SCIENTIFIC COMMITTEE
FIFTH REGULAR SESSION

10-21 August 2009
Port Vila, Vanuatu

Selectivity bias in grab samples and other factors affecting the analysis of species composition data collected by observers on purse seiners in the Western and Central Pacific Ocean

---

WCPFC-SC5-2009/ST-WP-03

Timothy Lawson[1]

---

[1] Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia.

SELECTIVITY BIAS IN GRAB SAMPLES AND OTHER FACTORS AFFECTING THE ANALYSIS OF SPECIES COMPOSITION DATA COLLECTED BY OBSERVERS ON PURSE SEINERS IN THE WESTERN AND CENTRAL PACIFIC OCEAN

Timothy Lawson
Oceanic Fisheries Programme
Secretariat of the Pacific Community
Noumea, New Caledonia

1. Introduction

Lawson (2008) examined biases in the species composition of the catch by purse seiners determined from grab samples collected by observers and port samplers during 1995–2007, and found considerable differences, particularly for schools associated with floating objects (Figure 1). For associated schools, the percentages of skipjack and yellowfin in the species composition determined from the observer data is 55.3% and 35.1% respectively, whereas for port sampling data, the percentages are 78.0% and 17.3%. That is, species compositions determined from observer data tend to have a suspiciously low proportion of skipjack and correspondingly high proportions of yellowfin and bigeye, while those determined from port sampling data tend to have a suspiciously high proportion of skipjack and correspondingly low proportions of yellowfin and bigeye.

Figure 1. Estimates of purse-seine species composition determined from observer data and port sampling data

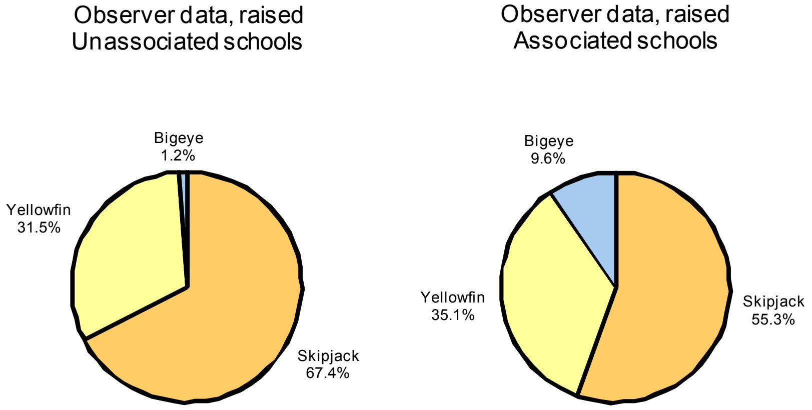
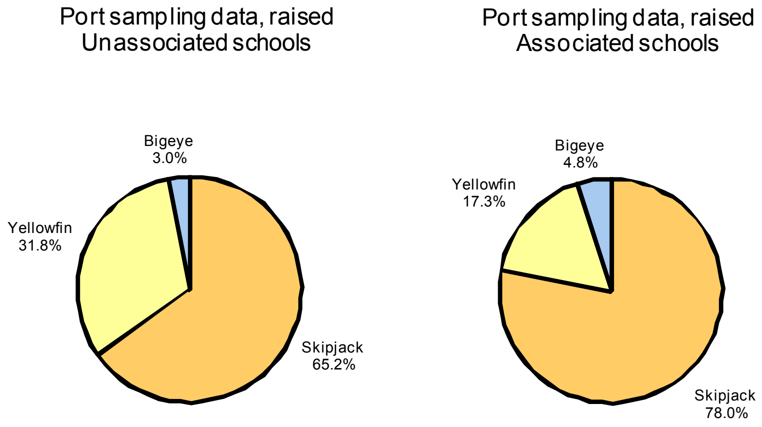


Figure 1 (continued)



Several sources of bias were examined, including:

- Bias related to set weight, wherein wells that are chosen for sampling in port because they contain a small number of large sets result in the over-estimation of the proportion of skipjack, since large associated schools contain a greater proportion of skipjack than small schools;
- Bias related to well mixing, particularly in transshipment ports, where fish are often sorted prior to unloading and which results in the selection for port sampling of an unreasonable number of wells that contain pure skipjack;
- Bias related to small sample sizes, wherein small samples (in terms of the number of fish selected per set or well) result in biased estimates of the species composition, even when the average weights determined from the grab sample data are unbiased; and
- Bias related to the non-random selection of fish, wherein observers and port samplers non-randomly select fish to sample, either due to physical constraints, such as layering in the set, brail or well, or to behaviour, such that samplers have a tendency to select certain species and/or sizes of fish more than others.

See Lawson (2008) for details. It is suspected that the main sources of bias in the observer data is size selection bias, while the main sources of bias in the port sampling data, particularly those collected during transshipment (in contrast to samples of categories of species and size recorded on cannery receipts), are set weight bias and bias related to well mixing. Species compositions determined from port samples may also be subject to bias resulting from the manner in which they are weighted (or 'raised') when determining the species composition of strata of time, area and school association (although this was not examined in Lawson, 2008).

Spill samples collected by observers were proposed as a sampling protocol that avoids each of the biases mentioned above. Since the samples are collected at sea during the brailing process, there are biases related to set weight or well mixing, as there are for port samples. And since in a spill sample, the fish are spilt into a bin, rather than grabbed by the observer, there are no grab sample or selection biases, as there are for grab samples collected by both observers and port samplers.

In 2008, paired spill and grab samples were collected by observers during four trips taken by purse seiners fishing anchored FADs in Papua New Guinea (Table 1). A total of 65 sets was sampled using both the grab sample and spill sample protocols. For grab samples, five fish were selected by the observer from each brail throughout the brailing process. For spill samples, fish from one or two brails were spilled into a bin; the sequential order of the brails selected for spilling was rotated from set to set to avoid layering related to the sizes of fish (Lawson 2008). For the spill samples, an average of 266 fish per set was sampled, whereas for grab samples, an average of 68 fish per set was sampled; thus, the spill samples were 3.1 times as large as grab samples.

Table 1. Number of sets per trip and the catch (tonnes) per trip sampled by observers conducting paired spill and grab samples

Trip	Vessel	Date of First Sample	Date of Last Sample	Sets Sampled	Catch Sampled	Spill Sample Observer	Grab Sample Observer
1	DOLORES 828	15-Mar-08	27-Mar-08	7	509	PSH	LPE
2	DOLORES 828	21-Jun-08	08-Aug-08	30	1,212	LKO	BWE
3	DOLORES 838	09-Jun-08	30-Jun-08	13	670	JTA	SUE
4	DOLORES 838	14-Jul-08	09-Aug-08	15	697	JTA	SUE

Analyses of the paired spill and grab samples are presented in sections 2 to 8 and focus on the estimation of the species composition per trip, rather than the species composition for strata of time period, geographic area and school association. This is due both to the nature of the analysis and to recent management measures adopted by the Western and Central Fisheries Commission that require monitoring of the species composition per trip. Estimates of selectivity bias are used to correct historical grab sample data collected by observers and the corrected data are applied to aggregated catch data covering the Western and Central Pacific Fisheries Commission (WCPFC) Statistical Area in sections 9 to 12. A model-based approach to estimating the species composition for strata used in the MULTIFAN-CL (MFCL) assessments with missing data is developed in section 13 and the sensitivity of estimates of species composition to length-weight parameters is briefly considered in section 14.

2. Species compositions determined from paired spill and grab samples

Figure 2 presents the species compositions determined from the paired spill and grab samples for each of the four trips and also for all trips combined. The species compositions were determined in terms of weight and raised by the set weight.

Figure 2. Estimates of purse-seine species composition determined from paired spill samples and grab samples collected during four trips in Papua New Guinea in 2008

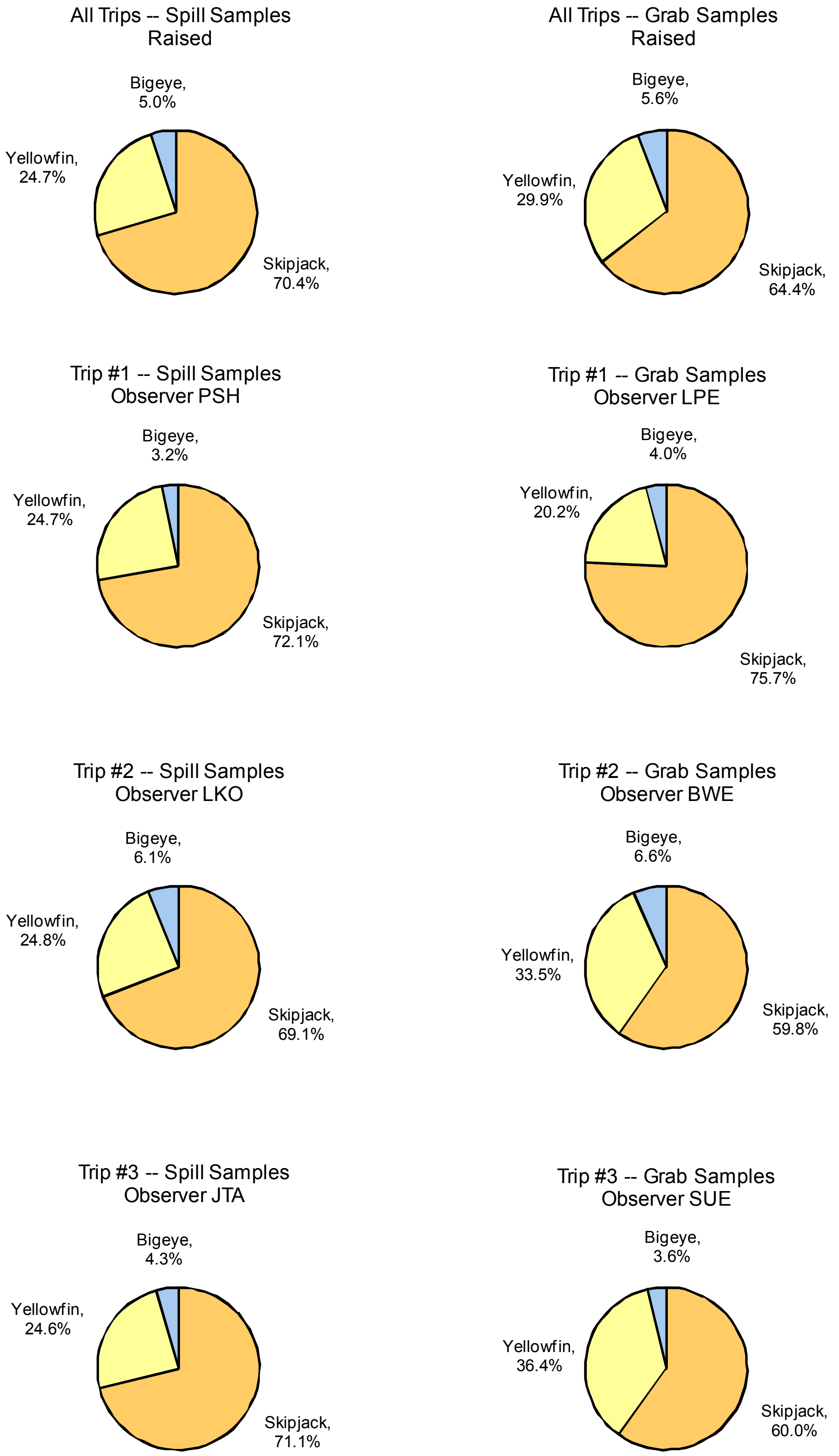
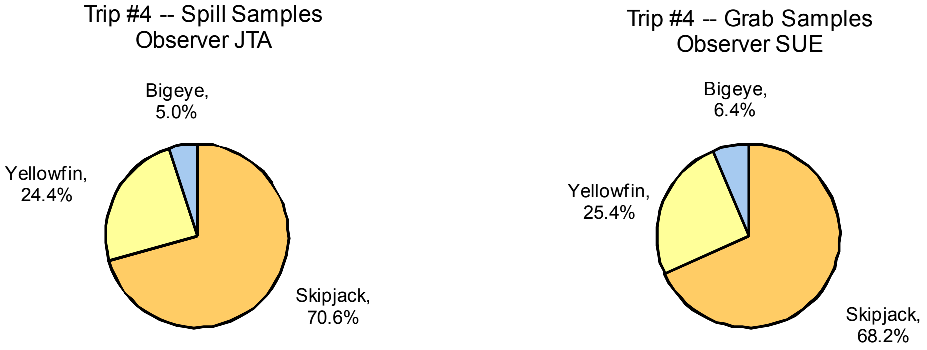


Figure 2 (continued)



For all trips combined, the species composition determined from the spill samples contains a greater amount of skipjack, 70.4%, and a lesser amount of yellowfin, 24.7%, than the species composition determined from the grab samples, 64.4% and 29.9% of skipjack and yellowfin respectively. The result for the grab samples is consistent with the top right-hand pie chart in Figure 1, which suggests that grab samples taken by observers taken from associated schools contain a suspiciously low amount of skipjack.

The extent of the differences in the species composition determined from spill and grab samples varies among trips. For trips #1 and #4, the spill and grab sample species compositions differ only slightly, whereas they are considerably different for trips #2 and #3. However, for each trip, the differences are consistent in direction, with less skipjack and more yellowfin in those determined from grab samples than from spill samples.

Trip #1 was taken in March 2008, while trips #2 to #4 were taken from June to August, with trip #2 overlapping trips #3 and #4. It is of interest to note that the species compositions determined from spill samples for trips #2 to #4 are almost identical, with the proportion of yellowfin varying by only 0.4%. This strongly suggests that the populations of fish that were associated with anchored FADs in Papua New Guinea during this relatively short time period were equally distributed in terms of species composition among the anchored FADs that were fished. If this supposition is correct and the populations fished during trips #2 to #4 were essentially the same, then we can interpret the species compositions for trips #2 to #4 as being replicates. That the variation in the species compositions determined from spill samples during these three trips was negligible suggests that they are reliable. That the proportions of skipjack and yellowfin in species composition determined from spill samples are more consistent with expectations than those determined from grab samples also suggests that they are, at the least, less biased, if not unbiased. In contrast, the species compositions determined from grab samples are less reliable and almost certainly biased.

Trips #3 and #4 had the same grab sampler. It is of interest to note that the species composition for trip #3 is biased, whereas that for trip #4 is almost unbiased. Roughly the same number of sets and the average number of fish per set were grab sampled during both trips, which may suggest that the bias was less related to the sample size and more to an inconsistency in size selection bias for the grab sampler.

3. Variance of estimates of the species composition per trip and per set

The species compositions determined from the spill samples for trips #2, #3 and #4 are almost identical; however, this does not necessarily imply that the species compositions determined for individual sets within a trip are uniform. Table 2 presents the average species composition per set for trips #2, #3 and #4 combined, and for each trip separately, and the standard deviations, minimums and maximums. The standard deviations for each species are quite high. The standard deviation, minimum and maximum depend to a certain extent on the number of fish in the spill sample. For example, there was one set during trip #2 for which there was only 16 fish in the spill sample; this is the sample that had 0% skipjack and 100% yellowfin. When Table 2 was calculated for spill samples that had at least 100, 200 and 300 fish, the standard deviation of the proportion of skipjack declined to 20.4% (56 sets), 18.7% (33 sets) and 16.0% (14 sets) respectively.

