Review of Age Precision Metrics with Recommendations for Future Use

##### Derek H. Ogle

###### Department of Mathematical Sciences, Northland College, 1411 Ellis Ave, Ashland, WI 54806

##### Joshua XX. Lyons

###### Department of Natural Resources, Northland College, 1411 Ellis Ave, Ashland, WI 54806

##### Gordon S. Scott

###### Department of Natural Resources, Northland College, 1411 Ellis Ave, Ashland, WI 54806

Running head: Age Precision Metrics Review

# Abstract

WILL BE <300 WORDS HERE

# Introduction

Age, along with length, is one of the most important characterisitics of individual fish recorded by fisheries scientists. Age data is used to estimate growth, mortality, and recruitment (Spurgeon et al. 2015; Kerns and Lombardi-Carlson 2017; Quist and Isermann 2017), which are primary rates of interest for fisheries scientists (Ricker 1975; Hilborn and Walters 1992; Haddon 2011; Paukert and Spurgeon 2017), and in age-structured population models to estimate yield and set harvest regulations (Beverton and Holt 1957; Hilborn and Walters 1992; Haddon 2011; Methot and Wetzel 2013). Age, unlike length, is not directly observed but is most often interpreted from periods of fast and slow growth recorded on calcified structures such as scales, otoliths, fin rays or spines, or vertebrae (Campana 2001; Spurgeon et al. 2015; Quist and Isermann 2017). Age data can be highly variable and prone to errors as age may not be reliably recorded on or interpreted from calcified structures, especially for fish that are more than only a few years old (Campana 2001; Quist et al. 2012; Phelps et al. 2017). Errors in age data have resulted in incorrect estimates of growth, mortality, and recruitment which have led to misunderstanding the population dynamics of some fish populations (Lai and Gunderson 1987; Reeves 2003; Bertignac and Pontual 2007; Yule et al. 2008; Koenigs et al. 2013; Hamel et al. 2016; Tyszko and Pritt 2017; Chang et al. 2019).

Sources of error in age estimation may be inherent to the calcified structure itself or part of the process of interpreting age from the calcified structure (Campana 2001; Morison et al. 2005; Buckmeier et al. 2017). Errors inherent to the structure are measurable, but cannot be controlled (Morison et al. 2005). Assessing this source of error is part of validating ages estimated from calcified structures, methods for which have been thoroughly reviewed (Maceina et al. 2007; Spurgeon et al. 2015; Buckmeier et al. 2017) and are not addressed further here. Errors related to the process of interpreting ages can be both measured and controlled (Morison et al. 2005) and consist of differences between the estimated and true age (i.e., accuracy) and differences among estimated ages from multiple interpretations of the same structure (either by multiple interpreters or the same interpreter multiple times). Differences among interpretations of the same structure may be systematic (e.g., one set of ages is always one year less than the other set of ages), patterned (e.g., one set of ages largely match the other set of ages up to a certain age and then become progressively more different after that age), or random. Identifying or measuring systematic or patterned differences among interpretations is referred to as identifying *bias* (or not) in the age estimates (Campana et al. 1995). Methods for assessing bias are described by Campana et al. (1995) and are also not discussed further here. In the absence of bias, *precision* is a measure of random differences among interpretations, or the repeatability of age estimates among interpreters (Campana 2001; Maceina et al. 2007) that do not display a bias (Campana et al. 1995). Precision and metrics for assessing precision are the focus this paper.

Precision metrics measure the repeatability of age estimates made on each of individual fish. Percent agreement, the simplest precision metric, is the percentage of fish for which the age estimates are the same (i.e., *exact percent agreement*) or differ by no more than a certain amount (e.g., differ by no more than one year). When , *percent partial agreement* is the percentage of fish for which the age estimates agree for some specific number of interpreters (e.g., at least two interpreters agree). For our purposes we do not differentiate between the percent agreement types and will only consider whether at least one of the *percent agreement* () metrics was used or not.

Other measures of precision result from computing various measures of variability among the age estimates within each of the fish and then averaging these values across the fish. The first two measures of variability are the familiar absolute deviation (AD) and standard deviation (SD) of estimated ages for each fish; i.e.,

where is the th age estimate for the th fish and is the mean of the age estimates for the th fish. The and may be scaled by to produce the percent error (PE) and coefficient of variation (CV) for the th fish, respectively; i.e.,

When and are averaged across all fish they produce the overall measures of precision called *average percent error* [; Beamish and Fournier (1981)] and *average coefficient of variation* [; Chang (1982)]. These measures are usually multiplied by 100 and expressed as a percentage; i.e.,

When , (Chang 1982; Kimura and Anderl 2005) and, thus, and only differ by a constant value and provide functionally the same information about precision among age estimates. For , is proportional to (Chang 1982). However, this proportionality depends on and, thus, there is no constant proportionality between and , though they are likely to be strongly correlated (Kimura and Anderl 2005).

