ISSN: 0579-3920

INTERNATIONAL PACIFIC HALIBUT COMMISSION

ESTABLISHED BY A CONVENTION BETWEEN
CANADA AND THE UNITED STATES OF AMERICA

Technical Report No. 42

A review of IPHC catch sampling for age and size composition from 1935 through 1999, including estimates for the years 1963-1990

by

William G. Clark, Bernard A. Vienneau, Calvin L. Blood,
and Joan E. Forsberg

SEATTLE, WASHINGTON
2000

The International Pacific Halibut Commission has three publications: Annual Reports (U.S. 0074-7238), Scientific Reports, and Technical Reports (U.S. ISSN 0579-3920). Until 1969, only one series was published (U.S. ISSN 0074-7426). The numbering of the original series has been continued with the Scientific Reports.

INTERNATIONAL PACIFIC HALIBUT COMMISSION
P.O. BOX 95009
SEATTLE, WASHINGTON 98145-2009, U.S.A.

A review of IPHC catch sampling for age and size composition from 1935 through 1999, including estimates for the years 1963-1990

William G. Clark, Bernard A. Vienneau, Calvin L. Blood, and Joan E. Forsberg

Contents

A review of IPHC catch sampling for age and size composition from 1935 through 1999, including estimates for the years 1963-1990

William G. Clark, Bernard A. Vienneau, Calvin L. Blood, and Joan E. Forsberg

Introduction

IPHC staff have been collecting size and age data on fish sampled from the commercial landings since 1935. For each sampled landing, the data include information on the vessel and trip (e.g., area and depth fished, gear used) and on the size and age of the sampled fish, numbering anywhere from a few to a few hundred specimens. Since 1964, all these data have been entered into computer files. They have always been called "market sample data" within IPHC. They are used to estimate the age composition of the landings, the average weight at age, and the total number of fish landed by age, which are essential ingredients of the annual stock assessment.

Generally speaking, the raw market sample data are not representative of the coastwide age composition of the commercial landings because geographic coverage was incomplete in some years and uneven in most years; and in many years the fish that were aged are a size-stratified subsample, and not necessarily a proportional subsample, of the fish that were measured. To calculate accurate estimates of age composition, therefore, it is essential to weight the raw age readings correctly, taking into account the catches by area and the sampling/subsampling procedures in use at the time. This review provides a detailed chronology of the sampling coverage, sampling strategies, and subsampling methods used in each year and region so that users can handle the raw data properly.

In most years the fork lengths of a random sample of fish were measured, but from 1963 through 1990 only the otoliths of a random sample were collected, and the lengths of those fish were estimated from one or another relationship between otolith size and fish size. Unfortunately the relationship between otolith size and fish size changed when the growth rates of halibut decreased dramatically during the 1980s (Clark 1992). Estimating the size composition of the catch in those years is therefore complicated. The second part of this paper reports a set of estimates for the years 1963-1990.

Chronology of sampling objectives and methods

IPHC staff began sampling commercial landings for catch composition in 1935. Over the years since then, there have been numerous changes in the objectives of the program, the extent of coverage, the kind of data collected (i.e., lengths or otoliths or both), the sampling strategy, and the method of estimating the age composition of the catch from sample data. The purpose of this paper is to provide a history of the sampling program that can be used as a companion reference to the raw sample data stored in the IPHC database for any of the years 1933-1999. This history draws heavily on earlier accounts by Hardman and Southward (1965), Southward (1976), Quinn et al. (1983), Blood (1989), and Gilroy et al. (1995).

Some features of the halibut fishery

For those not familiar with the halibut fishery, a short description of how the fish are caught, dressed at sea, and landed will be useful. The commercial halibut fishery is conducted almost entirely with longline gear, meaning a long weighted groundline with hooks attached (or ganged) at intervals of 3 to 25 ft. When fish are brought aboard a vessel at sea, they are normally eviscerated at once and then stored in the hold in ice or refrigerated seawater. To unload the fish at the end of the trip, a cargo sling is lowered into the hold and fish are pitched into the sling, which when full is hoisted up and the fish dumped onto a large table. Very large fish do not fit easily into a sling and are hoisted up singly or in small groups by straps attached to the tail. A typical sling holds about a thousand pounds. Many halibut vessels are small and land only a few thousand pounds, but larger vessels may bring in up to 100,000 lb and spend hours “under the hoist” unloading.

After being dumped onto the table, the fish are beheaded. In earlier years this was always done by members of the vessel crew wielding machetes, but nowadays it is usually done by a fish plant employee operating a hydraulic guillotine. The headed fish are graded by size and weighed, and the vessel operator is paid according to the headed (or “net”) weight of the landing.

Until 1940 there was no minimum size limit in the halibut fishery, but buyers preferred fish over five pounds net weight because they were less susceptible to damage and spoilage. In 1940 a 5 lb minimum size limit was adopted, and in 1944 it was qualified by requiring that fish be at least 26 inches long, which is the length corresponding to an average net weight of 5 lb. In 1973 the size limit was raised to 32 inches, with no corresponding minimum weight (Skud 1977). All IPHC regulations and data on body length refer to fork length (from the tip of the snout to the fork in the tail).

1935-1941: Monitoring indicator grounds in Seattle

Some pilot work was done in 1933 and 1934, but 1935 is regarded as the first year in which catch composition data were collected in earnest and in quantity. In the early years all sampling was done in Seattle, and the primary aim was to sample landings from two “indicator” grounds: the Goose Island grounds in Queen Charlotte Sound on the British Columbia coast, and Portlock and Albatross Banks off Kodiak Island in Alaska (Fig. 1). Some samples were taken from trips to other areas (especially the Yakutat region at about 140ºW and the region west of Kodiak), but the bulk of the data came from the indicator grounds. Quinn et al. (1983) describe the operations in those days:

> The original sampling plan called for obtaining a large number of length measurements and a smaller subsample of otoliths for age determination. Fish for measurement and otolith removal were selected by the “grab method” where as many fish as possible were grabbed, or partitioned off, as the fish were unloaded [from the sling onto the heading table]. The grab technique assumed that the sampler chose fish randomly, although in practice the potential for personal bias was considerable.
>
> ...Otoliths were collected in multiples of 70 which was the number of compartments in a box used to store the otoliths in sequence. Fish lengths were recorded sequentially to match the otolith collection. In most instances 140 otoliths were collected, although larger or smaller samples were occasionally taken.

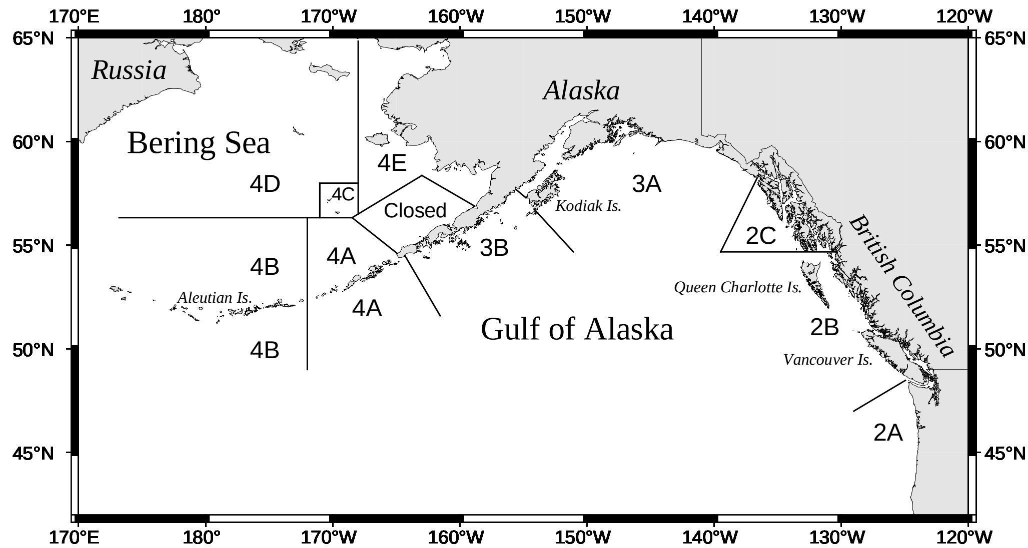


Figure 1. IPHC regulatory areas.

Southward (1965) explains that the fish in the otolith sample were not chosen randomly but were "selected as far as possible according to the range of sizes represented in the total random [length] sample. This stratification was aided by the recorder who kept a continuous tally of measured-only and measured-and-otolithed fish to assure that the frequency in each five-centimeter class of the former was adequately represented by age material from the latter."

At the time, IPHC estimated the age composition of every sampled landing. That is, the age subsample from each landing was used to key out the sample length composition of that landing. The age composition of the total landings from a ground or other defined area was then estimated by somehow combining the estimated age compositions of the sampled landings. The procedure is nowhere specified, but Southward (1976) was not enthusiastic about it:

> Age frequencies from vessels fishing the same grounds or subsections of the regulatory areas were combined by adding the individual age frequencies. A percentage age distribution was then obtained from the combined frequencies. The number of age *k* fish in the total catch was estimated by multiplying the "average" percentage of age *k* fish by the number of fish in the total catch. The latter was estimated from the overall average weight of the combined samples and the total catch. The concern for the primary sampling unit aspect of individual landings originally expressed in IPHC's sampling for age frequency data was ignored in the combining procedure.

Regardless of the propriety of the combining procedure used at the time, the basic sampling data are still useful for estimating the age composition of the landings in those years, at least for the chosen indicator grounds. It is true that the landings to be sampled were chosen arbitrarily or haphazardly; none of the accounts mentions any systematic or probabilistic method for choosing the landings. The grab sample of lengths is also suspect, because "the potential for bias was great, the randomness was questionable, and IPHC could not be sure that the procedure was standardized from crew to crew" (Southward 1976). Despite both these concerns, the data probably represent the landings reasonably well if weighted and expanded correctly.

1942-1948: Retrenchment during the war and postwar years

Sampling was severely curtailed during the war years and remained at a low level through 1948. The series of samples from the Goose Island grounds was maintained, but samples from all other areas were scant or nil. As in previous years, all sampling was done in Seattle following the procedures described above. This period is a major gap in the data series for most areas.

1949-1962: Resumption and extension of sampling

Sampling effort returned to near prewar levels in 1949 and increased thereafter. There was a gradual extension of fishing effort to grounds farther north, a shift of landings away from Seattle to northern ports, and a gradual change in IPHC sampling objectives from monitoring the indicator grounds to obtaining a set of samples that collectively represented the regional and seasonal distribution of the entire commercial landings. As Southward (1976) relates:

> By the late 1940s, grounds in [northern] Hecate Strait had become more productive. The trend toward progressively shorter seasons caused a shift in the distribution of fishing effort, which resulted in a change in the distribution of landings. The shortcomings of the two-indicator-ground design became apparent. Additional indicator grounds were established in Hecate Strait in 1949. Grounds in southeastern Alaska were added in 1958. Age data from these indicator grounds were combined and analyzed in monthly periods, as well as for the total season. In recent years, landings from other grounds have been sampled and these have been included in the combinations representing the various sections of the coast.

Ian McGregor (IPHC, pers. comm.) relates that in the mid-1950s the samplers began their work day by copying logbooks and collected samples thereafter, choosing landings from indicator grounds if any were available and sampling landings from other grounds if not.

To maintain sampling coverage when landings increased in the northern ports, IPHC sampled in Prince Rupert as well as Seattle beginning in 1949, and in Vancouver beginning in 1951. Petersburg, Alaska was added in 1958, and other Alaska ports were added in later years to maintain adequate coverage of increasing landings in Alaska. On a few occasions samplers were put aboard commercial vessels to obtain data on fish caught in the Bering Sea.

