CSIRO PUBLISHING

www.publish.csiro.au/journals/mfr *Marine and Freshwater Research*, 2006, **57**, 323–332

Corroborating the ages of walleye pollock (*Theragra chalcogramma*)

Daniel K. Kimura[A,C], *Craig R. Kastelle*[A], *Betty J. Goetz*[A], *Christopher M. Gburski*[A] *and Alexander V. Buslov*[B]

[A]Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, 7600 Sand Point Way N.E., Seattle, WA 98115-6349, USA.
[B]Kamchatka Research Institute of Fisheries and Oceanography, Kamchatniro, 18 Naberezhnaya, Petropavlovsk-Kamchatsky, Russia 683602.
[C]Corresponding author. Email: dan.kimura@noaa.gov

Abstract. Fish ageing researchers have long recognised the importance of validating age-reading methodologies. The strongest age validations require the acquisition of ageing structures from fish of known-ages, or specimens whose ages are appropriate for bomb carbon validation. Often such specimens are extremely difficult or impossible to acquire so researchers have sought alternatives to validation. The alternative to age validation is age corroboration. Corroboration of a fish ageing method occurs when fish ages are found to be consistent with some ancillary information when comparisons are made in an unbiased manner. The question pursued in this study is how desirable are such comparisons from a scientific viewpoint. Information is presented that corroborates otolith ages for walleye pollock (*Theragra chalcogramma*), one of the largest groundfish fisheries in the world. Walleye pollock ages were corroborated using marginal increment analysis, ages following the strong 1978 year class in the eastern Bering Sea, and a comparison of ages read from otoliths with ages read from vertebrae. A new statistical method is suggested for comparing otolith and vertebra age readings. The walleye pollock example demonstrated that corroborating evidence can improve confidence in fish ages and ageing techniques.

Extra keywords: age corroboration, age determination, age validation.

Introduction

Fishery biologists are well aware that fish ageing methodologies need to be validated (e.g. Beamish and McFarlane 1983). Age validation generally refers to confirming a method of age determination using fish of known ages (see Campana 2001). Validating fish age determination with any method can be extremely difficult. There are species where it is impossible to acquire known-age specimens, a species may be too delicate to survive tagging, specimens from the era of bomb carbon (^{14}C) increase may be non-existent, fish may be too old to be validated using $^{210}Pb/^{226}Ra$ radiometric validation, or marginal increment analysis and daily growth rings may be applicable only to younger age classes.

Under these circumstances should age researchers pursue corroboration of fish age determination methods? In the Webster's New World Dictionary (Guralnik 1972) 'corroborate' is defined as: (*1*) to support, (*2*) to make the validity more certain; confirm; bolster; support.

In the publication of the First International Symposium on Fish Otoliths, a glossary developed by Kalish *et al*. (1995) was presented. They defined 'corroboration' and 'validation' as follows:

(*1*) *Corroboration* – a measure of the consistency or repeatability of an age determination method. For example, if two different readers agree on the number of zones present in a hard part, or if two different age estimation structures are interpreted as having the same number of zones, corroboration (but not validation) has been accomplished. The term verification has been used in a similar sense; however, the term corroboration is preferred as verification implies that the age estimates were confirmed as true.

(*2*) *Validation* – the process of estimating the accuracy of an age estimation method. The concept of validation is one of degree and should not be considered in absolute terms. If the method involves counting zones, then part of the validation process involves confirming the temporal meaning of the zones being counted. Validation of an age estimation procedure indicates that the method is sound and based on fact.

 10.1071/MF05132 1323-1650/06/030323

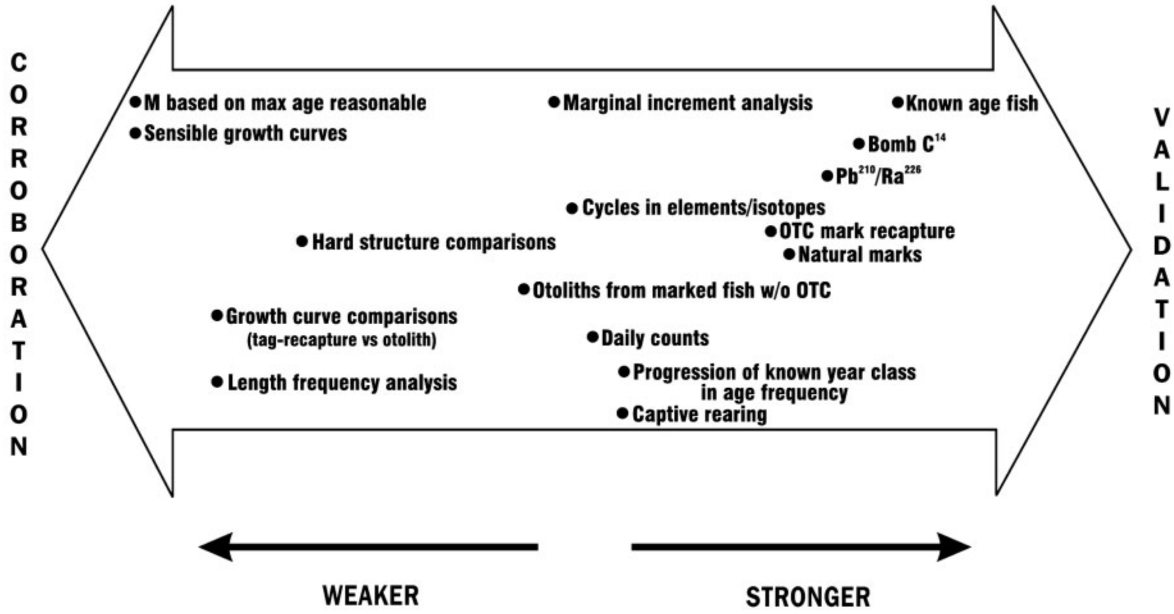


Fig. 1. This diagram describes how age corroboration and validation can be thought of as a continuum providing stronger or weaker evidence supporting the fish age determination criteria for a particular species. The placements of methodologies on this continuum (●) are meant only to be approximate because various researchers would likely want to place them differently.

