

Statistical basis for use of genetic information for cohort analysis.

J. R. Candy, T.B. Beacham, and R.E. Withler
Molecular Genetics Lab
Pacific Biological Station

Coded-wire tag workshop
Seattle, WA
June 7-10

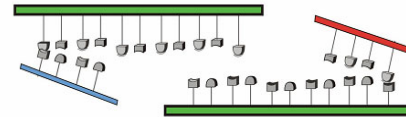
Using DNA to develop stock composition estimates

DNA Sample

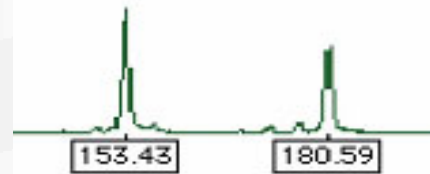
Extraction



Amplification



Sizing



Laboratory

DFO working baseline of ~260 populations from SE Alaska to Cal "inseason" turnaround 24-48 hrs.

Multi-locus Genotype



Mixture



Baseline

Analytical

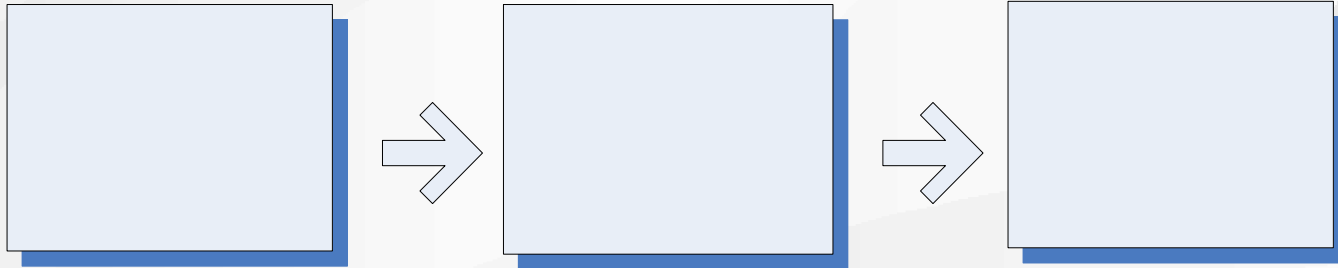
Mixture Estimate

Mixture model (Bayes*)

*Pella & Masuda 2001

Application of mixed stock genetic data now

Adaptive fisheries management



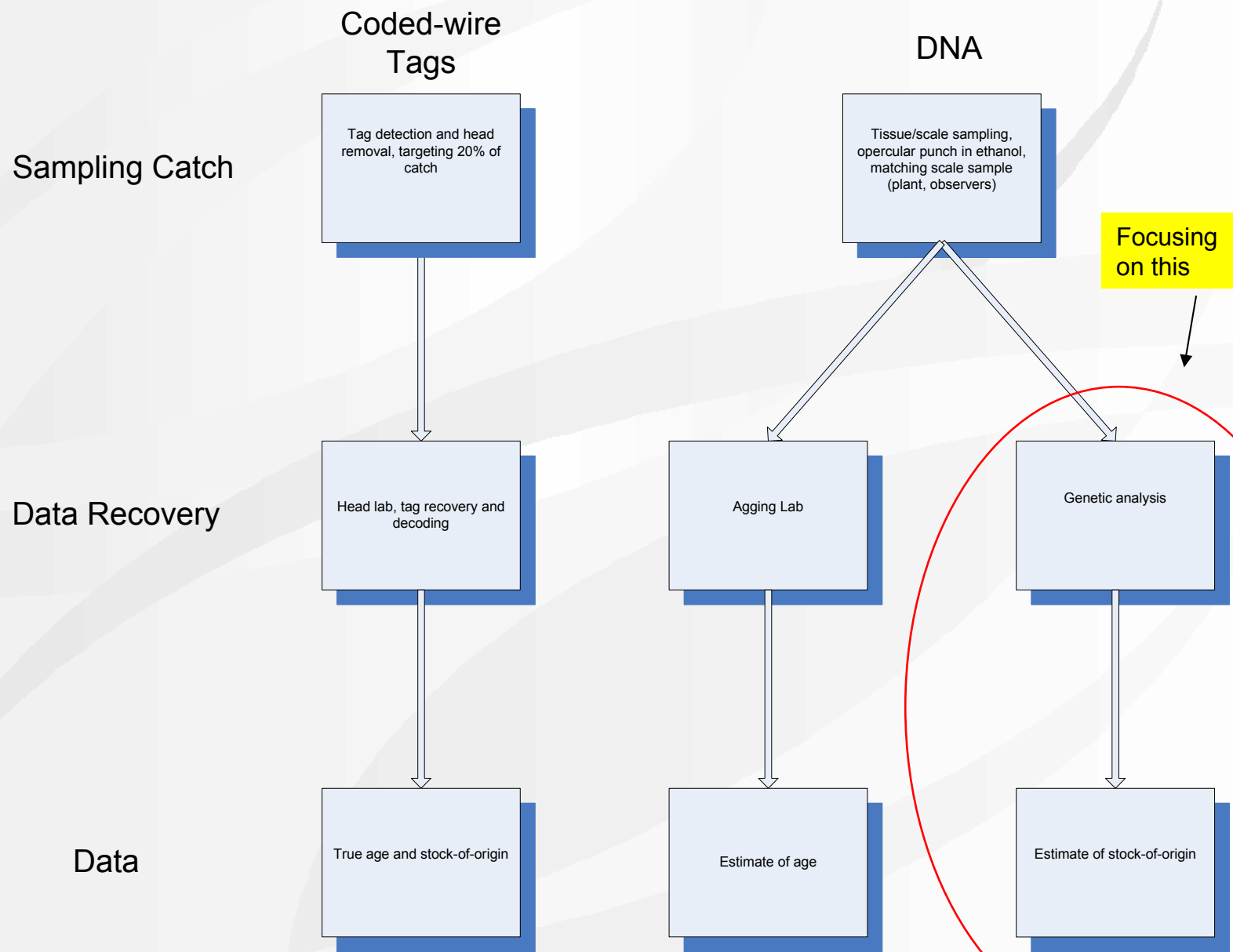
Can we use this data for more than fisheries management purposes?

