Alaska Groundfish Data Bank study to evaluate an electric monitoring program for estimating halibut discards: Statistical analysis of study data

Prepared for:
Marine Conservation Alliance Foundation
Ste 305, 431 N. Franklin St., Juneau, AK 99801

By:
Vivian Haist
1262 Marina Way
Nanoose Bay, BC. Canada V9P 9C1

This report was prepared by Vivian Haist under a subcontract from the Marine Conservation Alliance Foundation with funds from award NA06NOS4720020 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the National Oceanic and Atmospheric Administration or the U.S. Department of Commerce.

March, 2008

Table of Contents

1 Introduction

An experimental study to investigate the use of an electronic monitoring (EM) system for estimating at-sea halibut discards was conducted by the Alaska Groundfish Data Bank under exempted fishing permit (EFP) number 07-02. This work was motivated by planned changes to the management of the rockfish fishery (the Rockfish Pilot Program), which will require tracking of individual allocations of target and bycatch species. The use of EM for monitoring bycatch may be a cost effective approach relative to the proposed 100% observer coverage. The Alaska Groundfish Data Bank study was designed to investigate the feasibility of EM for monitoring at-sea halibut discards and to evaluate the relative accuracy and precision of EM accounting relative to current observer-based sampling methods.

This report presents results of statistical analyses of the data collected during the EM experimental study. Specifically, halibut estimates from observer sampling and EM methods are compared to halibut census and total halibut catch estimates to determine the relative accuracy and precision of the two monitoring approaches. Additionally, covariates are examined to see if they relate to errors in EM fish length measurements. Analyses and statistical tests are conducted at the haul level, fishing event level, trip level and study level, as appropriate. The analyses follow the specifications given in the *Statement of Work* for this project (Appendix A).

2 Experimental Fishing

The Electronic Monitoring study, conducted from September 16 to October 14, 2007, encompassed 27 hauls made over 6 fishing trips (Table 1). For each haul, halibut estimates (counts and weights) were made using observer sampling (OS), electronic monitoring video records of halibut (EM), and a census of the halibut discards (C). Additionally, halibut that were missed during at-sea sorting and were landed at the fishing plant were counted and weighed (L).

The observer samples were taken prior to any discarding, and hence the samples are representative of the total halibut catch. The data, collected for each haul, includes sub-sample weights and weights and counts of halibut in each sub-sample. For each haul, the halibut catch is estimated by extrapolating the ratio of halibut weight (or count) to sample weight (sum of the sub-sample weights) to the OTC estimate of the catch. On average observer samples represented 2.2% of the total catch (Table 1).

The electric monitoring data is a video-record of all halibut discards. Video-records were made of each haul and the records reviewed by two individuals (EMR1 and EMR2) to estimate the number of discarded halibut and the length of each halibut. Halibut lengths are converted to weight using standard IPHC conversion tables.

An on-board census was conducted to enumerate the halibut discards for each haul. Data includes the total number and individual lengths of each discarded halibut. Halibut lengths were measured sequentially as they were discarded so that individual lengths can be compared with those obtained from the EM method. Over all hauls, discarded halibut represented 1% of the total catch, by weight (Table 1).

Some halibut were inadvertently landed rather than discarded and information is available from plant landings records. For each fishing trip, the count and weight of landed halibut was recorded for each of three tanks on the fishing vessel. The catch from each haul was associated with a particular tank, so all landed halibut can be associated with either a single haul or a set of two hauls. The term “fishing event” is used to distinguish the units to which each landed halibut can be uniquely associated (i.e. either a single haul or a set of two hauls).

The true total halibut catch is taken to be the sum of the census and the landings. The count and weight of total halibut in the catch is available at the fishing event level.

Table 1. Summary of halibut catch (census and landings), halibut catch estimates (observer sampling and EM from reviewers EMR1 and EMR2), overall catch weight (OTC), haul information, and proportion (by weight) of catch in observer samples and proportion of halibut (by weight) in catch.

	Haul					Halibut weight (kg)						Halibut landed to plant	
Trip	#	Date (2007)	Set	Ret.	OTC (mt)	Cens.	Obs.	EMR1	EMR2	Prop. samp.	Prop. hal.	Tank	Number delivered (by tank)
1	1	16-Sep	17:45	19:20	36.58	3.6	0.0	2.5	3.6	0.009	0.000	F,M	Front: zero Middle: 16@129.7 kg Aft: 14@ 91.17 kg.
	2	17-Sep	8:00	8:45	15.35	172.2	0.0	176.9	188.2	0.018	0.011	M	
	3	17-Sep	11:00	12:00	12.97	129.5	0.0	120.7	118.1	0.020	0.010	A	
	4	17-Sep	13:35	14:16	9.98	54.3	0.0	54.9	57.5	0.032	0.005	A	
2	5	20-Sep	8:25	9:25	33.21	137.2	637.1	196.7	135.0	0.013	0.004	F,M	Front: 5 @ 34 kg. Middle: zero Aft: 2 @ 10.4 kg
	6	20-Sep	12:55	13:25	21.44	9.4	0.0	9.6	9.5	0.018	0.000	M	
	7	20-Sep	15:50	16:50	13.09	59.4	0.0	61.0	64.4	0.023	0.005	A	
	8	20-Sep	19:15	22:00	13.98	271.4	184.7	263.6	268.0	0.026	0.019	A	
3	9	24-Sep	13:35	14:01	5.79	0.0	0.0	0.0	0.0	0.052	0.000	F	Zero Halibut
	10	24-Sep	18:50	20:30	14.77	17.3	0.0	16.4	16.2	0.021	0.001	F	
	11	25-Sep	12:45	13:02	11.82	27.5	415.6	26.5	26.7	0.022	0.002	F,A	
	12	26-Sep	7:50	8:15	32.58	49.3	519.7	50.1	49.8	0.010	0.002	A	
4	13	28-Sep	21:00	23:50	8.71	37.2	0.0	39.7	41.2	0.030	0.004	F	Aft tank: 3 @ 9.98 kg.
	14	29-Sep	9:50	12:20	7.86	844.4	216.1	858.7	837.8	0.037	0.107	A	
	15	29-Sep	19:05	20:30	17.23	67.6	0.0	74.9	67.5	0.020	0.004	M	
5	16	8-Oct	13:37	14:05	18.15	391.9	238.1	411.3	387.4	0.016	0.022	F	Front tank: 1 @ 2.92 kg Middle tank: 1 @ 3.63 kg Aft tank: 2@ 13.61 kg
	17	8-Oct	18:30	19:02	10.18	94.4	0.0	92.5	93.1	0.027	0.009	F	
	18	9-Oct	8:10	8:59	16.85	33.9	0.0	33.4	30.3	0.018	0.002	M	
	19	9-Oct	10:55	12:11	21.67	122.4	0.0	124.2	122.7	0.017	0.006	M	
	20	9-Oct	14:40	16:05	19.13	152.7	0.0	164.0	145.9	0.016	0.008	A	
6	21	12-Oct	9:00	11:05	12.31	138.8	0.0	148.4	132.6	0.023	0.011	F	Middle tank: 6@ 22.23 kg.
	22	12-Oct	13:00	15:25	11.94	665.8	0.0	733.7	699.1	0.026	0.056	M	
	23	12-Oct	19:10	19:55	0.54	37.0	66.8	30.5	35.2	0.518	0.069	A	
	24	13-Oct	9:00	11:00	0.91	61.6	21.4	63.4	58.7	0.350	0.068	A	
	25	13-Oct	12:45	14:06	1.81	70.0	84.7	74.9	69.5	0.196	0.039	A	
	26	13-Oct	15:15	17:26	1.13	47.2	32.6	49.7	49.3	0.243	0.042	A	
	27	13-Oct	19:20	21:51	11.23	255.9	0.0	280.6	283.9	0.032	0.023	M	
Total					381.2	3951.	2416.	4158.	3991.	0.022	0.010		

3 EM Data

The EM data collected during the experimental study are video-records of all halibut moving through the discard chute. These records were reviewed independently by two individuals (EMR1 and EMR2) to: 1) obtain a count of halibut, and 2) to measure individual fish lengths. In cases where there was no clear image for an individual halibut a comment was noted in the data summary. For all analyses based on errors in individual fish length measurements (i.e. differences between the census length and EM reviewer length), only data from hauls where the census and EM halibut counts were the same are used. Hauls where the number of halibut enumerated differed between the census and the EM reviewer were excluded, as the association between census and reviewer lengths could not be established.

3.1 Length Conversion Factors

For each haul, factors for converting the video-record length measurements to actual fish size were estimated. The conversion factors were calculated from a known length grid system on the escape chute (Figure 1). There was some distortion in the video images such that conversion factors were different between the *top of grid* and *bottom of grid* measurements. Therefore, the *top of grid*, the *bottom of grid*, or an average of the two conversion factors could be used to calculate actual fish lengths from the raw video measurements.

For both EM reviewers, individual halibut lengths were estimated using conversion factors calculated for the *top of grid*, the *bottom of grid*, and for an average of the two sets of conversion factors. The mean halibut lengths from the census and from the alternative EM estimates are compared in Table 2. Error in the EM estimates, calculated as the proportional difference between the census and the EM estimates, range from -10.5% to 4.6% (Table 2). The mean error over all hauls ranged from -2.1% to 0.4%, depending on EM reviewer and length conversion method.