For the first time in 1958, the Commission's annual report refers to the age composition data as being representative of the commercial landings as a whole rather than just the indicator grounds, and this seems to be a fair characterization. From 1958 on, therefore, it is reasonable to regard the sample data as having been drawn from the commercial landings as a whole. The data from earlier years are affected by the emphasis on indicator grounds, although samples were nonetheless taken from other areas and the emphasis on indicator grounds declined during the 1950s.

Throughout this period the sampling procedure remained the same. The aim was to collect a large length sample and an age subsample of 140 or so otoliths well distributed across the range of the length sample so that the age composition of each sampled landing could be estimated. Like the earlier data, these samples are a little suspect because of the haphazard choice of landings to be sampled, the collection of the length sample by the "grab" procedure, and in some years the emphasis on indicator grounds. Nevertheless, the data are certainly usable if properly weighted and expanded.

1963-1972: Otolith-only sampling begins

The process of measuring lengths and collecting a subsample of otoliths required a sampling crew of four people: one moving fish, one measuring fish, one cutting otoliths, and one recording data. Clark (1992) describes the development of a one-person sampling method:

> Southward (1962) estimated a linear (log-log) relationship between otolith radius and body length that allowed the Commission to calculate a length corresponding to each sampled otolith. It was then possible for one person to sample a landing, cutting a large number of otoliths but not having to measure lengths or record any specimen data. Back at the office, an estimated length distribution was calculated from the otolith radii, and otoliths from each length interval were chosen for aging to construct a key that was applied to the estimated length distribution to estimate the age composition. Net weights of the fish in the sample were calculated from the estimated lengths using the Commission's length-weight relationship.

Hardman and Southward (1965) describe the field operation:

> When the fish are dumped on the table and before they are beheaded, the sampler draws 6 to 10 fish at random and opens the left auditory capsule of each... After all capsules have been opened the otoliths are removed and placed in a plastic container, usually strapped to the wrist. One person, working throughout the unloading of a vessel, can secure from 250 to 500 or more otoliths depending upon the average size of the fish, the size of the fare and the rapidity with which the vessel is unloaded.

Normally about 200 otoliths were collected from a landing and, as before, the aim was to estimate the age composition of that landing. This was done by aging a subsample of about 100 otoliths, typically up to six from each 5-cm length interval, and keying out the estimated length composition (based on otolith radii). The only change in the program, therefore, was to use otolith radius rather than body length as the primary measure of fish size.

Beginning in 1963 (and continuing through 1990), the sample data do not include measured length. Instead, the size measurement recorded is otolith radius from 1963 through 1967. Otolith length was adopted in place of otolith radius as the size measurement during the years 1968-1977, and otolith weight in 1978-1990. (In 1991 the Commission staff resumed body length measurements.)

1973-1986: Sling sampling

Concern over the arbitrariness of the "grab" method of choosing the size sample led the Commission to switch to whole-sling samples in the early 1970s. This involved selecting a systematic sample of trips and then selecting a sample of slings within each trip. From 1973 to 1979, the rule was to sample every tenth trip landing 1000-5000 pounds and every third trip landing over 5000 pounds. These trip sampling rates were chosen so as to approximate previous trip sampling rates. The trip sampling rate was lower for small trips because small vessels are under the hoist only a short time and it is simply not possible to sample a higher proportion of them consistently. Within a trip, the aim in 1973-1979 was to sample slings until a fixed sample size of 200 was obtained. Beginning in 1980, a systematic sampling of slings continued throughout the unloading. Quinn et al. (1983) recount the evolution of sling sampling procedures:

As an alternative to the grab sample, the “sling sample” technique was developed (Southward 1976), which utilizes the clustering effect and identifiable character of the sling and requires that every fish in the chosen sling be used. Southward (1976) tested the representativeness of the sling sample and grab sample techniques with respect to a vessel’s catch using data from 14 commercial landings plus 2 research cruises. Although the test did not show pronounced differences between sling and grab samples, the potential for bias was evident in the grab sample. Sling sampling was adopted in 1973 as the standard within-vessel sampling technique. Under this system, slings are selected for sampling at a specified frequency and all fish in the selected sling[s] are sampled.

Southward (1976) determined that 200 fish from the vessel would be an adequate sample if taken randomly using the sling sample method. However, this often resulted in the entire 200 fish being collected from the first part of the delivery.

A system of choosing slings was established in 1978, which distributed the sample over the entire load of fish. Up to 200 fish from a vessel were sampled as follows: the number of slings of fish to be sampled from each trip varied with the trip size and the expected average size of the fish. For vessels with trips originating in Regulatory Area 2, the following schedule was used to determine the frequency of slings to sample:

Trip size (pounds)	**Frequency of slings**
Under 6000	Every sling—all fish
6000-12,000	Every sling to 200 fish
12,000-18,000	Every other sling to 20
18,000+	Every third sling to 200

For vessels with trips from Regulatory Area 3, the schedule was:

Trip size (pounds)	**Frequency of slings**
Under 10,000	Every sling—all fish
10,000-20,000	Every sling to 200 fish
20,000-30,000	Every other sling to 200 fish
30,000+	Every third sling to 200 fish

The average size of fish in Area 2 is generally smaller than in Area 3, and hence a given sling contains more fish in Area 2. In both areas, sampling continued through the sling with the 200th fish in the sample or until the trip was completely unloaded.

In 1979, the sampling goal remained 200 fish but frequency of sampled slings was altered so that sampling continued until nearer the end of unloading. The size of the landing determined the frequency of sampled slings. This reduced the possibility of the sample not representing the trip in cases where the larger or smaller fish might have been unloaded first.

In our evaluation of fish selection, we uncovered some problems with this system. The proportion of fish sampled increased as trip size decreased and resulted in all the fish being sampled from the smaller trips. From larger trips some samples still came from the first part of the trip unloaded, which may not have represented the total trip. The major problem was that when the samples were combined the smaller trips contributed proportionally more fish measurements than their landed weight in pounds justified.

A common characteristic of the sampling programs prior to 1980 was the emphasis on obtaining a representative sample from each individual landing chosen for sampling. In the course of our evaluation, the emphasis shifted to obtaining a representative set of samples for a month-region stratum, which would be properly weighted when pooled together. The former goal of obtaining 200 fish per sample [trip] was dropped in favor of a sampling strategy which sampled trips proportionally to trip size. This involved adjusting the vessel and sling sampling rates for each trip size category within each stratum to obtain the identical proportion of catch sampled:

Trip size (pounds)	Vessel sampling rate	Sling sampling rate	Proportion of catch sampled
Under 1000	0	0	0
1000-4999	1/9	1/2	1/18
5000-14,999	1/3	1/6	1/18
15,000-39,999	1/3	1/6	1/18
40,000+	1/3	1/6	1/18

The overall sampling rate is one-eighteenth (5.6%) of the fish from trips over 1,000 pounds. The actual sampling rate, however, is about 3% of the fish in the total catch because some vessels unload at ports without samplers and trips of less than 1,000 pounds are not sampled.

Area 4 (Bering Sea and Aleutians) is treated as a special case, because total landings are small. The overall sampling rate is set at 1/3 to obtain adequate data, with the vessel sampling rate equal to 1/1 and the sling sampling rate equal to 1/3.

The change to sling sampling in 1973 improved the randomness of the sample. The change to systematic sling sampling in 1980 improved the representativeness of the sample while reducing the overall sample size. The practice of sampling at a uniform rate of 1/18 in all ports also had the advantage that the data could be combined straightforwardly to estimate the age composition of the landings from any combination of statistical areas or regions. In short, systematic sling sampling was a nearly ideal sampling procedure.

It broke down in 1985. The problem was that during the early 1980s the number of boats in the fleet increased, the length of open seasons decreased dramatically, and landings were concentrated into a few short periods of intense activity. Systematic sling sampling required that samplers remain with a chosen vessel throughout its unloading, and that a third of all landings over 5000 pounds be sampled. By 1985 the pace of unloading after a halibut opening made it impossible for the samplers to achieve the target vessel sampling rate.

The first remedy attempted, in 1986, was to halve the target vessel sampling rates in all areas except 2C. For 2C trips the sampling procedure in 1986 was unchanged. For trips in Areas 2B, 3A, and 3B, the schedule was:

Trip size (pounds)	Vessel sampling rate	Sling sampling rate	Proportion of catch in sample
Under 1000	0	0	0
1000-4999	1/18	1/2	1/36
5000+	1/6	1/6	1/36

In Area 2A the samplers were instructed to sample every available sling, and in Area 4 to sample every third sling from every available vessel. The expedient of lowering the target vessel sampling rates was not a satisfactory solution, because it resulted in a large part of the sample coming from only a few vessels in some ports. As a result, systematic sling sampling was discontinued after 1986, as described below.

Because of the difference in sampling rates among regulatory areas, the 1986 data cannot be combined straightforwardly to estimate the age composition of catches that overlap regulatory area boundaries (e.g., the total Alaska catch). That is also true of later years, when the practice was to set equal sample size targets for all regulatory areas, resulting in lower sampling rates in areas with higher quotas. In fact, it is only the few years of uniform and systematic sling sampling 1980-1985 that permit a simple recombination of sample data. For all other years, it is necessary to weight or partition the data to account for regional differences in sampling rates.

The method of estimating age compositions from sample data also changed during this period. In earlier years the practice was to age a subsample of the otoliths from each landing and use that to key out the size distribution of that landing and thereby obtain an estimate of the age composition of that landing, which was combined with the estimated age compositions of other landings to obtain an overall estimate for a time-area stratum, typically a region-month. Southward (1976) found that a more precise estimate could be obtained by first combining all the sample size distributions from a stratum and aging a subsample to key out the overall size distribution and thereby obtain an estimate of the overall age composition directly. In 1978 he also changed the subsampling strategy from fixed to proportional; that is, the number of otoliths to be aged from each length interval was made proportional to the number of otoliths in that length interval. The total size of the age subsample was 300 for each time-area stratum.

As noted above, Quinn et al. (1983) identified a bias in favor of small trips in the earlier sling sampling procedure and modified it so that the sample data could be used directly and correctly as it was in fact being used, to estimate the overall age composition of the landings in a stratum. He also did detailed studies of the age subsample size needed to achieve satisfactory precision, and different methods of estimating overall age compositions from combinations of primary strata. He concluded that the data from primary strata should be pooled for estimation purposes to obtain a minimum of 250 aged otoliths entering the calculation of each estimate, and that the target should be 600 aged otoliths per stratum.

Recall that at the beginning of this period the size measurement recorded in the data is otolith length; in 1978 that changes to otolith weight for the remainder of the period.

1987-1989: Spot sampling and port quotas

As explained above, by 1986 the systematic sling sampling procedure was producing a sample drawn from a relatively small number of vessels. In 1987, IPHC adopted a “spot” sampling strategy intended to distribute the sample over more landings and the entire landing period. The instructions to port samplers in the 1987 *Port Sampler Handbook* put it this way:

> Sampling procedures for the collection of otoliths from the commercial landings for age composition are changed for 1987. The old sampling routine is no longer suitable for short season openings. Acquiring representative samples has become increasingly difficult due to the intensity of the landings and lack of adequate sampling sites. The new sampling procedure will result in fewer otoliths collected but more otoliths read for age composition.

