A statistical method for evaluating differences between age-length keys with application to Georges Bank haddock, *Melanogrammus aeglefinus*

Daniel B. Hayes
Woods Hole Laboratory, Northeast Fisheries Science Center
National Marine Fisheries Service, NOAA
166 Water Street, Woods Hole, MA 02543-1097

Age-length keys are a critical component of many methods used to estimate catch at age (Southward, 1976; Kimura, 1977; Gavaris and Gavaris, 1983; Quinn et al., 1983; Lai, 1987; Martin and Cook, 1990). When formulating age-length keys for use in stock assessment, the appropriate time interval for data collection and aggregation is an important consideration. Potentially, an age-length key derived from data taken at a single time during the year could be used to compute catch at age for the entire year. This practice runs the risk of producing inaccurate estimates of catch at age, however. Finer time intervals for age-length keys are typically used (e.g., quarterly or monthly) but at a cost of increased sampling requirements. Generally, intensive sampling over brief time intervals provides the most accurate representation of age at length in the population (Kimura, 1977; Westrheim and Ricker, 1978). In practice, however, such a strategy is expensive to implement because of the labor-intensive nature of collecting and processing large numbers of fish for age data. For the efficient and cost-effective use of sampling resources, it is desirable to sample and age the fewest fish, giving adequate precision while avoiding biased results. The data needed can be considerably reduced if appropriate time intervals for age-length keys can be defined.

A further problem facing fishery scientists is the need for evaluation of samples collected with different gears. Although fishing gears may select for fish size, the question here is whether the gears select for fish with different proportions of age at length. For example, the Northeast Fisheries Science Center (NEFSC) of the National Marine Fisheries Service samples most of the commercially important fish stocks off the northeastern United States with bottom trawls during the spring and autumn. The fish collected in these surveys have a substantially different size composition than the commercial catch because the standard survey trawl has a small mesh codend liner. However, it is unknown whether fish of a given length caught in the research survey trawl differ in age composition from fish of the same length harvested commercially. If the age at length of fish captured is the same for these gears, then a pooled key, including data from research surveys, will increase precision by increasing sample size but will not bias catch at age estimates. However, if age-length keys differ, then pooling can introduce bias into estimates of catch at age (Westrheim and Ricker, 1978).

The primary objective of this study is to present a method for determining whether two age-length keys differ statistically. This method is then applied to Georges Bank haddock to determine what time intervals should be used for age-length keys for this stock and to determine if research trawl-survey age data can be combined with age data from commercially harvested haddock. The benefit (in terms of increased precision) of pooling age-length keys, where appropriate, is also determined.

Methods

Biological sampling

Scales were collected from commercially landed haddock in the NEFSC port sampling program. Because sampling requests to this program are filled quarterly, age-length keys and the age structure of landings were computed on a quarterly basis as in previous stock assessments (Clark et al., 1982; Gavaris and Van Eeckhaute, 1990). Length at age by month was averaged over several years (1980 to 1988) to determine growth on a monthly basis. Haddock scales were also collected in the NEFSC bottom trawl research survey. The NEFSC conducts this survey during the spring and autumn each year, providing a fishery independent source of age structures. These surveys are conducted from Cape Hatteras, North Carolina, to Nova Scotia, Canada, with the use of a stratified random sampling design. Further details concerning sampling procedures in the NEFSC survey are outlined in Grosslein (1969) and Azarovitz (1981).

Scales were removed from the lateral line region below the second dorsal fin. After removal, scales were dried and then impressed on a laminated plastic slide. Scales were viewed at approximately 40× magnification to determine age. Ages were determined following cri-

Manuscript accepted 2 April 1993.
Fishery Bulletin: 91:550–557 (1993).

teria outlined by Penttila (1988) and Jensen and Wise (1962). By convention, a 1 January birthdate was used.

