***Precision Metrics Summarized from 20+ years of Fish Age Estimation Studies***

Joshua F. Lyons, Natural Resources, Northland College

**Abstract:**

Fish age estimation is important for the analysis of growth, mortality, and recruitment in fish populations. Measuring precision of fish age estimation is important for validating various methodologies for ageing fish. The average coefficient of variation (ACV) has become the most used precision metric after the release of Campana (2001). The overall median ACV has increased from 5% in Campana (2001) to 9.3% in this study. Median ACV differed among the number of repeated readings per fish, the range of observed ages, and the calcified structure used, but not between between- and within-reader readings or classes of fish. Of the 12.5% of studies that checked for a relationship between a precision metric and age, 53% detected a relationship with age. Of the 41.5% of studies that checked for bias in the age estimates, 37.8% detected a bias. More specific comparisons can be made in the future to further our understanding of ageing precision.

**Introduction:**

Age estimation of fish is crucial for assessing stock structure and population dynamics as ages are needed to estimate growth, mortality, and recruitment in fish populations (Kerns and Lombardi-Carlson 2017; Quist and Isermann 2017). Growth is important to understand because it can affect a fish's vulnerability to predation and angling pressure. Growth can also be used as a metric of prey resource availability (Kerns and Lombardi-Carlson 2017). Estimates of mortality are important for setting good harvest regulations. A catch-curve is usually used to estimate mortality and they require estimates of age for individual fish (Ogle 2016). Recruitment is used to predict the future year-classes of a fish population. Fish recruitment to the stock population can be defined, for example, as when the fish reach a breeding age (Ogle 2016).

Calcified structures from fish often show seasonal patterns of growth which can be used to estimate the age of fish. Calcified structures commonly used for age estimation include otoliths, scales, spines, fin rays, and vertebrae. Traditionally, scales were the structure most used to estimate age of fish, as annuli on scales were validated as an ageing method in the late 1800's (Kerns and Lombardi-Carlson 2017). Scales provide one of the easiest and least invasive methods for age estimation as minimal harm comes to the fish when removing scale samples. As of the 20$^{th}$ century, scales have been found to commonly underestimate ages of fish, especially of longer-lived fishes. Thus, other calcified structures are more commonly used to estimate age of fish today. Otoliths tend to provide accurate ages for most fish, but extracting otoliths requires the fish to be sacrificed (Kerns and Lombardi-Carlson 2017). Spines are less invasive to sample and tend to provide more accurate ages than scales, but less accurate ages than otoliths for long-lived fishes.

Fisheries managers base many management decisions and regulations, in part, on estimates of rates of growth, mortality, and recruitment. Thus, accurate estimates of the age of fish are needed for proper management of fisheries. There are two major sources of error when estimating the age of fish (Campana 2001). The first source of error, called process error, is that some structures do not reliably record the true age of the fish. For example, due to resorption of calcified material on the margins of scales, scales often underestimate the true age of the fish. The second source of error is due to the subjectivity of age estimations by individual readers. Both errors may result in over- or under-estimates of the true ages of fish. These biases in fish age estimation may result in management decisions that are based on false conclusions, which may be detrimental to fish populations. For example, underestimating the age of Cisco (*Coregonus artedi*) using scales resulted in regulations that contribute to their major decline throughout the 20th century (Yule et al. 2008).

In addition to accuracy, it is important that age estimates are precise, or repeatable among multiple readings of a calcified structure. Precision can be measured from age estimations made among multiple readers or among multiple readings made by the same reader. Importantly, high precision (i.e., repeatability of age estimates) does not indicate high accuracy (i.e., estimating true age).

Commonly used measures of precision include percent agreement (PA), average percent error (APE), and average coefficient of variation (ACV). Percent agreement, which was the predominant precision metric prior to 1981, is the percentage of all fish for which the repeated readings perfectly agree. However, Beamish and Fournier (1981) stated that PA was an inadequate measure of precision because it varies widely both among species and among ages within a species (Campana 2001). For instance, 95% agreement might be considered poor precision for a short-lived species but good precision for a long-lived species. Beamish and Fournier (1981) recommended using APE instead of PA. Average percent error is defined as

$$APE = 100 * \frac{1}{nR}\sum_{j=1}^{n}\sum_{i=1}^{R}\frac{|x_{ij} - \overline{x_j}|}{\overline{x_j}}$$

where n is the number of aged fish in the sample, $R$ is the number of repeated age estimates for each fish, $x_{ij}$ is the $i$th age for the $j$th fish, and $\overline{x_j}$ is the mean age for the jth fish. Chang (1982) suggested using ACV instead of APE due to the inherent assumption in the APE that the standard deviation of age is proportional to the mean age for individual fish (Ogle 2016). Average coefficient of variation is defined as

$$ACV = 100 * \frac{1}{n}\sum_{j=1}^{n}\frac{s_j}{\overline{x_j}}$$

where $s_j$ is the standard deviation for the repeated age estimates of the $j$th fish. The $ACV = \sqrt{2} * APE$ when only two readings are made.