We require for cohort analysis:

- Accurate estimates to stock-of-origin.**
- Accurate estimates of age**

Pre-season planning

CWT "like" data from DNA sampling



Accuracy and Precision

- Estimation Error

- What is the *stock composition accuracy* in the sample?

- (accuracy ~ baseline build)

- What is the trade-off between sample size and precision in the model?

- Testing accuracy using known mixtures.

Model Accuracy and Precision - **baselines**

Standard baseline

- *Not all mixture populations may be present in the baseline. We rely on populations in the baseline to stand proxy for those populations not in the baseline. We rely on “regional structure” to make accurate assignments to a region. Estimate stock compositions to regions only. Reasonably accurate estimate to region (fisheries management applications).*

Optimized baseline

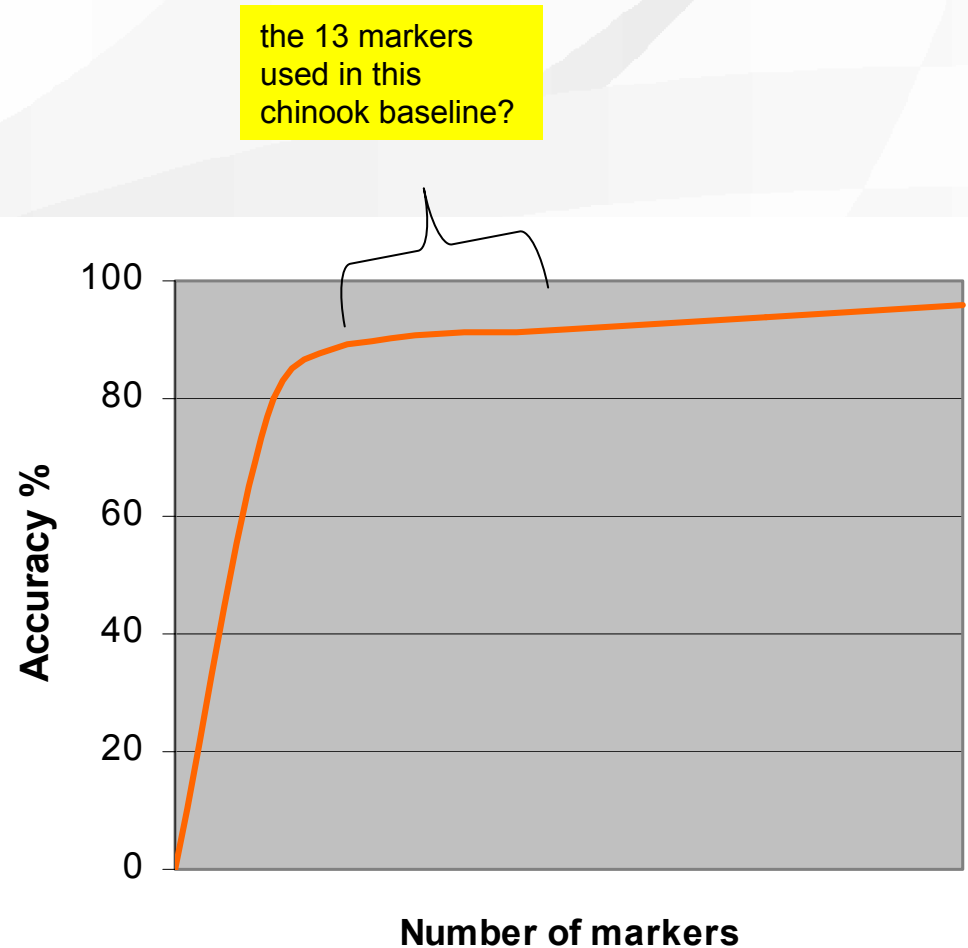
- *Three criteria met: 1) All mixture populations are represented in the baseline; 2) Enough genetic markers used to provide sufficient discrimination between populations; 3) baseline populations have large enough sample size.*
- *Should be able to accurately estimate to population.*

Model Accuracy and Precision - **baselines**

Optimized baseline – Sufficient number of markers to allow separation between populations.

