Estimating Status for CHaMP's Status and Trend Program

Leigh Ann Starcevich

Statistical consultant

November 27, 2012

Outline

- Design weights
  - How are they calculated?
  - When is weighting adjustment needed?
- Nonsampling error sources and impacts
- Estimating population extent and summary statistics
- On-going work

Design-based inference

- Model-based inference vs. design-based inference
- Design weights = sampling weights = inclusion weights
- Design weights are the number/extent of the population represented by the sampling units
  - Proper weighting required for unbiased estimation
  - Sum of the weights for the sampled units should equal the population number/extent

Weighting adjustment

- Required when:
  - Sample size is larger/smaller than intended
  - Certain nonsampling errors occur
- Failure to properly adjust weights may result in:
  - Biased inference
  - Confidence interval undercoverage

Nonsampling error

- The result of the imperfect execution of the sampling design
- **Nonresponse error**: a complete set of metrics is not obtained for every unit in the sample
  - Substituting oversample sites may not resolve the problem
- **Frame error**: Target population sites are omitted from the sampling frame or non-target sites are included in the frame
- **Measurement error**: We assume that this is not an issue

Frame error

- Adjusting for frame error is necessary for unbiased estimates of totals
- More frame error in Entiat and Wenatchee

Frame error (2011)

Watershed	Total Sites Evaluated	Non-Target Sites
Entiat	145	44 (30%)
John Day	107	6 (6%)
Lemhi	64	1 (2%)
SF Salmon	86	7 (8%)
Tucannon	43	0 (0%)
Upper Grande Ronde	101	6 (6%)
Wenatchee	73	31 (42%)

Frame error (2012)

Watershed	Total Sites Evaluated	Non-Target Sites
Entiat	154	51 (33%)
John Day	108	9 (8%)
Lemhi	72	3 (4%)
SF Salmon	64	6 (9%)
Tucannon	46	0 (0%)
Upper Grande Ronde	111	3 (3%)
Wenatchee	77	13 (17%)

Nonresponse error

- An issue for all watersheds
- EvalReason = “Provide Justification”
  - Need mutually exclusive categories to identify sites that are evaluated, visited, and successfully surveyed

Nonresponse error (2011)

Watershed	Total Sites where Surveys Attempted	Nonresponding Sites
Entiat	81	5 (6%)
John Day	77	14 (18%)
Lemhi	48	7 (15%)
SF Salmon	49	14 (29%)
Tucannon	32	7 (22%)
Upper Grande Ronde	75	19 (25%)
Wenatchee	35	11 (31%)

Nonresponse error (2012)

Watershed	Total Sites where Surveys Attempted	Nonresponding Sites
Entiat	59	7 (12%)
John Day	82	12 (15%)
Lemhi	66	19 (29%)
SF Salmon	37	12 (32%)
Tucannon	33	5 (15%)
Upper Grande Ronde	87	32 (37%)
Wenatchee	42	20 (48%)

Nonresponse reasons

- Three main reasons:
  - Landowner Denial
  - Not safe/Inaccessible
  - Provide Justification
- We may handle these differently based on the nature of the nonresponse
- For now, we are treating the missing data as unrelated to the design, other covariates, or the indicator of interest

Nonresponse reasons (2011)

Watershed	Landowner Denial	Not safe/Inaccessible	Provide Justification
Entiat	1	4	0
John Day	13	0	1
Lemhi	3	2	2
SF Salmon	2	2	10
Tucannon	7	0	0
Upper Grande Ronde	12	0	7
Wenatchee	7	0	4

Nonresponse reasons (2012)

Watershed	Landowner Denial	Not safe/Inaccessible	Provide Justification
Entiat	5	2	0
John Day	8	1	3
Lemhi	17	0	2
SF Salmon	2	1	9
Tucannon	4	0	1
Upper Grande Ronde	17	0	15
Wenatchee	14	3	3

CHaMP weights

- Valley class
- Stream order
- Priority drainage
- Land ownership

Post-hoc strata

- Sample size requirements by Land Ownership
  - Like a design stratum with a priori sample size
  - Different in that sampling not conducted within Ownership strata
  - Could affect spatial balance
  - Sample sizes within Ownership categories are often too small for inference
  - Weighting adjustment within levels of post-hoc strata when sample size is sufficient

Legacy sites

- Don Stevens recommends equal weighting for legacy and STM sites
  - Reasonable if legacy sites are randomly selected
  - Less ideal if legacy sites are subjectively chosen
- When sample sizes are sufficient, we can test for differences between legacy and STM sites
- 2012 legacy information not yet summarized