Percent agreement, , and are the most commonly used measures of precision (Campana 2001; Morison et al. 2005). Percent agreement has been criticized as a useful composite measure of precision, largely because it varies widely among species and among ages within a species (Beamish and Fournier 1981; Kimura and Lyons 1991; Campana et al. 1995; Campana 2001; Maceina et al. 2007). Despite these critiques and likely due to its simplicity, Morison et al. (2005) found that 49% of 53 laboratories from 23 countries used percent agreement as a measure of precision among estimated ages. Beamish and Fournier (1981) proposed the to address the shortcomings of the percent agreement. Chang (1982) introduced , though he called it , as a better alternative to the because the standard deviation is a better estimator (unbiased and consistent) of variability than the absolute deviation, though Kimura and Lyons (1991) noted that this was only true if the age estimates were normally distributed.

In a review of 131 publications that reported precision values for fish age estimates, Campana (2001) found that most authors used the (57%) to measure precision, though and were used almost equally for studies of annual age and was used predominantly (84%) when estimating daily age from otoliths. The median and modal across 117 studies deemed to have valid precision estimates were 7.6% and 5%, respectively. No significant differences in precision were found among calcified structures used to estimate annual age. Campana (2001) did, however, specifically mention differences between structures and between taxonomic classes of fish; e.g., “(v)irtually all studies reporting shark ages based on vertebrae did so with CV values exceeding 10%, while the most frequently reported CV for otoliths was 5%.” Campana (2001) reiterated that there is a direct conversion between and when and that both measures are “equally sensitive to precision differences among agers.” From this, Campana (2001) stated that “it is not self-evident that one measure [APE or ACV] is to be preferred over the other.” However, from his literature review and informal discussions with scientists from fish ageing laboratories, Campana (2001) concluded that an of 5% would represent reasonable precision for many fish of moderate longevity and complexity of structure interpretation.

Other metrics of precision have been proposed or could be derived using similar logic. Chang (1982) also introduced the index of precision for the th fish as . Kimura and Anderl (2005) suggested, based on distributional theory, that the median age estimate for the th fish () should be used instead of for . Presumably the same argument could be applied to . This change results in a modified percent error and coefficient of variation for the th fish; i.e.,

Bauerlien et al. (2018)] suggested that if the variability among age estimates is not related to the age of the fish then may be an appropriate measure of variability, as compared to . Similar arguments suggest that may be an appropriate measure of variability as compared to . Averaging these measures across all fish gives five alternative measures of precision among age estimates – *average index of precision* [; Chang (1982)], *average modified percent error* (), *average modified coefficient of variation* (), *average absolute deviation* [; Bauerlien et al. (2018)], and *average standard deviation* (). All but the and are typically multiplied by 100 and reported as a percentage.

Not suprisingly there are predictable relationships between several of these measures. When , , and because , and (Campana et al. 1995). Thus, when , , , , , and provide either the same or functionally the same information about precision among age estimates. Similarly, and provide functionally the same information about precision among age estimates. For , and are proportional and, thus, do not provide functionally different information about precision among age estimates.

It has been more than two decades since Campana (2001) reviewed the use of precision metrics in the published literature and made recommendation for their use. It has also been 15 years since Maceina et al. (2007) reported on a survey of fisheries scientists with respect to their use of precision metrics. In relatively recent years, Bauerlien et al. (2018) suggested that using a single summary precision metric may not be good practice and Buckmeier et al. (2017) described how the ASD may be more appropriate than the in some instances. Given the length of time since the last synthetic review of precision metrics and these newly raised questions, we report here on our comprehensive review of papers published since 1996 that reported age precision metrics for fish. Our overall goal with this review is to summarize the use of precision metrics since Campana (2001). Specifically, we aim to answer the following questions.

1. Has the frequency of published precision estimates changed since 1996?
2. Have there been any notable changes in the frequency of characteristics of the precision metric studies (e.g., in , between- or within-interpreter comparison types, type of calcified structure) since 1996?
3. How often did the authors check for no bias in age estimates among interpretations before computing the precision metric?
4. How often did the authors check for a relationship between the precision metric and age?
5. How often were each of the seven precision metrics introduced above used?
6. Has the frequency of use for the three main precision metrics (, , ) changed since 1996?
7. What is the distribution of values reported in the literature?
8. Do values reported in the literature differ by taxonomic class, calcified structure, number of repeated interpretations (i.e., ), type of comparison (between- or within-interpreter), maximum estimated age, number of age interpretations, or structure preparation method.