Table 2. Average species composition per set, standard deviation, minimum and maximum, determined from spill samples taken during trips #2, #3 and #4

Trip(s)	Number of Sets	Skipjack				Yellowfin				Bigeye			
		Avg	Std Dev	Min	Max	Avg	Std Dev	Min	Max	Avg	Std Dev	Min	Max
2, 3, 4	58	67.0%	22.1%	0.0%	98.5%	26.3%	19.0%	1.5%	100.0%	6.7%	9.9%	0.0%	47.5%
2	30	65.3%	21.6%	0.0%	94.8%	28.0%	20.3%	4.2%	100.0%	6.7%	10.4%	0.0%	47.5%
3	13	69.9%	20.9%	30.0%	95.5%	24.9%	18.3%	4.5%	60.3%	5.1%	5.6%	0.0%	21.2%
4	15	67.7%	23.7%	17.1%	98.5%	24.1%	16.4%	1.5%	54.8%	8.1%	11.6%	0.0%	38.4%

That these standard deviations did not decline more than they did with an increase in the sample size indicates that the schools fished during trips #2, #3 and #4 varied considerably in regard to their species compositions. While perhaps not surprising, this is an interesting result, given that the species compositions per trip for trips #2, #3 and #4 were almost identical. It suggests that even with considerable variation in the species composition among sets, estimates of the species composition per trip have a relatively low variance.

This idea can be tested by resampling the sets within a trip, such that one or more sets is deleted in each replicate. We would expect that the variance of the estimate of species composition per trip to be (a) higher when more sets are deleted per trip and (b) lower for trips with a greater number of sets. Figure 3 presents the species compositions for each trip determined from the spill samples, with 95% intervals determined from 1000 replicates of resampling of the sets within each trip, wherein one set (top histogram) and three sets (bottom) were randomly selected for deletion in each replicate. (These intervals should *not* be considered as *confidence* intervals, since all sets were sampled during each trip.) For each replicate, the species composition was calculated and then the 95% interval for each species was determined by ordering the 1000 values of the species composition for each species. The abscissas in Figure 3 shows the trip number and the species code for each 95% interval.

Regarding the top plot in Figure 3, the first point of interest is that, when one set is deleted, the intervals are, in all cases, narrow. This tends to confirm that the species composition per trip has low variance even when the species composition per set varies considerably. The second point is that, as expected, the intervals for trip #2, which had 30 sets, are narrower than the intervals for trips #3 and #4, which had 13 and 15 sets respectively. The intervals are widest for trip #1, which had 7 sets. Both of these points are further confirmed in the bottom plot. When three sets are deleted, the intervals are wider than when one set is deleted, but still narrow for those trips with a relatively large number of sets, i.e., trips #2, #3 and #4. For trip #1, which had only 7 sets, leaving only 4 sets after deleting three of them in each replicate, the intervals are much wider. The implications for observer programmes is that all sets during a trip should be sampled, except perhaps for long trips during which an unusually large number of sets are made.

Figure 3. Estimates of purse-seine species composition per trip determined from spill samples, with 95% intervals determined from resampling of the sets with deletions

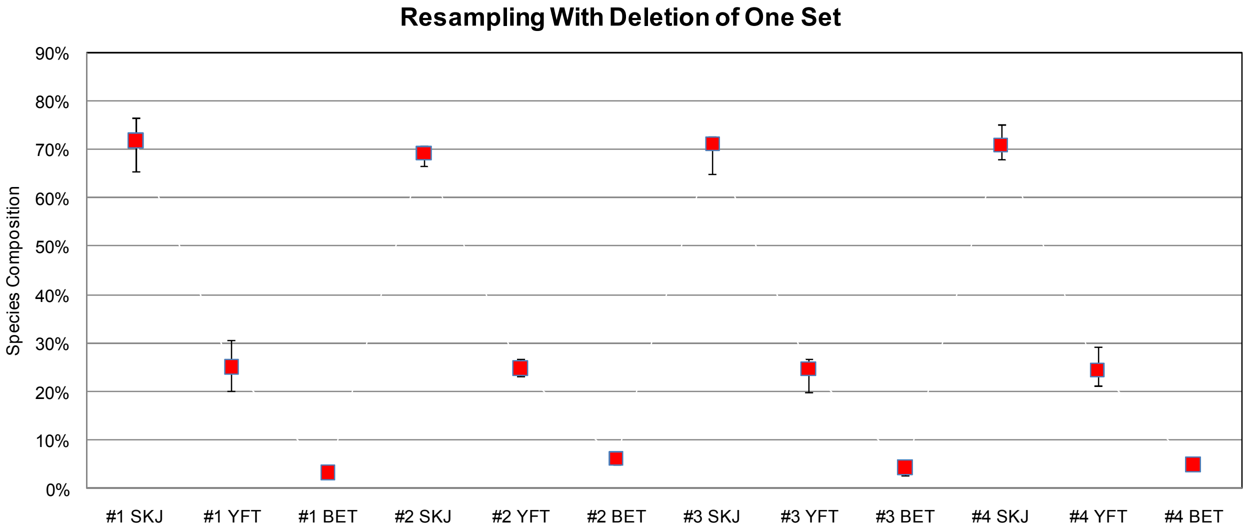
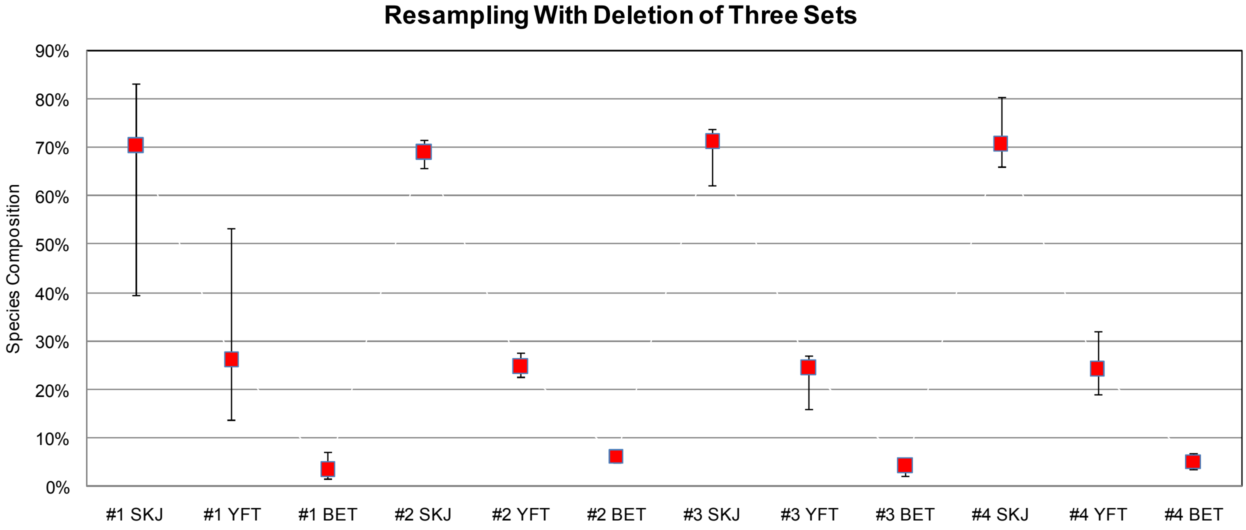


Figure 3 (continued)



For the estimates of the species composition per trip determined from spill samples, shown in Figure 2, all of the sets were sampled and thus none of the variance of the estimates is due to variation in the species composition among the sets.[1] Instead, the variance of the estimates of the species composition per trip is entirely due to variation within sets. Resampling of the fish from within a sample from a particular set, with replacement, will provide an estimate of the variance of the estimate of the species composition for that set. Simultaneously resampling the fish from samples of all sets within a trip will provide an estimate of the variance of the estimate of the species composition per trip.

First, we consider the variance of the estimate of the species composition per set. The spill samples from each of the 65 sets made during the four trips were resampled with replacement, in 1000 replicates. That is, for the spill sample taken from the second set in trip #2, for example, which consisted of 281 fish, 281 fish were selected randomly from the original 281 fish in each of the 1000 replicates; thus, in a given replicate, some of the original 281 fish may have been selected more than once and others not at all.[2] For each replicate, the estimate of the species composition of the set was calculated and then the 95% confidence interval for each species was determined by ordering the

---

[1] Assume for the moment that the standard error of the estimate of the species composition per trip is due entirely to variation among sets. Since the population of sets per trip is finite, the standard error must be multiplied by a finite population correction factor of the form $\sqrt{1-\frac{n}{N}}$, where n is the sample size and N is the population size (Cochran 1977). If all sets per trip are sampled, then the sample size will equal the population size and the finite population correction factor, and thus the corrected standard error, will be zero.

[2] Resampling with replacement of a spill sample of 281 fish that represents, say, 1% of the number of fish in the set is equivalent to simulating the sampling of a set that contains 28,100 fish. The only difference is that with resampling, a fish in the original spill sample could be selected a maximum of 281 times, whereas in a simulation, a fish could be selected a maximum of only 100 times. In either case, the odds of actually selecting a fish more than a couple of times are infinitesimal.

1000 values of the species composition for each species. Figure 4 shows the relationship between the variance and the number of fish in the spill sample. The variance for each species is expressed as the ratio of one-quarter of the width of the confidence interval to the mean value of the species composition; this value roughly corresponds to a coefficient of variation (i.e., the ratio of the standard error to the mean).

The "coefficients of variation" for skipjack are relatively low (most less than 10%), while those for yellowfin are moderate (most less than 30%) and those for bigeye are high (most greater than 30%). There is a clear relationship between the "coefficient of variation" and the sample size for skipjack and yellowfin, which suggests that spill sample sizes of about 300 to 400 fish are a reasonable compromise between sampling effort and reliability. There are so few bigeye in most of the spill samples examined in this study that much larger sample sizes would have been required in order to reduce the "coefficients of variation".

Figure 4. Relationship between "coefficients of variation" of estimates of the species composition per set and spill sample size

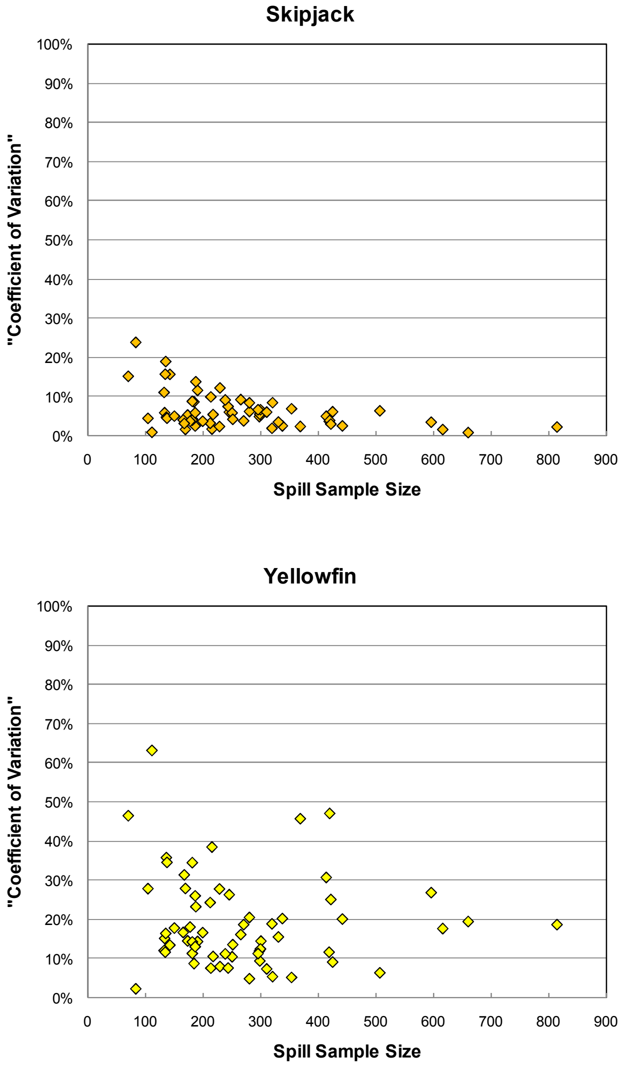
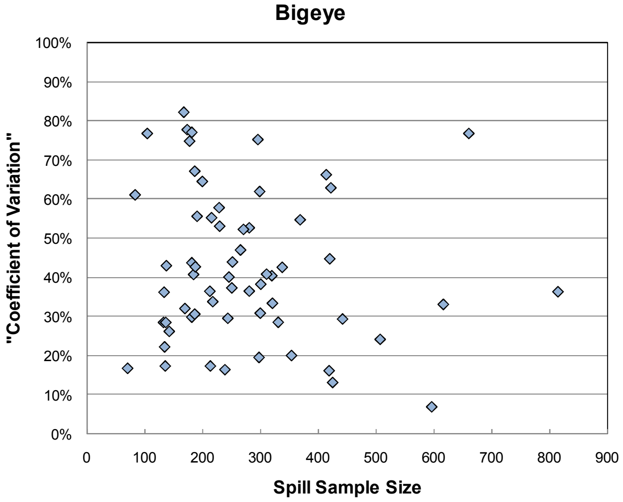
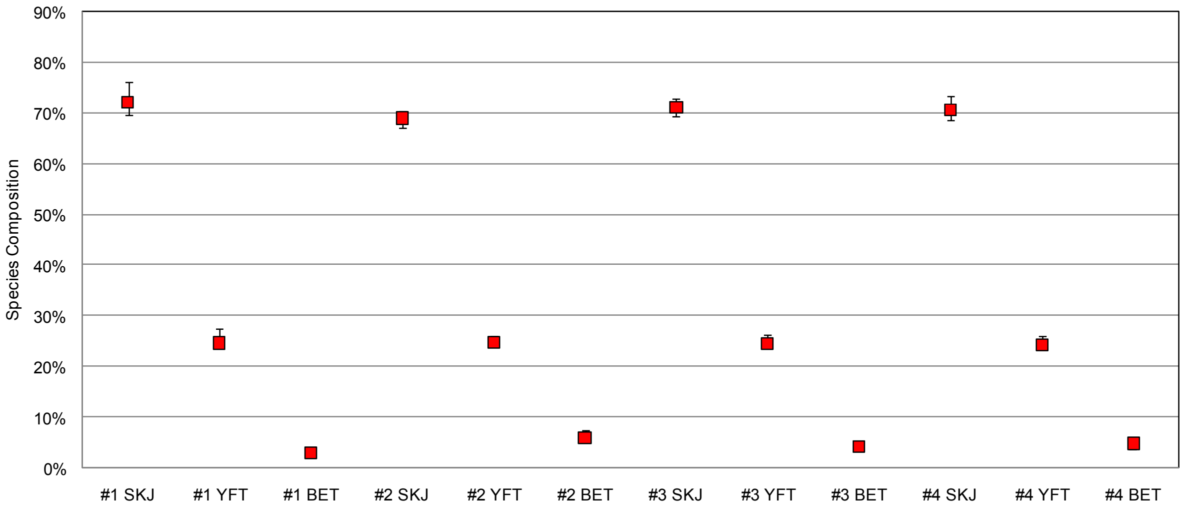


Figure 4 (continued)



Second, we consider the variance of the estimate of the species composition per trip. For each trip, the spill samples from all sets were resampled with replacement, in 1000 replicates. For each replicate, the estimate of the species composition of the trip was calculated, with the samples for each set raised by the set weight, and then the 95% confidence interval for each species was determined by ordering the 1000 values of the species composition for each species. As can be seen in Figure 5, almost all confidence intervals are negligible, even for estimates of the proportion of bigeye. The average "coefficients of variation" per trip for skipjack, yellowfin and bigeye are 1.6%, 2.2% and 10.0% respectively. Thus, while the variance of the estimates of the species composition per set can sometimes be moderate or high, particularly for yellowfin and bigeye, the variance of the species composition per trip determined from spill samples will usually be low.