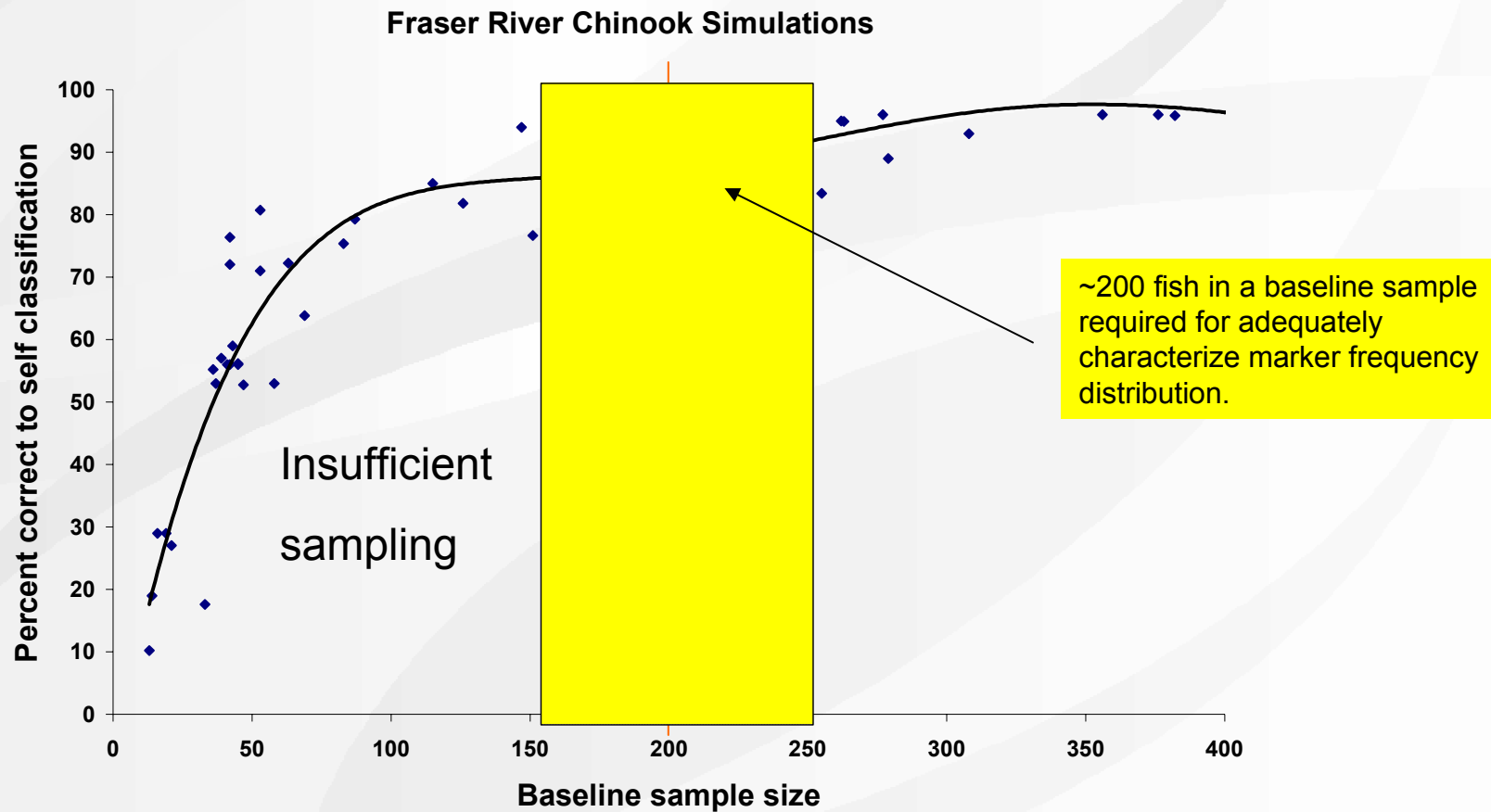
Some considerations:

- not all markers are equal (highly polymorphic markers provide more information)
- adding markers cost more work and more \$.
- certain markers can be run together (multiplex) saves \$.



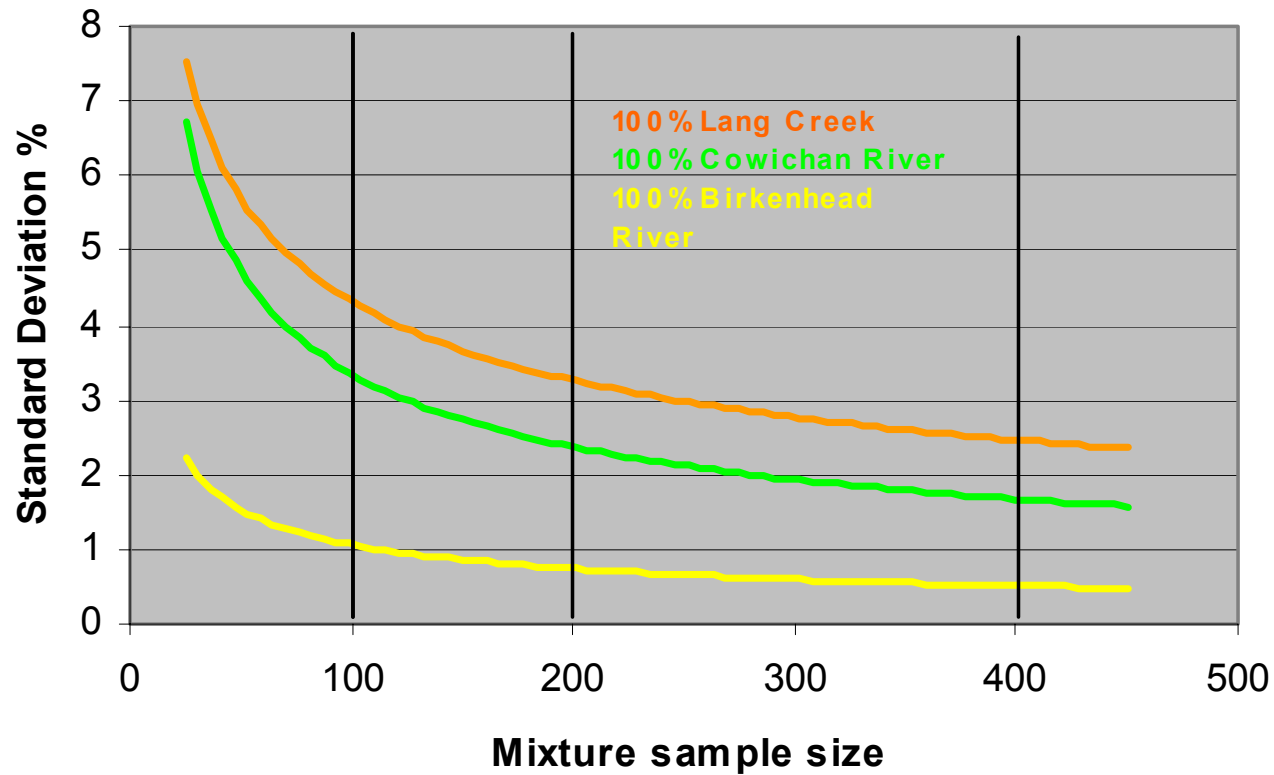
Model Accuracy and Precision - **baselines**

Optimized baseline – Baseline populations have sufficient sample size.