Statistical analysis

Age-length keys are commonly formed first by obtaining a matrix of numbers at age by length interval (Table 1), and then by converting this to a matrix of proportion at age for each length interval. For statistical tests between age-length keys, however, I used the matrix of numbers at age by length. Age-length keys were compared by making tests of significance separately for each length interval present in both keys where the sample size was greater than six for each age-length key. Fisher's exact test (Siegel, 1956) was used in these comparisons. Because of the large number of tests needed to compare age-length keys, experimentwise error was compensated for by adjusting the significance level for the individual tests. The significance level for n individual tests (α^*) needed to maintain a desired experimentwise error (α_{exp}) was determined by the following formula derived from Sokal and Rohlf (1981):

$$\alpha^* = 1 - e^{\frac{\ln(1-\alpha_{exp})}{n}}$$

This method can be quite conservative if the power of the individual tests is limited because of small sample sizes (Sokal and Rohlf, 1981). Owing to the conservative nature of this method, two subjective criteria for evaluating test results were used.

The first criterion was the number of tests exceeding the nominal significance level (i.e., α=0.05) in the set of comparisons of interest. A second criterion for evaluating the significance of individual tests was the pattern of nominally significant tests. If significant differences were observed between quarters for a given length interval, adjacent length intervals were also expected to show significant differences. Thus, tests suggesting nominally significant differences between quarters for a single length interval adjacent to length intervals where tests indicate no difference were suspected of occurring by random chance.

Age-length keys from the first and second, second and third, and third and fourth quarters were compared to determine which quarters could be combined into a single age-length key. Tests were conducted by comparing keys within a year because combining age-length keys between years can introduce bias in the resulting key (Westrheim and Ricker, 1978). Additional tests were conducted comparing NEFSC spring survey age-length keys with combined first and second quarter commercial age-length keys.

Table 1

Number sampled at age by 2-cm length groups from the commercial catch of Georges Bank haddock *Melanogrammus aeglefinus*, January to March 1983.

	Age								
Length (cm)	2	3	4	5	6	7	8	9+	Total
42	1	2	0	0	0	0	0	0	3
44	0	6	0	0	0	0	0	0	6
46	0	4	2	0	0	0	0	0	6
48	0	4	1	1	0	0	0	0	6
50	0	3	5	4	0	0	0	0	12
52	0	2	5	6	1	0	0	0	14
54	0	1	8	14	0	0	0	0	23
56	0	0	9	11	0	0	0	0	20
58	0	0	5	12	1	0	0	0	18
60	0	0	3	8	2	0	6	0	19
62	0	0	3	8	4	0	6	0	21
64	0	0	0	12	3	1	2	1	19
66	0	0	0	9	3	0	7	0	19
68	0	0	0	8	0	0	6	1	15
70	0	0	0	4	1	4	8	0	17
72	0	0	0	0	1	3	9	0	13
74	0	0	0	2	0	2	10	1	15
76	0	0	0	0	1	1	9	0	11
78	0	0	0	0	1	1	6	1	9
80	0	0	0	0	0	1	7	1	9
82	0	0	0	0	0	0	7	1	8
84	0	0	0	0	0	0	1	1	2
86	0	0	0	0	0	0	1	0	1

Estimation of catch at age and variance

As a check on the potential for introducing bias by combining age-length keys, estimates of the age composition of the commercial catch for each year from 1982 to 1988 were made by using age-length keys combined in four different ways. First, age-length keys were constructed for each quarter by using only fish sampled commercially during that quarter. These catch-at-age estimates served as a basis for comparison since they correspond to the level of temporal aggregation (quarterly) that has commonly been used in Georges Bank haddock assessments (Clark et al., 1982; Gavaris and Van Eeckhaute, 1990). The second set of catch-at-age estimates was formed by combining commercial age data from the first half of the year into a single age-length key that was then applied to the first and second quarter length-frequency distributions. The third set of catch-at-age estimates was based on quarterly age-length keys but included data from haddock sampled in the commercial catch and from the NEFSC bottom trawl surveys. The final set of age-length keys contained data from haddock caught both commercially and in trawl surveys but combined all data for the first half of the year.

I treated length-frequency samples as simple random samples from the entire stock area in the computation of the variance of catch at age. Following Gavaris and Gavaris (1983), the proportion of catch at age for unpooled samples was estimated as

$$\hat{A}_{iq} = \sum_j L_{jq} P_{ijq}$$

where,

$\hat{A}_{iq}$ = Estimated proportion at age i in quarter q
L_{jq} = Proportion of total individuals at length j in quarter q
P_{ijq} = Proportion of age i individuals at length j in quarter q.