Figure 5. Estimates of purse-seine species composition per trip determined from spill samples, with 95% confidence intervals determined from resampling of the fish within each sample



4. Size selection bias in grab samples

Figure 6 presents length frequencies for the fish collected in spill samples and grab samples; the length frequencies are for skipjack, yellowfin and bigeye combined, in all samples combined. The top histogram shows the length frequencies in terms of the number of fish sampled, while the bottom length frequency is in terms of the weight of the fish sampled. Lengths (cm) were converted to weights (kg) using the length-weight parameters below:

Species	a	b
Skipjack	0.8639E-05	3.2174
Yellowfin	2.5120E-05	2.9396
Bigeye	1.9729E-05	3.0247

Both histograms suggest that the main difference between spill samples and grab samples is the under-representation of small fish in grab samples. Since small fish consist of more skipjack than yellowfin, the species compositions determined from grab samples under-estimate the proportion of skipjack and over-estimate the proportion of yellowfin.

Figure 6. Length frequencies for paired spill samples and grab samples in terms of number of fish (top) and weight of fish (bottom), for all species combined

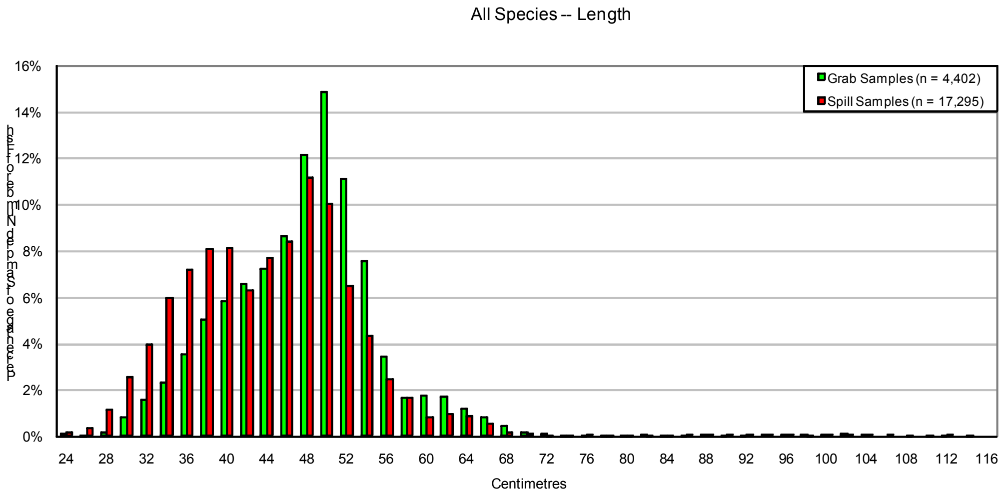


Figure 6 (continued)

All Species -- Weight

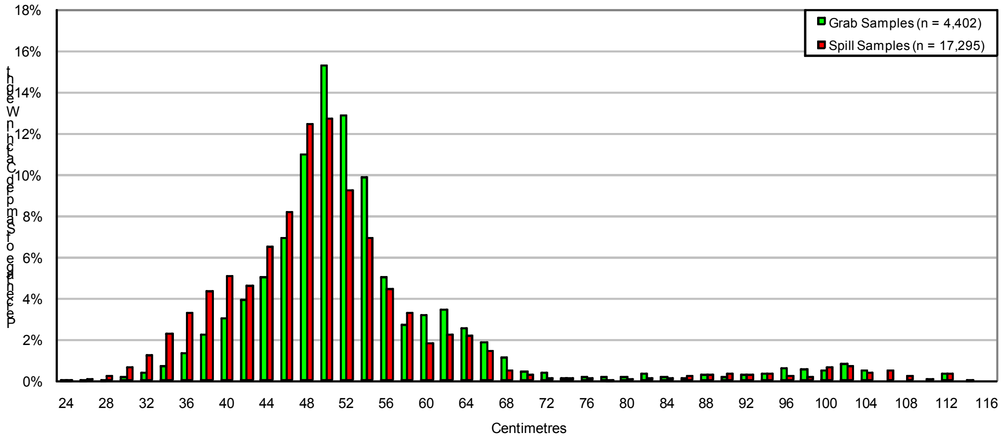


Figure 7 presents length frequencies in terms of number of fish for skipjack, yellowfin and bigeye separately. For skipjack and bigeye, the under-representation of small fish in grab samples is clear, while for yellowfin, it is somewhat less evident since there is a smaller proportion of small yellowfin than large yellowfin in the catch.

For yellowfin and bigeye greater than 80 centimetres, the differences between spill samples and grab samples appear minor (although see below). For yellowfin, the proportions of the number of fish greater than 80 centimetres is 3.0% and 2.9% for spill samples and grab samples respectively, while the proportions of the weight of fish are 17.9% and 15.1%. For bigeye, the proportions of the number of fish greater than 80 centimetres is 1.3% and 2.9% for spill samples and grab samples respectively, while the proportions of the weight of fish are 9.5% and 10.4%.

Figure 7. Length frequencies for paired spill samples and grab samples in terms of number of fish, by species

Skipjack

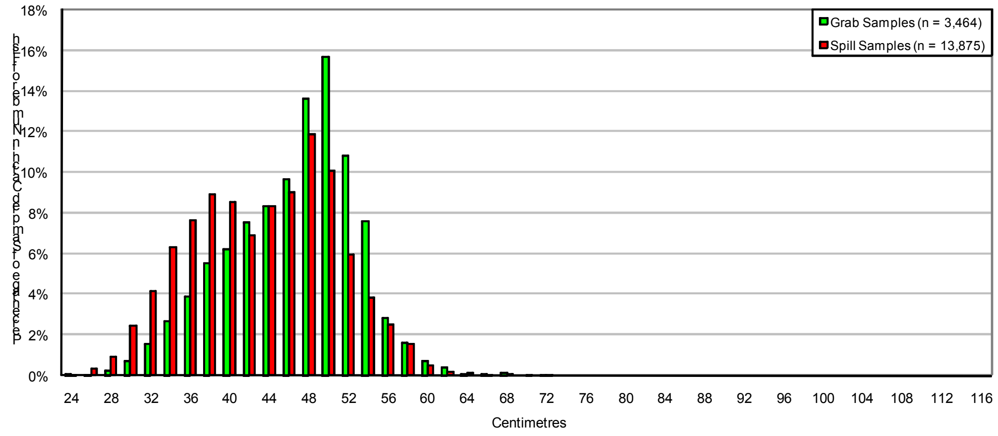
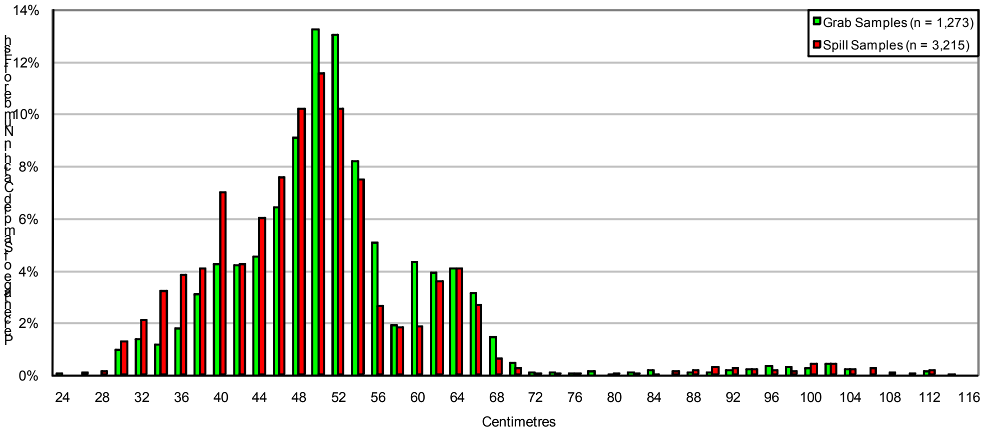
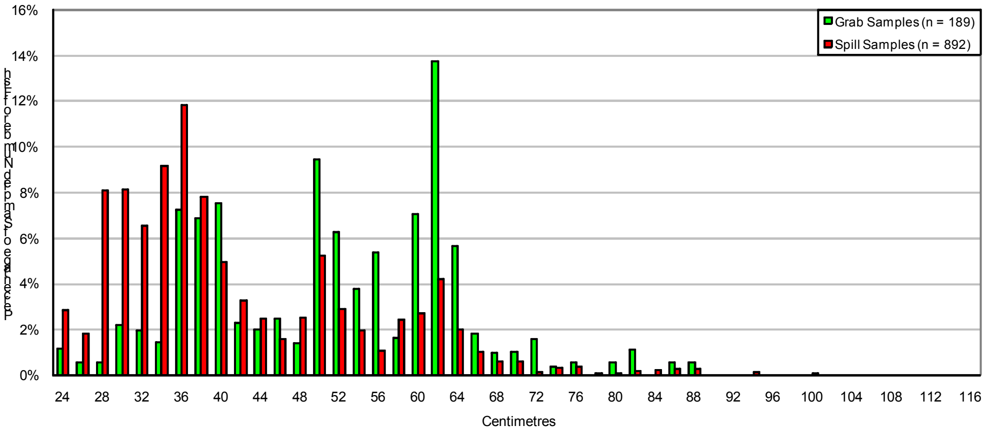


Figure 7 (continued)

Yellowfin

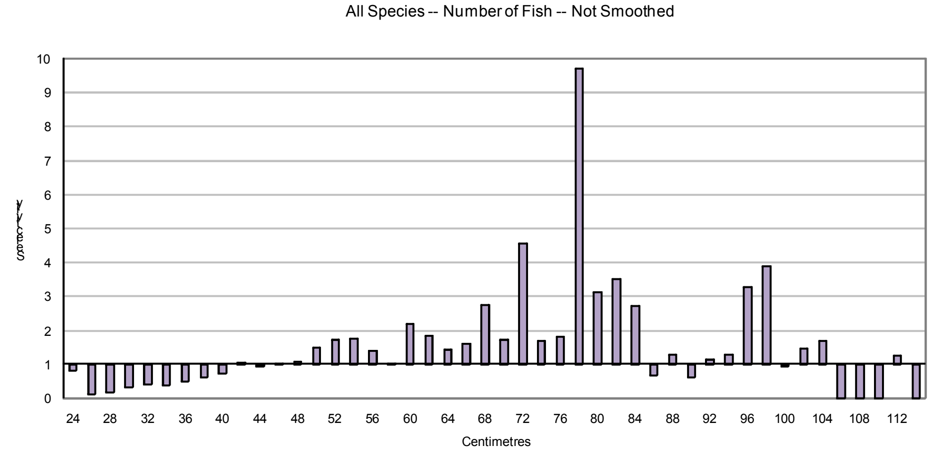


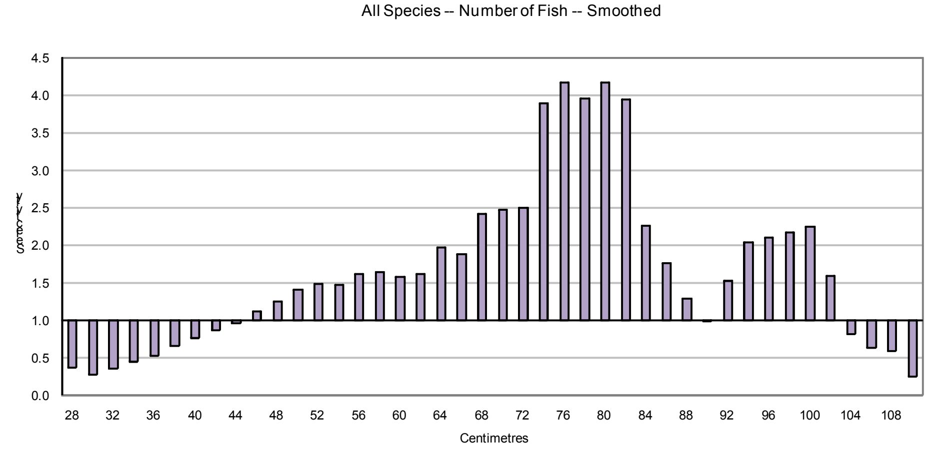
Bigeye



The size selection bias was examined in greater detail by plotting the ratio of the proportion at each length interval determined from the grab samples to that determined from the spill samples. For a particular length, if the proportion in the length frequency determined from the grab samples is greater than the proportion determined from the spill samples, then the ratio will be greater than one; if smaller, then the ratio will be less than one. Figure 8 shows the ratios of the proportions in terms of numbers of fish; both unsmoothed histograms and histograms smoothed with a moving average of five 2 centimetre intervals are presented. Assuming that the length frequencies determined from the spill samples are unbiased, then fish less than about 48 cm and greater than 106 cm appear to be under-selected by grab samplers, while fish between 50 cm and 104 cm appear to over-selected, except for fish around 90 cm, which appear to be neither under- nor over-selected. The pattern of selection bias is clearer in the smoothed histogram.

Figure 8. Grab sample selection bias in terms of number of fish





The selection histograms can also be constructed in terms of the weight of fish (Figure 9). The pattern of selection is the same; however, the magnitude of the over-selection is less in terms of weight than in terms of number of fish.