> The new procedure is based on “spot'' sampling at a set time and/or location. Each port leader will be assigned a quota of otoliths to obtain for his/her port. The goal is to spread the sampling over the most active landing period... If the number of otoliths assigned to your port is 500 and landings are estimated to occur over a five day period, then you will want to cut 100 otoliths a day. If your samplers work 10 hours per day, you will want to take 10 otoliths per hour. Oversampling is preferable to undersampling...
>
> It is the port leader’s responsibility to determine how best to take the sample. If possible, try to insert samplers in the processing line before the head is removed. If you are unable to take samples from the processing line, sample from slings or totes. If the slings or totes are especially large, you may want to sample every other hour to avoid gross oversampling. Avoid subsampling slings or totes. Do not under any circumstances size select your sample or grab sample.

Blood (1989) provides further details on the new system:

> In 1987, the staff modified the sampling program in four ways, following a major review which is reported in the 1988 *Stock Assessment Document* [Smith 1988]. First, it was determined by a cost-benefit analysis the number of otoliths collected in the field could be reduced from nearly 30,000 to 20,000. The number of otoliths aged increased from 12,000 to 16,000. Five thousand otoliths each were collected from the following regions: 2A-2B, 2C, 3A,3B-4. In 1988, subsampling of otoliths [i.e., aging only a size-stratified subsample] was eliminated and 16,000 otoliths were collected and read from all the regions [4000 from each of 2A-2B, 2C, 3A, and 3B] for the age composition. Second, each port was assigned a quota of otoliths for each fishing period based on prior landing history. The objective was to collect small samples from a large number of boats. In this way, the sampling program was tailored to the landings in each port to proportionally sample the 60-mile statistical areas. In other words if 10 percent of the landings come from a certain statistical area, then we should obtain 10 percent of the otolith sample from that area... Third, since it had become impossible to sample slings at some plants, other non-biased methods of “sling sampling” were developed. The preferred method was to intercept the fish between unloading and the heading operation and take a set number of fish that would be equivalent in number to a sling. This was called “line sampling.” Some plants unloaded slings directly into totes and trucked them into the plant for final processing. Alternatively, totes were selected randomly and sampled in the same manner as a sling. Standard sling sampling was still acceptable. Fourth, samplers would move from plant to plant, sampling whichever boat they encountered. A single sling was taken as a sample and it was assumed size-selection bias would not be a factor since the boats would be in various stages of unloading.

In practice, most samples were whole slings. A roving sampling crew would arrive at a hoist where an unloading was in progress and take the otoliths from all the fish in the next sling dumped on the table. This was a good procedure in that it preserved the randomness of sling sampling while distributing the sample over more landings. In effect, the samplers collected all the otoliths in a random sample of slings, which could be expected to represent the landings well.

The spot sampling system was used in 1987, 1988, and the first part of 1989. In the 1987 data, each sling sampled was treated as a distinct sample, even though a single vessel could be sampled more than once and in many cases was. Thus in the 1987 data, and only the 1987 data, there are multiple samples from the same vessel, with the total trip weight recorded for every sample. This poses no problems for estimating the age composition of the landings, but does require extra effort in calculating the total weight of trips sampled. In all other years, all the otoliths from a single landing are contained in a single sample.

1987 was also the last year in which the entire sample of otoliths was weighed and then only a stratified subsample was aged. In 1988 and later years, the entire sample was aged because of Smith's (1988) finding that double sampling provided only a slight improvement in the precision of the estimated age composition. Otolith weights continued to be recorded.

The major drawback of the spot sampling system was difficulty in setting and achieving the correct otolith quotas for each opening in each port. The overall target was a certain number of otoliths from each regulatory area over the course of the season (e.g., 5000 otoliths from 2C in 1987), and the port quotas for each opening were set by distributing that total in proportion to the predicted landings in each port during each opening. If the number of otoliths actually collected during an opening exceeded the quota, they were randomly subsampled after the opening to achieve the quota.

The subsampling method was based on the standard practice of storing all the otoliths from a single sling in a single vial at the time of collection. Usually, each vial was subsampled proportionally when it was necessary to reduce the number of otoliths. For a brief period early in 1988, whole vials were randomly chosen and discarded. Both methods produced a random subsample of a random sample, which is itself a random sample.

The subsampling scheme had the effect of making the sample sizes from each port and opening proportional to the predicted landings, but not to the actual landings. If landings in a port were higher or lower than the prediction, they were undersampled or oversampled. It was to avoid port quotas, subsampling, and disproportionate sampling that IPHC adopted proportional sampling by weight, described below.

1989 on: Proportional sampling by weight

Beginning in September 1989, samplers' instructions were to sample as many landings as possible, taking the otoliths from a random sample of fish equal in total weight to a certain percentage of the estimated total weight of each sampled landing. (An estimated or "hail" weight of a landing can be obtained from the skipper that is adequate for this purpose.) For example, in Kodiak the rule might be to take a 1% sample of landings from Area 3A, so the sample from a 12,000 pound trip would be collected by choosing fish randomly by some method until reaching a sample weight of 120 pounds. This might be a couple of large fish or several small ones. The port leader in each port was responsible for assuring that the fish were chosen in a truly random fashion. Clark (1990) explains the logic of the system:

> Every year we try to collect a certain number of otoliths from each regulatory area. In 1989, for example, we want about 3000 otoliths from Area 3A, which has a catch limit of 31 million lb or about 775,000 fish (at an average weight of 40 lb/fish). Approximately 0.4% of the catch should therefore be sampled. From Area 2B we also want 3000 otoliths from a catch limit of 10 million lb or about 400,000 fish (at 25 lb/fish), so about 0.9% of the catch should be sampled there.
>
> If we could sample every landing at every plant in every port, we could sample at those rates and come out right at the end of the season. But we do not sample in every port, nor do we sample in every plant in the ports we do cover, nor are we on hand for every landing even at the plants where we are active. We try to sample as many landings at as many plants in as many ports as is practical and sensible, but our coverage is necessarily incomplete. Before each opening, therefore, we estimate the proportion of the landings that will be sampled, and then

> raise the target sampling rate for landings that actually are sampled so as to achieve the desired percentage for the total landings from each area.
>
> The port leader's job is then to sample as many landings as possible at the prescribed rate, such as 1% or 2% or 4%. The basic sampling problem is therefore how to draw a "good" sample of a certain size from a landing. Owing to differences among plants, there is no standard solution to this problem, and it is up to the port leader in consultation with plant managers to devise an appropriate procedure for each plant.
>
> Note that there is no mention here of the *number* of fish to be selected from a landing. Rather the aim is to draw as the sample a certain percentage of the *weight* landed. The number of fish in the sample will then depend on the average size of fish in the landing, as it should. For equal total landing weights, a landing of large fish will be represented by a few large otoliths and a landing of small fish will be represented by many small otoliths, consistent with the relative numerical abundance of those sizes in the combined landings.

The target continued to be a certain number of otoliths from each regulatory area, beginning with 3000 in 1989. Since 1990, the target has been 1000 otoliths from Area 2A, and 2000 otoliths from each of Areas 2B, 2C, 3A, 3B, and 4. The lower targets were adopted on the basis of a study by Clark and Vienneau (1991) that found little loss in precision from reducing the sample size from 4000 to 2000 or even 1000, when the estimate in question was the overall age composition of the landings from an entire regulatory area for the entire year.

The sampling rates used in each port were determined as follows. A target of 2000 otoliths from Area 3A, for example, corresponded to about 80,000 pounds of fish at an average weight of 40 pounds each. If the quota was 25 million pounds, this implied an overall sampling rate of 80,000/25,000,000 = 0.32%. But only about 67% of the quota would go into ports with IPHC samplers, so the sampling rate in those ports should be 0.32/0.67 = 0.48%. Even in sampled ports, half or more of all landings could not be sampled because they occurred at night or at plants where sampling was impossible or at times when the samplers were busy. The percentage actually sampled varied from port to port, and the sampling rates were set accordingly. In Seward, for example, the percentage of 3A landings actually sampled was about 50%, so the working rate for sampled landings would be set at 0.48/0.50 = 0.96%, rounded up to 1%. Meanwhile in Homer only 30% of 3A landings were actually sampled, so the working rate there would be 0.48/0.3 = 1.6%, rounded up to 2.0%. (Sampling rates were normally rounded up to the next half percent.) The intent, and the usual outcome, was to sample the landings at the same effective rate in all ports and finish the season with about 2000 otoliths for Area 3A, regardless of how the landings were distributed among ports and openings.

This system has proved satisfactory, but it has a couple of drawbacks. Sampling rates are very low in most areas, so usually the sample from a landing consists of only a few fish, and these are selected individually rather than by the sling. Clearly this presents the danger of arbitrary rather than random selection, just as grab sampling did in the old days. This danger has been guarded against by requiring the port leaders personally to devise a truly random sampling method appropriate to each plant. This method is written up by the port leader and reviewed by the IPHC biometrician to assure randomness. Experience has shown the staff to be very conscientious and resourceful in devising random sampling procedures. The second drawback of the system is that small trips are undersampled, because all sampled trips are sampled at the same rate and small trips are less likely than large ones to be sampled at all. This has always been the case, but previous systems compensated one way or another for the lower probability of sampling small trips.

1991 on: Resumption of length measurements

Before 1963, IPHC samplers measured the lengths of a large number of fish, aged a subsample, and then keyed out the length distribution to estimate the age composition. Beginning in 1963, the samplers simply collected a large number of otoliths, aged a subsample, and then keyed out the *otolith* size distribution to estimate the age composition. Either way of handling the size distribution would produce an unbiased estimate of the age composition.

There was a second use of the size measurements, however, and that was to estimate the mean weight of fish in the landings and from that the number of fish at each age in the landings. Before 1963, weights were calculated from lengths using a length-weight relationship. From 1963 on, lengths were calculated from measured otolith sizes using an otolith size-body length relationship, and then weights were calculated from estimated lengths using the length-weight relationship. In the 1980s the staff noted systematic differences on a few occasions between actual fish weights and weights predicted from otolith size. A thorough review by Clark (1992) found that in fact the otolith size-body size relationship had changed substantially over time in Alaska, while the length-weight relationship held constant. In light of this finding, IPHC samplers resumed measuring the lengths of sampled fish in 1991. The routine weighing of otoliths from commercial catch samples was discontinued after 1992.

Summary of IPHC sampling for age and size composition of the catch

Geographic coverage

1933-1941	Emphasis on indicator grounds (Goose Island, Portlock, Albatross) with some sampling of landings from other areas.
1942-1948	Goose Island grounds only.
1949-1957	Declining emphasis on indicator grounds and increasing sampling in other areas, e.g. southeast Alaska.
1958-	Good geographic coverage of all commercial landings.

Sampling objectives and procedures

1933-1962	Objective was to estimate the age composition of individual landings. Large length sample and age subsample of 140 otoliths collected by the grab method from each landing.
1963-1972	Objective unchanged. Otolith-only sample of 200 collected by grab method from each sampled landing.
1973-1979	Objective unchanged. Otolith-only sample of about 200 collected by sling sampling from each sampled landing.
1980-1986	Objective was to estimate the age composition of all commercial landings in a month-region stratum. Systematic sample of slings taken, meaning that sample size was proportional to landing size.

1987-1989 Objective was to estimate the age composition of all commercial landings in a year-regulatory area stratum. Equal target sample sizes for all regulatory areas adopted. Otoliths collected mostly by spot sampling of slings. Targets achieved by setting quotas for each port.

1990- Objective unchanged. Equal but lower sample size targets for regulatory areas. Random otolith sample proportional to size of landing collected from as many landings as possible. Targets achieved by judicious setting of the sampling rates for sampled landings.

Primary size measurement

1933-1962 Fork length.

