Appendix II
Herring PDT Portside/Sea Sampling Data Analysis

A: Comparison of (landed) Bycatch Estimates from Portside and At-Sea Observer Sampling Programs in the Atlantic Herring Fishery (July 2010)

B: A Comparison of Porstide and At-Sea Sampling Methods of Estimating Bycatch in the Atlantic Herring Fishery

Amendment 5 Draft EIS

September 2011

Comparison of (Landed) Bycatch Estimates
from
Portside and At Sea Observer Sampling Programs
in the
Atlantic Herring Fishery

by

Herring Plan Development Team (PDT)
Matt Cieri (ME Department Marine Resource)
Steven Correia (MA Division of Marine Fisheries)

WORKING DRAFT

1.0 INTRODUCTION

In the past, members of the Herring PDT have estimated total removals of river herring in the Atlantic herring fishery by a combination of portside and at-sea observations. This analysis hinges, however, on the comparability between these two very different methods of documenting bycatch.

Estimates and frequency of occurrence of bycatch in the sea herring fishery is monitored by two independent programs: Maine DMR and Massachusetts Marine fisheries joint Portside sampling program and NOAA' National Observer Program. There are three estimates that are worth comparing:

1) Proportion of trips with occurrences of species
2) The amount of agreement on occurrences of species within trips
3) The amount of agreement on catch weight estimates between the two methodology

This analysis compares the total estimated catch weight for bycatch species for trips that were sampled by both a portside sampling program and the Northeast Fisheries Observer Program (NEFOP).

2.0 SAMPLING METHODS

2.1 AT-SEA OBSERVATIONS

During at-sea operations, NEFOP observers use basket sampling to document occurrence of other species during targeted Atlantic herring and mackerel trips on a haul by haul basis and during normal fishing operations. These non-target species are then included in the data as retained or "Kept" (http://www.nefsc.noaa.gov/fsb/Manuals/JANUARY%202010%20MANUALS/NEFOPM_010110_BOOKMARKS_LONG1.pdf). Normally, ten 50 lbs basket sub-samples are taken at regular intervals during the pumping process from net to hold. These samples are then checked for bycatch, weighed and measures, and the results expanded based on the captains' estimate of that hauls total weight. Because the Atlantic herring fishery is a high volume fishery, much of the bycatch is retained during the pumping process; particularly so for co-occurring pelagic species such as river herring. However, observers do hand select larger bycatch species. In these cases, these species are listed as "discarded" in the database if they are not retained by the crew

2.2 PORTSIDE OBSERVATIONS BY MA DMF

Sampling methodology in the MA DMF portside sampling program attempts to be consistent with NOAA Observer Program protocols, with some modifications to decrease variance in extrapolation of bycatch estimates and reduce potential sampling bias. Due to the large quantities of fish that are typically landed in these fisheries, sub-sampling will be required. Sub-sampling is used when the volume of fish that the sampler is attempting to quantify is too large to obtain actual weights or if the amount of by-catch is too abundant. During sub-sampling, the sampler will collect smaller batches of fish, sort and weigh by species and then extrapolate to the total catch. All sub-sample weights will be actually weighed (actual weight), and hail weights (for both truckloads and fishing vessels) will be acquired from the plant managers or vessel's captain and therefore estimated (estimate weight).

In most situations, sampling is conducted over the entire offloading period to capture any stratification that may occur throughout the entire fishing activity (e.g. while being pumped aboard while out at sea, due to the difference in species size and composition between tows, settling in the vessel's holding tanks, etc.). Because the catch is not unloaded the same way at every dealer and plant, sampling techniques will vary. Typically samples will be collected systematically at set intervals with predetermined sample sizes. All samples are sorted by species and actual weights will be taken. Lengths will be taken according to the NOAA Observer Program species priority list by statistical area. Haddock, alewife, blueback herring, and American shad have been specified as specific species of concern by MA DMF and therefore if available, the number of lengths taken will be 200 per trip. Two length frequency samples will be randomly selected, one during the first half and the second during the second half of the offloading period.

Below is MA DMF's description of the sampling protocol at a processing plant. The majority of sampling occurs at these types of off loading facilities for this project.

Processing Plant

Sampler should position himself at the discard vat where all bycatch and damaged fish are deposited. The sampler must position themselves in a location that is safe and will not disrupt plant operations. The name of the vessel should be recorded and hail weight, date landed, and general location fished (statistical area, known piece of bottom, etc.) should be collected from the plant manager or vessel captain. Hail weight should be confirmed after unloading process is complete and all fish have been processed. A processing rate (kg of catch processed/minute) should be calculated by dividing hail weight by the time it took to offload the vessel. When calculating time to off load catch, note time spent not pumping/processing, such as coffee or lunch breaks and processing hold-ups. To eliminate bias caused by periodicity, prior to the beginning of the offloading process, the sampler will use a random number table and pick a random start time between 1 and 30 minutes. Once the start time has been determined, a basket will be positioned in the discard vat and a sample will be collected. Once the basket has been filled, it will be weighed, sorted by species, and then weighed by species. Lengths will be collected according to NOAA Observer Program sampling protocols. This process will be repeated for thirty minutes until the sub-sampling period has been completed. If fish being sent to the bycatch vat is too abundant and sampler cannot weigh all fish being sent to the discard vat, then sub-sampling may be required to get an estimate of total bycatch per 30-minute sampling period. This sampling process will be repeated every other 30-minute interval during the entire pump offloading process. After the offload process, the sampler should consult with the plant's quality control personnel to obtain an accurate, by the box, quantification of species being processed. Lastly, to verify pump rates and landings estimates, the sampler should obtain a report of landings and processed fish from the plant manager after the off loading is complete.

2.3 PORTSIDE OBSERVATIONS BY ME DMR

For the ME DMR portside sampling program, the samplers collect and quantify all bycatch from individual lots of fish (transported by trucks or vessels) that enter the processing facilities. Samplers position themselves at the point of entry into the facility along an assembly line or at the base of the hoppers where the fish are unloaded. Sampling is conducted before grading or sorting of the catch occurs. All bycatch is removed from the assembly line or hopper and placed in bushel baskets or buckets specific to each species. The total weight of any observed bycatch is recorded along with species identification, total species weight, individual lengths and weights of all fish according to a NMFS and ACCSP specified protocol. If there is a large amount of one species, the total weight is recorded and then length frequencies and weight are gathered from a sub sample of n=50. The information collected for each bycatch study is recorded on the data sheets (see "Data Sheets" section of packet) and entered into the DMR biological database.

A sub-sampling protocol is sometimes utilized when sampling a large volume of catch. Instances where this is likely to occur include sampling sites where vessels land an entire catch (as much as one million pounds) to a single facility. Sub-sampling is also appropriate in instances when there is an overwhelming amount of bycatch and/or non targeted species mixed in with the lot of fish. In these cases it can be impossible to use the complete sampling protocol regardless of the amount inspected (< 80,000 lbs.). These situations are likely to occur when vessels are fishing mixed groups of herring and mackerel, some of which have a 50-50 composition.

Sub-samples are to be collected using bushel baskets at timed intervals during the pumping or unloading process following the NMFS at-sea observer sampling protocol. To accomplish this type of sub-sampling, one needs to know the total lot weight and the duration of time it will take to unload the catch. After sampling, the bushel basket of fish should be sorted by species, and total weight of each species and length frequencies should be recorded (sub sample n=50, for length frequencies if more than fifty of any species occurs).

Example:

Lot size = 120,000 lbs (3 Trucks)
Pumping or unloading time = 3 hours (180 minutes)

If a sample basket is to be collected for every 10,000 lbs of fish, then **12 sample baskets** need to be collected over the entire pumping or unloading process.

120,000 lbs/10,000 lbs = 12

If the entire pumping or unloading process takes an estimated 180 minutes, then **a basket sample needs to be taken every 15 mins.**

If the catch composition from the bushel baskets is 99% Atlantic Herring, then one can extrapolate that out of the 120,000 lbs unloaded, then 118,800lbs is Atlantic Herring.

99% Atlantic Herring = 120,000 lbs x 0.99 = 118,800lbs of Atlantic Herring

If the remaining 1% of the catch composition is Atlantic Mackerel, then one can extrapolate that out of the 120,000 lbs unloaded, 1,200lbs is Atlantic Mackerel

1% Atlantic Mackerel = 120,000lbs x 0.01 = 1,200lbs of Atlantic Mackerel

3.0 STATISTICAL ANALYSIS

For this analysis, data were gathered from the various projects by either request or direct querying of the data. In some cases, vessel trip report number was not available, and so trips between portside and at sea programs we matched by hand.

Several species were pooled into a species grouping because of potential for mis-identification or to make the analysis easier to understand. River herring group consisted of alewives, bluebacks and herring unknown were grouped as river herring. American shad and hickory shad were grouped as shad. Long-fin squid, short-fin squid and squid unknown were grouped as squid.

The analysis compares the number of occurrences of bycatch species by sampling method using a paired t-test. The binomial exact test was also used to check whether the probability of number of occurrence of bycatch in port sampling exceeding the number of occurrence in the observer sampling differed from 0.5.

The analysis compares the proportion of trips containing a particular species groups using Wald test with correction for continuity. Fisher's exact test was used to convert the differences into odds ratios. The test was conducted on the seven species groups with the highest percent occurrence: river herring, squid, silver hake, spiny dogfish, butterfish haddock, and shad. The family-wise error rate for multiple comparisons was not corrected.

The percent agreement for presence/ absence of species group was measured for both sampling methods using two indices of similarity. The first index was a simple matching index constructed by dividing the total the number of trips with joint presence and joint absence for both sampling methods by the total trips. In this index, joint absence (double zeros) contributes to similarity. However, the absence of a species group from both sampling methodology could be due to the trip occurring in an area or time where the species are not present, and inflating the index. To address joint absences, the Jaccard coefficient was used: the number of trips with joint presence divided by the number of trips with joint presence and the two unique combinations of present in one method and absent in the other. The joint absences do not contribute to similarity in the Jaccard index. This method was applied to seven species groups: river herring, squid, silver hake, spiny dogfish, butterfish, haddock and shad.

The relationship between the observer and portside estimates of landed weight of bycatch species was assessed using Pearson's product moment correlation coefficient. Agreement was tested between port and observer trip landings estimates using a paired t-test. T-tests were performed for all trips, trips without joint absences, and log transformed for trips without joint absences. Assumption that differences were distributed normally was assessed using quantile-quantile normal plots and Shapiro test for normality.