For samples where age-length keys were pooled across quarters, the proportion of catch at age was estimated as

$$\hat{A}_{iq} = \sum_j L^*_{jq} P^*_{ijq}$$

where P^*_{ijq} is the proportion of age i individuals at length j for the pooled quarters, and L^*_{jq} is the pooled proportion at length j in the pooled quarters. These proportions were calculated as

$$P^*_{ijq} = \frac{n_{ij1} + n_{ij2}}{n_{j1} + n_{j2}},$$

$$L^*_{jq} = \frac{(L_{j1})N_1 + (L_{j2})N_2}{N_1 + N_2},$$

where,

n_{ij1} = # of age i fish at length j in first time period
n_{ij2} = # of age i fish at length j in second time period
n_{j1} = # of fish at length j in first time period
n_{j2} = # of fish at length j in second time period
N_1 = total # of fish landed in first time period
N_2 = total # of fish landed in second time period.

This method of pooling age-length keys treats each observation as a random sample from a single population. No weighting factors were applied when age-length keys were pooled. A pooled proportion of total individuals at length was computed as a stratified random sample where the weighting for each quarter was the total number of individuals landed during that period. This allows for the possibility that the length composition of landings differed between quarters.

Estimates of the variance of proportion at age for unpooled data were computed following Gavaris and Gavaris (1983):

$$\mathrm{Var}(\hat{A}_{iq}) = \sum_j \frac{L^2_{jq} P_{ijq}(1-P_{ijq})}{n_{jq}-1} + \frac{L_{jq}(P_{ijq} - \hat{A}_{iq})^2}{n_q}$$

Where n_q is the number of fish aged in quarter q.

Estimates of the variance of proportion at age when age-length keys were pooled were computed with the above formulae, except P^*_{ijq} was substituted for P_{ijq}. Catch at age and variance for each quarter were computed following Gavaris and Gavaris (1983). Because these computations do not depend on whether pooled or unpooled age-length keys are used, the formulae are not repeated.

Results and discussion

Seasonal growth and comparison of age-length keys

Most of the annual growth of Georges Bank haddock takes place during the third quarter, from June through September (Fig. 1). From this pattern of annual growth, age-length keys would not be significantly different between the first and second quarters, but would differ significantly betweens the second and third quarters, and the third and fourth quarters. Accordingly, tests were conducted between these pairs of quarters to determine if age-length keys could be pooled across any adjacent quarters. Summary statistics of these tests are presented in Tables 2 through 4. Although little growth takes place between the fourth quarter and the first quarter in the following year, landings data are usually finalized on an annual basis; pooling between years is therefore of limited use.

Comparisons between the first and second quarters yielded 9 of 94 tests exceeding the 0.05 level (Table 2). None of these tests exceeded the adjusted significance level (α^*=0.00054) needed to maintain an error rate of 0.05 for this set of comparisons. The number of results exceeding 0.05 is not substantially greater than the number of significant results that would be expected based on random chance; further, these differences occurred sporadically among the length classes (Table 2). Thus, based on these tests, first and second quarter age-length keys within each year can be treated as samples drawn from the same population and can be pooled.

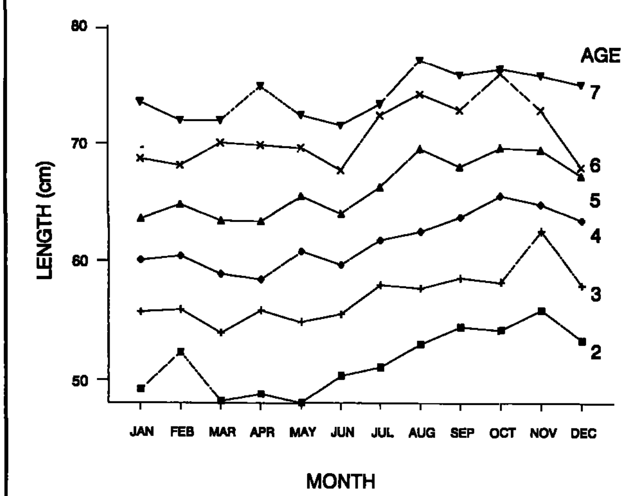

Figure 1
Monthly mean length at age of Georges Bank haddock *Melanogrammus aeglefinus* in the commercial catch, 1980–1988.