Under the *a priori* assumptions that only , , and were used extensively, few papers will have tested for bias prior to computing the precision metric, and few papers will have tested for a relationship between the precision metric and age, we performed additional analyses of multiple age estimates provided by publication authors. Specifically with these additional analyses we aim to answer the following questions:

1. How often is a bias detected (and, thus, an overall measure of precision should not be calculated)?
2. How are the various precision metrics related (beyond the known relationships identified above)?
3. How often is each precision metric related to age?
4. When a precision metric is related to age, what is the form of that relationship?

# [A]Methods

[C]*Literature Review.*– We collected precision metric information from a non-random set of published papers so as to maximize temporal, spatial, taxonomic, and methodological coverage. An initial sample of publications was gathered using GoogleScholar with the search terms “fish” and “precision” coupled with other terms such as “age,” “otolith,” and “vertebrae.” This sample was augmented by searching for publications in specific journals that often publish fish age precision studies (e.g., North American Journal of Fisheries Management, Journal of Ichthyology), searching ResearchGate, mining the references of previously selected publications, and including publications that we became aware of in other ways. Publications for which we could not access the full text, were not written in English, focused only on daily ages, only provided precision metrics across structures (e.g., between scales and otoliths), or did not report precision metric results were excluded.

Our primary sampling unit is the publication and our secondary sampling unit is an individual assessment of precision within each publication. Thus, more than one measure of precision may have been recorded from a single publication. For example, two sets of results (described below) were recorded if precision between interpreters was reported for each of two calcified structures (e.g., scales and otoliths), two sets of results were recorded if precision between interpreters was reported for one calcified structure for two species, or two sets of results were recorded if precision within one interpreter was recorded for one calcified structure for one species on two different water bodies. Results were not recorded if the precision metric was not reported, was clearly in error, or was a summary (i.e., average) across groups of fish.

For each precision assessment we recorded the precision metric(s) used (e.g., , , ), observed value(s) for each metric, taxonomic species and class of fish examined, calcified structure (e.g., otolith, vertebrae) used for estimating age, methodology for preparing the structure for ageing (e.g., sectioned/ground, whole), type of comparison made (between- or within-interpreters), number of repeated age interpretations (), number of fish with age estimates (), maximum estimated age, and range of estimated ages. Numbers of fish with age estimates, maximum estimated age, and range of estimated ages were approximated from figures for some assessments. Assessments were excluded from further analysis when =2 and both and were reported, but the to ratio was not within rounding. To increase the number of observed values for subsequent summarization, assessments with only results were converted to results when =2 with . For each publication we recorded the year of publication, the country of the lead author, whether or not was used alone or with either the or (assuming was used), whether or not the was used alone or with the (assuming was used), whether or not the was used alone or with the (assuming was used), whether or not the authors checked for bias among interpretations of estimated age and, if they checked for bias, which method they used to check for bias (e.g., age-bias plot, symmetry test), and whether or not they checked for a relationship between precision metric values and estimated age.

Tests with categorical response (e.g., ) and categorical explanatory (e.g., calcified structure) variables were conducted with a chi-square test with potential post-hoc comparisons made with multiple chi-squares tests for each pair of levels of the explanatory variable using the Benjamini-Holberg-Yekutieli (BHY) false discovery rate method to correct for multiple comparisons (Benjamini and Hochberg 1995; Benjamini and Yekutieli 2001). Differences in median values of quantitative response variables (e.g., ) among levels of categorical explanatory variables (e.g., calcified structure) were examined with a Kruskal-Wallis test followed by Dunn’s test (Dunn 1961, 1964) using the BHY method for multiple comparisons. Relationships between quantitative response (e.g., ) and quantitative explanatory variables (e.g., maximum age) were assessed with a general additive model (Wood 2017) using a smoothing term on the explanatory variable. All analyses were performed in R v4.1.0 (R Core Team 2021) using dunnTest() from the FSA v0.9.0 package (Ogle et al. 2021), gam() from the mgcv v1.8-36 package (Wood 2017), and =0.05.

[C]*Extended Analyses.*– XXX

# [A]Results

## [B]Literature Review

The results from the literature review consisted of 1256 individual precision assessments from 384 papers published between 1996 and 2020. These results contained assessments for 354 species from 100 families and were reported by lead authors from 40 countries with lead authors from the United States most represented (44.8%; Figure 1).

