

ATTAINABILITY OF ACCURATE AGE FREQUENCIES FOR OCEAN QUAHOGS (*ARCTICA ISLANDICA*) USING LARGE DATASETS: PROTOCOL, READER PRECISION, AND ERROR ASSESSMENT

KATHLEEN M. HEMEON,^{1*} ERIC N. POWELL,¹ ERIC ROBILLARD,² SARA M. PACE,¹ THERESA E. REDMOND³ AND ROGER MANN³

¹Gulf Coast Research Laboratory, University of Southern Mississippi, 703 East Beach Drive, Ocean Springs, MS 39564; ²National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northeast Fisheries Science Center, 166 Water Street, Woods Hole, MA 02543; ³Virginia Institute of Marine Science, College of William and Mary, 1370 Greate Road, Gloucester Point, VA 23062

ABSTRACT Ocean quahogs (*Arctica islandica*) are the longest lived bivalve on Earth. Individuals on the deep continental shelf off Georges Bank can survive for centuries, and in the colder, boreal waters of Iceland, ages over 500 y can be reached. Ocean quahog landings in the United States represent a \$24 million industry, yet assessment models operate with no age data because of the substantial sample size required to develop adequate population age distributions for such a long-lived species, the unknown error associated with age estimates, and the extensive time and financial investment required to create production-scale age datasets. Inclusion of age data for this species requires precision metrics to evaluate aging uncertainty such as percent agreement, percent error, coefficient of variation, and tests of bias. To move forward using error-validated age-composition data, a 3-fold error protocol was developed using a large dual-reader dataset ($n = 610$) from Georges Bank. First, a proxy age-validation study was performed to corroborate an aging method, followed by error evaluation in the context of age-reader bias, precision, and error frequency. Error thresholds were established for each of the three error methods. Georges Bank samples ranged from 33 to 261 y of age and met the predetermined error thresholds for bias (conditionally because of significant and nonsignificant results), precision (average coefficient of variation less than 7%), and error frequency (less than 10%). Consequently, age estimates were deemed acceptable to support age frequency analyses. Precision and bias error were greatest for the youngest animals and, in the context of age-reader bias, error rates were higher for young male ocean quahogs than for young females. Improved age validation of young, sex-differentiated *A. islandica* will constrain aging error and guide refinement of both aging and age-error protocols.

KEY WORDS: ocean quahog, *Arctica islandica*, error, precision, bias, age-reader

INTRODUCTION

Calcified structures (e.g., fish otoliths, vertebrae, coral skeletons, and bivalve shells) are commonly used to age animals in the marine realm (Hudson 1981, Pentilla & Dery 1988, Richardson 2001). As the animal grows, calcium carbonate is secreted in layers around the calcified structure and concentric growth rings are created. Growth rings reflect the rate of carbonate deposition correlated with seasonal and annual growth patterns and often retain information on environmental conditions such as temperature and available food (Schöne et al. 2011, Swart 2015, Purroy et al. 2018). The age of an individual can be determined by the sum of its annual growth rings and a collection of ages from a population sample can be extrapolated to construct an age distribution for the population. Age data are critical for managing fisheries as they are the cornerstone records used to estimate recruitment, spawning stock biomass (i.e., fecundity), and mortality rates (Brooks et al. 2008, Martell et al. 2008, Lee et al. 2011, Minte-Vera et al. 2019).

Age compositions are often estimated directly or by age-at-length keys (Mohn 1994, Harding et al. 2008, Stari et al. 2010). In either case, a sample size sufficient to resolve the age distribution at length in the population is essential (Kimura 1977, MacDonald & Pitcher 1979, Hoenig 2017, Hulson et al. 2017). The amalgamation of many ages within small-length divisions in adult animals (Weinberg 1999, Hofmann et al. 2006) poses

a particular challenge. Species that reach extremely old age, such as the ocean quahog *Arctica islandica*, provide an exceptional example (Ridgway et al. 2012, Pace et al. 2017a, 2017b, 2018). For long-lived species such as *A. islandica*, with upward of 200 possible age classes, the total number of aged animals required to provide a defensible age-at-length key is very large. Techniques used to create chronologies for this species, particularly cross dating and isotope dating (Butler et al. 2009, Schöne et al. 2011, Reynolds et al. 2017), often cannot provide the number of ages necessary for population age compositions at any affordable cost. Accordingly, traditional visual aging methods must be used (Ropes 1988). Given the life spans involved for this species, an inordinately large number of chances for reader error can occur and close attention must be paid to the precision at which ages can be determined under the constraint of high sample number and accurate age estimations where possible.

Error is a valued statistic used to appraise data quality and consistency across datasets, laboratories, researchers, and methodologies and error analysis is routinely used in the development of population age data for fisheries assessment purposes (Pentilla & Dery 1988, CARE 2006). Error is classically defined as the difference between an estimated value and the true, or accurate, value that is often categorized as either sampling error, observational error, or processing error. Sampling error influences data integrity and challenges typically stem from insufficient sample size or measurement bias (Duval & Tweedie 2000, Hjellvik et al. 2002, Pennington et al. 2002, Johnsen 2003, Jacobson et al. 2010, Costa et al. 2016, Ritter et al. 2016, Powell et al. 2017). Observational error can amass from the human

*Corresponding author. E-mail: kathleen.hemeon@usm.edu
DOI: 10.2983/035.040.0206

interpretation of sample data, whereas process error reflects variability in biological processes compelled by biotic and abiotic forces. In the case of determining the age of an animal such as the ocean quahog, age data are susceptible to observational error, as annual growth rings (i.e., annuli) need to be interpreted by different readers with varying experience levels, to distinguish annuli from subannual growth patterns where process errors can be substantial dependent on extreme oceanographic conditions (e.g., yearly, seasonally, and monthly) (Jones 1980, Campana et al. 1995). When sampling error is either negligible or inescapable, observational error is an important facet that can be improved and constrained to elevate data quality with high-precision and low-systematic bias. Reduction of observational error may not always drive the data toward the true value as that value is not known for many species; however, low observational error can improve precision and allow reproducible age data in future studies, which is a noteworthy alternative (Kimura & Lyons 1991).

Observational error is best evaluated by precision and bias metrics using paired blind age comparisons between two age readers. Precision is the scale of reproducibility, or agreement between readers, over time and is conventionally reported as average percent agreement (Beamish & Fournier 1981), average percent error (Beamish & Fournier 1981), and/or average coefficient of variation (ACV) (Chang 1982). Age bias is the systematic difference between paired age estimates and is the product of individual reader interpretation, aging methodology, age class of the animal, and the individual animal itself (processing error) (Kimura & Lyons 1991, Hoenig et al. 1995). Age-reader bias occurs when one set of age determinations is consistently higher or lower than a comparative set of age determinations for identical samples. Bias may be present even when precision is high, therefore, simple precision statistics alone are not sufficient to describe the quality of an age dataset (Campana et al. 1995, Hoenig et al. 1995, Kimura & Anderl 2005).

The evaluation of observational error is well described in fisheries literature by means of precision metrics (Campana et al. 1995, Campana 2001) and, more recently, tests of symmetry (McBride 2015). Many state and federally managed fisheries are aged at a production scale to inform population models used to set harvest limits, and precision in aging is a critical metric in establishing the degree of uncertainty present in age-composition data used in these population models. The bivalve *Arctica islandica* is an exceptionally valuable clam commercially harvested and managed at the federal level, but age-based models do not exist for this species (NEFSC 2017) because of its long lifespan (greater than 200 y in the Mid-Atlantic, United States) and the difficult interpretation of growth patterns for consistent aging. As a result, *A. islandica* is not aged at production scale because of the aforementioned constraints and eliminates any opportunity for managers to use quality-controlled age data. To produce age-composition data at production scale for potentially forthcoming *A. islandica* population age models, the same level of quality control must exist for this exceptionally long-lived species as it does for other commercially managed fisheries.

The objective of this paper was to assess three methods of observational error analysis, namely, age bias, age precision, and error frequency, for a large *Arctica islandica* age dataset ($n = 610$) created from a proxy age-validation study. Error thresholds for the sample population and both sexes were established and tested for each of the three error methods.

MATERIALS AND METHODS

Sample

In 2017, 706 live *Arctica islandica* clams were collected from Georges Bank (40.72767° N, 67.79850° W) at a depth of approximately 72 m by the *ESS F/V Pursuit* using a Dameron-Kubiak dredge (https://scemfis.org/wp-content/uploads/2021/01/DK_dredge1.pdf) that offered variable bar spacing to collect animals smaller than market size (i.e., less than 80 mm in shell length). Clams greater than 70 mm in shell length were retained for this study. Clams were measured for shell length, sex was identified by smear slide, and shell valves were cleaned by immersing in a bleach solution and stored dry for aging. A random subset of valves was chosen for age estimations and included as close to 100 animals per 5-mm size class as possible ($n = 645$) and equal numbers of males and females per size bin when possible. If 100 shells were not available per size class such as for rare size classes (<80 mm or >100 mm), all available shells were aged. Sizes ranged from 72.6 mm to 119.8 mm and resulted in 10 size-class groups based on the 5-mm delineations.

A single valve from each selected animal was sliced along the axis of greatest growth (largest height dimension) as close to the shell origin as possible using a Kobalt wet tile saw and the sectioned valve was progressively exfoliated with silicone carbide abrasive paper at 240, 320, 400 and 600 grit sizes (Pace et al. 2017a). Exfoliation removed excess shell to bring the cut edge as close to the shell origin as possible while also removing coarse shell texture. Shells were then polished to a reflective finish with a polycrystalline diamond suspension fluid (6 μ m and 1 μ m diamond sizes) to clearly display the annual growth lines. After processing, shells were imaged using a high-definition Olympus DP73 digital microscope camera. Segmented images of the hinge and umbo region were stitched together using Olympus CellSens microscopic imaging software. Stitched images created a single, comprehensive image of the entire hinge. Additional details on cleaning, processing, imaging, and aging *Arctica islandica* shells can be found at https://www.vims.edu/research/units/labgroups/molluscan_ecology/publications/topic/ocean_quahog_arctica/index.php.

