SCRS/1982/027 Rev. Col.Vol.Sci.Pap. ICCAT, 18 (3): 711-718 (1983)

A PRELIMINARY CRITICAL EXAMINATION OF THE ICCAT DATA BASE

J. P. Wise
ICCAT

SUMMARY

Considerable time and effort has been devoted by national offices and the ICCAT Secretariat to assembling a data base for use in studies on Atlantic tunas. The sources of information in the data base are national offices of member and non-member countries, and statistics and sampling carried out by the Secretariat on behalf of the Commission. It is inevitable that a base collected from heterogeneous sources will contain a certain amount of inconsistencies and conflicting data. This paper is a preliminary attempt to analyze some of these inconsistencies, with an eye to their eventual elimination. The parts of the data base examined here are longline length-frequency samples as collected by national offices and the Secretariat, and catch and effort data from longliners and baitboats similarly collected.

The results show considerable problems with the data. Many of them stem from insufficient monitoring and quality control inspection of the data as close to the source as possible. The most disturbing conflicts in the data examined are found in the Gulf of Guinea baitboat fisheries.

While the inconsistencies found are cause for genuine concern, they should not be taken as grounds for despair. Further analyses may show ways to compensate for them, and improvements in collection and processing will tend to eliminate them over time.

RESUME

Les administrations nationales et le Secrétariat de l'ICCAT ont consacré une somme considérable de temps et d'efforts au rassemblement d'une base de données destinée aux études sur les thonidés de l'Atlantique. Les sources d'obtention des données sont les administrations nationales de pays membres et non membres, ainsi que le recueil de statistiques et échantillonnage effectués par le Secrétariat pour le compte de la Commission. Il est inévitable qu'une base composée d'éléments d'origines diverses contienne un certain pourcentage de manques de concordance et de données contradictoires. Le présent document représente une première tentative d'analyse de certaines de ces imperfections, ayant pour but leur élimination éventuelle. Les sections de la base soumises à cet examen sont les échantillons palangriers de fréquence de taille rassemblés par les administrations nationales et le Secrétariat, et les données de capture et effort rassemblées de la même façon sur les palangriers et canneurs.

Les résultats signalent l'existence de graves problèmes concernant les données. Nombre d'entre eux découlent d'un certain manque de contrôle et d'inspection de la qualité des données le plus près possible du point d'origine. Les divergences les plus aigües observées concernent les pêcheries de canneurs du golfe de Guinée.

Bien que préoccupantes, ces manques de concordance ne doivent pas nous rebuter. Une analyse plus poussée devrait nous permettre de découvrir la façon d'y remédier, et toute amélioration des formules de recueil et de traitement tendra à les éliminer avec le temps.

RESUMEN

Las administraciones nacionales de pesca y la Secretaría de ICCAT, han dedicado mucho tiempo y esfuerzo a establecer una base de datos que pudiera utilizarse en estudios sobre túnidos del Atlántico. Las fuentes de información consultadas son las administraciones nacionales de diversos países, miembros o no miembros de ICCAT, y recopilación estadística y muestreo llevado a cabo por la Secretaría en nombre de la Comisión. Sin embargo, resulta inevitable el que una base de datos recopilados de diversas fuentes contenga una cierta proporción de faltas de concordancia y datos contradictorios. Este documento intenta un análisis preliminar de algunos de éstos, teniendo en cuenta su posible eliminación. Las partes de la base de datos que se examinan son las muestras de frecuencia de talla de capturas de palangre, y datos de captura y esfuerzo de palangre y barcos de cebo, recopilados por las administraciones pesqueras de los diversos países y la Secretaría.

Los resultados muestran la existencia de importantes problemas en cuanto a los datos. Muchos de ellos tienen su origen en una vigilancia insuficiente y escaso control de la calidad de los datos, que debería realizarse tan próximo como fuese posible de la fuente que los proporciona. De entre los examinados, los datos más contradictorios correspondieron a las pesquerías de cebo del Golfo de Guinea.

Si bien las faltas de concordancia de los datos son causa de preocupación no deberán, sin embargo, ser causa de desaliento. Análisis posteriores proporcionarían los medios necesarios para compensarlas, y la mejora en la recopilación y proceso de datos tenderán a eliminarlas con el tiempo.

