
| | ≤ 10 ha | >10 ha | | ≤ 10 ha | >10 ha | Total |
| Mountain | 11.6933 | 3.7467 | | 982.2 | 264.1 | 1246.3 |
| Lowland | 6.8400 | 2.1333 | | 441.2 | 172.8 | 614.0 |
| | | | | 1423.4 | 436.9 | 1860.3 |

| Ecoregion | Adjusted Weights | | | Sum of Adjusted Weights | | |
|---|---|---|---|---|---|---|
| | Lake Area | | | Lake Area | | |
| | ≤ 10 ha | >10 ha | | ≤ 10 ha | >10 ha | Total |
| Mountain | 15.660714 | 5.301887 | | 877 | 281 | 1158 |
| Lowland | 8.69.4915 | 3.33333 | | 513 | 160 | 673 |
| Total | | | | 1390 | 441 | 1831 |

# Comparison 1 vs 4 Class Adjustment

- **1 Class adjustment**

| Ecoregion | Adjusted Weights | | | Sum of Adjusted Weights | | |
|---|---|---|---|---|---|---|
| | Lake Area | | | Lake Area | | |
| | ≤ 10 ha | >10 ha | | ≤ 10 ha | >10 ha | Total |
| Mountain | 16.1981 | 5.19001 | | 939.49 | 264.69 | 1204.19 |
| Lowland | 9.4751 | 2.955192 | | 464.28 | 162.54 | 626.81 |
| Total | | | | 1403.77 | 427.23 | 1831 |

- **4 Class Adjustment**

| Ecoregion | Adjusted Weights | | | Sum of Adjusted Weights | | |
|---|---|---|---|---|---|---|
| | Lake Area | | | Lake Area | | |
| | ≤ 10 ha | >10 ha | | ≤ 10 ha | >10 ha | Total |
| Mountain | 15.660714 | 5.301887 | | 877 | 281 | 1158 |
| Lowland | 8.69.4915 | 3.33333 | | 513 | 160 | 673 |
| Total | | | | 1390 | 441 | 1831 |

# Why the difference?

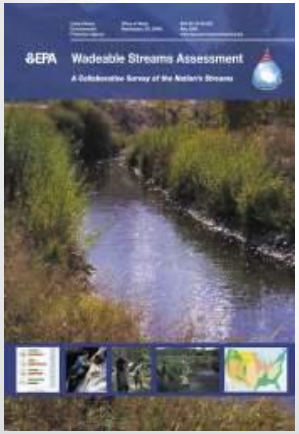
- Weight adjustment classes must match the classes used when selecting lakes to replace lakes that cannot be sampled.
  - Example 1: replaced ignoring ecoregion and lake area category
  - Example 2: replaced by ecoregion and lake area category
- The adjusted weights differ because different classes are used.
- The weight adjustment also depends on a random process since whether a lake must be replaced depends on it being selected for evaluation and on whether it can be sampled.
- Consequently, the adjusted weights are not fixed but are random in the sense that they are influenced by random events.
  - To minimize the impact that the random events have on the adjusted weights, the classes should be selected so that the number of sample units evaluated is at least greater than 10 and preferably much larger.
  - When 10 sample units are evaluated, the adjusted weights will change by less than 10 percent if one more or one less sample unit is evaluated.