These definitions have a degree of intended ambiguity that imply some overlap in their meanings. The term 'corroboration' includes the notion of consistency that seems to be a broader term than mere repeatability. If a 'validation' is not absolute, might it be corroborative?

This suggests that corroboration and validation really constitute a continuum of evidence supporting the ageing criteria being used on a particular species (Fig. 1). Not only does the strength of corroboration–validation methods differ intrinsically (i.e. by method), the quality of the study may vary owing to a myriad of factors including the quality of data or the experimental design. The result is that the evidence may vary greatly concerning how well the fish ageing criteria are being supported even if the corroboration–validation method being applied is nominally the same.

In fish age determination research, age corroboration is sometimes not viewed favourably because it is interpreted as a flawed attempt to validate ageing methodologies. Our feeling is that corroboration is a desirable goal in any field of science. We are always searching for 'corroboration' or consistency in theory or data, and it is often more important when we *don't* find it (i.e. when we reject the null hypothesis). This would mean our working assumptions are incorrect and should be changed. Especially in fisheries science where error bars are often large, and data sets difficult to interpret, the concept of corroboration seems especially important.

Walleye pollock (*Theragra chalcogramma*) are routinely aged at the Alaska Fisheries Science Center (AFSC) using a mixture of both whole otolith and cut-and-burn ageing methods (∼60% are aged using cut and burn). Munk (2001) suggested that the AFSC may be under ageing this species by not counting the finer marks. In this paper we present corroborative evidence that supports the AFSC ageing technique using marginal increment analysis, the unusually strong 1978 year class, and ages read from vertebrae. Our goal is to use the corroborative nature of these three lines of evidence and evaluate their combined usefulness in the context of questions regarding the validity of our walleye pollock ages. In addition to the age corroborations presented here, the AFSC is concurrently pursuing radiometric age validation using $^{210}Pb/^{226}Ra$.

Materials and methods

Age corroboration using marginal increment analysis (1989–2002)

Marginal increment analysis documents the seasonal nature of otolith growth by determining when what is believed to be the faster, summer opaque growth, and the slower winter growth actually occurs. It is this slower winter growth that results in a winter translucent zone. To do this we used 1989–2002 edge type data for Alaska walleye pollock (*Theragra chalcogramma*) collected by an experienced age reader. We categorised these data into four edge categories: (*1*) a full increment of opaque growth or a translucent zone on the edge; (*2*) slight opaque edge growth beyond the last translucent zone; (*3*) an opaque edge with up to 1/4 to 1/2 the opaque growth of the previous opaque increment; and (*4*) an opaque edge with 1/2 to a full year's opaque growth on the edge. By graphing the proportion of otoliths in category 1 and category 3, by month, the pattern of annual growth became apparent.

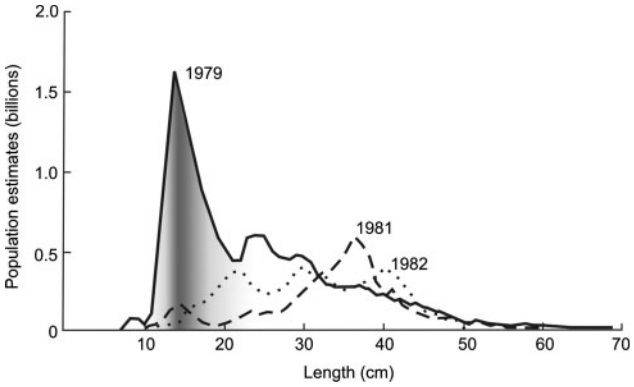


Fig. 2. Samples from the bottom trawl survey in the eastern Bering Sea in 1979 showed a strong length mode for the 1-year-old 1978 year class (shaded). In comparison, 1-year-old recruitment is much weaker in length frequencies sampled in 1981 and 1982.

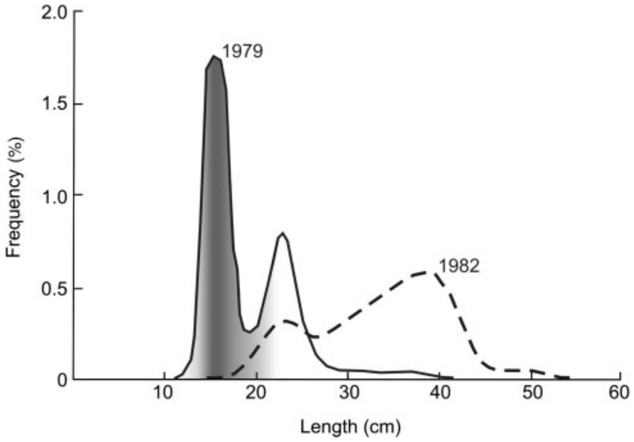


Fig. 3. Samples from the acoustic-midwater trawl survey in the eastern Bering Sea shelf and slope in 1979 also showed a strong length–frequency mode for the 1-year-old 1978 year class (shaded). In comparison, 1-year-old recruitment is much weaker in 1982 length frequencies.

Because of the subjective nature of marginal increment analysis, Campana (2001) calls this method of age validation 'one of the least rigorous methods'. Therefore, we consider this somewhat of a corroboration rather than a validation of the AFSC walleye pollock ageing methodology (Fig. 1).