The number of papers with precision assessments generally increased over the 25-year period (Figure 2). Otoliths were the most commonly assessed calcified structure (51% of all assessments for Actinopteri), though vertebrae were most commonly assessed among elasmobranchs (82% of all assessments for Elasmobranchii; Figure 3). The number of precision assessments for Actinopteri increased in the late 2000s, primarily due to increased assessments of scales, spines, and fin rays, and then again in the late 2010s due to more otolith assessments (Figure 3). The number of precision assessments for Elasmobranchii was fairly constant over the 25 years, though the assessment of spines increased beginning in the late 2000s (Figure 3). Most precision assessments were between-interpreters rather than within-interpreters for all calcified structures except vertebrae where only 47% of assessments were between-interpreters (Figure 4). The majority of precision assessments were between two interpretations (69%; Figure 5). The percentage of assessments using two interpretations increased to a peak in the late 2000s and decreased into the 2010s, primarily replaced with assessments among three interpretations, though assessments with four or more interpretations were more prevalent in the late 2010s (Figure 5).

Bias among interpretations was examined for at least one assessment in 45% of publications, with a slight increase from 38% of publications in 1996-2000 to 50% of publications in 2016-2020. Of those publications that examined bias, 72% used an age bias or age difference plot (Campana et al. 1995) and 29% used a symmetry test (Bowker 1948; McBride 2015), sometimes together or with various other methods. The relationship between the value of the precision metric and estimated age was mentioned in 11% of publications, though this relationship was not formally tested in nearly all of these publications.

The was the most used precision metric, though the and were also commonly used (Figure 6). The was most often used simultaneously with either the , , or and, thus, was used rarely by itself (17% of publications that used ). The and were commonly used together with 46% of publications that used also using and 58% of publications that used also using . The percent of publications that used the generally increased from 1996 to 2020, whereas the percent of publications that used the was fairly steady before increasing in the late 2010s (Figure 7). However, the percent of publications that used the without also using the declined to approximately 20% by the early to late 2000s and has stayed steady since then (Figure 7). The was used rarely and always with either the or (Figure 6). The , , and were each used in one study, and was not used in any publications.

The distribution of ACV values, including APEs converted to ACVs when =2, was strongly right-skewed (Figure 8) with a median of 8.9% and 75% of values between 0 and 13.5%. Median ACV was lowest when otoliths were assessed (6.7%) and highest when scales were assessed (12.2%) though the median ACV for scales was not significantly different than the median ACV for miscellaneous “other” structures (Figure 8). The median ACV for finrays, spines, and vertebrae were intermediate between the median ACV of otoliths and scales at 9.2-9.5% (Figure 8).

The effects of various explanatory variables on ACV values were inconsistent among combinations of calcified structure and taxonomic class. For otolith assessments of Actinopteri fish (Figure 9), median ACV was lower for two rather than more than two repeated interpretations (p=0.016), did not differ between between- and within-interpreter comparisons (p=0.739), showed a nonlinear relationship with maximum estimated age (p=0.001) primarily exhibited by a small peak for fish estimated at between approximately 4 and 9 years old, and was not related to the ratio of number of fish examined to range of estimated ages (p=0.247). For scale assessments of Actinopteri fish (Figure 10), median ACV was also lower for two rather than more than two repeated interpretations (p<0.0005), also did not differ between between- and within-interpreter comparisons (p=0.087), also showed a nonlinear relationship with maximum estimated age (p=0.034) but the small peak was later at between approximately 9 and 15 years old, and was generally negatively related to the ratio of number of fish examined to range of estimated ages (p<0.0005). For vertebrae assessments of Elasmobranchii fish (Figure 11), median ACV did not differ between two and more than two repeated interpretations (p=0.367), was lower for within- than between-interpreter comparisons (p=0.011), and was not significantly related to maximum estimated age (p=0.233) or the ratio of number of fish examined to range of estimated ages (p=0.632). Median ACV did not differ due to preparation methods for otoliths of Actinopteri (i.e., among whole, cracked, and ground or sectioned; p=0.094) or for vertebrae of Elasmobranchii (i.e., between whole and ground or sectioned; p=0.861).

[C]*Extended Analyses.*– XXX

WILL BE FLESHED OUT

# [A]Discussion

## [B]Subsection

WILL BE FLESHED OUT

Don’t lose this comment from Gordy … Just a thought - may be interesting to note the findings of Albequerque et al. 2019 here or elsewhere - where precision (APE) was linearly correlated with latitude in Micropogonias furnieri. This may need backing from other studies though.

# [A]Acknowledgments

WILL BE FLESHED OUT

# Figures

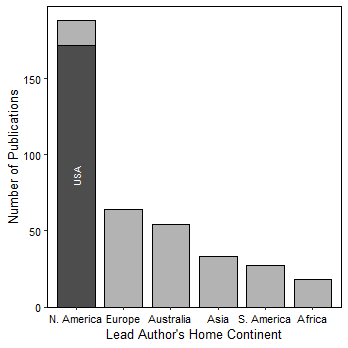


Figure 1. Number of publications in the literature review by the leaad author’s home continent. North American publications with lead authors from the USA are highlighted.