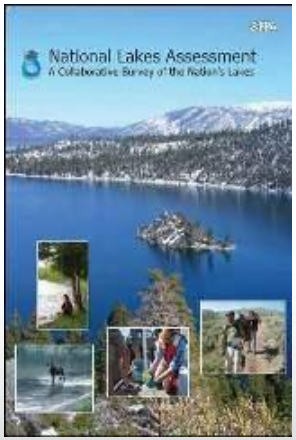
# Weight Adjustment Assumptions

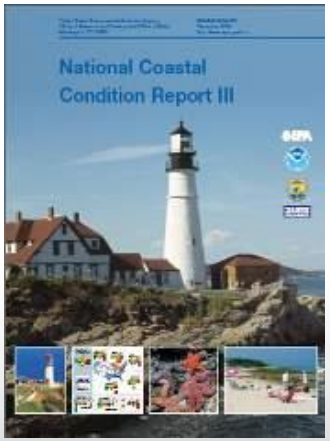
- **MCAR: missing completely at random**
  - **MCAR means that the sample unit's unknown eligibility or reason for not responding is independent of any information known about the sample unit or its response.**
  - **When sample units are MCAR, the eligible sampleable units are representative of the selected sample units.**
- **MAR: missing at random**
  - **MAR means that the sample unit's ineligibility or nonresponse does not depend on the response but depends only on information that is known about the sample unit. In this case, it is possible to model the dependence.**
  - **the information must be known for eligible sample units that are sampleable and that are not sampleable**
- **NINR: non-ignorable nonresponse**
  - **NINR means that the probability of a sample unit responding depends on one or more of the response variables and that this dependence cannot be removed based on information that is known for both sampleable and nonsampleable sample units.**

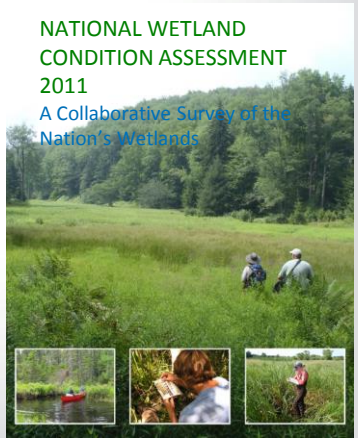
# Workshop Topics

- **Weights: What are they? Why are they important? How do I use them? How do I calculate them?**
- **Change and Trends: What are they? How do they impact survey design? How are they calculated?**
- **Designing a monitoring program: What is STRIDe and how can it help?**



# Change Design & Analysis

- **What are the definitions of trend and change?**
- **How does the choice of trend and change options influence the survey design?**
- **What options are available for change and trend estimation?**

# NARS Approach: National Consistency

- Four aquatic resources surveyed
  - Lakes and reservoirs
  - Rivers and streams
  - Coastal waters and Great Lakes
  - Wetlands
- Statistical survey design used as basis for estimates of the condition of an aquatic resource at national and regional scales
- Results are representative of the entire aquatic resource
- Long-term tracking of national and regional trends

# NARS Data Quality Objectives

- **National estimates**

  Estimate the proportion of aquatic resource (± 5%) in the conterminous U.S. that fall below the designated threshold for good conditions for selected measures with 95% confidence.

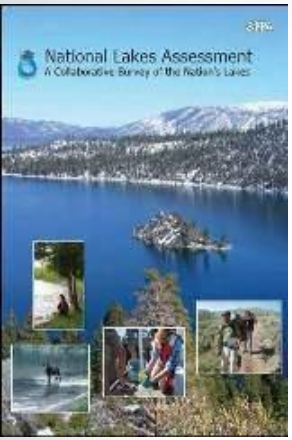
- **NARS Aggregated Ecoregions**

  Estimate the proportion of aquatic resource (± 15%) in a specific ecoregion that fall below the designated threshold for good conditions for selected measures with 95% confidence.

# NARS Survey Designs

- Spatial survey design applied once every five (5) years for each aquatic resource
- Each survey period has a sample size of 1,000 sites for the aquatic resource except 2,000 sites for rivers and streams
- Temporal survey design incorporates sampling with partial replacement over four (4) survey periods for an aquatic resource
- 10% of sites within a survey period revisited to obtain information on local measurement uncertainty and to provide quality assurance information

