16th Meeting of the Standing Committee on Tuna and Billfish

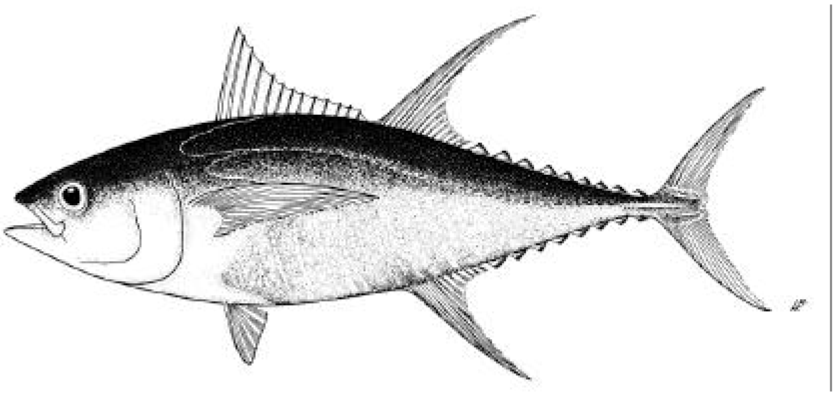
SCTB 16
Mooloolaba, Qld, Australia
9-16 July 2003

SCTB16 Working Paper

SWG–7

DESIGN OF OBSERVER PROGRAM
TO MONITOR BYCATCH SPECIES
IN THE EASTERN TUNA ANDBILLFISH FISHERY

Mark Bravington
Charis Burridge
Peter Toscas



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

June 2003

DESIGN OF OBSERVER PROGRAM TO MONITOR BYCATCH SPECIES IN THE EASTERN TUNA AND BILLFISH FISHERY

Mark Bravington

Charis Burridge

Peter Toscas

CSIRO
Mathematical
& Information
Sciences

December 2002

Introduction

In this report, we consider a range of possible designs for an observer program in the ETBF. For each design, and for a range of important by-catch species, we attempt to answer the question: how precise is the estimated annual by-catch likely to be? We also assess the utility of current logbook records for monitoring annual by-catch of these five species.

With around 60 species known to occur as by-catch[1], some summarizing is necessary. We chose five representative species or species groups, ranging from rare to common as by-catch, and with different spatial distributions: turtles (identified to family, not to species), black marlin (*Makaira indica*), blue marlin (*Makaira mazara*), dolphinfish (*Coryphaena hippurus*), and blue sharks (*Prionace glauca*). We considered five possible survey designs with between 5% and 20% overall coverage, and estimated the likely precision from each design for each species.

The survey designs considered (i.e. the breakdown of observer coverage by region and season) are only meant for exploring of the likely precision obtainable from a survey; they should not be treated as formal designs, let alone "optimal" designs. Because pre-existing data are limited, because there are multiple species with different patterns of by-catch in space and time, and most of all because there is likely to be inter-annual variability in local abundance, an "optimal" design is not feasible at this point; the possibilities for improved efficiency (i.e. equal precision for lower cost) can be considered once a few years of reliable data have been gathered.

Outline of source data and methods of analysis

Based on recent logbook records from the Eastern Tuna and Billfish Fishery (ETBF) we split the fishery into five regional strata and four seasons, to capture large-scale spatial and seasonal variation in abundance. Reported catch rates (number of animals per 1000 hooks) were evaluated in each stratum & season for the four fish species. The estimated annual catch rate for turtles was taken from Robins and Bache (2002)[2].

Because of concerns that logbook data may under- or over-record by-catch of some species, we used observer data from Japanese longline fishing on the east coast in 1991-1995 ("JOP") to provide a rough baseline for likely current by-catch rates (per 1000 hooks) in each stratum and season, and for variability across longline sets within a stratum/season. We combined this with 2001 logbook data on fishing effort, to produce multiple simulated datasets, each comprising a random sample of observed sets. From each simulated dataset, we calculated what the estimate of total by-catch would be for each species. To find the precision, we then looked at how much the estimate varied across all the simulated datasets.

The Japanese and Australian fisheries differ in terms of space/time distribution and targetting. Even in those times and places where the two fisheries overlapped, there were fairly substantial differences in catch rate for target species, of the order of 2- or 3-fold (Campbell 1999, p20-22, Table 3.6). It is plausible that there will be differences in by-catch rates, too, e.g. for species caught predominantly by attraction to light at night. Further, because of the

---

[1] See p70 of "Non-target species in Australia's Commonwealth fisheries: a critical review" by Aubrey Harris & Peter Ward, Bureau of Rural Sciences, Canberra, 1999.

[2] Robins, C. and Bache, S. (2002) Bycatch of Sea Turtles in Longline Fisheries – Australia. Bureau of Rural Sciences Final Report to the Fisheries Resources Research Fund, Agriculture, Fisheries and Forestry Australia, Canberra.

seasonal nature of the Japanese fishery, there is no JOP by-catch data for spring, and summer by-catch data is available only for the northernmost region. To fill these gaps in coverage, we have had to assume that the rate from an adjacent well-sampled season applies; this is clearly not an accurate description of seasonal fish movements, and adds some uncertainty to the results. In the most southerly region only, there are new data from a domestic seabird observer program; these were not available in time for this draft, but will be incorporated before 11th December meeting. However, for most of the ETBF region, there is at present no alternative data source to the JOP. The upshot is that the initial design of the program should not be fixed for all time. Once an observer program has run for a year or two, the design can and should be refined, based on the new data and consideration of priorities.

Definition of strata

Catch rates of species vary both in space and in time, and a survey that stratifies geographically and temporally to take account of this will produce a better estimate of the total catch. Effort levels also show spatial and temporal variation. For the exploratory analyses in this report, we split the fishery into five geographic strata: one region north of 22° S, one region south of 34° S, and a central region that was split into three components: a western inshore region extending out to 155° E, a middle region from 155° E to 158° E (approximately the Lord Howe Rise) and an eastern region east of this to about 170° E (east of Norfolk Island) (Figure 1).

