SCRS/2007/102 Collect. Vol. Sci. Pap. ICCAT, 62(3): 899-910 (2008)

AGE ESTIMATION IN CALCIFIED CALCAREOUS STRUCTURES; PRELIMINARY FINDINGS OF AN INTER-LABORATORY COMPARISON

Carys Ann Davies[1], Deirdre Brophy[1], Persefoni Megalofonou[2], Elizabeth Gosling[3], Nigel Griffin[4], Bruno Leroy[5] and Naomi Clear[6]

SUMMARY

Age estimates for albacore tuna are derived using a number of methods including the enumeration of growth increments in spines, scales and otoliths. Inconsistencies in interpretation between structures and between laboratories may introduce error into the age-data available for stock assessment. A comparative age estimation exercise was carried out to assess bias and precision of age estimates between readers and between structures. Six scientists from four laboratories, with varying levels of experience in albacore age estimation participated in the study. Images of the otoliths, scales and spines from 84 albacore tuna collected from the North east Atlantic and the Mediterranean were distributed to each reader. When the age estimates of an experienced reader were compared with those of an inexperienced reader who they had trained, moderate levels of precision were observed (mean % coefficient of variation 11.2, 7.7 and 6.5 for otoliths, scales and spines respectively). However, the wider comparison across laboratories yielded poor precision (% coefficient of variation 35.7, 25.7 and 24.0 for otoliths, scales and spines respectively). Linear bias was detected between readers for all structures, with readers systematically over- or underestimating age relative to each other. Analysis of variance indicates that inter-reader precision is higher for Mediterranean albacore than for Atlantic albacore. Agreement of age estimates derived using different structures from the same fish was low and scales appeared to provide lower estimates of age than spines. The results highlight the need for further inter-calibration of albacore tuna ageing methodologies.

RÉSUMÉ

Les estimations de l'âge du germon sont obtenues à l'aide d'un certain nombre de méthodes, dont l'énumération des augmentations de croissance des épines, des écailles et des otolithes. Des incohérences dans l'interprétation entre les structures et entre les laboratoires pourraient introduire des erreurs dans les données démographiques disponibles pour l'évaluation des stocks. Un exercice comparatif d'estimation de l'âge a été mené à bien afin d'évaluer les biais et la précision des estimations de l'âge entre les lecteurs et entre les structures. Six scientifiques de quatre laboratoires, dotés de divers niveaux d'expérience en matière d'estimation de l'âge du germon, ont participé à l'étude. Chaque lecteur a reçu des images d'otolithes, d'écailles et d'épines de 84 germons recueillis dans l'Atlantique Nord-Est et la Méditerranée. Lorsque les estimations de l'âge d'un lecteur expérimenté étaient comparées avec celles d'un lecteur inexpérimenté qu'ils avaient formé, des niveaux modérés de précision étaient observés (% du coefficient de variation 11,2 pour les otolithes, 7,7 pour les écailles et 6,5 pour les épines). Or, la comparaison plus large réalisée entre les laboratoires a donné lieu à une faible précision (% moyen du coefficient de variation 35,7 pour les otolithes, 25,7 pour les écailles et 24,0 pour les épines). Un biais linéaire a été détecté entre les lecteurs pour toutes les structures, les lecteurs surestimant ou sous-estimant systématiquement l'âge par rapport à chacun. L'analyse des variances indique que la précision entre les lecteurs est plus élevée pour le germon méditerranéen que pour celui de l'Atlantique. La concordance des estimations de l'âge obtenue en utilisant différentes structures du même poisson s'est avéré faible et les écailles paraissaient fournir des estimations de l'âge moins élevées que les épines. Les résultats soulignent la nécessité d'améliorer l'inter-calibrage des méthodologies de détermination de l'âge du germon.

[1] Commercial Fisheries Research Group, Galway-Mayo Institute of Technology, Dublin Road, Galway, Ireland. deirdre.brophy@gmit.ie
[2] Department of Biology, Section of Zoology-Marine Biology, University of Athens, Panepistimiopolis, Ilissia, Athens, Greece
[3] Molecular Ecology Research Group, Department of Life Sciences, Galway-Mayo Institute of Technology, Dublin Road, Galway, Ireland
[4] Irish Sea Fisheries Board, BIM, Galway, Ireland
[5] Secretariat of the Pacific Community, Noumea 98848, New Caledonia
[6] CSIRO Marine and Atmospheric Research, Hobart Tasmania 7001 Australia