The following summarizes the PDT's questions and methodology for statistical evaluation of the portside/at-sea data:

1. **Is the frequency of detection of bycatch species similar for portside and observer program?**
    a. Paired T-test for number of occurrences for portside and observer
    b. Exact binomial test for the probability of occurrence portside versus observer
2. **Does the estimate of percent occurrence differ between sampling methods for each bycatch species?**
    a. Test difference in proportions among methods using Wald's statistic with correction for continuity
    b. Get odds ratio using Fisher's exact test
3. **Describe similarity of occurrence of species by tows**
    a. Matching index (% agreement)
    b. Jaccard index (% agreement excluding joint absence)
4. **Does the estimation of bycatch weight differ by method?**
    a. Correlation between paired estimates by method
    b. Paired T-tests for differences in trip estimates by sampling methodology
    c. Provide estimates of total weight of landed bycatch with 95% confidence interval for each method

4.0 PRELIMINARY RESULTS (WORK IN PROGRESS)

A total of 52 trips were sampled with both portside and at sea observer sampling between 2005 and 2009 (Table 1). The number of trips containing bycatch species groups by sampling methodology is shown in Table 2, and the number of trips as a proportion of total trips is shown in Table 5.

The number of occurrences of bycatch species by methodology (at-sea versus portside) was significantly different (Table 3). Port sampling averaged 1.9 more occurrences than the observer program. The exact binomial test indicated that the probability of a species occurring portside versus at sea was significantly greater than 0.5, suggesting non-random effects (Table 4).

For the seven most frequently caught bycatch species, the Herring PDT compared the proportion of trips with observed bycatch by methodology using Wald test statistic without adjustment for multiple comparisons (Table 6). Overall, the proportions of trips with a particular species were significantly different for squid and for spiny dogfish only, with the portside sampling method having higher proportions than the observer.

Similarity index for presence/ absence of species is presented in Table 7. Similarity indices were relatively high for the simple matching coefficient (mean: 0.72, range: 0.54 to 0.87), but tended to be low for the Jaccard coefficient (mean: 0.30, range: 0.17 to 0.54). The joint absences influence the similarity indices, and the true similarity is bounded by these two values. Further work needs to be done to separate joint absences that reflect no occurrences in strata where the species occur from joint absences in strata where the species in not likely to occur.

Scatterplots of paired portside and observer estimates for eight species are shown in Figure 1. The paired comparisons indicate little relationship between weight estimates from the Portside and Observer projects. Correlation coefficients for these eight species are exhibited in Table 8. The correlations coefficients for 7 of the 8 species were low and not significantly different from zero. Correlation coefficient was moderately high (0.80, 0.79) and significantly different from zero for spiny dogfish. The correlation coefficient was highly influenced by one trip where both methods had high estimates of catch. The correlation coefficient estimated without this pair was low and not significantly different from zero.

Bland-Altman plots of the paired landings estimate between methods are shown in Figure 2. Variation is high, and differences are larger as might be expected given the low correlation between observer and paired estimates. The distribution of paired differences was significantly different from normal and was strongly leptokurtic with more observations in the middle and tails for the full dataset and for the dataset without joint absences. Only shad with removal of trips with joint absence were not significantly different from normal. A Bland-Altman plot of the log-transformed dataset is shown in Figure 3. This dataset does not include the joint absences. Distribution of paired differences for log transformed data were not significantly different from normal except for spiny dogfish ($p<0.01$). Paired T-test results are provided in Table 9 and Table 10. No differences were significant for untransformed data, which is not surprising given the large variances. Paired differences were not significant for the log transformed data except for spiny dogfish (0.02) and haddock (p=0.04). For non-significant tests, the confidence intervals were wide, indicating low power to detect differences. Spiny dogfish trip estimates from the observer sampling averaged 12% of the portside sampling estimates. Haddock trip estimates from the observer sampling averaged 5% of the portside sampling estimates.

Total estimates with 95% confidence intervals of landed catch by species and sampling method are shown in Table 11 and Table 12. Table 11 uses parametric statistic to derive 95% confidence interval and Table 12 uses bootstrap percentiles to estimate 95% confidence limits. These estimates were expanded using the trip estimates. They are only useful for comparing the estimates across sampling methods. As expected, confidence limits are wide. Note that estimates from the fishery would include stratification by month, area and gear types will improve precision.

5.0 PRELIMINARY CONCLUSIONS (WORK IN PROGRESS)

Portside and at sea sampling are two very different approaches to document bycatch in the directed Atlantic herring fish. During at-sea sampling observers have the ability to document discarded fish at sea and sample them. During portside operations, samplers cannot do so. However portside samplers have a much more stable platform, better working conditions and more time for a thorough examination.

The Herring PDT examined 52 trips which were sampled by both at sea and portside methods to test if both projects are similar in the amount and species composition detected. The PDT found large differences in retained bycatch between the two programs. More specifically, the portside sampling documented more occurrences of species, and a greater proportion of trips containing key bycatch species. However, at sea observation, when extrapolated to the entire retained weight, shows much higher weights of the more prevalent species. The lack of significant differences in many of the statically approaches taken here are a direct result of low sampling sizes. More co-occurring trips are needed by strata (gear type, sample mythology, area, quarter, and year) to detect significant differences; especially for species which occur infrequently in sampling. The analysis was further hampered by the number of co-occurring trips with either had no retained bycatch at all, or no bycatch of a particular species being tested.

It should be noted that the PDT is not suggesting one project or method is more useful or more accurate than the other. The PDT is, however, suggesting that pooling these two different methods of documenting bycatch may not be possible without further analysis and sampling. The PDT recommends a more thorough examination of both portside and at-sea observations to see if elucidation of these differences (and possible mathematical correction) is possible. By focusing on increasing the number of co-occurring trips statistical analysis may lead to increased comparability by analysis of the methods employed by both projects.

Summary of Herring PDT Conclusions to Date (Work in Progress)

1. Portside sampling method had more occurrences of bycatch than observer method. Proportion of occurrences in portside sampling is greater than at-sea observer sampling; and was significantly different from 0.5
2. The proportion of trips containing a bycatch species was not significantly different between Portside and Observer methods except for squid and spiny dogfish. Both of those species were significantly different
3. Relatively low levels of agreement of occurrences particularly with the Jaccard index.
4. No correlations between paired portside and observer trip estimates of weight
5. Paired T-test on log transformed estimates found no significant differences except for spiny dogfish and haddock. However, high variation in paired estimates lead to a loss of statistical power; and therefore the results cannot be taken as valid

Summary of Herring PDT Advice: Need to Examine Data to Find Sources of Variation (Work in Progress)

A. High variability in trip estimates in both the portside and observer sampling
B. Different methods for expanding within trip samples to trip estimates
C. Sampling design issues

Year	Quarter	Purse seine	Midwater trawl	Paired Midwater trawl
	1	0	0	3
	2	0	0	1
	3	1	0	2
2005	4	0	1	2
	1	0	0	0
	2	0	0	0
	3	0	1	1
2006	4	0	0	0
	1	0	0	1
	2	0	0	0
	3	0	0	0
2007	4	0	0	0
	1	0	1	2
	2	2	0	2
	3	3	0	1
2008	4	0	0	7
	1	0	0	4
	2	5	0	4
	3	3	0	4
2009	4	0	0	1
Total trips		14	3	35

Table 1 Count of trips sampled by both Portside and At Sea Observer Programs by gear type, year and quarter

Species Group	Purse seine		Midwater trawl		Paired Midwater trawl	
	Observer	Portside	Observer	Portside	Observer	Portside
River herring	2	3	2	1	15	20
Squid	2	6	1	2	10	19
Silver hake	3	6	0	2	12	15
Spiny dogfish	4	8	0	2	4	14
Butterfish	0	0	1	0	5	9
Haddock	0	0	0	1	4	10
Shad	0	0	0	1	5	8
Red hake	0	0	0	1	0	6
American plaice	0	1	0	1	0	3
Longhorn sculpin	1	0	0	0	0	2
Redfish	0	0	0	0	1	2
Cod	0	0	0	1	0	1
Fish unk	1	0	0	0	1	0
Lumpfish	1	0	0	0	0	1
Shrimp	0	1	0	0	0	1
Cunner	0	0	0	0	1	0
Cusk	0	0	0	1	0	0
Little skate	0	0	0	0	0	1
Menhaden	0	0	0	0	0	1
Pollock	0	0	0	0	0	1
Scup	0	0	0	0	0	1
Sea raven	0	0	0	0	0	1
Winter flounder	0	0	0	0	0	1
Number of trips	14	14	3	3	35	35

Table 2 Count of trips containing bycatch by species group, gear type and sampling program

Mean difference	95% confidence interval for mean difference	P-value	Degrees of Freedom
-1.9	-2.6 to -1.11	<0.001	41

Table 3 Summary of paired t-Test for number of occurrences of bycatch species by sampling methodology for in trips

Does not include trips with joint absence

Number of occurrences Port occurrences > observers	Proportion	95% confidence interval	Probability that Proportion is not different from 0.5
35	0.83	0.69-0.93	<0.001

Table 4 Summary for exact binomial test of number of occurrences of Port> Observer in number of occurrence of a bycatch species

Tests whether the true probability of Port occurrences > observer occurrences is not different from 0.5.