Hypothesis tests to investigate if the EM mean length estimates are biased were conducted for each EM reviewer and conversion method. A two-tailed t-test was used to test the null hypothesis that the mean error was equal to zero $(\alpha = 0.05)$. The null hypothesis was rejected for both EM reviewers using the *bottom of grid* conversion factors (p-values of 0.007 and <0.000), and for EMR2 using the *average* conversion factors (p-value <0.000, Table 2), suggesting the *bottom of grid* and *average* conversion factors may result in biased results. Using the *top of grid* conversions, the null hypothesis is not rejected for either reviewer, indicating this is the preferable conversion method. All further analyses of the EM data are based on the *top of grid* conversion factors.

3.2 EM image clarity

While reviewing the EM images, each reviewer noted when obtaining a length measurement was compromised due to either an unclear image or the fish not lying flat on the grid surface. Example comments are shown in Table 3. Overall, the fraction of images that were not noted as unclear or otherwise problematic was 71% and 66% for reviewers EMR1 and EMR2, respectively (Table 4).

Table 2 Estimates of mean halibut length from the census and from the EM review, and the error in EM estimates (proportional difference) for each haul. EM estimates are calculated for the two reviewers (EMR1 and EMR2), using conversion factors calculated for the *top of grid*, the *bottom of grid*, or an average of the two sets. For each EM reviewer and conversion method, a t-test was conducted to test the null hypothesis that the mean error, across hauls, was zero. Probability values for these tests are shown in the table.

	Mean Length (cm)							Error (EM-Census)/Census					
		Top of Grid		*Bottom of Grid*		*Average Grid*		*Top of Grid*		*Bottom of Grid*		*Average Grid*	
Haul	Census	EMR1	EMR2	EMR1	EMR2	EMR1	EMR2	EMR1	EMR2	EMR1	EMR2	EMR1	EMR2
1	68.0	60.8	67.6	62.9	65.0	61.9	66.3	-0.105	-0.006	-0.075	-0.045	-0.090	-0.025
2	85.3	84.6	86.3	83.2	84.8	83.9	85.5	-0.009	0.011	-0.024	-0.007	-0.017	0.002
3	80.2	78.4	78.0	78.5	77.3	78.5	77.6	-0.022	-0.027	-0.021	-0.036	-0.021	-0.032
4	78.7	78.3	79.6	76.5	77.2	77.4	78.4	-0.005	0.012	-0.028	-0.018	-0.016	-0.003
5	76.2	79.6	75.8	78.8	74.2	79.2	75.0	0.045	-0.004	0.035	-0.025	0.040	-0.015
6	73.5	74.1	73.8	73.0	71.5	73.6	72.7	0.008	0.004	-0.007	-0.027	0.001	-0.011
7	79.9	80.2	81.8	77.9	79.2	79.1	80.5	0.004	0.023	-0.025	-0.008	-0.010	0.008
8	80.5	79.4	79.6	79.5	78.5	79.4	79.1	-0.013	-0.011	-0.012	-0.024	-0.013	-0.017
9	-	-	-	-	-	-	-						
10	59.9	58.8	58.7	57.1	58.0	57.9	58.3	-0.017	-0.020	-0.047	-0.031	-0.032	-0.025
11	76.2	76.0	76.0	73.7	74.8	74.9	75.4	-0.002	-0.003	-0.033	-0.018	-0.018	-0.011
12	86.8	87.2	86.6	86.9	85.9	87.0	86.3	0.004	-0.002	0.001	-0.011	0.002	-0.007
13	84.0	86.0	86.7	86.4	83.9	86.2	85.3	0.024	0.032	0.028	-0.002	0.026	0.015
14	70.1	70.4	69.8	70.1	68.3	70.2	69.0	0.004	-0.005	0.000	-0.025	0.002	-0.015
15	84.3	87.0	84.3	83.4	83.3	85.2	83.8	0.031	0.000	-0.011	-0.012	0.010	-0.006
16	77.0	77.9	76.9	75.1	75.2	76.5	76.0	0.012	-0.002	-0.024	-0.023	-0.006	-0.012
17	76.4	75.4	75.6	74.4	74.3	74.9	75.0	-0.013	-0.010	-0.026	-0.027	-0.020	-0.019
18	82.4	82.1	79.6	80.1	78.8	81.1	79.2	-0.004	-0.033	-0.028	-0.044	-0.016	-0.039
19	76.1	76.0	75.9	75.0	74.1	75.5	75.0	-0.001	-0.002	-0.014	-0.026	-0.007	-0.014
20	77.2	78.5	76.0	75.3	76.1	76.9	76.0	0.017	-0.016	-0.024	-0.015	-0.003	-0.015
21	73.6	75.2	72.3	73.6	71.5	74.4	71.9	0.021	-0.017	0.000	-0.028	0.011	-0.023
22	69.9	71.5	70.4	69.3	68.7	70.4	69.6	0.022	0.007	-0.009	-0.017	0.007	-0.005
23	80.5	82.5	79.1	82.0	78.9	82.2	79.0	0.025	-0.018	0.018	-0.020	0.021	-0.019
24	66.0	66.4	64.7	65.3	63.8	65.9	64.2	0.006	-0.019	-0.010	-0.034	-0.002	-0.027
25	68.6	69.8	69.6	67.5	68.3	68.7	68.9	0.019	0.014	-0.016	-0.004	0.002	0.005
26	73.3	76.7	74.1	74.9	73.1	75.8	73.6	0.046	0.010	0.021	-0.003	0.033	0.004
27	73.2	74.2	74.6	72.3	72.8	73.2	73.7	0.014	0.019	-0.013	-0.006	0.001	0.006
						Mean		0.004	-0.002	-0.013	-0.021	-0.004	-0.012
						Std. Dev.		0.028	0.016	0.023	0.012	0.025	0.013
						Probability		0.451	0.449	0.007	0.000	0.361	0.000

The accuracy of EM mean fish length estimates may be increased by exclusion of images that were considered unclear by the reviewer. The estimates of mean halibut length when unclear images are excluded from the calculations are shown in Table 4. For both EM reviewers, exclusion of unclear images does not result in biased mean length estimates (t-test with null hypothesis that mean error across all hauls is zero, Table 4).

For both reviewers, the standard deviation of the error increased when unclear images were excluded from the mean halibut length calculations (compare Table 2 *top of grid* errors with Table 4 errors). That is, the accuracy of the mean halibut lengths for individual hauls decreased when only clear images were used in the estimation, although there was no indication that the estimates were biased. The standard deviation of the error increased from 2.8% to 5.0% for EMR1 and from 1.6% to 5.2% for EMR2 (Table 2, Table 4). Hence it appears that exclusion of unclear images is not warranted.

Table 3. Example comments from EMR1 and EMR2 notes.

- Fish not flat on grid
- Image very blurry - difficult to get an accurate measurement
- Fish head under cover of grid.
- Fish tail bent
- Deck lighting changed: Fish are blurry and difficult to get an accurate measurement
- Fish not flat on grid (piled on top of other fish)
- Fish tail bent up against grid rail.
- Fish tail not flat on grid and nose of head is hidden from view
- Half of fish bent up against grid rail
- Fish nose not in view
- Tail straight up

Figure 1. Discard chute for recording halibut discards showing the grid system for calibrating fish lengths.

Table 4 Estimates of mean halibut length from the census and from the EM reviews excluding unclear images, the error in EM estimates ([EM-Census]/Census), and the total number and number of clear images for each haul. EM estimates are calculated for the two reviewers (EMR1 and EMR2) based on the *top of grid* conversion factors. For each EM reviewer, a t-test was conducted to test the null hypothesis that the mean error across hauls was zero. Probability values for these tests are shown in the table.

	Census	EMR1					EMR2				
		Number					Number				
Haul	Mean length	All	Clear	Prop. clear	Mean length	Error	All	Clear	Prop. clear	Mean length	Error
1	68.0	1	0	0.00	-	-	1	1	1.00	67.6	-0.006
2	85.3	23	20	0.87	86.0	0.008	23	17	0.74	89.0	0.043
3	80.2	20	14	0.70	76.6	-0.044	20	12	0.60	76.9	-0.040
4	78.7	9	4	0.44	82.7	0.051	9	4	0.44	83.6	0.062
5	76.2	25	19	0.76	75.1	-0.014	25	18	0.72	75.7	-0.006
6	73.5	2	1	0.50	75.8	0.031	2	1	0.50	77.4	0.053
7	79.9	9	8	0.89	82.0	0.027	9	7	0.78	83.0	0.039
8	80.5	40	32	0.80	80.2	-0.003	40	28	0.70	80.5	0.001
9	-	0	0		-	-	0	-	-	-	-
10	59.9	7	4	0.57	56.6	-0.055	7	5	0.71	57.6	-0.037
11	76.2	5	4	0.80	73.8	-0.032	5	4	0.80	72.8	-0.044
12	86.8	6	5	0.83	89.0	0.025	6	5	0.83	89.7	0.033
13	84.0	5	3	0.60	89.7	0.068	5	5	1.00	86.7	0.032
14	70.1	199	164	0.82	70.6	0.008	199	150	0.75	69.7	-0.006
15	84.3	9	6	0.67	91.5	0.085	9	5	0.56	88.2	0.046
16	77.0	65	46	0.71	79.1	0.027	65	41	0.63	76.9	-0.002
17	76.4	16	11	0.69	77.7	0.018	16	9	0.56	79.8	0.045
18	82.4	5	5	1.00	82.1	-0.004	5	3	0.60	78.3	-0.050
19	76.1	21	19	0.90	76.6	0.007	21	16	0.76	77.9	0.025
20	77.2	26	22	0.85	76.4	-0.010	26	16	0.62	73.5	-0.048
21	73.6	27	16	0.59	79.3	0.077	27	8	0.30	73.5	-0.001
22	69.9	160	121	0.76	72.5	0.037	160	81	0.51	72.5	0.037
23	80.5	3	2	0.67	95.9	0.191	4	1	0.25	65.3	-0.189
24	66.0	18	12	0.67	67.0	0.015	18	11	0.61	64.1	-0.029
25	68.6	16	13	0.81	68.1	-0.007	15	11	0.73	67.2	-0.021
26	73.3	8	7	0.88	77.4	0.056	9	7	0.78	76.4	0.041
27	73.2	50	34	0.68	72.9	-0.004	50	32	0.64	73.3	0.001
Mean				0.71		0.022			0.66		-0.001
Std. dev				0.20		0.050			0.18		0.052
Probability						0.657					0.989