RESUMEN

Se derivan estimaciones de edad del atún blanco utilizando varios métodos, incluyendo la enumeración de incrementos de crecimiento en espinas, escamas y otolitos. Las incoherencias en la interpretación entre estructuras y entre laboratorios pueden introducir errores en los datos de edad disponibles para la evaluación de stock. Se llevó a cabo un ejercicio comparativo de estimación de edad para evaluar el sesgo y la precisión de las estimaciones de edad entre lectores y entre estructuras. Seis científicos de cuatro laboratorios, con diferentes niveles de experiencia en la estimación de edad del atún blanco participaron en el estudio. A cada lector se le distribuyeron imágenes de los otolitos, escamas y espinas de 84 atunes blancos recogidos en el Atlántico este norte y en el Mediterráneo. Cuando las estimaciones de edad de un lector experimentado se comparaban con las de un lector no experimentado a quién había entrenado, se observaron niveles moderados de precisión (% medio del coeficiente de variación de 11,2, 7,7 y 6,5 para otolitos, escamas y espinas respectivamente). Sin embargo, una mayor comparación entre laboratorios arrojaba peor precisión (un % del coeficiente de variación de 35,7, 25,7 y 24,0 para otolitos, escamas y espinas respectivamente). Se detectó sesgo lineal entre lectores para todas las estructuras, y los lectores sistemáticamente subestiman o sobreestimaban la edad en relación con los demás. El análisis de la varianza indica que la precisión entre lectores es mayor para el atún blanco del Mediterráneo que para el atún blanco del Atlántico. El acuerdo en las estimaciones de edad derivadas utilizando diferentes estructuras del mismo pez era bajo y las escamas parecían proporcionar menores estimaciones de edad que las espinas. Los resultados destacan la necesidad de realizar una mayor intercalibración de las metodologías de determinación de la edad para el atún blanco.

KEYWORDS

Thunnus alalunga*, albacore, age estimation, spines, scales, otoliths, precision, bias*

1. Introduction

Accurate assessment of the exploitable albacore stock depends on the accurate and precise estimation of catch-at-age. The catch-at-age matrix currently used in the assessment of Atlantic albacore is derived using likelihood based analysis (MULTIFAN) of the catch-at-length composition of the fishery (Santiago and Arrizabalaga 2005). Although the statistical approach has the advantage of objectivity, variability in mean length at age is not accounted for and can affect the precision of parameter estimation (Fournier *et al*. 1998). In contrast, ages estimated from periodic bands on calcified structures (otoliths, scales, spines and vertebrae) reflect individual variation in growth rates and are valuable for resolving catch-at –age estimation and validating MULTIFAN analysis (Ortiz de Zárate *et al*. 2007).

The calcified structures used to estimate age in albacore tuna include scales (Megalofonou *et al*. 2003) spines (Gonzalez-Garcés and Farina-Perez 1983, Santiago and Arrizabalaga 2005) otoliths (Fernadez 1992) and vertebrae (Fernadez 1992). Dorsal spines have emerged as a preferred structure (Anon 2004) due to relative ease of availability and preparation, however ageing protocols must take into account issues such as multiple banding (Compean-Jimenez and Bard 1983, Santiago and Arrizabalaga 2005) within each annulus and re-absorption of the core. Ageing using otoliths has been proposed as an alternative to ageing using dorsal spines (Anon 2004) and daily increment analysis in otoliths has been shown to be effective in albacore (Lee and Yeh 2006) and other species of tuna (Laurs *et al*. 1985, Doray *et al*. 2004) as a method of validating true age. However any data obtained from these age validation studies must be translatable to other structures which may be used as the preferred ageing structure.

Age estimates from all structures are affected by a degree of uncertainty. Reports of the timing of increment formation are sometimes contradictory, particularly in relation to the first increment (Santiago and Arrizabalaga 2005). Validation studies using mark-recapture techniques (Ortiz de Zárate *et al*. 1996) and daily increment counts in otoliths (Lee and Yeh 2006) are necessary to clarify the relationship between growth increments and absolute age and ensure accuracy of age estimates. Precision in age estimation is also of paramount importance. Poor reproducibility within and between readers, laboratories and ageing techniques will increase the margin of error associated with the parameters estimated using the catch-at-age matrix. Systematic ageing error can lead to misinterpretation of spatial and temporal growth patterns and can influence stock assessment outputs (Tyler *et al*. 1989, Reeves 2003). The potential influence of inconsistencies in ageing procedures on catch-at-age data is highlighted in the 2003 albacore stock assessment report. Regional comparison of ageing criteria and exchange of samples is recommended and the investigation of alternative structures such as otoliths for ageing suggested (Anon 2004).

The objective of this comparative age estimation study is to compare ageing criteria across laboratories and to assess bias and precision of age estimates obtained using three calcified structures (otoliths scales and spines). Six scientists from four laboratories, with varying levels of experience in albacore age estimation are participating in the study. Age estimates are compared between readers and between methods, statistically and graphically and the reproducibility of age estimates is assessed.

2. Materials and methods

2.1 Study Material

Albacore were obtained from Irish commercial vessels fishing in the North Atlantic (Bay of Biscay) and from fish merchants in the Alboran Sea in the Mediterranean during summer of 2005. The first dorsal spine and scales from under the pectoral fin were removed from each fish and stored frozen soon after collection. Sagittal otoliths were removed from the head, cleaned and stored dry at room temperature. Scales were dipped in water and cleaned of mucus by rubbing between the fingers. Suitable scales were placed on a glass slide with a drop of water and viewed under a binocular microscope (Megalofonou *et al*. 2003). Eight scales were selected on the basis of clarity and photographed at 2.5X magnification.