1963-1967 Otolith radius.

1968-1977 Otolith length.

1978-1990 Otolith weight.

1991- Fork length.

Subsampling procedure (of size distribution for age)

1935-1962 Age subsample of usually 140 otoliths deliberately distributed over the range of the length distribution of each sample landing.

1963-1977 Fixed-size age subsample totaling about 100 otoliths selected from otolith size distribution of each sampled landing. (I.e., approximately equal numbers of aged otoliths from each size interval.)

1978-1979 Proportional subsample of about 300 otoliths selected from the pooled otolith size distribution of a month-region stratum. (I.e., number of aged otoliths from a size interval proportional to the number of fish in the size interval in the measured sample.)

1980-1986 Proportional subsample of about 600 otoliths selected from the pooled otolith size distribution of a month-region stratum.

1987 Proportional subsample of 4000 otoliths selected from the pooled size distribution of the 5000 otoliths collected in each regulatory area.

1988- No subsampling; all otoliths read.

Estimates of age and size-at-age distributions, 1963-1990

From 1963 through 1990, port samplers collected otoliths but did not measure the lengths of fish in the market sample. Instead, the body lengths were estimated later from one or another predictive relationship between otolith size and fork length. The working measure of otolith size changed over the years from otolith radius (1963-1967) to otolith length (1968-1977) to otolith weight (1978-1990), but the database contains more than one kind of size measurement (some with associated age readings) in some years, so it is not always clear which otolith size distribution represents the size sample that should be keyed out to estimate the age composition and size at age. At time of writing there is also one year—1971—for which the data in the database are not decipherable at all and not usable. This section of the paper describes the data present in the database, identifies which otolith size measure represents the size distribution sample in each year, reviews the fork length predictors that were developed, and finally reports estimates of age and size compositions.

Data present in the database

Except for 1971 and some area-specific gaps listed below, the IPHC database contains usable otolith size samples of the expected kind, with age subsamples, for all areas and years. This was checked by plotting the frequency distributions of the size measurements in each year and looking for anomalous size samples.

But in addition to the expected data in each year, there are sometimes otolith size measurements of an unexpected kind, or actual fork lengths, usually with age subsamples. In 1978-1990 the standard measure of otolith size was weight, but otolith length was also measured for the subsample that was aged. The otolith length measurements in these years are therefore *not* a random sample of the otolith size distribution of the landings, but the otolith weight measurements are.

The other unexpected otolith measurements appear to be recording errors that may or may not be corrected in the future. The data for 1971 in particular are a mystery. There should be otolith length measurements and for most areas there are none. There are some otolith radius measurements but they have a strange frequency distribution. There are a large number of fork lengths but we know that fork lengths were not being measured by port samplers. The relatively small amount of otolith radius data recorded for 1972 in most areas appears to be part of the same problem. The most likely explanation of the 1971-72 anomalies is not any change in data collection procedures but in data processing procedures. The whole operation was being computerized at that time, and somehow the 1971 and 1972 data may have been scrambled (Dick Myhre, IPHC, pers. comm.) Whatever the reason, all of the 1971 data and the recorded otolith radius data from 1972 are unusable.

Tables 1a-f show the size and age data actually present in the database at time of writing. The list below explains the extra data and probable errors area by area.

Area 2A

Data are sparse and few before 1981. The otolith radius data from 1964-66 are quite consistent from year to year but have a very curious bimodal distribution. The otolith length data are highly variable from year to year, even in the two years with useful sample sizes (1970 and 1975). Only the otolith weight data from 1981-1990 appear to be useful for estimating age and size compositions.

Area 2B

The fork lengths recorded for 1964-68 may be measured fork lengths from commercial trips with observers aboard, or estimates calculated from otolith size, or misidentified (and rescaled) otolith radius measurements. They are not a subset of the fish with otolith radius measurements; no otolith has both a radius and a fork length recorded. The large number of "fork lengths" below 60 suggests that they were not actual fork lengths because commercial landings at the time were subject to a 66 cm minimum size limit. The frequency distribution of the numbers is different from the otolith radius frequencies when put on the same scale (multiplied by 10). The most likely possibility is therefore that these are calculated lengths. In any case these data are not usable for estimating age or size compositions (or really for anything else).

The expected otolith size measurements all have reasonable distributions. There is a shift in the otolith length distribution in 1973, but that is a natural result of the increase in the minimum size limit (from 66 to 81 cm).

Area 2C

The fork lengths recorded for 1967 are like those in 2B—unusable. All of the expected otolith size measurements look fine. There is a small but very noticeable upward shift in the otolith weight distribution in 1982 when the fleet switched from J-hooks to circle hooks, but only in this area.

Table 1a. Data on fork length, otolith size, and age of fish in samples of commercial landings in 1964-1990 for Area 2A as they appear in the IPHC market sample database in 1999. As explained in the text, some of the data are misidentified.

Area 2A	Fork length		Otolith radius		Otolith length		Otolith weight	
	Size	Age	Size	Age	Size	Age	Size	Age
1964	0	0	347	205	0	0	0	0
1965	0	0	521	217	0	0	0	0
1966	0	0	181	101	0	0	0	0
1967	0	0	0	0	0	0	0	0
1968	0	0	0	0	0	0	0	0
1969	0	0	0	0	0	0	0	0
1970	0	0	0	0	350	197	0	0
1971	933	291	206	100	0	0	0	0
1972	0	0	557	328	0	0	0	0
1973	0	0	0	0	92	36	0	0
1974	0	0	0	0	0	0	0	0
1975	0	0	0	0	636	379	0	0
1976	0	0	0	0	24	24	0	0
1977	0	0	0	0	0	0	0	0
1978	0	0	0	0	0	0	0	0
1979	0	0	0	0	0	0	0	0
1980	0	0	0	0	0	0	0	0
1981	0	0	0	0	539	539	545	545
1982	0	0	0	0	147	147	149	149
1983	0	0	0	0	284	284	284	284
1984	0	0	0	0	878	878	881	881
1985	0	0	0	0	893	893	894	894
1986	0	0	0	0	1549	1549	2843	1550
1987	0	0	0	0	215	215	216	216
1988	0	0	0	0	650	650	657	652
1989	0	0	0	0	813	813	819	816
1990	0	0	0	0	1384	1383	1385	1384

Table 1b. Data on fork length, otolith size, and age of fish in samples of commercial landings in 1964-1990 for Area 2B as they appear in the IPHC market sample database in 1999. As explained in the text, some of the data are misidentified.

Area 2B	Fork length		Otolith radius		Otolith length		Otolith weight	
	Size	Age	Size	Age	Size	Age	Size	Age
1964	2521	1055	13106	5892	0	0	0	0
1965	590	398	13717	7096	0	0	0	0
1966	1567	938	12339	5597	0	0	0	0
1967	1038	722	14844	7724	0	0	0	0
1968	1067	420	0	0	14548	7889	0	0
1969	0	0	0	0	14095	8153	0	0
1970	0	0	0	0	8348	4404	0	0
1971	12811	2530	2676	1653	0	0	0	0
1972	0	0	4304	2194	18181	6265	0	0
1973	0	0	0	0	11282	2897	0	0
1974	0	0	0	0	8076	3091	0	0
1975	0	0	0	0	14056	3086	1	0
1976	0	0	0	0	18255	4407	0	0
1977	0	0	0	0	11249	3892	0	0
1978	0	0	0	0	7563	2442	8947	3239
1979	0	0	0	0	0	0	14380	1921
1980	0	0	0	0	0	0	8256	4180
1981	0	0	0	0	3274	3274	6762	3283
1982	0	0	0	0	3953	3953	7917	3996
1983	0	0	0	0	2233	2233	5297	2259
1984	0	0	0	0	2385	2385	11162	2416
1985	0	0	0	0	2893	2893	10115	2924
1986	0	0	0	0	1773	1773	5686	1775
1987	0	0	0	0	4510	4510	5399	4515
1988	0	0	0	0	3910	3905	3955	3915
1989	0	0	0	0	3069	3064	3095	3080
1990	0	0	0	0	3575	3574	3581	3578

Table 1c. Data on fork length, otolith size, and age of fish in samples of commercial landings in 1964-1990 for Area 2C as they appear in the IPHC market sample database in 1999. As explained in the text, some of the data are misidentified.

Area 2C	Fork length		Otolith radius		Otolith length		Otolith weight	
	Size	Age	Size	Age	Size	Age	Size	Age
1964	0	0	11939	5908	0	0	0	0
1965	0	0	9038	4904	0	0	0	0
1966	0	0	9052	4029	0	0	0	0
1967	1029	119	9134	4434	0	0	0	0
1968	0	0	0	0	9714	4964	0	0
1969	0	0	0	0	3964	2586	0	0
1970	0	0	0	0	8576	4520	0	0
1971	1517	388	6432	3087	0	0	0	0
1972	0	0	590	513	15359	4720	0	0
1973	0	0	0	0	16501	5174	0	0
1974	0	0	0	0	7193	2999	0	0
1975	0	0	0	0	11233	4565	0	0
1976	0	0	0	0	8306	3213	0	0
1977	0	0	0	0	8817	3180	0	0
1978	0	0	0	0	6603	2133	8035	2831
1979	0	0	0	0	0	0	5857	1238
1980	0	0	0	0	0	0	2250	1310
1981	0	0	0	0	935	935	2042	936
1982	0	0	0	0	1124	1124	1748	1133
1983	0	0	0	0	1191	1191	2291	1202
1984	0	0	0	0	1097	1097	1957	1119
1985	0	0	0	0	2374	2374	3499	2399
1986	0	0	0	0	2322	2322	3502	2330
1987	0	0	0	0	3410	3409	4555	3414
1988	0	0	0	0	4008	4004	4052	4014
1989	0	0	0	0	2865	2863	2903	2874
1990	0	0	0	0	1681	1681	1681	1681

Table 1d. Data on fork length, otolith size, and age of fish in samples of commercial landings in 1964-1990 for Area 3A as they appear in the IPHC market sample database in 1999. As explained in the text, some of the data are misidentified.

Area 3A	**Fork length**		**Otolith radius**		**Otolith length**		**Otolith weight**	
	Size	**Age**	**Size**	**Age**	**Size**	**Age**	**Size**	**Age**
1964	3525	1249	18655	7301	0	0	0	0
1965	2110	654	16216	7086	0	0	0	0
1966	0	0	22198	8927	0	0	0	0
1967	0	0	22160	9941	0	0	0	0
1968	0	0	0	0	16245	7551	0	0
1969	0	0	0	0	14252	6427	0	0
1970	0	0	0	0	17734	8297	0	0
1971	5657	1782	4301	2151	0	0	0	0
1972	0	0	943	448	17043	5161	0	0
1973	0	0	0	0	22304	5900	0	0
1974	0	0	0	0	8259	3337	0	0
1975	0	0	0	0	11264	3142	0	0
1976	0	0	0	0	15195	3243	0	0
1977	0	0	0	0	14174	3150	0	0
1978	0	0	0	0	15315	2330	15315	2330
1979	0	0	0	0	0	0	14695	1395
1980	0	0	0	0	0	0	10816	2803
1981	0	0	0	0	1196	1196	7490	1201
1982	0	0	0	0	2203	2203	7674	2244
1983	0	0	0	0	1195	1195	7072	1201
1984	0	0	0	0	1685	1685	5941	1710
1985	0	0	0	0	2282	2282	7485	2309
1986	0	0	0	0	2375	2374	10355	2382
1987	0	0	0	0	4149	4149	5501	4150
1988	0	0	0	0	4122	4122	4151	4132
1989	0	0	0	0	4112	4110	4132	4117
1990	0	0	0	0	2440	2438	2441	2439

Table 1e. Data on fork length, otolith size, and age of fish in samples of commercial landings in 1964-1990 for Area 3B as they appear in the IPHC market sample database in 1999. As explained in the text, some of the data are misidentified.