Age Validation Proxy

Of the 645 clams that were processed for aging, 610 clams were used for final error analysis. The excluded were specimens with images that did not display consistently clear growth lines (Ropes et al. 1984a) or those aged by consensus for training using two age readers and consequently pairwise data did not exist. ImageJ software (ObjectJ plugin) was used to annotate annual growth lines on each comprehensive hinge image for aging. Annuli determination was vetted through a comparative aging-technique analysis using two strategies (Fig. 1). The first strategy applied a grouped hypothesis, where lighter gray lines or repeating patterns (e.g., doublets) were posited to represent periods of reduced growth within season and not terminal annual growth lines. Noticeably, repetitive patterns of lighter gray lines were more commonly observed when the animal was “young” and experiencing periods of rapid growth, but doublets and triplets were routinely observed through much of the growth history. The necessary ignorance of light gray lines in early years of life is a common occurrence in aging bivalves as these are routinely produced during periods of rapid juvenile

growth (Jacobson et al. 2006, Harding et al. 2008, Shirai et al. 2018, Huyghe et al. 2019), but such ignorance may be incorrect in later years. Hence, the second strategy applied a singular hypothesis, where observed repetitive growth patterns such as doublets were judged not to be seasonal, but a manifestation of annual periodicity. The singular hypothesis posits that growth lines, particularly those observed in the middle and later years of life, are true annuli.

Both hypotheses can be supported biologically, yet an arbitrary choice cannot be made because of the extreme differentiation in age estimates between the two aging strategies. To resolve this fundamental problem, shells from 20 of the oldest animals collected from both Georges Bank and Long Island were carbon-14 dated, in addition to two age readers (readers A and B) visually aging the samples using the grouped and singular strategies. The mean age was used for each sample for both visual aging strategies to compare with the carbon-14 results as it was not known which aging protocol was correct.

A Dremel tool removed between 0.018 g and 0.044 g of carbonate dust from the cut shell surface as close to the shell origin as possible (earliest carbonate deposited) without carbon contamination from the shell exterior. Samples were sent to the Keck Carbon Cycle AMS Facility at University of California Irvine for dating. Birth years were estimated by isotope analysis using “prebomb” carbon-dating techniques. Additional details on carbon-14 sampling can be found at https://www.vims.edu/research/units/labgroups/molluscan_ecology/_docs/lab_manuals/2020-4-carbon-14-quahog-protocol.pdf.

Carbon-14 ages were corrected for the marine reservoir effect using a 400-y correction factor. Animals used for carbon-14 dating were collected in the cold pool, south of Long Island and off Georges Bank (for cold pool, see Sha et al. 2015, Lentz 2017, Chen et al. 2018). This region of the continental shelf has been the site of relatively few reservoir age evaluations (Weidman & Jones 1993, Sherwood et al. 2008) in comparison with extensive work in the northeastern Atlantic (Tisnérat-Laborde et al. 2010, Heaton et al. 2020). The few values available approximate the average marine value; thus, the average marine value was used (Stuiver & Polach 1977, Heaton et al. 2020).

Age Bias

Observational error was redefined for each of the three subsequent error methods. Error, in the context of bias, is defined as the difference between age estimates of two age readers. A test of symmetry can identify systemic bias in ages between age readers when comparing aging methodologies (e.g., scales versus otoliths), or testing for age-reader drift over time (e.g., age reader A versus age reader B, age reader A versus reference dataset, and age reader A at start versus age reader A at end). The detection of age bias should be completed before precision estimates are made, as a bias will confound precision interpretations through artificial inflation of values (Campana & Jones 1992, Campana et al. 1995, Hoenig et al. 1995). Significant difference (asymmetry) is determined using the chi-square statistic for observations falling off the matrix diagonal (diagonal values represent 100% agreement between the two groups being tested) (Bowker 1948). The McNemar test maximally pools data on each side of the diagonal to create one group for chi-square analysis above and below the diagonal for a single comparison. In contrast, the Bowker test is an unpooled test that treats each

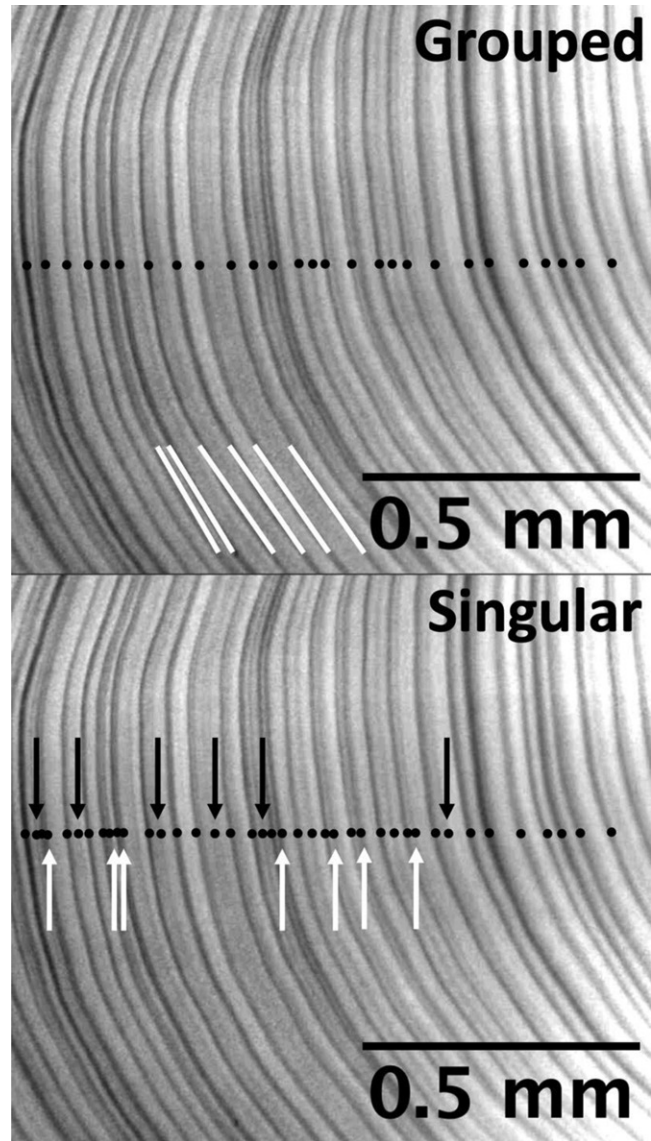


Figure 1. Comparative aging techniques. The grouped hypothesis assumed seasonal growth patterns, whereas the singular hypothesis assumed that growth lines are annual. Black circles indicate where an annulus is counted for each strategy. White vertical lines at the bottom of the grouped image (top) highlight dominant growth lines used to distinguish annuli because of their dark/bold appearance. The grouped aging strategy assumes pale growth lines are subannual as they often disappear at the lateral edges of the hinge (out of range of these images). Arrows in the singular image (bottom) designate additional annuli added when the singular aging strategy is used including doublets (white arrows below annuli) and weak annuli that appear fainter than surrounding annuli (black arrows above annuli). The singular strategy added 13 annuli.

pairwise comparison off the diagonal as an independent group, thereby using numerous comparisons. The Evans–Hoenig test pools (semipools) pairwise data immediately off the diagonal and compares these data with pooled groups at incremental levels off the diagonal (± 2 y, ± 3 y, etc.).

The AgeBias function from the “FSA” package (Ogle et al. 2021) in R (R Core Team 2018) was used to calculate tests of symmetry for the McNemar (Eq. 1) (McNemar 1947), Bowker (Eq. 2) (Bowker 1948), and Evans–Hoenig (Eq. 3) (Evans &

Hoenig 1998) equations (equation formatting taken from McBride 2015).

$$X^2_{\text{McNemar}} = \frac{\left(\sum_{i=1}^{m-1} \sum_{j=i+1}^m (n_{ij} - n_{ji}) \right)^2}{\sum_{i=1}^{m-1} \sum_{j=i+1}^m (n_{ij} + n_{ji})} \quad (1)$$

$$X^2_{\text{Bowker}} = \sum_{i=1}^{m-1} \sum_{j=i+1}^m \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}} \quad (2)$$

$$X^2_{\text{Evans-Hoenig}} = \sum_{p=1}^{m-1} \frac{\left(\sum_{j=1}^{m-p} (n_{p+j,j} - n_{j,p+j}) \right)^2}{\sum_{j=1}^{m-p} (n_{p+j,j} + n_{j,p+j})}, \quad (3)$$

where X^2 is the chi-square statistic, i is the reader A age (row), j is the reader B age (column), n is the frequency of age estimates at row i and column j , m is the number of readings, and p is $j - i$.

To better understand where a potential bias may exist, an age-bias plot was used to compare reader A ages as the reference ages to reader B. The designation of the reference reader is arbitrary when an age-validated reference collection is not being used, and age readers have similar experience levels (as is the case with readers A and B in this study), because the true ages are not known. If experience levels had differed, the expert reader would have been designated as the reference.

Raw absolute error would be expected to increase with age as error should accumulate with each additional annulus over the lifespan of an animal. If the absolute error is standardized by age to create an error rate (a similar statistic to CV), the slope of these data should be near 0 if no aging bias of this type exists (Kimura & Lyons 1991). Accordingly, the absolute value of the error, or the absolute difference between the age estimates between readers (presented in the age-bias plot on the Y axis), was standardized by age to understand how this type of error changed with age and thus create an error rate (specified as errors per year) (Eq. 4).

$$\text{Error rate} = \frac{|\text{Age difference}|}{\text{Age}_{\text{Reference Reader}}} = \frac{|\text{Error}|}{\text{Age}_{\text{Reference Reader}}} \quad (4)$$

As the number of animals aged per birth year was often sparse, once the error rate was determined for each reference age, the data were ordered by birth year and smoothed in 10 sample increments to refine any underlying pattern in error rate. The median for each 10 sample increment (i.e., rolling median) of error rate and reference age was used for error rate analysis and fitted to a trendline to elucidate any patterns of underlying bias.

Age Precision

Precision is an error metric represented by several statistics including the coefficient of variation (CV). Coefficient of variation is the more rigorous precision measurement when compared with the more traditional percent agreement and was thus chosen as the best statistic to validate age precision in this study (Beamish & Fournier 1981, Campana et al. 1995,

Campana 2001, Kimura & Anderl 2005). In the context of this project, precision error occurred when the ACV is greater than an accepted threshold for pairwise age comparisons (Eq. 5).

$$\text{ACV}(\%) = \frac{\sum \left(\frac{s}{\bar{x}} * 100\% \right)}{n}, \quad (5)$$

where s is the standard deviation, $\bar{x}$ is the mean for each set of pairwise ages, and n is the total number of samples. Coefficient of variation standardizes precision across size classes, which is valuable for a long-lived species such as *Arctica islandica*. Age analyses in marine fisheries often use age estimates that meet a precision error threshold of less than 7.6% ACV (Campana 2001). A 7% or less ACV threshold was chosen for *A. islandica* to mirror methods used by federal and state resource managers.