The first spine from the first dorsal fin was imbedded in epoxy resin. Three sections of 1mm in thickness were made using a Buelher Isomet® Low Speed Saw according to protocols in Santiago and Arrizabalaga (2005). The sections were placed on a slide and photographs of each section were taken at 1.6x magnification.

One sagittal otolith was chosen at random and imbedded in a 1.5cm^2 block of resin. Four 0.75mm transverse sections were made on the low speed (Doray *et al*. 2004). Each section was polished on 600, 2500 and 4000 grit silicon carbide wet-dry polishing paper with 3mm diamond paste; the final polish was on a Buehler micro cloth. Sections were mounted on a glass slide with PolyMount® and covered with a glass slip. Sections photographs were taken at 4 x magnifications.

Images of all structures were taken using a stereoscopic microscope with a digital camera and PC interface. Images are taken at the same magnification and light intensity parameters. The images for age estimation did not have scale bars attached.

2.2 Ageing procedure

Four laboratories participated in the study, two academic research laboratories and two international governmental fisheries laboratories. The age readers (hereafter referred to as age-readers A-F) had varying levels of expertise as outlined in **Table 1**. Age estimations were made from photograph images of the structures in question. A preliminary trial carried out by one age reader established that age estimates obtained using images of the structures compared well with estimates obtained from viewing the structures directly under a microscope (Coefficient of variation <3%), and age estimation using images was deemed appropriate for the comparison. Clear translucent marks, or groups of more closely packed marks were interpreted as annual bands according to procedures outlined in Santiago and Arrizabalaga (2005), Megalofonou (2003) and Fernández (1992) for spines, scales and otoliths respectively.

Six independent sets of each structure were collated from the available material. Each set contained images of the relevant structure from 14 randomly selected fish; 7 of Mediterranean origin and 7 of Atlantic origin. Images were randomly arranged within each set to ensure that images of spines, scales and otoliths were not presented in the same order to the reader. Sets 1-3 were read by readers B, C and F and Sets 4-6 were read by readers A, D and E. To facilitate a separate comparison of age estimation for two readers that worked closely together, reader A also read sets 1-3 after receiving training in the ageing techniques from reader B. In previous investigations of ageing error in albacore tuna ages were assigned with knowledge of the length of the fish. In order to examine the potential influence of length data on ageing error, three levels of length information were randomly assigned to the sets of images; no length information, true length information and false length information.

2.3 Data analysis

A combination of statistical and graphical methods was used to compare bias and precision between readers and between structures. The coefficient of variation (CV; calculated as the ratio of the standard deviation to the mean), and percent agreement (the proportion of readings for which paired estimates agree) were used as measures of precision across readers and across structures. ANOVAs and the non-parametric Kruskal Wallis test were used to compare precision (CV) between independent sets of age estimates. Three-way factorial ANOVAs were used to investigate the influence of three factors, stock (two levels; Mediterranean and Atlantic), information provided to reader (three levels; length data, no length data and false length data) and reader combination (two levels; A, D & E and B,C & F) on inter-reader precision expressed as the CV of age readings from three readers. Reader combination was included as a random factor, all other factors were fixed. When the interaction terms were non-significant at $p > 0.25$, they were pooled with the error term to increase the power of the test for the main effects (Winer *et al*. 1991). Where significant effects were detected, pair-wise comparisons were carried out using Tukey's post-hoc tests. Age estimates were analysed for each structure separately.

The occurrence of bias (systematic disagreement between readers or structures) in age estimates was visually assessed using age bias plots, where differences between two age-readers or age estimations are plotted as a function of one of the set of ages (Campana *et al*. 1995). Bias was also tested statistically using linear regression and the non-parametric equivalent of a paired t-test; the Wilcoxon's matched-pairs signed ranks test [A].

3. Results

The length of the fish used in the study ranged from 50.0 cm to 79.3 cm fork length with a mean length of 62.5 cm. Age estimates indicated that the samples were dominated by two and three year old fish. The assigned ages ranged from 0-5 years. The mean lengths of each age class assigned by reader A, who aged all structures, are shown in **Table 2**.

3.1 Inter-reader precision

The comparison of reader B (experienced reader) with reader A (inexperienced reader trained by reader B) yielded moderate levels of precision (**Table 3**). For the Mediterranean albacore, the mean CV was consistently below 6% and percent agreement was above 70% for all structures. The greatest between reader agreement was achieved using otoliths. For the Atlantic albacore, good precision was obtained using spines (CV; 5%, percent agreement; 86%) while the level of precision achieved using otoliths was unacceptably low (CV; 20%, percent agreement; 38.1%. Kruskal Wallis tests confirmed that the mean CV of otolith age estimates from readers A and B was significantly lower for the Mediterranean fish than the Atlantic fish ($p<0.001$), indicating better inter-reader precision for the Mediterranean fish. No difference in precision between Atlantic and Mediterranean age readings was detected for the scales and spines ($p>0.05$). For readers A and B, there was no significant difference in mean CV between sets read with length data, with no length data and with false length data for the scales, spines and otoliths for both areas ($p<0.05$). Therefore, the inclusion of length data did not influence the inter-reader precision of age estimates for these readers.