# NARS Five Year Cycle

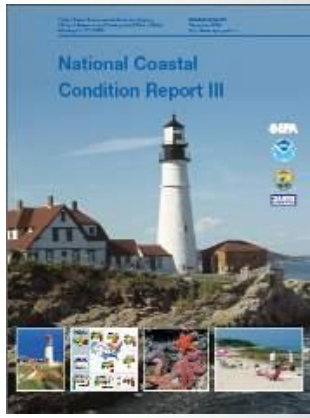



→2007 Lakes

→2008-9 Rivers & Streams

→2010 Coastal Waters

→2011 Wetlands

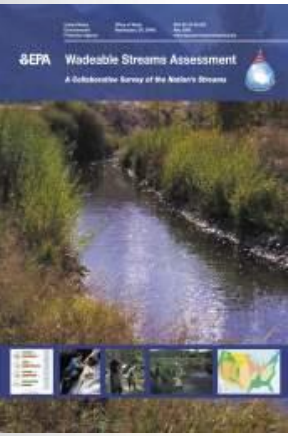



→2012 Lakes

→2013-14 Rivers & Streams

→2015 Coastal Waters

→2016 Wetlands

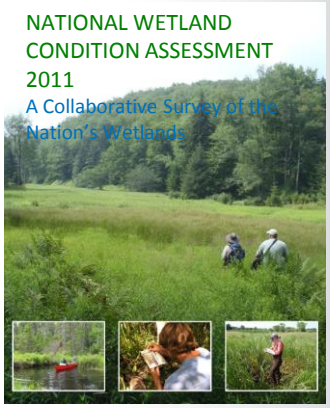



→2017 Lakes

→2018-19 Rivers & Streams

→2020 Coastal Waters

→2021 Wetlands




→2022 Lakes

→2023-24 Rivers & Streams

→2025 Coastal Waters

→2026 Wetlands

# NLA Temporal Design: Number of Lakes by Panel

| Panel | 2007 | 2012 | 2017 | 2022 | 2027 | 2032 | 2037 |
|---|---|---|---|---|---|---|---|
| NLA07_TS1 | 500 | | | | | | |
| NLA07_TS2 | 250 | 250 | | | | | |
| **NLA07_TS4** | **250** | **250** | **250** | **250** | | | |
| NLA12_TS1 | | 250 | | | | | |
| **NLA12_TS4** | | **250** | **250** | **250** | **250** | | |
| NLA17_TS1 | | | 250 | | | | |
| **NLA17_TS4** | | | **250** | **250** | **250** | **250** | |
| **NLA22_TS4** | | | | **250** | **250** | **250** | **250** |
| **NLA27_TS4** | | | | | **250** | **250** | **250** |
| **NLA32_TS4** | | | | | | **250** | **250** |
| **Total** | **1000** | **1000** | **1000** | **1000** | **1000** | **1000** | **…..** |

# NRSA Temporal Design
# Number of Sites by Panel

| Panel | 2008-9 | 2013-14 | 2018-19 | 2023-24 | 2028-29 | 2033-34 | 2038-9 |
|---|---|---|---|---|---|---|---|
| NRSA08_TS1 | 1000 | | | | | | |
| NRSA08_TS2 | 500 | 500 | | | | | |
| **NRSA08_TS4** | **500** | **500** | **500** | **500** | | | |
| NRSA13_TS1 | | 500 | | | | | |
| **NRSA13_TS4** | | **500** | **500** | **500** | **500** | | |
| NRSA18_TS1 | | | 500 | | | | |
| **NRSA18_TS4** | | | **500** | **500** | **500** | **500** | |
| **NRSA23_TS4** | | | | **500** | **500** | **500** | **500** |
| **NRSA28_TS4** | | | | | **500** | **500** | **500** |
| **NRSA33_TS4** | | | | | | **500** | **500** |
| **Total** | **2000** | **2000** | **2000** | **2000** | **2000** | **2000** | **….** |