Error Frequency

Error, in the context of error frequency as used in this study, is any sample with a dual-reader CV greater than 10%. Age estimates are deemed acceptable if the error frequency (i.e., number of samples with CV greater than 10%) is less than 10% of the total dataset (expected probability of error = 0.1) using a binomial test. A significant binomial test, or elevated frequency of samples with CVs greater than 10%, is an indication that at least one age reader is aging differently than another age reader and too many large errors are present in the age data. If the error frequency threshold is exceeded, samples with the highest CV can be aged by consensus (i.e., the sample can be aged jointly by at least two age readers) until the error frequency is less than 10%, but this approach is only useful if all specimens are aged by both readers.

RESULTS

Age Validation Proxy

Carbon-14 dating is a useful approach to validate aging techniques in animals of lifespans too long to easily follow the time course of growth from birth to death (Witbaard et al. 1994, Wanamaker et al. 2009, Shirai et al. 2018). A total of 20 shells were sampled for carbon-14 aging, but samples 4, 5, and 13 were contaminated with modern carbon and therefore not used in this analysis. Compared age estimates from the grouped and singular aging options, with the minimum and maximum error bounds for carbon-14 results (Fig. 2), indicated that the singular hypothesis better captured the validated results from the carbon-14 isotope analysis. Singular age estimates fall within the error bounds of carbon-14 ages more frequently ($n = 12$) than those of the grouped age estimates ($n = 3$). A one-way, repeated analysis of variance test coupled with pairwise comparison t -tests (Bonferroni correction), demonstrated that the grouped hypothesis was significantly different from both the singular ($P = 1.69\text{e-}9$) and carbon-14 age estimates ($P = 2.37\text{e-}5$). The singular and carbon-14 age estimates were not significantly different ($P > 0.05$) (Fig. 3).

Although isotope dating can be used for age validation for select samples, the carbon-14 data presented herein only apply to a small number of individuals of a similar age caste. Furthermore, the error bounds on the prebomb carbon-14 ages

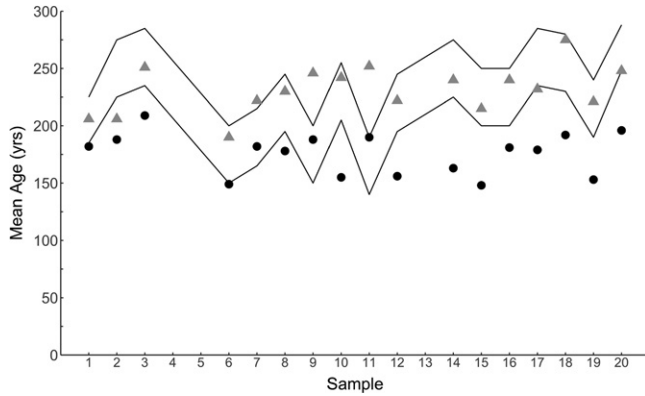


Figure 2. Age estimates using each of the two aging techniques. Solid black lines indicate the upper and lower error bounds inherent in carbon-14 ages. Age estimates from the grouped technique (dark, circle symbol) are consistently lower than estimates from the singular technique (medium gray, triangle symbol). The singular age estimates fell within carbon-14 error (black lines) more frequently than the grouped age estimates (12 and 3, respectively).

are too large to be used as definitive reference ages. Despite these shortcomings, these carbon-14 results are currently the best validation tool for this set of *Arctica islandica* samples and serve as a proxy age validation to support aging-technique selection. As a result, the singular aging technique was applied for all age estimates listed herein. This approach is consistent with conclusions of Butler et al. (2009, 2013) and Pace et al. (2017a) for *A. islandica* (Schöne et al. 2005, Harding et al. 2008) and for other long-lived species (Shirai et al. 2018) but diverges from other species frequently showing within-season growth checks of similar appearance to annuli (e.g., summer breaks and spawning breaks; Goodwin et al. 2001, Fan et al. 2011, Kubota et al. 2017).

Age Bias

Each of the three tests of symmetry pool age frequencies differently, resulting in varying degrees of freedom and significance levels (Table 1). The McNemar test produced the most significant results across all three sample types (Population $P = 1.37\text{e-}06$, Female $P = 0.03$, and Male $P = 3.27\text{e-}06$), followed by the Evans–Hoenig test (Population $P = 0.02$, Female $P = 0.18$, and Male $P = 0.02$) and finally the Bowker test that detected no significant bias (Population $P = 0.28$, Female $P = 0.47$, and Male $P = 0.45$). The gradient of significant test results suggested a slight bias that is not detected uniformly across pooling methods or sample type. The female age estimates were only significantly different with the McNemar test, whereas the male age estimates were significant for both the Evans–Hoenig and the McNemar tests and likely influenced the significant bias in the population sample results for the same two tests.

To better understand what differences are driving significant asymmetry, the age-bias plots were reviewed for error trends (Fig. 4). The X axis values at $y = 0$ represents 100% agreement between readers and a nonbiased dataset would demonstrate errors randomly distributed around the X axis. Between the ages of 60 and 100 y, errors are disproportionately distributed above or below the X axis on the Population and Male age-bias plots, indicating the likely age range driving significant test results. The mean error (difference) for the Population is +1.54

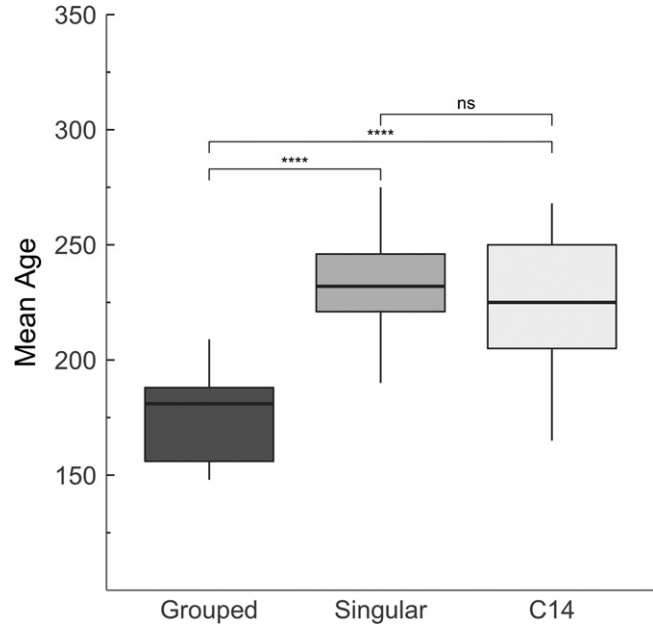


Figure 3. Comparison of aging strategies relative to carbon-14-estimated ages. A type III repeated analysis of variance identified significant difference in ages between hypotheses ($P = 5.16\text{e-}9$). Ages estimated using the grouped and singular strategies (Fig. 1) are significantly different (****), whereas no significant difference is observed between ages using the singular protocol and carbon-14 ages (ns) [posteriori pairwise comparison (pwc) t -test]. A Bonferroni correction was used to adjust for multiple comparisons.

y, Female is +0.40 y, and Male is +2.6 y using a standard deviation of 1.96. In other words, reader B, on average, ages 1.54 y higher than reader A on an animal that can live up to 261 y of age when the entire population sample is analyzed, but reader B ages, on average, 2.6 y higher than reader A when the male sample is analyzed. The higher mean error in the male data signified that young male samples may drive the bias results detected in the tests of symmetry, whereas error in the female dataset is evenly distributed around the agreement line (X axis, $y = 0$).

TABLE 1.

Test of symmetry results for pairwise age comparisons to identify bias.

Symmetry test	Sample	Degrees of freedom	Chi-square statistic	P value
Evans–Hoenig	Population	30	48.90	0.02*
	Female	29	35.80	0.18
	Male	26	42.10	0.02*
Bowker	Population	494	512	0.28
	Female	263	264	0.47
	Male	275	277	0.45
McNemar	Population	1	23.32	1.37e-06*
	Female	1	5.01	0.03*
	Male	1	21.65	3.27e-06*

Three tests were applied to an identical dataset. Significant P values (*) indicate a bias between the two age readers.

