CSIRO PUBLISHING

www.publish.csiro.au/journals/mfr *Marine and Freshwater Research*, 2005, **56**, 783–789

Quality control of age data at the Alaska Fisheries Science Center

Daniel K. Kimura[A,B] *and Delsa M. Anderl*[A]

[A]Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA,
7600 Sand Point Way N.E., Seattle, WA 98115-6349, USA.
[B]Corresponding author. Email: dan.kimura@noaa.gov

Abstract. The Alaska Fisheries Science Center operates a 14-person Age and Growth Program that specialises in the ageing of various groundfish species using otoliths. In 1983, a quality control programme was established whereby a random subsample of 20% of the total of aged samples is re-aged by a second age reader. The purpose of this programme is to assure, to the greatest extent possible, that ages used in stock assessment are based on consistent ageing criteria. This age data is entered into our AGEDATA Microsoft ACCESS™ database where it can be easily updated, corrected and analysed. VISUAL BASIC computer programmes AGREE (a precision estimating programme) and RANGES (an outlier searching programme) were written to routinely analyse age data before data are released to end-users. The statistical relationship between average percentage error and coefficient of variation is described, as well as an interpretation of Bowker's test for symmetry. Discrepancies between the reader and tester are reconciled while viewing the problematic otoliths using a dual-headed microscope, and reconciled ages are assigned. When necessary, all questionable otoliths in a troublesome sample may be re-aged.

Extra keywords: Bowker's test, fish ageing, precision measures, quality control.

Introduction

Alaska Fisheries Science Center (AFSC) age data quality control comes in two forms: validated ageing criteria and good precision that minimises between-reader biases. At the AFSC, age validation has proved difficult, with known-age fish usually occurring only at the youngest ages, with the exception of sablefish (Heifetz *et al*. 1998). Therefore, we have mainly pursued radiometric age validation (Kastelle *et al*. 1994, 2000). We also have had some success providing indirect age validation for Pacific cod (*Gadus macrocephalus*), when otoliths were available from tagged and recovered fish (Roberson *et al*. 2005).

Because most data generated by the Age and Growth Program are used for stock assessments, we require age data that are consistent from year to year so that year classes can be successfully modelled. It is important that strong year classes are not manufactured, but it is also important that they are not missed, because then the model fits will be poor and the ages will be unusable. Also, historically there has been a large turnover in age-reading personnel; therefore, a great effort has been applied towards optimising between-reader precision. Between-reader precision is the ideal tool for monitoring a new reader's ageing skills. Beginning in 1983, the AFSC's Age and Growth Program initiated a 20% 'test sample' for re-ageing for all ages released by age readers, even if the age reader was highly experienced. This 20% test sample is used for calculating precision statistics has been used for 20 years and is the basic tool that maintains our between-reader precision at high levels.

Materials and methods

The AGEDATA database

The heart of the Age and Growth Program is the AGEDATA Microsoft ACCESS™ database (ACCESS 2002), which is where all current age data (i.e. samples that are currently being aged) resides. The AGEDATA database is fairly complex with more than 16 macros that perform such tasks as entering age data, marking specimens for testing (i.e. re-ageing), calculating discrepancies (i.e. determining specimens aged differently by the reader and the tester), assigning codes, etc. In addition, two computational programmes, AGREE and RANGES, were written in Microsoft VISUAL BASIC™ and use AGEDATA tables for data input.

AGREE is a programme that calculates precision statistics such as percentage agreement (PA), average percentage error (APE), coefficient of variation (CV), as well as cross-tabulations and Bowker's test for symmetry. The RANGES programme plots fish size at age and is used to identify ages and lengths that are totally incompatible. If there are no apparent collector errors, errors in data or obvious reader errors, slow-growing and fast-growing individuals are allowed (i.e. growth outliers). However, specimens will be discarded when the otolith size *v.* fish size relationship clearly indicates collector errors, or other errors have been made that are not correctable.

Generating and processing age data

For experienced age readers at the AFSC, production ageing is performed using the following protocol based on the discipline of precision testing (Fig. 1).