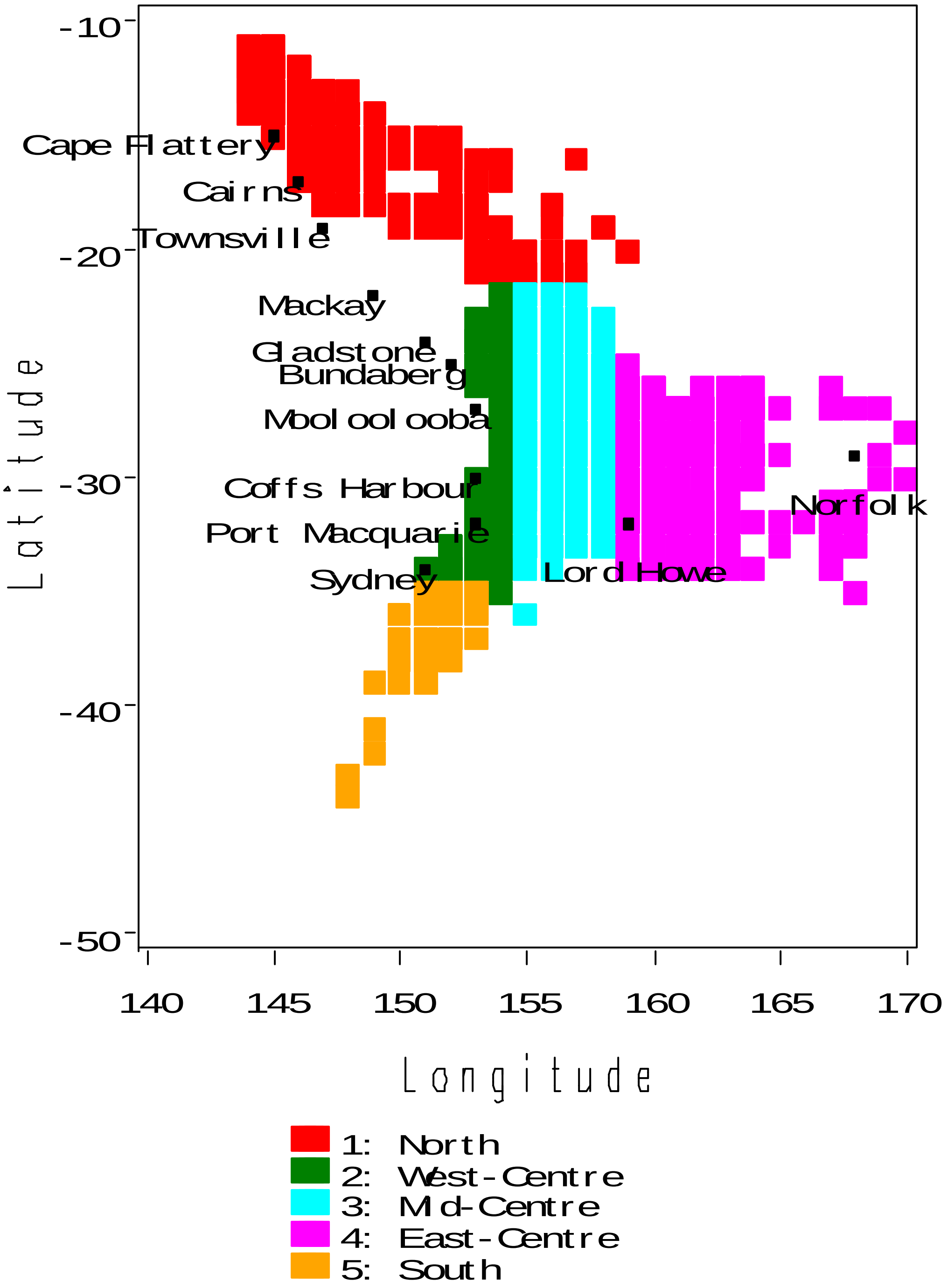


Figure 1. Stratified effort distribution in 2001.

Of the 11723 sets in the 2001 logbook data, a high proportion lay in a coastal strip roughly 60 n.mile wide (Figure 2), with one peak of effort around Cairns in the northern stratum and another stretching from Bundaberg to Coffs Harbour in the west-centre stratum. Substantial fishing also occurred in the mid-centre stratum.