3.3 Lighting Conditions

It is possible that ambient light conditions influence the quality of the video images and hence the accuracy of the halibut length measurements. To examine this possibility each haul was assigned to one of the categories – day, night, or dawn/dusk based on the time the net was retrieved. The dawn/dusk category included retrieval times that occurred within a half hour of sunrise or sunset.

Accuracy of individual fish length measurements was estimated by comparing the census length with EM reviewer length. Error in individual fish measurements was calculated as the proportional difference between the census and EM length.

A two-factor ANOVA was conducted to test for significant differences in the length measurement errors among the ambient light categories, between the two EM reviewers, and interactions between the two factors. Differences in the errors were not significant among the lighting condition categories, but were significant between the two EM reviewers (Table 5). The interaction of light conditions and reviewer was also not significant.

Table 5. Analysis of Variance summary table for the effect of EM reviewer and light conditions on error in length measurements.

Source of Variation	DF	SS	MS	F	P-value
EM reviewer	1	0.0246	0.0246	4.318	0.038
Light condition	1	0.0007	0.0007	0.131	0.718
Interaction	1	0.0135	0.0135	2.377	0.124
Residuals	757	4.3098	0.0057		

3.4 Between reader accuracy and precision

Accuracy of individual fish length measurements was estimated by comparing the census length with EM reviewer length. Error was measured as the proportional difference between the two lengths.

Both reviewers of the EM video images produced accurate estimates of halibut lengths, with no indication of bias in either reviewer's measurements (Table 6). In terms of precision, EMR2 length estimates were somewhat more precise than EMR1 estimates, with error standard deviations of 6.8% versus 8.3%. As shown in Section 3.3 differences in the errors in length measurements between the EM reviewers are statistically significant.

The EM and census counts of halibut show a high level of agreement between the two methods for both EM reviewers (Table 7). EMR1 halibut counts were correct for 22 hauls (81%) and EMR2 counts were correct for 23 hauls (85%). The error in the counts (difference between census and EM count) ranged from -1 to 2, and over all hauls the EMR1 halibut count was the same as the census and the EMR2 halibut count was one more than the census. There is no indication of bias in halibut counts from the EM method, and the standard deviation of the error in halibut counts (~ 0.5) gives an indication of the expected precision of EM halibut counts for individual hauls.

Table 6. Mean, standard deviation, and sample size (N) of the errors ([EM-census]/census) in individual length measurements for the two reviewers, EMR1 and EMR2. The null hypothesis that the mean is equal to zero is tested, and the associated p-value given.

	Error			
Reviewer	mean	Std. Dev.	N	p-value
EMR1	0.007	0.083	382	0.08
EMR1	-0.004	0.068	379	0.26

Table 7 Census and EM reader counts of halibut by haul and reader (EMR1 and EMR2).

	Count			Error (EM- Census)	
Haul	Census	EMR1	EMR2	EMR1	EMR2
1	1	1	1	0	0
2	22	23	23	1	1
3	20	20	20	0	0
4	9	9	9	0	0
5	25	25	25	0	0
6	2	2	2	0	0
7	9	9	9	0	0
8	40	40	40	0	0
9	0	0	0	0	0
10	7	7	7	0	0
11	5	5	5	0	0
12	6	6	6	0	0
13	5	5	5	0	0
14	200	199	199	-1	-1
15	9	9	9	0	0
16	65	65	65	0	0
17	16	16	16	0	0
18	5	5	5	0	0
19	21	21	21	0	0
20	26	26	26	0	0
21	27	27	27	0	0
22	158	160	160	2	2
23	4	3	4	-1	0
24	18	18	18	0	0
25	16	16	15	0	-1
26	9	8	9	-1	0
27	50	50	50	0	0
Total	775	775	776	0	1
Mean				0.000	0.037
Std. Dev.				0.555	0.518

3.5 Quantity of halibut in haul

It is possible that the rate that halibut are released through the discard chute influences the accuracy or precision of the EM length measurements. During the study there were no direct measurements of the rate at which halibut were released, so the number of halibut discarded in each haul is taken as a proxy for processing speed.

Both accuracy and precision of EM length measurements could be affected by the quantity of halibut processed in each haul. To test for potential effects on accuracy, a linear regression was fit to the haul-specific mean proportional error in length measurements (dependent variable) and the halibut counts (independent variable). Data for the two reviewers was combined as an analysis of covariance did not indicate any between reviewer differences in the relationship. The null hypothesis of no trend in the mean errors related to the quantity of halibut processed is not rejected at the 0.05 probability level (Table 8, slope and intercept of relationship not significantly different than zero).

With no indication of a relationship between the accuracy of EM length measurements and the quantity of halibut in the haul, the precision of the measurements was investigated with a linear regression of the absolute value of the proportional error in individual fish measurements (dependent variable) and the number of halibut discarded (independent variables). Again, there was no indication of differences in the regression relationships between the two EM reviewers (analysis of covariance), so the data for the two reviewers was combined to increase the sample size. The null hypothesis that the slope of the relationship is zero is rejected at the $\alpha = 0.05$ significance level (one tailed test that slope is > 0, Table 8). These results suggest that the accuracy of EM measurements are not affected by the quantity of halibut processed in each haul but the precision of the measurements decreases as the quantity of halibut increases.

Table 8 Summary statistics for linear regression fits for 1) mean length error vs halibut count and 2) absolute length errors vs halibut count. Note that for the first regression relationship the p-value is based on a two-tailed test and for the second regression relationship the p-value is based on a one-tailed test.

Hypothesis	Coefficient	Value	Std. error	T	p-value
Accuracy: mean length	Intercept	-0.0062	0.0053	-1.1617	0.2521
error	Slope	0.0002	0.0002	1.1049	0.2756
Precision: absolute	Intercept	0.0264	0.0047	5.5832	0.0000
value of length error	Slope	0.0002	0.0001	1.8555	0.0319

4 Precision and accuracy of halibut estimates

4.1 Observer Sampling

Observer sampling was conducted for each haul of the experimental fishing study. The objective of the observer sampling was to take 2 to 3 sub-samples per haul, but for some hauls only a single sub-sample was taken (Table 9). For each haul, the ratio of halibut (by weight and count) to total sample weight was calculated and this ratio extrapolated to the total haul weight to estimate halibut in the catch:

$$R_h^* = \frac{\sum_{i=1}^{n_h} h_{h,i}^*}{\sum_{i=1}^{n} s_{h,i}}; \qquad H_h^* = R_h^* O_h$$

where n_h is the number of sub-samples in haul *h*,

$h_{h,i}^*$ is the halibut by weight or number in sub-sample *i* of haul *h*,

$s_{h,i}$ is the weight of sub-sample *i* of haul *h*,

R_h^* is the ratio of halibut (by weight or number) to sample weight in haul *h*,

O_h is the OTC weight of halibut in haul *h*,

H_h^* is the estimate of halibut (by weight or number) in haul *h*.

Sub-sample details and the extrapolated halibut estimates by haul are shown in Table 9. Halibut were sampled in 10 of the 27 hauls, and for these hauls the estimated weight of halibut ranged from 21.4 kg to 637.1 kg. For hauls with halibut in the observer samples, the estimated number of halibut ranged from 8.2 to 99.9.

4.1.1 Variance of estimates

For ratio estimators, the formula for calculating the sampling variance of the estimate is an approximation that is valid only for large sample sizes (Cochran, 1977, page 153). This approximation is given by:

$$v\left(H_h^*\right) = O_h^2 v\left(R_h^*\right) \doteq O_h^2 \frac{\left(1-f_h\right)}{\overline{s}_h n_h \left(n_h - 1\right)} \sum_{i=1}^{n_h} \left(h_{h,i}^* - R_h^* s_{h,i}\right)^2$$

where f_h is the finite population correction $\left(f_h = \sum_{i=1}^{n_h} s_{h,i} \middle/ O_h \right)$. Note that no variance is assumed for the O_h estimates, which is not strictly correct.