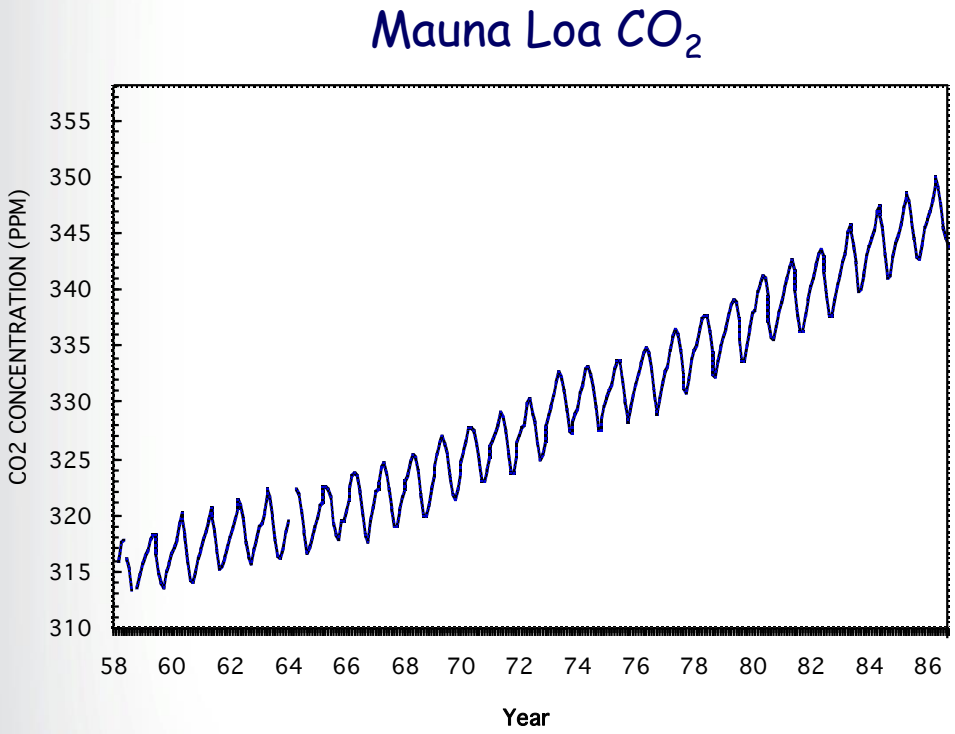
# Population Trend & Change

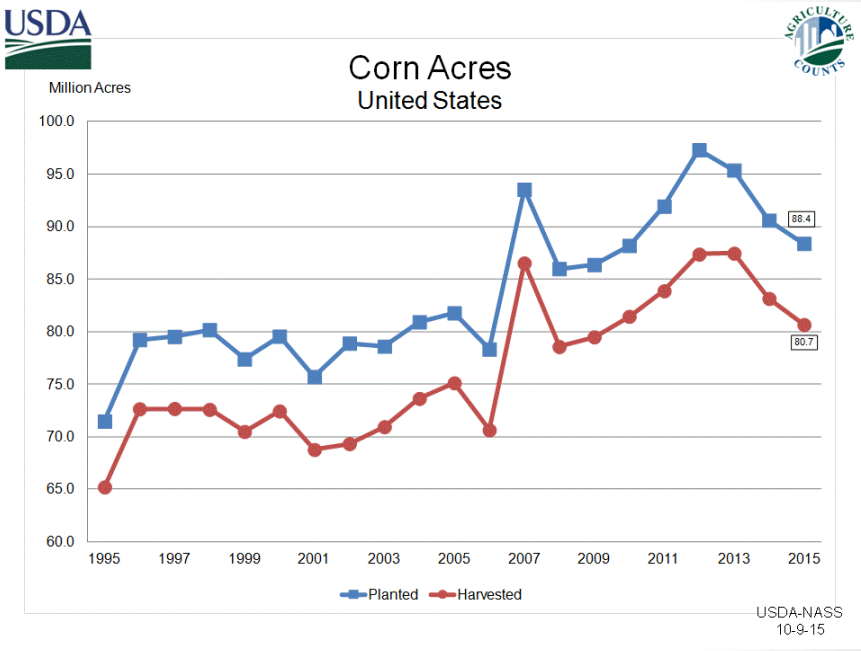
- **Trend trajectory: interest is in estimating the entire time series for all temporal units within the Temporal Period for population estimates of an indicator**
- **Trend pattern summary of trend trajectory**
  - **Linear change per year estimated from a regression model**
  - **Monotonic change per year estimated using Sen's median**
- **Change between two Temporal Units in estimates of population status for an indicator**
  - **Net Change: Overall change in status between two temporal units. Not information about change in indicator status for individual spatial units.**
  - **Gross Change: Change in indicator status for individual spatial units**