Age corroboration using the strong 1978 year class

The essence of fish age determination corroboration using a strong year class is that if strong year classes can be followed over time (i.e. advancing annually by one year), this would support that the ageing criteria being used is providing accurate age estimates. The AFSC has extensively sampled the early life history stages of walleye pollock (Brown and Bailey 1992). Extrapolation of 8–10 cm juveniles at age 150 days (in August–September) implies 1 year olds should be around 15 cm by the time of the next summer survey (June–July). The bottom trawl and hydroacoustic-midwater trawl surveys in 1979 provide evidence, independent of age data, of the strength of the strong 1978 year class (Figs 2 and 3) (Bakkala and Wespestad 1983). The 1978 year class in the eastern Bering Sea is perhaps the strongest year class of walleye pollock on record (Ianelli *et al*. 2002). These surveys also indicated that the 1979–1982 year classes were weak. The tracking of this strong 1978 year class in eastern Bering Sea bottom trawl surveys, over ten years, is our second way of corroborating walleye pollock ages.

Age corroboration by comparing otolith ages with vertebral ages

Vertebral ages have been widely used for ageing elasmobranches (see e.g. Cailliet *et al*. 1983; Prince and Pulos 1983; Moulton *et al*. 1992; Wintner and Cliff 1999), because elasmobranches do not have otoliths that can be used for ageing. The vertebral age of a leopard shark (*Triakis semifasciata*) was validated using oxytetracycline (OTC) marks to an age over 20 years (Smith *et al*. 2003). In teleosts, however, otoliths have been generally preferred over vertebrae, with occasional exceptions such as fugu (*Takifugu vermicularis*) (Matsui *et al*. 1987), whiting (*Merlangius merlangus euxinus*) (Polat and Gümüs 1996), shad (*Alosa pontica*) (Yilmaz and Polat 2002) and Atlantic bluefin tuna (*Thunnus thynnus*) (Mather and Schuck 1960; Caddy and Butler 1976; Lee *et al*. 1983). This is largely because otoliths are generally easier to excise, store, prepare, and age. Comparing vertebrae and otoliths is our third way to corroborate fish otolith ages.

The obvious basis of age corroboration using an additional hard structure is that if both structures provide similar ages, then the ageing criteria used on each structure is corroborated. In 2002, samples of vertebrae and otoliths from walleye pollock from areas of the USA Aleutian Islands ($n = 88$), the eastern portion of the Okhotsk Sea ($n = 97$), and the western Bering Sea ($n = 27$) were collected. Age structures from these samples were randomised so that otolith age was not known when reading vertebrae. Two age readers made independent readings of otoliths and vertebrae. Reader 1 was experienced in ageing walleye pollock using both otoliths and vertebrae. Reader 2 was extremely experienced in ageing walleye pollock using otoliths, but had never before aged vertebrae. The data set we generated consisted of four age readings: one using mostly otolith cut and burns (Chilton and Beamish 1982), and one using vertebrae (cleaned and read from the whole centrum), from each of the two readers.

The analysis of results consisted of the usual cross-tabulation and precision statistics (see Kimura and Lyons 1991; Campana 2001). In addition, we offer a new method of analysis based on the observation that if two sets of age readings $\{x_i, y_i\}$ compare well, then an X, Y plot should fall on the equal value line and the sum of squared residuals (SSR), $SSR = \sum_{i=1}^{n}(x_i - y_i)^2$, should be small. Assuming observations are normally distributed and unbiased, it is known that this sum of squares, scaled by the true variance, will have a chi-square distribution with n degrees of freedom (see Rao 1973). From there it is an easy step to derive an F-statistic for comparing within-structure precision between vertebra and otolith age readings, and a modified F-statistic to test whether vertebra ages are significantly different from otolith ages (see Appendix 1).

Results

Results of age corroboration using marginal increment analysis (1989–2002)

Using marginal increment analysis, the seasonal growth in the walleye pollock otolith was readily apparent (Table 1, Fig. 4). The predominance of the completed or near complete opaque zone (category 1) was clear for the months January through to March. It was also clear that during the summer (June–August) category 3, which was the category of substantial but incomplete opaque growth, predominates. During September and October, the expected completion of the current opaque zone occurred, setting the stage for the completed zones seen

Table 1. Proportion of otolith specimens having each of four edge category types
These categories are: 1, a full increment of opaque growth or a translucent zone on the edge; 2, slight opaque edge growth beyond the last translucent zone; 3, an opaque edge with up to 1/4 to 1/2 the opaque growth of the previous opaque increment; and 4, an opaque edge with 1/2 to one full-year's opaque growth on the edge

Month	Sample size	Edge category			
		1	2	3	4
Jan.	214	0.897	0.005	0.047	0.051
Feb.	537	0.885	0.019	0.039	0.058
March	717	0.852	0.038	0.086	0.024
April	0				
May	0				
June	436	0.275	0.055	0.656	0.014
July	1139	0.207	0.043	0.711	0.039
Aug.	860	0.259	0.016	0.615	0.109
Sept.	843	0.396	0.007	0.446	0.151
Oct.	505	0.592	0.004	0.281	0.123
Nov.	0				
Dec.	0				

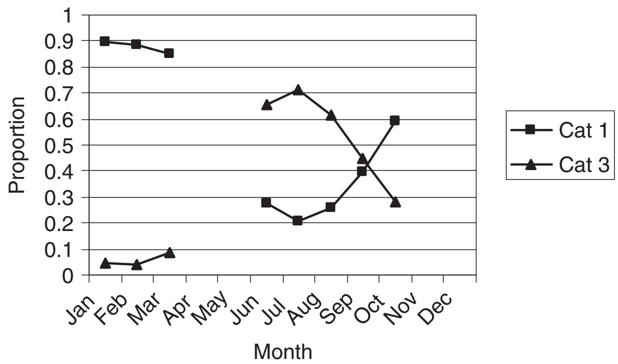


Fig. 4. Plot showing edge category by month as proportions of otolith specimens having a category of 1 or 3. Category 1 otoliths have a full increment of opaque growth or a translucent zone on the edge. Category 3 otoliths have an opaque edge with up to 1/4 to 1/2 the opaque growth of the previous opaque increment.

in winter. Edge categories 2 and 4 were transition categories that were most ambiguous.