An alternative approach is to estimate the number of halibut based on the mean per unit approach, $\tilde{H}_h^* = N_h \sum_{i=1}^{n_h} h_{h,i}^* \middle/ n_h$, where N_h is the number of potential sub-samples in the haul $\left(N_h = O_h n_h \middle/ \sum_{i=1}^{n_h} s_{h,i} \right)$. The two estimators, H_h^* and $\tilde{H}_h^*$, are equivalent. The variance of $\tilde{H}_h^*$ is given by:

$$v\left(\tilde{H}_h^*\right) = \frac{N_h^2 \left(1-f_h\right)}{n_h \left(n_h - 1\right)} \sum_{i=1}^{n_h} \left(h_{h,i}^* - \overline{h}_h^*\right).$$

Note that the second approach would be equivalent to the first if the size of all sub-samples was equal. The size of sub-samples is fairly uniform within hauls so the differences in variance estimates for the two approaches should be minor. The mean per unit approach for estimating the sampling variance of the halibut estimates is likely superior, given that the small samples sizes likely invalidate the large sample requirements of the approximate ratio variance estimator.

Standard errors of the halibut weight and count estimates were calculated using both the ratio and mean per unit formulas. The standard errors are large, generally of similar magnitude to the estimates. Standard errors are similar for the two methods, and neither method produced consistently larger estimates (Table 10). Note that for 17 hauls, there were no halibut in the observer samples resulting in standard error estimates of zero. For an additional 5 hauls only one observer sub-sample was taken so the standard errors are not defined for these hauls.

Table 9 Observer sampling data: number of sub-samples (n); weight and count of halibut and total weight by sub-sample; OTC weight, ratio of halibut weight and count to total sample weight, and extrapolated (Ext.) halibut weight and count by haul.

		OTC		Sub-sample wt (kg)			Halibut wt (kg)					Halibut count				
Trip	Haul	(mt)	n	S1	S2	S3	S1	S2	S3	Ratio	Ext.	S1	S2	S3	Ratio	Ext.
1	1	36.58	2	167.9	161.1		0.0	0.0		0.000	0.0	0	0		0.000	0.0
	2	15.35	2	127.4	147.4		0.0	0.0		0.000	0.0	0	0		0.000	0.0
	3	12.97	3	93.0	83.5	80.6	0.0	0.0	0.0	0.000	0.0	0	0	0	0.000	0.0
	4	9.98	2	172.6	149.4		0.0	0.0		0.000	0.0	0	0		0.000	0.0
2	5	33.21	3	153.2	141.7	148.2	0.0	0.0	8.5	0.019	637.1	0	0	1	0.002	75.0
	6	21.44	3	133.5	107.5	147.2	0.0	0.0	0.0	0.000	0.0	0	0	0	0.000	0.0
	7	13.09	2	144.3	154.2		0.0	0.0		0.000	0.0	0	0		0.000	0.0
	8	13.98	2	175.6	187.7		4.8	0.0		0.013	184.7	1	0		0.003	38.5
3	9	5.79	2	142.5	157.6		0.0	0.0		0.000	0.0	0	0		0.000	0.0
	10	14.77	2	140.4	170.7		0.0	0.0		0.000	0.0	0	0		0.000	0.0
	11	11.82	2	118.9	145.7		0.0	9.3		0.035	415.6	0	1		0.004	44.7
	12	32.58	3	103.0	117.2	105.9	0.0	5.2	0.0	0.016	519.7	0	1	0	0.003	99.9
4	13	8.71	2	147.3	116.9		0.0	0.0		0.000	0.0	0	0		0.000	0.0
	14	7.86	1	287.4			7.9			0.027	216.1	1			0.003	27.3
	15	17.23	3	117.7	124.0	106.7	0.0	0.0	0.0	0.000	0.0	0	0	0	0.000	0.0
5	16	18.15	2	147.9	134.2		0.0	3.7		0.013	238.1	0	1		0.004	64.4
	17	10.18	2	129.1	146.3		0.0	0.0		0.000	0.0	0	0		0.000	0.0
	18	16.85	2	142.8	154.9		0.0	0.0		0.000	0.0	0	0		0.000	0.0
	19	21.67	3	105.7	135.9	120.1	0.0	0.0	0.0	0.000	0.0	0	0	0	0.000	0.0
	20	19.13	3	95.1	105.6	113.5	0.0	0.0	0.0	0.000	0.0	0	0	0	0.000	0.0
6	21	12.31	2	145.8	139.5		0.0	0.0		0.000	0.0	0	0		0.000	0.0
	22	11.94	2	155.0	152.5		0.0	0.0		0.000	0.0	0	0		0.000	0.0
	23	0.54	1	279.7			34.6			0.124	66.8	2			0.007	3.9
	24	0.91	1	317.8			7.5			0.024	21.4	2			0.006	5.7
	25	1.81	1	354.9			16.6			0.047	84.7	4			0.011	20.4
	26	1.13	1	274.1			7.9			0.029	32.6	2			0.007	8.2
	27	11.23	2	182.4	174.1		0.0	0.0		0.000	0.0	0	0		0.000	0.0

Table 10 Estimates of halibut weight and number of halibut by haul, and their estimated standard errors based on ratio and mean per unit estimators.

	Halibut weight (kg)			Number of halibut		
		Standard Error			Standard Error	
Haul	Estimate	Ratio	Mean/unit	Estimate	Ratio	Mean/unit
1	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0
5	637.1	632.0	632.9	75.0	74.4	74.5
6	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0
8	184.7	188.4	182.3	38.5	39.2	38.0
9	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0
11	415.6	369.3	410.9	44.7	39.7	44.2
12	519.7	496.9	517.1	99.9	95.6	99.4
13	0.0	0.0	0.0	0.0	0.0	0.0
14	216.1	-	-	27.3	-	-
15	0.0	0.0	0.0	0.0	0.0	0.0
16	238.1	247.8	236.3	64.4	67.0	63.9
17	0.0	0.0	0.0	0.0	0.0	0.0
18	0.0	0.0	0.0	0.0	0.0	0.0
19	0.0	0.0	0.0	0.0	0.0	0.0
20	0.0	0.0	0.0	0.0	0.0	0.0
21	0.0	0.0	0.0	0.0	0.0	0.0
22	0.0	0.0	0.0	0.0	0.0	0.0
23	66.8	-	-	3.9	-	-
24	21.4	-	-	5.7	-	-
25	84.7	-	-	20.4	-	-
26	32.6	-	-	8.2	-	-
27	0.0	0.0	0.0	0.0	0.0	0.0

4.1.2 Hypothesis Tests

Hypothesis tests were conducted to assess whether the observer sampling estimates of the number and weight of halibut caught were significantly different from the census estimates. The null hypothesis tested was that the difference between the census and observer sampling estimate were zero. Two-tailed t-tests $(\alpha = 0.05)$ were conducted for individual hauls, trips, and over the study.

The calculation of the variance of the number and weight of halibut caught at the trip and study level are estimated following the study design. That is, hauls and trips are not random samples but rather are part of the study design. Thus the variance of the summed halibut estimates is the sum of the variances of the hauls included in the sum. The mean per unit estimators of standard errors were used for conducting the hypothesis tests.

Hypotheses tests could only be conducted for 5 of the 27 hauls; the remainder had either only one sub-sample or had standard error estimates of zero. The null hypotheses of no differences between the observer sampling estimates and the census halibut weights and counts were not rejected for these 5 hauls (Table 11). Hypotheses tests at the trip level, conducted for 3 of the 6 trips, rejected the null hypothesis in one case – the observer halibut weight estimate for trip 5 (Table 11). For the combined trips, 2, 3, and 5, the null hypotheses were also not rejected.

Additional hypotheses tests were conducted to test if the mean proportion error in the halibut estimates ([observer-census]/census) indicated bias in the observer estimates. Across both hauls and trips, the mean proportion errors were not significantly different from zero (Table 11).

The observer sampling should be representative of the total halibut catch, not just halibut enumerated in the census, so more appropriate hypotheses tests would compare observer estimates to the total halibut catch. Halibut that were landed (and therefore not in the census) could not always be attributed to a single haul but rather to a set of two hauls. Total halibut weights and counts (census plus landings) were calculated for hauls or sets of hauls for which the landings could be uniquely attributed (Table 12).

The hypotheses tests described above were repeated comparing observer estimates with the total halibut catch. Results were the same as for the tests comparing observer estimates with the census values. The null hypothesis of no difference between the observer estimates and total halibut weights and counts was rejected in just one case – the observer halibut weight estimate for trip 5 (Table 13).

Non-parametric statistical tests may be more appropriate to test for bias in the observer-based halibut estimates, as the distribution of errors is unlikely to be normal. The Wilcoxon signed-rank test is appropriate to test whether the proportional errors in halibut estimates are significantly different than zero. This test was conducted for proportional errors in the halibut counts and weights at the haul and fishing event level. For the haul-based errors, the null hypothesis that the proportional errors were equal to zero was rejected for both the halibut weights and counts (p-values of 0.019 and 0.015 for weights and counts, respectively). For the fishing event-based estimates, the null hypothesis that the proportional errors were equal to zero was not rejected for either the halibut weights or counts (p-values of 0.137 and 0.111 for weights and counts, respectively). For the haul-based proportional errors these test results are different than those from the parametric tests. The large number of hauls that had no halibut in the observer samples (16 of 27) results in a preponderance of negative proportional errors and hence rejection of the hypothesis that the median error is zero.

Table 11 Comparison of halibut weight (kg) and count estimates (N) from observer samples with census values, by haul and trip. Error is measured as the proportional difference between observer and census values ([observer-census]/census), and the null hypothesis that the mean error across hauls is equal to zero is tested. Where sample sizes are adequate and halibut were sampled, the null hypothesis that of no difference between the census value and observer estimates is tested for individual hauls and trips.