Area 3B	Fork length		Otolith radius		Otolith length		Otolith weight	
	Size	Age	Size	Age	Size	Age	Size	Age
1964	5426	1268	7574	2738	0	0	0	0
1965	1084	324	8567	3763	0	0	0	0
1966	997	517	6072	2377	0	0	0	0
1967	0	0	10403	3905	0	0	0	0
1968	1479	991	0	0	12192	5663	0	0
1969	0	0	0	0	8114	4065	0	0
1970	0	0	0	0	10346	4483	0	0
1971	1838	631	4952	2769	98	98	0	0
1972	0	0	0	0	11722	3177	0	0
1973	0	0	0	0	9113	2175	0	0
1974	0	0	0	0	1669	1334	0	0
1975	0	0	0	0	3375	1801	0	0
1976	0	0	0	0	4301	1877	0	0
1977	0	0	0	0	2777	1464	0	0
1978	0	0	0	0	1932	1426	1932	1426
1979	0	0	0	0	0	0	706	549
1980	0	0	0	0	0	0	0	0
1981	0	0	0	0	595	595	783	602
1982	0	0	0	0	965	965	3604	978
1983	0	0	0	0	1719	1719	3480	1725
1984	0	0	0	0	1378	1378	1559	1384
1985	0	0	0	0	1889	1889	3575	1922
1986	0	0	0	0	1628	1628	2765	1628
1987	0	0	0	0	2291	2291	2934	2297
1988	0	0	0	0	2367	2366	2377	2371
1989	0	0	0	0	1653	1652	1655	1653
1990	0	0	0	0	2845	2844	2849	2848

Table 1f. Data on fork length, otolith size, and age of fish in samples of commercial landings in 1964-1990 for Area 4 as they appear in the IPHC market sample database in 1999. As explained in the text, some of the data are misidentified.

Area 4	**Fork length**		**Otolith radius**		**Otolith length**		**Otolith weight**	
	Size	**Age**	**Size**	**Age**	**Size**	**Age**	**Size**	**Age**
1964	1139	393	3824	1543	0	0	0	0
1965	2538	1333	1890	1062	0	0	0	0
1966	641	582	2497	1137	0	0	0	0
1967	425	419	5119	1929	0	0	0	0
1968	0	0	0	0	4018	1539	0	0
1969	0	0	0	0	2981	1404	0	0
1970	214	212	0	0	4265	1754	0	0
1971	1420	467	1121	525	770	300	0	0
1972	0	0	599	311	288	100	0	0
1973	0	0	0	0	737	247	0	0
1974	0	0	0	0	1035	804	0	0
1975	0	0	0	0	1449	1069	0	0
1976	0	0	0	0	1546	1285	0	0
1977	0	0	0	0	2255	1718	0	0
1978	0	0	0	0	820	196	3250	1387
1979	0	0	0	0	0	0	2994	2116
1980	0	0	0	0	0	0	2553	2024
1981	0	0	0	0	1698	1698	3332	1701
1982	0	0	0	0	1133	1133	2730	1144
1983	0	0	0	0	2573	2573	4919	2591
1984	0	0	0	0	2671	2671	3990	2723
1985	0	0	0	0	2504	2504	5219	2538
1986	0	0	0	0	2761	2761	5641	2761
1987	0	0	0	0	1539	1539	1718	1545
1988	0	0	0	0	2066	2066	2072	2070
1989	0	0	0	0	1576	1576	1586	1582
1990	0	0	0	0	2871	2869	2872	2870

Area 3A

The recorded fork lengths in the early years are not usable. All of the expected otolith measurement data appear to be sound.

Area 3B

The early fork lengths are unusable. All of the expected data are good except for 1979, which has a very strange distribution of otolith weights. There are no data for 1980.

Area 4

The early fork lengths are unusable. The otolith radius distributions from 1964-67 vary from year to year much more than they should on the basis of the respectable sample sizes. The otolith length distributions from 1968-70 and 1974-77 are reasonable; the 1971 distribution is strange; and there are very small samples in 1972 and 1973. The otolith weight distributions are strange in 1978-80 but reasonable thereafter. For the whole period, the years with usable data are 1968-70, 1974-77, and 1981-90.

Outliers

The database does contain outliers—otolith sizes and ages that are clearly wrong. This requires screening the data prior to any use, e.g. by running a data smoother through a size-age scatterplot and removing points that deviate greatly from the trend. For otolith size measurements with no age reading, the following range checks can be used: otolith radius 500-2000; otolith length 750-2400; otolith weight 60-1200.

Fixed and proportional age subsamples

Through 1977, subsamples of equal size were drawn from each otolith size category and aged. As a result, a higher proportion of otoliths was aged among the smallest and largest size categories which contained the fewest otoliths. Beginning in 1978, the intent was to take proportional subsamples, so that an equal proportion of otoliths would be aged in each size category. The change is clear in year-by-year plots of sampling proportion as a function of otolith size (Fig. 2), but it is also clear that even in 1978 and later years, the aged subsample was not strictly proportional. As a result, it is necessary to key out the size distribution in every year to estimate age and size-at-age distributions.

Predictors of fork length from otolith size

In the 1960s and 1970s, data were collected on otolith size and fork length for a large number of fish, and predictive relationships were developed for use in estimating the average length and length-at-age of fish in the landings. Unfortunately, all of the basic data are now lost, so it is impossible to recompute or refine the original formulas. There are some survey data on otolith length, otolith weight, and fork length in the 1970s and 1980s (Table 2); these were used by Clark (1992) to develop corrected predictors of fork length from otolith weight during the 1980s when growth rates were falling and the otolith size-fork length relationship was changing. For the 1960s and 1970s the published predictors should still be valid.

Otolith radius

From 1963 through 1967, the radius of the otolith was measured, and mean fork length in cm (*FL*) was estimated from otolith radius (*OR*) from the regression (Southward 1962):

$$\ln(FL) = -1.32086 + 1.30795 \cdot \ln(OR)$$

The unit of measurement of otolith radius is not stated in the original paper, but it is 10^{-4} m, or 0.1 mm. This is how the data were originally recorded (Calvin Blood, IPHC, pers. comm.). It is a point of confusion because Southward (1962) generally states actual radius measurements in mm and Quinn et al. (1983) state that the unit of radius measurement in the equation is mm. As a further complication, the data in the database are in units of 10^{-5} m. To give some specific figures, the average value of otolith radius in the database is around 900, which is 9 mm. To use the equation, this figure has to be divided by 10; an

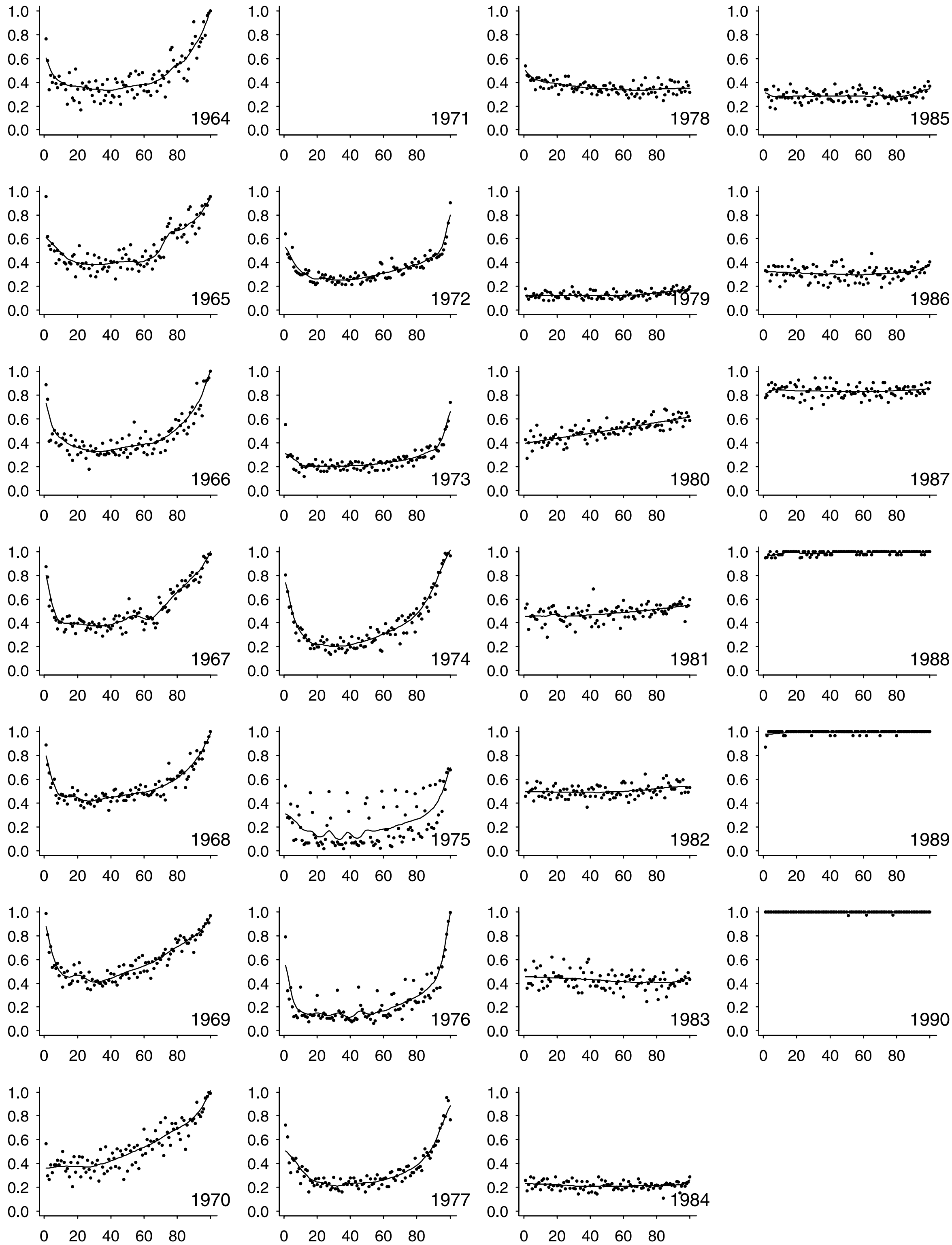


Figure 2. Proportion of measured 2B otoliths that were aged, plotted as a function of otolith size. (The x axis is an index of otolith size groups.)

otolith radius of 90 units corresponds to a predicted length of just under 100 cm. Alternatively, otolith radius measurements in mm can be used directly in the following equivalent transformation of the equation:

$$\ln(FL) = 1.690806 + 1.30795 \cdot \ln(OR)$$

The original paper states that the residual standard deviation about the regression line was 0.1284 (Table VII of Southward 1962).

Otolith length

From 1968 through 1977, the length of the otolith was measured instead of the radius. At first the fork length was estimated from a linear regression, but a set of area-specific

Table 2. Data on fork length (FL), otolith length (OL), and otolith weight (OW) from 1963-1990 present in the IPHC survey database in 1999.