The wider comparison of readers from across laboratories showed poor precision (**Table 3**). The mean CV was above 20% for all structures from both Mediterranean and Atlantic albacore and percent agreement ranged from 45% down to 25%.

In the case of the age estimates from otoliths, the three-way ANOVA (**Table 4**) revealed that inter-reader precision was influenced by the combination of readers involved in the comparison, the length information provided to the reader and the source of the otoliths (Mediterranean or Atlantic). Tukey post-hoc tests showed that the mean CV was significantly lower between readers B, C and F compared with readers A, D and E ($p<0.05$), indicating greater inter-reader precision in the first group of readers. Better inter-reader precision was achieved with the otoliths from Mediterranean albacore when compared to the Atlantic albacore, with the age estimates for the former group showing a significantly lower mean CV (Tukey post-hoc test, $p<0.05$). The influence of length information on inter-reader precision was difficult to interpret, with significantly lower CV's recorded for the sets that were read with false length information compared to those that were read with true length information (Tukey post-hoc test, $p<0.05$).

For the scales, inter-reader precision was influenced by just one of the main effects in the model; stock (**Table 4**). Better inter-reader precision was achieved with scales from Mediterranean albacore when compared to the

Atlantic albacore, with the age estimates for the former group showing a significantly lower mean CV (Tukey post-hoc test, $p<0.05$).

With respect to age estimates derived from spines, both stock and length information influenced inter-reader precision (**Table 4**). Better inter-reader precision was achieved with spines from Mediterranean albacore when compared to the Atlantic albacore, with the age estimates for the former group showing a significantly lower mean CV (Tukey post-hoc test, $p<0.05$). The mean CV of age estimates obtained with no length information was significantly higher than the mean CV of estimates made with the true length of the fish provided (Tukey post-hoc test, $p<0.05$), indicating that the provision of length information may artificially reduce inter-reader precision for spines age estimates.

3.2 Inter-reader bias

Linear regression analysis found no significant systematic differences in age estimates between readers A and B for otoliths, scales or spines. However, significant bias was detected for all other reader combinations and for all structures (**Table 5**). For the majority of reader pairings, the slope of the relationship between the two sets of age estimates was significantly different from one, indicating inconsistencies in the estimates of one reader. In addition, the intercept was significantly different than zero for most of the linear regressions. This signifies over-estimation or under-estimation of ages by one reader relative to the other. Relative bias is also apparent in the age-bias plots shown in **Figure 1**. It is important to note that as the absolute age of the fish is not known the actual direction of the bias can not be determined and readers' estimations are only assessed relative to each other. Within the first group of three readers (A,D,E), reader D showed a tendency to underestimate age relative to the other readers. The age estimates for reader E tended to be higher than the corresponding ages supplied by readers A and D, with the exception of the otolith ages which reader E appeared to underestimate relative to reader A. Within the second group of readers (B,C,F), the trend suggests over-estimation of age by reader C and underestimation by reader F relative to the other readers. Comparison of reader A with reader B indicates a slight tendency for reader A to overestimate age relative to reader B. Wilcoxon one-sample signed ranks tests showed that for most comparisons, the systematic inter-reader bias apparent in the age bias plots was statistically significant (**Table 5**).

3.3 Bias and precision between structures

Considerable inconsistencies were detected in the age estimates derived from two structures in the same fish (**Table 6**) for both Mediterranean and Atlantic albacore. The mean CV of the paired comparisons between structures (all readers combined) ranged from 16-25%. And mean percent agreement ranged from 39-52%. The most experienced reader, B, achieved high levels of precision between structures for the Mediterranean fish but relatively low precision for fish from the Atlantic. The non-parametric Wilcoxon's matched-pairs signed rank test showed significant systematic bias in the age estimates obtained from one structure relative to another (**Table 7**). The occurrence and direction of the bias was not consistent between readers. Reader A showed no tendency to over- or underestimate age from one structure relative to another while reader E showed some degree of bias for all structures. Higher age estimates were derived from scales compared to spines for three readers. The bias between otoliths and scales and between otoliths and spines was not consistent between readers.

4. Discussion

The fish used in this comparison were restricted to juveniles <80cm and age estimates indicate that 2 and 3 were the dominant age classes. The mean lengths of age classes 3 and 4 were lower than those reported in studies of age and growth of Mediterranean (Megalofonou 2000) and Atlantic albacore (Santiago and Arrizabalaga 2005). As the younger age classes were not affected, it is likely that this discrepancy reflects the restricted length distribution of the samples rather than a major difference in the interpretation of annuli. The authors intend to extend this comparison to a wider range of length classes using archived spines and otoliths. This will facilitate the assessment of bias and precision in older age classes and the construction of growth curves, improving comparability with other studies.