(*1*) Reader ages entire sample.
(*2*) Reader selects 20% test random sample: a random number from 1 to 5 is drawn and every 5th sample thereafter.

 10.1071/MF04141 1323-1650/05/050783

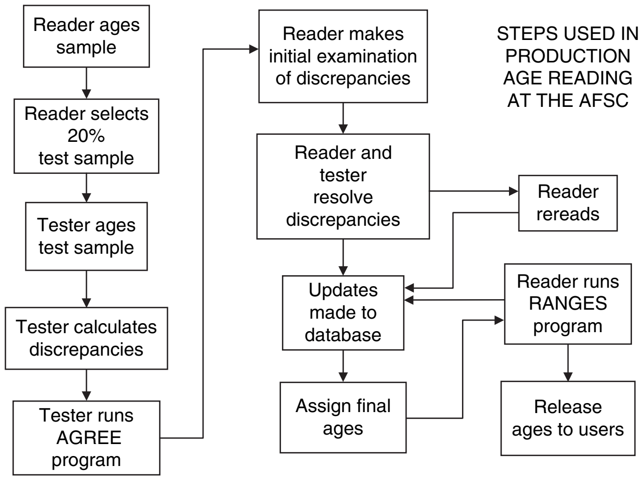

Fig. 1. Flow chart showing steps taken when generating production ages at the Alaska Fisheries Science Center's Age and Growth Program.

(*3*) An experienced tester reads the 20% sample without knowledge of the reader's original age.
(*4*) The tester runs macro that calculates discrepancies, which are generally those specimens where reader and tester ages differ (i.e. ages are not within ±0 years). However, this criterion may be too demanding for fish within the same species that attain older ages. For example, acceptable agreement is considered within ±1 years for rockfish aged 12–24 years, 2 years for rockfish aged 25+ years and ±1 year for pollock aged 15+ years. If the ages are not classified as discrepancies, then the reader age stands, and are not resolved by reader and tester.
(*5*) Tester runs AGREE to generate precision statistics and cross-tabulations.
(*6*) Reader initially reviews discrepancies to see if the tester's ages are preferable to the reader's original ages.
(*7*) Reader and tester resolve remaining discrepancies at the dual-headed scope.
(*8*) Reader performs rereads if necessary.
(*9*) Corrections are entered into the AGEDATA database.
(*10*) Final ages are assigned.
(*11*) Reader runs RANGES and output is examined for outliers. Corrections are made if necessary.
(*12*) Ages are released to users.

Readers are provided with the following information before ageing: specimen number, location in tray, fish length, sex and date of collection. Date of collection and the amount of marginal growth are used to assign an age based on the International Birthday Convention, so that all fish in the northern hemisphere have a birthday on January 1st regardless of the hatch date.

Supplying the age readers with fish length has been a more controversial issue. Obviously, knowledge of fish length may bias an age determination, so why do we provide fish length? There are several reasons, none as sinister as some might suspect.

(*1*) Half of our sample collections come through an at-sea Observer Program where collection errors can easily occur. We routinely find obvious errors such as three otoliths in a vial, or otoliths of the wrong species, so we know collection errors are occurring. We have a need to age large numbers for stock assessments, so if there are collector errors, then we would like to know this right away. Collection errors are easier to identify when fish length is available.
(*2*) We view fish length as useful ancillary information for species such as Pacific cod that are difficult to age even at young ages, but that often have strong modal lengths at these ages. Using ancillary information for these species probably reduces the mean square error of age estimation for these species.
(*3*) If the age readers did not have fish length, then they could simply use otolith size as a proxy because otolith size and fish size are strongly isometric.

Basically, we believe that our age readers are not negatively influenced by having fish length available during ageing. A new age reader overly depending on fish length would quickly be detected during precision testing, whereas an experienced age reader would only refer to fish length when annuli on the otolith are insufficient alone to provide an age. However, we understand that this is a controversial issue to some.

Following ageing, the age reader enters the age information described below in Table 1 into the AGEDATA database fields.

