North American Journal of Fisheries Management 21:374–380, 2001

Factors Influencing Precision of Age Estimation from Scales and Otoliths of Bluegills in Illinois Reservoirs

R. JOHN H. HOXMEIER,* D. DEREK ADAY, AND DAVID H. WAHL

Kaskaskia Biological Station, Center for Aquatic Ecology,
Illinois Natural History Survey,
Rural Route 1, Box 157, Sullivan, Illinois 61951, USA

Abstract.—We examined the effects of population-specific variation on age estimates from scales and otoliths of bluegills *Lepomis macrochirus* from Illinois reservoirs. We divided 12 reservoirs into four groups (north stunted, north quality, south stunted, and south quality) to examine the effects of latitude and fish size distribution on the precision of estimated ages. Ages of 40 bluegills from each reservoir (total $N = 480$) were independently estimated by two readers. Otoliths provided more precise age estimates than scales. Population size structure and sex had no effect on precision of ages estimated from either structure; however, latitude, age, and maturity stage all affected precision of ages estimated from scales. Age also affected precision of estimates from otoliths. Percent agreement, coefficient of variation, and age bias plots all provided useful interpretations of the data. Our results demonstrate the importance of examining population-specific sources of variation with multiple statistical methods when comparing ages of fish from different populations.

The assessment of ages within and among populations is a critical component of many fisheries investigations. Data on fish ages are necessary to examine growth, mortality, and reproductive characteristics (e.g., age of maturation) of a population. Although fish ages can be estimated through statistical analyses (e.g., length histograms, age–length keys), direct estimation from bony structures such as scales and otoliths is currently the most widely used method (DeVries and Frie 1996). Because biologists must often depend on multiple readers and multiple structures to evaluate fish ages in long-term data sets, the precision of age estimation has received much attention (e.g., Beamish and McFarlane 1987; Chilton and Stocker 1987; Hammers and Miranda 1991; Hoenig et al. 1995).

Previous studies have primarily addressed precision among multiple readers or multiple aging structures with fish collected from either single or a few populations (Boxrucker 1986; Sharp and Bernard 1988; Kruse et al. 1993; Robillard and Marsden 1996). Although these investigations have indicated which structures or methods are most precise for various fish species, they have not considered the effects of important external sources of variation associated with estimating ages from multiple populations. Latitude has been suggested to influence the precision of ages estimated from scales (Kruse et al. 1993); however, these effects have not been directly assessed. Population size structure (e.g., stunting) may also influence the precision of estimated ages, given that annuli are more compressed in slowly growing populations. In addition, growth rates can decline greatly with the onset of maturation (Roff 1983; Jennings et al. 1997). The effects of sex and maturation schedules on age estimates often have been overlooked (but see Sharp and Bernard 1988). Because management activities frequently cover large geographic scales with multiple populations, a synthetic examination of the effects of biotic and abiotic factors on precision between readers and structures is necessary.

There is currently considerable debate regarding appropriate statistical methods for examining precision among readers and structures. Percent agreement is considered a simple, intuitive statistic that provides valuable insight into differences in precision (Hoenig et al. 1995). Use of this statistic has been discouraged, however, because it is a relative measure of precision and does not allow for comparisons among species. The range of ages among species may be so different that percent agreement among multiple species becomes meaningless (Beamish and Fournier 1981; Chang 1982). Alternatives to percent agreement, such as average percent error (APE; Beamish and Fournier 1981) and coefficient of variation (CV; Chang 1982), have been suggested to compensate for this deficiency because being age-dependent, they allow for comparisons among species and populations. Construction of age bias plots also has been sug-

* Corresponding author: hoxmeier@mail.inhs.uiuc.edu
Received March 10, 2000; accepted October 1, 2000

gested for detecting both linear and nonlinear biases (Campana et al. 1995).