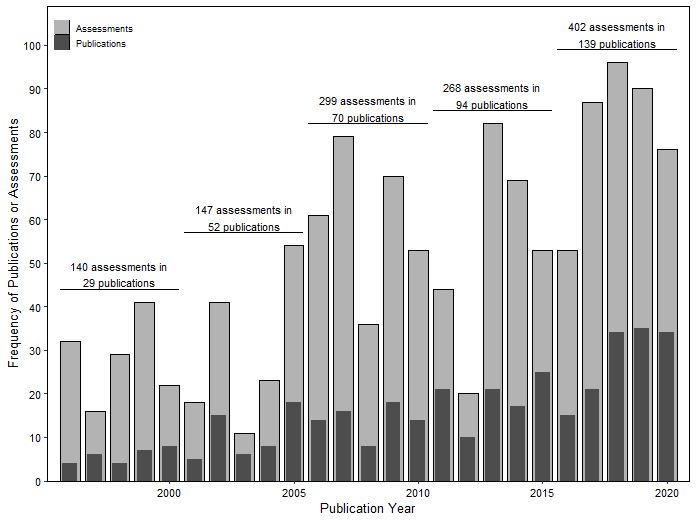


Figure 2. Number of individual assessments and number of publications with precision assessments used in the literature review by year from 1996-2020. Total number of assessments and publications are also shown for five-year periods.

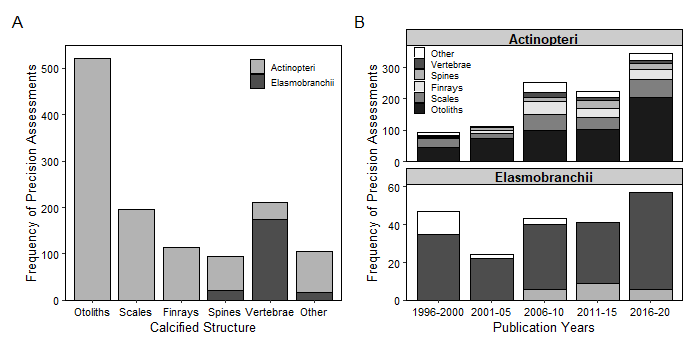


Figure 3. Number of individual precision assessments by primary calcified structure type separated by major taxonomic class (A) and further separated by five-year periods (B). Six assessments for Holocephali and ten assessments for Petromyzonti are not included.

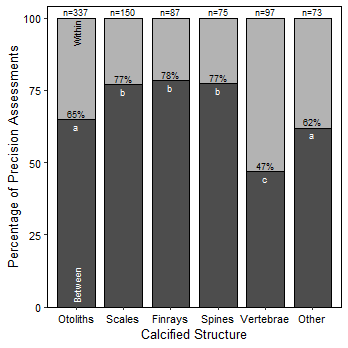


Figure 4. Percentage of individual precision assessments within calcified structure types that were within-interpreter assessments. Structures with the same letters have statistically equal percent of within-interpreter assessments. Overall sample size for each structure is shown above the bars. Eight observations that combined within- and between-interpreter assessments are not included.

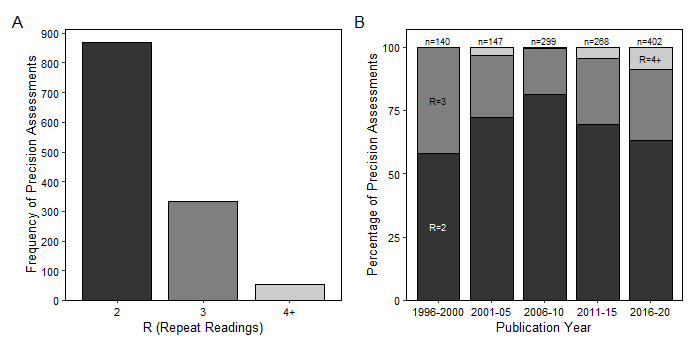


Figure 5. Number of individual precision assessments by the number of repeat interpretations (A) and further separated by five-year period (B). The maximum number of repeated interpretations was =9.

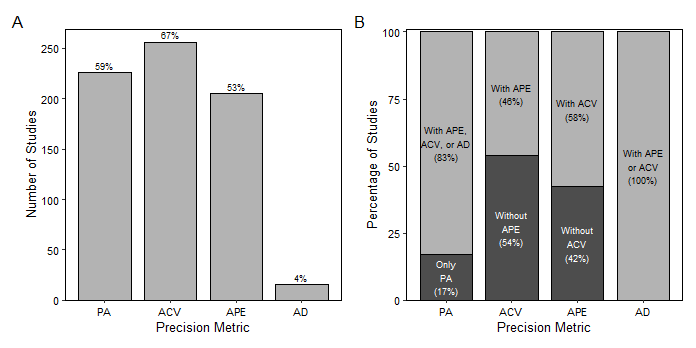


Figure 6. Number of publications that used each of the main precision metrics (A) and the percentage of publications using each metric separated by how the metric was used relative to the other main metrics (B). Note that =Percent Agreement, =Average Coefficient of Variation, =Average Percent Error, =Average Index of Precision.