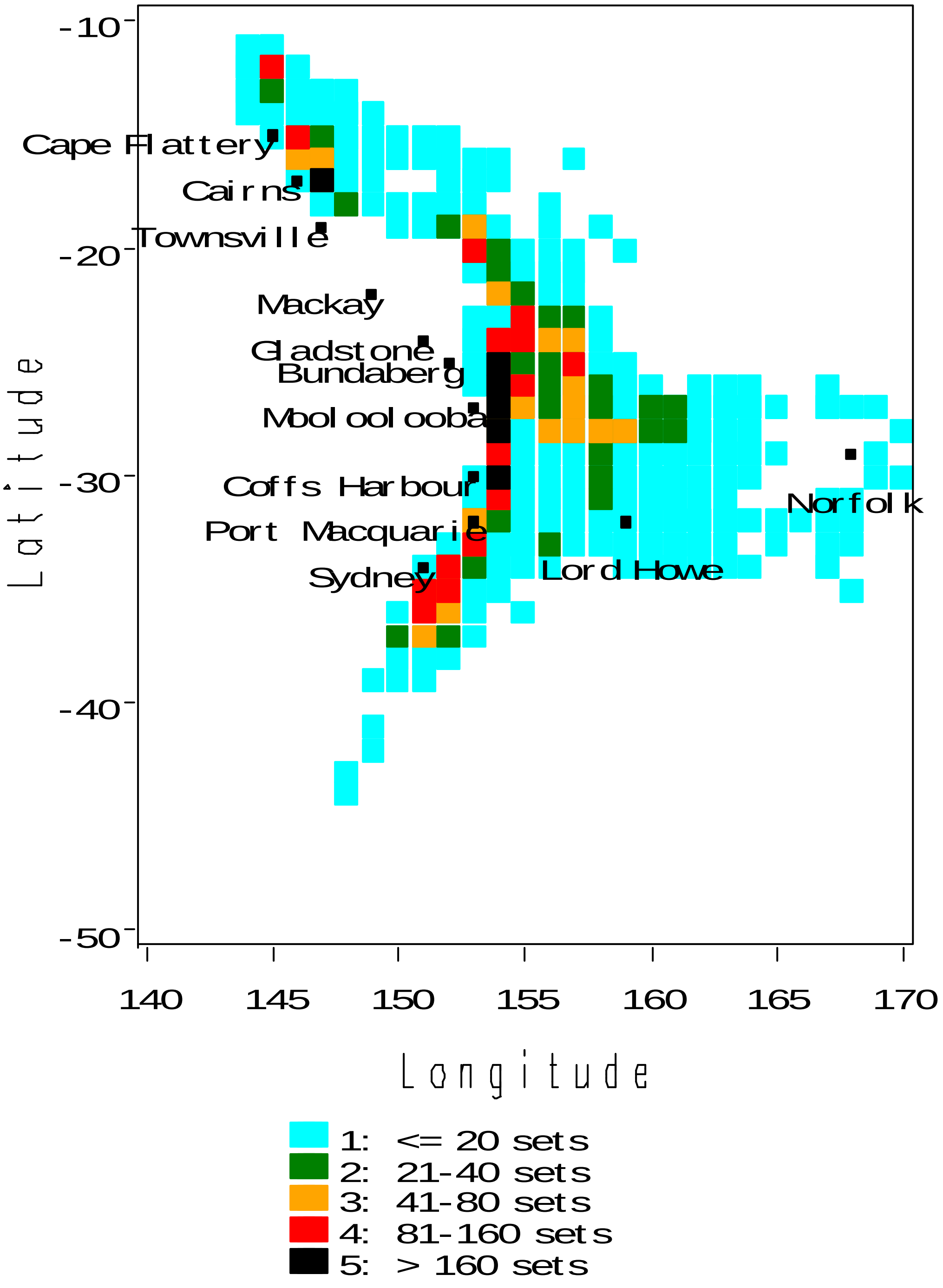


Figure 2. **Distribution of effort (number of sets) in 2001.**

In the northern and east-central strata, effort was fairly evenly distributed throughout the year (Table 1), but in the west-central and southern strata there was pronounced seasonal variation in effort. The number of sets ranged from 76 in the southern stratum in October-December 2001, to 1,498 sets in the west-central stratum in July-September 2001 (Table 1).

Table 1. Number of longline sets in the Eastern Tuna and Billfish Fishery in 2001.

Stratum	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec
North	515	608	535	495
West-centre	913	1176	1498	665
Mid-centre	609	730	891	1007
East-centre	188	171	278	236
South	471	487	174	76

Sets in the central and southern strata contained roughly 1,000-1,100 hooks on average, whereas those in the northern stratum had about 500-600 hooks (reflecting the effort limit in that region). Total number of hooks deployed ranged from 66,000 in October-December in the southern stratum up to 1,384,000 in July-September in the west-central stratum (Table 2).

Table 2. Number of longline hooks (thousands) in the East Coast Tuna and Billfish Fishery in 2001.

Stratum	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec
North	336	435	316	288
West-centre	880	1077	1390	626
Mid-centre	600	701	880	1018
East-centre	192	172	276	246
South	418	434	158	67

The overall level of effort and the distribution among strata and seasons was broadly similar for the three years 1999-2001. Effort levels fluctuated by about 20% of the mean from year to year. An exception to this was that the level of effort in the west-central stratum from April-June in 2001 was nearly twice that for the same period in 1999 and 2000.

In some fisheries, there are identifiable classes of boat or fishing operation, which allow stratification by variables other than space and time. We have not considered this possibility for the ETBF; to some extent, the inshore-offshore split will reflect this anyway, e.g. with only the bigger boats able to operate in the rich but distant waters around Norfolk Island.

Modelling the catch-effort relationship for bycatch species from the JOP

There are two key pieces of information required for a simulation study of precision: the mean catch rate in each stratum, and the variability of catch from set to set. The latter should reflect both the randomness due to presence-absence of by-catch in a particular set, and any "clustering" that arises due to several animals being caught together when there is a non-zero by-catch. This section describes the technical details of how we used existing data to model the mean and the variability.

Historical data were available on four of the selected bycatch species (excluding turtles) from the JOP. The period 1991-1995 was selected as being the most representative in terms of effort and catch. The data used were the hook count (from about 200 to more than 3000) for

each set, and the number of animals caught. We excluded from analysis the small number of sets with fewer than 500 hooks, so that the distribution of hook counts would be reasonably uniform and continuous.
We used a generalised linear model procedure to fit the following model to the number of animals per set:

$$E(Y_{ijk}) = a_{ij} H_{ijk} \qquad (1)$$

where Y_{ijk} is the number of animals in the k'th set in stratum i, season j

a_{ij} is the mean catch rate (per thousand hooks) in stratum i, season j

and H_{ijk} is the number of hooks (in thousands) in the k'th set in stratum i, season j.

This model describes how the mean varies between the strata, but needs extension in order to describe the variability between sets. The simplest assumption—and the best case, in terms of precision— is that this model captures all the important sources of variation: i.e. that there is little systematic variation between years, or between localities within a stratum, and not much tendency towards local aggregations. In such cases, a Poisson model for variability is sufficient. The data for blue and black marlin showed reasonable support for this assumption.

However, for blue sharks and dolphinfish (where catch rates were higher), the data showed significant between-set variability that the Poisson model could not account for ("over-dispersion"). To allow for this, we fitted a Negative Binomial model incorporating a dispersion paramater, ϕ, with a mean-variance relationship as follows:

$$V(Y_{ijk}) = m_{ijk}(1 + f m_{ijk}) \qquad (2)$$

The data were too sparse to allow a separate model to be fitted for every combination of species, stratum, and season, so some combining of strata was necessary. Stratum-season combinations with at least 9 sets were included in analysis. For the three central strata, sufficient data were available from three seasons. For the northern stratum, all four seasons were included, but in the southern stratum only two seasons were included. For April-June in the southern stratum there were enough sets per year to allow a separate model to be fitted for each year (at least 100 sets per year). For 1991 and 1992, there were also enough sets in the mid-central stratum. Otherwise, data were combined across the five years, producing a maximum of 21 stratum-by-season datasets for each species (Table 3). Blue sharks were very abundant in the southern stratum, but none of the other three species were reported in this stratum.

Table 3. Number of sets observed in the Japanese east coast longline fishery in 1991-1995, by season.

Stratum	Year(s)	Season	Number of sets
North	91-95	Jan-Mar	53
		Apr-Jun	19
		Jul-Sep	86
		Oct-Dec	14
West-centre	91-95	Jan-Mar	10
		Apr-Jun	9
		Jul-Sep	221
Mid-centre	91-95	Apr-Jun	31
	91	Jul-Sep	125
	92	Jul-Sep	166
	93-95	Jul-Sep	76
	91-95	Oct-Dec	13
East-centre	91-95	Jan-Mar	21
		Apr-Jun	49
		Jul-Oct	38
South	91	Apr-Jun	297
	92	Apr-Jun	190
	93	Apr-Jun	337
	94	Apr-Jun	133
	95	Apr-Jun	113
	91-95	Jul-Sep	198

Estimated underlying catch rates are shown in Table 4; these are very close to the empirical catch rates, but have been adjusted to allow for differences in variability caused by different numbers of hooks per set. An informal comparison between-year variances in the point estimates, and measurement variance obtained from the standard errors, suggests that there is genuine year-to-year variability, at least for some species in some strata and seasons (e.g. blue marlin and blue shark, mid-centre, July-September).