Significant differences were observed between second and third quarter age-length keys; 45 of 95 comparisons exceeded the 0.05 level, and 19 tests of these tests had probabilities of less than 0.00054 (α^*) (Table 3). Tests between the third and fourth were not as definitive; 21 of 85 comparisons exceeded the 0.05 level, but no tests exceeded 0.00060 (α^*). Several tests, however, were close to α^* (Table 4). The large number of tests exceeding 0.05 and the occurrence of contiguous blocks of tests with low probability (Table 4) indicate that age-length keys differ between the third and fourth quarters, but the power to detect these differences may have been limited by small sample sizes. In general, the number of fish aged from the fourth quarter was lower than those from other quarters as indicated by the smaller number of length intervals with sample sizes greater than six. From these results, age-length keys should not be combined across the second and third quarters or third and fourth quarters for haddock. The consequence of pooling age-length keys across quarters when statistical tests indicate significant differences would be to bias catch-at-age estimates (Westrheim and Ricker, 1978). In the case of Georges Bank haddock, the effect of pooling third quarter age-length keys with second quarter keys (for example) would be to bias catch-at-age estimates for the second quarter towards younger fish. This occurs because growth during the third quarter tends to increase the mean length at age, and conversely, fish of a given length tend to be younger during the third quarter than during the second quarter.

Comparison of pooled first and second quarter age-length keys with spring NEFSC bottom trawl surveys keys resulted in 7 of 71 tests having probabilities less than 0.05 (Table 5), but none exceeding the α^* level of 0.00072. Although more "significant" results were obtained than would be expected by random chance, no consistent pattern among these differences was apparent (Table 5). Further, several of these differences occurred in the 60 to 66 cm size classes where the sample size from the bottom trawl survey was generally small (<10 fish), while sample sizes from the commercial catch were often large (>50 fish). Examination of the contingency tables for these size classes indicated that the difference in proportion at age between commercial and survey age-length keys was small and was often due to a broader representation of age classes in the commercial data. A broader representation in the commercial samples would be expected, however, given the

Table 2
Results of Fisher's exact test comparing first and second quarter age-length keys from Georges Bank haddock *Melanogrammus aeglefinus*, 1982–1988. Entries are probability of difference observed occurring by random chance. Dashes indicate comparisons where the sample size was less than seven within one of the two quarters being compared.

Length (cm)	Year						
	1982	1983	1984	1985	1986	1987	1988
42	—	—	—	—	—	—	—
44	0.307	0.676	—	—	—	—	—
46	0.782	0.117	0.559	1.000	—	—	—
48	1.000	0.804	0.892	0.661	—	—	1.000
50	0.888	0.033*	0.653	0.289	—	—	—
52	0.786	0.819	0.022*	0.363	—	—	0.724
54	0.176	0.434	0.143	0.008*	—	0.018*	0.026*
56	0.051	0.284	0.207	0.451	—	1.000	0.067
58	0.562	0.749	0.002*	0.074	0.753	0.434	0.744
60	0.177	0.002*	0.950	0.889	0.432	0.303	0.870
62	0.153	0.123	0.163	—	1.000	0.799	0.406
64	0.136	0.372	0.223	—	0.644	0.389	0.464
66	0.300	0.694	0.849	—	0.230	0.950	0.811
68	0.352	0.497	0.373	—	0.418	0.461	0.704
70	0.182	0.032*	0.676	—	1.000	1.000	0.799
72	0.097	0.119	0.584	—	0.143	1.000	0.439
74	0.730	0.808	0.036*	—	0.128	—	—
76	0.488	0.589	0.096	—	—	—	—
78	—	0.852	0.249	—	—	—	—
80	—	1.000	0.161	—	—	—	—
82	—	0.466	—	—	—	—	—

$^*P<0.05$

Table 3
Results of Fisher's exact test comparing second and third quarter age-length keys from Georges Bank haddock *Melanogrammus aeglefinus*, 1982–1988. Entries are probability of difference observed occurring by random chance. Dashes indicate comparisons where the sample size was less than seven within one of the two quarters being compared.