Model Accuracy and Precision - mixtures

- Relationship between mixture sample size and SD
- Simulation for three populations at 100%



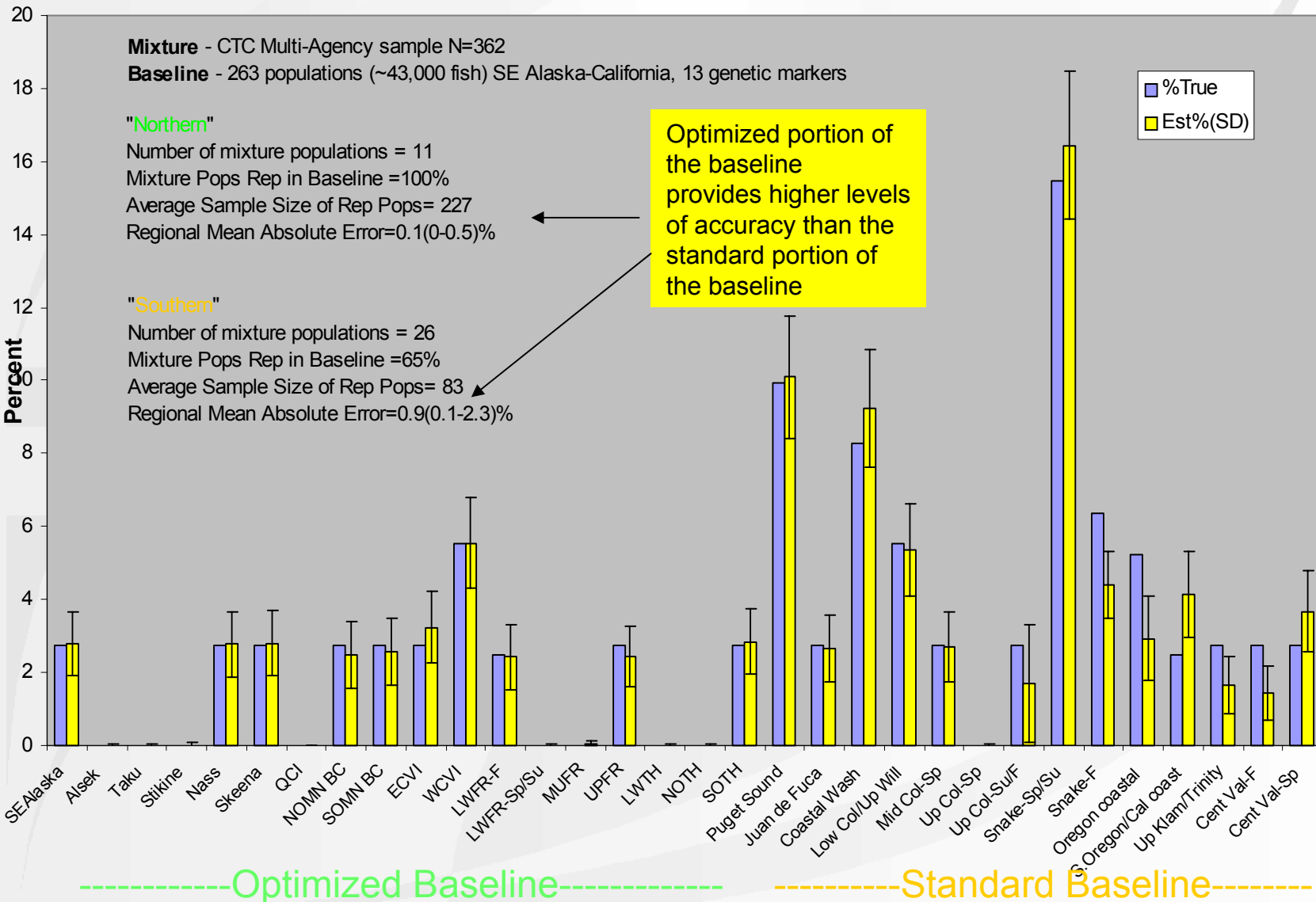
1% decrease in SD with doubling sample size

Known Sample – CTC Multi-Agency Sampling plan

Use 37 populations from SE Alaska to California as known mixture sample.

Region	Numbers of populations	Number of individuals	Source laboratory
Kamchatka to SE Alaska	10	100	ADFG/ABL
British Columbia Puget Sound	10 5	100 50	CDFO WDFW
Columbia River Washington/Oregon Coast	15 5	150 50	WDFW/ NWFSC/CRITFC/ UI/IDFG
California	5	50	WDFW/OSU SWFSC

Example 1 - Model Accuracy (standard and optimized baseline)



Example 1 - Individual assignments from “optimized” portion of baseline – Multi-Agency mixture sample .