Results of age corroboration using the strong 1978 year class

The age frequency samples from the eastern Bering Sea bottom trawl survey from 1982 to 1991 showed the predominance of the 1978 year class (Fig. 5). The progression of this year class was evident from age 4 to 13 years. In interpreting these data, note that age samples were length stratified on a vessel basis so that the tails of the distribution were over sampled.

Results of age corroboration by comparing otolith ages with vertebral ages

A comparison of ages read from 212 specimens aged by two age readers (subscripted 1 and 2) using both otoliths (O) and vertebrae (V) clearly showed the similarity of otolith ages and vertebral ages taken from the same fish. In what follows the sum on specimen i in the SSR equation is made implicit, and the variables x, y are replaced by variables O_1, O_2, V_1, V_2 referring to structure and reader. The SSR and mean square error (MSE, see Appendix 1) for ages from different age structures and age readers generally showed which pairs of ages compared the best (Table 2). From rows 2, 4, and 5, it was evident V_2 ages compared poorest with the other ages. This was not surprising because reader 2 had never aged vertebra before the present study.

Precision statistics from rows 1 and 2 (Table 2) show that otolith ages were more repeatable than vertebral ages when comparisons were made between readers. Because both age readers were experienced with ageing from otoliths, the agreement in Table 2 (O_1 *v.* O_2) of 60.0% (± 0) and 92.4% (± 1) was quite good. For vertebrae, the agreement in Table 2 (V_1 *v.* V_2) was less than that for otoliths with 48.3% (± 0) and 86.1% (± 1). The F-statistics described in Appendix 1, comparing within-structure precision, can be used to compare precision between otolith readings, and precision between vertebral readings, from readers 1 and 2. Using the F-statistic calculated from the MSEs in rows 1 and 2, $\hat{F} = 1.938/0.824 = 2.35$ with d.f. $= 209, 210$. This F-statistic was highly significant ($\alpha = 0.001$) so we concluded that vertebral ages are not as repeatable as otolith ages between these readers.

Cross-tabulations and precision statistics based on comparisons O_1 *v.* V_1, and O_2 *v.* V_2 (Tables 2–4) showed a

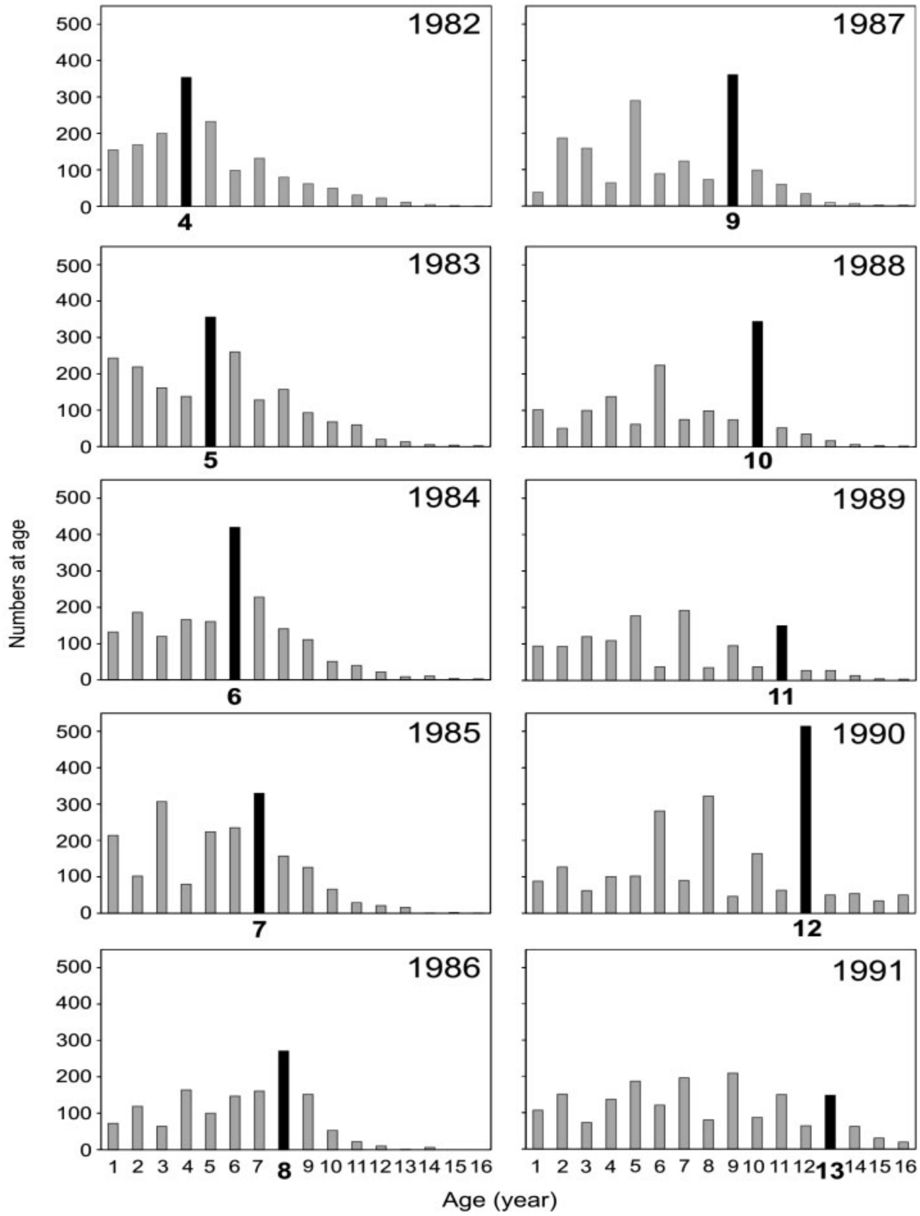


Fig. 5. A series of annual age distribution histograms from the eastern Bering Sea bottom trawl surveys showing progression of the strong 1978 year class clearly visible from age 4 years to age 13 years.