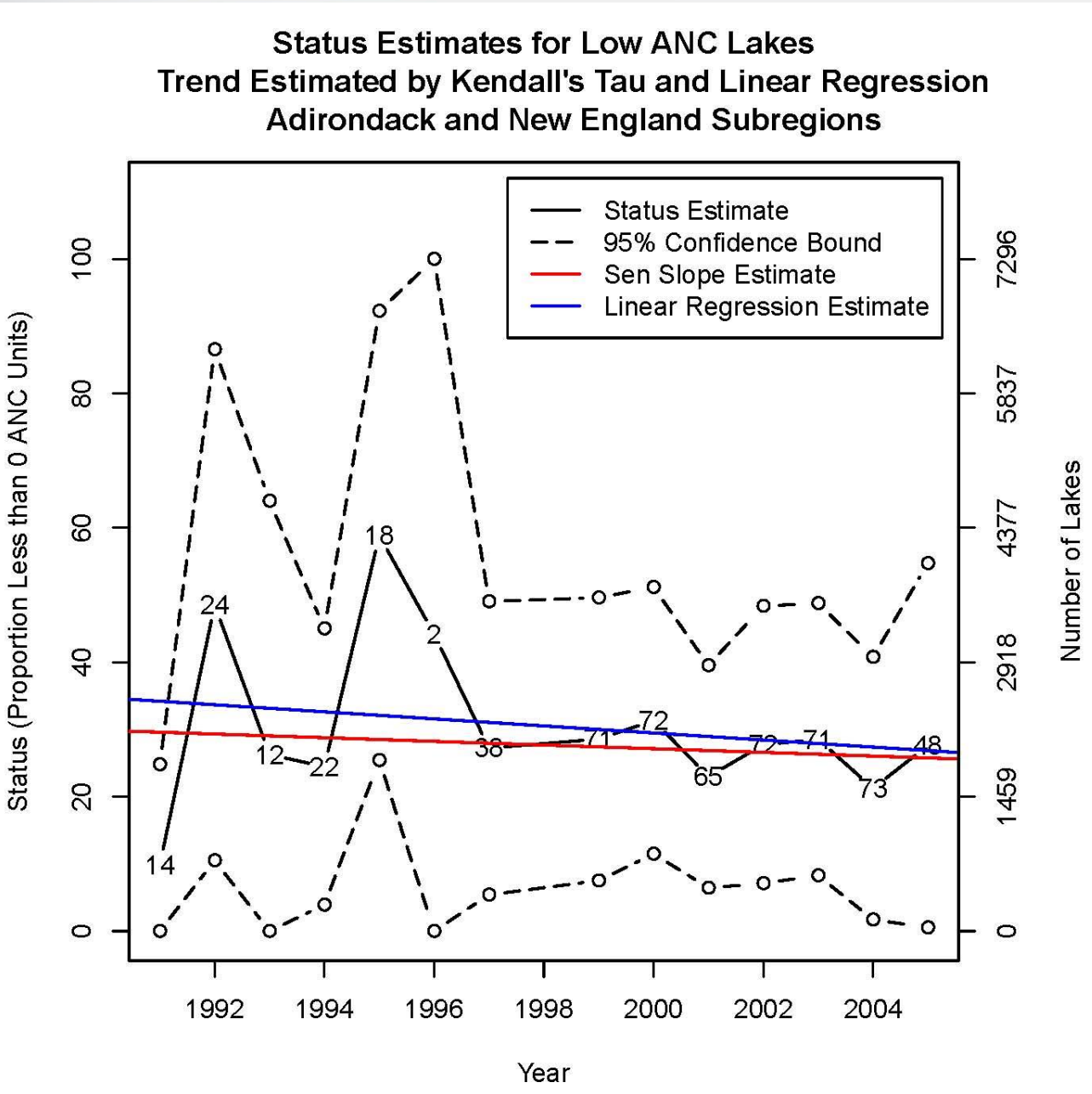
# Population Trends: Types

- Time trend in population status
  - **Trend trajectory**: entire time series
  - Summary of trend pattern
    - Linear change/yr from linear model
    - Monotonic change/yr (e.g. Sen's estimate)
- Change in population status between two time periods
  - **Net change**: Change in population status (e.g. %G period 2 minus %G in period 1)
  - **Gross change**: Two-way population status (e.g. G/F/P) table between time periods 1 and 2