Length (cm)	Year 1982	1983	1984	1985	1986	1987	1988
42	0.219	0.170	—	—	—	—	—
44	<0.001**	0.075	—	0.592	—	—	—
46	<0.001**	0.017*	—	0.020*	—	—	—
48	<0.001**	0.110	0.620	0.013*	—	<0.001**	0.832
50	<0.001**	0.026*	0.747	0.108	—	<0.001**	0.773
52	<0.001**	<0.001**	1.000	0.115	1.000	0.003*	<0.001**
54	<0.001**	<0.001**	0.688	0.011*	0.582	0.012*	<0.001**
56	0.042*	<0.001**	1.000	0.321	0.429	0.488	<0.001**
58	0.046*	<0.001**	0.146	0.053	0.115	0.335	<0.001**
60	0.038*	<0.001**	0.863	—	<0.001**	0.075	<0.001**
62	0.011*	0.011*	0.106	—	0.002*	0.055	<0.001**
64	0.523	0.015*	0.662	—	0.004*	0.010*	—
66	0.246	0.032*	0.354	—	0.292	0.004*	—
68	0.068	0.016*	0.192	—	0.046*	0.122	—
70	0.544	0.002*	0.700	—	0.011*	0.046*	—
72	0.379	0.022*	0.738	—	0.590	0.058	—
74	0.185	0.544	—	—	0.020*	0.755	—
76	0.686	0.407	—	—	0.569	—	—
78	—	0.427	—	—	—	—	—
80	—	0.206	—	—	—	—	—
82	—	0.537	—	—	—	—	—

*P<0.05
**P<0.00054=α^*

Table 4
Results of Fisher's exact test comparing third and fourth quarter age-length keys from Georges Bank haddock *Melanogrammus aeglefinus*, 1982–1988. Entries are probability of difference observed occurring by random chance. Dashes indicate comparisons where the sample size was less than seven within one of the two quarters being compared.

Length (cm)	Year 1982	1983	1984	1985	1986	1987	1988
42	0.286	1.000	—	—	—	—	—
44	0.298	0.014*	—	—	—	—	—
46	0.158	0.003*	—	—	—	1.000	—
48	0.051	0.001*	—	—	—	0.192	0.124
50	<0.001*	0.123	0.123	—	—	0.053	1.000
52	0.159	0.315	0.212	—	—	0.024*	0.832
54	0.146	0.049*	0.720	—	—	0.754	0.211
56	0.205	0.765	1.000	—	—	0.050*	0.791
58	0.143	0.009*	0.154	—	0.477	0.675	0.016*
60	0.292	0.174	0.939	—	0.794	0.540	0.014*
62	0.028*	0.032*	0.758	—	0.272	0.244	<0.001*
64	0.265	0.042*	0.132	—	0.085	0.022*	—
66	0.014*	0.034*	0.079	—	0.785	<0.001*	—
68	0.484	0.937	0.759	—	0.068	0.107	—
70	0.011*	0.038*	0.218	—	0.207	0.046*	—
72	1.000	0.813	0.550	—	1.000	0.408	—
74	0.756	0.162	—	—	0.054	0.922	—
76	0.282	0.612	—	—	0.713	—	—
78	0.355	0.499	—	—	1.000	—	—
80	0.765	—	—	—	—	—	—
82	—	—	—	—	—	—	—

*P<0.05

larger sample size available from the commercial fishery for fish in these length classes. Based on the lack of pattern among the significance tests and the lack of tests exceeding α^*, it appears that survey gear does not more consistently select for fish of a different age at a given length than do commercial fishing gears. Accordingly data on age at length obtained from fish collected in the research surveys can be combined with data from fish sampled commercially.

Catch at age and precision

Catch-at-age estimates obtained with the various age-length keys showed no indication of systematic bias (Fig. 2). This was expected, as the results of statistical tests indicated no significant difference between the age-length keys that were pooled. The precision of catch-at-age estimates, however, did show a trend among the different levels of aggregation (Fig. 3). Combining first and second quarter age-length keys derived solely from commercially caught haddock increased precision for all age groups, particularly for older age classes that typically had small sample sizes within a single quarter (e.g., the 82 and 84 cm length classes; Table 1). The inclusion of survey age data had a relatively smaller effect on the precision of estimates for most age classes except for age-2 fish (Fig. 3). This occurred because the sample size of small haddock was generally less in the commercial samples than in the research survey samples.