The level of precision observed in the comparison spine age estimates of readers B (experienced reader) and A (inexperienced reader trained by reader B) is similar to previous reports of inter-reader precision within a laboratory (Oritz de Zarate *et al*. 2005). This corroborates the assertion that reasonable precision can be achieved with albacore spine sections when the ageing methodology is standardised between readers. The precision achieved for the scales and otoliths of Mediterranean fish was also acceptable and supports the suitability of these structures for estimating age in this stock. The low precision observed in the otolith age estimates for the Atlantic fish indicates that the interpretation of otolith annuli may be prone to a degree of subjectivity for this stock, even with cross-calibration of methodologies.

Although acceptable precision was achieved by readers A and B, poor precision observed in all other comparisons between readers, especially for otoliths. In addition, most inter-reader comparisons (including readers A and B) showed some degree of bias with one reader over- or underestimating age relative to the other. There were also considerable inconsistencies and systematic differences between age estimates derived using different structures from the same fish. In some cases the choice of one ageing structure over another could cause bias in the estimation of length at age. In particular, scales appear to produce lower estimates of age compared to scales. The results emphasize the need for much greater standardisation of ageing methodologies across laboratories through further exchange of material and consultation between readers. In particular, uncertainty surrounding the interpretation of multiple banding needs to be resolved. Inter-calibration of the ageing methodologies used for each structure may help to reduce variability between estimates derived from spines, scales and otoliths and ensure that the annual bands identified in one structure correspond to those identified in another. This is especially pertinent where features in one structure are used to validate annual marks in another. Inter-calibration could increase precision between laboratories to the levels observed within laboratories as reported above. This is critical to ensure that age information from different jurisdictions is reliable and comparable and that spatial and temporal variations in growth rates are not an artefact of the ageing methodologies employed.

Higher precision was achieved using material from Mediterranean albacore than from Atlantic albacore. The apparent difference in increment clarity may reflect a real stock-related difference in the conditions experienced across migration pathways that influence the deposition of annual growth bands in spines scales and otoliths. This highlights the importance of verifying ageing procedures for each stock and for putative components within stocks that may have experienced different environmental conditions. Further comparison of material from each area across a wider range of length classes is needed to confirm if annual growth bands are more variable and more difficult to interpret in Atlantic albacore.

This study deals only with precision and relative bias and not with accuracy of age estimates. The development and standardisation of ageing protocols should be supported by validation studies using daily increment counts or tag recapture studies to ensure that age estimates reflect the absolute age of the fish and are accurate as well as precise.

5. Acknowledgements

This work was funded by funded by the Irish Government and part-financed by the European Union under the National Development Plan 2000-2006 through the Supporting Measures in the Fisheries Sector.

NDP
NATIONAL DEVELOPMENT PLAN
FINANCIAL INSTRUMENT FOR FISHERIES GUIDANCE
BIM
Bord Iascaigh Mhara
Irish Sea Fisheries Board

References

ANON. 2004. 2003 ICCAT Albacore stock assessment session. Collect. Vol. Sci. Pap. ICCAT 56(4): 1223-1311.

CAMPANA. Vol., S.E., M.C. Annand, and J.I. McMillan. 1995. Graphical and statistical methods for determining the consistency of age-determinations. Trans. Am. Fish. Soc. 124:131-138.

COMPEAN-JIMENEZ, G. and F.X. Bard. 1983. Growth increments on dorsal spines of eastern Atlantic bluefin tuna, Thunnus thynnus, and their possible relation to migration patterns. NOAA Tech. Rep. NMFS 8:77-86.

DORAY, M., B. Stequert and M. Taquet. 2004. Age and growth of blackfin tuna *(Thunnus atlanticus)* caught under moored fish aggregating devices, around Martinique Island. Aq. Liv. Res. 17:13-18.

FERNANDEZ, M. 1992. Revision des me´thodes d'aˆgeage du germon (*Thunnus alalunga*, Bonn. 1788) nord-est Atlantique par l'e´tude des pie`ces anatomique´s calcifiee´s. Collect. Vol. Sci. Pap. ICCAT, 39(1): 225-240.

FOURNIER, D.A., J. Hampton, and J.R. Sibert. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. Can. J Fish. Aquat. Sci. 55: 2105-2116.

GONZALEZ-GARCÉZ, A., and A. C. Farina-Perez. 1983. Determining age of young albacore, Thunnus alalunga, using dorsal spines. NOAA Tech. Rep. NMFS, 8. 117-122.

LAURS, R. M., R. Nishimoto, and J. A. Wetherall. 1985. Frequency of increment formation on sagittae of north Pacific albacore *(Thunnus alalunga)*. Can. J Fish. Aquat. Sci. 42: 1552-1555.

LEE, L.-K., and S.-Y. Yeh. 2006. Morphology of rings on otolith and spine characters from North Atlantic albacore of 40-44 cm fork length. ICCAT, Col. Vol. Sci. Pap 60. 443-456.

MEGALOFONOU, P. 2000. Age and growth of Mediterranean albacore. J Fish Bio. 57: 700-715.