Table 4. Estimated catch rates (animals per 1000 hooks) for four bycatch species from Japanese observer data for east coast longline fishing. Standard error is in parentheses. Asterisk denotes combinations where a ratio estimate was obtained, either because there were less than 5 animals in total or because there were less than 15 sets in total.

Stratum	Year(s)	Season	Black marlin	Blue marlin	Dolphinfish	Blue shark
North	91-95	Jan-Mar	0.145 (0.036)	1.200 (0.104)	0.277 (0.064)	1.641 (0.311)
		Apr-Jun	1.275 (0.159)	0.259 (0.072)	1.453 (0.364)	0.717 (0.120)
		Jul-Sep	0.190 (0.031)	0.077 (0.020)	1.436 (0.158)	1.223 (0.173)
		Oct-Dec	0.687*	0.315*	1.088*	0.601*
West-centre	91-95	Jan-Mar	0.337*	1.878*	1.300*	0.578*
		Apr-Jun	0.112*	0.112*	0.279*	3.300*
		Jul-Sep	0.045 (0.010)	0.018 (0.006)	0.202 (0.030)	0.606 (0.052)
Mid-centre	91-95	Apr-Jun	0.123 (0.044)	0.139 (0.046)	0.267 (0.084)	1.151 (0.256)
	91	Jul-Sep	0.024 (0.009)	0.014 (0.007)	0.222 (0.039)	0.473 (0.062)
	92	Jul-Sep	0.018 (0.007)	0.039 (0.010)	0.168 (0.032)	1.006 (0.103)
	93-95	Jul-Sep	0.012*	0.035 (0.014)	0.183 (0.049)	1.357 (0.210)
	91-95	Oct-Dec	0.345*	0.421*	0.460*	0.498*
East-centre	91-95	Jan-Mar	0.064*	0.021*	0.277 (0.082)	2.182 (0.317)
		Apr-Jun	0.088 (0.028)	0.079 (0.026)	0.123 (0.035)	1.907 (0.295)
		Jul-Oct	0.032*	0	0.145 (0.063)	2.141 (0.289)
South	91	Apr-Jun	0	0	0	9.964 (0.572)
	92	Apr-Jun	0	0	0	6.431 (0.493)
	93	Apr-Jun	0	0	0	6.035 (0.325)
	94	Apr-Jun	0	0	0	7.688 (0.877)
	95	Apr-Jun	0	0	0	10.772 (1.018)
	91-95	Jul-Sep	0	0	0	12.413 (0.983)

The two marlin species were modelled by the Poisson distribution, for which the variance is equal to the mean. Dolphinfish and blue sharks showed over-dispersion relative to the

Poisson, and were modelled by the Negative Binomial distribution. For dolphinfish the dispersion parameter estimates ranged from 0.49 to 3.48. For most estimates, the standard error was about 25-30% of the estimate. This confirms the need to model extra-Poisson variation. For blue sharks, the dispersion parameter estimates ranged from 0.89 to 1.67, and were generally better defined than the dolphinfish dispersion parameter (standard error 10-20% of the estimate). This is not surprising, given that the catch rate for blue sharks is about 10 times higher than that of dolphinfish.

Turtles were seen only rarely in the JOP data. Instead, we used estimates from Robins and Bache, who based their results on recent interview surveys with Australian fishers. Robins and Bache found relatively little space/time pattern in ETBF turtle catch rates. In the absence of a clear pattern, and given the substantial uncertainty always associated with patterning of rare events such as turtle by-catch, we assumed a constant rate of 0.024/1000 hooks in all seasons and areas, compared to Robins and Bache's estimates of 0.012, 0.033, 0.024 in the north, centre, and south respectively. We did not consider the possibility of clustering for turtles; with such a low overall by-catch rate, the dominant source of uncertainty is perhaps more likely to be sampling variability (which we have modelled) rather than clustering within sets (which we have not).

For all species, if no data for a particular season were available, we extrapolated by-catch rates from adjacent seasons. The final set of rates used for simulations is shown in Table 5.

Table 5. Catch rates used for simulating data from an observer program (Negative Binomial dispersion parameter in parentheses for blue sharks and dolphinfish).

Species	Stratum	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec
Black marlin	North	0.165	0.165	0.183	0.183
	West-centre	0.045	0.045	0.045	0.045
	Mid-centre	0.122	0.122	0.018	0.018
	East-centre	0.091	0.091	0.091	0.091
	South	0.091	0.091	0.091	0.091
Blue marlin	North	1.221	1.221	0.074	0.074
	West-centre	0.091	0.091	0.091	0.091
	Mid-centre	0.135	0.135	0.037	0.037
	East-centre	0.082	0.082	0.082	0.082
	South	0.091	0.091	0.091	0.091
Dolphinfish	North	0.183 (1.1)	0.183 (1.1)	1.492 (0.7)	1.492 (0.7)
	West-centre	0.202 (2.6)	0.202 (2.6)	0.202 (2.6)	0.202 (2.6)
	Mid-centre	0.273 (1.2)	0.273 (1.2)	0.223 (1.2)	0.223 (1.2)
	East-centre	0.122 (0.5)	0.122 (0.5)	0.122 (0.5)	0.122 (0.5)
	South	0.091 (0.5)	0.091 (0.5)	0.091 (0.5)	0.091 (0.5)
Blue shark	North	1.822 (1.5)	1.822 (1.5)	1.221 (1.4)	1.221 (1.4)
	West-centre	0.607 (0.9)	0.607 (0.9)	0.607 (0.9)	0.607 (0.9)
	Mid-centre	1.105 (1.1)	1.105 (1.1)	0.497 (1.2)	0.497 (1.2)
	East-centre	1.822 (0.9)	1.822 (0.9)	1.822 (0.9)	1.822 (0.9)
	South	9.974 (0.9)	9.974 (0.9)	12.182 (1.2)	12.182 (1.2)
Turtles	(all)	0.024	0.024	0.024	0.024

Comparison of by-catch rates from logbooks and from the observer program

Table 6 shows comparative by-catch rates for the four fish species, based on JOP data and 2001 logbook data. For dolphinfish, logbook by-catch rates are 5-10 times higher than in the JOP data for corresponding strata. For all other species, logbook rates are generally much lower than JOP rates; typically 3-4 times for the marlins, and up to 20 times for blue shark. Apart from the overall reduced rate in logbooks, there is considerable variation across strata and season, but with no obvious pattern for the marlins. For blue sharks, though, there is a consistent and striking variation in the JOP-to-logbook ratio with stratum; JOP rates are about 2 times higher in the north, roughly equal in the centre, and about 20 times higher in the south. Although north-south differences in *comparative* fishing practices may account for some of this difference, such a large spatial difference in JOP-to-logbook ratio is suggestive of spatial variation in under-reporting.