The relationship between precision statistics, average percentage error and coefficient of variation

Reader–tester data can be used to measure between-reader precision (Kimura and Lyons 1991; Campana 2001). The most traditional method of estimating precision by age readers is through the 'percentage agreement' (PA). Simply put, PA is the percentage of structures that are aged the same on two different occasions by either the same or different age readers. If known-age specimens are available, PA is the percentage of structures that were aged correctly. Also, PA might refer to agreement between ageing methods.

Unless qualified, PA refers to exact agreement (±0). This is the 'official' agreement that we generally report. However, sometimes the definition is extended so that ages within a year or two are considered in agreement, as described earlier in the definition of discrepancies. When ages are in agreement, they are not resolved between reader and tester.

In this section, we will not scale APE or CV to percentages, but keep them as proportions. The mathematical discussion is less confusing without this conversion. However, in the results section, the results are presented in the more traditional percentages, which are just the proportions of this section multiplied by 100.

Beamish and Fournier (1981) recognised that PA did not take into account the absolute age of the fish. For example, ±1 year is poor agreement for a 2-year-old fish, but quite outstanding agreement for a 30-year-old fish. To adjust for the age of fish, Beamish and Fournier (1981) proposed using the average percentage error or APE. Let

$$x_{ij} = \text{be the age of the } j\text{-th fish determined}$$

on the i-th occasion, $i = 1, \ldots, r$, $j = 1, \ldots, n$.

$$\bar{x}_j = \Sigma_i x_{ij}/r$$
$$APE_j = \Sigma_i |x_{ij} - \bar{x}_j|/(r\bar{x}_j).$$
Averaged over all fish, $APE = \Sigma_j APE_j/n$.

Chang (1982) argued that statistical efficiency and the possibilities for statistical hypothesis testing made the CV more attractive than APE. Here,

$$sd_j = \sqrt{\Sigma_i (x_{ij} - \bar{x}_j)^2/(r-1)},$$
$$cv_j = sd_j/\bar{x}_j,$$

and averaged over all fish $cv = \Sigma_j cv_j/n$.

It should be recognised that Chang's argument for the superiority of the CV rests on the assumption of normally distributed error. For the normal distribution, the sample mean and variance are the generally accepted estimates of the theoretical mean and variance.

Table 1. Information entered by age readers into the AGEDATA database

Age	Otolith age estimate. If an age cannot be determined for a specimen, the assigned value in this field is '−1'. If an age reader is only comfortable assigning an age range, minimum age is assigned in this field.
Method	If age was determined from the otolith surface pattern, this space is left blank. If age was determined using the break-and-burn method, 'B' is written in this column. See Appendix 1 for other method codes.
Read code	The readability code provides data reviewers a sense of the difficulty of an otolith pattern or if problems were associated with a specimen. Readers are required to assign a readability code for each specimen. Appendix 1 lists the readability codes.
Edge type	An important factor in determining the age of an otolith, especially from a young fish, is the amount of opaque growth between the last translucent zone and the edge of the otolith. Edge type is easier to determine in some fish than others and determining edge types of very old fish is sometimes not useful. Edge types may vary depending on the reading axis and reading method. Assigning an edge type designation is optional but highly recommended, especially for young otoliths. Appendix 1 lists the edge type codes.
Maximum age	There will be otoliths where a single age is difficult to assign because of the ambiguity of the otolith pattern. However, if a reader is only comfortable assigning an age range, the minimum age is assigned to age field (above) and the determined maximum aged entered under the maximum age field. When an age range is provided the read code is 3.
Comment	Writing comments by hand on reading sheets pertinent to the ageing of a specimen is highly encouraged, especially for new age readers.