Statistical considerations

Fisher's exact test tests the hypothesis that the proportion at age within each length class is no different among keys than would be obtained by random chance.

Formally, the hypothesis tested

for length class j is

Ho: $p_{ij1} = p_{ij2}$ for all age classes i from source 1 and source 2

Ha: $p_{ij1} \neq p_{ij2}$ for all age classes i from source 1 and source 2.

Consider as an example haddock (*Melanogrammus aeglefinus*) in the 54-cm size class sampled from the commercial catch from the first and second quarters of 1983. Numbers at age from each of these samples are

	Age									
	0	1	2	3	4	5	6	7	8	Total
Quarter 1	0	0	0	1	8	14	0	0	0	23
Quarter 2	0	0	0	7	10	24	1	0	0	42

The question asked is whether the two samples are likely to be drawn from the same population or whether they differ sufficiently to indicate that the sampled populations are different. Assuming fixed marginal totals, an appropriate test of this hypothesis is Fisher's exact test (Siegel, 1956). Previously this test was impractical for contingency tables greater than 2 × 2 because of the amount of computational power required by the algorithms available for its solution. Recent improvements (Pagano and Halvorsen, 1981; Mehta and Patel, 1983; implemented in Version 6 of SAS [SAS Institute, 1990]) allow problems of this size to be readily solved.

Alternative tests exist in the chi-square test of homogeneity (Hennemuth, 1965) and the G^2 test (Bishop et al., 1975). These tests have the advantage that age-length keys can be compared in their entirety. Some studies indicate that these tests may also have greater power than Fisher's exact test (D'Agostino et al., 1988; Storer and Kim, 1990), but others have disputed the validity of this these assertions (Little, 1989). Also, the chi-square and G^2 tests are often viewed as inappropriate when some of the expected values are less than 5 (e.g., Sokal and Rohlf, 1981; Haberman, 1988); a situation that commonly occurs in comparisons of age-length keys. Although grouping data across age or length classes is a way of increasing the expected values for each cell in the contingency table, such a procedure results in a loss of statistical power (Cochran, 1952).

With Fisher's exact test, there are no restrictions on the expected values for any cell within the contingency table (Siegel, 1956). In practice, however, each source (i.e., time period) should contain at least six observations because smaller sample sizes do not have sufficient power to resolve even major discrepancies between sources (Bennett and Hsu, 1960).

In summary, Fisher's exact test provides a means of testing differences between age-length keys derived from different sources or from different time periods. When age-length keys are pooled across time periods or sources that do not differ significantly, the resulting estimates of catch at age are more precise and importantly do not appear to be biased by the pooling procedure. For Georges Bank haddock, age-length keys from commercially harvested haddock from the first and second quarters can be combined, as well as keys from haddock sampled in the NEFSC research survey. In the future, allocation of sampling effort should consider the benefits of pooling age-length keys.

Acknowledgments

I thank S. Clark, J. Forrester, S. Gavaris, and F. Serchuk for their constructive comments on earlier drafts of this manuscript. Thanks are also expressed to J. Brodziak for his helpful discussions on this subject.

Table 5

Results of Fisher's exact test comparing age-length keys for Georges Bank haddock *Melanogrammus aeglefinus* derived from fish sampled in NEFSC bottom trawl surveys to commercially sampled fish from the first and second quarters, 1982–1988. Entries are probability of difference observed occurring by random chance. Dashes indicate comparisons where the sample size was less than seven within one of the two sources being compared.

Length (cm)	Year						
	1982	1983	1984	1985	1986	1987	1988
38	0.209	—	—	—	—	—	—
40	0.214	—	—	—	—	—	—
42	—	—	0.054	—	0.350	—	—
44	—	—	0.151	—	1.000	—	—
46	0.087	—	—	0.699	0.232	—	1.000
48	0.038*	1.000	—	—	0.243	—	1.000
50	0.388	1.000	0.344	—	0.095	—	—
52	0.151	0.175	—	—	0.220	0.241	—
54	0.107	0.479	0.547	0.123	—	0.360	—
56	0.002*	0.014*	0.146	0.093	0.033*	0.465	1.000
58	0.139	0.127	0.057	0.549	—	—	—
60	0.019*	0.839	1.000	0.102	0.273	—	—
62	0.268	0.470	0.554	0.165	0.179	0.079	—
64	—	0.526	0.163	0.391	0.641	0.302	—
66	0.821	0.302	0.026*	0.768	0.003*	—	—
68	—	0.191	0.128	0.725	—	—	—
70	0.625	0.139	0.423	0.745	—	—	—
72	0.777	0.682	—	—	—	—	—
74	—	—	—	0.066	—	—	—
76	0.927	0.194	—	—	—	—	—