Table 6. Mean by-catch rates (number per thousand hooks) for two species in the Eastern Tuna and Billfish Fishery. Upright type = 2001 logbooks, italics = Japanese observer program, 1991-1995; ** =small sample size in JOP.

Species	Stratum	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec
Black marlin	North	0.11 *0.14*	0.08 *1.28*	0.05 *0.19*	1.36 *0.69*
	West-centre	0.07 *0.34*	0.03 *0.11*	0.01 *0.05*	<0.01 **
	Mid-centre	0.22 **	0.02 *0.12*	0.01 0.02	0.01 *0.35*
	East-centre	0.02 *0.06*	0.01 *0.09*	<0.01 *0.03*	0.01 **
	South	0.01 **	<0.01*<0.01*	<0.01*<0.01*	<0.01 **
Blue marlin	North	0.41 *1.20*	0.06 *0.26*	0.08 *0.08*	1.03 *0.32*
	West-centre	0.19 *1.88*	0.04 *0.11*	0.01 *0.02*	0.03 **
	Mid-centre	0.13 **	0.05 *0.14*	0.02 *0.03*	0.01 *0.42*
	East-centre	0.03 *0.02*	0.03 *0.08*	<0.01 *<0.01*	0.01 **
	South	0.01 **	<0.01 *<0.01*	<0.01*<0.01*	<0.01 **
Blue sharks	North	0.45 *1.64*	0.38 *0.72*	0.53 *1.22*	0.48 *0.60*
	West-centre	0.34 *0.58*	0.35 *3.30*	0.12 *0.61*	0.05 **
	Mid-centre	0.98 **	0.86 *1.15*	0.59 *1.00*	0.21 *0.50*
	East-centre	2.80 *2.18*	1.82 *1.91*	2.44 *2.14*	0.74 **
	South	0.37 **	0.43 *7.13*	0.21 *12.41*	0.21 **
Dolphinfish	North	2.93 *0.28*	4.74 *1.45*	5.89 *1.44*	1.62 *1.09*
	West-centre	3.96 *1.30*	1.33 *0.28*	0.55 *0.20*	7.71 **
	Mid-centre	1.77 **	2.23 *0.27*	0.79 *0.19*	2.83 *0.46*
	East-centre	1.35 *0.28*	0.86 *0.12*	0.32 *0.15*	1.23 **
	South	1.44 *<0.01*	0.78 *<0.01*	0.31 *<0.01*	0.21 *<0.01*

On the face of it, it might seem attractive to combine logbook and observer data on by-catch rates. Even if there is under-reporting in the logbooks, it might seem possible to use a limited observer program to calibrate the extent of under-reporting, and then scale up the logbook records accordingly. However, the data above suggest that this would not work reliably. As the blue shark data show, the extent of under-reporting can vary spatially. Also, recent logbook data show a 3-fold change in recorded catches of blue marlin between 1999 and 2001[3] suggesting variability in reporting rates over time as well. To cope with variability in space and time, it would still be necessary to have an observer program that covered the different strata, seasons, and years. And if a full observer program is necessary anyway, then there is minimal statistical benefit in trying to combine logbook and observer data on by-catch rates. The most reliable estimates will be obtained by using a stratified ratio estimator: take the *observed rates* from the observer program, and scale up by the *total fishery effort* recorded in the logbooks.

Design and analysis of an observer program

In the simulations presented here, we have assumed that the program will operate by placing observers on a predetermined number of trips in each stratum/season. Also, although the total

---

[3] See Hartmann & Lynch, AFMA ETBF Data Summary, 2002.

number of trips is fixed, the trips actually observed are assumed to form representative random sample; in practice, this means that trips within a stratum should as far as possible be balanced (sampled proportionally) across ports and boat types.

Analysing the data from such a program is simple: calculate the observed catch rate in each stratum/season, and scale up by the total effort. There will also be another analysis to evaluate the post-hoc precision, but this too is straightforward.) Designing the program is more complicated: how much observer coverage should be there be in total, and how should it be apportioned in each stratum and season?

The total amount of observer coverage is determined largely by non-statistical considerations: how much precision is enough, and how much will the program cost? But even calculating what precision can be obtained for a given total, entails consideration of the allocation between strata. In principle, this is a well-understood statistical question, discussed for example in Cochran (1963)[4]. The classical answer is that the optimal number of observed trips in a stratum should be proportional to

$$\text{total trips in stratum} \times \sqrt{\text{by-catch rate in stratum} \times \text{variance-mean ratio of sets in stratum}}$$

In practice, this generally means that strata with more fishing effort should receive proportionally more observer trips. Also, strata with higher by-catch rates should ideally receive *higher* coverage in absolute terms, but *lower* coverage in proportional terms. These remain useful rules of thumb. However, if taken too literally, the classical theory immediately runs into several basic problems:

(i) How should strata be chosen? In principle, it is desirable to make strata small, because this reduces the component of variance that is caused by systematic differences in by-catch rate within each stratum. However, although *total* variance will tend to be smaller, smaller strata will show *individually* more variance. This makes it difficult to improve design efficiency in subsequent surveys, because of high uncertainty about the real by-catch rate in each small stratum.

(ii) It may not be possible to forecast accurately what the total effort in a stratum will be, at the time that the number trips to be observed is being decided. This could lead to designs that turn out post-hoc to be highly non-proportional to actual effort, and thus quite inefficient. This is a particular risk if strata are chosen too small.

(iii) What if there is interest not just in the overall total, but also in the space/time breakdown of by-catch? If the primary spatial interest is in absolute by-catch levels, then the classical rules-of-thumb will work well. But if the interest is in by-catch *rate* (e.g. in order to map the abundance of a species, rather than its catch), then the classical design may undersample some strata.

(iv) What to do when there is uncertainty about catch rates and variance-mean ratios? Obviously, no observer program would be necessary if catch rates were known perfectly and were fixed over time.

(v) What exactly is to be optimised? Designs optimal for one species, will be sub-optimal for another. For example, blue shark catch rates are highest in the south, so an optimal design for blue shark would place more observers there. But an optimal design for blue

---

[4] See W.G. Cochran "Sampling techniques", Edition 2 John Wiley and Sons, 1963.

marlin, which are mostly caught in the north at least during autumn/winter, would do precisely the opposite in those seasons.