# Trend Trajectory: Site versus Population



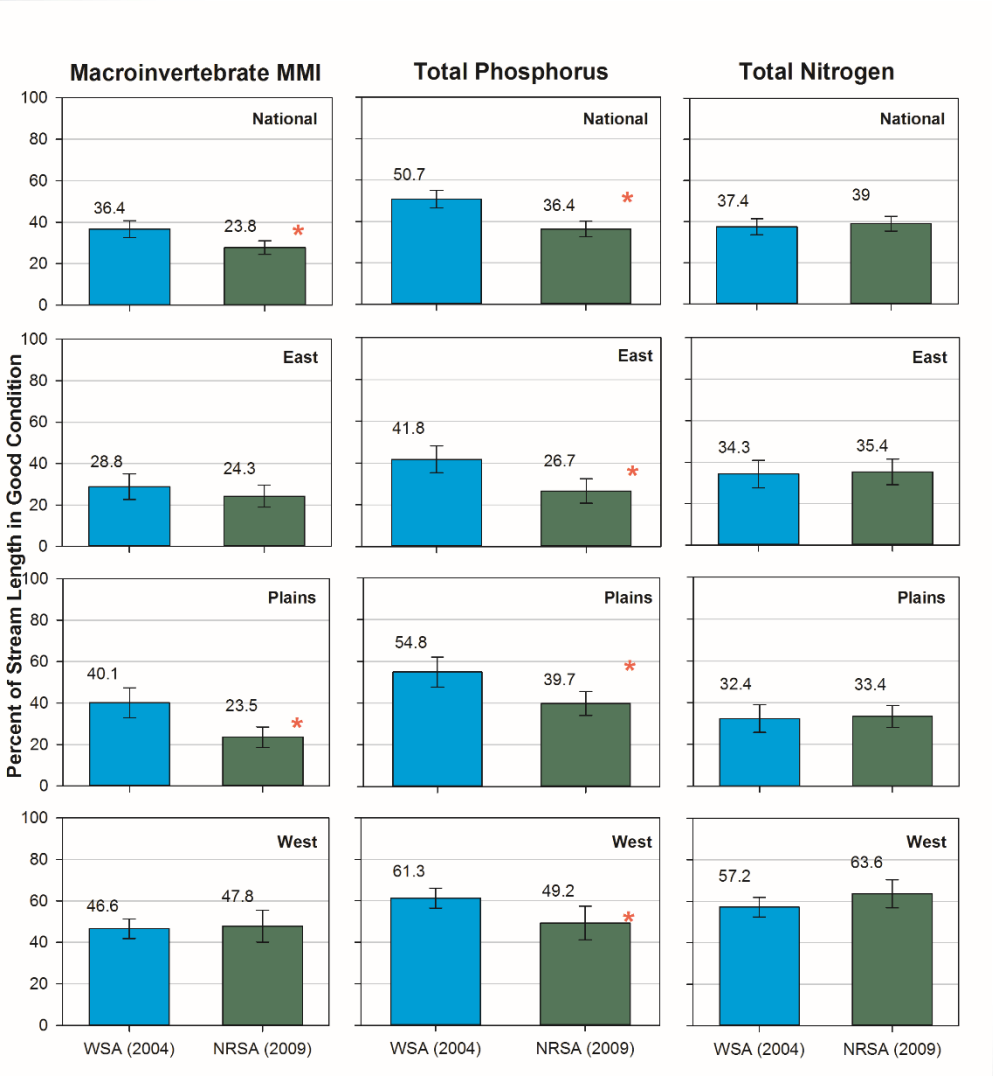


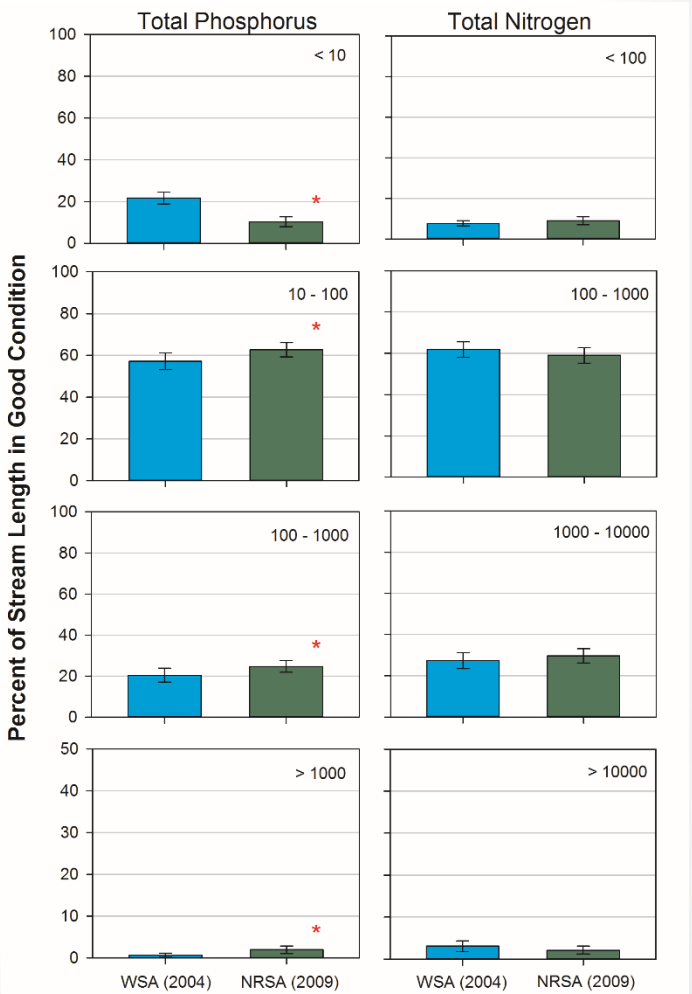

Population net change per year estimate