Our goal was to expand the scope of previous studies and quantify the effects of potential sources of variation on the precision of age estimation. Specifically, we examined the effects of latitude, population size structure, sex, maturity level, and age on the precision of age estimation between readers and structures (scales and otoliths) of bluegills *Lepomis macrochirus.* We chose bluegills because use of both their scales and otoliths has been validated (Regier 1962; Schramm 1989), but no previous studies have examined precision of bony structures for this common sport fish species. Additionally, bluegills exhibit complex life history strategies that often result in differences among populations in characteristics such as growth rate, population size structure, longevity, and maturation schedule (Claussen 1991). This variability makes the bluegill particularly suited for examination of precision of age estimation in multiple populations. Additionally, we have compared statistical methods to determine which, if any, is most appropriate for identifying and quantifying differences in precision between both readers and structures.

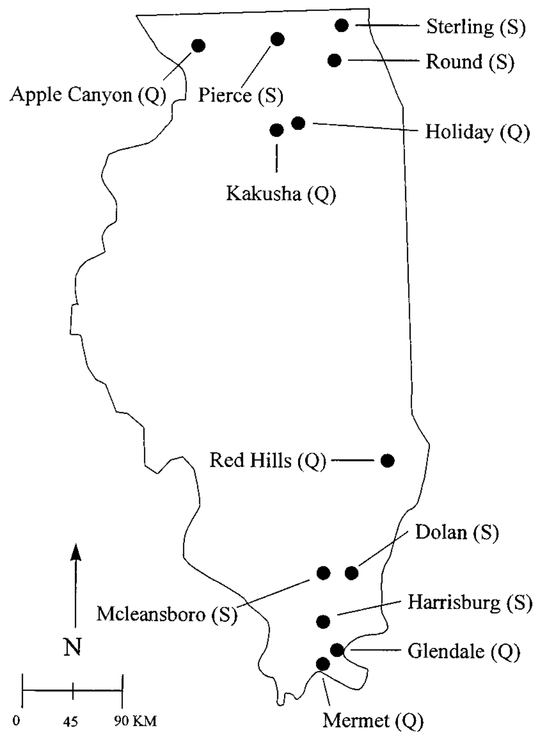

FIGURE 1.—Location of bluegill study lakes used to compare the effects of latitude and population size structure located in northern and southern Illinois. Quality populations (Q) were those in which mature, male bluegills greater than 180 mm total length (TL) were abundant, whereas in stunted populations (S) the majority of mature male bluegills were less than 160 mm TL.

Methods

Bluegills were collected from May through July in 12 Illinois reservoirs that ranged from 12 to 220 ha in size and from 2 to 9 m in average depth. Despite the size range, lake characteristics were fairly homogeneous; each is a shallow, eutrophic reservoir with large littoral areas dominated by pondweed *Potomogeton* spp. and fish communities dominated by centrarchids.

To examine the effects of latitude and population size structure on precision of age estimates, we divided the lakes into four groups: north stunted, north quality, south stunted, and south quality. The six northern lakes were above latitude 42°N and the six southern lakes were below latitude 39°N (Figure 1). The two closest lakes between the latitude groups were 340 km apart. Because of the large latitudinal gradient, northern and southern lakes were subject to different temperature regimes and growing seasons. Summer water temperatures (June–August) averaged 24.6 ± 0.6°C (±SE) in the northern lakes and 26.7 ± 0.3°C in the southern lakes. Additionally, all northern lakes typically had ice cover from late December through February, but ice was rarely present on the southern lakes.

To identify stunted and quality lakes, we sampled 60 lakes across Illinois for bluegill size structure. From these we chose six lakes with high-quality fish and six with the most stunted fish in each latitude group. Mature male bluegills averaged 178 ± 4.34 mm total length (TL) in high-quality populations and 147 ± 8.96 mm in stunted populations. Proportional stock density was also greater in the lakes with high-quality fish (48 ± 6.2) than in those with stunted fish (14 ± 3.9).