MEGALOFONOU, P., C. Yannopoulos, and J.M. Dean. 2003. The potential use of scales for estimating age and growth of Mediterranean albacore *(Thunnus alalunga)*. J Appl. Ichthyol. 19: 189-194.

ORITZ DE ZÁRTE, V., J. Landa, M. Ruiz, and C. Rodriguez-Cabello. 2005. Ageing based on spine section reading of north Atlantic albacore *(Thunnus alalunga)*: precision, accuracy and agreement. ICCAT Collect. Vol. Sci. Pap. ICCAT, 58(4): 1235-1248.

ORITZ DE ZÁRTE, V., P. Megalofonou, G. De Metrio, and C. Rodríguez-Cabello. 1996. Preliminary age validation results from tagged-recaptured fluorochrome label albacore in North East Atlantic. Collect. Vol. Sci. Pap. ICCAT, 43: 331-338.

ORITZ DE ZÁRTE, V., X. Valeiras, C. Rodríguez-Cabello, and M. Ruiz. 2007. Application of age-length-keys to estimate catch-at-age for north Atlantic albacore *(Thunnus alalunga)* stock. Collect. Vol. Sci. Pap. ICCAT, 60(2): 428-436.

REEVES, S. A. 2003. A simulation study of the implications of age-reading errors for stock assessment and management advice. ICES J Mar. Sci. 60:314-328.

SANTIAGO, J., and H. Arrizabalaga. 2005. An integrated growth study for North Atlantic albacore (*Thunnus alalunga* Bonn. 1788). ICES J Mar. Sci. 62:740-749.

TYLER, A. V., R. J. Beamish, and G. A. McFarlane. 1989. Implications of age determination errors to yield estimates. Can. Sp. Pub. Fish. Aquat. Sci. 108:27–35.

WINER, B. J., D. R. Brown, and K. M. Michels. 1991. Statistical principles in experimental design, 3rd edition. McGraw-Hill, New York (1991).

Electronic citations

[A] Department of humanities, University of Amsterdam, Institute of Phonetic Sciences <http://fonsg3.let.uva.nl/Service/Statistics.html> Accessed 2007, June 20th

Table 1. Details of experience level for participating readers.

Reader	*Experience Level*
A	Inexperienced reader
B	Experience of reading all three structures for albacore
C	Inexperienced reader
D	Some experience of reading otoliths for other fish species
E	Some experience of reading otoliths and spines for other tuna species
F	Some experience of reading otoliths and spines for other tuna species

Table 2. Mean length of each age class assigned by reader A for otoliths scales and spines, ± 95% confidence intervals. The numbers in each class are shown in parenthesis.

	Atlantic			*Mediterranean*		
Age Class	*Otoliths*	*Scales*	*Spines*	*Otoliths*	*Scales*	*Spines*
1		54.3 ± 8.2 (5)	51.2 ± 1.0 (4)			
2	55.7 ± 4.7 (11)	55.8 ± 2.5 (7)	62.0 ±4.1 (15)	62.2±0.82 (12)	62.4±0.64 (25)	62.4±0.64 (20)
3	62.1 ± 2.3 (24)	62.2 ± 2.4 (15)	61.6 ± 3.6 (16)	63.0 ± 1.1 (22)	63.4 ± 1.3 (15)	63.1 ± 1.3 (18)
4	71.6 ± 7.4 (7)	67.5 ± 4.6 (15)	69.9 ± 6.9 (4)	65.6 ± 5.9 (4)	66.8 ± 47.6 (2)	64.8 ± 6.3 (4)

Table 3. Inter-reader precision of age-estimates for scales, spines and otoliths from Mediterranean (Med) and Atlantic (Atl) albacore. Standard errors are shown in parenthesis.

	Readers A and B					
	Otoliths		*Scales*		*Spines*	
	Med	*Atl*	*Med*	*Atl*	*Med*	*Atl*
% Agreement	90.5	38.1	81.0	61.9	71.4	85.7
Mean CV (%)	2.7 (± 1.9)	19.8 (± 3.8)	5.4 (± 2.5)	10.0 (± 2.9)	8.1 (± 2.9)	4.9 (± 2.8)
N	21	21	21	21	21	21
	All readers					
	Otoliths		*Scales*		*Spines*	
	Med	*Atl*	*Med*	*Atl*	*Med*	*Atl*
% Agreement (averaged across comparisons)	24.7 (± 4.3)	28.2 (± 5.7)	45.3 (± 2.2)	35.6 (± 2.0)	38.7 (± 2.6)	25.4 (± 1.3)
Mean CV (%)	31.8 (± 2.1)	39.6 (± 2.6)	22.9 (± 2.5)	28.4 (± 2.9)	21.2 (± 2.9)	26.8 (± 2.5)
N	42	40	41	37	41	38

Table 4. Three-way ANOVA summary table showing the influence of inter-reader precision (%CV) achieved from age-estimations for Mediterranean and Atlantic albacore.