reasonable relationship between otolith ages and vertebral ages within the same reader. In this notation, the first variable was the row variable (e.g. in Table 3, the row variable was the otolith age from reader 1). For comparison of otolith and vertebral ages within readers, both readers had agreement statistics similar to the vertebra-to-vertebra statistics (Table 2, rows 2, 3 and 4). For reader 1, O_1 *v.* V_1 was 57.2% (± 0) and 87.5% (± 1). For reader 2, O_2 *v.* V_2 was 48.3% (± 0) and 77.8% (± 1). For comparing otolith ages and vertebral ages within reader 1, a modified F-statistic (see Appendix 1) was calculated using MSEs from rows 1 and 3, $\hat{F} = 1.389/0.824 = 1.69$ with d.f. $= 208, 210$ indicating that reader 1 vertebral ages were significantly different ($\alpha = 0.001$) from reader 1 otolith ages.

Results of this analysis indicated that while vertebral ages were statistically different from otolith ages, the ages were generally in a similar range and showed little relative bias (Tables 3 and 4). Comparing positive and negative residuals between otolith and vertebral ages using a binomial model showed neither reader had a significant excess of either residuals ($\alpha = 0.05$). Cross-tabulations in Tables 3 and 4 indicated that the growth marks found on otoliths and vertebrae were

Table 2. Sum of squared residuals (SSR) and mean square error (MSE) used to compare various age types (V = vertebra, O = otolith) and age readers (subscripts 1 and 2)
Agreement statistics refer to the two age estimates occurring in the SSR formula

SSR formula	SSR estimate	n	MSE estimate	Percentage agreement ± 0	Percentage agreement ± 1
1 $\sum(O_1 - O_2)^2$	173	210	0.824	60.0	92.4
2 $\sum(V_1 - V_2)^2$	405	209	1.938	48.3	86.1
3 $\sum(O_1 - V_1)^2$	289	208	1.389	57.2	87.5
4 $\sum(O_2 - V_2)^2$	586	207	2.831	48.3	77.8
5 $\sum(O_1 - V_2)^2$	637	208	3.063	49.0	76.4
6 $\sum(O_2 - V_1)^2$	360	207	1.739	49.8	82.6

Table 3. A cross-tabulation of walleye pollock ages generated by reader 1
Row ages were read from otoliths and column ages were read from vertebrae

		VERTEBRA																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total
	1	**25**																				25
	2		**11**	12																		23
	3			**6**	1																	7
	4				**18**	8																26
	5					**15**	6															21
	6				1	2	**3**	2	1	1												10
O	7						3	**15**	3													21
T	8					1		3	**6**	1	1											12
O	9							1	3	**4**	2		1	1								12
L	10							1	1		**6**		1									9
I	11								2	2	2	**1**		1	1							9
T	12												**1**	1	1							3
H	13											1	2	**2**	1							6
	14								1					3	**1**	1						6
	15											1		1		**1**						3
	16											1		1	1	2	**1**					6
	17														1		1	**2**				4
	18																	1	**0**			1
	19																		1	**1**	2	4
	20																				**0**	0
	Total	25	11	18	20	26	12	22	17	8	11	4	5	10	6	4	2	3	1	1	2	208

sufficiently similar so that vertebral ages generally corroborated the ages obtained from otoliths. As age readers get more experienced in ageing vertebrae, it is probable that measures of precision will improve.

Discussion

Campana (2005) noted that fish ageing remains the most common discipline in published otolith studies. The reason is that fishery management needs accurate age data for stock assessment modelling, and many individuals are employed to assist fishery management (in the broad sense) in whatever capacity they can.

This paper brings together two interconnected themes of considerable importance. The first is that the accuracy of fish ages can only be assured by age corroboration and validation studies (Campana 2001). The importance of accurate fish ages to fishery stock assessments has been well established by numerous papers investigating the effects of inaccurate ages (see e.g. Beamish and McFarlane 1983, 1995; Tyler *et al*. 1989; Coggins and Quinn 1998; Reeves 2003). We have argued that age corroboration-validation is best understood as a continuum of methods that can all contribute in varying degrees to understanding the accuracy of ageing criteria. We have also argued that age researchers should pursue age corroboration until stronger age validation studies can be made. Age validation is necessary, but not always possible or practical. Managers and age-reading professionals are often forced to make the best of a situation without strong validation. The corroboration described here is using the best

Table 4. A cross-tabulation of walleye pollock ages generated by reader 2
Row ages were read from otoliths and column ages were read from vertebrae

		VERTEBRA																				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total
	1	**25**																				25
	2		**15**	2	2																	19
	3		2	**9**	3																	14
	4			2	**15**	4	1															22
	5				4	**14**	8															26
	6				3	2	**4**	6	2	1												18
O	7						4	**1**	7	1			1									14
T	8						1	5	**8**	2		1										17
O	9					1	1	2	2	**0**												6
L	10							3	1		**2**		1		1							8
I	11											**1**	1	1	1							4
T	12									1	2		**0**	1				1				5
H	13										1			**3**								4
	14								1						**2**							3
	15											2			2	**0**	1	1				6
	16										1	2	1	1			**1**	1				7
	17												1		1			**0**				2
	18											1					1	1	**0**		1	4
	19															1		1		**0**		2
	20																			1	**0**	1
	Total	25	17	13	27	21	19	17	21	5	6	7	5	6	7	1	3	5	0	1	1	207

information available, not the best information wished for, or recommended. Managers and age-reading professionals worldwide face this problem. Our corroboration of walleye pollock ages was partly initiated because of this situation.