		Census		Observer weight estimate				Observer count estimate			
Trip	Haul	Kg.	N	Est.	St. Error	p-value	Error	Est.	St. Error	p-value	Error
1	1	3.6	1	0.0	0.0		-1.00	0.0	0.0		-1.00
	2	172.2	22	0.0	0.0		-1.00	0.0	0.0		-1.00
	3	129.5	20	0.0	0.0		-1.00	0.0	0.0		-1.00
	4	54.3	9	0.0	0.0		-1.00	0.0	0.0		-1.00
	Trip total	359.5	52	0.0	0.0		-1.00	0.0	0.0		-1.00
2	5	137.2	25	637.1	632.9	0.51	3.64	75.0	74.5	0.57	2.00
	6	9.4	2	0.0	0.0		-1.00	0.0	0.0		-1.00
	7	59.4	9	0.0	0.0		-1.00	0.0	0.0		-1.00
	8	271.4	40	184.7	182.3	0.72	-0.32	38.5	38.0	0.97	-0.04
	Trip total	477.4	76	821.8	658.6	0.62	0.72	113.4	83.6	0.67	0.49
3	9	0.0	0	0.0	0.0			0.0	0.0		
	10	17.3	7	0.0	0.0		-1.00	0.0	0.0		-1.00
	11	27.5	5	415.6	410.9	0.52	14.11	44.7	44.2	0.53	7.94
	12	49.3	6	519.7	517.1	0.46	9.54	99.9	99.4	0.44	15.66
	Trip total	94.1	18	935.3	660.5	0.26	8.94	144.6	108.8	0.30	7.03
4	13	37.2	5	0.0	0.0		-1.00	0.0	0.0		-1.00
	14	844.4	200	216.1	-		-0.74	27.3	-		-0.86
	15	67.6	9	0.0	0.0		-1.00	0.0	0.0		-1.00
	Trip total	949.2	214	216.1	-		-0.77	27.3	-		-0.87
5	16	391.9	65	238.1	236.3	0.63	-0.39	64.4	63.9	0.99	-0.01
	17	94.4	16	0.0	0		-1.00	0.0	0.0		-1.00
	18	33.9	5	0.0	0.0		-1.00	0.0	0.0		-1.00
	19	122.4	21	0.0	0.0		-1.00	0.0	0.0		-1.00
	20	152.7	26	0.0	0.0		-1.00	0.0	0.0		-1.00
	Trip total	795.2	133	238.1	236.3	0.05	-0.70	64.4	63.9	0.32	-0.52
6	21	138.8	27	0.0	0.0		-1.00	0.0	0.0		-1.00
	22	665.8	158	0.0	0.0		-1.00	0.0	0.0		-1.00
	23	37.0	4	66.8	-		0.81	3.9	-		-0.03
	24	61.6	18	21.4	-		-0.65	5.7	-		-0.68
	25	70.0	16	84.7	-		0.21	20.4	-		0.28
	26	47.2	9	32.6	-		-0.31	8.2	-		-0.08
	27	255.9	50	0.0	0.0		-1.00	0.0	0.0		-1.00
	Trip total	1276.3	282	205.5	-		-0.84	38.2	-		-0.86
	Trips 2+3+5	1366.7	227	1995.2	962.2	0.52	0.46	322.4	151.3	0.54	0.42
	All trips	3951.7	775	2416.7			-0.39	388.0			-0.50
		Mean over hauls					0.380				0.314
		Std. error					0.699				0.709
		P-value					0.59				0.66
		Mean over trips					1.058				0.712
		Std error					1.597				1.284
		P-value					0.54				0.60

Table 12 Number (N) and weight (kg) of halibut in census, landings records, and total (census plus landings) by haul or sets of hauls for which landings are uniquely defined.

	Census		Landing		Total	
Haul	Kg.	N	Kg.	N	Kg.	N
1,2	129.7	16	175.8	23	305.5	39
3,4	91.2	14	183.7	29	274.9	43
5	34.0	5	137.2	25	171.2	30
6	0.0	0	9.4	2	9.4	2
7,8	10.4	2	330.8	49	341.2	51
9	0.0	0	0.0	0	0.0	0
10	0.0	0	17.3	7	17.3	7
11	0.0	0	27.5	5	27.5	5
12	0.0	0	49.3	6	49.3	6
13	0.0	0	37.2	5	37.2	5
14	10.0	3	844.4	200	854.4	203
15	0.0	0	67.6	9	67.6	9
16,17	2.9	1	486.3	81	489.2	82
18,19	3.6	1	156.3	26	159.9	27
20	13.6	2	152.7	26	166.3	28
21	0.0	0	138.8	27	138.8	27
22,27	22.2	6	921.7	208	944.0	214
23	0.0	0	37.0	4	37.0	4
24	0.0	0	61.6	18	61.6	18
25	0.0	0	70.0	16	70.0	16
26	0.0	0	47.2	9	47.2	9

Table 13 Comparison of halibut weight (kg) and count estimates (N) from observer samples with total values, by haul and trip. Error is measured as the proportional difference between observer and total values ([observer-total]/total), and the null hypothesis that the mean error across hauls is equal to zero is tested. Where sample sizes are adequate and halibut were sampled, the null hypothesis that of no difference between the total value and observer estimates is tested for individual hauls and trips.

	Fishing event	Total (census + landings)		Observer weight estimate				Observer count estimate			
Trip	(hauls)	Kg.	N	Kg.	Std. error	p-value	error	N	Std. error	p-value	Error
1	1,2	305.5	39	0.0	0.0		-1.00	0.0	0.0		-1.00
	3,4	274.9	43	0.0	0.0		-1.00	0.0	0.0		-1.00
	Trip total	580.4	82	0.0	0.0		-1.00	0.0	0.0		-1.00
2	5	171.2	30	637.1	632.9	0.54	2.72	75.0	74.5	0.61	1.50
	6	9.4	2	0.0	0.0		-1.00	0.0	0.0		-1.00
	7,8	341.2	51	184.7	182.3	0.48	-0.46	38.5	38.0	0.77	-0.25
	Trip total	521.8	83	821.8	658.6	0.66	0.58	113.4	83.6	0.73	0.37
3											
	9	0.0	0	0.0	0.0			0.0	0.0		
	10	17.3	7	0.0	0.0		-1.00	0.0	0.0		-1.00
	11	27.5	5	415.6	410.9	0.52	14.11	44.7	44.2	0.53	7.94
	12	49.3	6	519.7	517.1	0.46	9.54	99.9	99.4	0.44	15.66
	Trip total	94.1	18	935.3	660.5	0.26	8.94	144.6	108.8	0.30	7.03
4	13	37.2	5	0.0	0.0		-1.00	0.0	0.0		-1.00
	14	854.4	203	216.1	-		-0.75	27.3	-		-0.87
	15	67.6	9	0.0	0.0		-1.00	0.0	0.0		-1.00
	Trip total	959.2	217	216.1	-		-0.77	27.3	-		-0.87
5	16,17	489.2	82	238.1	236.3	0.40	-0.51	64.4	63.9	0.81	-0.22
	18,19	159.9	27	0.0	0.0		-1.00	0.0	0.0		-1.00
	20	166.3	28	0.0	0.0		-1.00	0.0	0.0		-1.00
	Trip total	815.4	137	238.1	236.3	0.04	-0.71	64.4	63.9	0.29	-0.53
6	21	138.8	27	0.0	0.0		-1.00	0.0	0.0		-1.00
	22,27	944.0	214	0.0	0.0		-1.00	0.0	0.0		-1.00
	23	37.0	4	66.8	-		0.81	3.9	-		-0.03
	24	61.6	18	21.4	-		-0.65	5.7	-		-0.68
	25	70.0	16	84.7	-		0.21	20.4	-		0.28
	26	47.2	9	32.6	-		-0.31	8.2	-		-0.08
	Trip total	1298.5	288	205.5	-		-1.00	38.2	-		-0.87
	Trip 2+3+5	1431.3	238	1995.2	962.2	0.57	0.39	322.4	151.3	0.59	0.35
	All trips	4269.4	825	2416.7			-0.43	388.0			-0.53
Mean over fishing events							0.735				0.662
		Std. error					0.887				0.908
		P-value					0.42				0.47
		Mean over trips					1.031				0.688
		Std error					1.598				1.286
		P-value					0.55				0.62

4.1.3 Bootstrap distribution

An alternative to the analytical survey-based method for estimating uncertainty in halibut estimates is the bootstrap approach (Efron and Tibshirani, 1998). Bootstrap methods do not require assumptions about the underlying distribution of the population being sampled (normal distribution with constant homogeneous variance), so can be superior to classical statistical methods when those assumptions are not met.

The bootstrap procedure follows the experimental sampling design. The trips and hauls are treated as fixed design elements, rather than part of a random selection process. Therefore, for each bootstrap replicate each haul is sampled and then sub-samples are randomly selected with replacement. The ratio of halibut to sample weight for the selected sub-samples is calculated and prorated to the OTC weight for the haul. Halibut weights are summed across all hauls to generate each bootstrap replicate and the resultant distribution of halibut weight and numbers reflect the uncertainty due to the sampling process. The bootstrap procedure was applied to the entire study (6 trips) and to combined trips 2, 3, and 5. 2000 bootstrap replicates were generated to represent the uncertainty in the halibut estimates. Note that while the number of unique combinations of sub-samples in each haul is small (3 combinations when there are 2 sub-samples and 10 combinations when there are 3 sub-samples), the number of unique combinations across the study (27 hauls) is large (order of $1e^{14}$).