The APE estimate itself is nearly a maximum likelihood estimate under the double exponential (or Laplace) distribution. The double exponential (DE) distribution (Kendall and Stuart 1973)

$$f(x) = [1/(2\beta)]\exp[-|x-\alpha|/\beta],$$

has maximum likelihood estimates

$$\hat{\alpha}_j = \text{sample median.}$$

$$\hat{\beta}_j = \Sigma_i |x_{ij} - \hat{\alpha}_j|/r.$$

An alternative APE that differs from the original APE only in that the median has replaced the mean, can be defined as

$$APE(2)_j = \hat{\beta}_j/\hat{\alpha}_j.$$

$APE(2)_j$ can be averaged over all fish so that

$$APE(2) = \Sigma_j APE(2)_j/n.$$

The difference between Beamish and Fournier's APE and Chang's CV estimate is mostly the difference in parameterisation between the normal and DE distributions. When we compare APE and the normal distribution CV, we are comparing the DE parameter $APE = \beta/\alpha$ with the normal theory $CV = \sigma/\mu$. The problem is that APE is not the CV of the DE distribution. Since the variance of the DE distribution is $VAR(X) = 2\beta^2$, the CV of this distribution is $CV(X) = \sqrt{2}\beta/\alpha = \sqrt{2}APE$. For most samples the $\sqrt{2}$ factor almost completely converts APE to CV. When only two observations are made on each otolith, this conversion is exact:

$$CV = \sqrt{2}\sqrt{(x_1/2 - x_2/2)^2}/\bar{x}$$

$$APE = |x_1/2 - x_2/2|/\bar{x}, \text{ so that}$$

$$CV = \sqrt{2} \times APE.$$

In order to compare the normal and DE distributions, we will parameterise both distributions so they both have a variance of 1.0. We can then compare the normal distribution $N(0,1)$ and the DE distribution $DE(\alpha,\beta) = DE(0,1/\sqrt{2})$. The DE distribution differs from the normal distribution mainly in that it has heavier concentration close to the mean, and heavier tails as well.

Proportion of the distribution between $\mu \pm k(\sigma/\mu)\mu$

K	Normal	Double Exponential
1	0.6827	0.7569
2	0.9545	0.9409
3	0.9973	0.9856
4	0.9999	0.9965

The tabulated values for the DE distribution is $g(k) = 1.0 - \exp(-k/\beta)$, where $\beta = 1/\sqrt{2}$.

At the AFSC, we use both CV and APE, but recognise that CV has greater meaning (i.e. it is easier to interpret) in broader statistical contexts.

Looking at cross-tabulations

Perhaps the most informative look at reader–tester data is simply to cross-tabulate the ages so that the *x*-axis is reader age and the *y*-axis is tester age. From such a table, it can be seen precisely where the reader and tester differ in their age determination. For example, several specimens that were aged 10 years by the tester might be aged 11 years and 12 years by the reader. This cross-tabulation is perhaps the single most important tool for comparing between-reader ages.

Although reader precision says a lot about reader performance, it does not say much about between-reader biases. Experiences with age-structured stock assessment models indicate that age biases can be much more detrimental to model performance than low precision. Although cross-tabulation plots illustrate biases, how can this plot be interpreted? There are two ways that appear preferable.

Cross-tabulations with the tester as authority

Under this interpretation the tester age is a proxy for true age. The reader's ages are expected to be symmetrically distributed about the tester age. For example, if the tester age is 5 years, the reader's ages of these same specimens should ideally have a bell-shaped distribution around the 5-year age (Table 2*a*). This interpretation has been the usual one at the AFSC. The irony here is that, as shown below, the reader can appear biased relative to the tester using this interpretation, even if both age readers are 'statistically' ageing in exactly the same way (see the following section).

Cross-tabulations with the reader and tester as equal partners (Bowker's test)

Consider the cross-tabulation of ages presented in Table 2*b*. In this table, given the same nominal age class of reader, the age distributions of the tester is identical. For example, if the nominal age class is given to be 5 years, the age distribution of the 5th row or 5th column is the same for both age readers. Bowker's test is designed to test for departures from this type of symmetry. However, note that this type symmetry does not necessarily provide 'good' results for the criteria when tester is the standard as in Table 2*a*.