*$P<0.05$

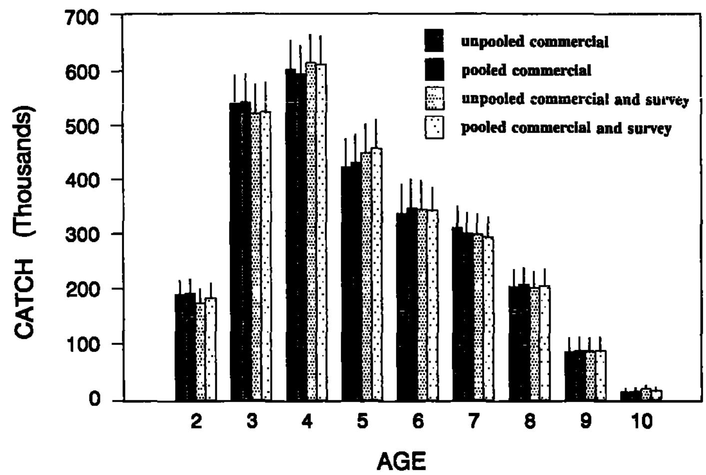

Figure 2
Mean catch at age of Georges Bank haddock *Melanogrammus aeglefinus*, 1982–1988, based on different levels of aggregation in the age-length key. Unpooled commercial refers to quarterly age-length keys derived solely from commercial samples. Pooled commercial refers to age-length keys derived solely from commercial samples, but with quarter-1 and quarter-2 data pooled into a single age-length key. Pooled and unpooled commercial and survey refer to the same levels of temporal aggregation as above but include age-at-length data obtained from fish collected in the NEFSC bottom trawl surveys. Vertical lines indicate 2 SE of the mean catch at age.

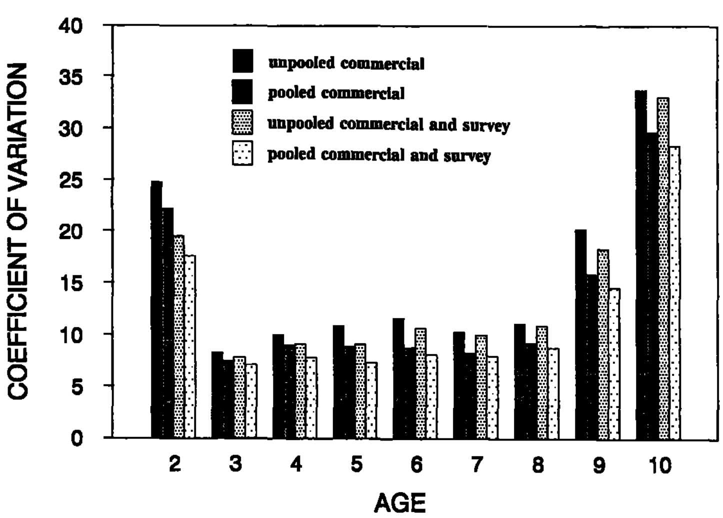

Figure 3
Mean coefficient of variation of estimates of catch at age for Georges Bank haddock *Melanogrammus aeglefinus*, 1982–1988, based on different levels of aggregation in the age-length key. Symbol definitions as in Figure 2.

Literature cited

Azarovitz, T. R.
1981. A brief historical review of the Woods Hole Laboratory trawl survey time series. *In* W. G. Doubleday and D. Rivard (eds.), p. 62–67. Canadian Spec. Publ. Fish. Aquat. Sci. 58.

Bennett, B. M., and P. Hsu.
1960. On the power function of the exact test for the 2×2 contingency table. Biometrika 47:393–397.

Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland.
1975. Discrete multivariate analysis: theory and practice. MIT Press, Cambridge, MA.

Clark, S. H., W. J. Overholtz, and R. C. Hennemuth.
1982. Review and assessment of the Georges Bank and Gulf of Maine haddock fishery. J. Northwest Atl. Fish. Sci. 3:1–27.

Cochran, W. G.
1952. The χ^2 test of goodness of fit. Annals of Mathematical Statistics 23:315–345.

D'Agostino, R. B., W. Chase, and A. Belanger.
1988. The appropriateness of some common procedures for testing the equality of two independent binomial populations. Am. Statistician 42:198–202.

Gavaris, S., and C. A. Gavaris.
1983. Estimation of catch at age and its variance for groundfish stocks in the Newfoundland region, p. 178–182. *In* W. G. Doubleday and D. Rivard (eds.), Sampling commercial catches of marine fish and invertebrates. Canadian Spec. Publ. Fish. Aquat. Sci. 66.

Gavaris, S., and L. Van Eeckhaute.
1990. Assessment of haddock on Eastern Georges Bank. Canadian Atl. Fish. Sci. Advisory Comm. Res. Doc. 90/86, 37 p.

Grosslein, M. D.
1969. Groundfish survey program of BCF Woods Hole. Commer. Fish. Rev. 31:22–30.

Haberman, S. J.
1988. A warning on the use of Chi-squared statistics with frequency tables with small expected cell counts. J. Am. Statistical Assoc. 83:555–560.

Hennemuth, R. C.
1965. Homogeneity of age-length frequencies among months and quarters of the year for haddock caught on Georges Bank, 1962. Int. Comm. Northwest Atl. Fish. Res. Bull. 2:76–76.

Jensen, A. C., and J. P. Wise.
1962. Determining age of young haddock from their scales. U.S. Fish Wildl. Serv. Fisheries Bull. 195(61):439–450.

Kimura, D. K.
1977. Statistical assessment of the age-length key. J. Fish. Res. Board Canada 34:317–324.

Lai, H.
1987. Optimum allocation for estimating age composition using age-length key. Fish. Bull. 85:179–183.

Little, R., J. A.
1989. Testing the equality of two independent binomial proportions. Am. Statistician 43:283–288.

Martin, I., and R. M. Cook.
1990. Combined analysis of length and age-at-length data. J. Cons. Cons. Int. Explor. Mer 46:178–186.

Mehta, C. R., and N. R. Patel.
1983. A network algorithm for performing Fisher's exact test in r × c contingency tables. J. Am. Statistical Assoc. 78:427–434.

Pagano, M, and K. T. Halvorsen.
1981. An algorithm for finding the exact significance levels of r × c contingency tables. J. Am. Statistical Assoc. 76:931–934.

Penttila, J.
1988. Haddock *Melanogrammus aeglefinus*. *In* J. Penttila and L. M. Dery (eds.), Age determination methods for Northwest Atlantic species, p. 23–29. U.S. Dep. Commer., NOAA Technical Report NMFS 72.

Quinn, T. J. II, E. A. Best, L. Bijsterveld, and I. R. McGregor.
1983. Sampling Pacific halibut (*Hippoglossus stenolepis*) landings for age composition: history, evaluation and estimation. Int. Pacific Halibut Comm. Sci. Rep. 68, Seattle, WA.

SAS Institute Incorporated.
1990. SAS/STAT users guide, Version 6, fourth edition. SAS Inst., Carey, NC, 1686 p.

Siegel, S.
1956. Nonparametric statistics for the behavioral sciences. McGraw-Hill, NY.

Sokal, R. R., and F. J. Rohlf.
1981. Biometry, second edition. W. H. Freeman, NY.

Southward, G. M.
1976. Sampling landings of halibut for age composition. Int. Pac. Halibut Comm. Sci. Rep. 8, Seattle, WA.

Storer, B. E., and C. Kim.
1990. Exact properties of some exact test statistics for comparing two binomial proportions. J. Am. Statistical Assoc. 85:146–155.

Westrheim, S. J., and W. E. Ricker.
1978. Bias in using an age-length key to estimate age-frequency distributions. J. Fish. Res. Board Canada 35:184–189.