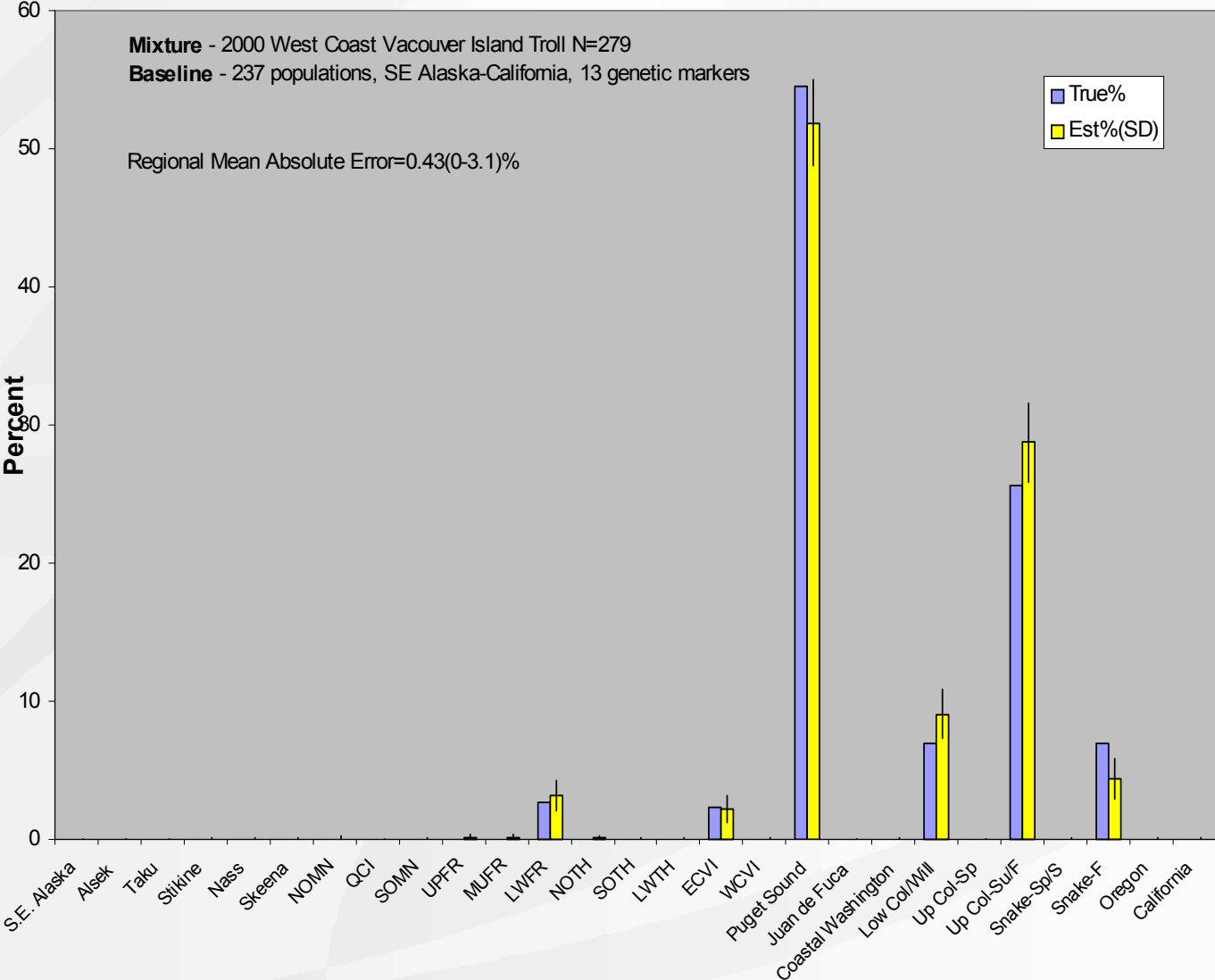
Region	Population	Correct to Population	Correct to Region	Total
SEAlaska	King Salmon	10	10	10
Skeena	Sustut	10	10	10
Nass	Tseax	10	10	10
CCST BC	Klinaklini	8	8	10
NCST BC	Dean	8	8	10
WCVI	Sarita	9	10	10
WCVI	Tahsis	10	10	10
ECVI	Nanaimo	10	10	10
LWFR	Chilliwack	7	7	9
SOTH	Bessette	10	10	10
UPFR	Holmes	9	9	10
Total		101	102	109
Percent		92.7%	93.6%	

For individual assignments take the highest bayes probability associated with a population for each fish.

Example 1 - Individual assignments from “standard” portion of baseline where baseline contains mixture population – Multi-Agency mixture sample .

Region	Population	Correct to Population	Correct to Region	Total
PugetSnd	Skagit	5	5	6
PugetSnd	Green	9	10	10
PugetSnd	White	9	10	10
Coast Wash	Hoh	8	10	10
Coast Wash	Quinault	8	9	10
Low Col/Will	Abernathy	9	10	10
Snake-F	Deschutes_F	6	8	10
UpCol-Su/F	Handford Reach	5	9	10
Snake-Sp/S	South_F_Salmon	10	10	10
Snake-Sp/S	Mid_F_Salmon	8	10	10
Snake-Sp/S	Rapid	10	10	10
Snake-Sp/S	Imnaha	1	10	10
Snake-F	Lyons Ferry	2	2	9
MidCol-Sp	Granite	8	8	10
Low Col/Will	Mackenzie	0	9	10
CentVal_F	Sacr_F	2	4	10
Cen_Val_Sp	Bute	7	9	10
Total		107	143	165
Percent		64.8%	86.7%	

Example 2 – Model accuracy – standard baseline -known CWT mixture sample



Example 3 - Optimized baseline

Fraser chinook migration timing at Albion.