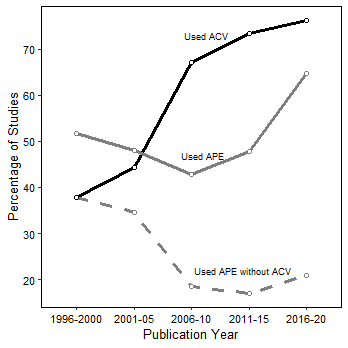


Figure 7. Percentage of publications that used the Average Coefficient of Variation (), Average Percent Error (), or the alone without also using the by five-year period.

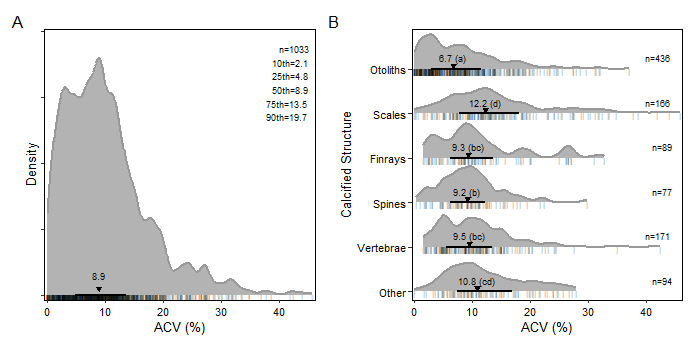


Figure 8. Density distribution and individual values (vertical marks) of Average Coefficient of Variation () values for all assessments (A) and for assessments by main calcified structure type (B). The median value is shown and marked with a triangle. The inter-quartile range (first to third quartile) is shown by the black horizontal bar. Calcified structures with different letters have significantly different medians. These values include all Average Percent Error () values for two interpretations (=2) converted to values.

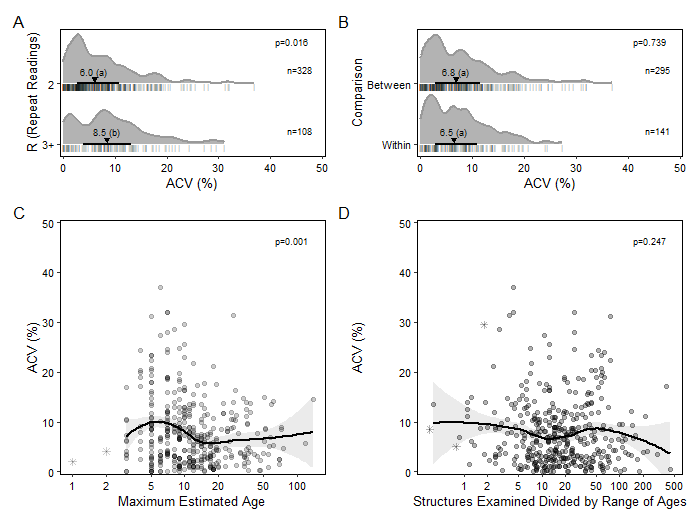


Figure 9. Density distribution and individual values (vertical marks) of Average Coefficient of Variation () values for all assessments by number of repeated interpretations (A) and comparison type (B), and the relationship between and maximum estimated age (C) and sample size (D) for otoliths of Actinopteri. The median value is shown and marked with a triangle and the inter-quartile range (first to third quartile) is shown by the black horizontal bar in panels A and B. Assessments with different letters in the same panel have significantly different medians. The significance of the GAM smoother is shown in panels C and D. Estimated ages less than 3 and number of structures examined less than 20 were excluded from the respective GAM models and are marked with asterisks. A log scale is used for the x-axis in panels C and D.