Because of these issues, and other logistical complexities which inevitably cloud the picture, it does not make sense to think about strictly-optimal designs at the start of an observer program, if ever. Probably the worst thing statistically that can happen to a program, is to find out that a stratum expected to have very low catch (and therefore receiving very low coverage), actually turns out post-hoc to have apparently very high catches (but with high uncertainty, because of the low coverage). Rather than optimality, it is better to aim for a design that guards against too much undersampling in strata that might hold a surprise, while otherwise adhering roughly to proportional coverage.

In this study, we considered five different designs, with overall coverage ranging from about 5% to 20% (Table 7). Design A has equal numbers of sets sampled in all strata/seasons; the allocation and total coverage varies through to design E, which has highest overall coverage and highest variability in absolute numbers of samples per stratum. Designs B and E are closest to proportional coverage.

Table 7. Number of observed sets drawn per stratum-season combination, for five survey designs ranging from 5.1% coverage (600 sets) to 18.8% coverage (2200 sets).

Design (coverage)	Stratum	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec
(A) 5.1%	North	30	30	30	30
	West-centre	30	30	30	30
	Mid-centre	30	30	30	30
	East-centre	30	30	30	30
	South	30	30	30	30
(B) 5.1%	North	30	30	30	30
	West-centre	50	50	50	30
	Mid-centre	30	30	50	50
	East-centre	10	10	20	20
	South	30	30	10	10
(C) 8.5%	North	50	50	50	50
	West-centre	50	50	50	50
	Mid-centre	50	50	50	50
	East-centre	50	50	50	50
	South	50	50	50	50
(D) 13.6%	North	50	50	50	50
	West-centre	100	200	200	100
	Mid-centre	100	100	100	100
	East-centre	50	50	50	50
	South	50	50	50	50
(E) 18.8%	North	100	100	100	100
	West-centre	200	200	200	100
	Mid-centre	100	100	200	200
	East-centre	50	50	50	50
	South	100	100	50	50

Note that the table shows numbers of SETS observed, not number of TRIPS. We were unable to allocate sets to trips in the time available for this study, so we used the set as the unit of observation. [Nevertheless, all scaling-up of observed by-catch rates is on the basis of hooks, not sets.] However, in practice the actual unit of observation will be the trip. Because there is likely to be some between-trip variability (e.g. because different trips tend to concentrate on different parts of a stratum), the precisions reported here will be slightly optimistic. The extent cannot be quantified from the results presented here, but presumably will not be large for marlin or turtles, because there was no general evidence of over-dispersion. For dolphinfish and blue sharks, the effect may be partly countered by our having included unmodelled year-to-year variability into the estimated between-set variability.

Coverage levels in individual strata/seasons are not particularly high in any of the designs, except perhaps with 50 sets in the October-December in the southerns stratum. When the number of sets observed in a stratum is low, there is a high probability that zero by-catch will be observed for some species in some strata. With 50 samples per stratum, zero catch is unlikely unless the average by-catch rate is much less than about 20 animals per million hooks; however, this will be true for turtles in some strata. Very low observed by-catches in a stratum do cause some minor problems in analysis (it becomes harder to assess uncertainty of the total) but, more significantly, it becomes hard to make a design more efficient.

The estimates of likely precision for the different designs, based on 5000 simulated datasets, are shown in Table 8. Roughly, the estimated total by-catch will be within +/- twice the CV of the true by-catch 95% of the time. The 5th and 95th percentiles show in more detail the likely range of variation of the estimate as a percentage of the true value.

Table 8. Distribution of estimated total number of animals caught, using a stratified ratio estimator. Estimates obtained from 5000 simulated surveys for each of five survey designs.

Species/ group	Coverage	Statistics for estimated total count		
		5th per-centile ÷ true value (%)	95th per-centile ÷ true value (%)	C.V. (%)
Turtle	(A) 5.1%	47	162	35
	(B) 5.1%	53	152	30
	(C) 8.5%	58	147	27
	(D) 13.6%	71	131	18
	(E) 18.8%	76	124	14
Black marlin	(A) 5.1%	72	131	18
	(B) 5.1%	73	127	17
	(C) 8.5%	78	123	14
	(D) 13.6%	84	117	10
	(E) 18.8%	86	114	8
Blue marlin	(A) 5.1%	81	119	12
	(B) 5.1%	81	119	12
	(C) 8.5%	86	115	9
	(D) 13.6%	87	113	8
	(E) 18.8%	90	110	6
Dolphinfish	(A) 5.1%	82	120	11
	(B) 5.1%	84	118	10
	(C) 8.5%	86	115	9
	(D) 13.6%	90	111	7
	(E) 18.8%	92	108	5
Blue shark	(A) 5.1%	90	111	6
	(B) 5.1%	88	112	7
	(C) 8.5%	92	108	5
	(D) 13.6%	93	108	5
	(E) 18.8%	95	106	3

Discussion: overall coverage and precision

Precision will be much higher for species that are more commonly caught. Even at 5% coverage, for example, blue shark by-catch is estimated very precisely (CV ~ 7%). With a species caught as frequently as blue sharks, it is fairly difficult to go wrong (unless there is extreme clustering, which the data do not suggest). However, when only a handful of trips are observed per stratum/season (e.g. design B, with just 10 sets in some cases), it can be difficult to ensure that sampling is really representative. If not, there will be legitimate concerns about possible bias and about whether the estimated uncertainty is truly realistic.

It is more informative to consider what precision is obtainable for rarely-caught animals, such as turtles. At 5% coverage, the CV for turtles is likely to be 30-35%, depending on the design. Roughly speaking, this would make the annual estimate of turtle by-catch precise to ±70% (a 95% confidence interval). Note that this conclusion, and all the other simulation results, are based on point estimates ("best estimates") of likely actual by-catch rates. The real by-catch rate may be lower or higher than the estimates used here, both because the estimate is based on limited sampling and is thus intrinsically uncertain, and also because of differences between the Japanese and domestic fisheries (though not for turtles, where we used domestic interview data). If the true by-catch rate is lower than we have assumed, precision will also be lower.