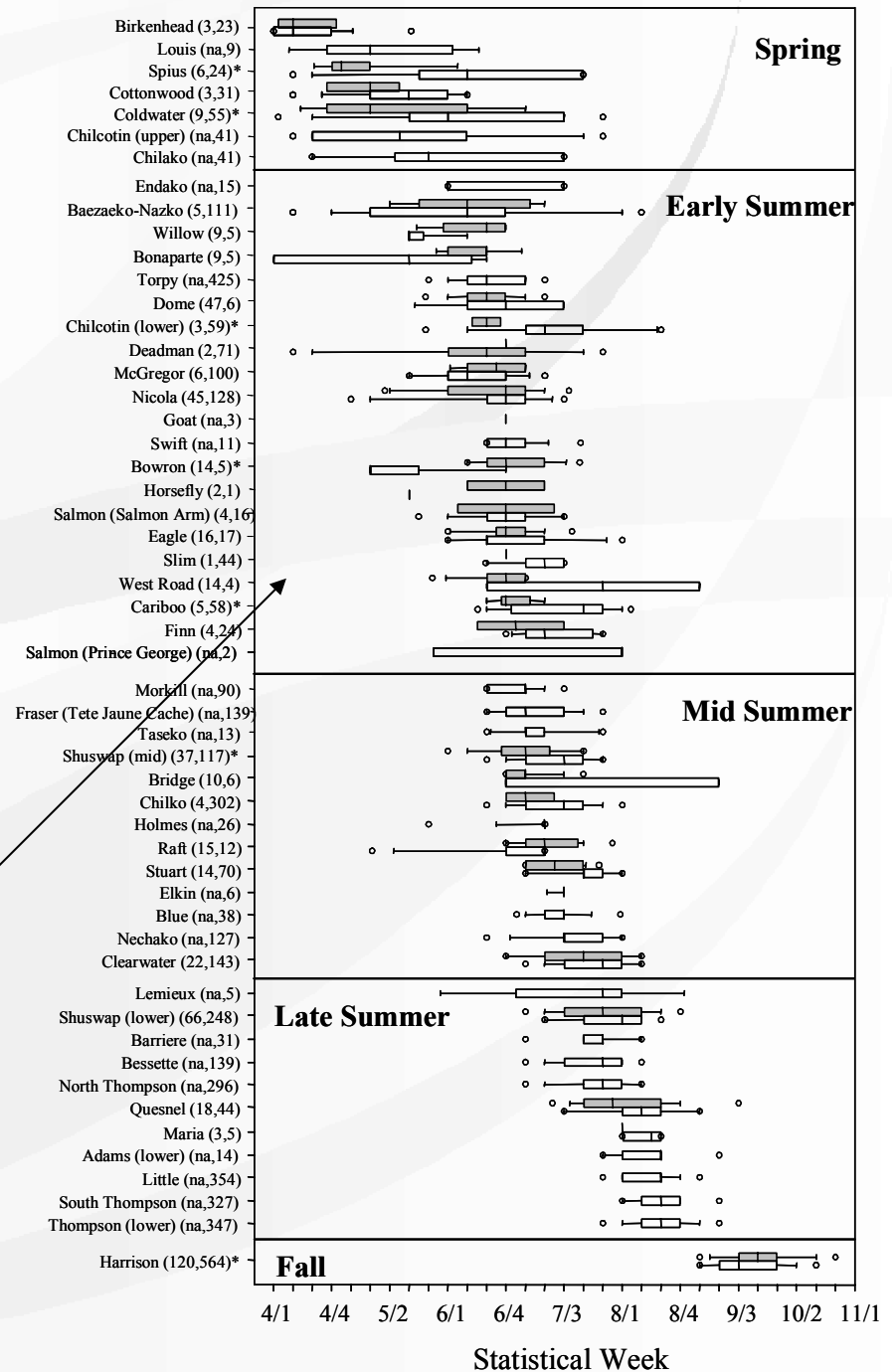
Baseline Fraser 56 populations

Open boxes = 4,822 DNA individual assignments (2000-2001)

Dark boxes = 516 CWT (1987-2003)

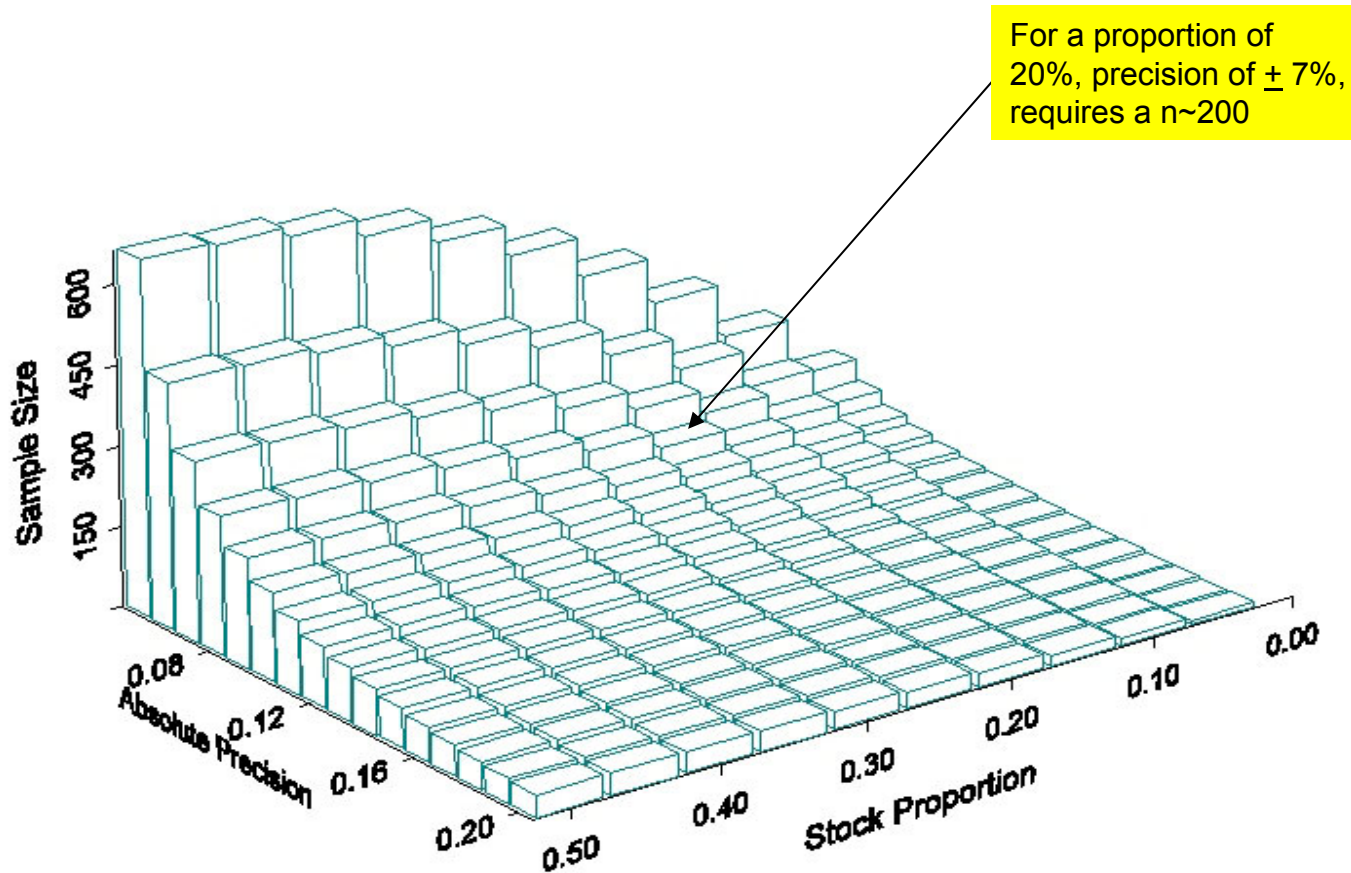
Parken et al. (in prep)

Most of the comparisons between CWT and DNA show no significant difference between median migration timing, where they do we can usually identify inadequacies in the baseline.