Source of variation	*d.f.*	*MS*	*F-ratio*	*p*
Otoliths				
Stock (S)	1	0.13	6.6	0.012*
Length information (L)	1	0.09	4.5	0.014*
Reader combination (R)	2	0.09	4.8	0.032*
S*L				>0.25
S*R				>0.25
L*R				>0.25
S*L*R				>0.25
Error	79	0.019		
Scales				
Stock (S)	1	0.10	4.19	0.044*
Length information (L)	2	0.87	1.24	0.45
Reader combination (R)	1	0.0043	0.06	0.83
S*L	2	0.066		0.072
S*R				>0.25
L*R	2	0.070		0.062
S*L*R				>0.25
Error	75	0.024		
Spines				
Stock (S)	1	0.070	4.29	0.041*
Length information (L)	2	0.056	0.01	0.036*
Reader combination (R)	1	0.00014	3.47	0.926
S*L				>0.25
S*R				>0.25
L*R				>0.25
S*L*R				>0.25
Error	83	1.29		

Table 5. Results of statistical analysis of Inter-reader bias.

	Reader A vs. Reader B			*Reader A vs. Reader D*			*Reader A vs. Reader E*			*Reader D vs. Reader E*		
Statistic	*Otolith*	*Scale*	*Spine*	*Otolith*	*Scale*	*Spine*	*Otolith*	*Scale*	*Spine*	*Otolith*	*Scale*	*Spine*
Slope	1.149±0.45	1.003±0.25	1.122±0.37	0.214±0.29	0.729±0.43	0.376±0.34	0.284±0.13	0.497±0.22	0.250±0.20	0.100±0.17	0.003±0.18	0.197±0.17
p value	0.000	0.000	0.000	0.139	0.001	0.033	0.000	0.000	0.017	0.236	0.785	0.022
Intercept	-0.050±0.88	0.232±0.55	-0.041±0.73	2.669±0.60	1.538±0.91	2.192±0.83	2.429±0.34	1.556±0.71	2.200±0.73	1.760±0.44	1.998±0.59	1.684±0.61
p value	0.910	0.397	0.911	<0.001	0.002	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Test of Slope =1	p>0.05	p>0.05	p>0.05	p<0.05	p>0.05	p<0.05	p<0.05	p<0.05	p<0.05	p<0.05	p<0.05	p<0.05
Wilcoxon test	p=0.002 A>B	p=0.006 A>B	p=0.004 A>B	p<0.001 A>D	p<0.001 A>D	p<0.001 A>D	p<0.001 A>E	p>0.05	p=0.02 A<E	p>0.05	p<0.001 D<E	p<0.001 D<E
	Reader B vs. Reader C			*Reader B vs. Reader F*			*Reader C vs. Reader F*					
	Otolith	*Scale*	*Spine*	*Otolith*	*Scale*	*Spine*	*Otolith*	*Scale*	*Spine*			
Slope	0.697±0.25	0.554±0.23	0.246±0.13	0.391±0.40	0.278±0.29	0.323±0.14	0.490±0.36	0.361±0.35	0.506±0.34			
p value	<0.001	<0.001	<0.001	0.056	0.058	<0.001	0.009	0.044	0.005			
Intercept	-0.032±0.83	0.682±0.60	1.249±0.38	1.609±0.66	1.604±0.47	1.266±0.30	2.4546±0.59	1.905±0.58	1.790±0.75			
p value	0.939	0.028	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001			
Test of Slope =1	p<0.05	p<0.05	p<0.05	p<0.05	p<0.05	p<0.05	p<0.05	p<0.05	p<0.05			
Wilcoxon test	p<0.001 C>B	p<0.001 C>B	p<0.001 C>B	p<0.001 B>F	p<0.001 B>F	p>0.05	p<0.001 C>F	p<0.001 C>F	p<0.001 C>F			

Table 6. Precision of age-estimates between structures for Mediterranean (Med) and Atlantic (Atl) albacore.

	Otoliths versus scales		*Otoliths versus spines*		*Scales versus spines*	
	Med	*Atl*	*Med*	*Atl*	*Med*	*Atl*
% Agreement						
Reader A	47.6	47.6	57.1	52.4	52.4	42.9
Reader B	95.2	42.9	85.7	61.9	95.2	57.1
Reader C	40.0	36.8	55.0	44.4	42.1	41.2
Reader D	52.4	50.0	37.0	35.3	38.7	61.1
Reader E	14.3	9.5	28.6	15.8	21.1	33.3
Reader F	61.1	47.1	42.9	45.0	42.1	44.4
Mean % agreement	51.8	39.0	51.1	42.5	45.1	41.2
CV (%)						
Reader A	14.0	13.9	10.6	12.8	15.3	13.1
Reader B	5.39	21.2	4.0	17.1	1.3	18.0
Reader C	19.0	21.1	13.0	17.1	17.4	21.7
Reader D	19.1	16.8	18.6	24.1	13.9	15.5
Reader E	30.7	41.9	27.5	41.6	26.4	18.8
Reader F	18.3	36.0	23.8	25.0	30.9	41.9
Mean CV (%)	17.8	25.2	16.3	22.9	17.5	21.5
N	21	21	21	21	21	21

Table 7. Results from Wilcoxon paired tests used to detect systematic bias in age readings between structures for each reader. * indicates statistical significance at the alpha = 0.05 level, ** indicates statistical significance after Bonferroni correction for multiple comparisons, ns indicates non-significance.