There may of course be by-catch species which are caught even less frequently than turtles. If it became overwhelmingly important to get precise estimates for such a species, then higher coverage even than design E might be needed. For rarely-caught species, CVs will tend to be higher; thus, in order to achieve a pre-specified CV for *all* species, coverage might have to be very high. However, achieving a specific CV may not always be the appropriate objective. For some by-catch species, it may only be important to make sure that average by-catch is below a certain pre-specified *level* (e.g. a pre-defined sustainability threshold). If the upper confidence limit on by-catch turns out to be below that level, then the precise value of the by-catch is irrelevant, and improving the CV becomes unnecessary. This argument should not be over-interpreted; in practice, if the CV is very high, the upper confidence limit will automatically be much higher than the mean, making it unlikely that a "clean bill of health" can be given. Also, for many species, the sustainability threshold will be unclear. For those species, it will remain important to get reasonably precise estimates of by-catch, against the day when the impact of different by-catch levels can be predicted.

The acceptable level of precision for a particular by-catch species, depends on the level of concern for that species, and also on the time frame over which an answer is needed. In terms of assessing population impact for a long-lived species, individual years are not very important; it would be more meaningful to consider how precisely a 5- or 10-year average by-catch could be estimated. There are two answers to this.

The simple answer is that the annual CV in Table 8 will reduced by a factor of sqrt(5) ~2.2 or sqrt(10)~3.1 after 5 or 10 years (see Appendix). For turtles, the CV of the retrospective average after 5 years would be a respectable16%, even at the lowest coverage considered here. However, even 5 years is not long-term from the point of view of turtle population dynamics. To properly assess the possible long-term impact of by-catch, it is necessary to take account of general inter-annual fluctuations in animal distribution. Even if we could accurately assess the mean annual by-catch over the last 5 years, say, that period is unlikely to have been exactly representative of long-term animal distribution. Further, the space/time distribution of the fishery is also important, and can vary just as much. The complicated answer— admittedly to a more complicated question: "how precisely can long-term impact be forecast'?'— therefore cannot be given until more data on inter-annual variation is available. Even if a perfect 100% coverage program ran for just one or two years, the complexities of interannual variation would lead to uncertainty in assessing the possible long-term ecological impact of by-catch.

Another longer-term objective might be to monitor trends in by-catch *rate* rather than total, in order to monitor population status. Again, the task is rendered more complex by inter-annual shifts in distribution; if these shifts are large, it is hard to measure trends precisely *even with 100% observer coverage*. Until several years of data have been collected, so that the extent of inter-annual variability can be assessed, it would be premature to explicitly consider survey design and coverage in terms of trend estimation. Nevertheless, trend estimation could turn out to be a very valuable by-product of a long-term observer program.

Despite the complexities introduced by inter-annual variability, it is important to emphasize that the *main* uncertainty at the moment, is the total by-catch within each year. The only way to address both types of uncertainty, is through a multi-year observer program. If there is reasonable coverage early on, it becomes more feasible to improve the efficiency of the design in subsequent years.

Discussion: allocation of coverage to strata and seasons

The detailed design does make a difference to attainable precision. For turtles, where by-catch rates have (for example purposes only) been assumed constant across strata and seasons, there is a 15% gain in precision in moving from the equal-number-of-sets design A to the proportional-almost-to-fishing-effort design B. However, as above, design B is risky, because of the very low absolute numbers of trips observed in some strata. Scaling up a near-proportional design to higher overall coverage, and slightly beefing up strata/seasons that would otherwise receive very low absolute numbers of observed trips, leads to quite an efficient program (e.g. design D or E).

It is obviously possible to consider modified designs in more detail. However, at present there really are not enough data yet to try for an "optimal" design, especially since there are multiple species with different spatial distributions and hence different "optimal" designs. Also, the stratification requires further consideration: does it make logistical sense? A reasonable strategy would be to pick an overall level of coverage based on the CVs suggested by these analyses, then redefine the strata if necessary, and then select a roughly-proportional design, with some reinforcement of weakly-sampled strata. After a few years of data collection, it would be possible to consider whether efficiency gains are achievable through a changed design.

Acknowledgements

Our thanks go to Rob Campbell and Thim Skousen, for assistance with data preparation and interpretation.

Appendix: derivation of formulae

In stratum i, let the true by-catch rate be r_i, the number of observed sets e_i, the total number of sets E_i, the observed by-catch n_i, the total by-catch N_i, and the variance-to-mean ratio of by-catch between sets k_i. Then, since n_i is the sum of e_i independent identically-distributed random variables:

$$\hat{N}_i = E_i \frac{n_i}{e_i}$$

$$\Rightarrow \operatorname{var}[\hat{N}_i] = \frac{E_i^{\,2}}{e_i^{\,2}} \operatorname{var}[n_i]$$

$$E[n_i] = e_i r_i$$

$\operatorname{var}[n_i] = e_i k_i r_i$ since n_i is the sum of e_i independent random variables

$$\Rightarrow \operatorname{var}[\hat{N}_i] = \frac{E_i^{\,2} k_i r_i}{e_i}$$

$$\Rightarrow \operatorname{var}[\hat{N} = \sum_i \hat{N}_i] = \sum_i \frac{E_i^{\,2} k_i r_i}{e_i}$$

To minimize $\operatorname{var}[\hat{N}]$ subject to the constraint $\sum e_i = e$, include a Lagrange multiplier and minimize

$$S_l = \sum_i \frac{E_i^{\,2} k_i r_i}{e_i} - l\left(\sum_i e_i - e\right)$$

$$\frac{\partial S_l}{\partial e_i} = -\frac{E_i^{\,2} k_i r_i}{e_i^{\,2}} - l$$

$$\frac{\partial S_l}{\partial e_i} = 0 \Rightarrow e_i \propto E_i \sqrt{k_i r_i}$$

This ignores finite sample corrections, which would become significant if the observed sets in a stratum formed a high proportion of the total (say above 30%). Also, no adjustment is made for variable numbers of hooks per set. In practice, scaling-up is done on the basis of hooks rather than sets. However, provided sampling is representative, this is a minor refinement which does not affect the conclusions.

To examine the CV of total by-catch over a period of T years, assume that $\bar{N}$ is the true annual mean by-catch over that period, and that (consistent with the above) $\operatorname{var}[\hat{N}_t] = oN_t$ for the estimated by-catch in year t. Then

$$\text{var}[\hat{N} = \sum_t \hat{N}_t] = o \sum_t N_t = oT\bar{N}$$

$$E[\hat{N}] = T\bar{N}$$

$$CV[\hat{N}] = \sqrt{\frac{o}{T\bar{N}}}$$

$$CV[\hat{N}_t] = \sqrt{\frac{o}{N_t}}$$

$$\Rightarrow CV[\hat{N}] = \frac{1}{\sqrt{T}} \sqrt{\frac{N_t}{\bar{N}}} CV[\hat{N}_t] \approx \frac{1}{\sqrt{T}} CV[\hat{N}_t] \text{ if year } t \text{ is "typical"}$$

The estimated annual average is just $\hat{N}/T$, and has the same CV as the total $\hat{N}$.