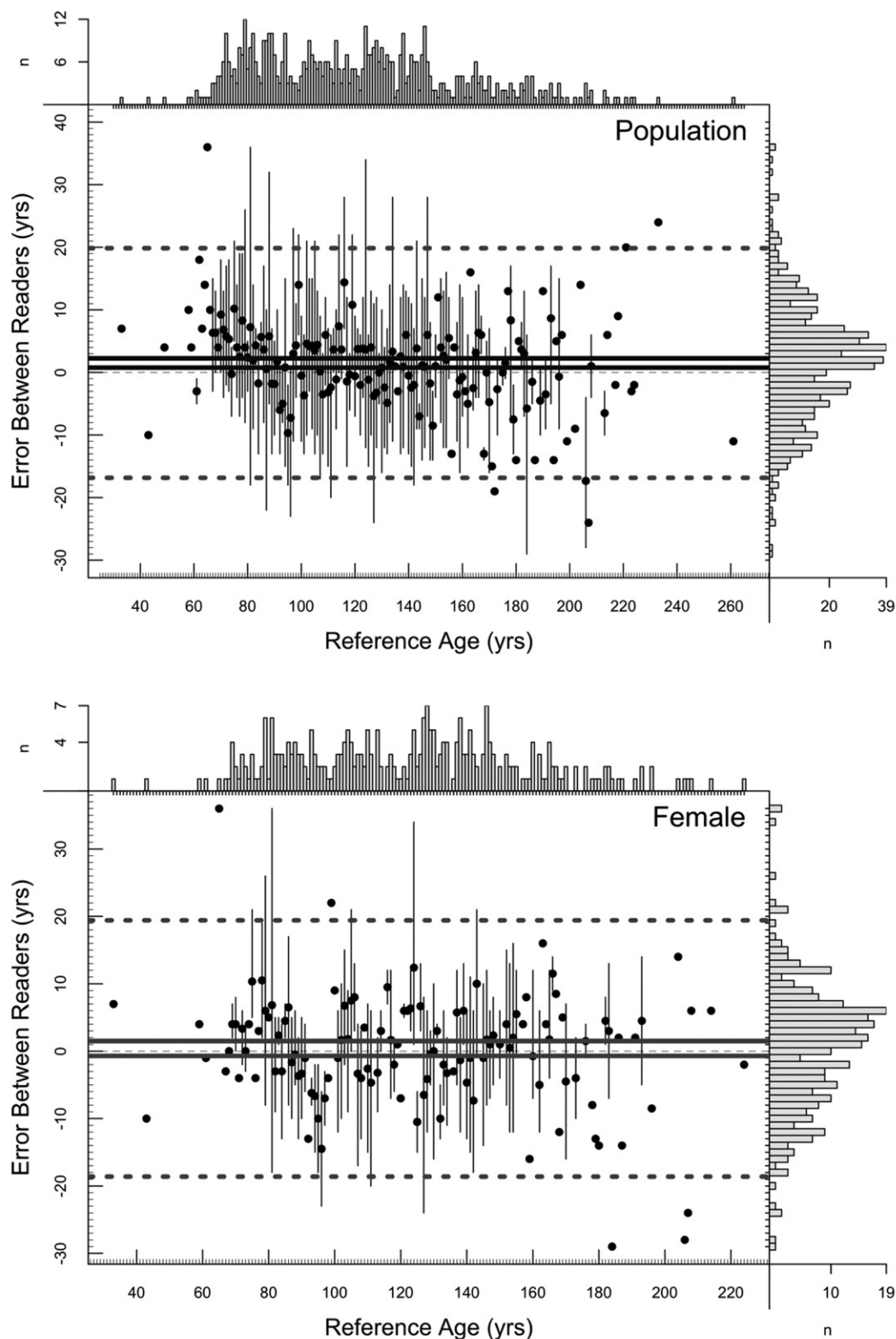


Figure 4. Difference in age estimates of a second reader (reader B) from the age estimates of the reference reader (i.e., reference age). Black points represent the mean difference in age between the two readers at a reference age and the vertical black lines represent the range of values if more than one error exists for that age (i.e., multiple samples). The two horizontal solid lines represent the 95% confidence intervals of the mean difference using a 1.96 standard deviation. The two horizontal dashed lines represent the 95% agreement bounds using a 1.96 standard deviation. Reader agreement is 100% at $y = 0$. The histogram on the Y axis denotes the frequency of difference values and the histogram on the X axis denotes the number of reader B ages at a given reference age. Mean Population error is +1.5 y ($n = 610$), mean Female error is +0.40 y ($n = 298$), and mean Male error is +2.6 y ($n = 312$).

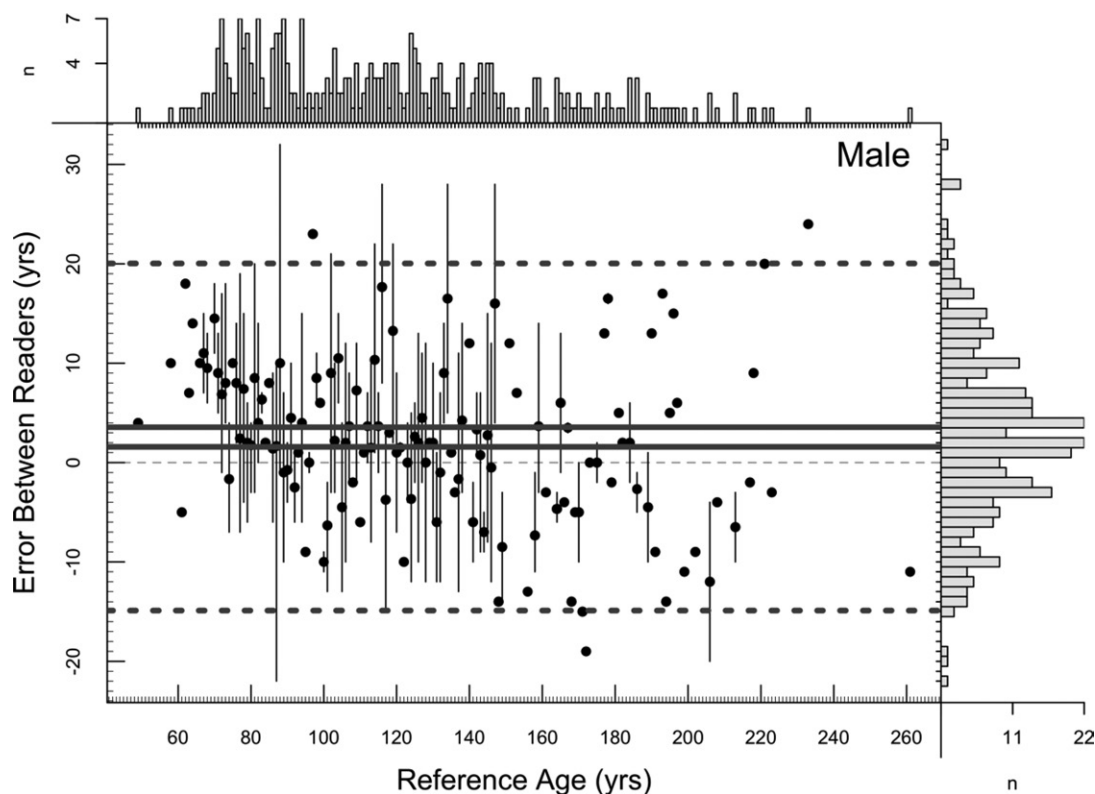


Figure 4. Continued.

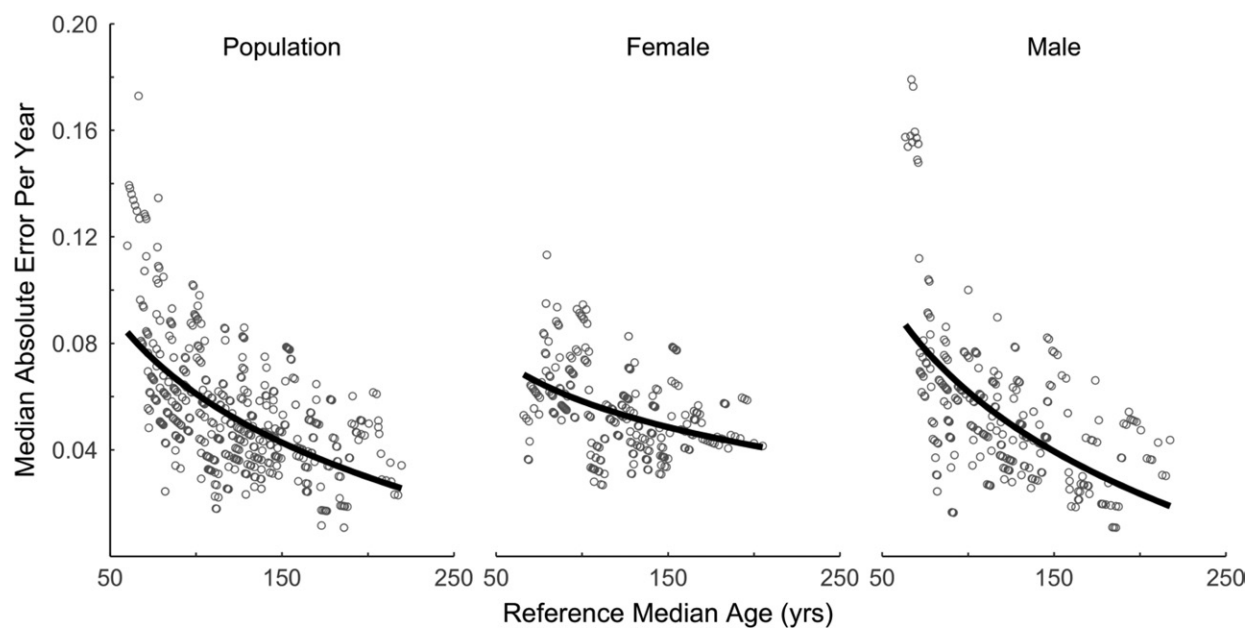


Figure 5. Rolling median absolute error per year smoothed over 10 sample increments (*Y* axis) versus the rolling median age of reference ages smoothed over the same 10 sample increments (*X* axis). A type III analysis of variance indicated a significant difference between median error rate (median absolute error per year) and median age (Population: $P = 3.0\text{e-}44$, Female: $P = 6.48\text{e-}12$, and Male: $P = 1.79\text{e-}23$). A logarithmic relationship provided the best fit for all sample groups. Population $f(x) = (-0.0452)\ln(x) + 0.27$ ($R^2 = 0.33$, $P = 2.2\text{e-}16$), Female $f(x) = (-0.0241)\ln(x) + 0.17$ ($R^2 = 0.16$, $P = 8.423\text{e-}13$), and Male $f(x) = (-0.0553)\ln(x) + 0.32$ ($R^2 = 0.34$, $P = 2.2\text{e-}16$).

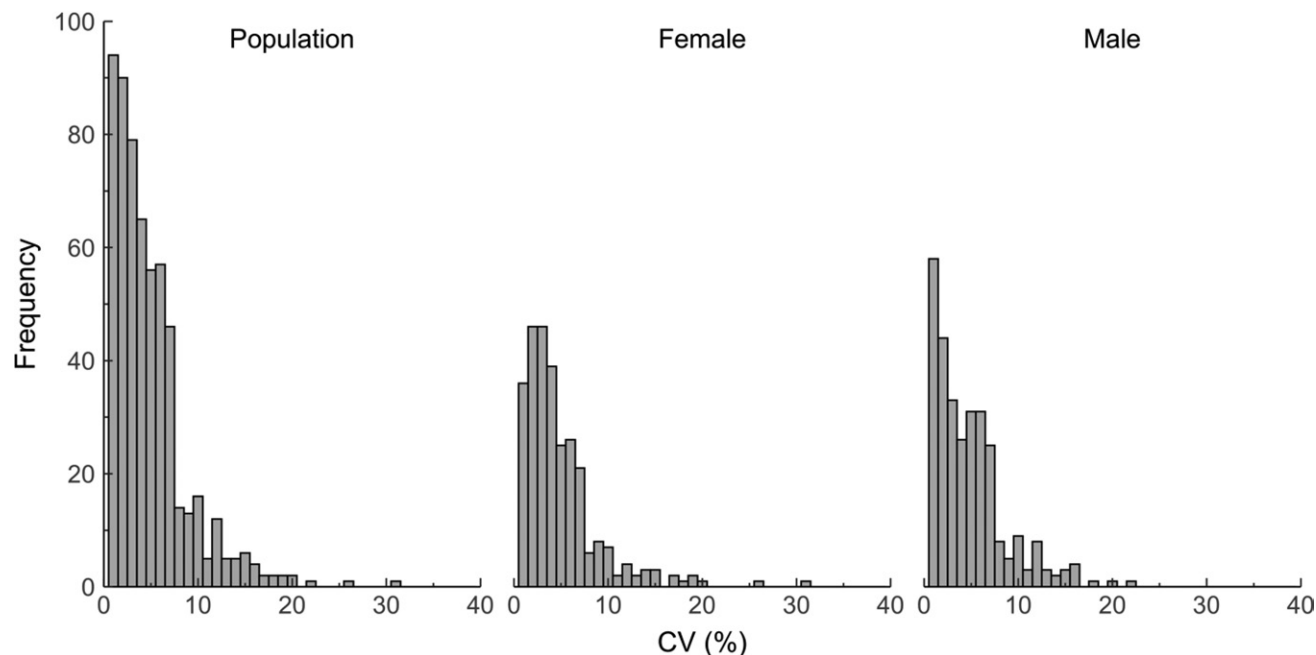


Figure 6. Frequency of coefficient of variation (CV) results. Average coefficient of variation (ACV) is 4.6% for the entire population and both the female and male subsets.