Area		2B			2C			3A	
Datum	FL	OL	OW	FL	OL	OW	FL	OL	OW
1963	—	—	—	—	—	—	2705	622	622
1964	—	—	—	—	—	—	3852	0	325
1965	—	—	—	—	—	—	1915	0	443
1966	1248	79	79	—	—	—	—	—	—
1967	—	—	—	—	—	—	—	—	—
1968	—	—	—	—	—	—	—	—	—
1969	—	—	—	—	—	—	—	—	—
1970	—	—	—	—	—	—	—	—	—
1971	—	—	—	—	—	—	—	—	—
1972	—	—	—	—	—	—	—	—	—
1973	—	—	—	—	—	—	—	—	—
1974	—	—	—	—	—	—	—	—	—
1975	—	—	—	—	—	—	—	—	—
1976	142	124	127	—	—	—	431	333	339
1977	134	104	113	—	—	—	227	202	215
1978	165	149	149	—	—	—	289	255	255
1979	—	—	—	—	—	—	597	485	485
1980	—	—	—	747	143	447	1047	945	947
1981	331	303	302	—	—	—	2193	1954	1957
1982	238	226	223	643	585	585	1872	1622	1616
1983	202	146	147	1513	1380	1265	1160	1010	1011
1984	741	660	660	1549	1381	1364	914	770	766
1985	444	407	405	1626	1431	1424	916	837	835
1986	499	456	456	1669	1614	1614	932	888	889
1987	—	—	—	—	—	—	—	—	—
1988	—	—	—	—	—	—	—	—	—
1989	547	546	546	735	735	735	3139	3136	3139
1990	—	—	—	—	—	—	—	—	—

cubic equations was adopted when it was found that the linear regression "overestimated the lengths of the fish from the larger otoliths and underestimated the fish lengths from the smaller otoliths" (Southward and Hardman 1973). The revised predictors of average fork length *FL* in cm from otolith length *OL* in mm are:

British Columbia (Area 2B)

$$\ln(FL) = 2.06035 + 0.27736 \cdot \ln(OL) + 0.26648 \cdot [\ln(OL)]^2 + 0.00160 \cdot [\ln(OL)]^3$$

Southeast Alaska (Area 2C)

$$\ln(FL) = 1.62676 + 0.90838 \cdot \ln(OL) - 0.03469 \cdot [\ln(OL)]^2 + 0.04949 \cdot [\ln(OL)]^3$$

Gulf of Alaska (Area 3)

$$\ln(FL) = 3.46510 - 2.30676 \cdot \ln(OL) + 1.68946 \cdot [\ln(OL)]^2 - 0.23942 \cdot [\ln(OL)]^3$$

Bering Sea (Area 4)

$$\ln(FL) = 2.29027 - 0.27978 \cdot \ln(OL) + 0.61843 \cdot [\ln(OL)]^2 - 0.06415 \cdot [\ln(OL)]^3$$

The standard deviation about the cubics varied somewhat among areas but was close to 0.11 in all of them (Table 1 of Southward and Hardman 1973).

In the database, the otolith lengths are recorded in units of 10^{-5} m, like the otolith radii, so they have to be divided by 100 for use in the predictive equations.

Otolith weight

From 1978 through 1990 otolith weight was measured, and average fork length *FL* in cm was calculated from otolith weight *OW* in mg using the equations (Quinn et al. 1983):

All of Area 2

$$FL = 16.3 + 0.499 \cdot OW - 0.528 \times 10^{-3} \cdot OW^2 + 0.242 \times 10^{-6} \cdot OW^3$$

Areas 3 & 4

$$FL = 21.0 + 0.409 \cdot OW - 0.373 \times 10^{-3} \cdot OW^2 + 0.153 \times 10^{-6} \cdot OW^3$$

When halibut growth rates declined in the 1980s, it was found that these equations overestimated fork length because otolith growth rates decreased much less than somatic growth rates. Clark (1992) found that year-by-year fits of the relationship between otolith weight and fork length fell into two clusters corresponding to years through 1984 and years since 1984. He also reported the coefficients of cubic predictive equations by area for each of the two periods. These coefficients were estimated by fitting the equations to data on legal sized fish only, i.e. fish 81 cm and larger, so the predictor never estimates a fork length below 81 cm.

All of the other predictors reported in the present paper were fitted to all sizes of fish. For consistency, coefficients of Clark's (1992) cubic predictors have been re-estimated in the same way for use in recomputing the fork length estimates (Table 3). The form of the predictor is the same as in Clark (1992): a pair of cubic equations relating log fork length *FL* in cm to log otolith weight *OW* in mg, one equation for log otolith weights up to and including 5.4, and the other for larger log otolith weights:

$$\ln(FL) = b_0 + b_1 \cdot \ln(OW) + b_2 \cdot [\ln(OW)]^2 + b_3 \cdot [\ln(OW)]^3$$

The predictor was fitted to survey data from each area (Table 2). In every area, the standard deviation about the predictor was close to 0.080. (The residual variance could be reduced by fitting a separate equation for each year, but only very slightly.)

Calculation of estimates

Because of the need to key out the otolith size distribution and to allow for the effect of the size limit on the length distribution of younger age groups, the bootstrap is the most

Table 3. Coefficients of double cubic equations for predicting mean log fork length *FL* in cm from log otolith weight *OW* in mg:

$$\ln(FL) = b_0 + b_1 \cdot \ln(OW) + b_2 \cdot [\ln(OW)]^2 + b_3 \cdot [\ln(OW)]^3$$

	Coefficients of cubic equations			
Area, period, otolith size	b_0	b_1	b_2	b_3
Areas 2A and 2B				
Years through 1984				
Log otolith weight ≤ 5.4	0.168911	1.689535	-0.287523	0.022966
Log otolith weight > 5.4	18.822443	-8.721450	1.648698	-0.097052
Years after 1984				
Log otolith weight ≤ 5.4	-4.411993	4.520551	-0.864738	0.061690
Log otolith weight > 5.4	-16.806213	9.031324	-1.298920	0.065994
Area 2C				
Years through 1984				
Log otolith weight ≤ 5.4	-7.227097	6.305791	-1.259729	0.091862
Log otolith weight > 5.4	-3.486363	2.412906	-0.198836	0.005093
Years after 1984				
Log otolith weight ≤ 5.4	-6.726892	6.000806	-1.195514	0.087095
Log otolith weight > 5.4	-5.682871	3.493782	-0.374362	0.014347
Areas 3 and 4				
Years through 1984				
Log otolith weight ≤ 5.4	-6.550696	6.143548	-1.286100	0.098550
Log otolith weight > 5.4	-15.068282	8.349328	-1.198068	0.060632
Years after 1984				
Log otolith weight ≤ 5.4	-3.289715	3.624590	-0.652838	0.046311
Log otolith weight > 5.4	-8.566704	4.781904	-0.567859	0.024426

practical way to calculate point estimates and variances of age and size at age distributions from the market sample data. That has been done for the years 1964-1990. Specifically, for each IPHC regulatory area and year with usable data, the following steps were performed:

1. Outliers were removed by running range checks on the distribution of otolith sizes and by running a data smoother through plots of otolith size against age and removing the 0.5% of points farthest from the trend line. Inspection of the plots showed that this fraction was sufficient to remove all of the suspicious points.

2. An artificial sample of otoliths (some with ages, some not) equal in size to the actual data set was drawn by resampling the data with replacement.

3. A fork length was estimated for each otolith by applying the appropriate predictor to the otolith size and adding a random normal deviate with mean zero and standard deviation equal to that of the residuals of log fork length about log otolith size. (The distribution of this deviate was truncated below so that the predicted fork length could not fall below the minimum size limit. It was also truncated above so that the predicted length could not exceed 250 cm, which would have happened in some cases when using the predictors based on otolith length.)

4. The otolith size distribution (not the artificial fork length distribution) was keyed out with the aged subsample to estimate proportion at age, mean length at age, variance of length at age, and so on.

5. Steps 2-4 were repeated many times to obtain empirical values for the mean and variance of all the estimates. For example, the mean and variance of the estimated standard deviation of length at age 10 was obtained as the mean and variance of the values calculated in all of the bootstrap samples. (Only for the proportion at age could a point estimate be calculated from the data. It was always very close to the bootstrap mean, but it is the point estimate that is reported.)

Comparison with earlier estimates

Estimates of age composition and mean weight at age in the commercial landings for the years 1935-1982 were calculated by T.J. Quinn II of the IPHC staff in the early 1980s. For the years through 1962 they are based on measured fork lengths. For later years the estimates of mean weight are based on estimates of fork length calculated from the predictors in use at the time.

Quinn’s estimates of age composition are nearly identical to those calculated by the author for years after 1964, as they should be because in both cases the age composition was estimated by keying out the otolith size distribution.

Figure 3 shows a comparison of Quinn’s estimates of mean weight at age with the ones computed by the author for Areas 3A and 2B. Also shown are the sample values of mean weight at age in the commercial landings in the years when fork lengths were actually measured (1935-62 and 1991-98), and the mean weight at age of fish above the commercial size limit that were caught in IPHC setline surveys. During the years 1963-90 surveys were conducted irregularly, almost all of them in Areas 3A and 2B. The commercial fishery tends to select larger fish than the survey, so among the younger age groups one would

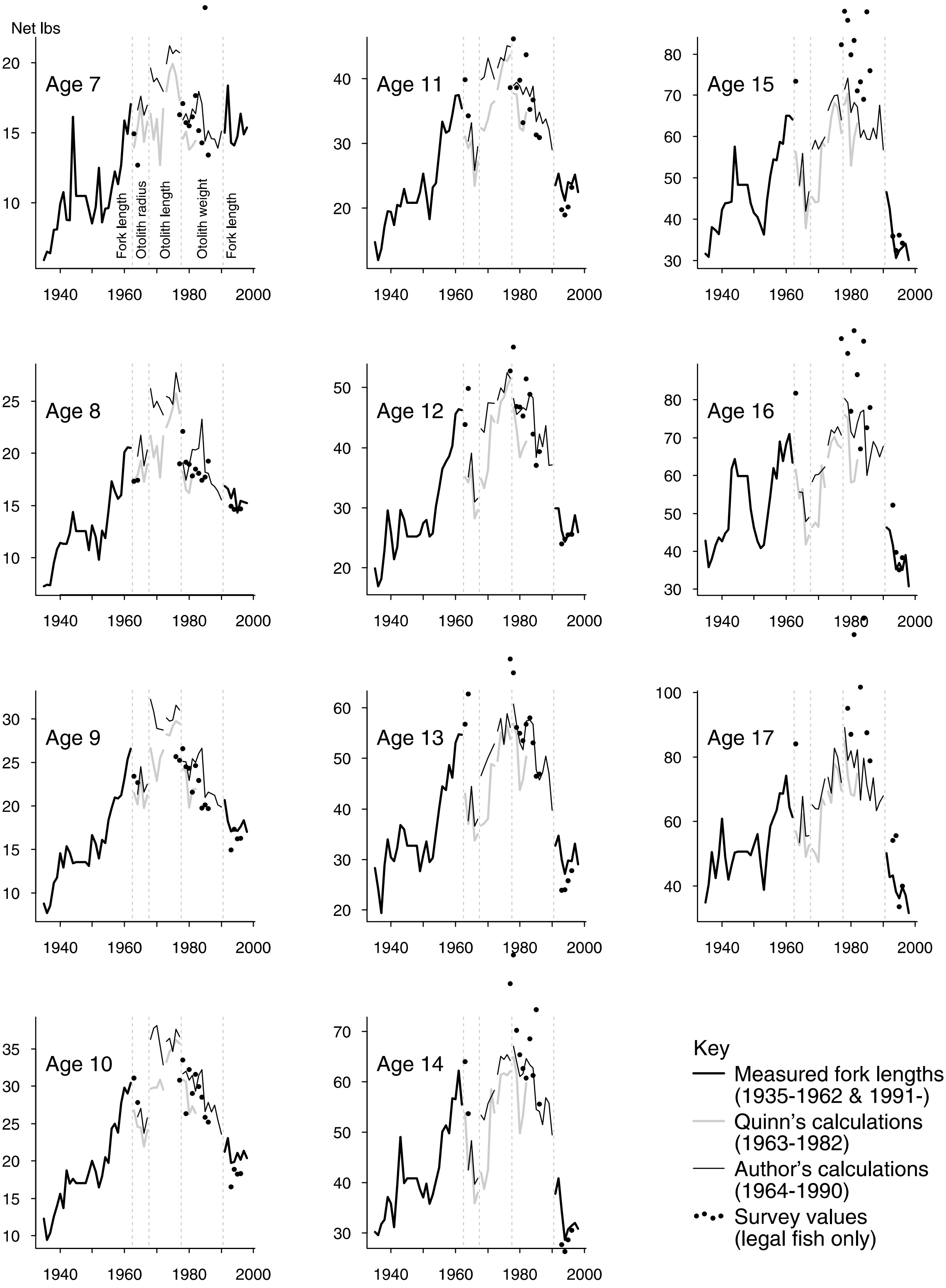


Figure 3a. Various estimates of mean weight at age over time in Area 3A.