DNA implementation considerations – sample size

Use sampling theory to determine precision levels for given sample sizes.

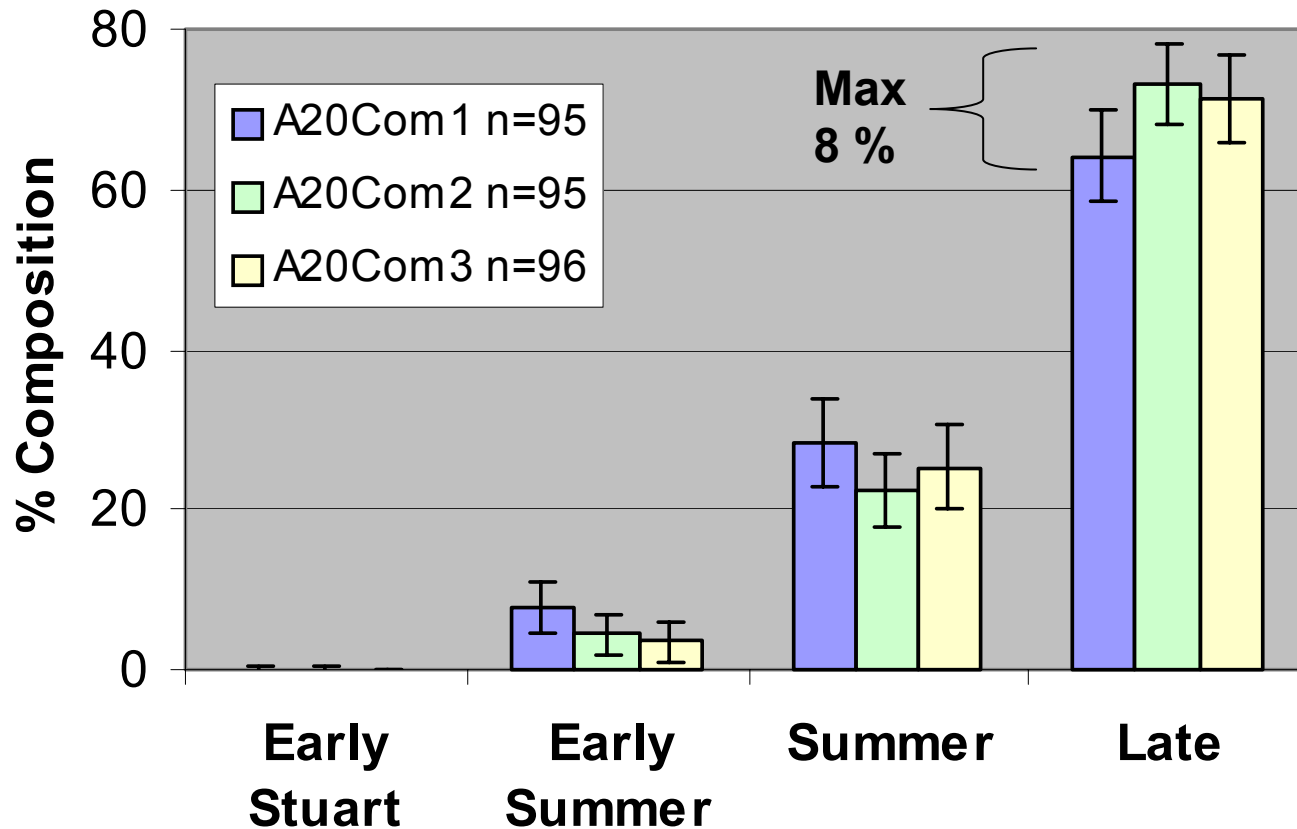


DNA implementation considerations – random vs. non-random sampling.

“Boat Effect” Sockeye example

Three seine boats, all fishing Area20 on August 12, 2003

Chi-squared value = 3.33 (Critical value $p < 0.05 = 9.49$)

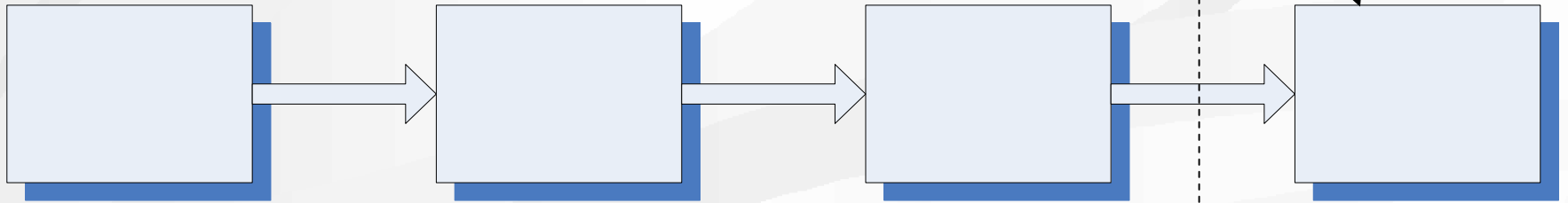


Ease of sampling may allow better sampling design

DNA implementation considerations – information systems.