ADDENDUM: Summary of chute trial data

An up-to-date source of reliable information on by-catch rates, is AFMA's recent chute trial (referred to here as SOP: seabird observer program). Although this is seabird-oriented, observers also record by-catch of other non-target species. Both the JOP and SOP data give unbiassed estimates of by-catch rates in the respective fisheries. Because fishing practices and locations differ between the Japanese and Australian fisheries, the JOP data is not ideal for predicting precision of an observer program in the Australian fishery, and in this respect the SOP data would in theory be preferable. However, overall coverage in the SOP is low, at around 2%; the series is short; and there has been no coverage in the northern stratum (Table A1).

Table A1. Number of sets observed in the Australian seabird east coast longline observer program from April 2001 to April 2002, by season.

Stratum	Season	Number of sets
West-centre	Jan-Mar	46
	Apr-Jun	27
	Oct-Dec	25
Mid-centre	Jan-Mar	3
	Apr-Jun	2
	Oct-Dec	15
East-centre	Oct-Dec	5
South	Jan-Mar	7
	Apr-Jun	7
	Jul-Sep	3
	Oct-Dec	7

For most stratum/season combinations, the number of SOP sets was too low to merit a formal modelling exercise (such as we did for the JOP data), but it is still instructive to compare rates (Table A2; there were also two turtles, not shown). Although sample sizes are small, SOP rates do seem consistently higher than logbook rates. Out of 24 comparisons in Table A2, SOP rates are higher 19 times and logbooks rates are higher 5 times (about 80%). The magnitude of the difference varies, and is obviously subject to some uncertainty because of limited sample sizes in the SOP, but in 14 of the 24 comparisons, the SOP rate is more than double the logbook rate.

For some species and regions, there is still an enormous difference between JOP and SOP rates, which obviously cannot be due to under-reporting and is presumably due mostly to differences in fishing practice. The principal example is blue sharks, in the Southern stratum and in April-June. The main text suggests that differences between logbook and JOP rates for blue sharks in the south might partly reflect differences in proportions reported across strata, but this cannot be the whole story as the SOP and JOP rates for blue sharks in the south are also very different. However, at least for marlin, there are other indications of differences in proportion reported across strata (subject to small sample size concerns). For both marlin species, the SOP rates in the south are *much* higher than logbook rates in the early part of the year, and the difference is substantially more than in other strata.

It is worth considering how the estimated precisions might change if SOP-like rates instead of JOP-like rates applied. In general, JOP rates are slightly higher than SOP rates for the marlins, so precision would be slightly lower. For blue sharks, some of the JOP rates seem much too high, so precision would again be lower; however, precision is likely to be so good for blue sharks that this hardly matters. For dolphinfish, SOP rates are considerably higher than JOP rates, and precision is likely to be rather better than shown in the main report. The two turtles reported in the SOP data (not shown in Table A2) are consistent with the by-catch estimates in Bache and Robins (2002), given the low coverage, and the SOP data provide no reason to change the expected precision for turtles.

Table A2. Mean by-catch rates (number per thousand hooks) for four species in the Eastern Tuna and Billfish Fishery. Bold type = Seabird observer program (SOP), April 2001 to April 2002; upright type = 2001 logbooks (LB); italics = Japanese observer program, 1991-1995; ** =small sample size in JOP (< 9 sets) and SOP (< 7 sets).

Species	Stratum	Jan-Mar			Apr-Jun			Jul-Sep			Oct-Dec		
		SOP	LB	JOP	SOP	LB	JOP	SOP	LB	JOP	SOP	LB	JOP
Black marlin	North	**	0.11	*0.14*	**	0.08	*1.28*	**	0.05	*0.19*	**	1.36	*0.69*
	West-centre	**0.32**	0.07	*0.34*	**0.09**	0.03	*0.11*	**	0.01	*0.05*	**<0.01**	<0.01	**
	Mid-centre	**	0.22	**	**	0.02	*0.12*	**	0.01	0.02	**<0.01**	0.01	*0.35*
	East-centre	**	0.02	*0.06*	**	0.01	*0.09*	**	<0.01	*0.03*	**	0.01	**
	South	**0.68**	0.01	**	**0.20**	<0.01	*<0.01*	**	<0.01	*<0.01*	**<0.01**	<0.01	**
Blue marlin	North	**	0.41	*1.20*	**	0.06	*0.26*	**	0.08	*0.08*	**	1.03	*0.32*
	West-centre	**0.25**	0.19	*1.88*	**0.09**	0.04	*0.11*	**	0.01	*0.02*	**0.05**	0.03	**
	Mid-centre	**	0.13	**	**	0.05	*0.14*	**	0.02	*0.03*	**0.08**	0.01	*0.42*
	East-centre	**	0.03	*0.02*	**	0.03	*0.08*	**	<0.01	*<0.01*	**	0.01	**
	South	**0.23**	0.01	**	**<0.01**	<0.01	*<0.01*	**	<0.01	*<0.01*	**<0.01**	<0.01	**
Blue sharks	North	**	0.45	*1.64*	**	0.38	*0.72*	**	0.53	*1.22*	**	0.48	*0.60*
	West-centre	**0.69**	0.34	*0.58*	**0.26**	0.35	*3.30*	**	0.12	*0.61*	**0.31**	0.05	**
	Mid-centre	**	0.98	**	**	0.86	*1.15*	**	0.59	*1.00*	**0.32**	0.21	*0.50*

	East-centre	**	2.80	*2.18*	**	1.82	*1.91*	**	2.44	*2.14*	**	0.74	**
	South	**0.45**	0.37	**	**0.40**	0.43	*7.13*	**	0.21	*12.41*	**0.43**	0.21	**
Dolphinfish	North	**	2.93	*0.28*	**	4.74	*1.45*	**	5.89	*1.44*	**	1.62	*1.09*
	West-centre	**8.29**	3.96	*1.30*	**1.42**	1.33	*0.28*	**	0.55	*0.20*	**6.21**	7.71	**
	Mid-centre	**	1.77	**	**	2.23	*0.27*	**	0.79	*0.19*	**1.05**	2.83	*0.46*
	East-centre	**	1.35	*0.28*	**	0.86	*0.12*	**	0.32	*0.15*	**	1.23	**
	South	**4.99**	1.44	*<0.01*	**6.26**	0.78	*<0.01*	**	0.31	*<0.01*	**0.43**	0.21	*<0.01*