Species Group	Purse seine		Midwater trawl		Paired Midwater trawl	
	Observer	Portside	Observer	Portside	Observer	Portside
River herring	0.14	0.21	0.67	0.33	0.43	0.57
Squid	0.14	0.43	0.33	0.67	0.29	0.54
Silver hake	0.21	0.43	0.00	0.67	0.34	0.43
Spiny dogfish	0.29	0.57	0.00	0.67	0.11	0.40
Butterfish	0.00	0.00	0.33	0.00	0.14	0.26
Haddock	0.00	0.00	0.00	0.33	0.11	0.29
Shad	0.00	0.00	0.00	0.33	0.14	0.23
Red hake	0.00	0.00	0.00	0.33	0.00	0.17
American plaice	0.00	0.07	0.00	0.33	0.00	0.09
Longhorn sculpin	0.07	0.00	0.00	0.00	0.00	0.06
Redfish	0.00	0.00	0.00	0.00	0.03	0.06
Cod	0.00	0.00	0.00	0.33	0.00	0.03
Fish unk	0.07	0.00	0.00	0.00	0.03	0.00
Lumpfish	0.07	0.00	0.00	0.00	0.00	0.03
Shrimp	0.00	0.07	0.00	0.00	0.00	0.03
Cunner	0.00	0.00	0.00	0.00	0.03	0.00
Cusk	0.00	0.00	0.00	0.33	0.00	0.00
Little skate	0.00	0.00	0.00	0.00	0.00	0.03
Menhaden	0.00	0.00	0.00	0.00	0.00	0.03
Pollock	0.00	0.00	0.00	0.00	0.00	0.03
Scup	0.00	0.00	0.00	0.00	0.00	0.03
Sea raven	0.00	0.00	0.00	0.00	0.00	0.03
Winter flounder	0.00	0.00	0.00	0.00	0.00	0.03
Number of trips	14	14	3	3	35	35

Table 5 Counts of trips with occurrence of bycatch as proportion of total trips by species group, gear type and sampling method

Species group	Port sampling Proportion	Observer Proportion	95% confidence interval on difference	Odds ratio	95% confidence interval on odds ratio	Probability of odds ratio
River herring	0.46	0.37	-0.11 - 0.30	1.48	0.63 - 3.52	0.42
Butterfish	0.17	0.12	-0.10 - 0.21	0.58	0.46 - 5.94	0.58
Squid	0.52	0.25	0.07 - 0.47	3.20	1.31 - 8.14	<0.01
Silver hake	0.44	0.29	-0.05 - 0.36	1.94	0.81 - 4.79	0.15
Spiny dogfish	0.46	0.15	0.12 - 0.49	5.08	1.72 - 13.71	<0.01
Haddock	0.21	0.08	-0.02 - 0.29	3.18	0.86 - 14.7	0.09
Shad	0.17	0.10	-0.72 - 0.23	1.95	0.54 - 8.03	0.39

Table 6 Comparing the differences in proportion of trips with species in observer and portside trips for all gear types

Test is two sided.

Species group			Observer sampling +	Observer sampling -	Matching coefficient	Jaccard coefficient
River herring	Port	+	15	9	0.75	0.54
		-	4	24		
Butterfish	Port	+	4	5	0.87	0.36
		-	2	41		
Squid	Port	+	8	19	0.54	0.25
		-	5	20		
Silver hake	Port	+	9	14	0.62	0.31
		-	6	23		
Spiny dogfish	Port	+	6	18	0.62	0.23
		-	2	26		
Haddock	Port	+	3	8	0.83	0.25
		-	1	40		
Shad	Port	+	2	7	0.81	0.17
		-	3	40		

Table 7 Count of trips with species groups present (+) or absent (-) by sampling method and two measures of percent agreement between methods

Species group	All trips	Excludes trips with double-zeros.
River herring	-0.04	-0.13
Squid	0.06	-0.01
Silver hake	0.22	0.17
haddock	-0.02	-0.23
Spiny dogfish[1]	0.80	0.79
Spiny dogfish[2]	0.06	-0.08
Butterfish	0.25	0.12
Shad	-0.04	-0.30

Table 8 Pearson's product moment correlation coefficients for observer and portside estimates of landed weight

[1.] Correlation coefficients are significantly different from 0 at P=0.05, but correlation coefficients are highly influenced by one trip.

[2] Removing influential points lowers correlation coefficients to not significantly different from zero.

Species group	Mean difference	95% Confidence interval for mean difference	P-value	Degrees of freedom
		All trips		
River herring	1242.9	-131.4 - 2,617.2	0.08	51
Squid	-4.3	-98.1 - 89.6	0.93	51
Silver hake	57.7	-176.1 - 291.6	0.62	51
Spiny dogfish	57.8	-94.7 - 210.4	0.45	51
Butterfish	-158.1	-480.2 - 164.0	0.33	51
Haddock	-22.2	-206.9 - 162.6	0.81	51
Shad	21.1	-39.9 - 82.2	0.49	51
		Without trips with joint absence (double zeros)		
River herring	2308.3	-248.5 - 4,865.1	0.07	27
Squid	-7.0	-162.9 - 148.9	0.93	31
Silver hake	103.5	-326.8 - 533.9	0.63	28
Spiny dogfish	118.0	-203.2 - 439.2	0.46	24
Butterfish	-747.3	-2,439.2 - 944.5	0.35	10
Haddock	-96.1	-1,002.2 - 809.9	0.82	11
Shad	91.6	-203.7 - 386.9	0.51	11

Table 9 Summary of paired T-test for estimates of trip catch by sampling method (observer-port)

Upper table uses all 52 trips. Bottom table does not include trips with joint absence.

Species group	Mean difference	95% Confidence interval for mean difference	P-value	Degrees of freedom
		All trips		
River herring	2.68	0.46 - 15.58	0.26	27
Squid	0.78	0.23 - 2.64	0.69	31
Silver hake	0.62	0.17 - 2.26	0.45	28
Spiny dogfish	0.12	0.02 - 0.68	0.02	24
Butterfish	1.21	0.15 - 9.80	0.84	10
Haddock	0.05	0.00 - 0.91	0.04	11
Shad	0.79	0.04 - 15.70	0.86	11

Table 10 Back-transformed summary of paired T-test for estimates of log trip catch by sampling method (observer-port)

Analysis does not include trips with joint absence by both sampling methods. Back transformed values are ratio of observer estimate to port sampling estimate.

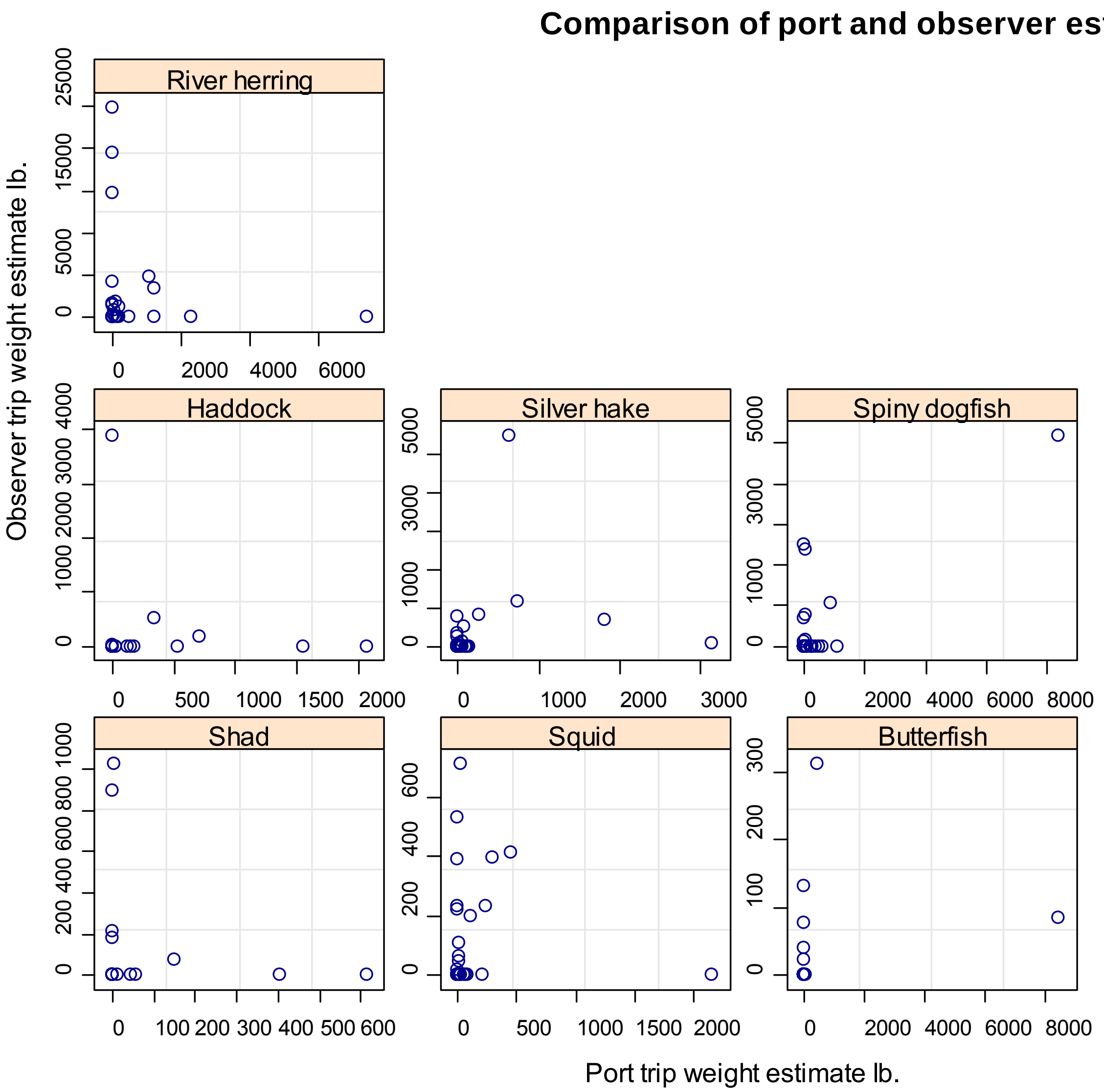

Figure 1 Scatterplot of Observer weight against Portside weight

Note that x and y scales differ among panels.
Plot includes estimates where both port and observer estimates are zero.