Campana (2001) reviewed the scientific literature that reported precision metrics for the age estimation of fish. From this review, he suggested that the ACV should be used to measure precision of age estimates in most cases because it is statistically more rigorous and more flexible. Furthermore, he reported that the modal observed ACV across all studies he examined was 5%, and suggested that a 5% ACV could be used as a threshold for identifying “acceptable precision.”

Despite the suggestion from Campana (2001) to use the ACV as the measure for estimating precision among age estimates, there are situations where computing the ACV may be inappropriate. Bauerlien et al. (2018) said that if CVs are related to age, then it may be inappropriate to report the overall ACV because that overall value will not pertain to fish of all ages. For example, one may conclude that precision is “good” if the overall ACV is less than 5%, but this may be misleading if the CV is much greater than 5% for some ages (and concomitantly lower for other ages). McBride (2015) suggested that it may be difficult to detect bias among age estimates if precision among age estimates is low. However, I also suggest that a synthetic measure of precision may be inflated by bias among age estimates. In other words, ACV could be related to age due to an age-related bias among age estimates or due to an age-related decrease in precision with no bias.

The literature with respect to the precision of fish age estimates has not been systematically reviewed since Campana's review nearly twenty years ago. In this paper, we reviewed published papers since Campana's review to determine if (i) researchers have followed Campana's advice and primarily used ACV as the measure of precision, (ii) the observed modal ACV across all studies has changed from the 5% that Campana reported, (iii) if ACV differs among a variety of characteristics of the study (e.g., type of comparison, class of fish, number of readings, range of observed ages, and type of calcified structure), (iv) if authors examined the relationship between ACV and age, and (v) if authors examined bias in addition to precision.

**Methods:**

We obtained a non-random sample of papers to review by first searching GoogleScholar and ResearchGate with '"age precision" AND fish' and then including relevant papers cited in those papers. Papers for which we could not access a PDF (or for which authors did not return our request for a PDF), were not written in English, did not provide a summary precision metric, or provided only precision metrics across structures (e.g., precision between scales and otoliths) were immediately rejected from the initial sample. After these exclusions we reviewed and extracted information from 313 papers from as early as 1983, though 288 (92.0%) were from after 1995.

Our sampling unit was each comparison of age estimates presented in a paper. For example, information regarding two precision results was recorded for a paper where age estimates from two readers was compared for each of two structures (e.g., scales and otoliths). As another example, information regarding ten precision results was recorded for a paper where age estimates from one structure was compared between two readings by the same individual but separately for ten different lakes. We did not record information about precision metrics if there were obvious errors in the results, if the precision metric did not represent a single sample of fish (i.e., averaged across populations), or if it was unclear if the metric represented within- or between-readings precision.

The information recorded from each paper included: precision metric(s) used (e.g., PA, APE, ACV), observed metric value(s), publication year, country and study site of the study, species of fish aged, calcified structure used for ageing, methodology for preparing the structure for ageing (e.g., sectioned, whole, ground), type of precision analysis (between- or within-readers), number of repeated age readings done (R), number of fish aged in the sample (n), range of estimated ages, whether the authors checked for a bias in ages and, if so, the methodology for checking bias (e.g., age-bias plot, symmetry test), and whether the authors checked for a relationship between ACV and age and, if so, how they examined that relationship (e.g., linear regression, quadratic regression). To increase the number of observed ACV values for when we summarized ACV values, APE-only results were converted to ACV results when two repeated readings were made with $ACV = \sqrt{2} * APE$.

The Kruskal-Wallis and Dunn's post-hoc tests were used to test for differences in median ACV by comparison type, class of fish, number of readings, range of ages, and calcified structure. These non-parametric tests were used because the assumptions for an ANOVA test could not be met with the distribution of the data. We used R v3.5.1 (R Core Team 2018) for the Krusal-Wallis test and dunnTest() from the FSA v0.8.22 package (Ogle et al. 2018) for the Dunn's test.

**Results:**

Our sample contained 255 studies published after the release of Campana (2001). Of those 255 studies, 170 (66.7%) used the ACV, 118 (46.3%) used the APE, 152 (59.6%) used perfect PA, and 63 (24.7%) used both the APE and ACV. Of the 197 studies since 2001 with only two readings per fish (i.e., R=2), 124 (62.9%) used the ACV, 88 (44.7%) used the APE, 130 (66%) used perfect PA, and 46 (23.4%) used both the APE and ACV.