Port and at-sea length frequency sampling of longline catches

The earliest length-frequency data available in the ICCAT data base for longline-caught yellowfin, bigeye and albacore are for 1965, for the Japanese and USSR fisheries. It was recognized early on that sampling of distant-water longline fleets presented special problems, since the catches are usually transshipped in foreign ports.

Although attempts had been made previously by the Secretariat to collect these data, the Sub-Committee on Statistics recommended, and the Commission agreed at the Third Regular meeting in 1973 that the Secretariat should make special efforts at certain key locations to collect these length-frequency data. Sampling has been carried out on a regular basis at transshipment ports since late 1974. Some data for the last few months of 1974 are presented in the ICCAT Statistical Series, No. 1 (1976), but the systematic inclusion of Secretariat-collected data in the data base begins with 1975.

Sampling of length frequencies of longline-caught yellowfin, bigeye and albacore has been carried out by the Secretariat at:

Abidjan, Ivory Coast
Cape Town, South Africa
Cumaná, Venezuela
Las Palmas, Las Palmas, Spain
Santa Cruz, Tenerife, Spain
St. Maarten, Neth. Antilles

Sampling at Cumaná was begun in 1980. Sampling at Abidjan was terminated after 1977 because the longliners stopped transshipping from there. Other data on transshipped longline-caught fish have been provided by Brazil (1977) and by Bermuda starting in 1980.

Size sampling on longliners at sea by Korean fishermen started in 1976; data for 1976 and 1977 are available only as weight frequencies. Sampling of length frequencies began in 1978; the 1978 and 1979 data are available only in five-centimeter groups. Sampling of length frequencies at sea by Taiwanese fishermen began in 1979.

The ICCAT data file covering sampling of length frequencies of yellowfin, bigeye and albacore caught by Korean and Taiwanese fishermen using longlines for the years 1975-80 consists of some 3,200 records. Each record contains information on the species sampled, the nationality of the vessel from which the sample was taken, the area in which the fish were caught, and the distribution of the lengths sampled.

The weight samples taken by Korean fishermen in 1976 and 1977 are not included in the data base. Samples are recorded in the data base in two centimeter intervals except for the data collected by Korean fishermen in 1978 and 1979, which are in five centimeter intervals.

Quality of longline length frequency samples

All of the length frequency samples from longliners in the ICCAT data base for 1978-80 were combined by quarter of the year, ICCAT sampling area, and by Korean or Taiwanese government or by ICCAT port sampling, and inspected by eye.

While it is difficult to develop strictly objective criteria for evaluation of the quality of length frequency samples, the experienced eye can usually detect samples which appear implausible. Some of the samples in the ICCAT data base taken at sea did not appear particularly credible. There were cases in which there were only one or two fish in a quarterly sample, or a total of seven fish of neatly-spaced different lengths, for instance.

Port samples taken at St. Maarten showed peculiar problems. Nearly all of the samples taken in 1979 and 1980 were "square," that is, they more nearly resembled stratified samples than random samples. I decided to omit all of the St. Maarten samples in comparing port sampling with at-sea sampling rather than evaluating them on an individual basis. All other samples were included in the following analyses.

Distribution of yellowfin longline samples

There were in theory 220 cells for the yellowfin samples-- 10 areas (plus unknown) by 12 quarters for Korean and 8 quarters for Taiwanese catches. Only 21 cells contained data, limited principally to Korean data, to areas LLYF 14 and LLYF 15, and to the third and fourth quarters (Table 5).

Comparable yellowfin sampling may be summarized:

Korean catch

	LLYF 14	LLYF 15
1978 - Quarters	1,2,3,4	2,3
1979 - Quarters	3,4	1,3,4
1980 - Quarters	1,2,3,4	2,3,4
(+ LLYF 16, 1979, Quarter 1)		

Taiwanese catch

1980 LLYF 11, Quarter 3 and LLYF 16, Quarter 3

Distribution of bigeye longline samples

There were in theory 260 cells for comparable bigeye catch samples-- 12 areas (plus unknown) by 12 quarters for Korean and 8 for Taiwanese catches. Only 25 cells have data, limited principally to Korean catches, to areas LLBE 44 and LLBE 46, and to the first, third and fourth quarters (Table 6)