Figure 9. Grab sample selection bias in terms of weight

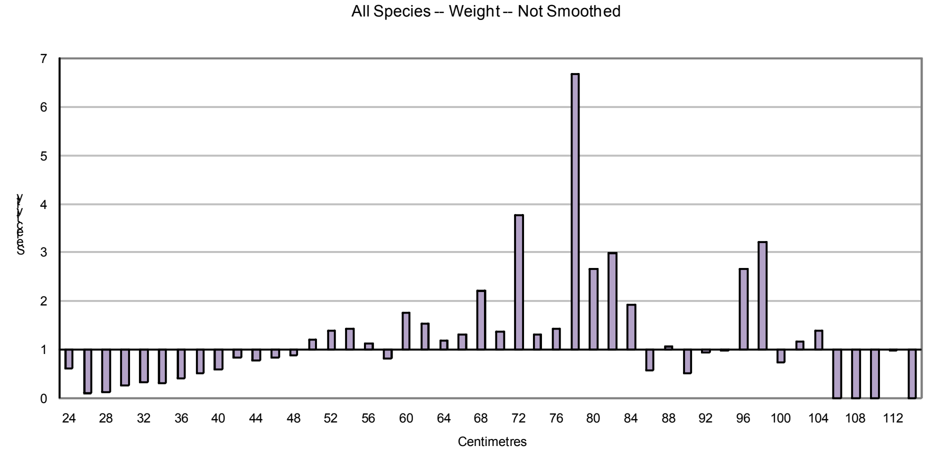
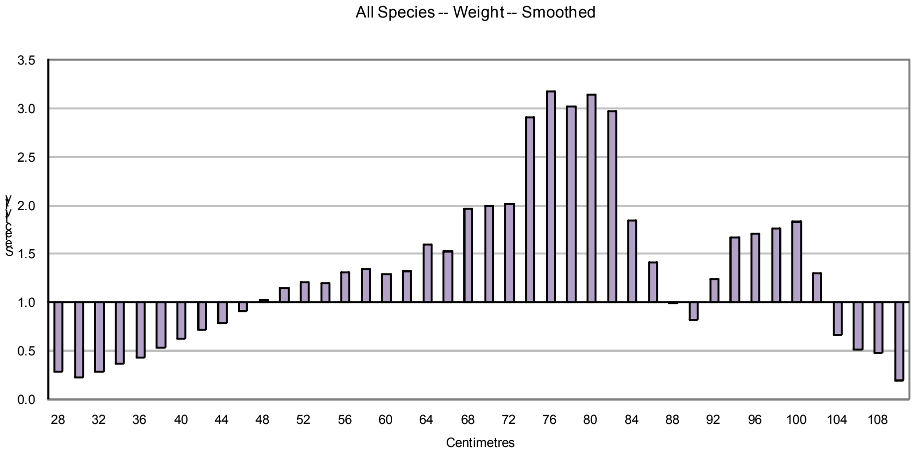


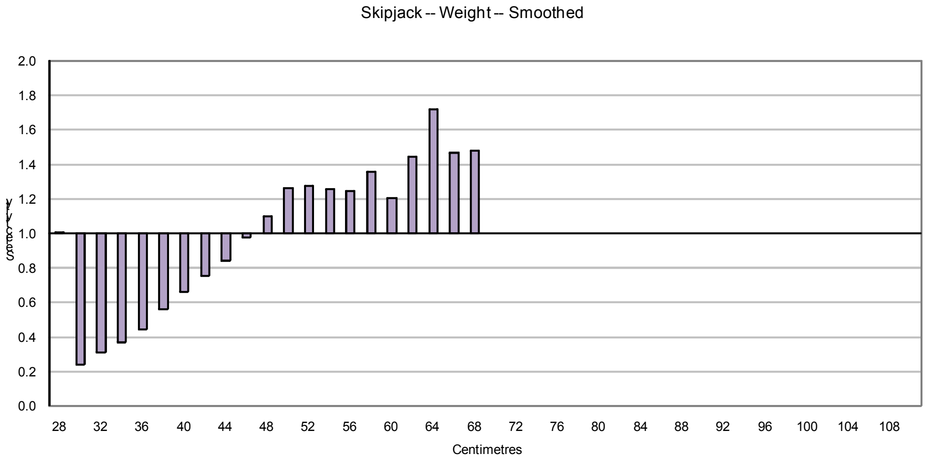
Figure 9 (continued)

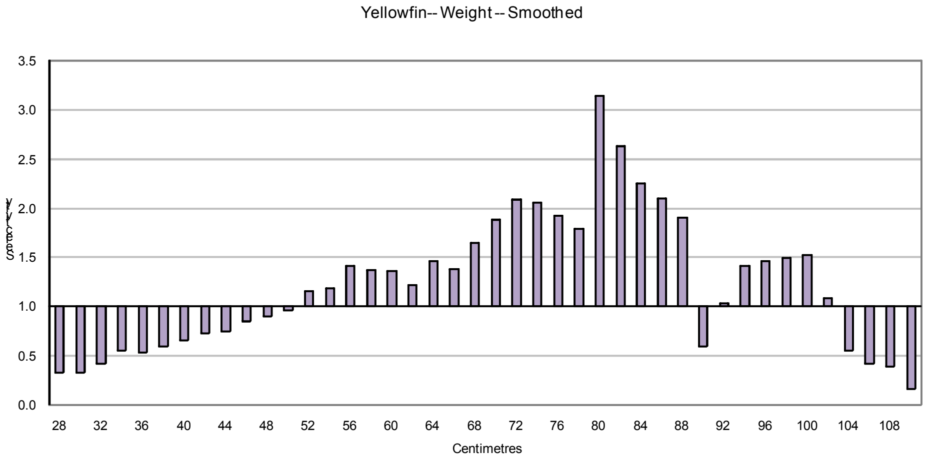

All Species -- Weight -- Smoothed

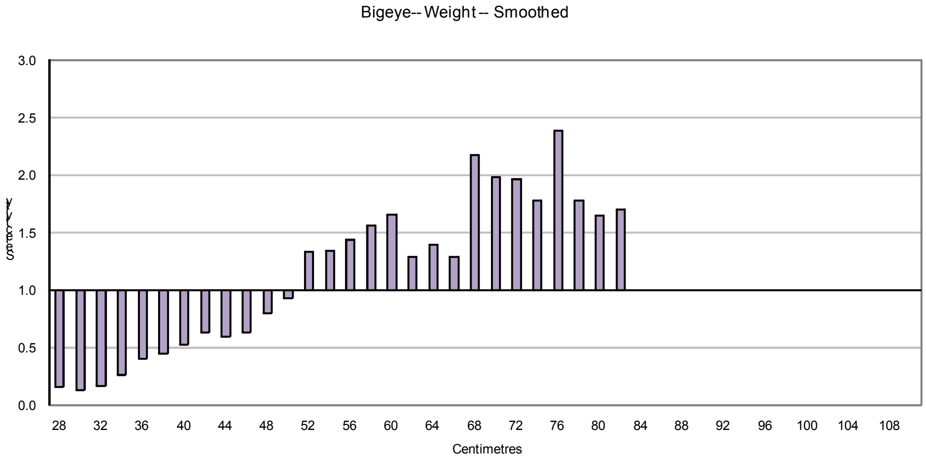
This pattern of selection is somewhat puzzling. It might be expected that grab samplers, for one reason or another, tend to miss very small and very large fish, and this would explain the under-selection of those fish and the over-selection of the fish of lengths in between. However, it is not clear why the over-selection appears as two modes, with negligible selection bias in the middle, at around 90 cm, rather than as a single mode. It is not related to the values of the length frequencies at around 90 cm; the length frequencies in terms of both numbers of fish and weight drop off at about 70 cm and then remain at more or less the same low magnitude for all sizes above 70 cm, other than a small mode at around 102 cm (Figure 6). Instead, it may be an artefact related to the sample size. The number of fish sampled by the grab samplers, and used to construct the histograms, was large, 4,402 fish, but the number of fish greater than 70 cm is small relative to the number below 70 cm and so the selection bias shown in Figures 8 and 9 may be less accurate for lengths greater than 70 cm than for those below 70 cm.

Figure 10 presents the grab sample selection bias separately for skipjack, yellowfin and bigeye. The pattern of bias for each species is consistent with the pattern for all species combined; however, the magnitude of the negative bias for small skipjack is much greater than for small yellowfin and bigeye, and the magnitude of the positive bias for mid-sized fish is much greater for yellowfin and bigeye than for skipjack. This would suggest that a method of correcting the grab samples for size selection bias based on these empirical results should take both species and length into account, rather than length alone.

Figure 10. Grab sample selection bias in terms of weight, by species







5. An empirical approach to correcting species compositions determined from grab samples for size selection bias

The simplest method of correcting the species compositions is to assume that the bias determined from the paired spill and grab samples discussed above is applicable to grab samples in general.

Assuming the species compositions determined from spill samples are unbiased, the bias in the species composition determined from grab samples is –5.94% for skipjack, +5.27% for yellowfin, and +0.66% for bigeye (see Figure 2, top row).

However, the applicability of these biases to grab samples in general may be limited. The biases were determined from paired samples collected under specific circumstances, i.e., from schools associated with anchored FADs in the waters of Papua New Guinea during March–August 2008. The species composition of the tuna that were fished appears to have been remarkably uniform. The paired samples were collected during only four trips and by only three grab samplers. Obviously, many more paired samples will need to be collected to correct the species compositions determined from historical grab samples in a rigorous manner. In particular, paired samples from unassociated schools and schools associated with logs and drifting FADs will be required. For each school association, paired samples should be collected in different parts of the region, since species composition varies by geographic area (Lawson 2008), and preferably from several grab samplers. The extent to which the bias depends on the true species composition, which may be affected by school association, school size and geographic area, needs to be determined. It should be noted that the limitations in the applicability of the currently available paired samples is common to all of the methods of correcting species compositions proposed below.

Another empirical approach to correcting the species compositions is to apply correction factors that are specific to a particular species and length interval. If w_{ij} is the sum of the weight of fish of species i and length interval j in a grab sample, then the corrected proportion of species i, P_i', in a grab sample is given by

$$P_i' = \frac{\sum_j w_{ij} \cdot f_{ij}}{\sum_i \sum_j w_{ij} \cdot f_{ij}} \tag{1}$$

$$f_{ij} = \frac{P_{ij}^S}{P_{ij}^G} \tag{2}$$

where f_{ij} is the correction factor for fish of species i and length interval j, and P_{ij}^S and P_{ij}^G are the uncorrected proportions of fish of species i and length interval j in spill samples and grab samples respectively.

Table 3 presents the uncorrected species compositions determined from the spill and grab samples, together with species compositions determined from all grab samples combined that were corrected on the basis of equations (1) and (2). However, in addition to using correction factors for one-centimetre length intervals, various groupings of length intervals were also considered, including intervals of 2 cm, 5 cm, 10 cm and 20 cm, and three intervals of lengths ≤ 47 cm, between 48 cm and 103 cm, and ≥ 104 cm. The latter grouping of length corresponds to the very small and very large lengths that are under-selected by grab samplers and the lengths in between that are over-selected (Figures 7 and 8).

The uncorrected species compositions in Table 3 are those shown in the top row of Figure 2; the species composition determined from the spill samples is assumed to be unbiased, while the species composition determined from the grab samples is subject to size selection bias. If the method of correcting for the size selection bias given by equations (1) and (2) is adequate, then the corrected species composition should be similar to that determined from the spill samples. As expected, this is indeed the case. The corrections are slightly less accurate using lengths grouped by 1 cm and 2 cm intervals, because of the effect of missing values for certain lengths in the spill samples. The corrections using lengths grouped by 5 cm, 10 cm and 20 cm are very close and the correction using the three length intervals corresponding to very small, mid-sized and very large fish is perfect.

Table 3. Uncorrected species compositions determined from spill samples and grab samples, and the species composition determined from grab samples corrected with empirical factors by species and various groupings of length intervals

Method	Skipjack	Yellowfin	Bigeye
Spill samples, uncorrected	70.36%	24.67%	4.97%
Grab samples, uncorrected	64.42%	29.94%	5.64%
Grab samples, corrected, 1 cm	72.23%	23.49%	4.28%
Grab samples, corrected, 2 cm	71.29%	23.90%	4.81%
Grab samples, corrected, 5 cm	70.55%	24.58%	4.87%
Grab samples, corrected, 10 cm	70.55%	24.59%	4.87%
Grab samples, corrected, 20 cm	70.40%	24.68%	4.92%
Grab samples, corrected, three intervals	70.36%	24.67%	4.97%

Table 4 presents the species compositions determined from correction factors for length only, rather than species and length. Also as expected, while the corrected species compositions in Table 4 are less biased than the species composition determined from uncorrected grab samples, they are less accurate than those corrected on the basis of both species and length (Table 3).

Table 4. Species compositions determined from grab samples corrected with empirical factors for various groupings of length intervals only

Method	Skipjack	Yellowfin	Bigeye
Grab samples, corrected, 1 cm	68.08%	25.34%	4.79%
Grab samples, corrected, 2 cm	68.11%	25.98%	4.93%
Grab samples, corrected, 5 cm	68.28%	26.75%	4.83%
Grab samples, corrected, 10 cm	68.38%	26.70%	4.78%
Grab samples, corrected, 20 cm	67.65%	27.45%	4.89%
Grab samples, corrected, three intervals	66.91%	27.95%	5.14%

6. A model-based approach to correcting species compositions determined from grab samples for size selection bias

A simple model of the grab sampling process can be derived as follows. Let the number of fish of species *i* and length interval *j* in set *k* be N_{ijk}. The probability of a grab sampler selecting a fish of species *i* and length interval *j* depends on (a) the sampling protocol, notably the number of fish grabbed per brail, (b) physical factors, such as the possible layering by species and/or size of fish in the set and/or in the brail, and (c) on behavioural factors, such as the tendency of samplers to non-randomly select fish of certain species and/or sizes, rather than others. Let this probability, which can be thought of as the *availability* of a fish to be sampled, be A_{ijk}. The term *availability* is borrowed from the literature on gear selectivity — e.g., Millar & Fryer (1999) — to which grab sample selectivity is analogous. The number of fish selected by a grab sampler, n_{ijk}, is then given by

$$n_{ijk} = N_{ijk} \cdot A_{ijk} + \varepsilon \tag{3}$$

where ε is a random variable of mean zero, such that the expected value of n_{ijk} is $N_{ijk} \cdot A_{ijk}$.

The N_{ijk} can be written in terms of the total weight of the set, W_k, and the average weight of fish of species *i* and length interval *j* in the set, $\overline{w}_{ijk}$, and the true proportion of fish of species *i* and length interval *j*, in terms of weight, T_{ijk}:

$$N_{ijk} = \frac{W_k \cdot T_{ijk}}{\overline{w}_{ijk}}. \tag{4}$$

Substituting equation (4) into equation (3), we have

$$n_{ijk} = \frac{W_k \cdot T_{ijk}}{\overline{w}_{ijk}} \cdot A_{ijk} + \varepsilon \,. \tag{5}$$

The total weight of the set, W_k, is known and the average weights, $\overline{w}_{ijk}$, can be estimated from the samples and treated as known, whereas the true proportions of fish in the set, T_{ijk}, are unknown. The n_{ijk} can thus be written in terms of the ratio of two constants, which are known or treated as known, and two proportions, which are unknown, i.e., the true proportions and the availabilities.

Any log-likelihood function based on equation (5) would have many unknown parameters. In the spill and grab samples examined above, there are 93 one-centimetre length intervals (from 24 cm to 116 cm). With three species and two unknown proportions for each species and length interval, there are a maximum of 93 * 3 * 2 = 558 parameters for each set. However, the number of parameters can be reduced considerably.

Firstly, if we assume that the species composition determined from spill samples is unbiased, then the true proportions of fish of species *i* and length interval *j* in set *k*, in terms of weight, T_{ijk}, can be estimated from the spill sample and treated as known. If the T_{ijk} for one-centimetre length intervals in a set are not considered to be well estimated, perhaps because data for certain lengths are missing, then the model can be formulated in terms of wider length intervals. The length intervals could, for example, be 5 cm, 10 cm or 20 cm intervals, or three intervals of very small, mid-sized and very large fish, as considered in section 5. If wider length intervals are used, the average weights, $\overline{w}_{ijk}$, in the model must be determined accordingly.

Secondly, rather than estimating the availability of a fish to be sampled for each species *i*, length interval *j* and set *k*, availability can be assumed constant over all sets:

$$n_{ijk} = \frac{W_k \cdot T_{ijk}}{\overline{w}_{ijk}} \cdot A_{ij} + \varepsilon \,. \tag{6}$$

where the only difference between equations (5) and (6) is that availability no longer depends on set in equation (6).