Sen's slope -0.27%/yr

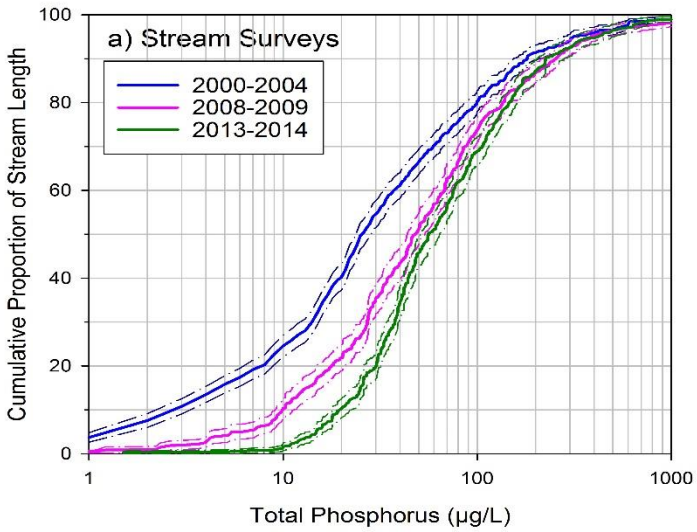
Regr slope -0.53%/yr

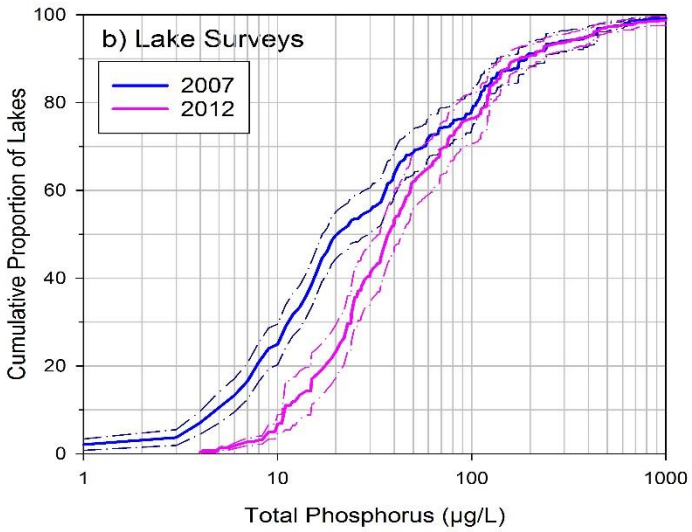
# Net Change Estimate between NRSA 2008-9 and WSA 2004