Extreme errors (outside the 95% agreement bounds using a 1.96 standard deviation) occurred across the entirety of the reference age range, and the absence of a trend indicated that these errors are the result of particularly challenging samples to age and not an underlying bias (i.e., processing error).

The age-bias plot indicated that a bias may be present in the younger animals of this study and the error rate was examined for similar trends in errors at age (Fig. 5). A negative logarithmic model best fit the data (Population: $R^2 = 0.33$, $P = 2.2\text{e-}16$; Female: $R^2 = 0.16$, $P = 8.423\text{e-}13$; Male: $R^2 = 0.34$, $P = 2.2\text{e-}16$) and the level of significance indicated a decline in error rate with increasing age. A type III one-way analysis of variance was performed to test the significant effect of specimen age at death on the error rate and all three sample types had a significant effect (Population: $P = 3.0\text{e-}44$, Female: $P = 6.48\text{e-}12$, and Male: $P = 1.79\text{e-}23$). The highest rate of error occurred at approximately 60 y of age and declined steadily with age and was highest for the male and population datasets. The error rate of a 60-y-old animal (0.12 errors/year) was 3.5 times higher than the error rate of a 220-y-old animal (0.034 errors/year). A significant relationship existed between female age at death and error rate, but both linear and negative logarithmic models fit the data similarly (linear: $R^2 = 0.15$, $P = 6.48\text{e-}12$) with a linear slope of nearly 0 ($-1.96\text{e-}04$). The combined data from the age-bias plots and the smoothed error rate plots revealed that the underlying bias is likely manifested in the youngest animals and males produced higher error rates in these young animals. Bias results indicated that age estimates can be accepted conditionally, where the greatest error rate and bias error occurred for the youngest, male animals in the population.

Age Precision

The Georges Bank samples had a population and sex-based ACV of 5% (Fig. 6); therefore, Georges Bank precision was high

and met ACV precision thresholds. Linear regression depicts a declining CV with the mean age for each set of pairwise ages (Fig. 7). The R^2 values are low (0.05–0.07) indicating that mean age may not be the primary source of variability in the data. Conversely, the relationship between mean age and CV is significant (Population: $P = 2.09\text{e-}10$, Female: $P = 4.45\text{e-}05$, and Male: $P = 9.03\text{e-}07$) where for every year increase in mean age, on average, CV declines by 0.03%.

Error Frequency

Samples aged from Georges Bank met the conditions of a 10% error frequency; 54 samples had CVs greater than 10%, a number fewer than expected by chance (binomial test, $P = 0.19$). The female dataset ($n = 298$) contained 24 errors and the male dataset ($n = 312$) contained 30 errors, both sexes fell within the 10% error frequency. When binomial tests were calculated using a range of expected probabilities (expected error frequency) between 0.01 and 0.2, the 53 population errors are significant with an error frequency set at less than or equal to 7%, the 24 female errors with an error frequency at less than or equal to 6%, and the 30 male errors with an error frequency at less than or equal to 8%.

DISCUSSION

The development of *Arctica islandica* age compositions for applications in standard fisheries assessment models requires solutions for the challenging nature of this species age-at-length data including the necessity for large sample sizes, constraining age precision and accuracy for an animal with greater than 200 age classes, and the time commitment and cost of aging such substantial sample sizes. Pace et al. (2017b) identified that number of cohorts in an *A. islandica* population, and the number

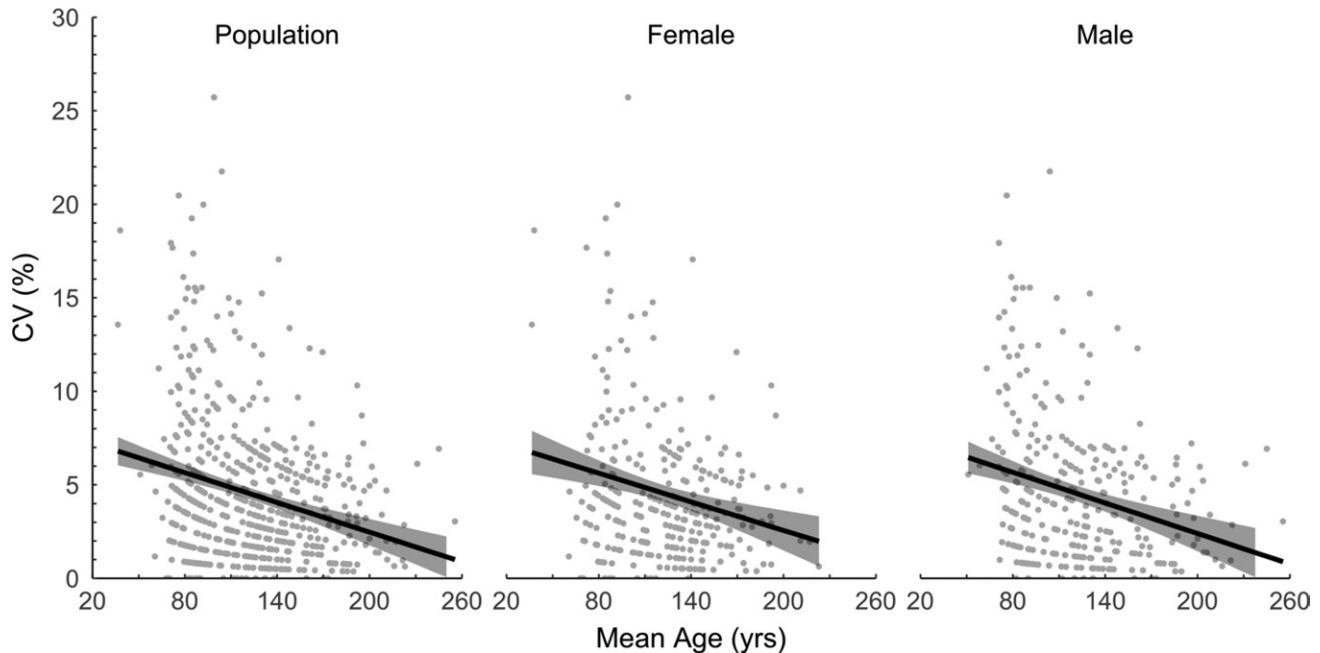


Figure 7. Relationship of coefficient of variation (CV) versus mean age of two age-reader age estimates. Linear regressions of the three samples have identical slopes (-0.03) and significant regressions: Population $y = -0.03x + 7.95$ ($R^2 = 0.06$, $P = 2.09\text{e-}10$), Female $y = -0.03x + 8.08$ ($R^2 = 0.05$, $P = 4.45\text{e-}05$), and Male $y = -0.03x + 7.85$ ($R^2 = 0.07$, $P = 9.03\text{e-}07$). R^2 values are low, an indication of a significant but poorly defined relationship between CV and mean age.

of cohorts within a narrow length class, required aging more than 20 animals per size class (i.e., greater than 200 animals) to construct robust population age compositions. Given the sample numbers needed and the time and cost commitment (Ropes 1984), maximizing precision in age determination is essential, as the employment of multiple readers to continuously age by consensus is infeasible.

Age determinations require levels of interpretation that inherently introduce error into the data. Historically, precision statistics including percent agreement, percent error, and CV were the only methods to assess error in fisheries age data (Campana et al. 1995, Campana 2001, McBride 2015). Precision statistics do not account for age effects and therefore can change based on the age of the animal (Hoenig et al. 1995). Early analyses regarding age precision in fisheries assumed that variability in age determinations was homogenous across a sample and consequently could be averaged across all age classes (Beamish & Fournier 1981, Chang 1982), yet it is now apparent that the precision of age determinations varies with the age of an animal and that age effect is an important variable to consider. Species as taxonomically divergent as *Arctica islandica* (Fig. 7) and the lemon shark (*Negaprion brevirostris*) (Brown & Gruber 1988) both demonstrate high CV and percent error (i.e., low precision) at young ages, whereas species such as walleye pollack often show low precision in the older individuals (Kimura & Lyons 1991, Hoenig et al. 1995). Precision variability within a species proves that precision is highly dependent on the species themselves and the age distribution of the sample. In other words, a sample dominated with young *A. islandica* will likely have lower precision than a sample primarily composed of older clams.

For many species without validated reference age collections, constraining precision and bias of age estimates is the

best strategy to improve the quality of the ages when accuracy is unknown, and a single test of error (e.g., CV) is not sufficient to accept age data (Beamish & McFarlane 1983, Campana et al. 1995). Error frequency was introduced in this paper as an additional method for evaluating error, but it is not a protocol used in many evaluations of aging precision. Because of the longevity of *Arctica islandica*, and the tendency of CV to obscure large differences in age estimates between readers for old animals because of age standardization, an option to evaluate exceptionally large CVs was desirable. The imposed error frequency threshold of 10% defined a limit on how many exceptionally large precision errors could arise in a dataset before the dataset is deemed to be unacceptable. A binomial test can then be applied to investigate alternative aging scenarios and determine the maximum number of errors a dataset can incur before it significantly exceeds a 10% error frequency. One such scenario is to identify how many errors would create an error frequency larger than 10%. For a 610-sample dataset, 74 population errors (or 40 female errors and 41 male errors) or more would exceed the designated acceptable error frequency. An alternative scenario is to test a more typical aging strategy where a second age reader only aged a random 20% subset of the sample (Kimura & Anderl 2005). In such a case, only 18 errors or fewer can be made within the 122-sample subset to maintain a 10% error frequency (or fewer than 10 errors in each of the female and male datasets). This approach to error also allows flexibility in implementation, either by changing the definition of an error to be more conservative, such as to match our acceptable ACV cutoff of 7%, or to allow more errors to occur so that the probability of error frequency surpasses 10%.