The bootstrap and analytical estimates of halibut weight and numbers, and their 95% confidence intervals, are compared with the total values (census plus landings) in Table 14. The bootstrap frequency distributions of weight estimates are shown in Figure 2. The bootstrap confidence intervals are narrower than the analytical estimates, but for the "all trips combined" the 95% limits do not encompass the total weight or numbers of halibut caught. For combined trips 2,3, and 5, both the bootstrap and analytical 95% confidence limits encompass the census plus landings values. For the bootstrap distributions of weight estimates the coefficients of variation (c.v.s) are 32% for all trips combined and 38% for trips 2, 3, and 5 combined.

Table 14 Halibut weight (kg) and numbers from the total estimate (census plus landings) and from observer sampling using bootstrap and analytical approachs for all trips and combined trips 2, 3, and 5. 95% confidence intervals are given in parentheses.

	Weight (kg)			Number		
		Observer-based estimates			Observer-based estimates	
	Total	Bootstrap	Analytical	Total	Bootstrap	Analytical
All trips	4269.4	2388.3 (1021.8 - 3928.6)	2416.7	825	384.5 (168.4 - 628.4)	388.0
Trips 2,3,5	1431.3	1942.0 (600.3 - 3429.8)	1995.2 (109.6 - 3881.4)	238	315.9 (102.8 - 558.1)	322.4 (25.9 - 618.9)

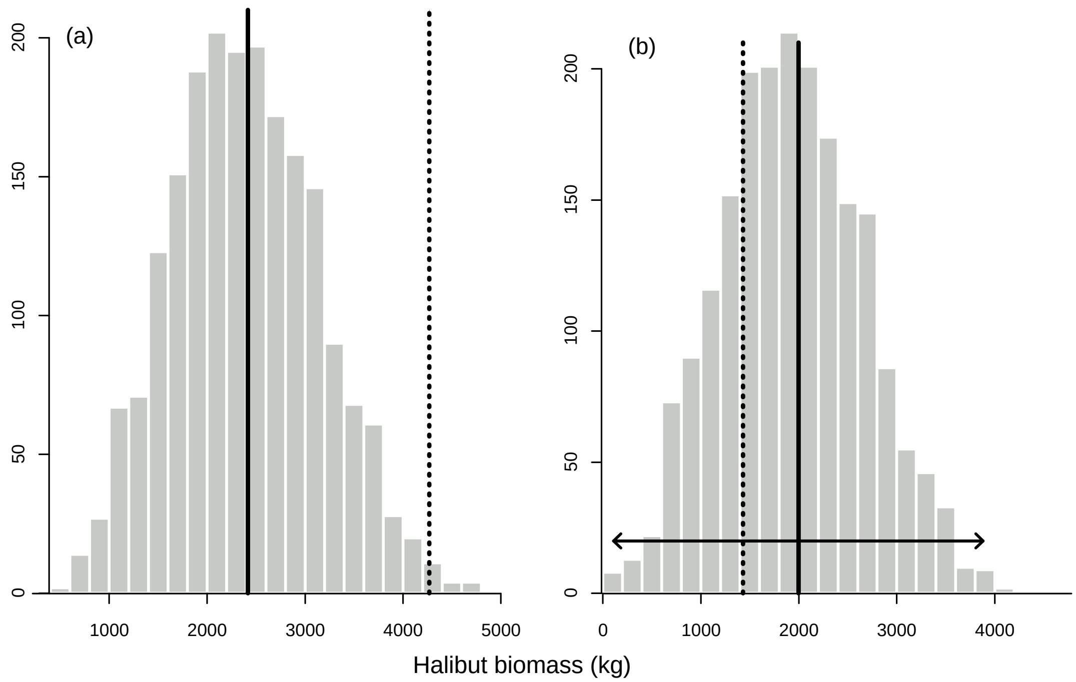

Figure 2 Bootstrap frequency distribution of observer-based halibut weight (kg) for: (a) all trips and (b) combined trips 2, 3, and 5. Actual values (census plus landings) are shown by the dashed vertical lines; observer-based estimates are shown by solid vertical lines with the 95% confidence interval shown by the horizontal arrow.

4.2 EM method

The EM method for estimating halibut catch is not based on sampling theory. All halibut are enumerated and their images measured to estimate their length. Lengths are converted to weight (based on standard INPFC conversions) and weights summed to estimate the halibut weight for each haul (Table 15). Uncertainty arises because counts and length estimates have associated errors, rather than through a sampling process. It is not possible to calculate the uncertainty of the EM halibut weight and count estimates. Accuracy and precision estimates for this method are based on the error of the EM relative to the census.

Error in EM estimates are calculated as proportional differences, $(EM - census)/census$, for halibut weight and numbers. The error in EM weight estimates ranges from –30% to 43% across all hauls, with means of 2.2% and 0.1% for EMR1 and EMR2, respectively (Table 15). For halibut count estimates, the errors range from –25% to 2% with means of –1.1% and 0.0% for EMR1 and EMR2, respectively. When weight and count estimates are summed across hauls within fishing trips, the mean error for weight estimates increases slightly and the mean error for count estimates decreases (Table 15). The null hypothesis that the mean errors are equal to zero is not rejected for any of the tests conducted (weight and counts across hauls and trips for EMR1 and EMR2, Table 15). This supports that conclusion that there is no evidence of bias in the EM estimates of halibut discard weights and numbers.

Table 15 Comparison of EM estimates (EMR1 and EMR2) with census values for halibut count and weight by haul and trip. EM errors are calculated as proportional differences ([EM-census]/census), and the null hypothesis that the mean error across hauls and trips is equal to zero is tested.

		Census		EM weight				EM number			
				Kg.		Error		N		Error	
Trip	Haul	Kg.	N	EMR1	EMR2	EMR1	EMR2	EMR1	EMR2	EMR1	EMR2
1	1	3.6	1	2.5	3.6	-0.30	-0.02	1	1	0.00	0.00
	2	172.2	22	176.9	188.2	0.03	0.09	23	23	0.05	0.05
	3	129.5	20	120.7	118.1	-0.07	-0.09	20	20	0.00	0.00
	4	54.3	9	54.9	57.5	0.01	0.06	9	9	0.00	0.00
	Trip total	359.5	52	355.1	367.2	-0.01	0.02	53	53	0.02	0.02
2	5	137.2	25	196.7	135.0	0.43	-0.02	25	25	0.00	0.00
	6	9.4	2	9.6	9.5	0.02	0.01	2	2	0.00	0.00
	7	59.4	9	61.0	64.4	0.03	0.08	9	9	0.00	0.00
	8	271.4	40	263.6	268.0	-0.03	-0.01	40	40	0.00	0.00
	Trip total	477.4	76	530.8	476.9	0.11	0.00	76	76	0.00	0.00
3	9	0.0	0	0.0	0.0						
	10	17.3	7	16.4	16.2	-0.05	-0.06	7	7	0.00	0.00
	11	27.5	5	26.5	26.7	-0.04	-0.03	5	5	0.00	0.00
	12	49.3	6	50.1	49.8	0.02	0.01	6	6	0.00	0.00
	Trip total	94.1	18	92.9	92.8	-0.01	-0.01	18	18	0.00	0.00
4	13	37.2	5	39.7	41.2	0.07	0.11	5	5	0.00	0.00
	14	844.4	200	858.7	837.8	0.02	-0.01	199	199	-0.01	-0.01
	15	67.6	9	74.9	67.5	0.11	0.00	9	9	0.00	0.00
	Trip total	949.2	214	973.3	946.5	0.03	0.00	213	213	0.00	0.00
5	16	391.9	65	411.3	387.4	0.05	-0.01	65	65	0.00	0.00
	17	94.4	16	92.5	93.1	-0.02	-0.01	16	16	0.00	0.00
	18	33.9	5	33.4	30.3	-0.01	-0.11	5	5	0.00	0.00
	19	122.4	21	124.2	122.7	0.02	0.00	21	21	0.00	0.00
	20	152.7	26	164.0	145.9	0.07	-0.04	26	26	0.00	0.00
	Trip total	795.2	133	825.4	779.5	0.04	-0.02	133	133	0.00	0.00
6	21	138.8	27	148.4	132.6	0.07	-0.04	27	27	0.00	0.00
	22	665.8	158	733.7	699.1	0.10	0.05	160	160	0.01	0.01
	23	37.0	4	30.5	35.2	-0.18	-0.05	3	4	-0.25	0.00
	24	61.6	18	63.4	58.7	0.03	-0.05	18	18	0.00	0.00
	25	70.0	16	74.9	69.5	0.07	-0.01	16	15	0.00	-0.06
	26	47.2	9	49.7	49.3	0.05	0.05	8	9	-0.11	0.00
	27	255.9	50	280.6	283.9	0.10	0.11	50	50	0.00	0.00
	Trip total	1276.3	282	1381.1	1328.3	0.08	0.04	282	283	0.00	0.00
	All trips	3951.7	775	4158.7	3991.1	0.05	0.01	775	776	0.00	0.00
	Mean over hauls					0.022	0.001			-0.011	0.000
	Std. dev.					0.119	0.056			0.053	0.015
	p-value					0.36	0.95			0.28	0.91
	Mean over trips					0.039	0.004			0.002	0.003
	Std. dev.					0.050	0.023			0.008	0.008
	p-value					0.12	0.68			0.51	0.41

Table 16 Comparison of EM estimates (EMR1 and EMR2) with total values (census plus landings) for halibut count and weight by haul and trip. EM errors are calculated as proportional differences ([EM-census]/census), and the null hypothesis that the mean error across hauls and trips is equal to zero is tested.