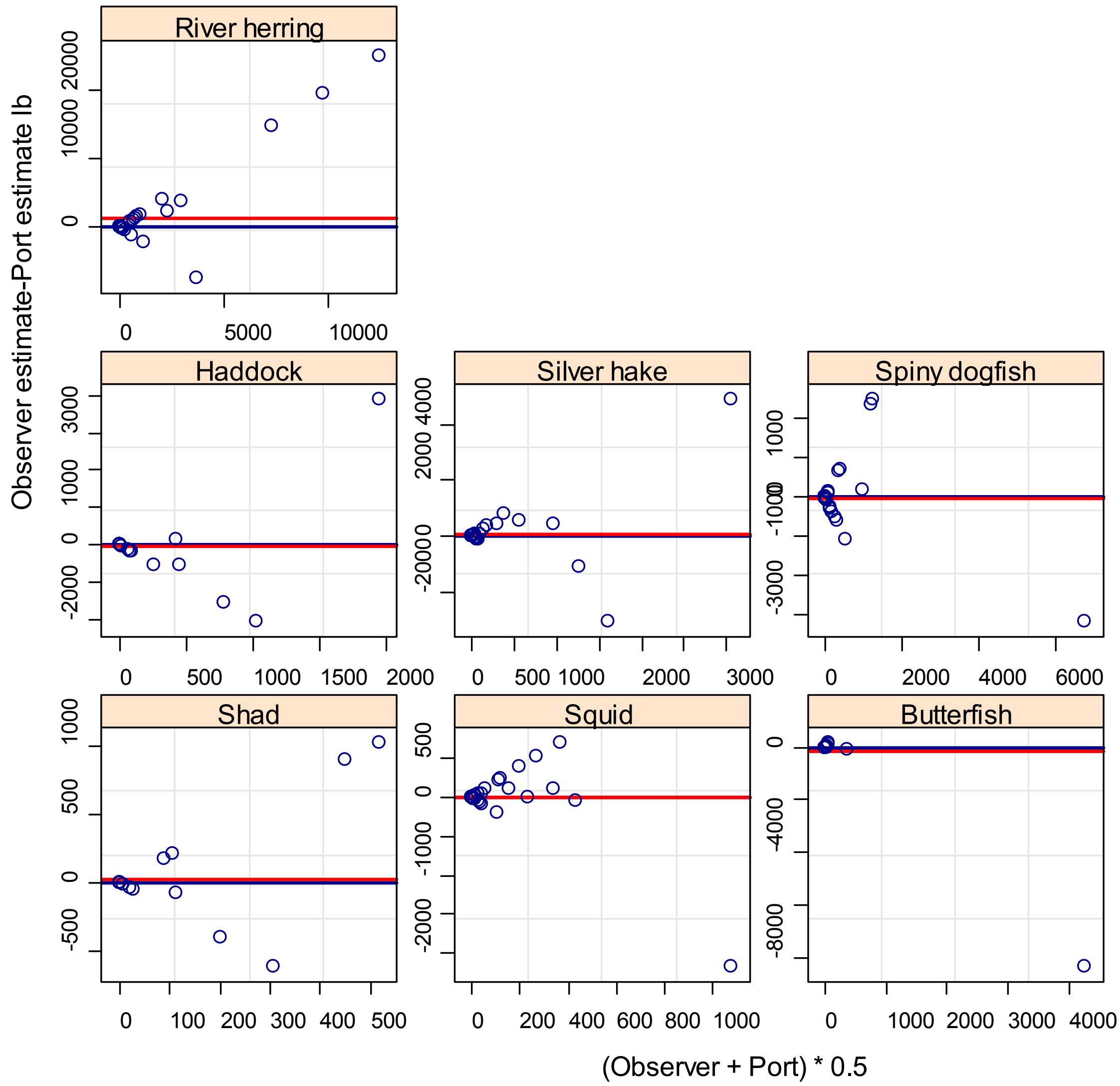

Figure 2 Bland-Altman plot of paired estimates of landings

Redline is average difference. Blue line indicates 0. Dataset includes all trips including joint absence.

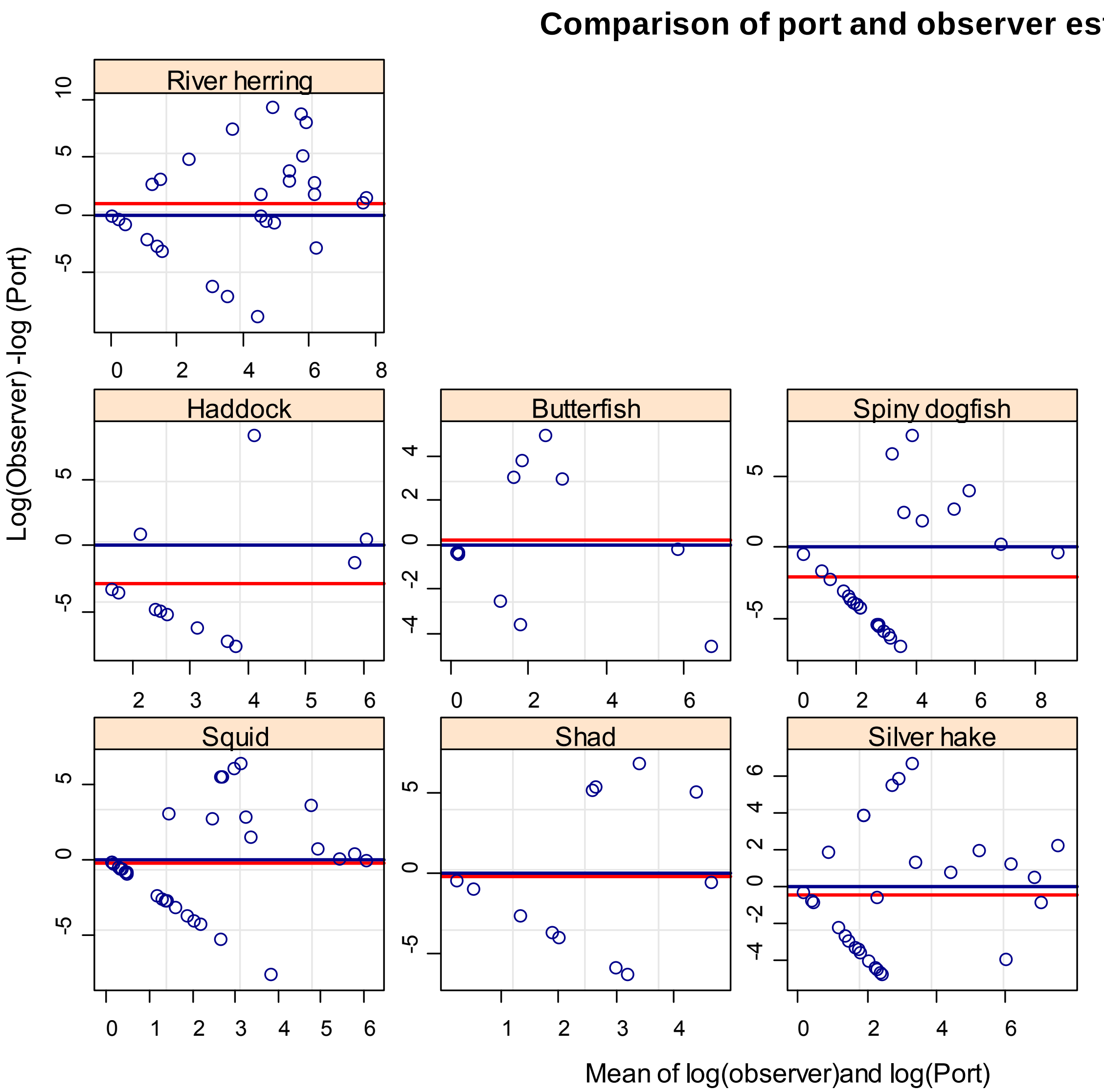

Figure 3 Bland-Altman plot of paired estimates of log landings

Redline is average difference. Blue line indicates 0.   Dataset does not include trips with joint absence.

Species grouping	Total portside estimate (lb)	95% confidence interval		Total observer estimate lb	95% confidence interval	
Fish unk	0	0	0	100,000	-64,784	264,784
River herring	14,695	-1,030	30,420	79,327	10,313	148,341
Spiny dogfish	13,076	-3,821	29,973	12,852	379	25,325
Silver hake	7,372	5	14,739	10,375	-1,057	21,806
Haddock	5,743	364	11,122	4,590	-3,264	12,443
Butterfish	8,888	-8,023	25,798	667	-39	1,373
Squid	3,769	-687	8,225	3,546	1,295	5,797
Cunner	0			4,864	-4,901	14,629
Shad	1,288	-193	2,769	2,387	-359	5,133
Scup	1,667	-1,679	5,012	0		
Redfish	43	-38	124	210	-212	632
Red hake	238	-36	512	0		
Pollock	160	-161	482	0		
Longhorn sculpin	6	-5	17	54	-54	162
American plaice	35	-5	76	0		
Cod	17	-7	41	0		
Lumpfish	9	-9	27	6	-6	18
Winter flounder	12	-12	36	0		
Shrimp	4	-4	12	0		
Menhaden	3	-4	11	0		
Sea Raven	3	-3	10	0		
Cusk	3	-3	8	0		
Little skate	2	-2	5	0		

Table 11 Estimates of total landings in weight with 95% confidence intervals derived from Observer and Portside sampling for 52 trips

Total Estimate based on expansion of mean landings per individual trip.

Species grouping	Total portside estimate (lb)	95% confidence interval		Total observer estimate lb	95% confidence interval	
Fish unk	0			100,000	0	280,020
River herring	14,695	3,250	32,614	79,331	23,348	154,440
Spiny dogfish	13,076	2,777	36,156	12,849	2,621	26,140
Silver hake	7,372	1,560	15,220	10,375	2,444	23,322
Haddock	5,743	1,243	11,627	4,590	12	13,055
Butterfish	10,375	16	26,083	667	132	1,452
Squid	3,770	760	9,282	3,546	1,550	5,793
Cunner	0	0	0	4,864	0	14,592
Shad	1,288	142	3,124	2387	176	5,514
Scup	1,667	0	5,000	0		
Redfish	43	0	129	0		
Red hake	238	38	541	0		
Pollock	160	0	481	0		
Longhorn sculpin	6	0	21	54	0	162
American plaice	35	3	83	0		
Cod	17	0	46	0		
Lumpfish	9	0	27	6	0	24
Winter flounder	12	0	36	0		
Shrimp	4	0	12	0		
Menhaden	3	0	10	0		
Sea Raven	3	0	10	0		
Cusk	3	0	8	0		
Little skate	2	0	5	0		

Table 12 Estimates of total landings in weight with 95% confidence intervals based on bootstrap percentiles derived from Observer and Portside sampling for 52 trips

Total Estimate based on expansion of mean landings per individual trip.

Set type	species	r	95% confidence interval		P-value	Bootstrap problem indicator	95% confidence interval based on bootstrap	
All pairs	River herring	-0.04	-0.31	0.23	0.75		-0.11	0.19
Non-zero pairs	River herring	-0.13	-0.48	0.26	0.52		-0.37	0.98
All pairs	Squid	0.06	-0.22	0.32	0.70	B	-0.48	0.264
Non-zero pairs	Squid	-0.01	-0.36	0.34	0.95		-0.17	0.6196
All pairs	Silver hake	0.22	-0.06	0.47	0.12	B	0.04	0.91
Non-zero pairs	Silver hake	0.17	-0.21	0.51	0.37	B	-0.03	0.91
All pairs	Spiny dogfish[1]	0.80	0.68	0.88	<0.001	B	-0.10	0.98
Non-zero pairs	spiny dogfish[1]	0.79	-0.45	0.32	0.70	B	-0.27	0.98
All pairs	Butterfish-all	0.25	-0.03	0.49	0.08	B	-0.04	0.99
Non-zero pairs	Butterfish-pos	0.12	-0.51	0.67	0.72	B	-0.37	0.98
All pairs	Haddock	-0.02	-0.29	0.25	0.89	B,S	-0.08	0.64
Non-zero pairs	Haddock	-0.23	-0.71	0.40	0.47	B,S	-0.48	0.26

Table 13 Pearson's product moment Correlation coefficients, 95% confidence interval and 95% confidence interval from bootstrap for paired catches

[1.]Correlation coefficients are significantly different from 0 at P=0.05, but correlation coefficients are highly influenced by one trip.