Our second theme was to demonstrate the value of successful age corroboration for the case of Alaska walleye pollock. Walleye pollock in the eastern Bering Sea represent an enormous resource with an age 3+ year biomass of around 11 million tonnes. This stock has sustained average annual catches from 1970 to 2004 of 1.245 million tonnes, and in 2004 catches were 1.332 million tonnes (NPFMC 2005). Munk (2001) believed that ages generated at the AFSC could be two times older than reported. If walleye pollock were miss-aged to this degree, the recommended catch levels would be far below those recommended (Beamish and McFarlane 1995; Reeves 2003). Because of the biological and economic impact that such an ageing error would have on the walleye pollock resource, it became paramount to corroborate the AFSC walleye pollock ageing criteria, until a stronger age validation becomes available.

We presented three lines of evidence in which AFSC walleye pollock ages read from otoliths were corroborated. The first used classic marginal increment analysis that showed that the growth patterns in otoliths were annual. The second corroboration was that ages obtained from readings of survey samples were able to follow an unusually strong 1978 year class. The final corroboration was that ages read from vertebrae were reasonably similar to ages read from otoliths. None of these were strong age validations as described by Campana (2001) because they did not involve the ageing of known-age materials, or the sampling of physical processes directly related to fish age (e.g. radiometric validation or bomb carbon). However, taken together, they seem to indicate that there is reason for confidence in the validity of the walleye pollock ages analysed here.

To go further and answer the question of whether successful tracking of the 1978 walleye pollock year class was simply a result of 'knowledge' that it was there, an additional study can be cited. Kimura *et al.* (1992) re-aged a randomised sample consisting of fish aged 9–11 years from the 1978 year class sampled (surveyed) in 1987–1989. These ages from the 1978 year class were successfully resolved by three age readers.

We feel that the information presented in these studies corroborated AFSC walleye pollock otolith ages in the dictionary sense of 'confirm; bolster; support' or making the 'validity more certain'. We especially feel that these studies fulfilled the Otolith Symposium glossary definition of corroboration (Kalish *et al.* 1995) in that they tested the 'consistency' of otolith ages. Hence our goal appears to be satisfied. There does appear to be value in using a combination of individual corroborations to provide an overall level of confidence.

The dictionary definition of corroboration, including the idea that it makes 'validity more certain' is probably an area that bothers critics. This certainly sounds like 'corroboration' can be used as a type of 'false validation'. Indeed, the fact that age validation is 'not absolute' and that corroboration 'makes validity more certain' means that there is room for confusion.

Therefore, semantics may be one of the real issues here because the definitions of Kalish *et al*. (1995) are overlapping and are not mutually exclusive. Earlier we suggested that corroboration and validation constitute a continuum (Fig. 1) and not a classification. This continuum would be based on the methods used, the quality of the data and the quality of the experimental design.

Campana (2001) stated that, 'Comparison of multiple ageing structures within each fish is also a form of age non-corroboration' because 'consistency among within-fish growth structures is the rule rather than the exception'. In other words, it is irrelevant if the ages from two hard structures compare in age because this is usually the case. However, in the search for corroboration, the researcher may find something even more important, namely non-corroboration. Beamish and McFarlane (1995) describe ageing walleye pollock from the Central Bering Sea using three hard structures: scales, finrays and otoliths. They found that cut-and-burn otolith readings gave ages distinctly older than scales and fin-rays. Any researcher ageing walleye pollock with scales would have found it enlightening to try to corroborate scale ages using otolith cut and burns. Beamish and McFarlane (1995) favoured the older ages from cut and burn otoliths, but did not provide corroborating or validating evidence concerning their accuracy.

Often researchers pursue corroboration–validation when there is disagreement concerning ageing criteria for a species. Depending on the magnitude of age differences resulting from the disputed ageing criteria, corroboration studies can be useful. A corroborative study might be more useful in arguing basic reasonableness of the age data where large differences in ageing criteria are being proposed, rather than finer interpretation of annual zones. Nevertheless, we feel that corroborative studies can contribute greatly to a scientific evaluation of age determination criteria, as demonstrated here for walleye pollock.

Because the ocean environment changes over time and human interpretation that produces fish ages can vary, age validation is never absolute and therefore corroborating age data should be occasionally reexamined. At the AFSC we have found that the difficulty of ageing walleye pollock otoliths collected from Alaska waters have increased over time (probably owing to changes in the ocean environment), while employee turnover has also affected the quality of ages. These factors make it possible that what are thought of as valid ages this year may not be so next year.

It should be understood without saying that the preferred methods of determining the accuracy of ageing methods are those that are the strongest in Fig. 1. Several weak corroborative studies are far less valuable than a single strong validation study. The literature should maintain the distinction between age validation and age corroboration (Campana 2001) that has proved conceptually useful, and highlights methods that should be preferred by researchers. Ageing criteria confirmation is not completed for a species until strong validation methods have been successfully applied.

Acknowledgments

We thank Dr Gregor Cailliet and Mr Allen Andrews of the Moss Landing Marine Laboratories for their thoughtful, insightful, reviews. We also thank Ms Delsa Anderl of the Alaska Fisheries Science Center for several helpful comments.

References

Bakkala, R. G., and Wespestad, V. G. (1983). Walleye pollock. In 'Condition of Groundfish Resources in the Eastern Bering Sea and Aleutian Islands Region in 1982'. (Eds R. G. Bakkala and L. L. Low.) pp. 1–37. *NOAA Technical Memorandum NMFS F/NWC*-**42**.

Beamish, R. J., and McFarlane, G. A. (1983). The forgotten requirement for age validation in fisheries biology. *Transactions of the American Fisheries Society* **112**, 735–743. doi:10.1577/1548-8659(1983)112<735:TFRFAV>2.0.CO;2

Beamish, R. J., and McFarlane, G. A. (1995). A discussion of the importance of aging errors, and an application to walleye pollock: the world's largest fishery. In 'Recent Developments in Fish Otolith Research'. (Eds D. H. Secor, J. M. Dean and S. E. Campana) pp. 545–565. (University of South Carolina Press: Columbia, SC.)