Accepting that a designated number of large precision errors can exist in a dataset, the identification of systematic patterns of error across age classes is critical to account for age effects. In

the age-bias plot (Fig. 4), the 95% agreement bounds are seemingly large (± 19 y from the mean), yet when compared with the data on spiny dogfish (*Squalus acanthias*) provided by Beamish and Fournier (1981) for which 95% agreement was within 8.3% of the total lifespan of the species (± 5 y for 95% agreement; 60-y lifespan), the 95% agreement for *Arctica islandica* appears reasonable. For *A. islandica*, the 95% agreement was within 7.3% of the total lifespan of the oldest individual from Georges Bank (± 19 y for 95% agreement; 261-y lifespan) or within 3.8% of the lifespan of the species (approximately 500 y).

Tests of symmetry to identify age bias have only recently been adopted in fisheries science and are often not reported alongside precision results (McBride 2015). Three tests of symmetry are easily calculated in contemporary age analyses, yet test selection and interpretation are easily confounded as evidenced by Table 1 where results are vastly different. The McNemar test was designed to perform a single paired test for the entire dataset, and in the case of age contingency tables, age is never accounted for. Regardless of whether the species has five age classes or 500 age classes, only one degree of freedom exists. The Bowker test is a pairwise comparison, where every cell is compared with its mirror image across the diagonal and no pooling occurs. When the Bowker test is used to analyze a species with many age classes, the degrees of freedom (or number of paired comparisons) will be high, as evidenced by *Arctica islandica*, whereas when a species with few age classes is analyzed, the degrees of freedom will be low and potentially similar to that of the Evans–Hoenig test. The Evans–Hoenig test pools comparisons based on the degree of difference from the diagonal (or 100% agreement). The Evans–Hoenig test is the only test of symmetry specifically designed for fisheries science to evaluate how age differences are dispersed around the agreement age (i.e., the diagonal) (Evans & Hoenig 1998).

For *Arctica islandica*, age-bias plots (Fig. 4) and error rate (Fig. 5) demonstrated that deviations in age estimates occurred more frequently in the youngest animals and particularly in young male animals and that a bias may be present in those samples. The bias is not necessarily large, but an underlying trend is observed across all three error methods. The McNemar test is the most sensitive and always detected a bias in this dataset. Conversely, the Bowker test was the least sensitive. Arguably, the Evans–Hoenig test was the most reliable test for bias detection as female data were not significantly biased in this test of symmetry, which is supported by error rate, error frequency, and age-bias plot results. Interpretation of these three tests would be extremely difficult if multiple representations of error were not available. Diverse methods to describe error are critical to identify the origin of uncertainty, and to implement procedures to target the most significant sources.

In the case of *Arctica islandica*, uncertainty in age estimates appears to originate primarily during the first decades of life as clearly shown by the ascending error rate (Fig. 5) and CV (Fig. 7) with younger and younger ages. The CV was developed to standardize error by age and error rate was also calculated on a per year basis, whereby an older animal will need larger errors [standard deviation (CV), age differences (error rate)] than a young animal to manifest the same magnitude of standardized error. Furthermore, as is common across many sclerochronological datasets from otoliths to bivalves, as growth rates decline with age, the ability to observe intraannual (subannual) growth lines is diminished. As a result, age readers tend to agree more in the

latter years when every line is viewed as a clear annulus, whereas early growth increments are large enough to display subannual changes in growth rates that manifest as repetitive growth lines that are not true annuli (Pannella 1971). The presence of intraannual growth lines in rapidly growing bivalves is well known and their discrimination is normally a challenge (Jacobson et al. 2006). Reducing this source of uncertainty is clearly the primary challenge in aging *A. islandica* (Harding et al. 2008).

Growth lines are created when shell carbonate production slows, and more protein is secreted into the shell matrix. These dark, protein-dense growth lines reflect seasonal depressions in carbonate production because of reduced food supply, spawning events, or unfavorable stratification/mixing that result in suppressed metabolic functions required for growth. Normally, the winter cessation of growth generates the strongest growth line (annulus) in most bivalve species because of cold water temperatures that mark the end of the annual growth period (Jones & Quitmyer 1996, Fan et al. 2011, Chute et al. 2016). A high temperature-induced growth line may produce the primary annulus in some species, however (Peterson et al. 1985, Goodwin et al. 2001). The transition zone from fast juvenile growth to slower adult growth is the most challenging section of the hinge to age when a reader must decide when each growth line is a true annulus. For this reason, the expectation, clearly demonstrated by this *Arctica islandica* dataset, is that precision will be low (i.e., high CV) for young animals where many of the annuli are intermixed with subannual growth lines and in which increased scope for growth permits growth over a longer season than observed in the adult animal (Hofmann et al. 2006, Munroe et al. 2013).

Bias error was higher for male clams (Table 1, Figs. 4 and 5) despite an identical ACV between sexes and for the entire population sample (4.6%) (Figs. 6 and 7); clear evidence that precision alone is not a sufficient metric to describe the quality of age estimates. Ropes et al. (1984a) noted that gametogenesis was initiated in males at a smaller size and younger age than females. Possibly, the earlier onset of maturity might increase the number of subannual growth lines in young males, though the physiological mechanism is unclear. The expression of additional subannual growth lines in males relative to females is thus unexplained, but clearly present and results in an increase in contrasting interpretations of true annuli between age readers and an increased occurrence of aging error in males.

The carbon-14 dating used a selective sample of old animals to illuminate what an accurate age for an old *Arctica islandica* may be. A critical realization is that high precision does not necessarily mean high accuracy. Independent validations of accuracy are important. For *A. islandica* and other bivalves, age validations have generally been provided by carbon-14 dating, amino acid racemization dating, cross dating (as used in dendrochronology), or oxygen isotopes (Weidman et al. 1994, Machitto et al. 2000, Schöne et al. 2011, Wanamaker et al. 2011, Mette et al. 2016, Reynolds et al. 2016). None of these methods can provide an adequate sample size for fisheries assessment purposes; hence, continued focus on reader precision in determining age, whereas accuracy validation from small subsamples will remain essential (Beamish & McFarlane 1983). As growth rates continue to accelerate over time in portions of the *A. islandica* range (Pace et al. 2018), younger and younger animals will be available to the fishery and these young animals will bring higher rates of error. Thus, both precision and accuracy

can be best improved by focusing on the shell growth dynamics of young (but sexually determined) male and female animals, thereby improving the discrimination of subannual increments from annuli.

CONCLUSIONS

A 3-fold error study indicated that the 610-sample *Arctica islandica* age dataset from Georges Bank met the predetermined error thresholds for bias (conditionally because of significant and nonsignificant results), precision (ACV less than 7%), and error frequency (less than 10%). Pending improved age-validation data for this species, particularly for the younger animals entering the fishery, these age data are within acceptable error bounds proposed in this paper to be used for age compositions and suggest that the reader aging protocol can be used in future age-structure studies. These analyses also establish the degree of uncertainty to attach to age compositions derived for implementation in fisheries assessment models. The representativeness of the Georges Bank population for error applications generally is, as yet, unknown, though published growth rates on Georges Bank are thought to be higher than other locations at similar latitudes (Ropes & Pyoas 1982, Lewis et al. 2001, but see Pace et al. 2018). Also unclear is the degree to which these higher growth rates might provide reduced precision relative to animals aged from other regions, as lower precision in this study was associated with periods of higher growth rate. Regardless, the degree of uncertainty places a detection limit on identifying the shortest detectable period of low recruitment, a consideration of some importance, given the population dynamics of this species. Age frequencies derived are, in effect, smoothed by this degree of error and pose a limitation on the interpretation of fine-scale variations in the inferred cohort dynamics within the population.

Given the cost of processing and the number of aged animals required to provide an adequate age-at-length relationship across many ages wherein high variability exists in age at

length (Pace et al. 2017a, 2017b), attention to increasing precision is necessary and potentially result in the reduction of required sample size. The differential error rate between males and females provides a possible opportunity to reduce age determination bias by focusing on the females. Precision was clearly greater and bias less for females. The tendency for females to be larger than males (Ropes et al. 1984a, Thórarinsdóttir & Steingrímsson, 2000) would suggest some bias in the sex ratio of landings as well, which would support the preferential use of female SSB in an assessment (Wilderbuer & Turnock 2009, Powell et al. 2013, Okamura et al. 2014). Although a number of studies have examined *Arctica islandica* aging methods in the Mid-Atlantic region (Jones 1980, Murawski et al. 1982, Ropes 1984, Ropes et al. 1984b, Weidman et al. 1994), a focus on the increased uncertainty in age determination at small size, which is conflated with the number of males in those size classes, has not occurred (Harding et al. 2008). Nonetheless, this study suggests that a focus on females would reduce uncertainty in the age frequency and possibly reduce the required sample number to produce a reliable age-at-length key and subsequent population age-frequency distribution.

ACKNOWLEDGMENTS

The authors would like to thank the Northeast Fisheries Science Center, particularly the Fishery Biology Program and the Population Dynamics Branch, for their expert mentorship and hospitality during this research project. The Washington Department of Fish and Wildlife provided additional correspondence regarding their long-lived species and precision methodologies used by them, which guided the determination of reference points for ocean quahog error thresholds. This research was supported by the National Science Foundation (NSF) Science Center for Marine Fisheries (SCeMFIS) under NSF awards 1266057 and 1841112, through membership fees provided by the SCeMFIS Industry Advisory Board and an NSF nonacademic research internship.