The distribution of ACV in our sample was right skewed (Figure 2A), with a median of 9.3%. Median ACV was not significantly different between between-reader and within-reader comparisons (Figure 2B) or between Elasmobranchii and Actinopterygii fish classes (Figure 2C). Median ACV for two repeated readings was significantly smaller than for three repeated readings, though median ACV from both two and three repeated readings were statistically the same as for 4+ repeated readings (Figure 2D). The median ACV for observed age ranges of less than 10 years was significantly greater than median ACV for observed age ranges of 10-20 years and more than 20 years (Figure 2E). Finally, otoliths had a significantly smaller median ACV than spines, finrays, vertebrae, and scales, with scales having a significantly larger median ACV than the other structures (Figure 2F).

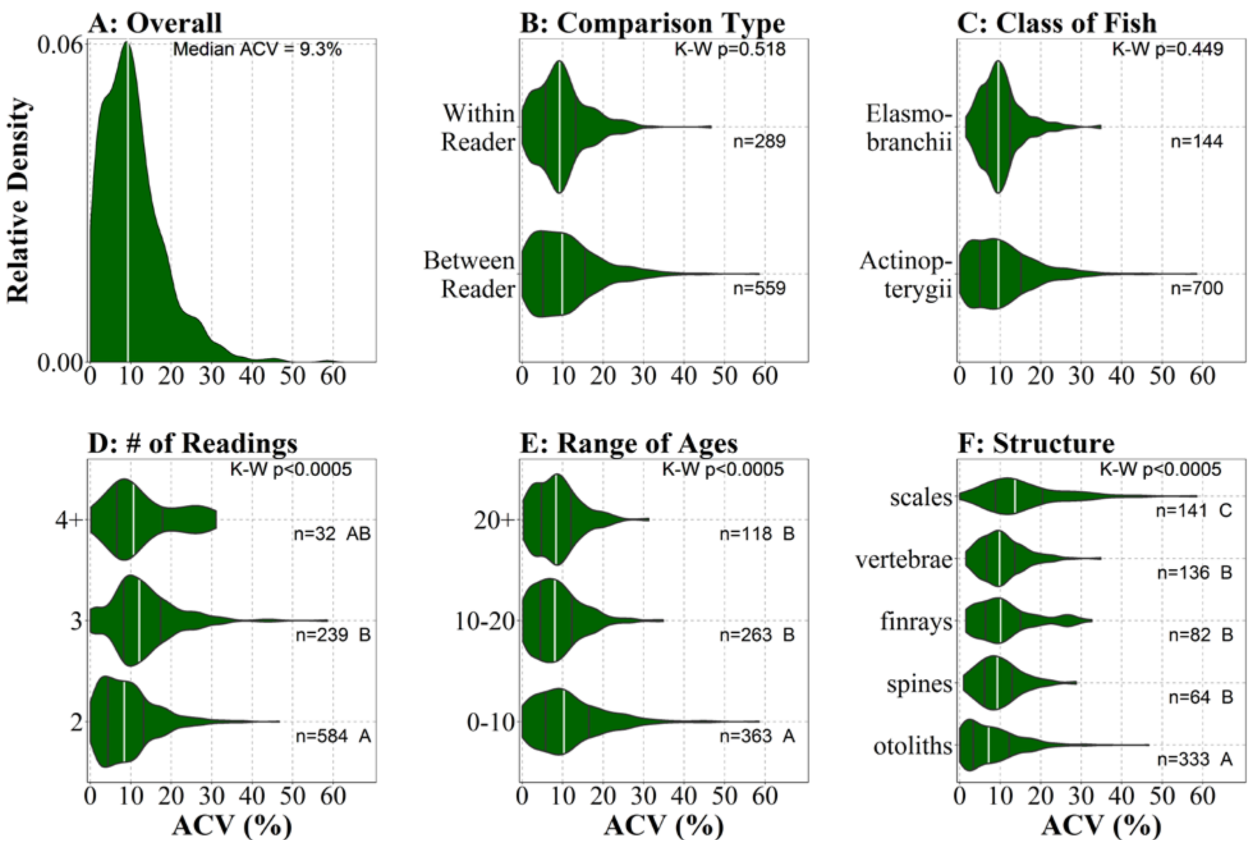


Figure 2. Density of observed ACV values for all studies (A) and by comparison type (B), class of fish (C), number of repeated readings (D), range of observed ages (E), and type of calcified structure (F). Vertical white lines are the median, whereas vertical black lines are Q1 and Q3. The Kruskal-Wallis (K-W) p-values and group sample sizes (n) are shown. Groups with different letters have significantly different median ACVs.

Only 39 of the 313 studies (12.5%) checked for a relationship between a precision metric and age. Of those 39 studies, 22 (56.4%) used the ACV, 11 (28.2%) used the APE, and 24 (61.5%) used perfect PA. A relationship with age was found for 61 (55.5%) of the precision metric results while 49 (44.5%) of the records did not find a relationship with age. There were 5 precision metric results that checked for a relationship with age but did not report any conclusions.