[2] Removing influential points lowers correlation coefficients to not significantly different from zero. Bootstrap indicator: B= high bias, A= some bootstrap samples had zero standard deviations.

A comparison of portside and at-sea sampling methods of estimating bycatch in the Atlantic herring fishery

Micah Dean
Herring PDT
May 2011

MarineFisheries
Commonwealth of Massachusetts

Introduction

There are three primary gear types used to target Atlantic herring in the United States: bottom trawls, midwater trawls and purse seines (Table 1). Estimates of bycatch in this fishery are primarily derived from data collected at-sea under the Northeast Fisheries Observer Program (NEFOP). On bottom trawl vessels, bycatch species are typically sorted from the catch on deck and are discarded at sea. NEFOP samplers often achieve a census of the bycatch before it is thrown overboard and as a result there is essentially no variance surrounding the estimate of bycatch on a trip. The other two gear types are considered "high-volume" fisheries and bycatch species usually remain mixed with the catch as it is pumped from the net to the hold, as it is impractical to sort through such large catches at sea. As such, sea samplers estimate bycatch by taking a systematic sample of the catch as it is pumped onboard. While this method of sampling provides a less precise estimate of bycatch than a census of discards, the fact that bycatch species are retained presents an opportunity to also sample the catch when it is offloaded at port.

Table 1. Average trip information for vessels participating* in the Atlantic herring fishery in 2010, as reported to the National Marine Fisheries Services (NMFS) on Vessel Trip Reports (VTRs).

Gear Type	Total number of trips	Median catch per trip (Kg all species)	Median tows per trip	Median trip length (days)
Purse seine	163	31,752	2	1
Midwater trawl	350	145,423	2	3
Bottom trawl	207	2,472	3	1

** vessels with A,B,C or D herring permits not declared out of the fishery via VMS.*

In an effort to increase the number of trips sampled and thereby reduce the uncertainty surrounding fishery-wide estimates of bycatch, portside sampling programs have been initiated in Massachusetts and Maine. An obvious prerequisite to combining these portside and sea-sampling data is the comparability of the sampling programs. An initial comparison found relatively poor agreement between the two methods and raised concerns over the ability of either program to estimate bycatch in this fishery (Appendix A). The objective of this report is to explain the source of this previous disagreement, and to provide an updated comparison of the different programs. At-sea and portside sampling protocols were compared using a simulation model as well as with empirical data from herring trips sampled under both programs. This comparison focuses on midwater trawl vessels, as they present the greater challenge in sampling at-sea would therefore benefit the most from additional portside sampled trips.

Methods

Simulation

The contents of a typical, yet simplified hold of a midwater trawl herring vessel were simulated in the R software package by assembling an array of individual fish caught from three tows, totaling 150 mt in weight. The tows were of equal size (50 mt), but to evaluate the sensitivity of each sampling protocol to non-randomly distributed bycatch, two different scenarios were evaluated: 1) similar bycatch per tow and 2) dissimilar bycatch per tow. To simplify the comparison between protocols, the simulated hold contained only 3 species: Atlantic herring (target species), river herring[1] (higher abundance bycatch species) and whiting (lower abundance bycatch species). The hold under each scenario contained a similar amount of each species and differed only in the concentration of bycatch species in each tow (Table 2).

Table 2. Percent of target and bycatch species by weight in each tow under each simulation scenario.

Similar Tows Scenario

	Atlantic herring	river herring	whiting
Tow 1	98.9%	1.0%	0.1%
Tow 2	98.9%	1.0%	0.1%
Tow 3	98.9%	1.0%	0.1%
Total	98.9%	1.0%	0.1%

Dissimilar Tows Scenario

	Atlantic herring	river herring	whiting
Tow 1	99.89%	0.10%	0.01%
Tow 2	97.40%	2.40%	0.24%
Tow 3	99.41%	0.50%	0.05%
Total	98.9%	1.0%	0.1%

The number of individuals of each species (n_s) in a given tow was determined by multiplying the percent contribution of that species (p_s) by the weight of the tow (w_{tow}), and dividing by the average individual size ($\overline{w}_s$ - Figure 1).

[1] For the purposes of estimating bycatch in the Atlantic herring fishery, alewife (*Alosa pseudoharengus*) and blueback herring (*Alosa aestivalis*) are typically grouped together and referred to as "river herring."

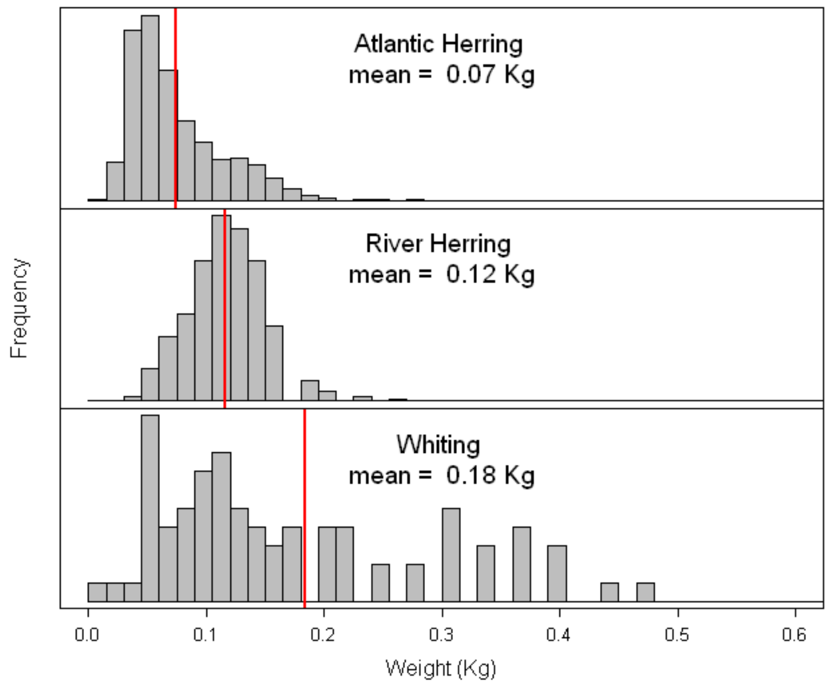

Figure 1. Frequency distributions of individual fish weights from observed herring trips in 2010.

A tow was represented by creating an array of individual fish of length n_{tow} ($\sum n_s$). For each bycatch species, a random sample of n_s individuals was selected from this array and designated as species *s*. The remaining individuals were designated as the target species, Atlantic herring. Individual fish were assigned random weights by sampling with replacement from the distribution of weights observed at-sea in 2010 for that species (Figure 1). The three tow arrays were joined into a single array to represent the hold. The bycatch distribution from each tow was kept intact when assembled into the hold array (i.e. no mixing) to mimic the process of pumping each tow's catch into a separate tank onboard the vessel. The total weight of each bycatch species was estimated by sampling from the hold array using four different protocols: 1) at-sea sampling: (AS) 2) portside unsorted sampling: (PU) 3) portside sorted sampling: (PS) and 4) portside lot sampling: (PL).

At-Sea Sampling (AS)

In high-volume fisheries, at-sea samplers typically take a systematic sample of 10 "baskets" from each tow to describe the species composition of the catch. The contents of each ~30 kg basket are sorted by species and weighed. The proportion of each species from the basket sample is then multiplied by the captain's estimate of tow size to achieve the amount of each species per tow. The total amount of bycatch for the trip is the sum of the bycatch estimates from each tow (Table 3). This sampling process was simulated by taking a systematic sample of 10 individuals from each tow array, with a random starting point. Each selected individual represented the beginning of a basket, and every fish following the initial one was added to the basket until the weight of the basket exceeded 30 kg.

Portside Unsorted Sampling (PU)
During the offload process, vessels typically pump their catch through a "de-watering box" as it is transferred from the hold to trucks or vats onshore. Portside samplers take a systematic basket sample of the catch as it is pumped through the de-watering box. In this case, the sampling interval is determined by the amount of time required to sort and weigh the contents of a basket, which on average is about seven minutes. Since the amount of time required to offload a vessel can be 24 or more hours, samplers take occasional breaks from sampling and are typically working for approximately 75% of the pumpout. The total amount of each bycatch species from the basket sample is expanded to the entire hold using the captain's estimate of the trip size (Table 3). This sampling protocol was simulated in a manner similar to the method of creating baskets under the AS protocol; However, in this case a single systematic sample of 32 baskets was taken from the hold array, based on an estimated 5 hour pumpout for a trip of this size (5 hours x 75% sampling time / 7 min per basket = 32 baskets).

Portside Sorted Sampling (PS)
The majority of Atlantic herring landings are sold as bait for the American lobster fishery, and as such are typically transferred directly from vessels to waiting trucks or vats. Alternatively, some herring are packaged, frozen and sold as food, often in international markets. Herring destined to be food are pumped from the vessel onto a conveyor belt at a shoreside facility where as many as 8 to 12 personnel (i.e. "pickers") separate bycatch species from the catch prior to packaging. The bycatch, as well as any damaged Atlantic herring, are sent to a central vat via a series of chutes. Portside samplers systematically sample the flow of bycatch to this vat via baskets in a manner similar to unsorted sampling. However, since the bycatch is more concentrated at this sampling location, processing the basket contents is more laborious and the average sampling interval is approximately 15 minutes. Likewise, samplers often take longer breaks and are working for approximately 50% of the pumpout. The total amount of each bycatch species from the basket sample is expanded to the entire hold using the total amount of bycatch (and damaged Atlantic herring) sorted to the vat (Table 3). To simulate this protocol, 5% of the target species in the hold array were randomly designated as damaged. The bycatch vat was represented by extracting a subset of the hold array that contained all bycatch species and damaged Atlantic herring, maintaining the original order from the hold array. A single 10-basket systematic sample from this vat array was taken, using the method described under the AS protocol (5 hours x 50% sampling time / 15 min per basket = 10 baskets).