		Total		EM weight (Kg.)				EM number			
	Fishing events			Estimate		Error		Estimate		Error	
Trip	(hauls)	Kg.	N	EMR1	EMR2	EMR1	EMR2	EMR1	EMR2	EMR1	EMR2
1	1,2	305.5	39	179.5	191.7	-0.41	-0.37	24	24	-0.38	-0.38
	3,4	274.9	43	175.6	175.5	-0.36	-0.36	29	29	-0.33	-0.33
	trip total	580.4	82	355.1	367.2	-0.39	-0.37	53	53	-0.35	-0.35
2	5	171.2	30	196.7	135.0	0.15	-0.21	25	25	-0.17	-0.17
	6	9.4	2	9.6	9.5	0.02	0.01	2	2	0.00	0.00
	7,8	341.2	51	324.5	332.3	-0.05	-0.03	49	49	-0.04	-0.04
	trip total	521.8	83	530.8	476.9	0.02	-0.09	76	76	-0.08	-0.08
3											
	9	0.0	0	0.0	0.0			0	0		
	10	17.3	7	16.4	16.2	-0.05	-0.06	7	7	0.00	0.00
	11	27.5	5	26.5	26.7	-0.04	-0.03	5	5	0.00	0.00
	12	49.3	6	50.1	49.8	0.02	0.01	6	6	0.00	0.00
	trip total	94.1	18	92.9	92.8	-0.01	-0.01	18	18	0.00	0.00
4	13	37.2	5	39.7	41.2	0.07	0.11	5	5	0.00	0.00
	14	854.4	203	858.7	837.8	0.01	-0.02	199	199	-0.02	-0.02
	15	67.6	9	74.9	67.5	0.11	0.00	9	9	0.00	0.00
	trip total	959.2	217	973.3	946.5	0.01	-0.01	213	213	-0.02	-0.02
5	16,17	489.2	82	503.7	480.6	0.03	-0.02	81	81	-0.01	-0.01
	18,19	159.9	27	157.7	153.0	-0.01	-0.04	26	26	-0.04	-0.04
	20	166.3	28	164.0	145.9	-0.01	-0.12	26	26	-0.07	-0.07
	trip total	815.4	137	825.4	779.5	0.01	-0.04	133	133	-0.03	-0.03
6	21	138.8	27	148.4	132.6	0.07	-0.04	27	27	0.00	0.00
	22,27	944.0	214	1014.2	983.0	0.07	0.04	210	210	-0.02	-0.02
	23	37.0	4	30.5	35.2	-0.18	-0.05	3	4	-0.25	0.00
	24	61.6	18	63.4	58.7	0.03	-0.05	18	18	0.00	0.00
	25	70.0	16	74.9	69.5	0.07	-0.01	16	15	0.00	-0.06
	26	47.2	9	49.7	49.3	0.05	0.05	8	9	-0.11	0.00
	Trip total	1298.5	288	1381.1	1328.3	0.06	0.02	282	283	-0.02	-0.02
	All trips	4269.4	825	4158.7	3991.1	-0.03	-0.07	775	776	-0.06	-0.06
Mean over fishing events						-0.021	-0.060			-0.072	-0.057
Std. dev.						0.143	0.123			0.117	0.110
p-value						0.05	0.04			0.01	0.03
Mean over trips						-0.071	-0.105			-0.097	-0.097
Std. dev.						0.178	0.150			0.147	0.147
p-value						0.37	0.15			0.17	0.17

The standard deviations of the errors in the EM halibut weight and count estimates provide an indication of the expected precision of this methodology. On a haul basis, the standard deviation of the error in weight estimates was 11.9% and 5.6% for EMR1 and EMR2, respectively. On a trip basis, these values decreased to 5.0% and 2.3% for the two reviewers (Table 15). For the halibut count estimates, the standard deviation of the error was substantially lower at the trip level (0.8% and 0.8% for EMR1 and EMR2, respectively) than at the haul level (5.3% and 1.5% for EMR1 and EMR2, respectively).

Equivalent error calculations and hypotheses tests were conducted comparing EM-based halibut estimates to the total halibut estimates (census plus landings), as was done for the observer sampling. The EM method is designed to measure discards, so it should provide estimates that are closer to the census values than the total catch values.

The errors in halibut weight and number estimates generally increase when EM-based estimates are compared with the total (census plus landings values) rather than the census values (Table 16). The standard deviation of the weight errors across hauls increases from 11.9% to 14.3% for EMR1 and from 5.6% to 12.3% for EMR2. The standard deviations estimated across fishing trips increase to an even greater degree (Table 15, Table 16).

The null hypothesis that the error in EM weight and number estimates (relative to the total values) is zero, is rejected at the $\alpha = 0.05$ level for both EMR1 and EMR2 when evaluated across hauls (Table 16). This result is not surprising, given the EM method surveyed discards only, not the entire catch and substantial numbers of halibut were landed for some trips. Bias in the EM estimates relative to total catch would be expected.

4.2.1 Bootstrap distribution

A bootstrap-type approach was also applied to the EM data. As for the observer sampling data, the bootstrap procedure followed the experimental design of the EM component of the study. However, because the EM method did not involve random sampling, a random component was introduced by including random error in each of the halibut length measurements and in the count of the number of fish.

Errors were calculated as follows:

For each haul, the error in the number of enumerated fish was calculated (EM count – Census count, *count error*). The error in individual fish measurements was calculated as the proportional difference between the EM and census length ([EM length – census length]/census length). Hauls where the number of halibut enumerated differed between the census and the EM reviewer were not included in these measurements, as the association between census and reviewer lengths could not be established. For each haul where length errors could be determined, the *haul mean length error* was calculated. Finally, the *haul mean length error* was subtracted from individual fish length errors to generate a series of *fish length errors* with mean zero across each haul

To generate bootstrap replicates, each of the hauls was re-sampled. For each haul a *count error* and a *haul mean length error* was randomly selected. Then the lengths of each of the fish in the census was adjusted based on the *haul mean length error* and a randomly selected *fish length error*. Each modified length was translated to weight, and the mean weight for the haul calculated. Finally, the haul weight was calculated based on the count of fish in the census adjusted by the randomly selected *count error*. 2000 bootstrap replicates were generated to represent the bootstrap distribution of uncertainty in the weight of halibut discards for the study.

The bootstrap and analytical estimates of halibut weight and numbers, and 95% confidence intervals for the bootstrap distribution, are compared with the census values in Table 14. The bootstrap frequency distributions of weight estimates are shown in Figure 3. For both EM reviewers, the 95% bootstrap confidence intervals include the census values. The distribution of weight estimates is tighter for EMR2 than EMR1, with c.v.s of 2.4% and 3.5%, respectively, The c.v.s for the estimates of halibut numbers are even smaller at 0.33% and 0.36% for EMR2 and EMR1, respectively.

Table 17 Halibut weight (kg) and numbers from the census and from the EM analytical and bootstrap methods for EMR1 and EMR2. 95% confidence intervals are given in parentheses.

	Weight (kg)			Number		
		EM-based estimates			EM-based estimates	
	Census	Bootstrap	Analytical	Census	Bootstrap	Analytical
EMR1		4093.3	4158.7		775.0	775
		(3788.2 - 4384.7)			(770.0 – 781.0)	
	3951.7			775		
EMR2		3984.2	3991.1		775.9	776
		(3820.8 – 4193.7)			(771.0 – 181.0)	

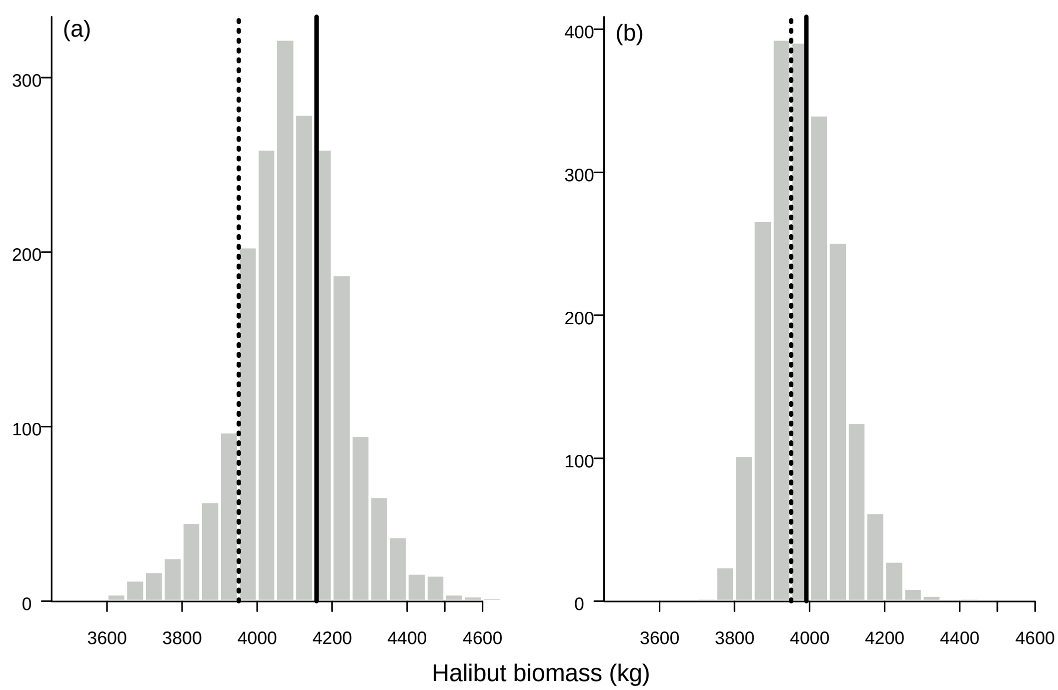

Figure 3 EM-based bootstrap frequency distribution of study total halibut weight (kg) for: (a) EMR1 and (b) EMR2. Census values are shown by dashed vertical lines; EM point estimates are shown by solid vertical lines.