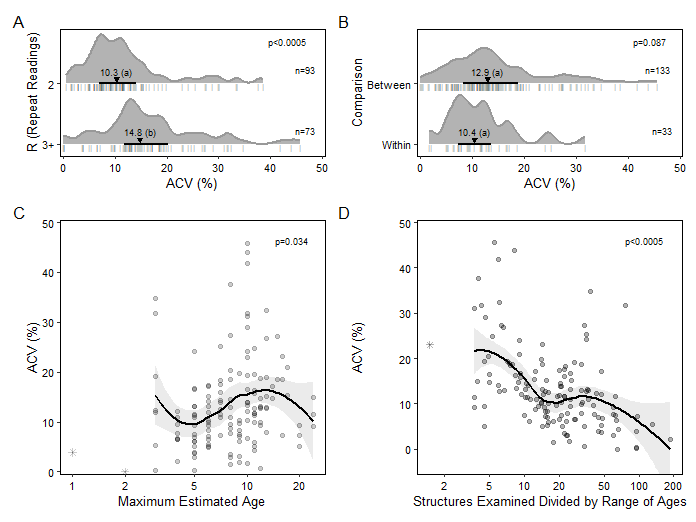


Figure 10. Density distribution and individual values (vertical marks) of Average Coefficient of Variation () values for all assessments by number of repeated interpretations (A) and comparison type (B), and the relationship between and maximum estimated age (C) and sample size (D) for scales of Actinopteri. The median value is shown and marked with a triangle and the inter-quartile range (first to third quartile) is shown by the black horizontal bar in panels A and B. Assessments with different letters in the same panel have significantly different medians. The significance of the GAM smoother is shown in panels C and D. Estimated ages less than 3 and number of structures examined less than 20 were excluded from the respective GAM models and are marked with asterisks. A log scale is used for the x-axis in panels C and D.

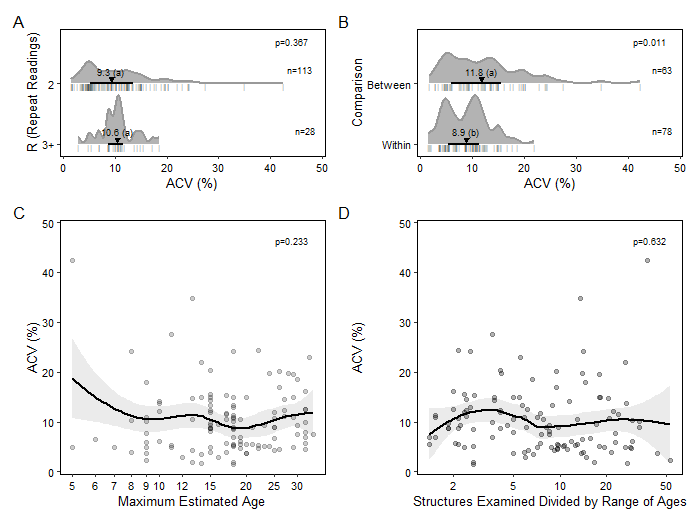


Figure 11. Density distribution and individual values (vertical marks) of Average Coefficient of Variation () values for all assessments by number of repeated interpretations (A) and comparison type (B), and the relationship between and maximum estimated age (C) and sample size (D) for vertebrae of Elasmobranchii. The median value is shown and marked with a triangle and the inter-quartile range (first to third quartile) is shown by the black horizontal bar in panels A and B. Assessments with different letters in the same panel have significantly different medians. The significance of the GAM smoother is shown in panels C and D. Number of structures examined less than 20 were excluded from the respective GAM models and are marked with asterisks. A log scale is used for the x-axis in panels C and D.

# [A]References

Need to fix species capitalization

Bauerlien, C. J., M. R. Cornett, E. A. Zielonka, D. P. Crane, and J. S. Bulak. 2018. Precision of calcified structures used for estimating age of Chain Pickerel *Esox niger*. North American Journal of Fisheries Management 38.

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

Benjamini, Y., and Y. Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 57:289–300.

Benjamini, Y., and D. Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependency. Annals of Statistics 29:1165–1188.

Bertignac, M., and H. de Pontual. 2007. Consequences of bias in age estimation on assessment of the northern stock of European Hake (*Merluccius merluccius*) and on management advice. ICES Journal of Marine Science 64(5):981–988.

Beverton, R. J. H., and S. J. Holt. 1957. On the dynamics of exploited fish populations. United Kingdom Ministry of Agriculture, Fisheries, London.

Bowker, A. H. 1948. A test for symmetry in contingency tables. Journal of the American Statistical Association 43:572–574.

Buckmeier, D. L., P. C. Sakaris, and D. J. Schill. 2017. Validation of annual and daily increments in calcified structures and verification of age estimates. Pages 33–80 *in* M. C. Quist and D. A. Isermann, editors. Age and growth of fishes: Principles and techniques. American Fisheries Society, Bethesda, Maryland.

Campana, S. 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.

Campana, S. E., M. C. Annand, and J. I. McMillan. 1995. Graphical and statistical methods for determining the consistency of age determinations. Transactions of the American Fisheries Society 124:131–138.

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.

Chang, Y.-J., J. Hsu, J.-C. Shiao, and S.-K. Chang. 2019. Evaluation of the effects of otolith sampling strategies and ageing error on estimation of the age composition and growth curve for Pacific Bluefin Tuna *Thunnus orientalis*. Marine and Freshwater Research.