Thirdly, availability can be modelled with an appropriate distribution. If the A_{ij} (now assumed constant over sets) were thought to be normally distributed in terms of length, for example, then there would be two parameters (i.e., the mean μ and the standard deviation σ) for three normal distributions (i.e., one for each species) to give a total of six unknown parameters in the model.

The appropriateness of a given distribution for modelling availability will depend on the length intervals used in the model. If wide intervals are used, such as 10 cm or 20 cm intervals or three intervals of very small, mid-sized and very large fish, then the A_{ij} can be modelled as a step function, where each step represents the value of the A_{ij} for species *i* and length interval *j*. If, for example, three intervals are used, then there would be a maximum of nine unknown parameters in the model, i.e., three steps for each of the three species.

The average weights of species *i* and length interval *j* in equation (6) can be determined either for each set or for all sets combined, as given in equations (7) and (8) respectively:

$$\overline{w}_{ijk} = \frac{\sum_{l} a_i \cdot L_{ijkl}^{b_i}}{n_{ijk}^{s}} \tag{7}$$

$$\overline{w}_{ij} = \frac{\sum_{k} \sum_{l} a_i \cdot L_{ijkl}^{b_i}}{\sum_{k} n_{ijk}^{s}} \tag{8}$$

where L_{ijkl} is the length of fish *l* in the category of species *i* and length interval *j* in the spill sample taken from set *k*; a_i and b_i are the weight-length parameters for species i (see section 4); and n_{ijk}^{s} is the number of fish of species *i* and length interval *j* in the spill sample taken from set *k* .

The estimates of availability can be used to correct the species composition for an individual set determined from the grab samples as follows:

$$\widehat{P}_{ik} = \frac{\sum_{j} \frac{W_{ijk}}{A_{ij}}}{\sum_{i} \sum_{j} \frac{W_{ijk}}{A_{ij}}} \tag{9}$$

$$W_{ijk} = W_k \cdot \frac{\sum_{l} a_i \cdot L_{ijkl}^{b_i}}{\sum_{i} \sum_{j} \sum_{l} a_i \cdot L_{ijkl}^{b_i}} \tag{10}$$

where $\widehat{P}_{ik}$ is the estimated proportion of species *i* in set *k* and W_{ijk} is the weight of fish of species *i* and length interval *j*, raised by the set weight W_k , that were selected by the grab sampler from set *k*.

The estimates of availability can also be used to correct the species composition for a group of sets, such as for a trip or a time-area stratum, as follows:

$$\widehat{P}_{i} = \frac{\sum_{j} \sum_{k} \frac{W_{ijk}}{A_{ij}}}{\sum_{i} \sum_{j} \sum_{k} \frac{W_{ijk}}{A_{ij}}} \tag{11}$$

where $\widehat{P}_{i}$ is the estimated proportion of species *i* in the group of sets.

If one is willing to assume that availability depends on length interval only, rather than on both species and length interval, the model can be formulated as follows:

$$n_{jk} = \frac{W_k \cdot T_{jk}}{\overline{w}_{jk}} \cdot A_j + \varepsilon \tag{12}$$

$$\widehat{P}_i = \frac{\sum_j \sum_k \frac{W_{ijk}}{A_j}}{\sum_i \sum_j \sum_k \frac{W_{ijk}}{A_j}} . \tag{13}$$

7. Estimation of availability

The model was first applied using the equation (12), which assumes that availability depends only on the length interval. Various definitions of length intervals were used, including (i) three length intervals of very small, mid-sized and very large fish, with break points at 48 cm and 104 cm; (ii) four intervals similar to (i) but with an additional break point at 90 cm; (iii) groupings of 20 cm; (iv) groupings of 15 cm and (v) groupings of 10 cm. General linear models with various families of distribution and transformations of the n_{jk} were explored. The best fit was obtained using the Normal distribution and untransformed n_{jk}. The estimates of the corrected species compositions for all grab samples combined, determined from equation (13), are presented in Table 5. The first two rows of Table 5 show the uncorrected species compositions determined from spill samples and grab samples. Each subsequent row shows the species compositions determined from the A_j estimated using models with increasingly narrower length intervals. Table 5 also gives the bias of the estimated species compositions as a percentage of the uncorrected species composition determined from spill samples; these values thus represent the percentage bias in an estimate of a catch of species *i*, if it were determined from the corrected species composition.

Table 5. Uncorrected species compositions determined from spill samples and grab samples, and the species composition determined from grab samples corrected with estimates of availability that depend on length interval only

Method	Number of Strata		Number of Parameters		Species Composition			Bias		
	Total	Zero Grab	Total	Zero Grab	Skipjack	Yellowfin	Bigeye	Skipjack	Yellowfin	Bigeye
Spill samples, uncorrected					70.36%	24.67%	4.97%			
Grab samples, uncorrected					64.42%	29.94%	5.64%	-8.4%	+21.4%	+13.4%
Grab samples, corrected, three intervals	145	12	3	0	68.26%	26.74%	5.00%	-3.0%	+8.4%	+0.5%
Grab samples, corrected, four intervals	159	19	4	0	67.92%	27.10%	4.99%	-3.5%	+9.8%	+0.3%
Grab samples, corrected, 20 cm	227	39	8	2	68.47%	26.66%	4.88%	-2.7%	+8.1%	-2.0%
Grab samples, corrected, 15 cm	269	71	10	2	68.76%	26.48%	4.77%	-2.3%	+7.3%	-4.2%
Grab samples, corrected, 10 cm	352	90	14	3	66.56%	28.80%	4.63%	-5.4%	+16.8%	-6.8%

The second and third columns of Table 5 give the total number of strata of set and length interval for which the n_{jk} were fitted and the number of strata for which the n_{jk} were zero, respectively. The fourth and fifth columns give the number of A_j that were estimated and the number of A_j for which all n_{jk} were zero, respectively. The number of strata and parameters, and the number of strata and parameters for which the n_{jk} were zero, increases as the length intervals narrow. If all n_{jk} for a particular A_j are zero, then the estimate of the A_j is zero and that length interval is effectively excluded from the model and the estimation of the corrected species composition. There is thus a trade-off between the number of length intervals in the model and the number of length intervals excluded from the model because of lack of data. This trade-off should become less important in future analyses, as data from more paired spill samples and grab samples become available.

The n_{jk} were stratified by set and length interval; however, the number of strata of set and length interval for each model is less than the potential number of strata because of the lack of data for certain length intervals in some sets. For example, for the model with three length intervals of very small, mid-sized and very large fish, there were 145 n_{jk}, much less than the maximum of 65 sets * 3 length intervals = 195 strata. For length intervals grouped by 20 cm, 15 cm and 10 cm, there were 2, 2, and 3 A_j respectively for which all n_{jk} were zero and hence those numbers of length intervals were excluded from the models.

There were no very large skipjack nor very large bigeye in the spill and grab samples, so there were only two A_{ij} for skipjack and bigeye; with the three A_{ij} for yellowfin, there were a total of seven unknown A_{ij}. The number of n_{ijk} was still less than 65 sets * 7 categories = 455 strata because not all sets contained all seven categories of species and length interval. For example, only 17 sets contained very large yellowfin.

The species composition for the model with three length intervals (third row in Table 5) is relatively close in absolute value to the uncorrected species composition determined from spill samples. However, the bias as a proportion of the species composition determined from spill samples is quite large for yellowfin, +8.4%. A slight improvement in the bias for yellowfin is found for length intervals grouped by 15 cm, but at the expense of a larger bias in bigeye. The biases for the model with four length intervals is similar to those for the model with three intervals, but the model with three intervals should be considered a better model since these is one fewer parameter.

The estimates of availability for the model with three length intervals, with bars of plus or minus two standard errors, are shown in Figure 11. The estimates in Figure 11 are consistent with the grab sample size selection bias shown in Figure 10, with the availability of very small fish and very large fish less than that of mid-sized fish. The availability of each length interval is less than 0.5% of the number of fish in the set, and the average of the non-zero estimates is 0.36%. The standard errors of the estimates of availability for very small and mid-sized fish are negligible, while the standard error for very large fish is large, 1.14%; presumably, the standard errors of the estimates for very large fish will decline as more data from paired samples become available and are included in the analysis. The model explained 96.5% of the deviance. Residuals are plotted against the fitted values of the n_{jk} in Figure 12.

Figure 11. Estimates of grab sample availability for categories of species and length interval for the model with three length intervals

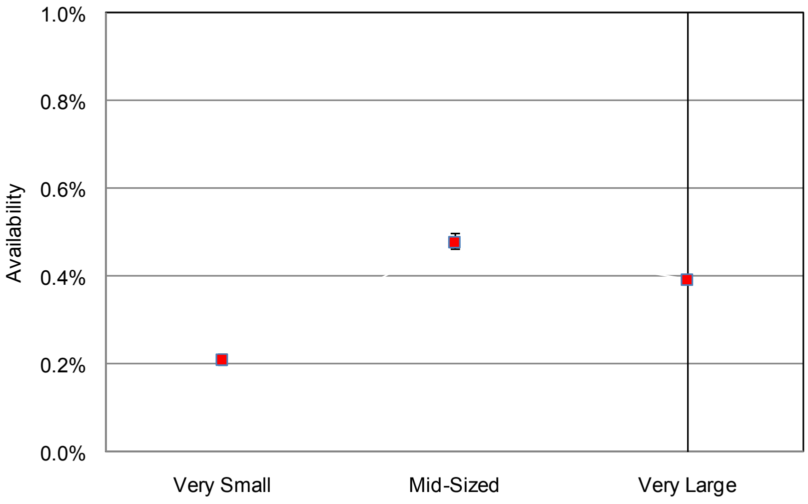
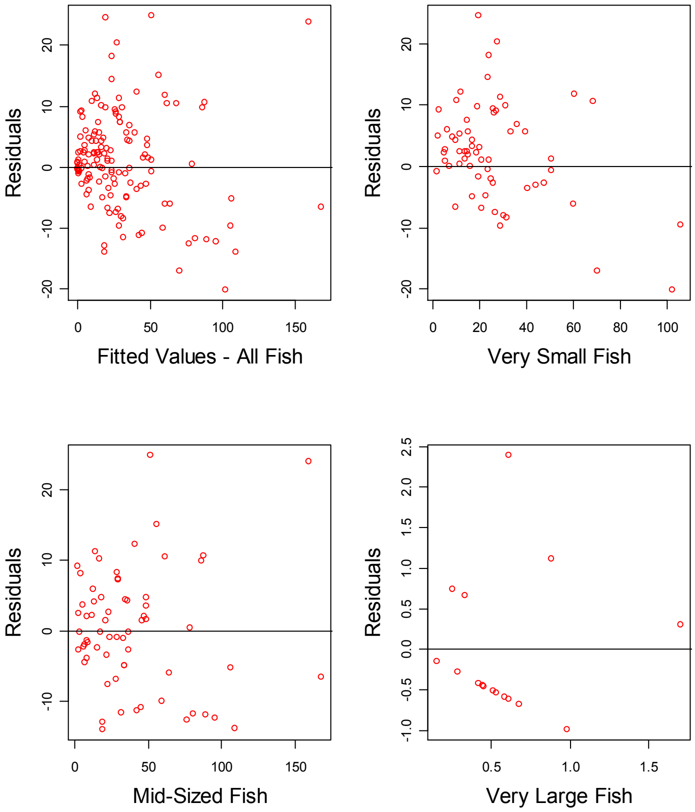


Figure 12. Residuals for the model with three length intervals



The model was then applied assuming that availability depends on both species and length interval, using equation (6). The best fit was obtained using the normal distribution and untransformed n_{ijk}, and under-weighting large values of n_{ijk} for bigeye. The results are shown in Table 6. With the n_{ijk} stratified by both species and length interval, in addition to set, there are many more strata and parameters. The model with the length intervals of 20 cm could be considered the best model, with low bias and a smaller number of parameters than the model with length intervals of 15 cm.

Table 6. Uncorrected species compositions determined from spill samples and grab samples, and the species composition determined from grab samples corrected with estimates of availability

Method	Number of Strata		Number of Parameters		Species Composition			Bias		
	Total	Zero Grab	Total	Zero Grab	Skipjack	Yellowfin	Bigeye	Skipjack	Yellowfin	Bigeye
Spill samples, uncorrected					70.36%	24.67%	4.97%			
Grab samples, uncorrected					64.42%	29.94%	5.64%	-8.4%	+21.4%	+13.4%
Grab samples, corrected, three intervals	367	68	7	0	72.32%	21.98%	5.70%	+2.8%	-10.9%	+14.5%
Grab samples, corrected, four intervals	382	77	9	1	71.52%	23.14%	5.34%	+1.7%	-6.2%	+7.2%
Grab samples, corrected, 20 cm	498	150	16	3	71.66%	23.26%	5.08%	+1.9%	-5.7%	+2.2%
Grab samples, corrected, 15 cm	559	192	20	4	71.93%	23.18%	4.89%	+2.2%	-6.0%	-1.7%
Grab samples, corrected, 10 cm	715	257	29	6	68.51%	26.39%	5.10%	-2.6%	+7.0%	+2.6%

The estimates of availability for the model with length intervals grouped by 20 cm, with bars of plus or minus two standard errors, are shown in Figure 13. The categories for which all n_{ijk} were zero, and thus the estimated A_{ij} were zero, are not shown; these categories include bigeye from 100 cm to 120 cm, skipjack from 0 cm to 20 cm and yellowfin from 140 cm to 160 cm. The estimates in Figure 13 are consistent with the grab sample size selection bias shown in Figures 8 and 9, with the availability of very small fish and, for yellowfin, very large fish, less than that of mid-sized fish. The availability of each category of species and length interval is less than 1% of the number of fish in the set, and the average of the non-zero estimates is 0.41%. The standard error of the estimate of availability for certain categories is much greater than for others. The value for "BET 80", i.e., bigeye from 80 cm to 100 cm, is more than twice as large as the average; this estimate is supported by only one positive n_{ijk} and hence is very uncertain. The model explained 92.1% of the deviance. Residuals are plotted against the fitted values of the n_{ijk} in Figure 14. The striations in Figure 14 are due to the fact that for many strata of species, length interval and set, the numbers of fish selected by a grab sampler are small integer values (e.g., 0, 1, 2, 3, ...).