The authors checked for bias in age estimates in 130 of the 313 studies (41.5%). In these 130 studies, bias was examined for 392 comparisons. Of the comparisons that examined bias, 298 (76.0%) used age-bias plots, 76 (19.4%) used tests of symmetry, 54 (13.8%) used t-tests, and 52 (13.3%) used regression. Many studies used more than one method when looking for bias (e.g., age-bias plots and regression, age-bias plots and t-tests). A bias in age estimates was detected in 141 of the 373 comparisons (37.8%) that reported whether or not bias was found.

**Discussion:**

Most but not all publications after Campana (2001) used the ACV as a precision metric, which suggests that some scientists are either not aware of Campana's advice or are not heeding it. The median ACV (9.3%), which is very near the modal ACV, in this study appears to be greater than the modal ACV (5%) reported by Campana (2001). Two possibilities may explain this increase in ACV values. First, it is possible that there has been an increase in studies that compare ages from experienced readers to ages from inexperienced readers. If this were the case, then these studies would likely have lower precision than studies that compared ages only among experienced readers. We attempted to record the experience level of readers when extracting results from the sampled studies, but this information appeared infrequently in the papers. Secondly, more recent studies may be examining ages estimated from more calcified structures within a fish species and from more fish species that fisheries technicians have less experience with. In these cases, some of the structures or fish may be inherently difficult to reliably age, which would lead to larger ACV values.

Variability in median ACV was explained by the number of repeated readings, the range of observed ages, and the calcified structure used. Two repeated readings of a structure had a significantly higher precision (i.e., lower ACV) than three readings of a structure. This is probably best explained by the fact that there is more room for error when more repeated readings are made of the same structure. There was probably no significant difference detected between when two repeated readings were made and when 4+ repeated readings were made due to the relatively low sample size of 4+ repeated readings (n=32). Younger fish (0-10 age range) had significantly worse precision than older fish (10+ age range). A small disparity in age estimation within a small range of ages will have a larger impact on precision than a larger disparity in age estimation within a large range of ages. Age precision was highest for otoliths, followed by spines, fin rays, and vertebrae, and then scales. Thus, otoliths are most precise but they require sacrificing the fish and scales are easiest to use but lack precision. Overall, spines and finrays may be alternatives that balance precision and invasiveness. No differences in ACV values were observed according to comparison type or class of fish.

Species of fish will probably have an impact on the quality of the calcified structure as an ageing method. It would be a good idea to look at more nuanced comparisons in the future. These findings can help determine a good overall methodology for ageing fish by selecting for predicted outcomes with a lower median ACV. In the future, more specific comparisons can be made to further our understanding of ageing precision and how it interacts with various aspects of the analysis. It would be good to look at how these findings compare with a specific species of fish. In many cases, more data will need to be collected for a species due to a low sample size.

Less than half of the studies checked for bias in their age estimates and bias was detected in approximately one-third of comparisons. A bias in the age estimates can serve to inflate the overall ACV, though this issue was not addressed in any of the studies with these comparisons. Very few papers checked for a relationship between a precision metric and age. Of the comparisons that did check for this relationship, more than half of them found a relationship between a precision metric and age. For example, Kotas et al. (2011) found that APE was positively related with age for Scalloped Hammerhead Sharks (*Sphyrna lewini*). If a relationship is found with age, then the age distribution should be considered when using the precision metric (Bauerlien et al. 2018).

**Future Directions:**

More nuanced and specific comparisons should be made with these data in the future. It would be helpful to know how these findings compare among specific species of fish. Further comparisons between the various precision metrics should also be done. It was hard to compare precision metrics in this analysis due to the low sample sizes of many of the precision metrics. We have asked for raw data from many different fish ageing precision studies. We will be able to compare all of the different precision metrics using the raw data.

## References


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. North American Journal of Fisheries Management. 38:1548-8675.

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.

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.

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.

Kerns, J. A., and L. A. Lombardi-Carlson, 2017. History and importance of age and growth information. Pages 1-5 in M. C. Quist and D. A. Isermann, editors. Age and growth of fishes: principles and techniques. American Fisheries Society, Bethesda, Maryland.

Kotas, J. E., V. Mastrochirico, and J. M. Petrere, 2011. Age and growth of the Scalloped Hammerhead Shark, *Sphyrna lewini* (Griffith and Smith, 1834), from the southern Brazilian coast. Brazilian Journal of Biology. 71:755-761.

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

Ogle, D. H. (2016) Introductory Fisheries Analyses with R. Boca Raton, FL: CRC Press.

Ogle, D. H., P. Wheeler, and A. Dinno. 2018. FSA: Fisheries Stock Analysis. R Package version 0.8.22, https://github.com/droglenc/FSA.

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

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. URL http://www.R-project.org/.

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.