2011 Legacy Sites by Watershed

Watershed	Total Sites Evaluated	Total Sites Sampled	Legacy Sites
Entiat	95	17	5 (29%)
John Day	107	63	37 (59%)
Lemhi	64	41	29 (71%)
SF Salmon	86	35	2 (6%)
Tucannon	43	25	0 (0%)
Upper Grande Ronde	101	56	15 (27%)
Wenatchee	73	24	21 (88%)

Notation for weighting

n_{eval} = Number of evaluated sites

n_T = Number of target sites

n_S = Number of sites at which surveys were attempted

n_R = Number of surveyed sites

$|R|$ = Extent of the resource (e.g. stream km)

Effective sample size

$$n' = n_{eval} \times \frac{n_T}{n_{eval}} \times \frac{n_S}{n_T} \times \frac{n_R}{n_S}$$

$$n' = n_{eval} \times \frac{n_R}{n_T}$$

Weighting adjustment

$$\text{Adj. weight} = w_i = \frac{|R|}{n'} = \frac{|R|}{n_{eval} \times \frac{n_R}{n_T}}$$

Estimate frame extent:

$$\sum_{i=1}^{n_R} w_i = \sum_{i=1}^{n_R} \frac{|R|}{n_{eval} \times \frac{n_R}{n_T}} = \frac{n_R |R|}{n_{eval} \times \frac{n_R}{n_T}} = \frac{n_T}{n_{eval}} |R|$$

An example - Entiat

Stratum	n	n_{eval}	n_T	n_S	n_R	n'	$\|R\|$ *(km)*	w_i
Dep. Public	30	26	16	12	9	14.6	50.8	3.5
Source Private	6	2	2	1	1	1	24.8	24.8
Source Public	51	37	10	7	6	22.2	110.6	5.0
Trans. Private	1	1	1	1	1	1	2.0	2.0
Trans. Public	7	7	0	0	0	-	6.5	-

Estimate Entiat frame extent

- Obtain temporary weight = $|R|/n_{eval}$
- Use *cat.analysis* in *spsurvey* package
  - Use evaluated sites
  - Temporary weight
  - Indicator = EvalStatus != "Non-Target"
- Can also calculate the adjusted weights as:

$$\frac{\left|\hat{R}\right|}{n_R}$$

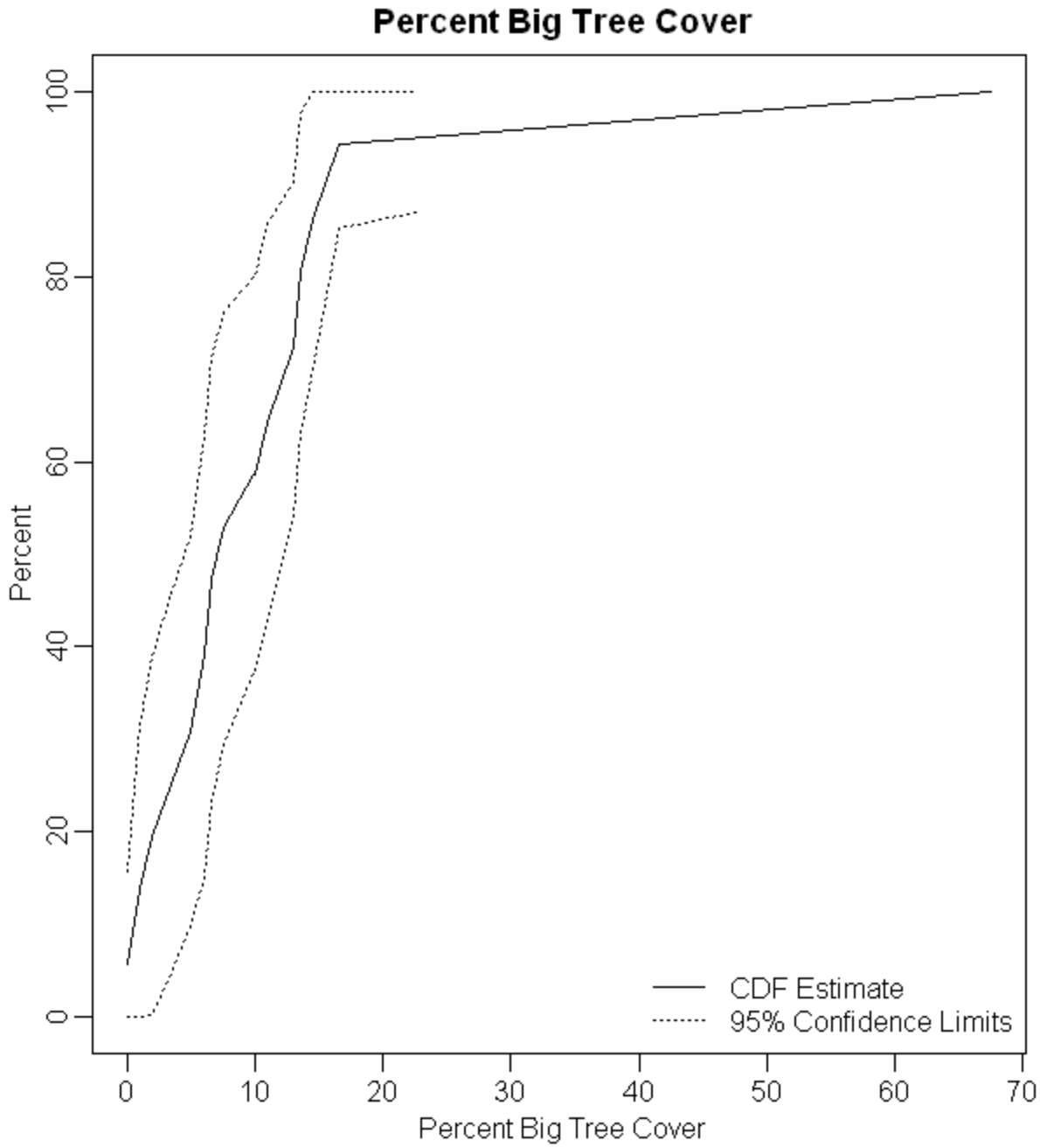
Frame estimates

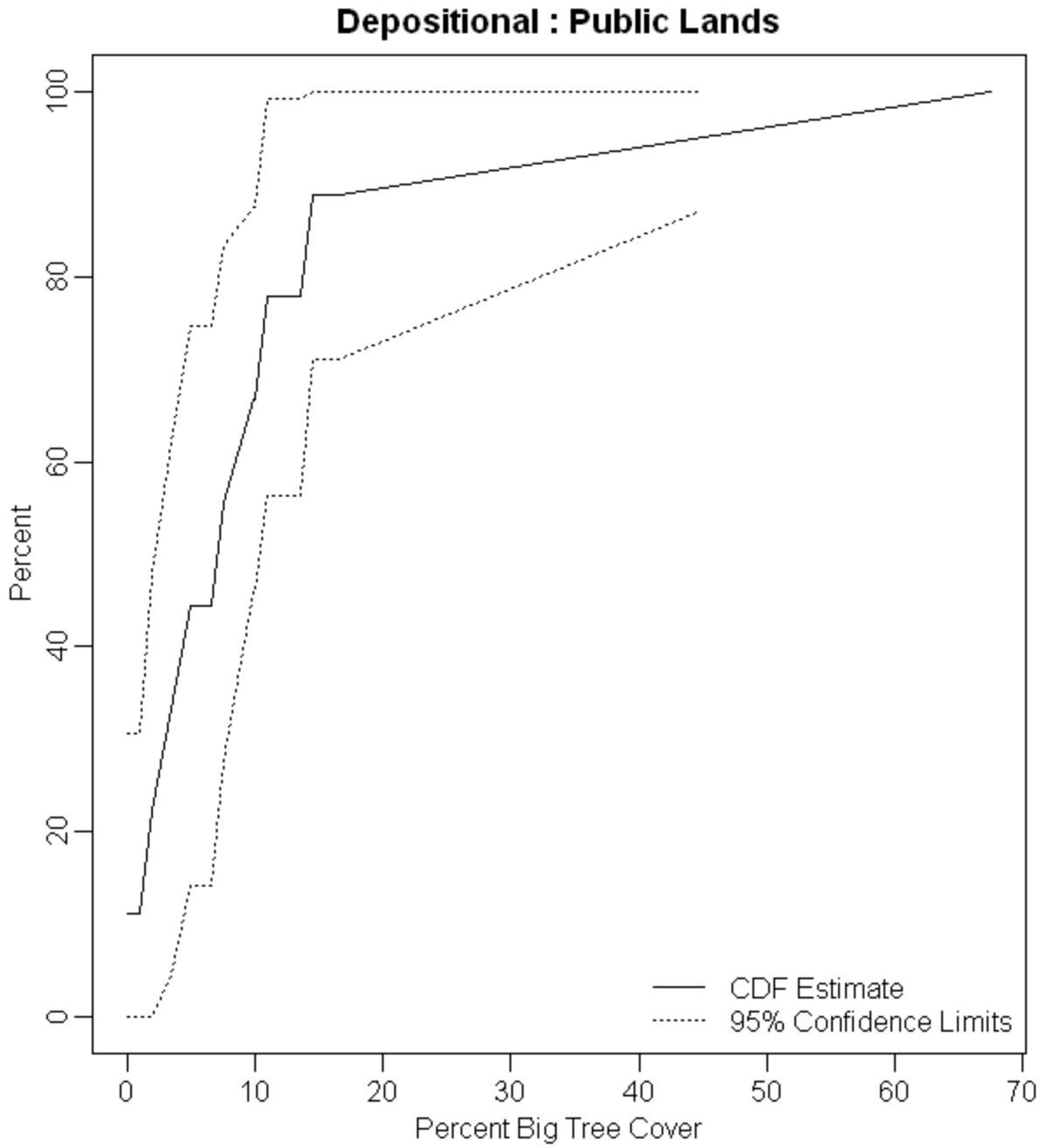
Subpop	Num Resp.	Est. Frame Extent	SE	95% CI Low	95% CI High
Dep. Public	16	31.27	2.30	26.76	35.77
Source Private	2	24.82	0.00	24.82	24.82
Source Public	10	29.88	5.95	18.22	41.54