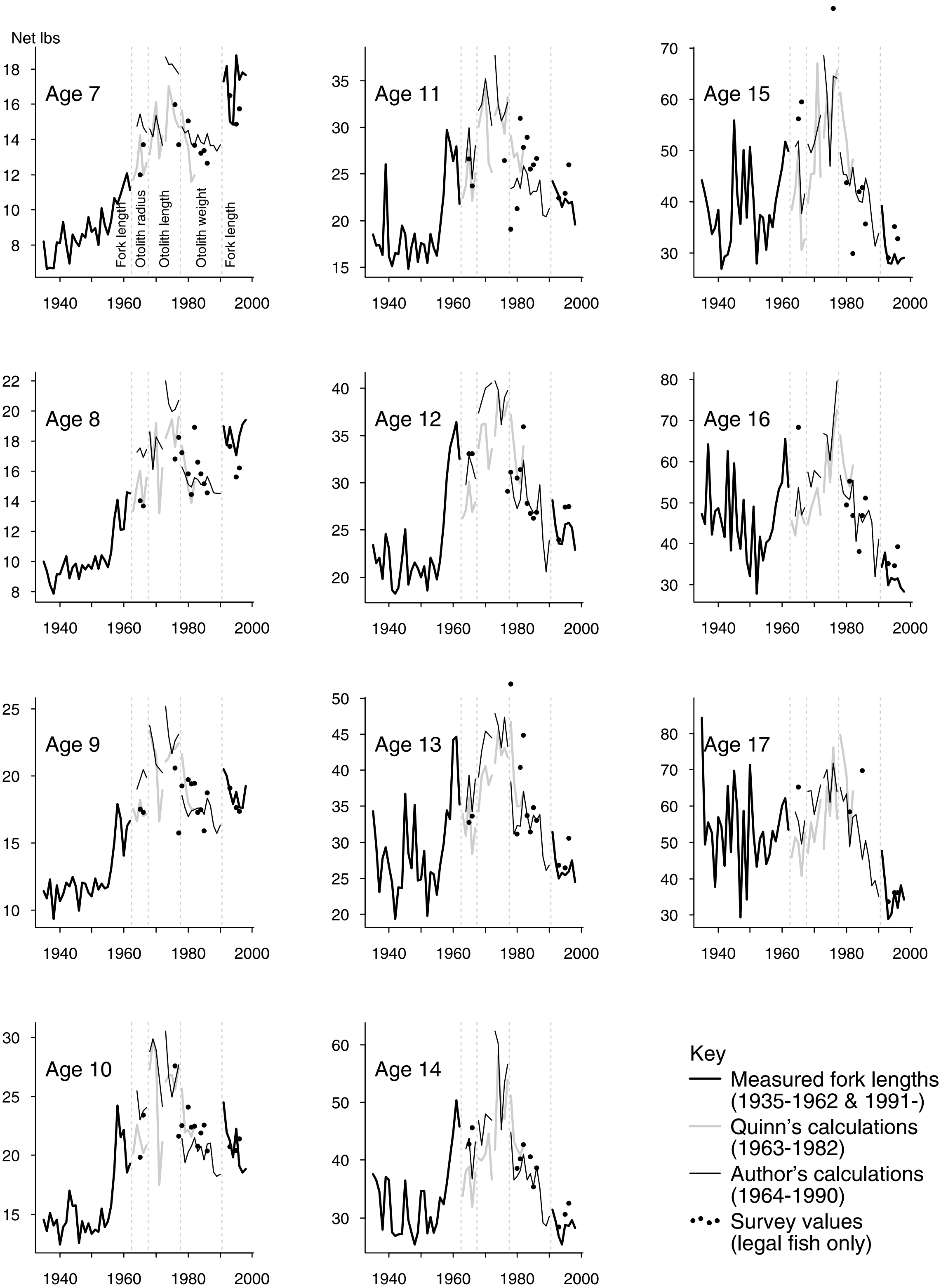


Figure 3b. Various estimates of mean weight at age over time in Area 2B.

expect some difference and this is in fact seen in the measured survey and commercial values for 1991-98. Among older age groups the mean weight at age is similar in survey and commercial catches. There are a number of other features in the figures that require explanation:

(i) The author's estimates of mean weight at age based on otolith size-fork length relationships are generally higher than Quinn's, especially for younger fish. This is because the author's estimates take account of the effect of the commercial size limit on the size distribution of the landings, while Quinn's do not.
(ii) In many cases the estimates for Area 3A in the years 1963-67, based on otolith radii, are much lower than both the survey estimates and the surrounding commercial values. It seems clear that they are too low, probably the result of using a single coastwide predictor based on long time series of data. By the mid-1960s halibut in Area 3A were larger at each age than they had been earlier in the century, and larger than fish in Area 2B.
(iii) The author's estimates for Area 3A are much higher than Quinn's for the years 1968-70. This is because Quinn's estimates for those years are based on the original otolith length-fork length predictor, while the author's are based on the revised formula given by Southward and Hardman (1973).
(iv) There were a number of surveys during the years 1978-90, when fork lengths in commercial samples were being predicted from otolith weight. Both the survey average weights at age and the predicted commercial averages vary a good deal from year to year, but for the most part the author's estimates track the trend of the survey values better than Quinn's, as they should because they were based on the survey data (Clark 1992).

Smoothed estimates of mean weight at age

The halibut stock assessment requires an estimate of mean weight at age for every age group in the catch so that catch in number can be calculated from catch in weight. Calculation of exploitable biomass by present methods requires an estimate of mean weight for all ages in the stock, even those not seen in the landings. As explained above, there are problems with the data for 1971 in all areas and some other years in some areas. It also appears that there are problems with the fork length predictor for 1963-67. To calculate a single, complete set of estimates, a smoother was run through all of the trustworthy year-by-year estimates of mean weight at age from 1935 through 1998. For the period 1935-1963, these were Quinn's estimates (based on measured fork lengths through 1962), excluding 1963 in all areas and 1942-1947 north of Area 2B because it is known that no real data were collected for those areas in that period. Weights from Area 3B were used for Area 4 in the years 1935-1951, when there was no catch in Area 4 under the area definitions used by Quinn. For the period 1964-1990, the trusted estimates were those calculated by the author, i.e. excluding 1964-1967 and 1971 in all areas, and all other years without usable data in each area. Weights from Area 2B were used for Area 2A in the years 1964-1980. For the period 1991-1998, the recent estimates based on measured fork lengths were used (Fig. 4). The smoothed weights still had some gaps, which were filled as follows:

(i) In two areas, the earliest smooth weight at age 5 was extended back to 1935.
(ii) The mean weight for the 25+ age group in 1935-1967 was scaled to the average of ages 21-24 in each year, the scaler being the overall 1968-1998 ratio.

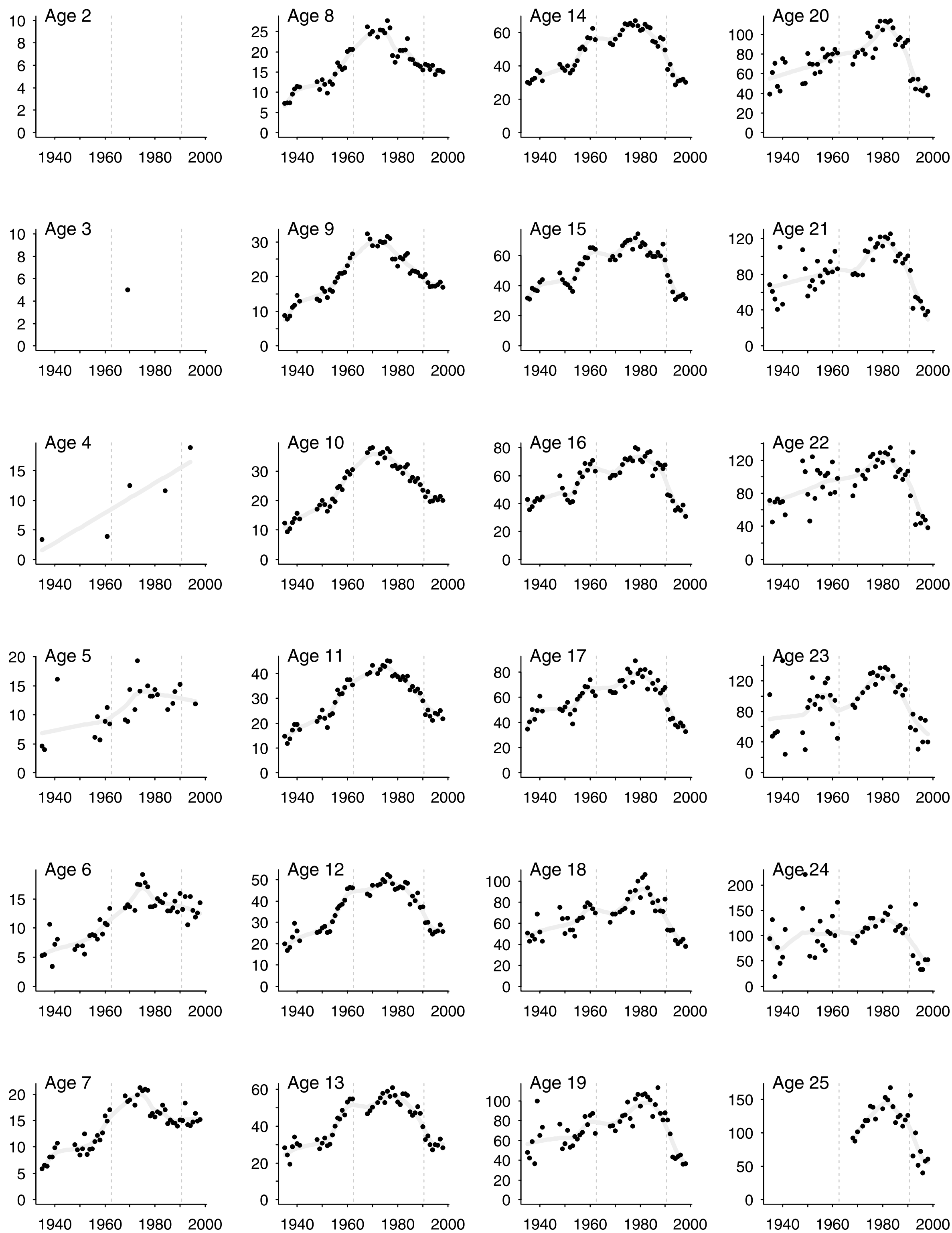


Figure 4a. Year-by year and smoothed estimates of mean weight at age in Area 3A.