Figure 13. Estimates of grab sample availability for categories of species and length interval for the model with length intervals of 20 cm

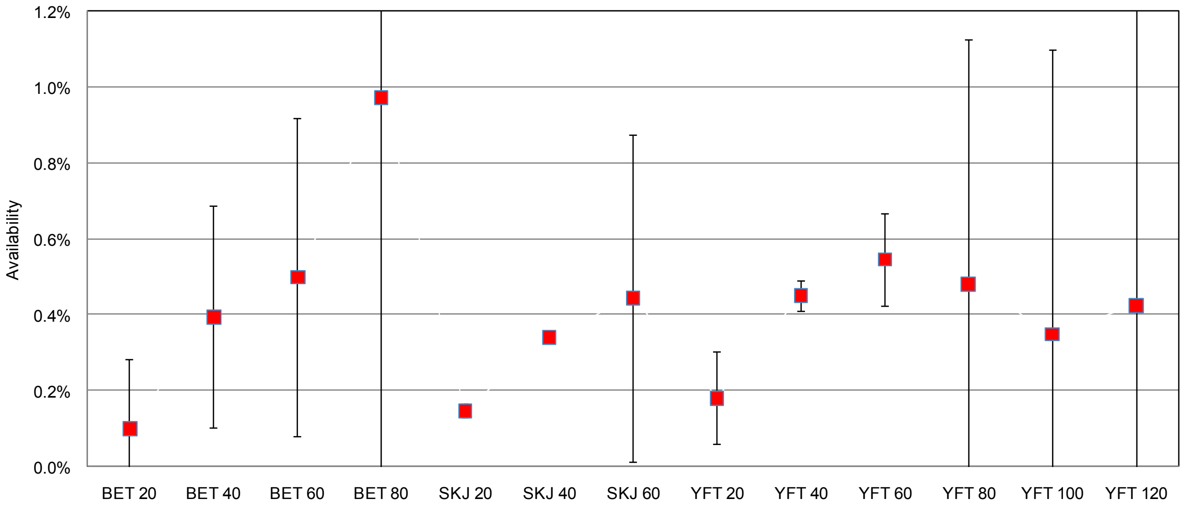
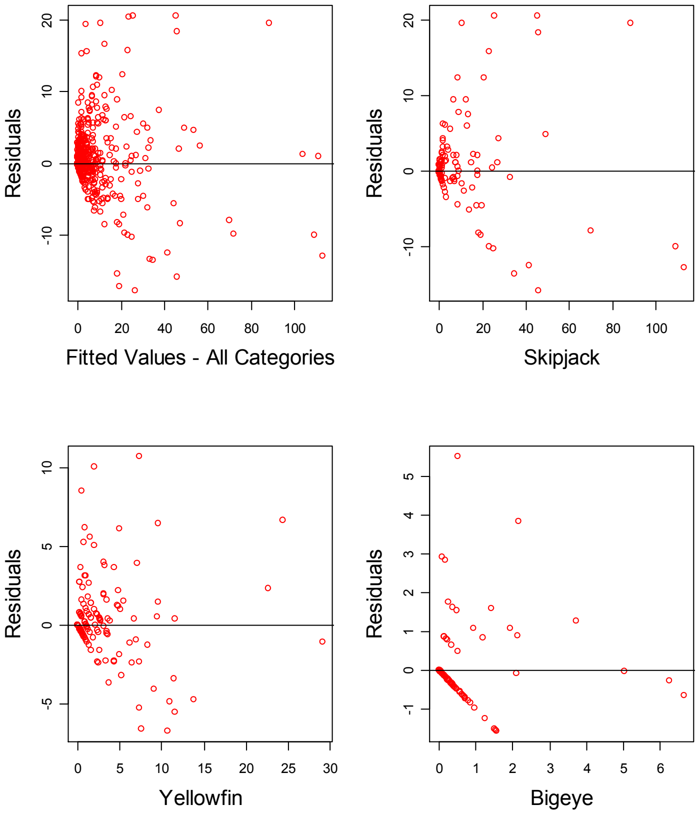


Figure 14. Residuals for the model with species and length intervals of 20 cm



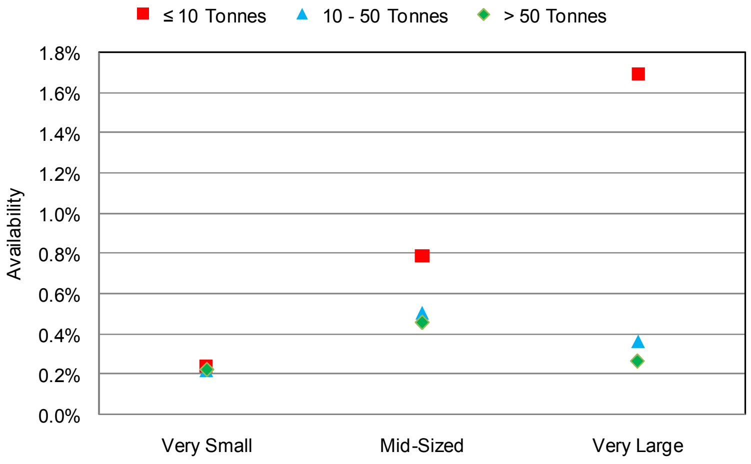
The model with species and length intervals of 20 cm was also formulated with average weights, $\overline{w}_{ijk}$, determined for each set and with the true proportions, T_{ijk}, assumed constant over all sets combined, but in both cases the fit of the model was not improved.

8. Availability and set weight

The number of fish in most sets is sufficiently high that the availability of a category of species and/or length interval should be independent of the size of the set. However, for very small sets this may not be the case. At the extreme, the availability of a fish in a set containing, say, 100 fish will be much greater than the availability of a fish in a set containing 10,000 fish. Figure 15 presents estimates of availability determined from sets of (a) less than 10 tonnes, (b) 10 to 50 tonnes and (c) greater than 50 tonnes; there were 5, 42 and 18 sets in each strata respectively. The model with length intervals of very small, mid-sized and very large fish was used. The average number of fish per set in the three strata of set weight were 3,772 fish, 11,959 fish and 25,797 fish respectively.

Figure 15 suggests that the availability of very small fish is independent of set weight, but that the availability of mid-sized fish and very large fish is greater in sets less than 10 tonnes than in larger sets. However, the number of sets less than 10 tonnes in the analysis is small and while all five contained mid-sized fish, only one of the five sets contained very large fish, and so these results should be considered provisional.

Figure 15. Estimates of grab sample availability for three strata of set weight



9. Correction of species compositions determined from grab samples with estimates of availability

The estimates of availability for (a) three length intervals of very small, mid-sized and very large fish (Figure 11) and (b) categories of species and 20 cm length intervals (Figure 13) were used to correct the species compositions determined from all grab samples collected in the WCPO from 1995 to 2008 combined. For categories of species and 20 cm length intervals for which estimates of availability were unavailable, appropriate substitutions were made. The results are presented in Table 7; the tonnage represents the total of samples raised by the set weight.

For all school associations combined, the corrections differ from the uncorrected species composition by only 2% or 3% for skipjack and yellowfin, and less than 1% for bigeye. The largest corrections are for anchored FADs, which contain more very small fish than the other school associations, while the smallest corrections are for unassociated schools, which contain less. The differences between the corrections for (a) the three length intervals of very small, mid-sized and very large fish, and (b) categories of species and 20 cm length intervals, is only about 1% or less.

Table 7. **Species compositions, uncorrected and corrected with estimates of availability, determined from grab samples taken by observers from 1995 to 2008**

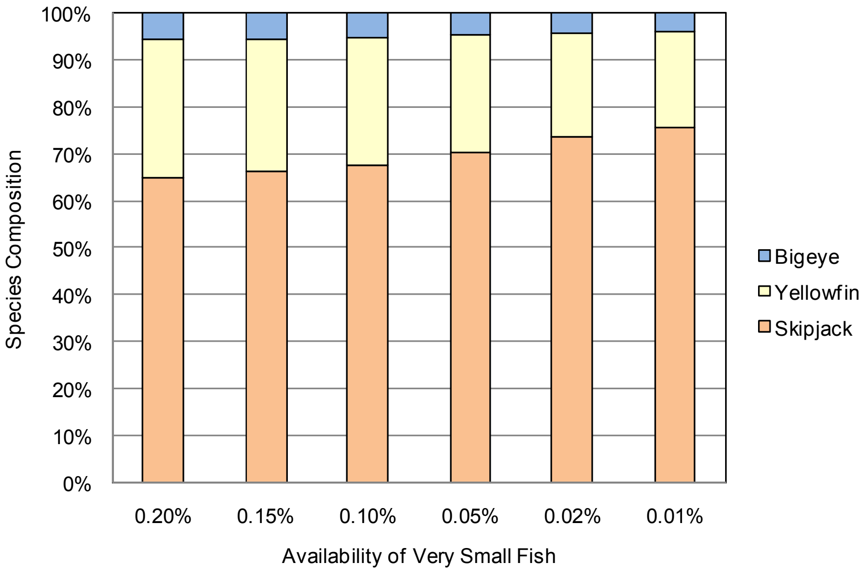
School Association	Skipjack		Yellowfin		Bigeye		Total
	Tonnes	%	Tonnes	%	Tonnes	%	Tonnes
Uncorrected							
Logs	129,444	59.6%	71,842	33.1%	15,738	7.3%	217,024
Drifting FADs	179,443	63.9%	72,661	25.9%	28,783	10.2%	280,887
Anchored FADs	92,261	51.8%	70,132	39.4%	15,739	8.8%	178,133
Unassociated	217,767	69.6%	91,939	29.4%	2,993	1.0%	312,699
All Associations	618,915	62.6%	306,573	31.0%	63,254	6.4%	988,743
Corrected with availability for very small, mid-sized and very large fish							
Logs	137,508	63.4%	65,182	30.0%	14,334	6.6%	217,024
Drifting FADs	186,102	66.3%	68,269	24.3%	26,515	9.4%	280,887
Anchored FADs	100,798	56.6%	63,512	35.7%	13,823	7.8%	178,133
Unassociated	219,123	70.1%	90,720	29.0%	2,856	0.9%	312,699
All Associations	643,532	65.1%	287,683	29.1%	57,528	5.8%	988,743
Corrected with availability for categories of species and 20 cm length intervals							
Logs	139,351	64.2%	62,772	28.9%	14,901	6.9%	217,024
Drifting FADs	189,253	67.4%	64,661	23.0%	26,973	9.6%	280,887
Anchored FADs	102,383	57.5%	61,420	34.5%	14,330	8.0%	178,133
Unassociated	221,197	70.7%	88,675	28.4%	2,827	0.9%	312,699
All Associations	652,184	66.0%	277,528	28.1%	59,031	6.0%	988,743

Given that the estimates of availability were based on paired samples from only four trips in Papua New Guinea during which only anchored FADs were fished, the corrected species compositions in Table 7 should be considered provisional. The corrected species compositions will become more meaningful as additional paired samples are conducted, particularly for the school associations other than anchored FADs.

The sensitivity of the corrected species compositions to the estimates of availability is examined in Figure 16. Each bar represents the species composition determined from all grab samples collected in the WCPO during 1995–2008, corrected by holding the estimate of availability of mid-sized and very large fish at the values shown in Figure 11 and varying the estimate of availability of very small fish. The left-hand bar represents the species composition corrected with the value of the availability of very small fish shown in Figure 11, while the bars to the right represent species compositions corrected with increasingly smaller estimates of availability of very small fish.

The proportions of skipjack and yellowfin are relatively sensitive to the estimate of availability of very small fish, with the proportion of skipjack increasing from 65.0% to 75.6% and the proportion of yellowfin decreasing from 29.3% to 20.4%. An estimate of availability of very small fish of 0.01% (the far right-hand bar) is extremely low, but intermediate values of 0.15%, 0.10% and 0.05% may not be unreasonable, particularly if the availability of very small fish is greater in sets on schools associated with anchored FADs than for other school associations. The proportion of bigeye in Figure 16 decreases from 5.7% to 4.0%; while this is a small change in absolute value, it represents a decrease of (5.7% – 4.0%) / 5.7% = 29.8% in relative terms. The sensitivity to changes in the availability of mid-sized or very large fish was not examined.

Figure 16. Sensitivity of corrected species compositions to the estimate of availability of very small fish



10. Correction of species compositions determined from grab samples with empirical factors

The species composition determined from all grab samples collected in the WCPO from 1995 to 2008 combined were corrected with empirical factors using equation (1). Empirical factors were determined from the paired spill and grab samples for categories of species and length intervals of very small, mid-sized and very large fish. The results presented in Table 8 are similar to those obtained with estimates of availability (Table 7).

Table 8. Species compositions, uncorrected and corrected with empirical factors, determined from grab samples taken by observers from 1995 to 2008

School Association	Skipjack		Yellowfin		Bigeye		Total
	Tonnes	%	Tonnes	%	Tonnes	%	Tonnes
Uncorrected							
Logs	129,444	59.6%	71,842	33.1%	15,738	7.3%	217,024
Drifting FADs	179,443	63.9%	72,661	25.9%	28,783	10.2%	280,887
Anchored FADs	92,261	51.8%	70,132	39.4%	15,740	8.8%	178,133
Unassociated	217,767	69.6%	91,939	29.4%	2,993	1.0%	312,699
All Associations	618,915	62.6%	306,573	31.0%	63,254	6.4%	988,743
Corrected with factors for categories of species and very small, mid-sized and very large fish							
Logs	144,522	66.6%	60,581	27.9%	11,920	5.5%	217,024
Drifting FADs	193,018	68.7%	64,837	23.1%	23,032	8.2%	280,887
Anchored FADs	107,417	60.3%	59,168	33.2%	11,547	6.5%	178,133
Unassociated	215,374	68.9%	95,572	30.6%	1,752	0.6%	312,699
All Associations	660,333	66.8%	280,159	28.3%	48,252	4.9%	988,743

11. Comparison of species compositions determined from observer data and port sampling data

Table 9 compares species compositions determined from port sampling data collected by the National Marine Fisheries Service (NMFS) from United States vessels unloading to canneries in Pago Pago, American Samoa, to observer data corrected with availability using three length intervals. The port sampling and observer data cover the period from 1996 to 2007. The United States vessels do not fish schools associated with anchored FADs and so observer data covering schools associated with anchored FADs were not included in Table 9.

For a detailed description of port sampling in Pago Pago, see Appendix A in Lawson (2008). In most other ports in the region, wells are sampled during transshipment, whereas in Pago Pago, landing categories of species and size, within wells, are sampled and the samples used to correct cannery receipts. Port sampling during transshipment is usually subject to problems of well mixing and non-representative selection of wells in regard to set weights, such that the proportion of skipjack in the species composition is over-estimated and the proportion of yellowfin under-estimated to large degrees (see sections 5 and 10 respectively in Lawson, 2008). The port sampling data from ports other than Pago Pago were therefore not included in this analysis.

The species compositions determined from the NMFS port sampling data and the corrected observer data in Table 9 are relatively similar for unassociated schools, but somewhat divergent for associated schools, with a higher proportion of skipjack and a lower proportion of yellowfin in the species compositions determined from the port sampling data.

Table 9. Species compositions determined from NMFS port sampling data and observer data corrected with availability, collected from 1996 to 2007

School Association	Skipjack		Yellowfin		Bigeye		Total
	Tonnes	%	Tonnes	%	Tonnes	%	Tonnes
NMFS port sampling data, 1996-2007							
Associated	79,405	69.1%	24,274	21.1%	11,262	9.8%	114,941
Unassociated	34,027	65.7%	17,561	33.9%	186	0.4%	51,775
Total	113,432	68.0%	41,836	25.1%	11,449	6.9%	166,716
Observer data corrected with availability for very small, mid-sized and very large fish, 1996-2007							
Associated	300,383	64.5%	126,472	27.1%	39,085	8.4%	465,939
Unassociated	194,279	68.3%	87,348	30.7%	2,820	1.0%	284,447
Total	494,662	65.9%	213,820	28.5%	41,904	5.6%	750,386

Table 10 presents the same comparison as in Table 9, except that only strata of year, 5° latitude, 5° longitude and school association with both port sampling data and observer data were considered; there were 159 and 92 overlapping strata for associated and unassociated schools respectively. The results are similar to those in Table 9.