Comparable bigeye sampling may be summarized:

Korean catch

	LLBE 44	LLBE 46
1978 - Quarters	1,2,3,4	2,4
1979 - Quarters	1,3,4	1,4
1980 - Quarters	1,2,3,4	1,2,3,4

(+ LLBE 45, 2 Quarters and LLBE 42 and LLBE 50, 1 Quarter each)

Taiwanese catch

1980 LLBE 42, Quarter 1 and LLBE 47, Quarter 3

Distribution of albacore longline samples

There were in theory 100 cells for comparable albacore catch samples-- 4 areas (plus unknown) by 8 quarters for Taiwanese and 12 for Korean catches. In fact, only 35 cells have data, limited principally to areas AL 33 and AL 34 and to the first, third and fourth quarters (Table 7).

Comparable albacore sampling may be summarized:

Korean catch

	AL 31	AL 32	AL 33	AL 34
1978 - Quarters	1,4	1,3,4	1,2,3,4	-
1979 - Quarters	-	-	1,3,4	-
1980 - Quarters	-	3	1,4	-

Taiwanese catch

	AL 31	AL 32	AL 33	AL 34
1979 - Quarters	1	1	1,3,4	1,2,3,4
1980 - Quarters	1	2,3	1,2,3	1,2,3,4

(+ unknown area, 1979, Quarter 1)

Comparison of port sampling and at-sea sampling

The statistical test chosen for comparison of the samples was the Kolmogorov-Smirnov (K-S) two-sample test. The K-S two-sample test is a non-parametric test of whether two samples have been drawn from the same population or from populations with the same distribution. Cumulative frequency distributions are made from each of the samples, using the same intervals, and compared to find the largest observed deviation. The significance of the largest deviation is evaluated by:

$$\chi^2 = \frac{4 D^2 (n_1 n_2)}{(n_1 + n_2)}$$

where D is the largest deviation and n_1 and n_2 are the numbers in the two samples.

See Siegel (1956) for a more detailed description. This test was used by Herrick (1982) to compare length frequency samples taken in Ghana and in Puerto Rico.

As mentioned above, the 1978 and 1979 length frequency samples taken by Korean fishermen are reported in five-centimeter groups, whereas all the others are repoted in two-centimeter groups. The Korean data were adjusted before calculation by reducing the odd-centimeter groups by one centimeter-- that is, 75 cm adjusted to 74 cm, 85 cm to 84 cm, etc.

While the K-S test per se considers only the magnitude of the largest deviation of the two distributions, the number of deviations in each direction was calculated and is shown in Tables 5-7* This calculation was initiated when it became obvious that the correspondences between the length frequencies were in general not very good. It was suspected that systematic differences between at-sea and port sampling measurements might be part of the problem. There does not seem, however, to be any systematic tendency for more negative or positive differences in either the Taiwanese or Korean data. (The number of zero differences is largely an artifact of the method.)

Results of the Kolmogorov-Smirnov test

On first examination, the results of the K-S test applied to compare at-sea versus port sampling (Tables 5 - 7)* are very disappointing. Values of chi-square run from a little over 1 to about 800, with more than three quarters of them over 7.8, the value that is acceptable as indicating that the samples are drawn from the same or comparable populations at the p= .05 leval. This is not much better than would be expected with randomly drawn length-frequency samples.

A second look, however, reveals that nearly three quarters of the acceptable values are those for Korean catches in 1979-80, and half of them are for Korean catches in 1980. This way of looking at the results suggests that there has been

*Tables 5-7 (29 pages) are omitted in the interest of economy. They may may be consulted in the meeting document, and are available from the Secretariat.

a considerable improvement in the Korean samples with time, and that a certain amount of time must be allowed for a reasonable at-sea sampling scheme to be developed.

It is by no means a reason for complacency, however. Considerable effort must be put forth by those responsible for at-sea and port sampling to improve the quality of the data collected. This objective can be achieved only through constant monitoring of the sampling process and quality control inspection of the data as soon as possible after they have been collected.

Catch and effort data in the ICCAT data base

Longline catch and effort data

The ICCAT data base for 1978-80 was chosen to compare longline catch and effort. About 20 port sampling records of catches taken east of 20° E. were eliminated as being outside the Atlantic.