LITERATURE CITED

- Beamish, R. J. & D. A. Fournier. 1981. A method for comparing the precision of a set of age determinations. *Can. J. Fish. Aquat. Sci.* 38:982–983.
- Beamish, R. J. & G. A. McFarlane. 1983. The forgotten requirement for age validation in fisheries biology. *Trans. Am. Fish. Soc.* 112:735–743.
- Bowker, A. H. 1948. A test for symmetry in contingency tables. *J. Am. Stat. Assoc.* 43:572–574.
- Brooks, E. N., K. W. Shertzer, T. Gedanke & D. S. Vaughan. 2008. Stock assessment of protogynous fish: evaluating measures of spawning biomass used to estimate biological reference points. *Fish. Bull.* 106:12–23.
- Brown, C. & S. Gruber. 1988. Age assessment of the lemon shark, *Negaprion brevirostris*, using tetracycline validated vertebral centra. *Copeia* 3:747–753.
- Butler, P. G., C. A. Richardson, J. D. Scourse, R. Witbaard, B. R. Schöne, N. M. Fraser, A. D. Wanamaker, Jr., C. L. Bryant, I. Harris & I. Robertson. 2009. Accurate increments identification and the spatial extent of the common signal in five *Arctica islandica* chronologies from the Fladen Ground, northern Baltic Sea. *Paleoceanography* 24:1–18.
- Butler, P. G., A. D. Wanamaker, Jr., J. D. Scourse, C. A. Richardson & D. J. Reynolds. 2013. Variability in marine climate on the North Icelandic Shelf in a 1357-year proxy archive based on growth increments in the bivalve *Arctica islandica*. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 373:141–151.
- Campana, S. E. 2001. Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *J. Fish Biol.* 59:197–242.
- Campana, S. E., M. C. Annand & J. K. McMillan. 1995. Graphical and statistical methods for determining the consistency of age determinations. *Trans. Am. Fish. Soc.* 124:131–138.
- Campana, S. E. & C. M. Jones. 1992. Analysis of otolith microstructure data. In: Stevenson, D. K. & S. E. Campana, editors. Otolith microstructure examination and analysis. *Can. Spec. Publ. Fish. Aquat. Sci.* 117:73–100.
- CARE. 2006. Committee of age reading experts manual on generalized age determination procedures for groundfish. Seattle, WA: The Canada/U.S. Groundfish Committee. 52 pp.
- Chang, W. Y. B. 1982. A statistical method for evaluating the reproducibility of age determination. *Can. J. Fish. Aquat. Sci.* 39:1208–1210.
- Chen, Z., E. Curchitser, R. Chant & D. Kang. 2018. Seasonal variability of the cold pool over the Mid-Atlantic Bight continental shelf. *J. Geophys. Res. Oceans* 123:8203–8226.
- Chute, A. S., R. S. McBride, S. J. Emery & E. Robillard. 2016. Annulus formation and growth of Atlantic surfclam (*Spisula solidissima*)

- along a latitudinal gradient in the western North Atlantic Ocean. *J. Shellfish Res.* 35:729–737.
- Costa, E. G., R. M. Lopes & J. M. Singer. 2016. Sample size for estimating the mean concentration of organisms in ballast water. *J. Environ. Manage.* 180:433–438.
- Duval, S. & R. Tweedie. 2000. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56:455–463.
- Evans, G. T. & J. M. Hoenig. 1998. Testing and viewing symmetry in contingency tables, with application to readers of fish ages. *Biometrics* 54:620.
- Fan, C., P. Koeniger, H. Wang & M. Frechen. 2011. Ligamental increments of the mid-Holocene Pacific oyster *Crassostrea gigas* are reliable independent proxies for seasonality in the western Bohai Sea, China. *Paleogeogr. Palaeoclimatol. Palaeoecol.* 299:437–448.
- Goodwin, D. H., K. W. Flessa, B. R. Schöne & D. L. Dettman. 2001. Cross-calibration of daily growth increments, stable isotope variation, and temperature in the Gulf of California bivalve mollusk *Chione cortezii*: implications for paleoenvironmental analysis. *Palaio* 16:387–398.
- Harding, J. M., S. E. King, E. N. Powell & R. Mann. 2008. Decadal trends in age structure and recruitment patterns of ocean quahogs *Arctica islandica* from the Mid-Atlantic Bight in relation to water temperatures. *J. Shellfish Res.* 27:667–690.
- Heaton, T. J., P. Köhler, M. Butzin, E. Bard, R. W. Reimer, W. E. N. Austin, C. B. Ramsey, P. M. Grootes, K. A. Hughen, B. Kromer, P. J. Reimer, J. Adkins, A. Burke, M. S. Cook, J. Olsen & L. C. Skinner. 2020. Marine20—the marine radiocarbon age calibration curve (0–55,000 Cal BP). *Radiocarbon* 62:779–820.
- Hjellvik, V., O. R. Godø & D. Tjøstheim. 2002. The measurement error in marine survey catches: the bottom trawl case. *Fish. Bull.* 100:720–726.
- Hoenig, J. M. 2017. Should natural mortality estimators based on maximum age also consider sample size. *Trans. Am. Fish. Soc.* 146:136–146.
- Hoenig, J. M., M. J. Morgan & C. A. Brown. 1995. Analysing differences between two age determination methods by tests of symmetry. *Can. J. Fish. Aquat. Sci.* 52:364–368.
- Hofmann, E. E., J. M. Klinck, J. N. Kraeuter, E. N. Powell, R. E. Grizzle, S. C. Buckner & V. M. Bricelj. 2006. A population dynamics model of the hard clam, *Mercenaria*: development of the age- and length-frequency structure of the population. *J. Shellfish Res.* 25:417–444.
- Hudson, J. H. 1981. Growth rates in *Montastrea annularis*: a record of environmental change in Key Largo Coral Reef Marine Sanctuary, Florida. *Bull. Mar. Sci.* 31:444–459.
- Hulson, P. J. F., D. H. Hanseltnan & S. K. Shotwell. 2017. Investigations into the distribution of sample sizes for determining age composition of multiple species. *Fish. Bull.* 115:326–342.
- Huyghe, D., M. de Rafelis, M. Ropert, V. Mouchi, L. Emmanuel, M. Renard & F. Lartaud. 2019. New insights into oyster high-resolution hinge growth patterns. *Mar. Biol.* 166:48.
- Jacobson, L. D., D. Hart, K. D. E. Stokesbury, T. Jaffarian, M. A. Allard, M. C. Marino, II, A. Chute, J. I. Nogueira, B. P. Harris & P. Rago. 2010. Measurement errors in body size of sea scallops (*Placopecten magellanicus*) and their effect on stock assessment models. *Fish. Bull.* 108:233–247.
- Jacobson, L., S. Sutherland, J. Burnett, M. Davidson, J. Harding, J. Normant, A. Picariello & E. Powell. 2006. Report from the Atlantic surfclam (*Spisula solidissima*) aging workshop. Woods Hole, MA: Northeast Fisheries Science Center. NEFSC Ref. Doc. 06-12. 24 pp.
- Johnsen, E. 2003. Improving the precision of length frequency distribution estimates from trawl surveys by including spatial covariance—using Namibian *Merluccius capensis* as an example. *Fish. Res.* 62:7–20.
- Jones, D. S. 1980. Annual cycle of shell growth increment formation in two continental shelf bivalves and its paleoecologic significance. *Palaio* 6:331–340.
- Jones, D. S. & I. R. Quitmyer. 1996. Marking time with bivalve shells: oxygen isotopes and season of annual increment formation. *Palaio* 11:340–346.
- Kimura, D. K. 1977. Statistical assessment of the age-length key. *J. Fish. Res. Board Can.* 34:317–324.
- Kimura, D. K. & D. M. Anderl. 2005. Quality control of age data at the Alaska Fisheries Science Center. *Mar. Freshw. Res.* 56:783–789.
- Kimura, D. K. & J. J. Lyons. 1991. Between-reader bias and variability in the age-determination process. *Fish. Bull.* 89:53–60.
- Kubota, K., K. Shirai, N. Muralcami-Sugihara, K. Seike, M. Hori & K. Tanabe. 2017. Annual shell growth pattern of the Stimpson's hard clam *Mercenaria stimpsoni* as revealed by sclerochronological and oxygen stable isotope measurements. *Palaeoecogr. Palaeoclimatol. Palaeoecol.* 465:307–315.
- Lee, H.-H., M. N. Maunder, K. R. Piner & R. D. Methot. 2011. Estimating natural mortality within a fisheries stock assessment model: an evaluation using simulation analysis based on twelve stock assessments. *Fish. Res.* 109:89–94.
- Lentz, S. J. 2017. Seasonal warming of the Middle Atlantic Bight cold pool. *J. Geophys. Res. Oceans* 122:941–954.
- Lewis, C. V. W., J. R. Weinberg & C. S. Davis. 2001. Population structure and recruitment of the bivalve *Arctica islandica* (Linnaeus, 1767) on Georges Bank from 1980–1999. *J. Shellfish Res.* 20:1135–1144.
- MacDonald, P. D. M. & T. J. Pitcher. 1979. Age-groups from size-frequency data: a versatile and efficient method of analyzing distribution mixtures. *J. Fish. Res. Board Can.* 36:987–1001.
- Machitto, T. M., Jr., G. A. Jones, G. A. Goodfriend & C. R. Weidman. 2000. Precise temporal correlation of Holocene mollusk shells using sclerochronology. *Quat. Res.* 53:236–246.
- Martell, S. J. D., W. E. Pine, III & C. J. Walters. 2008. Parameterizing age-structured models from a fisheries management perspective. *Can. J. Fish. Aquat. Sci.* 65:1586–1600.
- McBride, R. S. 2015. Diagnosis of paired age agreement: a simulation of accuracy and precision effects. *ICES J. Mar. Sci.* 72:2149–2167.
- McNemar, Q. 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 12:153–157.
- Mette, M. J., A. D. Wanamaker, Jr., M. L. Carroll, W. G. Ambrose, Jr. & M. J. Retella. 2016. Linking large-scale climate variability with *Arctica islandica* shell growth and geochemistry in northern Norway. *Limnol. Oceanogr.* 61:748–764.
- Minte-Vera, C. V., M. N. Maunder, K. M. Schaefer & A. M. Aires-da-Silva. 2019. The influence of metrics for spawning output on stock assessment results and evaluation of reference points: an illustration with yellowfin tuna in the eastern Pacific ocean. *Fish. Res.* 217:35–45.
- Mohn, R. 1994. A comparison of three methods to convert catch at length data into catch at age. *Int. Comm. Conserv. Atl. Tuna* 42:110–119.
- Munroe, D. M., E. N. Powell, R. Mann, J. M. Klinck & E. E. Hofmann. 2013. Underestimation of primary productivity on continental shelves: evidence from maximum size of extant surfclam (*Spisula solidissima*) populations. *Fish. Oceanogr.* 22:220–233.
- Murawski, S., J. W. Ropes & F. M. Serchuk. 1982. Growth of the ocean quahog, *Arctica islandica*, in the Middle Atlantic Bight. *Fish. Bull.* 80:21–34.
- NEFSC. 2017. 63rd northeast regional stock assessment workshop (63rd SAW) assessment report. Woods Hole, MA: Northeast Fisheries Science Center. NEFSC Ref. Doc. 17-10. 414 pp.
- Ogle, D. H., P. Wheeler & A. Dinno. 2021. FSA: fisheries stock analysis. R package version 0.8.23. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://github.com/droglenc/FSA>.
- Okamura, H., M. K. McAllister, M. Ichinokawa, L. Yamanoka & K. Holt. 2014. Evaluation of the sensitivity of biological reference points to the spatio-temporal distribution of fishing effort when seasonal migrations are sex-specific. *Fish. Res.* 158:116–123.
- Pace, S. M., E. N. Powell & R. Mann. 2018. Two-hundred year record of increasing growth rates for ocean quahogs (*Arctica islandica*) from the northwestern Atlantic Ocean. *J. Exp. Mar. Biol. Ecol.* 503:8–22.
- Pace, S. M., E. N. Powell, R. Mann & M. C. Long. 2017a. Comparison of age-frequency distributions for ocean quahogs *Arctica islandica* on the western Atlantic US continental shelf. *Mar. Ecol. Prog. Ser.* 585:81–98.