Table 10. Species compositions determined from NMFS port sampling data and observer data corrected with availability, collected from 1996 to 2007, for overlapping strata of year, 5° latitude, 5° longitude and school association

School Association	Skipjack		Yellowfin		Bigeye		Total
	Tonnes	%	Tonnes	%	Tonnes	%	Tonnes
NMFS port sampling data, 1996-2007							
Associated	58,132	70.3%	16,602	20.1%	7,976	9.6%	82,710
Unassociated	18,378	66.7%	9,023	32.7%	152	0.6%	27,553
Total	76,510	69.4%	25,625	23.2%	8,128	7.4%	110,263
Observer data corrected with availability for very small, mid-sized and very large fish, 1996-2007							
Associated	105,670	61.1%	49,682	28.7%	17,651	10.2%	173,004
Unassociated	46,563	70.8%	18,522	28.2%	671	1.0%	65,756
Total	152,233	63.8%	68,204	28.6%	18,323	7.7%	238,759

The NMFS port sampling data are also subject to non-representative selection of wells in regard to set weights, and the species compositions may be subject to bias due to small sample sizes (see sections 11 and 12 in Lawson, 2008), both of which result in the over-estimation of the proportion of skipjack and the under-estimation of the proportion of yellowfin. On the other hand, port sampling is probably subject to selection bias, which results in the under-estimation of the proportion of skipjack and the over-estimation of the proportion of yellowfin. (There may also be bias due to raising the species compositions with the amount unloaded from the well, rather than the set weight, although it is not known what effect this may have on the species composition; see section 4 in Lawson, 2008.) If selection bias is less of a factor in port samples than observer data, then the species compositions determined from the NMFS port sampling data may over-estimate the proportion of skipjack and under-estimate the proportion of yellowfin because of set weight bias.

Another explanation for the discrepancies in Tables 9 and 10 may be that the availability of very small fish is less in sets on schools associated with logs and drifting FADs than schools associated with anchored FADs, such that the observer data for 1996–2007 have not been corrected appropriately.

12. Application of species compositions determined from corrected observer data to aggregated catch and effort data

Observer data that were corrected for availability using three length intervals of very small, mid-sized and very large fish were used to adjust estimates of annual catches by species determined from aggregated catch data for 1996–2007 in the *s_best* database maintained by the OFP. Annual species compositions for each type of school association — logs, drifting FADs, anchored FADs and unassociated schools — were derived by grouping all samples, corrected for availability and raised by the set weight, for each stratum of year and school association. The annual species compositions were then applied to total catches for each stratum of year and school association determined from *s_best*. That is, the catch data in *s_best* were not adjusted at a finer level of time-area, such as Multifan-CL area and quarter or 1° longitude by 1° latitude and month. Catch estimates were determined for the WCPFC Statistical Area; however, catches taken by the domestic fleets of Indonesia and the Philippines were ignored.

The catch data in *s_best* that were the basis for the adjustment are determined from logsheet data, with an adjustment of the proportions of bigeye and yellowfin in the combined catch of yellowfin plus bigeye reported on the logsheets, using observer data (Lawson 2007). The estimates of skipjack catches in *s_best* are not adjusted and, since they are over-reported on logsheets, are therefore biased upwards.

The catch data in *s_best* were also adjusted using the NMFS port sampling data, with a procedure similar to the adjustment with corrected observer data.

The estimates of annual catches are presented in the Appendix. Table 11 presents a summary of the average annual catches for 1996–2007. As expected, the species compositions determined from *s_best* adjusted with corrected observer data and with NMFS port sampling data contain a smaller percentage of skipjack and greater percentages of yellowfin and bigeye than the species composition determined from *s_best*. The species compositions determined from the two adjustments are relatively consistent, although the adjustment with observer data corrected for availability results in less skipjack and more yellowfin than the adjustment with NMFS port sampling data. The percentages of bigeye determined from the two adjustments are similar.

Table 11. Estimates of average annual catches, 1996–2007

School Association	Skipjack		Yellowfin		Bigeye		Total
	Tonnes	%	Tonnes	%	Tonnes	%	Tonnes
s_best							
Associated	484,111	83.6%	74,090	12.8%	20,652	3.6%	578,853
Unassociated	318,967	76.6%	94,251	22.6%	3,159	0.8%	416,377
All	803,078	80.7%	168,341	16.9%	23,811	2.4%	995,230
s_best adjusted with observer data corrected for availability							
Associated	346,522	59.9%	178,694	30.9%	53,637	9.3%	578,853
Unassociated	282,705	67.9%	127,865	30.7%	5,807	1.4%	416,377
All	629,226	63.2%	306,559	30.8%	59,445	6.0%	995,230
s_best adjusted with NMFS port sampling data							
Associated	392,432	67.8%	125,690	21.7%	60,731	10.5%	578,853
Unassociated	280,836	67.4%	134,018	32.2%	1,523	0.4%	416,377
All	673,268	67.6%	259,707	26.1%	62,255	6.3%	995,230

13. A model-based approach to estimating the species composition for strata with missing data

Observer data corrected for availability were also used to adjust the input data for the MFCL assessments of bigeye and yellowfin. The purse-seine catch and effort data are stratified by year, quarter, MFCL areas 3 and 4, and school association (associated or unassociated). For 1996–2007, there are 12 * 4 * 2 * 2 = 192 strata. However, only 163 strata are covered by observer data, and 22 strata are covered by less than 20 sets. There are, therefore, 29 strata that are missing observer data and many others with a low level of coverage.

The usual approach to dealing with strata that are missing species composition data is to substitute species compositions estimated for neighbouring strata (e.g., Pianet et al. 2000). A statistically more rigorous approach would be to model the species composition on the basis of the variables on which the data are stratified and then predict the species composition for strata with no or low coverage. Equations (14) to (16) model the species composition as a linear function of year *i* (YY_i), quarter *j* (QQ_j), MFCL area *k* (AR_k) and school association *l* (AS_l), where each variable is categorical.

$$p_{ijkl}^{SKJ} = \beta_{Intercept}^{SKJ} + \beta_i^{SKJ} \cdot YY_i + \beta_j^{SKJ} \cdot QQ_j + \beta_k^{SKJ} \cdot AR_k + \beta_l^{SKJ} \cdot AS_l \qquad (14)$$

$$p_{ijkl}^{YFT} = \beta_{Intercept}^{YFT} + \beta_i^{YFT} \cdot YY_i + \beta_j^{YFT} \cdot QQ_j + \beta_k^{YFT} \cdot AR_k + \beta_l^{YFT} \cdot AS_l \qquad (15)$$

$$p_{ijkl}^{BET} = \beta_{Intercept}^{BET} + \beta_{i}^{BET} \cdot YY_{i} + \beta_{j}^{BET} \cdot QQ_{j} + \beta_{k}^{BET} \cdot AR_{k} + \beta_{l}^{BET} \cdot AS_{l} \qquad (16)$$

For each of the 163 strata for which observer data were available, the species composition was estimated from the observer data, corrected for availability and with the species compositions per set raised by the set weight. For each species, the parameters β were estimated from the 163 observed species compositions using a linear regression, weighted by the number of sets used to calculate the species composition in each stratum. Thus, those strata for which there were only a small number of observed sets were still used in the estimation, but their influence on the estimated parameters was small because of the weighting. The deviance explained by the models for skipjack, yellowfin and bigeye was 32.0%, 21.2% and 60.8% respectively.

When the terms in each of the three models are identical, including interaction terms (see below), it can be shown empirically that the species compositions predicted by linear categorical models, such as equations (14) to (16), have the property that the predicted proportions of the three species in a stratum sum to unity, for each and every stratum. This property is, of course, very convenient.

However, there is a small problem. When the proportion of bigeye in a stratum is near zero, the predicted value can sometimes be negative. This was the case for five strata of unassociated schools in MFCL area 4; the range of the predicted values was –1.76% to –0.04%.

There are three solutions to this problem of negative predicted values for near-zero proportions of bigeye. First, the estimation of the model parameters and the predictions could be formulated such that negative values are not permitted; however, this would be complicated and would detract from the simplicity and convenience of the unconstrained, untransformed linear regressions and predictions. Indeed, both an arcsine square-root transformation, often used with proportions (Snedecor & Cochran 1989), and a logit transformation resulted in predicted species compositions that did not sum to unity. Second, the negative values could simply be set to zero and the predicted proportions of skipjack and yellowfin normalised. This would be an acceptable fudge, given the small magnitude of the negative values. In fact, the first solution, i.e., developing a procedure to ensure that the predicted values to sum to unity by *constraining* them to do so, and this second solution, i.e., a simple fudge, are more or less equivalent, with the difference being that the latter is certainly easier to implement. The third solution, which is the most elegant, would be to introduce the concept of *anti-bigeye* [3], which is defined such that equal amounts of bigeye and *anti-bigeye* sum to zero. This third solution will necessitate the complete reformulation of the MFCL bigeye assessment model, however, and so may not be practical.

---

[3] Following Spencer-Brown (1969), it is proposed that the scientific name for this new species be *Thunnus imaginarius* and that the AFSIS three-alpha code be TIM.

The prediction errors were examined by comparing the catches estimated from the species composition determined from observer data corrected with availability, for strata that had at least 50 observed sets, to catches estimated from the species composition predicted from the model. For estimated catches of 10,000, 5,000 and 1,000 tonnes or greater for skipjack, yellowfin and bigeye respectively, the average prediction errors, as a percentage of the former, were 7.16% (101 strata), 1.50% (95 strata) and 0.54% (85 strata), which are low. The standard deviations of the prediction errors, however, were relatively high, i.e., 33.2%, 36.2% and 58.2% for skipjack, yellowfin and bigeye respectively.

The low percentage of deviance explained and the high standard deviations of the prediction errors suggest that the models given by equations (14) to (16) are perhaps too simple. Models containing interaction terms were therefore examined. When all second-order terms — i.e., *YY:QQ*, *YY:AR*, *YY:AS*, *QQ:AR*, *QQ:AS* and *AR:AS* — were included, the deviance explained by the models for skipjack, yellowfin and bigeye was 70.7%, 63.5% and 81.9% respectively, which is a considerable improvement over the models without interactions. However, two parameters were not estimated because the models were rank deficient. From a total of 162 degrees of freedom, the residual degrees of freedom for the model with interactions was 86, still moderately high, but much lower compared to 146 for the model without interactions.

When the interaction between year and area, *YY:AR*, which caused the deficiency, was removed, the deviance explained by the models for skipjack, yellowfin and bigeye was 66.5%, 55.9% and 79.2% respectively, which is still a considerable improvement. There were 95 residual degrees of freedom. Compared to the models without interactions, the average catch prediction errors were lower for skipjack, similar for yellowfin and higher for bigeye, i.e., 1.02%, 2.74% and 17.03%, while the standard deviations of the catch prediction errors were also lower for skipjack, similar for yellowfin and higher for bigeye, i.e., 18.2%, 33.2% and 83.0%. Given that skipjack accounts for a much larger proportion of the catch than yellowfin, and that yellowfin accounts for a much larger proportion than bigeye, these results can be considered more acceptable than the results for the models without interactions.

Except for bigeye, the prediction errors are not appreciably different when presented by school association. For bigeye, the average and standard deviation of the prediction errors for associated schools were 13.91% and 63.67% respectively, while for unassociated schools they were 26.06% and 122.35%. The standard deviation expressed as a percentage of the predicted catch is much higher for unassociated schools because the average catch of bigeye from unassociated schools is much lower.

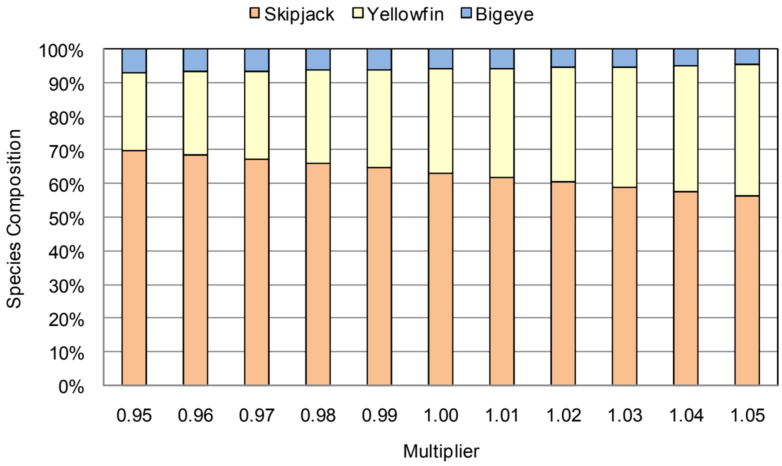
The proportions of the catch per species for 1996–2007 that were estimated from species compositions predicted from the model, i.e., catches in strata covered by less than 20 observed sets, were 8.0%, 8.1% and 9.0% for skipjack, yellowfin and bigeye respectively. These results suggest that even though the reliability of catch estimates based on the predicted species compositions may be low for yellowfin and bigeye, the overall effect on the MFCL assessments should be acceptably small.

An advantage of the model-based approach is that standard errors of the predictions of the species composition for each stratum can be estimated. For the models with all interactions except year–area, the average standard errors of the predicted species compositions for all 192 strata were 9.37%, 8.53% and 3.75% for skipjack, yellowfin and bigeye respectively, compared to the average predicted species composition of 62.95%, 27.83% and 9.21%.

14. Sensitivity of estimates of species composition to length-weight parameters

An aspect of the estimation of purse-seine species composition that has not been discussed above, but which is basic to the estimation, is the use of length-weight parameters to convert the length measurements recorded by the samplers to weights. Figure 16 presents the relationship of the average annual species composition determined from *s_best*, adjusted with observer data corrected for availability, to incremental 1% variations in the length-weight parameters for yellowfin listed in section 3. For each bar in Figure 17, the *a* and *b* parameters for yellowfin were both multiplied by the value shown on the abscissa; the species composition was then determined from the modified yellowfin length-weight parameters, while the skipjack and bigeye length-weight parameters remained unchanged. The species composition is somewhat sensitive to small changes in the length-weight parameters and very sensitive to large changes. Similar results were obtained for skipjack length-weight parameters. The species composition is less sensitive to bigeye length-weight parameters, since bigeye account for a much smaller percentage of the catch, but the estimates of the bigeye catches are still affected proportionately. These results suggest that the accuracy and reliability of length-weight parameters, and their possible variation in time and across geographic areas, should be examined.

Figure 17. Sensitivity of estimates of purse-seine species composition to yellowfin length-weight parameters



The sensitivity of species composition to length-weight parameters also suggests that consideration should be given to measuring weights directly during sampling, rather than estimating weights from lengths. This could be done onboard the vessel by weighing fish in a spill sample with a motion-compensated scale. The scale shown in Figure 18 has been developed specifically for fishing vessels and is available with a variety of load cells, including a dual scale with maximums of 30 kg and 60 kg, and accuracies of 10 gm and 20 gm respectively, which would be appropriate for weighing individual fish. Load cells with a maximum of up to 1500 kg, which would be appropriate for weighing groups of fish sorted by species, are also available.