# Change in Population CDFs between time periods: NRSA & NLA



# Change Between Two Time Period Estimation

- R revisited sites in common for two time periods
- Q period 1 unique sites
- U period 2 unique sites
- $N_1$ & $N_2$ estimated stream length periods 1 & 2
- $S_{1R}$ & $S_{2R}$ estimated stream length from revisit sites for periods 1 & 2
- Estimate $\Delta = P_2 - P_1$ where $P_i$ is estimated proportion
- $Var(\Delta) = Var(P_2) + Var(P_1) - 2(S_{1R}*S_{2R})/(N_1*N_2) * Cov(P_{1R}, P_{2R})$
- In most cases Cov() is positive so reduces variance estimate

# STRIDe: What is it?

- **STRIDe: Spatial, Temporal, Response, and Inference Design approach to designing an aquatic resource monitoring program.**
- **How do monitoring questions influence STRIDe?**
- **Why should I begin the design of a monitoring program by defining what reports will be produced by the program?**
- **Is it useful to agree on a set of terms to clarify the process?**

# Important terms to know

- **Reporting domains**
  - **Specific collections of spatial-temporal units in the spatial and temporal domain for which estimates are required**
- **Spatial domain & spatial unit**
  - **Geographic region over which the study will be conducted**
  - **Domain usually consists of a collection of spatial units**
- **Temporal domain & temporal unit**
  - **Entire length of time over which the study will be conducted**
  - **Domain usually consists of a sequence of temporal units**
- **Summary estimates, indicators, metrics & measurements**

# Estimates, Indicators, Measurements

- **A measurement is a value resulting from a data collection event which is taken on or within a spatial-temporal unit using protocols described in what is called the response design**
- **An indicator is the spatial-temporal unit value resulting from the reduction or processing of measurements as described in the response design**
- **An estimate is the reporting domain value resulting from the processing of indicators across spatial-temporal units as described by what is called the inference design**

# STRIDe

- **S**patial design: how we select what spatial units to monitor within the spatial domain
- **T**emporal design: how we select what temporal units to monitor within the temporal domain
- **R**esponse design: what measurements we make, how we take them & how we calculate indicators on spatial-temporal units based on the measurements
- **I**nference **De**sign: how we summarize indicators across spatial-temporal units within a temporal domain to obtain estimated value for a reporting domain

# Building Blocks Spatial Design

- **Spatial domain and its spatial units define the target population**
- **Target population**
  - **Requires a clear, precise written definition**
    - **Must be understandable to users**
    - **Field crews must be able to determine if a particular site is included**
  - **More difficult to define than most expect.**
  - **Includes definition of what the spatial units (elements) are that make up the spatial domain**
  - **Definition is written and usually not given in terms of a GIS layer**
- **Spatial survey design specifies how spatial units are selected to meet monitoring program objectives**

# Spatial Design & Representative Sample

- **Goal is to obtain a "representative sample" of the target population that can be used to make inferences from the metric values on the sampled spatial units to indicator values for a reporting domain**
- **Problem: At least 9 definitions for representative sample**
  - **General acclaim for data**
  - **Absence of selective forces**
  - **Miniature of the population**
  - **Typical or ideal case(s)**
  - **Coverage of the population**
  - **Vague term, to be made precise**
  - **Representative sampling as a specific sampling method**
  - **Representative sampling as permitting good estimation**
  - **Representative sampling as good enough for a particular purpose**

# Representative Spatial Sample

- Goal is for the spatial survey design to result in a sample that is a "miniature" of the indicator value in the population
- Example: IBI indicator cumulative distributions for
  - Entire target population of lakes
  - Estimated cumulative distribution based on a spatial sample of size 25
- A representative sampling process (spatial survey design) enables margins of error to accompany the estimates
- Generalized Random Tessellation Stratified (GRTS) survey design options available in R package spsurvey produce spatial survey designs appropriate for aquatic resources