Dunn, O. J. 1961. Multiple Comparisons among Means. Journal of the American Statistical Association 56:52–64.

Dunn, O. J. 1964. Multiple Comparisons Using Rank Sums. Technometrics 6:241–252.

Haddon, M. 2011. Modelling and quantitative methods in fisheries, 2nd edition. CRC Press, Boca Raton, FL.

Hamel, M. J., J. J. Spurgeon, C. J. Chizinski, K. D. Steffensen, and M. A. Pegg. 2016. Variability in age estimation results in ambiguity and false understanding of population persistence. North American Journal of Fisheries Management 36:514–522.

Hilborn, R., and C. J. Walters. 1992. Quantitative fisheries stock assessment: Choice, dynamics and uncertainty. Chapman; Hall, New York.

Kerns, J. A., and L. Lombardi-Carlson. 2017. History and importance of age and growth information. Pages 1–8 *in* M. C. Quist and D. A. Isermann, editors. Age and growth of fishes: Principles and techniques. American Fisheries Society, Bethesda, Maryland.

Kimura, D. K., and D. M. Anderl. 2005. Quality control of age data at the Alaska Fisheries Science Center. Marine and Freshwater Research 56:783–789.

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

Koenigs, R. P., R. M. Bruch, and K. K. Kamke. 2013. Impacts of aging error on Walleye management in the Winnebago System. North American Journal of Fisheries Management 33:900–908.

Lai, H. L., and D. R. Gunderson. 1987. Effects of ageing errors on estimates of growth, mortality and yield per recruit for Walleye Pollock (*Theragra chalcogramma*). Fisheries Research 5:287–302.

Maceina, M. J., J. Boxrucker, D. L. Buckmeier, R. S. Gangl, D. O. Lucchesi, D. A. Isermann, J. R. Jackson, and P. J. Martinez. 2007. Current status and review of freshwater fish aging procedures used by state and provincial fisheries agencies with recommendations for future directions. Fisheries 32(7):329–340.

McBride, R. S. 2015. Diagnosis of paired age agreement: A simulation of accuracy and precision effects. ICES Journal of Marine Science 72:2149–2167.

Methot, R. D., and C. R. Wetzel. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142:86–99.

Morison, A. K., J. Burnett, W. J. McCurdy, and E. Moksness. 2005. Quality issues in the use of otoliths for fish age estimation. Marine and Freshwater Research 56:773–782.

Ogle, D. H., J. C. Doll, P. Wheeler, and A. Dinno. 2021. FSA: Fisheries Stock Analysis.

Paukert, C. P., and J. J. Spurgeon. 2017. Age structure. Pages 221–232 *in* M. C. Quist and D. A. Isermann, editors. Age and growth of fishes: Principles and techniques. American Fisheries Society, Bethesda, Maryland.

Phelps, Q. E., S. J. Tripp, M. J. Hamel, R. P. Koenigs, and Z. J. Jackson. 2017. Choice of structure for estimating fish age and growth. Pages 81–106 *in* M. C. Quist and D. A. Isermann, editors. Age and growth of fishes: Principles and techniques. American Fisheries Society, Bethesda, Maryland.

Quist, M. C., and D. A. Isermann. 2017. Age and growth of fishes: Principles and techniques. American Fisheries Society, Bethesda, Maryland.

Quist, M. C., M. A. Pegg, and D. R. DeVries. 2012. Age and growth. Pages 677–732 *in* A. V. Zale, D. L. Parrish, and T. M. Sutton, editors. Fisheries techniques, 3rd edition. American Fisheries Society, Bethesda, Maryland.

R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reeves, S. 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.

Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Bulletin of the Fisheries Research Board of Canada 191:1–382.

Spurgeon, J. J., M. J. Hamel, K. L. Pope, and M. A. Pegg. 2015. The global status of freshwater fish age validation studies and a prioritization framework for further research. Reviews in Fisheries Science & Aquaculture 23:329–345.

Tyszko, S. M., and J. J. Pritt. 2017. Comparing otoliths and scales as structures used to estimate ages of Largemouth Bass: Consequences of biased age estimates. North American Journal of Fisheries Management 37:1075–1082.

Wood, S. N. 2017. Generalized Additive Models: An Introduction with R, 2nd edition. Chapman; Hall/CRC, Boca Raton.

Yule, D. L., J. D. Stockwell, J. A. Black, K. I. Cullis, G. A. Cholwek, and J. T. Myers. 2008. How systematic age underestimation can impede understanding of fish population dynamics: Lessons learned from a Lake Superior Cisco stock. Transactions of the American Fisheries Society 137:481–495.