Brown, A. L., and Bailey, K. M. (1992). Otolith analysis of juvenile walleye pollock *Theragra chalcogramma* from the western Gulf of Alaska. *Marine Biology* **112**, 23–30. doi:10.1007/BF00349723

Caddy, J. F., and Butler, M. J. A. (1976). Recent catch trends and age composition in Canadian coastal fisheries for giant bluefin tuna (*Thunnus thynnus*) and their relevance to assessment of the northwest Atlantic large fish stocks. *International Commission for the Conservation of Atlantic Tunas, Collected Volume of Scientific Papers, Madrid* **5**(2), 244–257.

Cailliet, G. M., Martin, L. K., Kusher, D., Wolf, P., and Welden, B. A. (1983). Techniques for enhancing vertebral bands in age estimation of California elasmobranchs. In 'Proceedings of the International Workshop on Age Determination of Oceanic Pelagic Fishes: Tunas, Billfishes, and Sharks'. (Eds E. D. Prince and L. M. Pulos.) pp. 157–165. NOAA *Technical Report NMFS* **8**.

Campana, S. E. (2001). Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *Journal of Fish Biology* **59**, 197–242. doi:10.1111/J.1095-8649.2001.TB00127.X

Campana, S. E. (2005). Otolith science entering the 21st century. *Marine and Freshwater Research* **56**, 485–495. doi:10.1071/MF04147

Chilton, D. E., and Beamish, R. J. (1982). 'Age Determination Methods for Fishes Studied by the Groundfish Program at the Pacific Biological Station.' *Canadian Special Publication of Fisheries and Aquatic Sciences* **60**.

Coggins, L. G., and Quinn II, T. J. (1998). A simulation study of the effects of aging error and sample size on sustained yield estimates. In 'Fishery Stock Assessment Models'. (Eds F. Funk II, T. J. Quinn, J. Heifetz, J. N. Ianelli, J. E. Powers, J. F. Schweigert, P. J. Sullivan, and C.-I. Zhang.) pp. 955–975. Alaska Sea Grant College Program Report No. AK-SG-98–01. (University of Alaska Fairbanks: Fairbanks, AK.)

Guralnik, D. B. (Ed.) (1972). 'Webster's New World Dictionary.' (The World Publishing Company: New York.)

Ianelli, J. N., Barbeaux, S., Honkalehto, T., Walters, G., and Williamson, N. (2002). Eastern Bering Sea walleye pollock stock assessment. In 'Appendix A, Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Eastern Bering Sea/Aleutian Islands Regions'. (Compiled by The Plan Team for the Groundfish Fisheries of the Bering Sea and Aleutian Islands.) pp. 33–120. (North Pacific Fishery Management Council: Anchorage, AK.)

Kalish, J. M., Beamish, R. J., Brothers, E. B., Casselman, J. M., Francis, R. I. C. C., Mosegaard, H., Panfili, J., Prince, E. D., Thresher, R. E., Wilson, C. A., and Wright, P. J. (1995). Glossary for otolith studies. In 'Recent Developments in Fish Otolith Research'. (Eds D. H. Secor, J. M. Dean and S. E. Campana.) pp. 723–729. (University of South Carolina Press: Columbia, SC.)

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

Kimura, D. K., Lyons, J. J., and MacLellan, S. E. (1992). Effects of year-class strength on age determination. In 'Age Determination and Growth in Fish and Other Aquatic Animals'. (Ed. D. C. Smith.) *Australian Journal of Marine and Freshwater Research* **43**, 1221–1228. doi:10.1071/MF9921221

Lee, D. W., Prince, E. D., and Crow, M. E. (1983). Interpretation of growth bands on vertebrae and otoliths of Atlantic bluefin tuna, *Thunnus thynnus*. In 'Proceedings of the International Workshop on Age Determination of Oceanic Pelagic Fishes: Tunas, Billfishes, and Sharks'. (Eds E. D. Prince and L. M. Pulos.) pp. 61–69. *NOAA Technical Report NMFS* **8**.

Mather III, F. J., and Schuck, H. A. (1960). Growth of bluefin tuna of the western North Atlantic. *Fishery Bulletin* **61**, 39–52.

Matsui, S., Fukumoto, K., and Tsukahara, H. (1987). The age and growth of *Takifugu vermicularis. Science Bulletin of the Faculty of Agriculture, Kyushu University* **41**(3–4), 105–110.

Moulton, P. L., Walker, T. I., and Saddlier, S. R. (1992). Age and growth studies of gummy shark, *Mustelus antarcticus* Günther, and school shark, *Galeorhinus galeus* (Linnaeus), from southern Australian waters. In 'Age Determination and Growth in Fish and Other Aquatic Animals'. (Ed. D. C. Smith.) *Australian Journal of Marine and Freshwater Research* **43**, 1241–1267. doi:10.1071/MF9921241

Munk, K. M. (2001). Walleye pollock otolith aging: comparison of the techniques used by two agencies. Regional Information Report No. 5J01–06. Alaska Department of Fish and Game, Juneau, AK.

NPFMC (2005). 'Fishery Management Plan for Groundfish in the Bering Sea and Aleutian Islands Management Area. January, 2005.' (North Pacific Fishery Management Council: Anchorage, AK.)

Polat, N., and Gümüs, A. (1996). Ageing of whiting (*Merlangius merlangus euxinus*, Nord., 1840) based on broken and burnt otoliths. *Fisheries Research* **28**, 231–236. doi:10.1016/0165-7836(96)00475-4

Prince, E. D., and Pulos, L. M. (Eds) (1983). 'Proceedings of the International Workshop on Age Determination of Oceanic Pelagic Fishes: Tunas, Billfishes, and Sharks. U.S. Department of Commerce.' *NOAA Technical Report* NMFS **8**.