Figure 18. Marel M1100 motion-compensated crane scale [4]

15. Conclusion

This analysis of selectivity bias in grab samples is based on paired spill and grab samples collected during four trips in Papua New Guinea in 2008, during which only anchored FADs were fished. The data have proved adequate for the development of the methodology, but it may not be appropriate to apply the estimates of availability presented above to samples taken from other school associations. Likewise, the estimates of annual catches determined from the adjustment to *s_best* based on observer data corrected for availability may not be accurate if the estimates of availability based on data from anchored FADs are not applicable to the other school associations. Greater confidence in the results of these analyses will have to await the collection of additional paired spill and grab samples.

---

[4] The SPC Oceanic Fisheries Programme is conducting trials on purse seiners and longliners with a Marel M1100 crane scale with dual 60 kg / 300 kg scales. For enquiries, contact Mr Peter Bullock, National Sales Manager, Marel Food Systems, Brisbane, Australia. Tel: +61 7 3900 3000. Mobile: +61 407 736 729. Fax: +61 7 3900 3033. peter.bullock@marel.com | www.marel.com

Nevertheless, we can still have confidence in three conclusions. First, the remarkable consistency in the species compositions determined from the spill samples for the four trips suggests that this sampling protocol results in data that can be used to estimate the species composition per trip to a high degree of accuracy. Second, grab samples are indeed subject to selectivity bias. Third, the fact that the average annual species compositions determined from adjusting *s_best* with (a) observer data, corrected for availability, and (b) port sampling data, used in conjunction with cannery receipts, are relatively consistent is encouraging. Additional work will be done on the adjustment of the purse-seine catch data as more paired spill and grab samples are conducted and as the spill sample protocol becomes more widely adopted by observer programmes in the region.

References

Cochran, W.G. 1977. Sampling Techniques (Third Edition). Wiley & Sons, New York.

Lawson, T.A. 2007. Analysis of the proportion of bigeye in 'yellowfin plus bigeye' caught by purse seiners in the WCPFC Statistical Area. Working Paper SC3–ST–IP5. Third Regular Session of the Scientific Committee of the Western and Central Pacific Fisheries Commission, 13–24 August 2007, Honolulu, Hawaii, United States of America. Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia
http://www.wcpfc.int/sc3/pdf/ST%20IP-5.pdf

Lawson, T.A. 2008. Factors affecting the use of species composition data collected by observers and port samplers from purse seiners in the Western and Central Pacific Ocean. Working Paper SC4–ST–WP3. Fourth Regular Session of the Scientific Committee of the Western and Central Pacific Fisheries Commission, 11–22 August 2008, Port Moresby, Papua New Guinea. Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia
http://www.wcpfc.int/sc4/pdf/sc4-st-wp3%20lawson.pdf

Millar, R.B. & R.J. Fryer. 1999. Estimating the size-selection curves of towed gears, traps, nets and hooks. Reviews in Fish Biology and Fisheries 9: 89–116.
http://www.springerlink.com/content/h3m60050674005wr/

Pianet, R., P. Pallarés & C. Petit. 2000. New sampling and data processing strategy for estimating the composition of catches by species and sizes in the European purse seine tropical tuna fisheries. IOTC Proceedings No. 3 (2000) : 104-139. Indian Ocean Tuna Commission, Seychelles.
http://www.iotc.org/files/proceedings/2000/wpdcs/IOTC-2000-WPDCS-10.pdf

Snedecor, G.W. & W.G. Cochran. 1989. Statistical Methods (Eighth Edition). Iowa State University Press, Ames, Iowa.

Spencer-Brown, G. 1969. Laws of Form. Allen and Unwin, London.
http://books.google.com/books?id=JbtKAAAAMAAJ&q=%22laws+of+form%22&dq=%22laws+of+form%22&pgis=1

APPENDIX

Table A1. Estimates of annual catches by purse seiners in the WCPFC Statistical Area determined from aggregated data in *s_best*

Associated Schools

Year	Skipjack		Yellowfin		Bigeye		Total
	Tonnes	%	Tonnes	%	Tonnes	%	Tonnes
1996	360,469	84.6%	49,805	11.7%	15,956	3.7%	426,230
1997	294,160	65.9%	111,210	24.9%	40,835	9.2%	446,205
1998	433,650	85.6%	59,381	11.7%	13,847	2.7%	506,878
1999	493,332	77.8%	115,645	18.2%	25,131	4.0%	634,109
2000	421,801	83.0%	68,636	13.5%	18,005	3.5%	508,441
2001	331,745	80.8%	62,849	15.3%	16,206	3.9%	410,799
2002	485,344	85.9%	62,497	11.1%	17,104	3.0%	564,945
2003	381,354	81.8%	68,002	14.6%	16,622	3.6%	465,978
2004	704,301	88.9%	67,955	8.6%	19,576	2.5%	791,832
2005	527,483	81.7%	93,374	14.5%	24,657	3.8%	645,514
2006	701,700	89.1%	65,112	8.3%	20,634	2.6%	787,445
2007	673,994	88.9%	64,618	8.5%	19,248	2.5%	757,860
Average	484,111	83.6%	74,090	12.8%	20,652	3.6%	578,853

Unassociated Schools

Year	Skipjack		Yellowfin		Bigeye		Total
	Tonnes	%	Tonnes	%	Tonnes	%	Tonnes
1996	256,829	86.5%	38,160	12.8%	2,007	0.7%	296,996
1997	195,927	63.8%	102,715	33.4%	8,472	2.8%	307,114
1998	308,569	63.6%	170,118	35.1%	6,197	1.3%	484,884
1999	179,082	75.5%	57,152	24.1%	979	0.4%	237,213
2000	318,098	77.5%	91,295	22.2%	1,160	0.3%	410,553
2001	359,059	73.6%	124,748	25.6%	3,836	0.8%	487,643
2002	399,168	83.7%	75,345	15.8%	2,565	0.5%	477,078
2003	416,153	79.2%	107,611	20.5%	1,582	0.3%	525,346
2004	202,681	79.2%	51,602	20.2%	1,549	0.6%	255,833
2005	436,465	79.4%	109,955	20.0%	2,972	0.5%	549,392
2006	328,780	77.7%	90,736	21.4%	3,797	0.9%	423,312
2007	426,793	78.9%	111,573	20.6%	2,795	0.5%	541,161
Average	318,967	76.6%	94,251	22.6%	3,159	0.8%	416,377

Table A1 (continued)

All Schools							
Year	Skipjack		Yellowfin		Bigeye		Total
	Tonnes	%	Tonnes	%	Tonnes	%	Tonnes
1996	617,298	85.4%	87,965	12.2%	17,963	2.5%	723,226
1997	490,087	65.1%	213,925	28.4%	49,307	6.5%	753,319
1998	742,219	74.8%	229,499	23.1%	20,045	2.0%	991,762
1999	672,414	77.2%	172,798	19.8%	26,110	3.0%	871,322
2000	739,898	80.5%	159,931	17.4%	19,165	2.1%	918,994
2001	690,804	76.9%	187,596	20.9%	20,042	2.2%	898,442
2002	884,512	84.9%	137,842	13.2%	19,669	1.9%	1,042,023
2003	797,507	80.4%	175,613	17.7%	18,204	1.8%	991,325
2004	906,982	86.6%	119,557	11.4%	21,125	2.0%	1,047,665
2005	963,948	80.7%	203,329	17.0%	27,628	2.3%	1,194,906
2006	1,030,479	85.1%	155,847	12.9%	24,431	2.0%	1,210,757
2007	1,100,787	84.7%	176,191	13.6%	22,043	1.7%	1,299,021
Average	803,078	80.7%	168,341	16.9%	23,811	2.4%	995,230

TableA2. Estimates of annual catches by purse seiners in the WCPFC Statistical Area determined from aggregated data in *s_best* adjusted using observer data corrected with availability

Associated Schools							
Year	Skipjack		Yellowfin		Bigeye		Total
	Tonnes	%	Tonnes	%	Tonnes	%	Tonnes
1996	243,109	57.0%	138,526	32.5%	44,595	10.5%	426,230
1997	199,028	44.6%	159,494	35.7%	87,683	19.7%	446,205
1998	300,189	59.2%	154,376	30.5%	52,313	10.3%	506,878
1999	320,237	50.5%	235,643	37.2%	78,228	12.3%	634,109
2000	276,658	54.4%	174,780	34.4%	57,004	11.2%	508,441
2001	222,308	54.1%	144,444	35.2%	44,047	10.7%	410,799
2002	327,120	57.9%	186,753	33.1%	51,073	9.0%	564,945
2003	281,216	60.3%	152,697	32.8%	32,066	6.9%	465,978
2004	519,485	65.6%	210,809	26.6%	61,538	7.8%	791,832
2005	358,761	55.6%	238,240	36.9%	48,513	7.5%	645,514
2006	566,407	71.9%	172,952	22.0%	48,087	6.1%	787,445
2007	543,743	71.7%	175,614	23.2%	38,502	5.1%	757,860
Average	346,522	59.9%	178,694	30.9%	53,637	9.3%	578,853

Table A2 (continued)

Unassociated Schools							
Year	Skipjack		Yellowfin		Bigeye		Total
	Tonnes	%	Tonnes	%	Tonnes	%	Tonnes
1996	246,874	83.1%	43,737	14.7%	6,384	2.1%	296,996
1997	179,508	58.5%	105,498	34.4%	22,108	7.2%	307,114
1998	269,983	55.7%	208,400	43.0%	6,501	1.3%	484,884
1999	144,053	60.7%	89,615	37.8%	3,545	1.5%	237,213
2000	305,586	74.4%	104,037	25.3%	930	0.2%	410,553
2001	317,162	65.0%	165,263	33.9%	5,217	1.1%	487,643
2002	377,155	79.1%	94,490	19.8%	5,434	1.1%	477,078
2003	331,731	63.1%	189,658	36.1%	3,958	0.8%	525,346
2004	190,063	74.3%	64,032	25.0%	1,738	0.7%	255,833
2005	365,853	66.6%	177,732	32.4%	5,807	1.1%	549,392
2006	287,280	67.9%	131,298	31.0%	4,735	1.1%	423,312
2007	377,208	69.7%	160,622	29.7%	3,331	0.6%	541,161
Average	282,705	67.9%	127,865	30.7%	5,807	1.4%	416,377

All Schools							
Year	Skipjack		Yellowfin		Bigeye		Total
	Tonnes	%	Tonnes	%	Tonnes	%	Tonnes
1996	489,983	67.7%	182,264	25.2%	50,980	7.0%	723,226
1997	378,536	50.2%	264,992	35.2%	109,791	14.6%	753,319
1998	570,173	57.5%	362,776	36.6%	58,813	5.9%	991,762
1999	464,290	53.3%	325,258	37.3%	81,773	9.4%	871,322
2000	582,244	63.4%	278,817	30.3%	57,933	6.3%	918,994
2001	539,470	60.0%	309,708	34.5%	49,264	5.5%	898,442
2002	704,275	67.6%	281,242	27.0%	56,507	5.4%	1,042,023
2003	612,946	61.8%	342,355	34.5%	36,024	3.6%	991,325
2004	709,548	67.7%	274,841	26.2%	63,276	6.0%	1,047,665
2005	724,614	60.6%	415,972	34.8%	54,320	4.5%	1,194,906
2006	853,687	70.5%	304,249	25.1%	52,822	4.4%	1,210,757
2007	920,952	70.9%	336,237	25.9%	41,833	3.2%	1,299,021
Average	629,226	63.2%	306,559	30.8%	59,445	6.0%	995,230

Table A3. Estimates of annual catches by purse seiners in the WCPFC Statistical Area determined from aggregated data in *s_best* adjusted using NMFS port sampling data

Associated Schools							
Year	Skipjack		Yellowfin		Bigeye		Total
	Tonnes	%	Tonnes	%	Tonnes	%	Tonnes
1996	320,823	75.3%	47,776	11.2%	57,631	13.5%	426,230
1997	303,153	67.9%	105,060	23.5%	37,993	8.5%	446,205
1998	425,610	84.0%	59,793	11.8%	21,476	4.2%	506,878
1999	485,568	76.6%	99,427	15.7%	49,114	7.7%	634,109
2000	302,392	59.5%	141,002	27.7%	65,047	12.8%	508,441
2001	299,779	73.0%	67,648	16.5%	43,372	10.6%	410,799
2002	397,482	70.4%	111,564	19.7%	55,899	9.9%	564,945
2003	266,347	57.2%	151,576	32.5%	48,055	10.3%	465,978
2004	463,914	58.6%	237,468	30.0%	90,449	11.4%	791,832
2005	383,095	59.3%	174,926	27.1%	87,493	13.6%	645,514
2006	523,445	66.5%	161,676	20.5%	102,325	13.0%	787,445
2007	537,576	70.9%	150,363	19.8%	69,921	9.2%	757,860
Average	392,432	67.8%	125,690	21.7%	60,731	10.5%	578,853

Unassociated Schools							
Year	Skipjack		Yellowfin		Bigeye		Total
	Tonnes	%	Tonnes	%	Tonnes	%	Tonnes
1996	265,758	89.5%	31,238	10.5%	0	0.0%	296,996
1997	189,715	61.8%	117,399	38.2%	0	0.0%	307,114
1998	255,791	52.8%	225,511	46.5%	3,582	0.7%	484,884
1999	191,750	80.8%	45,464	19.2%	0	0.0%	237,213
2000	337,554	82.2%	70,877	17.3%	2,121	0.5%	410,553
2001	371,374	76.2%	115,597	23.7%	672	0.1%	487,643
2002	298,725	62.6%	176,032	36.9%	2,322	0.5%	477,078
2003	317,179	60.4%	206,840	39.4%	1,327	0.3%	525,346
2004	142,475	55.7%	112,871	44.1%	486	0.2%	255,833
2005	346,499	63.1%	199,886	36.4%	3,007	0.5%	549,392
2006	322,964	76.3%	100,024	23.6%	324	0.1%	423,312
2007	330,248	61.0%	206,474	38.2%	4,440	0.8%	541,161
Average	280,836	67.4%	134,018	32.2%	1,523	0.4%	416,377

Table A3 (continued)

All Schools							
Year	Skipjack		Yellowfin		Bigeye		Total
	Tonnes	%	Tonnes	%	Tonnes	%	Tonnes
1996	586,581	81.1%	79,014	10.9%	57,631	8.0%	723,226
1997	492,868	65.4%	222,458	29.5%	37,993	5.0%	753,319
1998	681,401	68.7%	285,304	28.8%	25,057	2.5%	991,762
1999	677,317	77.7%	144,890	16.6%	49,114	5.6%	871,322
2000	639,947	69.6%	211,879	23.1%	67,168	7.3%	918,994
2001	671,153	74.7%	183,244	20.4%	44,044	4.9%	898,442
2002	696,207	66.8%	287,595	27.6%	58,221	5.6%	1,042,023
2003	583,527	58.9%	358,416	36.2%	49,382	5.0%	991,325
2004	606,389	57.9%	350,340	33.4%	90,936	8.7%	1,047,665
2005	729,594	61.1%	374,811	31.4%	90,500	7.6%	1,194,906
2006	846,409	69.9%	261,699	21.6%	102,649	8.5%	1,210,757
2007	867,824	66.8%	356,837	27.5%	74,361	5.7%	1,299,021
Average	673,268	67.6%	259,707	26.1%	62,255	6.3%	995,230