	Otoliths versus scales		*Otoliths versus spines*		*Scales versus spines*	
	p value	*Bias*	*p value*	*Bias*	*p value*	*Bias*
Reader A	ns	-	ns	-	ns	-
Reader B	ns	-	0.03*	ot>sp	ns	-
Reader C	0.0002**	ot>sc	ns	-	ns	-
Reader D	ns	-	ns	-	0.02*	sp>sc
Reader E	0.03*	sc>ot	0.0008**	sp>ot	0.03*	sp>sc
Reader F	ns		0.0004**	sp>ot	0.003**	sp>sc

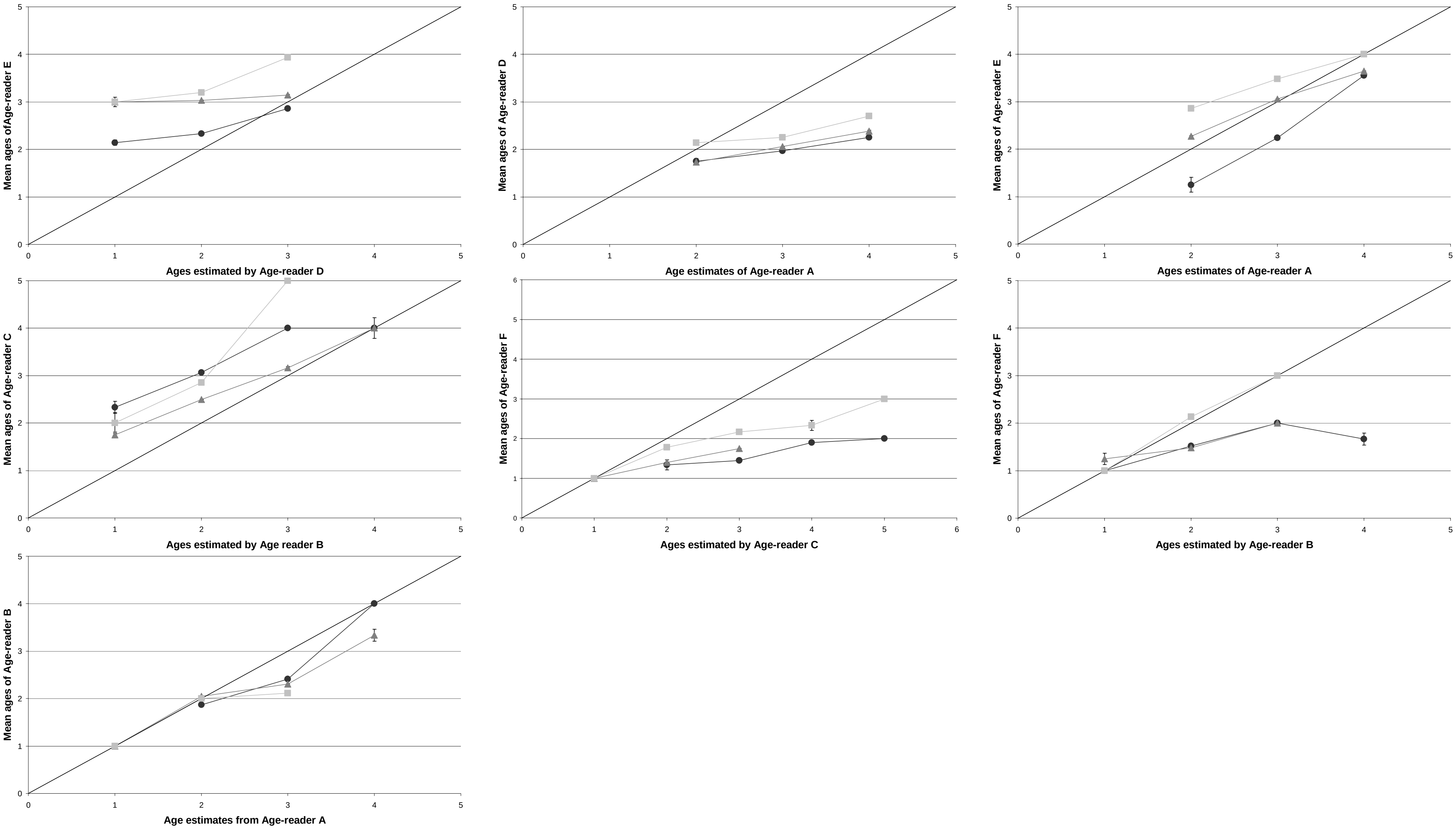

Figure 1. Age bias plots for each of the seven paired age comparisons for the three structures; otoliths (●), scales (▲) and spines (■). Error bars represent the 95% confidence intervals about the average age assigned by one age-reader for all fish assigned a certain age by another age-reader. The 1:1, or zero difference line (solid line) is indicated.