Portside Lot Sampling (PL)
Herring that are sold as lobster bait are often pumped directly into trucks and driven over land to dealers in Maine. When it is not possible for portside samplers to intercept and sample the vessel during the pumpout, the contents of bait trucks are often sampled as they are unloaded at the retail dealer in Maine. In this case, a systematic sample of

baskets is taken from the contents of the truck as it is emptied. Typically, 1-3 trucks from a vessel are sampled and together are referred to as a "lot." On average, a 16-basket sample is taken from each truck. The total amount of bycatch from a trip is calculated by expanding the basket sample first to the entire lot, and then from the lot to the entire hold (Table 3). At times, conditions permit sorting all bycatch species from the lot as it is offloaded. In these cases, a census of bycatch from the lot is achieved. However, to simplify the comparison with other protocols, the systematic basket-sampling approach was used to represent PL sampling in the simulation. The hold array was broken up into eight sections to represent individual trucks, two of which were randomly selected for sampling. A 16-basket systematic sample was made from each selected truck, using the method described under the AS protocol.

A total of 1,000 iterations of each sampling protocol were made from the simulated hold under each scenario. The distribution of bycatch estimates from all sampling iterations was used to describe the accuracy and precision of each protocol. The mean estimate and coefficient of variation (CV) for each bycatch species were compared across the four protocols and two scenarios

A second simulation experiment was conducted to illustrate the effect of sample size and bycatch rarity on the precision of bycatch estimates as well as the ability to detect a bycatch species. In this case, a hold was simulated that contained four randomly-distributed bycatch species at various densities (1%, 0.1%, 0.05%, and 0.01%). A single systematic sample with random starting point was taken from this hold at various sample sizes (10, 25, 50 and 100 baskets). This sampling routine was iterated 1000 times, and the CV of the estimates, as well as the percent of estimates that were zero were compared.

Table 3.

Sampling protocol	Method used to estimate bycatch and variance	
At-Sea Sampling (AS)	$r_{s,tow} = \frac{\sum w_{s,i}}{\sum w_i}$ $w_{s,tow} = r_{s,tow} w_{tow}$ $w_s = \sum w_{s,tow}$ $V(w_{s,tow}) = N^2 \left(\frac{N-n}{N}\right)\left(\frac{s^2}{n}\right)$ $s^2 = \frac{\sum (w_{s,i} - r_{s,tow} w_i)^2}{n-1}$ $V(w_s) = \sum V(w_{s,tow})$	Where, $r_{s,tow}$ = ratio of species *s* in tow $w_{s,i}$ = weight of species *s* from basket *i* $w_{s,tow}$ = weight of species *s* from tow w_s = weight of species *s* in the hold w_i = weight of basket *i* *N* = number of possible baskets from tow *n* = number baskets sampled from tow
Portside Unsorted (PU)	$w_s = \frac{N}{n} \sum w_{s,i}$ $V(w_s) = N^2 \left(\frac{N-n}{N}\right)\left(\frac{s^2}{n}\right)$	Where, w_s = weight of species *s* in the hold $w_{s,i}$ = weight of species *s* from basket *i* *N* = number of possible baskets from hold *n* = number of baskets sampled from hold s^2 = sample variance of species *s* from baskets
Portside Sorted (PS)	$w_s = \frac{N}{n} \sum w_{s,i}$ $V(w_s) = N^2 \left(\frac{N-n}{N}\right)\left(\frac{s^2}{n}\right)$	Where, w_s = weight of species *s* in the hold $w_{s,i}$ = weight of species *s* from basket *i* *N* = number of possible baskets from bycatch vat *n* = number baskets sampled from bycatch vat s^2 = sample variance of species *s* from baskets
Portside Lot (PL)	$w_{s,lot} = \frac{N}{n} \sum w_{s,i}$ $w_s = \left(\frac{M}{m}\right) w_{s,lot}$ $V(w_{s,lot}) = N^2 \left(\frac{N-n}{N}\right)\left(\frac{s^2}{n}\right)$ $V(w_s) = \left(\frac{M}{m}\right)^2 V(w_{s,lot})$	Where, w_s = weight of species *s* in the hold $w_{s,i}$ = weight of species *s* from basket *i* *N* = number of possible baskets in lot *n* = number of baskets sampled from lot *M* = number of possible trucks from hold *m* = number of trucks sampled from hold s^2 = sample variance of species *s* from baskets

Empirical Data
A total of 30 midwater-trawl herring trips from 2010-2011 were identified as being sampled by both at-sea and portside methods. Twenty-four trips were sampled by PU methods and six trips were sampled by PS methods. Five trips were sampled by more than one portside method (PU and PL). Estimates of bycatch for six common species (river herring, whiting, American shad, butterfish, haddock and spiny dogfish) were calculated for each trip and compared across sampling methods. The variability of the bycatch observed in the basket sample data was used to estimate the variance surrounding the bycatch estimate for each trip (Table 3). The amount of agreement between at-sea sampling and port sampling was evaluated in two different ways: 1) the ability to detect bycatch species (i.e. presence-absence) and 2) significant differences in bycatch estimates.

There are four possible outcomes when comparing two sampling protocols' ability to detect a bycatch species: 1) present in both (++); 2) present in neither (--); 3) present in portside sampling, absent from sea sampling (P+); and 4) present in sea sampling, absent from portside sampling (S+). For each of the six species, a matching coefficient was calculated by dividing the number of trips that fell into the first two categories (++,--) by the total number of trips sampled. Because some species are infrequently encountered by this fishery, a high proportion of "double negative" (--) trips could yield an unrealistically high matching coefficient. To account for this, a second matching coefficient was calculated that omits the "--" trips from both the numerator and denominator.

A modified two sample *t*-test assuming unequal variances (i.e. Welch's test) was used to test for a significant difference ($\alpha = 0.05$) between portside and at-sea sampling estimates of bycatch. This test typically relies on the sample means ($\bar{x}$) and sample variances (s^2) to calculate the *t* statistic:

Eq. 2) $$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

However, since we are comparing a variety of protocols with different *types* of samples (e.g. sorted vs unsorted), the sample means and variances were replaced with the total estimated bycatch per trip (*w*), and the variance of those estimates (*V(w)*):

Eq. 3) $$t = \frac{w_1 - w_2}{\sqrt{V(w_1) + V(w_2)}}$$

To evaluate how often bycatch is non-randomly distributed in the catch, a one-sided "runs" test was performed on the series of basket observations for a single species (river herring) for each sampled trip. This test relies on the calculation of an expected number of "runs" given the number of observations at two levels, a "run" being a sequence of adjacent observations at the same level. In this case, the two levels were defined as: 1) above the mean value and 2) below the mean value. If the number of observed runs for a trip was significantly lower than the expected value ($\alpha = 0.05$), it was considered to have non-randomly distributed bycatch.

Results

Simulation

AS, PU and PL sampling achieved comparable levels of precision under the similar tows scenario, with a CV of approximately 0.1 for river herring and 0.6 for whiting (Figure 2). PS sampling was identified as being the most precise, with a CV roughly half of the other protocols. AS, PU and PS sampling performed equally well under the dissimilar tows scenario, indicating these protocols are robust to non-random distributions of bycatch (Figure 3). On the other hand, PL sampling performed very poorly under the dissimilar tows scenario, yielding a CV of 0.59 for river herring (a 500% increase) and a CV of 0.91 for whiting (an 80% increase). Additionally, the PL protocol failed to detect any of the less abundant whiting 9% of the time under the dissimilar tows scenario. None of the protocols were found to be biased, achieving mean estimates within 3% of the true value under either scenario.

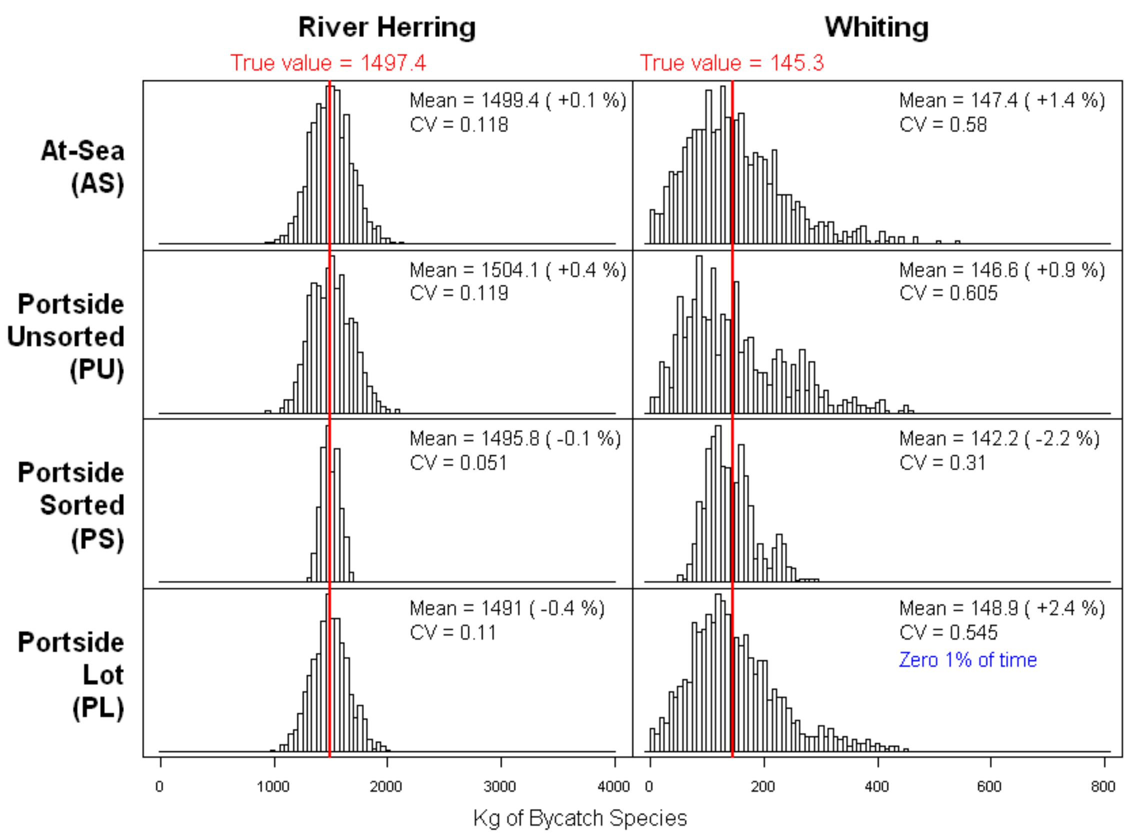

Figure 2. Distribution of bycatch estimates from 1000 iterations of sampling the simulated hold under the similar tows scenario.