Bluegills were collected with an AC electrofishing boat; attempts were made to collect all sizes of mature and immature bluegills from each lake. The bluegills were returned to the laboratory and stored frozen for later processing, which consisted of measurement of length (mm, TL), weight (g), removal of gonads for sex and maturity determination, and removal of scales and sagittal otoliths. For each lake we divided the bluegill length histogram into quartiles to get a representative sample from each size group. We then estimated ages of

10 randomly selected bluegills from each length quartile (N = 40).

To determine bluegill ages, otoliths were submersed in ethanol and examined with a dissecting scope (12× magnification) in whole view on a black background with reflected light. Annuli were identified as opaque bands separated by dark, translucent bands (regions of fast growth; Schramm 1989). Because of the ease of annulus identification in whole view, otoliths were not sectioned. This is consistent with other investigations of centrarchids (e.g., Boxrucker 1986; Maceina and Betsill 1987). For scale analysis, 6–10 scales were removed from just below the lateral line underneath the pectoral fin (DeVries and Frie 1996). Scale impressions were made on acetate slides and examined with a microfiche reader. Only nonregenerated scales were examined during the study. Annuli were identified by crowding and crossing over of circuli present on both the right and left side of the focus (DeVries and Frie 1996).

Statistical analyses.—Ages were estimated independently by two readers equally experienced with scales and otoliths. Precision was estimated by percent agreement, coefficient of variation (as defined by Chang 1982), and age bias plots (Campana et al. 1995). Percent agreement was estimated as the proportion of each age on which both readers agreed. Three-way analysis of variance (ANOVA) was used to test for differences across bony structures, latitude, and population size structure for both percent agreement and CV. Multiple comparison tests were performed by using least-squares means in SAS (SAS Institute 1989). Effects of sex and maturity stage were also determined between structures with three-way ANOVA. Age bias plots were created for each reader using otoliths as the benchmark for comparison with scales. Age bias plots were also constructed to examine potential biases between readers. We used t-tests to determine if the intercept of the age bias plot differed from 0.0 and if the slope of the age bias plot differed from 1.0. Regression analysis and analysis of covariance (ANCOVA) was used to examine the effect of age on percent agreement with both scales and otoliths.

Results

Otoliths versus Scales

Otoliths were more precise than scales for estimating age of the bluegill, based on both percent agreement (three-way ANOVA; $F_{1,23}$ = 33.8; P < 0.01) and CV ($F_{1,23}$ = 23.9; P < 0.01). In addition,

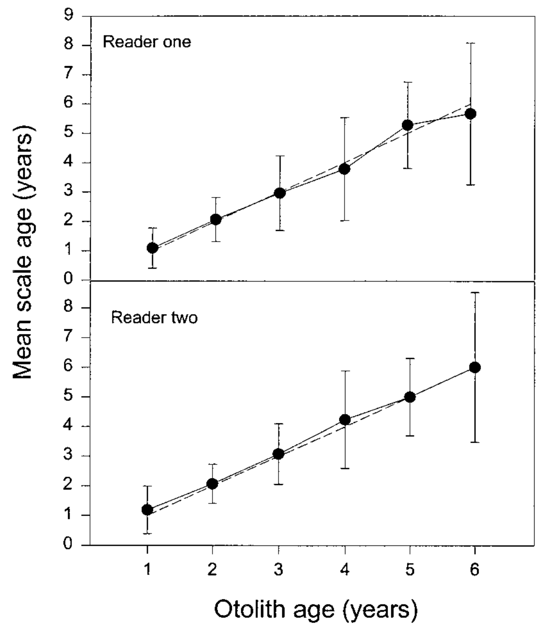

FIGURE 2.—Age bias plots for each reader using scales and otoliths for bluegills from 12 study lakes. The dashed line represents the 1:1 line, which was derived from otolith-estimated ages. Values (solid circles) are mean ages (±SE) determined from scales. Points above the 1:1 line indicate ages that were overestimated when based on scales, whereas points below the 1:1 line indicate ages that were underestimated.