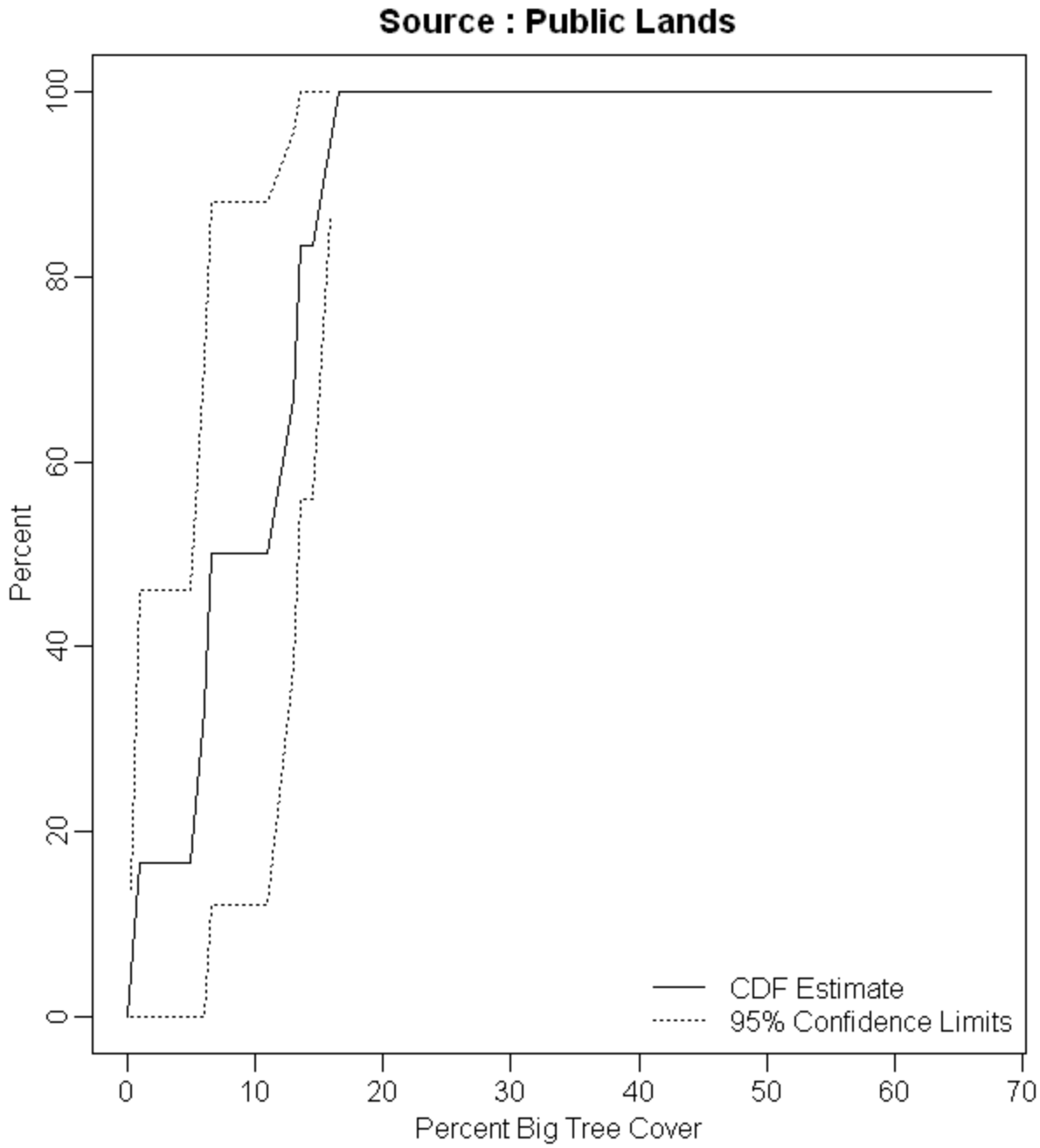
Status estimation

- Use adjusted weights with *cont.analysis*
- Variance estimate and confidence intervals do not reflect the nonresponse adjustment

Stratum	**Est. Mean**	**SE**	**95%-CI Low**	**95%-CI High**
Dep. Public	13.43	5.54	2.58	24.29
Source Public	9.43	2.04	5.44	13.43
ALL	11.48	3.00	5.60	17.36


Percent Big Tree Cover
Percent
0
20
40
60
80
100
0
10
20
30
40
50
60
70
Percent Big Tree Cover
CDF Estimate
95% Confidence Limits


Depositional : Public Lands
Percent
0
20
40
60
80
100
0
10
20
30
40
50
60
70
Percent Big Tree Cover
CDF Estimate
95% Confidence Limits


Source : Public Lands
Percent
0
20
40
60
80
100
0
10
20
30
40
50
60
70
Percent Big Tree Cover
CDF Estimate
95% Confidence Limits

Pool Frequency

Watershed	2011 Est. Mean	95%-CI	2012 Est. Mean	95%-CI
Entiat (STM only)	1.25	(0.75, 1.75)	1.39	(0.80, 1.98)
Lemhi	3.29	(1.75, 4.84)	2.23	(1.70, 2.76)
Secesh	2.61	(1.70, 3.51)	3.26	(2.38, 4.13)
SF Salmon	2.00	(1.30, 2.70)	-	-
Wenatchee	0.70	(0.38, 1.02)	1.97	(0.51, 3.43)

Thalweg Depth Profile Filtered CV

Watershed	2011 Est. Mean	95%-CI	2012 Est. Mean	95%-CI
Entiat (STM only)	0.31	(0.28, 0.35)	0.28	(0.25, 0.31)
Lemhi	0.37	(0.33, 0.40)	0.43	(0.40, 0.46)
Secesh	0.30	(0.27, 0.34)	0.32	(0.29, 0.35)
SF Salmon	0.34	(0.27, 0.42)	-	-
Wenatchee	0.36	(0.32, 0.39)	0.31	(0.27, 0.35)

Wetted Large Wood Volume By Site

Watershed	2011 Est. Mean	95%-CI	2012 Est. Mean	95%-CI
Entiat (STM only)	53.30	(29.98, 76.62)	12.94	(5.29, 20.59)
Lemhi	7.49	(1.56, 13.42)	1.53	(0.96, 2.10)
Secesh	104.85	(52.62, 157.08)	13.67	(6.08, 21.26)
SF Salmon	87.12	(18.84, 155.39)	-	-
Wenatchee	29.32	(6.50, 52.15)	7.94	(2.65, 13.24)

Measurement of D50

Watershed	2011 Est. Mean	95%-CI	2012 Est. Mean	95%-CI
Entiat (STM only)	71.42	(60.32, 82.52)	71.22	(55.28, 87.17)
Lemhi	42.15	(33.17, 51.13)	37.80	(32.76, 42.84)
Secesh	113.96	(79.69, 148.23)	73.20	(49.84, 96.56)
SF Salmon	54.55	(41.00, 68.10)	-	-
Wenatchee	39.84	(28.74, 50.95)	53.63	(39.35, 67.92)

Percent Big Tree Cover

Watershed	2011 Est. Mean	95%-CI	2012 Est. Mean	95%-CI
Entiat (STM only)	11.48	(5.60, 17.36)	11.29	(7.46, 15.12)
Lemhi	4.36	(1.54, 7.17)	3.20	(1.74, 4.66)
Secesh	13.33	(9.85, 16.83)	5.43	(4.00, 6.86)
SF Salmon	11.32	(4.10, 18.53)	-	-
Wenatchee	13.62	(8.75, 18.50)	7.05	(4.81, 9.29)

Ongoing work

- Complete weighting adjustment
- Compile estimates and CDF plots
- Nonresponse adjustments
  - Variance adjustment for nonresponse
  - Accounting for item nonresponse
- Archival of weights
- Test assumptions of legacy sites
- Trend modeling
  - Complications from design changes