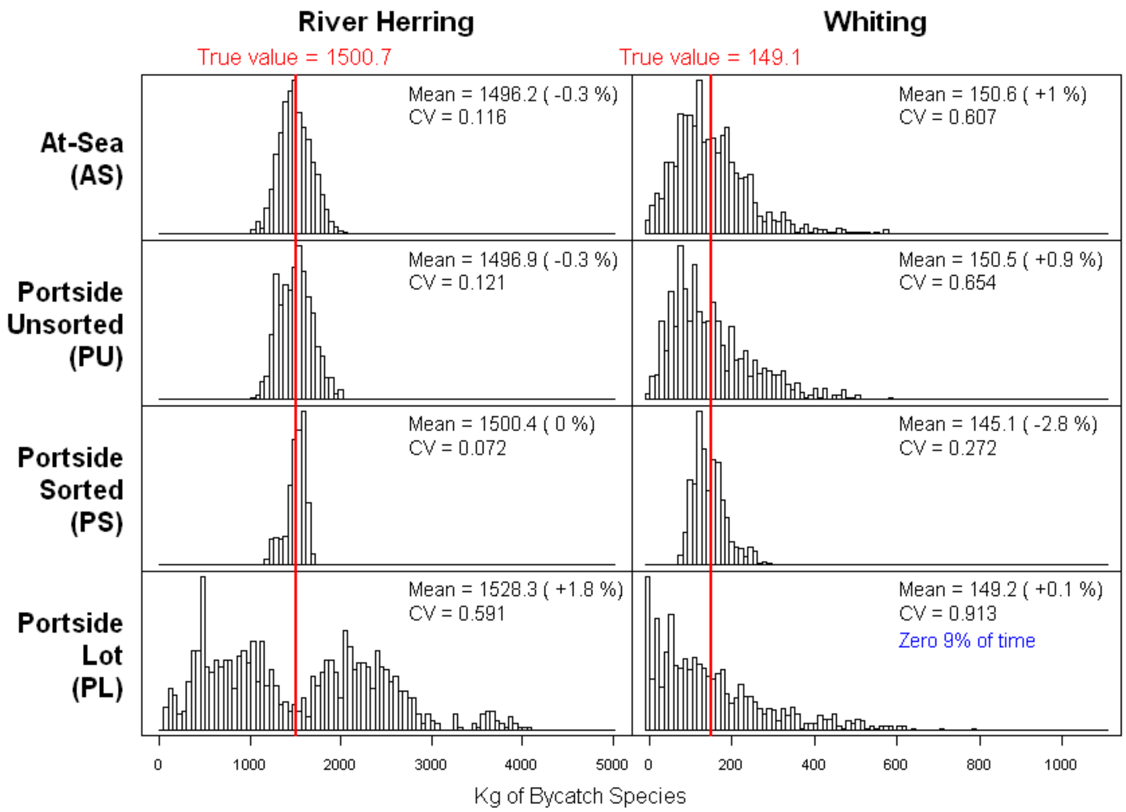

Figure 3. Distribution of bycatch estimates from 1000 iterations of sampling the simulated hold under the dissimilar tows scenario.

For the second simulation experiment, results indicated that both sample size and bycatch rarity have a strong influence on the precision of the estimate and the ability to detect bycatch species (Figure 4). For all levels of bycatch rarity, the CV of the estimate increased by a factor of 3 when the sample size was dropped from 100 baskets to 10 baskets. Likewise, the CV of the estimate increased by a factor of 10 when the rarity of the bycatch dropped from 1% of the catch to 0.01%. The ability to detect the rarest bycatch was very low at the smallest sample size, with 71% of the sampling iterations failing to detect a single bycatch individual.

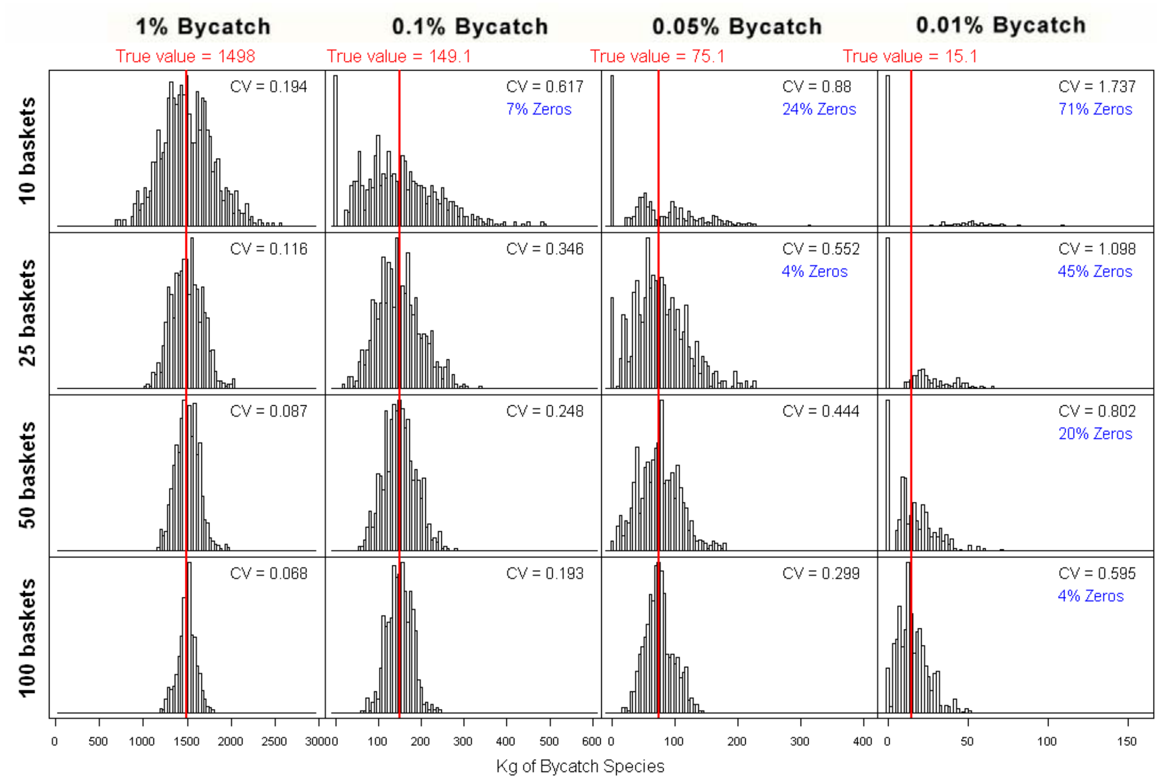

Figure 4. Distribution of bycatch estimates at four levels of bycatch rarity from 1000 iterations of a single systematic sample of various sizes.

Empirical Data

PU sampling achieved a 75% to 92% agreement with AS in the ability to detect the six bycatch species (Table 4). When the "double negative" (--) tows were omitted, the amount of agreement dropped substantially for some species (e.g. spiny dogfish, butterfish), yet remained high for others (e.g. river herring, silver hake). The *amount* of bycatch estimated also agreed well between PU and AS sampling. For the six species considered, significant differences in bycatch amount were detected on only 8 to 25% of the trips (75 to 92% agreement). All of these significant differences resulted from one protocol "missing" the bycatch species (i.e. P+ or S+).

Table 4. Number of trips that detected bycatch species under PU and AS and the percent agreement between the two protocols (top). Number of trips with a significant difference (alpha = 0.05) in bycatch amount of six species, between AS and PU methods (bottom).

Presence-Absence

	Agree		Disagree		% Agree	% Agree
	++	- -	P+	S+	% Agree	(omit --)
River Herring	18	3	1	2	88%	86%
Butterfish	5	15	3	1	83%	56%
Silver Hake	12	9	1	2	88%	80%
Spiny Dogfish	2	19	3	0	88%	40%
Haddock	6	16	2	0	92%	75%
American Shad	7	11	4	2	75%	54%

Significant Differences

	No Sig Diff		Sig Diff			% No Sig	% No Sig Diff
	++	- -	++	P+	S+	Diff	(omit --)
River Herring	18	3	0	1	2	88%	86%
Butterfish	5	15	0	3	1	83%	56%
Silver Hake	12	9	0	1	2	88%	80%
Spiny Dogfish	2	19	0	3	0	88%	40%
Haddock	6	16	0	2	0	92%	75%
American Shad	7	11	0	4	2	75%	54%

PS sampling achieved 50% to 83% agreement with AS in the detection of bycatch species (Table 5). However, for some species the agreement dropped to 0% when "--" trips were omitted (i.e. spiny dogfish, haddock). In terms of the quantity of bycatch estimated, PS sampling had limited agreement with AS sampling: 33%-83% of the trips sampled had significant differences between the protocols (17%-67% agreement). However, much of that agreement came from "double negative" trips and once they were removed, the amount of agreement dropped to 0%-50%.

Table 5. Number of trips that detected bycatch species under PS and AS and the percent agreement between the two protocols (top). Number of trips with a significant difference (alpha = 0.05) in bycatch amount of six species, between AS and PS methods (bottom).

Presence-Absence

	Agree		Disagree		% Agree	% Agree
	++	- -	P+	S+		(omit --)
River Herring	5	0	1	0	83%	83%
Butterfish	2	2	2	0	67%	50%
Silver Hake	4	0	2	0	67%	67%
Spiny Dogfish	0	3	3	0	50%	0%
Haddock	0	4	2	0	67%	0%
American Shad	3	0	3	0	50%	50%

Significant Differences

	No Sig Diff		Sig Diff			% No Sig Diff	% No Sig Diff
	++	- -	++	P+	S+		(omit --)
River Herring	2	0	3	1	0	33%	33%
Butterfish	1	2	1	2	0	50%	25%
Silver Hake	1	0	3	2	0	17%	17%
Spiny Dogfish	0	3	0	3	0	50%	0%
Haddock	0	4	0	2	0	67%	0%
American Shad	3	0	0	3	0	50%	50%

The comparison of PL to AS sampling was restricted to only three species, as none of the other common bycatch species were encountered under either protocol. PL sampling achieved 20% to 80% agreement with AS in the detection of those three bycatch species (Table 6). Similarly, significant differences between the two protocols were found on 20% to 80% of the trips. While it is difficult to draw conclusions from so few trips sampled, it appears that the amount of agreement between PL and AS sampling was highly variable.

Table 6. Number of trips that detected bycatch species under PL and AS and the percent agreement between the two protocols (top). Number of trips with a significant difference (alpha = 0.05) in bycatch amount of six species, between AS and PS methods (bottom).