Hoenig *et al*. (1995) may have been first to suggest that the Bowker's test (see Bishop *et al*. 1975) for symmetry could be used to determine whether bias exists between age readers. If the test statistic was significant, this would indicate that reader and tester are interpreting otolith ages differently. The Bowker test statistic is

Table 2. Hypothetical cross tabulation of tester and reader ages where (*a*) tester is the standard and reader ages are expected to by symmetric about the tester age (Bowker's $\chi^2 = 6.01$, d.f. = 6), or (*b*) tester and reader are considered equal partners and tester and reader are expected to have the same age distribution at each nominal age (Bowker's $\chi^2 = 0$, d.f. = 6)

(*a*) Tester as standard

		Reader						
	Age	1	2	3	4	5	6	7
Tester	1	**2**	1					
	2	3	**5**	3				
	3		4	**7**	4			
	4			6	**7**	6		
	5				2	**9**	2	
	6					4	**5**	4
	7						1	**2**

(*b*) Reader and tester as equal partners

		Reader						
	Age	1	2	3	4	5	6	7
Tester	1	**2**	10					
	2	10	**5**	6				
	3		6	**7**	2			
	4			2	**7**	10		
	5				10	**9**	1	
	6					1	**5**	6
	7						6	**2**

$\chi^2 = \sum_{i=1}^{m} \sum_{j=i+1}^{m} \frac{(n_{ij} - n_{ji})^2}{(n_{ij} + n_{ji})^2}$, where n_{ij} is the number of specimens aged *i* by the tester and *j* by the reader (here n_{ji} is the number of specimens aged *j* by tester and *i* by the reader), and *m* is the maximum age. The summation is made over all terms in which the denominator in the χ^2 expression is greater than zero. If there are *r* such terms then d.f. = *r*. Birnbaum (1962) describes how the χ^2 distribution can be approximated by the normal deviate $Z = \sqrt{2\chi^2} - \sqrt{2r-1}$ so that the approximate significance of Bowker's test can be seen without reference to the χ^2 degrees of freedom.

It should be clear that χ^2 will be small when n_{ij} is approximately n_{ji}, for all i, j (Table 2*b*). What plausible hypothesis concerning ageing might this be testing? Suppose the sample consists of $N = \sum_k n_k$ fish, where each sample of n_k fish are of true age *k*. Suppose the age distribution generated by age readers, when ageing age *k* fish are $\{t_{ki}\}$ for the tester and $\{r_{kj}\}$ for the reader, $\sum_i t_{ki} = 1$ and $\sum_j r_{kj} = 1$. Then the expected values for $E(n_{ij}) = \sum_k n_k t_{ki} r_{kj}$. Therefore, if $t_{ki} = r_{ki}$ for all *k*, *i*, it follows that $n_{ij} \approx n_{ji}$ and Bowker's test will tend to be non-significant. Therefore, the hypothesis that is being tested by Bowker's test is that the age distribution, generated at each true age *k*, is the same for both age readers. The age readers can be biased from the true age, but as long as they age similarly, Bowker's test will tend to be non-significant.

Because of the global nature of Bowker's test, one would think that it would tend to be statistically significant a lot of the time. However, the test statistic may tend not to be significant in small sample sizes, or situations where the age range is broad. In small samples the χ^2 approximation to the distribution of the test statistic may be poor, and when the age range is wide, the degrees of freedom will tend to be large and may dilute the significance of the test statistic. There may be reader–tester discrepancies that are apparent in the cross-tabulations that are not significant according to Bowker's test. For this reason Bowker's test is a convenient index of overall similarity between reader and tester (i.e. the similarity between $\{t_{ki}\}$ and $\{r_{kj}\}$), but may not be a sufficient guide for determining differences in individual small samples, or when age ranges are broad.

Results

We analysed 1990–2003 precision data generated at the Alaska Fisheries Science Center (Table 3). These are the years following our move from large mainframe computers to small desktop computers. During this time period we aged ~22 species, some in very large numbers. We grouped them into three categories: high volume round fishes (the first five species), rockfishes of the genus *Sebastes* (species 6–10), and flatfishes (species 11–22). Species within each group are in the order of increasing percentage agreement.

An examination of percentage agreement (Table 3) shows each category contains difficult species, but rockfish appear to be the most difficult. This is due to the fact that rockfish tend to be older and older specimens will have a lower PA. The coefficient of variation (Table 3), compensates for this greater age, and all species groups appear to have a similar range of CV. However, some species within each group are seen to be inherently more difficult than others. More dramatically, we can plot PA (Fig. 2), and CV against average age (Fig. 3). Clearly the effect of average age can be seen in Fig. 2, but not Fig. 3.