Catches expressed in numbers of fish were converted to weights by use of mean weights for 5x5 degree square areas and quarters, with substitutions made as necessary.

Catches in tons per 10,000 hooks were calculated for yellowfin, bigeye and albacore for each 5x5 degree square and quarter for Korean and Taiwanese government reported data and for Korean and Taiwanese data from ICCAT port sampling (recorded as Kor.Pan and Chi.ICAT). Correlations and regressions were run for catch per unit effort for each of the three species for government data vs. ICCAT port sampling data, and for Korean government vs. Taiwanese government data (Table 1).

Nearly all of the correlations between longline catches per unit effort shown in Table 1 are significant at the p= .o5 level or better. When samples with effort of less than 10,000 hooks per quarter/square are eliminated (about 20 percent of all the data), eight of the nine correlations are improved and all correlations are significant at the p = .01 level.

Although the correlations are statistically significant, they are not as high as might have been expected. Also, there are some puzzles in the differences between the correlations by species. The Taiwanese longline fishery has in recent years been directed principally at albacore, with smaller catches of bigeye and yellowfin catches almost at the incidental level. The correlations in Table 2 for the China vs. Chi.ICAT sampling are highest for yellowfin, intermediate for bigeye, and lowest for albacore. These results appear to be just the opposite of what might have been expected.

A similar phenomenon may be seen in the Korea vs. Kor.Pan comparison in Table 2. While albacore is the species least sought after (least caught) by Korean longline fishermen, the correlation for albacore is by a considerable amount the highest for the three species.

And in the comparison between Korean and Taiwanese catch per unit effort data based on government-reported catch and effort data, the correlation for yellowfin in Table 2 is clearly the highest.

The reasons for these results are by no means clear at this time.

Baitboat catch and effort data

The ICCAT data base for 1978-80 was used to compare baitboat catch and effort. All records between 10° S. and the African coast and 10° W. and the African coast were included. More than 90 percent of the records in this area were found north of the Equator and east of 5° W., that is in the Gulf of Guinea baitboat fishery.

Catches in tons per day were calculated for yellowfin, skipjack and bigeye for each 1x1 degree square and quarter for Ghanaian, Japanese and Korean government data, and for ICCAT-sponsored port sampling of Japanese (Jap.ICAT) and Korean or Panamanian baitboats (Kor.Pan). Correlations and regressions were run for catch per day (CPUE) for each of the three species for Korean and Japanese government data vs. ICCAT sampling data, for Japanese vs. Ghanaian government data, and for Japanese vs. Korean government data (Table 3).

Only a little over half of the correlations between baitboat CPUE figures shown in Table 3 are significant at the p = .05 level or better. Three of the four correlations for skipjack, the principal species, are significant at p = .05 or better.

Correlations were calculated a second time to test whether these low values were due primarily to using 1x1 square/quarter cells with low fishing effort by eliminating cells with less than 5 days effort (a little more than half of the total data). Nine of the twelve correlations improved, but the number significant at the p = .05 level or better decreased. All of the correlations for skipjack CPUE improved, and three of the four became significant at the p = .01 level.

Although most of the correlations are statistically significant, the correlation coefficients are much lower than might have been expected. The highest correlation coefficient for skipjack is only .399 for Korea vs. Kor.Pan. The coefficient of determination (r^2) in this case is only about 16 percent. That is, only 16 percent of the variation in Kor.Pan CPUE can be predicted from Korean CPUE, or vice versa.

The A values, representing the intersections of the regression lines on the Y axis, are disturbing. It is difficult to believe that, for instance, Korean baitboats catch on the average 5 tons more per day of skipjack than do Japanese bait boats. The B values, the slopes of the regressions, are also hard to believe. One would expect to find values appraoching 1 in many of these cases, but the average B for skipjack is less than .5 and the highest is only about .7.

Further investigations are obviously necessary to determine if possible the reasons for the unexpected values found in the regressions and correlations. One approach that is obviously indicated is a study of the variances involved. Another is partitioning of the data in various ways to see if the effects of time and space can be measured.

Until such investigations are completed, researchers would be prudent to exercise extreme caution in combining the data from which these CPUEs are derived for purposes such as estimating catches of undersized yellowfin and bigeye in Gulf of Guinea baitboat catches.