5 EM simulations

The bootstrap approach used to calculate confidence intervals for the EM-based estimates of halibut discards was extended to examine the question of how reliable the estimates would be if not all halibut in the video record were measured. The sampling regimes examined included measuring every 3rd, every 5th, or every 7th halibut. The approach used was as follows: Let *k* be the frequency for sampling halibut. For each of the 27 hauls, a random number between 1 and *k* was selected. This represents the first fish to measure, and subsequently every *k*th fish is measured. In cases where there were fewer fish than the first randomly selected one, the first fish was measured.

Errors in fish lengths and in the enumeration of fish in each haul were introduced as for the bootstrap analysis (Section 4.2.1). For each sampling regime, 2000 replicates were generated to represent the sampling distribution.

Decreasing the frequency at which fish lengths are measured has little effect on the average estimate of halibut weight, indicating it does not result in biased estimates (Table 18). The variance of the distribution of weight estimates does increase, with c.v.s increasing from 3.5% to 7.6% for EMR1 and from 2.4% to 6.9% for EMR2 when fish measurements change from every fish to every 7th fish.

Table 18 Mean, standard deviation and c.v. of halibut weight estimates, under alternative frequencies of measuring fish length.

	1		3rd		5th		7th	
	EMR1	EMR2	EMR1	EMR2	EMR1	EMR2	EMR1	EMR2
Mean	4094.3	3984.2	4106.9	3987.7	4093.5	3976.2	4062.5	3969.1
Std. Dev.	144.2	96.7	233.6	191.6	279.6	222.1	307.4	272.6
c.v.	0.035	0.024	0.057	0.048	0.068	0.056	0.076	0.069

6 Summary and Conclusions

In general, halibut estimates from observer sampling and EM methods are not biased. The observer-based estimates of halibut weight and numbers were not significantly different from the total (census plus landings) values. However, the precision of the observer estimates was low, in particular at the haul level. This is not surprising, given the observer program is not designed to estimate catch abundance at the haul level. For the observer estimates, the bootstrap distributions of halibut weight and numbers were narrower than the analytical estimates, and are likely more reliable given they do not require adherence to assumptions of normality. For the bootstrap approach, the c.v. of the estimated weight of halibut caught over all trips in the study was 32%.

For the EM method, the null hypotheses of no difference between halibut estimates and census values were not rejected, indicating no bias in the EM method. Differences between EM estimates and total values (census plus landings) were significant, but this is expected given the EM method enumerates discards, not total catch. Precision of the EM estimates was high. The bootstrap c.v.s of the weight of halibut discarded over all fishing trips was 3.5% and 2.4% for EM reviewers EMR1 and EMR2, respectively.

Although both the observer and EM methods appear to produce unbiased halibut estimates, precision is much higher for the EM method. The most appropriate comparison of precision for the two methods is the c.v.s from the bootstrap distribution of halibut estimates for the entire study: 32% for the observer method and 3.5% (EMR1) and 2.4% (EMR2) for the EM method. At the haul or fishing event level, the distribution of proportional errors provides a comparison of the relative precision of the two methods. The fishing event level errors in observer weight estimates ranged from -100% to +1400%, with an average absolute error of 190%. The haul level errors in EM weight estimates ranged from -30% to 43%, with an average absolute error of 6%. These comparisons reflect errors relative to the quantities surveyed by each method; total catch for the observer estimates (Table 13), and discards for the EM estimates (Table 15),

A simulation was conducted to investigate how uncertainty in EM halibut estimates would increase if only a sub-set of the halibut EM images were measured. Measuring a sub-set of the halibut images could increase the speed of processing the EM data, and hence potentially decrease the associated costs. Measuring every image, then c.v.s of the estimates of halibut discard weight (over all trips) were 3.5% and 2.4% for EMR1 and EMR2, respectively. When every 7th image was measured the c.v.s increased to 7.6% and 6.9% for the two reviewers. The c.v.s describe the expected distribution of errors in the halibut estimates: the halibut estimate would be within $\pm 3.5\%$ of the true number with every 3rd fish measured and within $\pm 7.6\%$ of the true number with every 7th fish measured, 65% of the time, given the precision of the EMR1 measurements.

For the EM method a number of covariates were examined to see if they influenced the accuracy of the estimates of halibut discards. Excluding length measurements for images that were unclear or otherwise obstructed, did not increase the accuracy or precision in the haul level mean length estimates. Also, there was no apparent difference in the accuracy of fish measurements related to the time of day (ambient light conditions) that the video record was made. The number of fish processed each haul did not influence the accuracy of the EM length measurements, but precision of the measurements appear to have decreased as the quantity of halibut increased.

The EM video images were somewhat distorted so that factors for converting image size to fish length differed between the *top of grid* and *bottom of grid* conversions. Using *bottom of grid* conversions resulted in biased length estimates, while using *top of grid* conversions did not. This could result from the majority of fish lying along the top of the discard chute grid, but this is unknown. Given the difference between *top of grid* and *bottom of grid* conversions was significant, efforts to ensure the video images are not distorted in future EM trials is warranted.

7 References

Efron, B, and R.J. Tibshirani. 1998. An introduction to the bootstrap. Chapman & Hall. New York. 436 p.

Cochran, W. G. 1977. Sampling techniques. 3rd edition. John Wiley & Sons. New York. 428 p.

Appendix "A". Statement of Work, dated December 13, 2007

Analysis of the Alaska Groundfish Data Bank exempted fishing permit study to evaluate an electric monitoring program for estimating the weight of halibut discards

The Electronic Monitoring (EM) study encompassed 27 hauls made over 6 fishing trips. For the study, halibut estimates were made using observer sampling (OS), electronic monitoring videotape records (EM), a census of halibut discards (C), and records of halibut landed at the fishing plant (L).

The observer samples were taken prior to any discarding, and hence the samples are representative of the total halibut catch. This data was collected for each haul and includes the weight of sub-samples and the weight and count of halibut in the sub-samples. The halibut catch is estimated by extrapolating the proportion of halibut in the sum of the sub-samples (by weight) to the OTC estimate of the catch.

The electric monitoring data is a video-record of all halibut discards. Video-records were made of each haul and the records reviewed by two individuals (EMR1 and EMR2) to estimate the number of discarded halibut and the length of each halibut. Halibut lengths are converted to weight using standard IPHC conversion tables.

An on-board census was conducted to enumerate the halibut discards for each haul. Data includes the total number and individual lengths of each discarded halibut. Halibut lengths were measured sequentially as they were discarded so that individual lengths can be compared with those obtained from the EM method.

Some halibut were inadvertently landed rather than discarded and information is available from plant landings records. For each fishing trip, the count and weight of landed halibut was recorded for each of three tanks on the fishing vessel. The catch from each haul was associated with a particular tank, so all landed halibut can be associated with either a single haul or a set of two hauls. The term "fishing event" is used to distinguish the units to which each landed halibut can be uniquely associated (i.e. either a single haul or a set of two hauls).

The true total halibut catch is taken to be the sum of the census and the landings. The count and weight of total halibut in the catch is available at the fishing event level.

A number of analyses will be conducted to investigate the accuracy and precision of halibut estimates from the OS and EM protocols. Estimates from both the OS and the EM methods will be compared to the halibut census and the total halibut catch (census and landings). Analyses and statistical tests will be conducted at the haul level, fishing event level, trip level and study level, as appropriate.

1) Using sampling theory-based analytical methods estimate the number and weight of halibut (and the precision of estimates) by haul, fishing event, trip and study for:
    a. Observer sampling
    b. EM method (EMR1 and EMR2)

2) Conduct statistical tests to compare the halibut estimates from 1) above:
    a. Observer sampling with census and total catch (census and landings)
    b. EM with census and total catch

3) Use bootstrap methods to compare study-level estimates of:
    a. Observer sampling with total catch
    b. EM with census

4) For EM data
    a. Compare accuracy and precision of EMR1 with EMR2 estimates of halibut counts and average halibut length.
    b. Compare accuracy and precision of average lengths estimated using all measured halibut with average lengths estimated using only clear halibut images (ie. not recoded as "fish not flat on grid", etc.).
    c. Compare accuracy and precision of converting fish length using "top of grid", "bottom of grid", or average of both estimates.

5) Conduct a simulation using the study data to investigate the accuracy and precision of EM halibut biomass estimates obtained from enumerating all halibut but measuring only a subset of them (eg. every 3rd or 5th halibut).