Plotting APE against CV (Fig. 4) shows the theoretical result $CV = \sqrt{2} \times APE$ that was established in theory and derived as exact when only one reader and tester observation was made on each specimen.

Discussion

The Alaska Fisheries Science Center takes an extremely disciplined approach towards production ageing. This approach involves carefully outlined steps concerning how ages are generated by a reader, how a test sample is randomly selected for the tester, how discrepancies are generated and resolved and when precision statistics and growth outliers are to be examined. This method of quality control requires a great commitment of time and energy, but helps assure that ageing criteria are applied as consistently as possible.

Examining reader–tester results may indicate if a particular sample or age reader is problematic and requires more attention (e.g. re-examining ageing criteria or re-ageing samples). Analysing recent or historic samples provides information concerning the status of ageing skills among readers ageing a new species, and provides a quantitative basis for describing relatively easy *v.* difficult species. For a more detailed look at differences in ageing criteria between reader and tester, we use the more subjective cross-tabulation plot that is especially useful for pinpointing where ageing biases are occurring. When severe bias is occurring, the reader may be asked to re-age portions of the sample.

Table 3. Precision statistics from 1990 to 2003 from the Alaska Fisheries Science Center's Age and Growth Program
Percentage agreement is ±0. Groupings are for roundfish (species 1–5), rockfish (species 6–10), and flatfish (species 11–22)

	Scientific name	Common name	Number aged	Number tested	Percentage tested	Average age	Percentage agreement	Percentage CV	Percentage APE
1	*Anoplopoma fimbria*	Sablefish (black cod)	27 872	5977	21.4	10.77	37.3	11.01	7.79
2	*Gadus macrocephalus*	Pacific cod	14 000	5035	36.0	5.00	59.9	7.58	5.36
3	*Theragra chalcogramma*	Walleye pollock	158 850	45 544	28.7	6.52	66.8	4.36	3.09
4	*Merluccius productus*	Pacific hake	28 219	9360	33.2	6.35	74.4	4.65	3.29
5	*Pleurogrammus monopterygius*	Atka mackerel	15 437	3900	25.3	4.17	83.8	2.78	1.96
6	*Sebastes aleutianus*	Rougheye rockfish	239	235	98.3	19.45	22.6	9.87	6.98
7	*Sebastes alutus*	Pacific ocean perch	19 061	6334	33.2	12.64	39.9	6.66	4.71
8	*Sebastes ciliatus*	Dark dusky rockfish	922	345	37.4	23.61	44.3	3.57	2.52
9	*Sebastes polyspinis*	Northern rockfish	7329	1746	23.8	15.46	45.5	4.39	3.11
10	*Sebastes* sp.	Light dusky rockfish	1315	383	29.1	13.22	52.0	3.55	2.51
11	*Microstomus pacificus*	Dover sole	2148	495	23.0	17.37	25.3	9.64	6.82
12	*Reinhardtius hippoglossoides*	Greenland turbot	169	55	32.5	12.01	29.1	6.71	4.75
13	*Hippoglossoides elassodon*	Flathead sole	6478	1525	23.5	10.38	38.8	7.23	5.11
14	*Atheresthes stomias*	Arrowtooth flounder	2753	614	22.3	6.80	40.4	11.69	8.27
15	*Atheresthes evermanni*	Kamchatka flounder	120	24	20.0	8.93	41.7	9.23	6.53
16	*Glyptocephalus zachirus*	Rex sole	1869	404	21.6	7.55	49.0	8.49	6.00
17	*Lepidopsetta bilineata*	Southern rock sole	471	101	21.4	10.16	53.5	5.43	3.84
18	*Limanda aspera*	Yellowfin sole	17 202	4550	26.5	12.44	62.9	2.88	2.04
19	*Pleuronectes quadrituberculatus*	Alaska plaice	3082	663	21.5	13.29	62.9	2.76	1.95
20	*Lepidopsetta bilineata*	Rock sole	8062	1750	21.7	8.33	68.6	3.25	2.30
21	*Lepidopsetta peracuata*	Northern rock sole	5486	1976	36.0	10.50	69.5	2.52	1.78
22	*Hippoglossoides robustus*	Bering flounder	237	48	20.3	8.61	79.2	1.94	1.37