REFERENCES

Herrick, S.
1982

Size and species compositions of Atlantic tunas from imports landed in Puerto Rico during 1980. ICCAT Collective Volume of Scientific Papers, Vol. XVII, pp. 53-67 (SCRS/81/52).

Siegel, S.
1956

Non-parametric statistics for the behavioral sciences. McGraw-Hill, New York: 312 p.

Table 1 - Comparisons of 1978-80 longline catch per unit effort data, 5x5 degree square by quarter.

	S	R	A	B	N
China vs. Chi.ICAT	YFT	.454**	.154	.462	279
	BET	.241**	.412	.174	279
	ALB	.104	4.157	.174	279
Korea vs. Kor.Pan	YFT	.440**	1.737	.244	297
	BET	.521**	1.179	.746	297
	ALB	.740**	.266	.764	297
Korea vs. China	YFT	.715**	-.001	.552	94
	BET	.255*	.408	.457	91
	ALB	.517**	2.519	.508	94

Key:

S = Species

R = Coefficient of correlation

A = Intersection of regression line on Y axis

B = Slope of regression line

N = Number of pairs of observations

* = Significant at p = .05

** = Significant at p = .01

Table 2 - Comparisons of 1978-80 longline catch per unit effort data, 5x5 degree squares by quarter, squares with less than 10,000 hooks removed.

	S	R	A	B	N
China vs. Chi.ICAT	YFT	.524**	[illegible]	[illegible]	220
	BET	.364**	[illegible]	[illegible]	220
	ALB	.265**	3.064	[illegible]	221
Korea vs. Kor.Pan	YFT	.490**	1.241	.475	265
	BET	.545**	1.076	[illegible]	265
	ALB	.750**	[illegible]	[illegible]	265
Korea vs. China	YFT	.765**	.012	.655	61
	BET	.782**	.087	.507	61
	ALB	.646**	3.141	.652	61

Key:

S = Species

R = Coefficient of correlation

A = Intersection of regression line on Y axis

B = Slope of regression line

N = Number of pairs of observations

* = Significant at p = .05

** = Significant at p = .01

Table 3 - Comparisons of 1978-80 baitboat catch per unit effort data, Gulf of Guinea, 1x1 degree square by quarter.

	S	R	A	B	N
Korea vs. Kor.Pan	YFT	.240*	.656	.094	119
	SKJ	.312*	3.505	.252	119
	BET	.112	.922	-.074	119
Japan vs. Jap.ICAT	YFT	.419**	.334	.554	110
	SKJ	.747**	1.511	.441	110
	BET	.000	.479	-.011	110
Japan vs. Ghana	YFT	.113	.967	-.276	36
	SKJ	.289	1.799	.350	36
	BET	.149	.273	.369	36
Japan vs. Korea	YFT	.145**	.280	.055	399
	SKJ	.181**	3.553	.203	399
	BET	.100*	.261	.067	399

Key:

S = Species

R = Coefficient of correlation

A = Intersection of regression line on Y axis

B = Slope of regression line

N = Number of pairs of observations

* = Significant at p = .05

** = Significant at p = .01

Table 4 - Comparisons of 1978-80 baitboat catch per unit effort data, Gulf of Guinea, 1x1 degree square by quarter, squares with less than 5 days fishing removed.

	S	R	A	B	N
Korea vs. Kor.Pan	YFT	.019	.754	.013	57
	SKJ	.399**	3.103	.710	57
	BET	.191	.926	-.313	57
Japan vs. Jap.ICAT	YFT	.432**	.297	.763	83
	SKJ	.394**	.953	.675	83
	BET	.038	.451	.031	83
Japan vs. Ghana	YFT	.188	1.059	-.175	[illegible]
	SKJ	.366	1.107	.511	[illegible]
	BET	.075	.350	.211	[illegible]
Japan vs. Korea	YFT	.085	.336	.044	[illegible]
	SKJ	.298**	5.580	.102	[illegible]
	BET	.340**	.207	.147	[illegible]

Key:

S = Species

R = Coefficient of correlation

A = Intersection of regression line on Y axis

B = Slope of regression line

N = Number of pairs of observations

* = Significant at p = .05

** = Significant at p = .01