- Pace, S. M., E. N. Powell, R. Mann, M. C. Long & J. M. Klinck. 2017b. Development of an age-frequency distribution for ocean quahogs (*Arctica islandica*) on Georges Bank. *J. Shellfish Res.* 36:41–53.
- Pannella, G. 1971. Fish otoliths: daily growth layers and periodical patterns. *Science* 173:1124–1127.
- Pennington, M., L.-M. Burmeister & V. Hjellvik. 2002. Assessing the precision of frequency distributions estimated from trawl-survey samples. *Fish. Bull.* 100:74–80.
- Pentilla, J. & L. M. Dery, editors. 1988. Age determination methods for Northwest Atlantic species. Woods Hole, MA: National Marine Fisheries Service. NMFS Report No. 72. 135 pp.
- Peterson, C. H., P. B. Duncan, H. C. Summerson & B. F. Beal. 1985. Annual band deposition within shells of the hard clam, *Mercenaria*: consistency across habitat near Cape Lookout, North Carolina. *Fish. Bull.* 88:671–677.
- Powell, E. N., R. Mann, K. A. Ashton-Alcox, K. M. Kuykendall & M. C. Long. 2017. Can we estimate molluscan abundance and biomass on the continental shelf? *Estuar. Coast. Shelf Sci.* 198:213–224.
- Powell, E. N., J. M. Morson, K. A. Ashton-Alcox & Y. Kim. 2013. Accommodation of the sex-ratio in eastern oysters *Crassostrea virginica* to variation in growth and mortality across the estuarine salinity gradient. *J. Mar. Biol. Ass. U.K.* 93:533–555.
- Purroy, A., S. Milano, B. R. Schöne, J. Thébault & M. Peharda. 2018. Drivers of shell growth of the bivalve, *Callista chione* (L. 1758)—combined environmental and biological factors. *Mar. Environ. Res.* 134:138–149.
- R Core Team. 2018. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: <http://www.R-project.org>.
- Reynolds, D. J., C. A. Richardson, J. D. Scourse, P. E. Butler, P. Hollyman, A. Pomán-González & I. R. Hall. 2017. Reconstructing North Atlantic marine climate variability using an absolutely-dated sclerochronological network. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 465:333–346.
- Reynolds, D. J., J. D. Scourse, P. R. Halloran, A. J. Niederbragt, A. D. Wanamaker, P. G. Butler, C. A. Richardson, J. Heinemeier, J. Eiriksson, K. Knudsen & I. R. Hall. 2016. Annually resolved North Atlantic marine climate over the last millennium. *Nat. Commun.* 7:13502.
- Richardson, C. A. 2001. Molluscs as archives of environmental change. *Oceanogr. Mar. Biol. Annu. Rev.* 39:103–164.
- Ridgway, I. D., C. A. Richardson, J. D. Scourse, P. G. Butler & D. J. Reynolds. 2012. The population structure and biology of the ocean quahog, *Arctica islandica*, in Belfast Lough, northern Ireland. *J. Mar. Biol. Ass. U.K.* 92:539–546.
- Ritter, M. N., H. Francischini, L. A. Kuhn, N. C. da Luz, F. H. Michels, A. L. M. de Moraes, P. A. V. Paim & P. L. A. Xavier. 2016. El sesgo del operador en la replicabilidad de los estudios tafonómicos comparativos. *Rev. Bras. Paleontol.* 19:449–464.
- Ropes, J. W. 1984. Procedures for preparing acetate peels and evidence validating the annual periodicity of growth lines formed in the shells of ocean quahogs, *Arctica islandica*. *Mar. Fish. Rev.* 46:27–35.
- Ropes, J. W. 1988. Ocean quahog, *Arctica islandica*. In: Pentilla, J. & L. M. Dery, editors. Age determination methods for Northwest Atlantic species. Woods Hole, MA: National Marine Fisheries Service. NMFS Report No. 72. pp. 129–130.
- Ropes, J. W., D. S. Jones, S. A. Murawski, F. M. Serchuk & A. Jearld, Jr. 1984a. Documentation of annual growth lines in ocean quahogs, *Arctica islandica* Linné. *Fish. Bull.* 82:1–19.
- Ropes, J. W., S. A. Murawski & F. M. Serchuk. 1984b. Size, age, sexual maturity, and sex ratio in ocean quahog, *Arctica islandica* Linné, off Long Island, New York. *Fish. Bull.* 82:253–266.
- Ropes, J. W. & D. Pyoas. 1982. Preliminary age and growth observations of ocean quahogs, *Arctica islandica* Linné, from Georges Bank. San Diego, CA: ICES C.M. Report No. 1982/K:15. 6 pp.
- Schöne, B. R., J. Fiebig, M. Pfeiffer, R. Gleß, J. Hickson, A. L. A. Johnson, W. Dreyer & W. Oschmann. 2005. Climate records from bivalved Methuselah (*Arctica islandica*, Mollusca; Iceland). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 228:130–148.
- Schöne, B. R., A. D. Wanamaker, Jr., J. Fiebig, J. Thébault & K. Krentz. 2011. Annually resolved $\delta^{13}\text{C}$ shell chronologies of long-lived bivalve mollusks (*Arctica islandica*) reveal oceanic carbon dynamics in the temperate North Atlantic during recent centuries. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 302:31–42.
- Sha, J., Y.-H. Jo, X.-H. Yan & W. T. Liu. 2015. The modulation of the seasonal cross-shelf sea level variation by the cold pool in the Middle Atlantic Bight. *J. Geophys. Res. Oceans* 120:7182–7194.
- Sherwood, O. A., E. N. Edinger, T. P. Guilderson, B. Ghaleb, M. J. Risk & D. B. Scott. 2008. Late Holocene radiocarbon variability in Northwest Atlantic slope waters. *Earth Planet. Sci. Lett.* 275:146–153.
- Shirai, K., K. Kubota, N. Murakami-Sugihara, R. Seike, M. Hakozaiki & K. Tanabe. 2018. Stimpson's hard clam *Mercenaria stimpsoni*: a multi-decadal climate recorder for the northwest Pacific coast. *Mar. Environ. Res.* 133:49–56.
- Stari, T., K. F. Preedy, E. McKenzie, W. S. C. Gurney, M. R. Heuth, P. A. Kunzlik & D. C. Speirs. 2010. Smooth age length keys: observations and implications for data collection on North Sea haddock. *Fish. Res.* 105:2–12.
- Stuiver, M. & H. A. Polach. 1977. Discussion reporting of ^{14}C data. *Radiocarbon* 19:355–363.
- Swart, P. K. 2015. The geochemistry of carbonate diagenesis: the past, present and future. *Sedimentology* 62:1233–1304.
- Thórarinsdóttir, G. G. & S. A. Steingrímsson. 2000. Size and age at sexual maturity and sex ratio in ocean quahog, *Arctica islandica* (Linnaeus, 1767), off northwest Iceland. *J. Shellfish Res.* 19:943–947.
- Tisnérat-Laborde, N., M. Paterne, B. Bétivier, M. Arnold, P. Yiou, D. Blamart & S. Raynaud. 2010. Variability of the northeast Atlantic sea surface $\Delta^{14}\text{C}$ and marine reservoir age and the North Atlantic Oscillation (NAO). *Quat. Sci. Rev.* 29:2633–2646.
- Wanamaker, A. D., Jr., K. J. Kreutz, B. R. Schöne & D. S. Introne. 2011. Gulf of Maine shells reveal changes in seawater temperature seasonality during the medieval climate anomaly and the Little Ice Age. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 302:47–51.
- Wanamaker, A. D., Jr., K. J. Kreutz, B. R. Schöne, K. A. Maasch, A. J. Pershing, H. W. Burns, D. S. Introne & S. Feindel. 2009. A late Holocene paleo-productivity record in the western Gulf of Maine, USA, inferred from growth histories of the long-lived ocean quahog (*Arctica islandica*). *Int. J. Earth Sci.* 98:19–29.
- Weidman, C. R. & G. A. Jones. 1993. A shell-derived time history of bomb ^{14}C on Georges Bank and its Labrador Sea implications. *J. Geophys. Res. Oceans* 98:14577–14588.
- Weidman, C. R., G. A. Jones & K. C. Lohmann. 1994. The long-lived mollusc *Arctica islandica*: a new paleoceanographic tool for the reconstruction of bottom temperatures for the continental shelves of the northern North Atlantic Ocean. *J. Geophys. Res. Oceans* 99:18305–18314.
- Weinberg, J. R. 1999. Age-structure, recruitment, and adult mortality in populations of the Atlantic surfclam, *Spisula solidissima*, from 1978–1997. *Mar. Biol.* 134:113–125.
- Wilderbuer, T. K. & B. J. Turnock. 2009. Sex-specific natural mortality of arrowtooth flounder in Alaska: implications of a skewed sex ratio on exploitation and management. *N. Am. J. Fish. Manage.* 29:306–322.
- Witbaard, R., M. I. Jenness, K. van der Borg & G. Gaussen. 1994. Verification of annual growth increments in *Arctica islandica* L. from the North Sea by means of oxygen and carbon isotopes. *Neth. J. Sea Res.* 33:91–101.