Rao, C. R. (1973). 'Linear Statistical Inference and Its Applications.' 2nd edn. (John Wiley and Sons: New York.)

Reeves, S. A. (2003). A simulation study of the implications of age-reading errors for stock assessment and management advice. *ICES Journal of Marine Science* **60**, 314–328. doi:10.1016/S1054-3139(03)00011-0

Smith, S. E., Mitchell, R. A., and Fuller, D. (2003). Age validation of leopard shark (*Triakis semifasciata*) recaptured after 20 years. *Fishery Bulletin* **101**, 194–198.

Tyler, A. V., Beamish, R. J., and McFarlane, G. A. (1989). Implications of age determination errors to yield estimates. In 'Effects of Ocean Variability on Recruitment and an Evaluation of Parameters used in Stock Assessment Models'. (Eds R. J. Beamish and G. A. McFarlane.) pp. 27–35. *Canadian Special Publication in Fisheries and Aquatic Sciences* **108**.

Wintner, S. P., and Cliff, G. (1999). Age and growth determination of the white shark, *Carcharodon carcharias*, from the east coast of South Africa. *Fishery Bulletin* **97**, 153–169.

Yilmaz, S., and Polat, N. (2002). Age determination of shad (*Alosa pontica* Eichwald, 1838) inhabiting the Black Sea. *Turkish Journal of Zoology* **26**, 393–398.

Manuscript received 2 July 2005; revised 18 November 2005; and accepted 17 January 2006.

Appendix 1. Statistical method for comparing two different ageing methods from two different age readers

The observations from the present study are essentially the different ages read from a particular fish by two different age readers using otoliths and vertebrae. These data were analysed using the standard percentage agreement statistic [(*#agree*/*#aged*) × 100]. Also, since ages from a particular fish, using either structure, should ideally lie on the line $y = x$, these data can be nicely summarised in cross-tabulations or on simple plots with the line $y = x$. We use cross-tabulations that avoid the problem of over plotting integer data.

To quantify the differences in the fits to the line of equality, we use the simple sum of squares of residuals, $SSR = \sum(y_i - x_i)^2$; the larger the SSR, the poorer the fit. The mean square error (MSE) is defined as $MSE = \sum(y_i - x_i)^2/n$, where n is the number of fish aged in the comparison. Let the subscripts 1 and 2 refer to the age readers and let O and V refer to otolith and vertebra ages respectively. In what follows the sum on specimen i is made implicit, and the variables x, y are replaced by variables O_1, O_2, V_1, V_2 referring to structure and reader. We can consider three types of SSR:

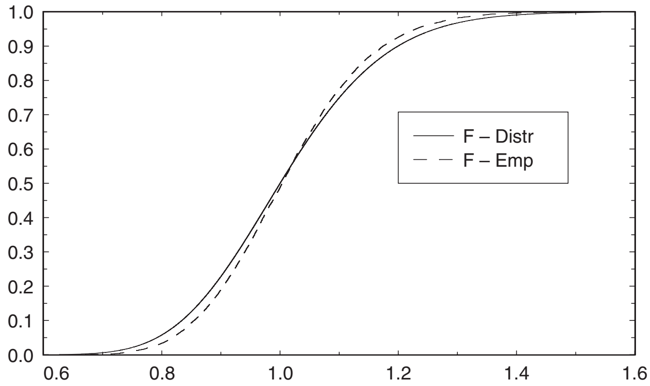
(*1*) between readers: $\sum(O_1 - O_2)^2$ and $\sum(V_1 - V_2)^2$;
(*2*) between ageing structures: $\sum(O_1 - V_1)^2$ and $\sum(O_2 - V_2)^2$; and
(*3*) between readers and ageing structures: $\sum(O_1 - V_2)^2$ and $\sum(O_2 - V_1)^2$.

The third SSR should have the largest values since they should include error introduced by both the age structures and the age readers.

Generally these sums of squares, scaled by the true variance, can be treated as independent chi-square random variables for the purpose of constructing F-tests and testing hypotheses (see Rao 1973). To do this, all we need under the null hypothesis is that each ageing method is unbiased and has an independent random normal error of the same magnitude. If this is true, F-tests such as $F = \frac{\sum(V_1 - V_2)^2/n_v}{\sum(O_1 - O_2)^2/n_o}$ where n fish are aged will have d.f. $= n_v, n_o$. This test can be used to test if vertebra agreement is as good as otolith agreement. If the between vertebra ages variance is greater than between otolith ages variance, then the F-test will tend to be significant.

If we consider the modified F-test $F = \frac{\sum(O_1 - V_1)^2/n_{ov}}{\sum(O_1 - O_2)^2/n_o}$, this test can be used to test whether vertebra ages are significantly different to otolith ages. Because O_1 occurs in the numerator and the denominator, the null distribution does not have the correct central F-distribution. However, simulation using normally distributed deviates (Appendix Fig. 1) indicated that this failure in assumption caused only a modest departure from the central F-distribution under the null hypothesis of no difference in structures. In addition, the change in distribution made the test statistic more conservative. That is, the modified statistic would be even more significant than indicated by the central F-distribution (i.e. the nominal P-values). Alternatively, the significance of the modified test statistic can be estimated from the empirical modified F-distribution (Appendix Fig. 1).

Another possibility is to read another set of independent otolith readings from either age reader so that O_1 does not appear in both the numerator and denominator. In this case it would appear that the standard F-test, and not the modified F-test, could be applied. We should not lose sight that the residual $SSR = \sum(y_i - x_i)^2$ provide a simple and fairly intuitive statistical way of analysing different ageing methods.



Appendix Fig. 1. The cumulative distribution function of the central F-distribution compared with an empirical distribution of a modified F-statistic calculated when the numerator and denominator contain identical age sequences, and $n = 200$ for each sample (see text).