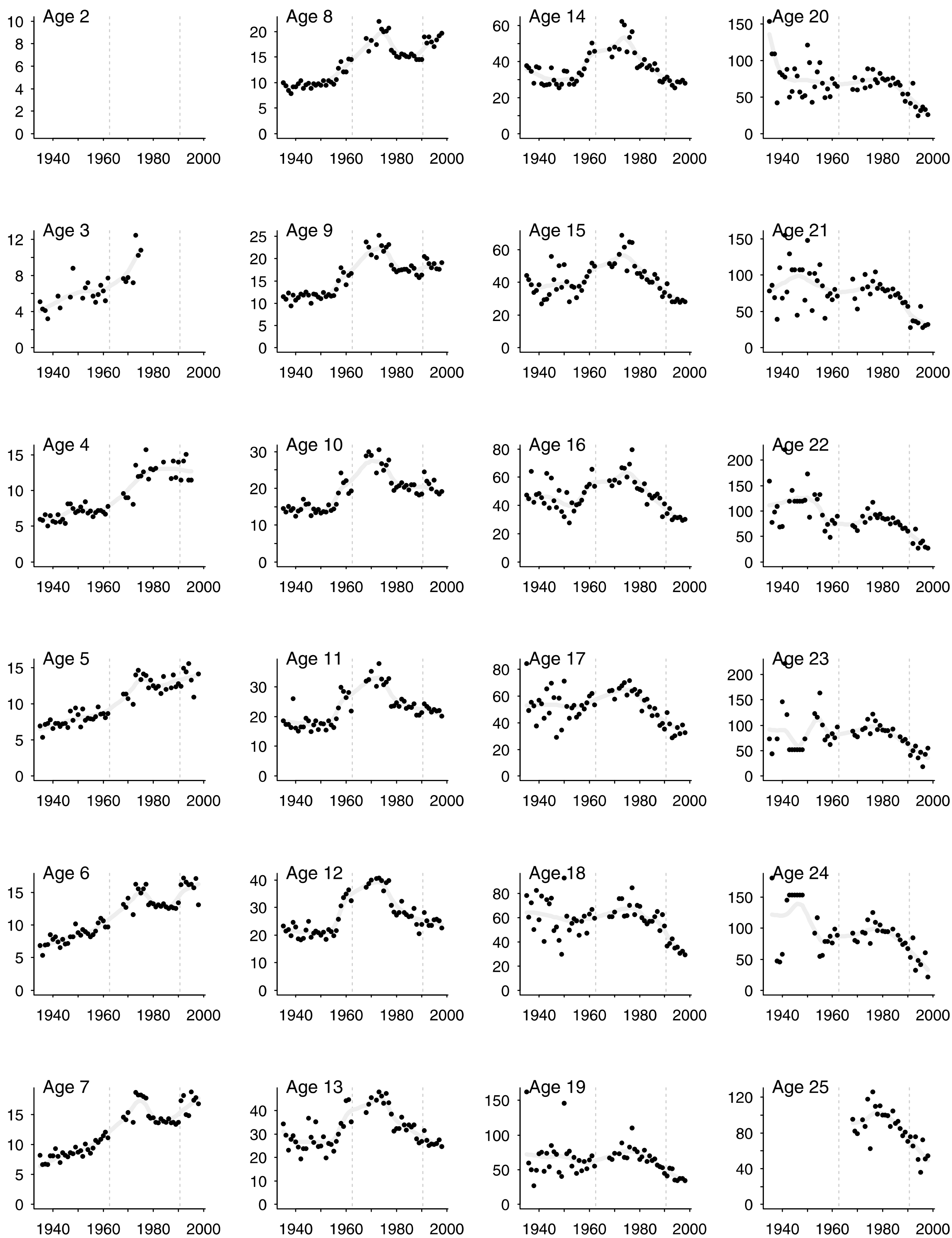


Figure 4b. Year-by year and smoothed estimates of mean weight at age in Area 2B.

By these assorted devices, a complete set of smoothed weights for ages 5 through 25+ was produced for all areas in all years 1935-1990.

Estimates of age composition in years without usable data

Almost all of the area/year cells for which there are no usable data in the present database are before 1981 and therefore are included in Quinn's estimates for the period 1935-1982. In some cases his estimates were obtained by substituting data from adjacent areas or years (Quinn et al. 1983), but it is also possible that he had a bigger data set to work with, before some of it was lost or scrambled in migrating among computer systems. The author's estimates of age composition agree closely with Quinn's for area/year cells with usable data, as they should because in both cases the age composition is estimated by keying out the otolith size distribution with no reference to any fork length predictor. There is no reason not to trust Quinn's judgment and computation for the area/year cells that are now problematical, so his estimates are simply copied for those cells, even though they are not reproducible now.

The only gap not covered by Quinn's estimates was Area 4 in the years 1945-1951. Under present area definitions there was some catch in Area 4 then, but it was all in Area 3B under the definitions used by Quinn, so his estimates of age composition in Area 3B in those years were taken for Area 4 as well.

With these borrowings, a set of age composition estimates was produced covering all area/year cells that produced some catch according to present area definitions.

Conversion from estimates of standard deviation to estimates of variance

The stored estimates include a point estimate of the standard deviation of length at age and the coefficient of variation of that estimate; likewise for the standard deviation of log length. For some purposes, users will want estimates that refer to the variance rather than the standard deviation. For clarity, let d denote the standard deviation of, say, log length, and $D = d^2$ the variance. If $\hat{d}$ is the point estimate (bootstrap mean) of the standard deviation, the point estimate of the variance $\hat{D}$ is:

$$\hat{D} = \hat{d}^2 + \mathrm{V}(d) = \hat{d}^2 \cdot \left\{1 + \left[\mathrm{CV}(d)\right]^2\right\}$$

and by the delta method the variance of D is (to a good approximation):

$$\mathrm{V}(D) \approx \left[2\hat{d}\right]^2 \mathrm{V}(d)$$

The coefficient of variation of D is then:

$$\mathrm{CV}(D) = \frac{\sqrt{\mathrm{V}(D)}}{\hat{D}} = \frac{2 \cdot \hat{d} \cdot \sqrt{\mathrm{V}(d)}}{\hat{d}^2 \cdot \left\{1 + \left[\mathrm{CV}(d)\right]^2\right\}} = \frac{2 \cdot \mathrm{CV}(d)}{1 + \left[\mathrm{CV}(d)\right]^2}$$

Availability of files containing the estimates

The estimates themselves are not tabulated in this report, but are available from the International Pacific Halibut Commission. At time of writing they are posted on the Commission's website, which is: www.iphc.washington.edu.

Discussion

IPHC has a long series of biological sampling data from the commercial fishery. The estimates of age composition based on these data are straightforward and reliable except for the likely bias resulting from spotty sampling in the 1930s and greatly reduced sampling in the 1940s. The estimates of size at age are similarly reliable for the years when fork lengths were measured (1935-62 and 1991-), but they are greatly complicated by the use of otolith-only sampling during the years 1963-90, and for some years by an apparent scrambling of otolith size measurements in the present database. The stored estimates of size at age are, it is hoped, the best that can be done in the circumstances, but they cannot be regarded as accurate for the years 1963-90 and therefore cannot be used for purposes of analysis in the same way as estimates based on actual fork length measurements.

References

Blood, C.L. 1989. Catch sampling. IPHC *1989 Stock Assessment Document I*: 60-69. Int. Pac. Halibut Comm., Seattle (mimeo).

Clark, W.G. 1990. IPHC otolith sampling guidelines. IPHC *1990 Stock Assessment Document I*: 48-57. Int. Pac. Halibut Comm., Seattle (mimeo).

Clark, W.G. 1992. Estimation of halibut body size from otolith size. Int. Pac. Halibut Comm. Sci. Rep. 75.

Clark, W.G., and Vienneau, B.A. 1991. Evaluation of otolith sample size and double reading. IPHC *Report of Commission Activities 1991*: 229-231. Int. Pac. Halibut Comm., Seattle (mimeo).

Gilroy, H.L., Forsberg, J.E., and Clark, W.G. 1995. Changes in commercial catch sampling and age determination procedures for Pacific halibut, 1982 to 1993. Int. Pac. Halibut Comm. Tech. Rep. 32.

Hardman, W.H., and Southward, G.M. 1965. Sampling the commercial catch and use of calculated lengths in stock composition studies of Pacific halibut. Int. Pac. Halibut Comm. Rep. 37.

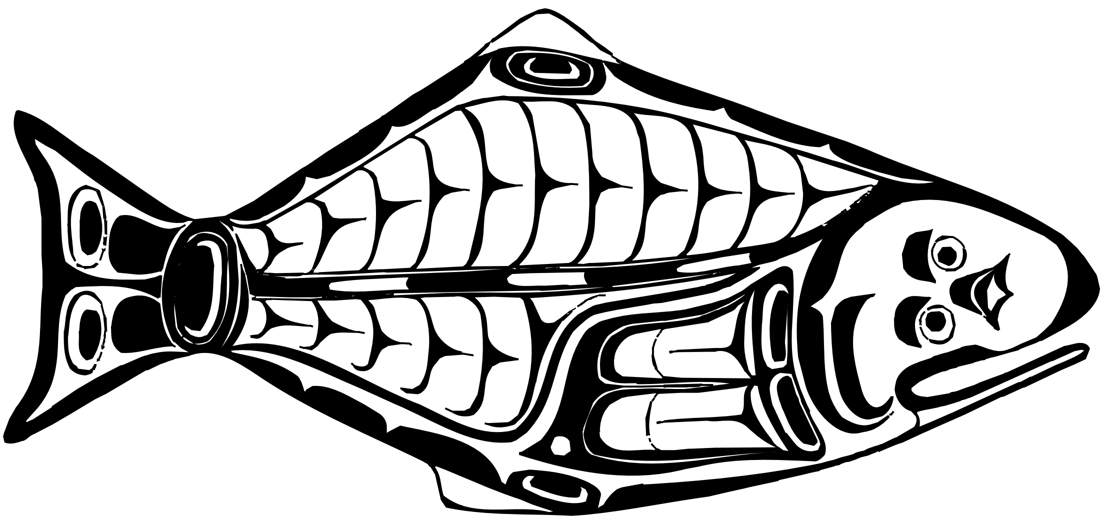
Quinn, T.J. II, Best, E.A., Bijsterveld, L., and McGregor, I.R. 1983. Sampling Pacific halibut (*Hippoglossus stenolepis*) landings for age composition: history, evaluation, and estimation. Int. Pac. Halibut Comm. Sci. Rep. 68.

Skud, B.E. 1977. Regulations of the Pacific halibut fishery, 1924-1976. Int. Pac. Halibut Comm. Tech. Rep. 15.

Smith, P.J. 1988. Optimal two phase sampling for estimating the exploitable biomass of halibut accounting for nonsampling errors. IPHC *1988 Stock Assessment Document II*:1-10. Int. Pac. Halibut Comm., Seattle (mimeo).

Southward, G.M. 1976. Sampling landings of halibut for age composition. Int. Pac. Halibut Comm. Sci. Rep. 58.

Southward, G.M., and Hardman, W.H. 1973. Otolith length and fish length of Pacific halibut. Int. Pac. Halibut Comm. Tech. Rep. 10.



HALIBUT CREST - adapted from designs used by Tlingit, Tsimshian and Haida Indians