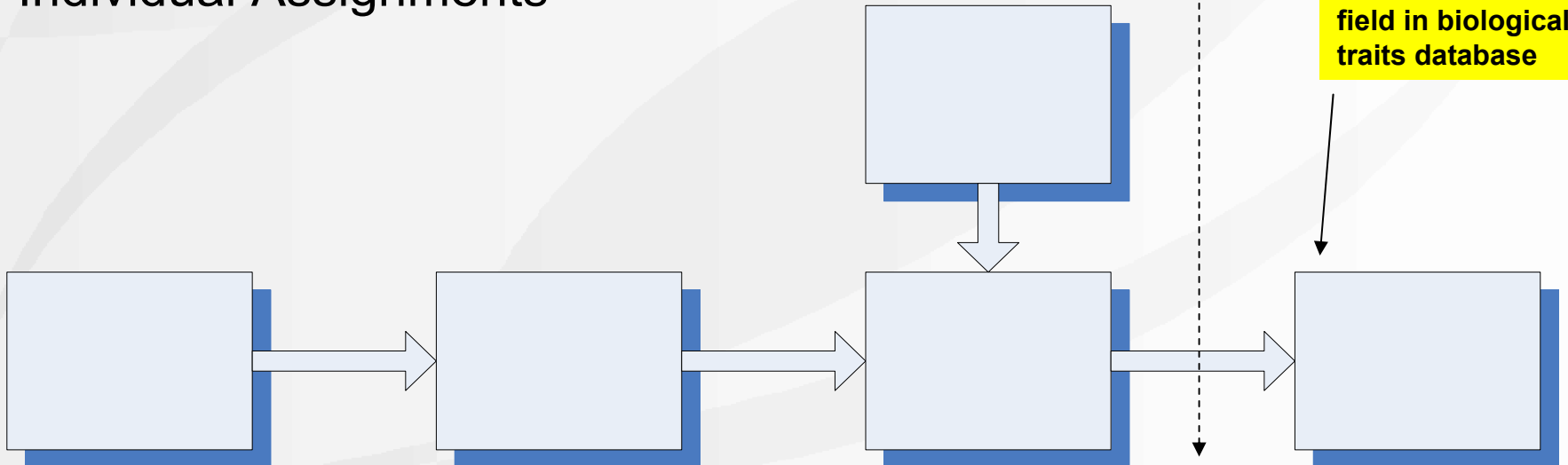
User interface

Mixture Analysis



User may need access to rerun different mixtures

Individual Assignments



Can be another field in biological traits database

DNA cost.

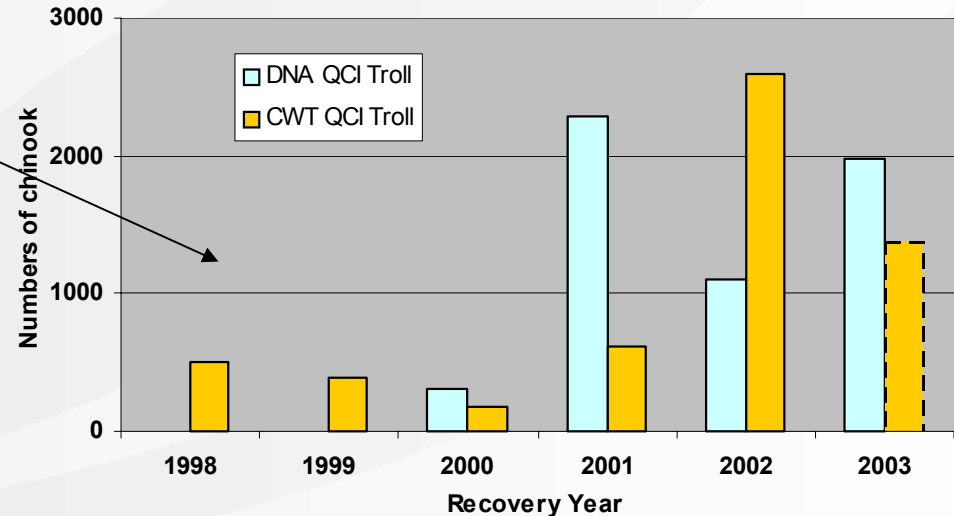
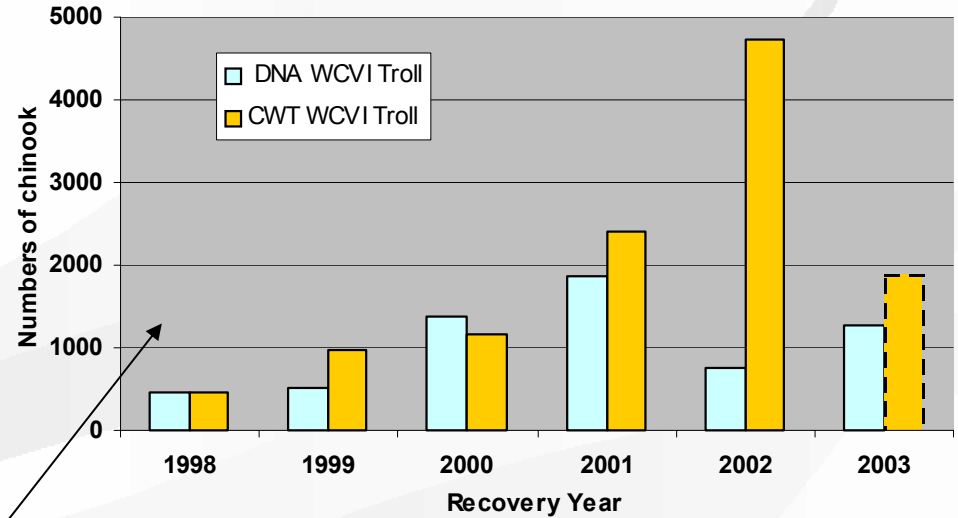
Depends on numbers of samples analyzed:

- Levels of stratification required (area, gear, age, time, marked/unmarked)
- Sampling precision (sampling theory)
- Coordination with Fisheries Management sampling?
- Only interested marked fish and indicator stocks?

DNA data available from BC fisheries – a starting point?

Some years there are more chinook DNA samples analyzed than CWT recovered, most years there is around 1/3 - 1/2 as many DNA samples.

Not all this data has corresponding age



Summary

Benefits/Opportunities using DNA are:

- Collecting information on all populations in the catch, not just the tagged component.
- Regional groups correspond to genetic ESU\MU designations or rolled up into PSC regional groups.
- Optimized baseline will provide accuracy of ~90% for individual populations (individual assignments) (consistent with what we see in sockeye).
- Room for improvement if the baseline can be reduced to indicator populations only.
- Ease of DNA sampling should allow for better random sampling of catch (sampling does not devalue fish).
- Get a handle on sub-legals, non-retention fish by non-lethal DNA sampling.
- Reconstruct historical fisheries using DNA from scales.

Disadvantages using DNA are:

- Cannot separate hatchery and wild from the same population
- Must deal with an estimate of age and stock-of-origin if we want provide a CWT equivalent.