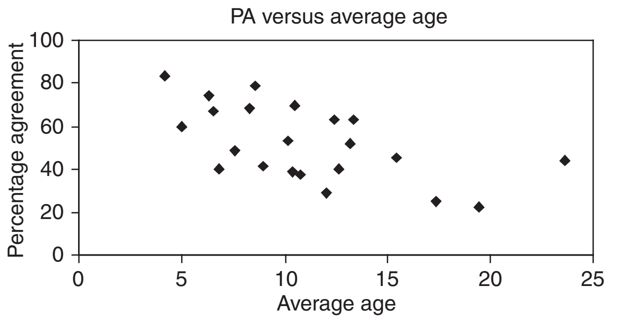

Fig. 2. Plot of average age *v.* percentage agreement (PA) showing that percentage agreement declines as average age increases.

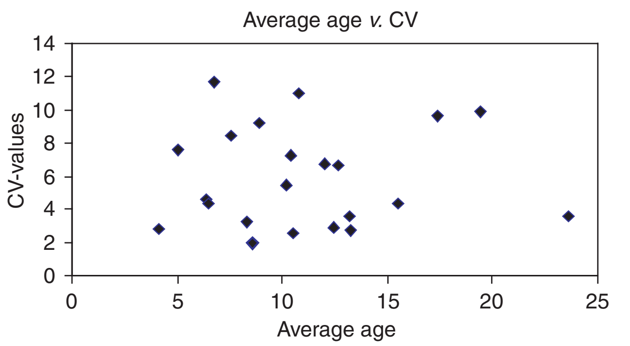

Fig. 3. Plot of average age *v.* coefficient of variation (CV) showing that CV does not vary as average age increases.

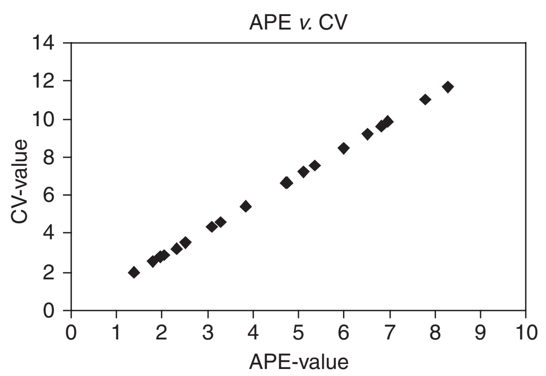

Fig. 4. Plot of coefficient of variation (CV) *v.* average percentage error (APE) showing the relationship $CV = \sqrt{2} \times APE$.

Campana (2001) noted that CV was 41% larger than APE, which is virtually identical to our conclusion that $CV = \sqrt{2} \times APE$. Our analysis emphasises that this difference is due to implicit statistical assumptions underlying these estimates.

In interpreting cross-tabulated data, we often use the tester as the authority method described in the text. Bowker's test of symmetry seems to challenge this approach because one interpretation of Bowker's test indicates that reader and tester can appear significantly different (when the tester is selected as the authority) even when age readers are ageing with statistically similar criteria. The best interpretation may depend on the relative experience levels of the reader and tester. More experience with Bowker's test is needed before we can resolve the dilemma between 'tester as authority' and 'reader and tester as equal partners' views of cross-tabulated data. Bowker's symmetry test seems useful for gauging the amount of between-reader differences in ageing criteria in large samples, but seems less useful for detecting biases in smaller samples.

In small samples, we are often interested in the growth characteristics in particular sets of otoliths. We may wish to examine these under the dual-headed scope even if statistically significant differences between reader and tester have not been detected. This has been our approach to precision testing at the AFSC.

Acknowledgments

We thank Betty Goetz, Craig Kastelle, and two anonymous reviewers for their helpful comments.