Presence-Absence

	Agree		Disagree		% Agree	% Agree
	++	- -	P+	S+		(omit --)
River Herring	3	1	0	1	80%	75%
American Shad	0	1	1	3	20%	0%
Spiny Dogfish	1	3	1	0	80%	50%

Significant Differences

	No Sig Diff		Sig Diff			% No Sig Diff	% No Sig Diff
	++	- -	++	P+	S+		(omit --)
River Herring	2	1	1	0	1	60%	50%
American Shad	0	1	0	1	3	20%	0%
Spiny Dogfish	1	3	0	1	0	80%	50%

Of the 30 trips that were sampled both at-sea and portside, 11 trips (37%) were found to have a significantly non-random distribution of river herring in the catch when sampled at-sea (Figure 5). Similarly, nine trips (30%) were found to have non-randomly distributed river herring when sampled portside. Six of these portside trips were also found to be non-random at-sea, indicating that the distribution of bycatch at-sea often determines the distribution of bycatch seen during the offload at port.

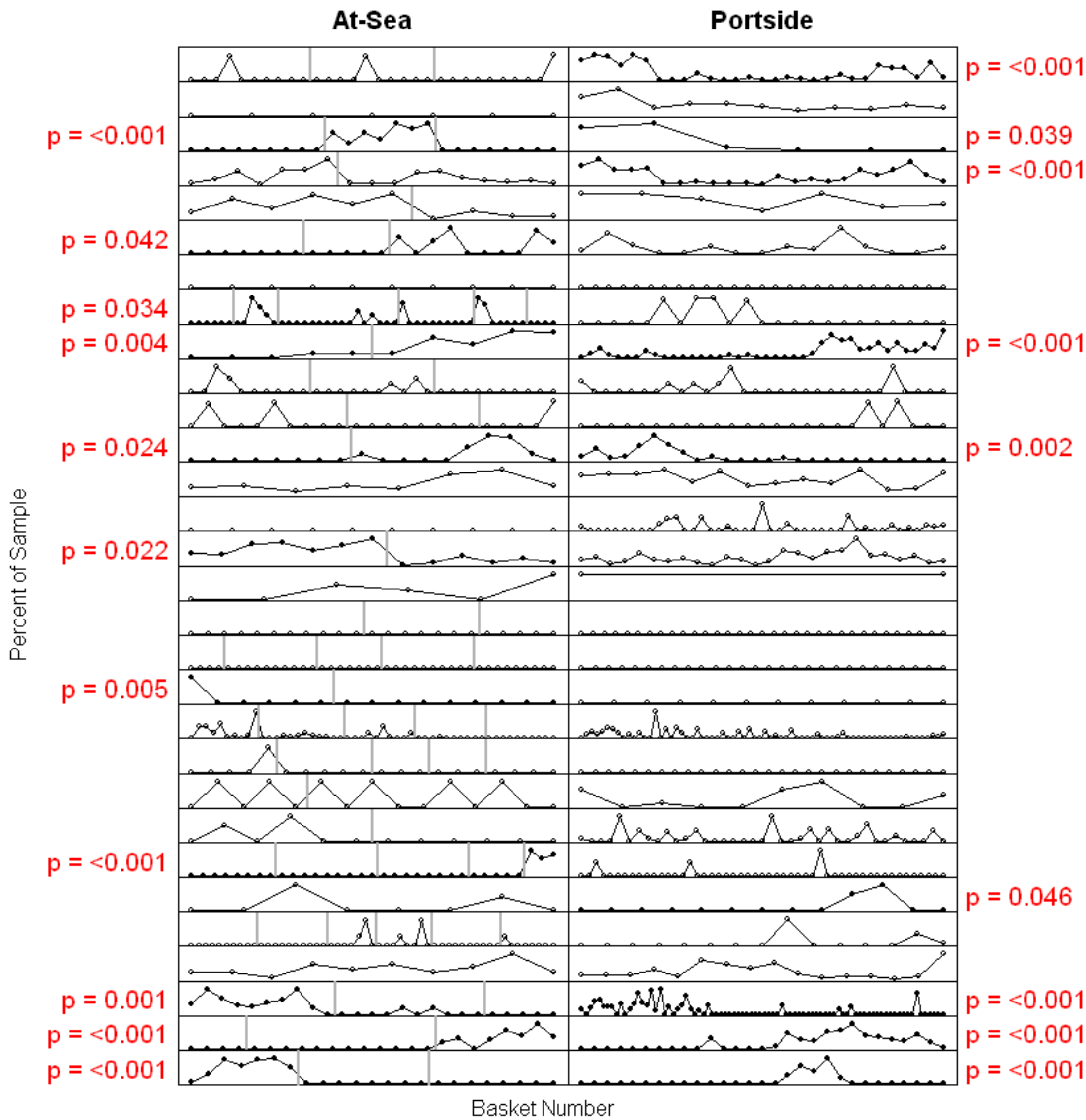

Figure 5. Sequence of river herring observations from the basket samples taken at-sea (left) and portside (right) for each co-sampled trip. The vertical gray lines on the left pane indicate the separation between tows. Trips that were found to have a non-random distribution of river herring are identified by the *p*-value of the runs test in the margin.

Discussion

The results of the simulation experiment indicate that AS and PU sampling should estimate the amount of bycatch on a herring trip with equal precision. However, the precision surrounding these bycatch estimates is primarily determined by the sample size (i.e. the number of baskets), and each protocol arrives at this number differently. The configuration of the hold in the simulation experiment (3 tows with 150 mt total landings) was intentionally designed to create roughly equivalent sample sizes under each protocol. In reality, most herring trips would likely end up with different sample sizes if sampled at-sea or portside. On trips with relatively low landings from a high number of tows, AS sampling will achieve the larger sample size, since AS protocol requires a fixed number of baskets per tow. On the other hand, trips with high landings from few tows will yield a higher sample size under PU sampling, as number of samples taken is a function of pumpout time under the PU protocol. For the 24 trips that were sampled by both methods, 75% had a higher sample size under PU sampling. As a result, the average CV from PU sampling was 42% less than that achieved under AS on the same trips. Despite these dissimilar sample sizes, the empirical data appear to corroborate the results of the simulation, showing little disagreement between PU and AS sampling. None of the trips where both protocols detected a species had significant differences in bycatch estimates. Most of the trips where PU and AS did disagree fell into the "P+" category, or where bycatch was observed in PU but not observed in AS. This is likely a symptom of the larger sample sizes under portside sampling. From the second simulation experiment, it is clear that the likelihood of not detecting a rare bycatch species is far greater at lower sample sizes.

The simulation identified PS sampling as having the highest precision, while being just as accurate as PU and AS methods. Unfortunately, this did not appear to be the case when the empirical data were examined. Although only six trips were sampled by both PS and AS methods, it was clear that the two methods had very poor agreement. Four trips (67%) showed a significant difference in the estimated amount of river herring bycatch. This high amount of disagreement was a surprising find and indicates the presence of a strong bias in PS sampling. It is unlikely that AS protocol is biased, since it is corroborated by the good agreement with PU sampling. A potential source of bias from PS sampling could be that pickers are "missing" bycatch species as they pass by on the conveyor belt. The PS method of sampling assumes that all bycatch are separated from catch and sent to a vat to be sampled from. If a large portion of a particular bycatch species is missed on the picking line, the PS method will significantly underestimate the amount of bycatch. In fact, three of the four trips that had a significant difference in river herring bycatch had a lower estimate under PS sampling. Occasionally, a few random boxes of packaged Atlantic herring are opened for quality control purposes and examined for missed bycatch species. However, since bycatch species such as river herring often account for less than 1% of the weight of trip, the chance of encountering the occasional 'missed' river herring in a randomly selected box is very small. Likewise, if a fraction of 1% of the packaged product is of a different herring species, it unlikely that it would lead to customer complaints.

Although apparently biased, the average precision surrounding the PS estimates of bycatch was 40% less than that of AS sampling on the same trips. Also, PS sampling was far better at identifying the presence of rare bycatch species than AS, with 13 instances of a species being identified under PS sampling but not under AS (P+). There were no cases of a species being identified under AS but not under PS (S+). If it is possible to identify and correct the source of the sampling bias, the PS protocol has the potential to provide the most precise estimate of bycatch for this fishery.

The simulation experiment identified PL sampling as being roughly equivalent to PU and AS sampling for trips that had randomly distributed bycatch in the hold (i.e. similar tows scenario). However, PL sampling was particularly vulnerable to non-random bycatch distribution. Intuitively, this vulnerability is caused by focusing all of the sampling on a small portion of the hold. If bycatch are more concentrated in the trucks selected for sampling, the resulting bycatch estimate will be too high. Conversely, if trucks with less bycatch are selected, the bycatch estimate will be too low, or none will be detected at all. The other protocols distribute the sampling effort across the entire catch, and are therefore more robust to non-randomly distributed bycatch. From the results of the runs test, it appears that non-random bycatch is fairly common in this fishery, with more than a third of the examined trips identified as having river herring non-randomly distributed in the catch. A previous comparison of portside and at-sea bycatch estimates relied heavily on data from PL sampling (Appendix A) and yielded relatively poor agreement between portside and at-sea estimates of bycatch. It appears the frequency of non-random bycatch coupled with the vulnerability of PL sampling to this phenomenon is the primary cause of the disagreement found in that investigation. It is important to note that sampling the offload from trucks is not inherently flawed, and could provide a reliable estimate of bycatch if it is possible to sample *all* the trucks from a particular trip. In many cases this may not possible, as trucks filled from a single hold are often destined for multiple locations. In any case, portside resources should be directed to ensure that the entire catch from a trip is available for sampling.

Currently, estimates of bycatch in this fishery are derived from AS sampling alone. The high amount of agreement between PU and AS sampling found in this investigation lends credibility to both programs and it seems reasonable to combine bycatch estimates from trips sampled under either protocol. Incorporating PU sampled trips could reduce the CV on the fishery-wide estimate of bycatch by dramatically increasing the sample size, particularly for areas and gears with limited AS sampling coverage. In addition, PU sampling can be a far more efficient use of limited resources than AS sampling: Consider the midwater-trawl trip from the simulation experiment, with 150 mt of landings that takes approximately 5 hours to offload. Depending on the fishing area and port of landing, this might be a four to six day trip. The total cost of sampling that trip portside would be approximately $350 (2 samplers at ~$35 / hr for 5 hrs), whereas the AS cost would likely be $5000 to $7000 (1 sampler at ~$1200 / day). If additional sampling resources are to be directed at this fishery beyond those already required by the NEFOP allocation algorithm (SBRM), portside unsorted sampling should be considered as it is a reliable cost-effective method of estimating retained bycatch.