References

Beamish, R. J., and Fournier, D. A. (1981). A method for comparing the precision of a set of age determinations. *Canadian Journal of Fisheries and Aquatic Sciences* **38**, 982–983.

Birnbaum, Z. W. (1962). 'Introduction to Probability and Mathematical Statistics.' (Harper and Brothers: New York.)

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). 'Discrete Multivariate Analysis.' (The MIT Press: Cambridge, MA.)

Campana, S. E. (2001). Accuracy, precision, and quality control in age determination, including review of the use and abuse of age validation methods. *Journal of Fish Biology* **59**, 197–242.

Chang, W. Y. B. (1982). A statistical method for evaluating the reproducibility of age determination. *Canadian Journal of Fisheries and Aquatic Sciences* **39**, 1208–1210.

Heifetz, J., Anderl, D., Maloney, N. E., and Rutecki, T. L. (1998). Age validation and analysis of ageing error from marked and recaptured sablefish, *Anoplopoma fimbria. U.S. Fishery Bulletin* **97**, 256–263.

Hoenig, J. M., Morgan, M. J., and Brown, C. A. (1995). Analysing differences between two age determination methods by testing symmetry. *Canadian Journal of Fisheries and Aquatic Sciences* **52**, 364–368.

Kastelle, C. R., Kimura, D. K., Nevissi, A. E., and Gunderson, D. R. (1994). Using Pb-210/Ra-226 disequilibria for sablefish (*Anoplopoma fimbria*) age validation. *U.S. Fishery Bulletin* **92**, 292–301.

Kastelle, C. R., Kimura, D. K., and Jay, S. R. (2000). Using Pb-210/Ra-226 disequilibrium to validate conventional ages in Scorpaenids (genera *Sebastes* and *Sebastolobus*). In '2nd International Symposium on fish Otolith Research & Application, Bergen, Norway, 20–25 June 1998'. *Fisheries Research* **46**, 299–312. doi:10.1016/S0165-7836(00)00155-7

Kendall, M. G., and Stuart, A. (1973). 'The Advanced Theory of Statistics. Vol. 2.' (Hafner Publishing: New York.)

Kimura, D. K., and Lyons, J. J. (1991). Between-reader bias and variability in the age determination process. *U.S. Fishery Bulletin* **89**, 53–60.

Roberson, N. R., Kimura, D. K., Gunderson, D. R., and Shimada, A. M. (2005). Indirect validation of the age reading method for Pacific cod, *Gadus macrocephalus*, using otoliths from marked and recaptured fish. *U.S. Fishery Bulletin* **103**, 153–160.

Manuscript received 12 July 2004; revised 5 January 2005; and accepted 2 March 2005.

Appendix 1. Age and otolith descriptions (i.e. codes) entered by age readers into the AGEDATA database

Age method code (structure preparation codes)

Method	Code
Surface age	(empty)
Break and burn	B
Cut and stain	C
Oven bake and flame burn	M
Thin section	T
Unburnt cut or 'snap'	U
Break and oven bake	V

Readability codes

Code	Description
1	Clear otolith
2	A single age can be generated with variable level of confidence
3	Very difficult; can only assign minimum age or age range
4	Unageable due to problems in interpretation of the growth patterns such as extremely faint or checky. No age range can be assigned
5	Unageable due to physical characteristics of the otolith such as crystallisation, chalkiness or broken into unageable fragments
6	Unageable due to reasons not related to otolith condition such as collector error, more than 2 otoliths in a vial, otolith pair collected from different fish, otolith size not appropriate for recorded fish length, etc. Ages may be assigned but associated biological data is in question

Note: The last 3 codes describe situations where Readers assigned an age of '−1' (unageable).

Edge type codes

Code	Description
0	Strong annulus on the otolith edge
1	Strong annulus with halenation (slight halo of growth)
2	Up to 1/4 of marginal opaque growth as compared to previously deposited opaque zone
3	Up to 1/2 of marginal opaque growth as compared to previously deposited opaque zone
4	Full year of marginal opaque growth as compared to previously deposited opaque zone
5	Full year of marginal opaque growth with an annulus appearing to form along the otolith margins