ages estimated from scales and otoliths showed no significant bias for either reader one (t-test; df = 4; slope, t = 0.93; P = 0.41; intercept, t = 0.66; P = 0.55) or reader two (slope, t = 1.36; P = 0.24; intercept, t = 2.26; P = 0.09; Figure 2). Disagreement increased with age between scales and otoliths, as indicated by the larger standard errors for older ages, but there was no consistent pattern of over- or underestimating ages from scales (Figure 2). Similarly, bias was not significant between readers for scales (slope, t = 2.42; P = 0.07; intercept, t = 1.81; P = 0.14) or otoliths (slope, t = 1.37; P = 0.24; intercept, t = 1.16; P = 0.31; Figure 3).

Treatment Effects

Population size structure did not significantly affect the precision of ages estimated from either otoliths or scales (CV, three-way ANOVA; $F_{1,23}$ = 0.01; P = 0.94; percent agreement, $F_{1,23}$ = 0.01; P = 0.91; Table 1). Latitude, too, did not affect the percent agreement of ages estimated from otoliths, although it did affect the precision of ages estimated from scales (structure × location inter-

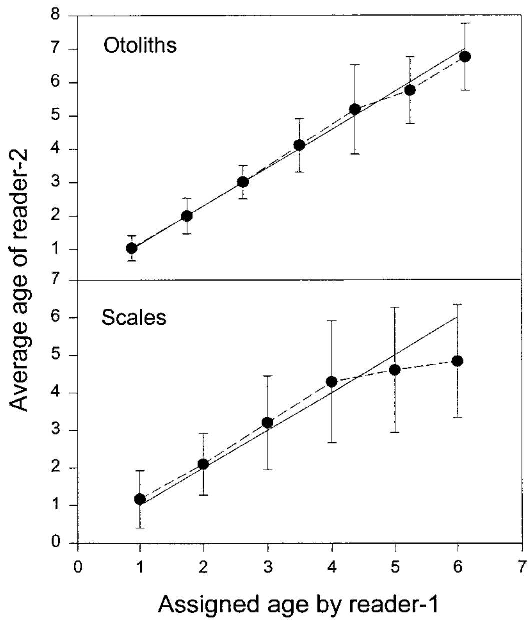

FIGURE 3.—Age bias plots between readers for otoliths and scales of bluegills from 12 study lakes in Illinois. Ages determined by reader one were plotted as the 1:1 line. Mean ages (±SE, solid circles) are from reader two; biases between readers are represented by deviations from the 1:1 line.

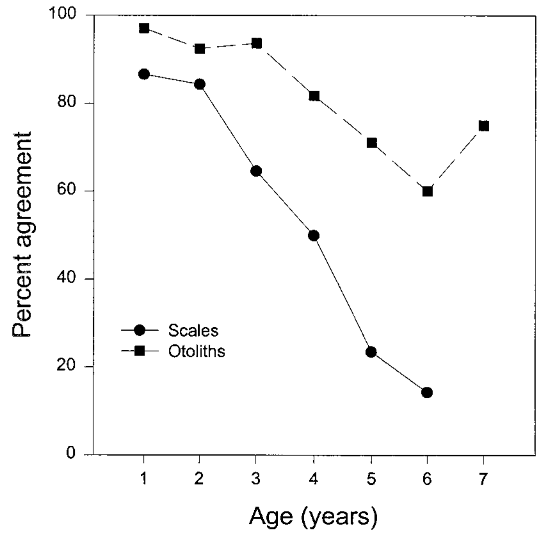

FIGURE 4.—Relation between percent agreement and age for bluegill scales and otoliths across 12 study lakes in Illinois. Percent agreement is the number of scales or otoliths that two independent readers agreed on divided by the total number of fish evaluated.

action; $F_{1,23} = 4.89$; $P = 0.04$). Percent agreement of ages estimated from scales of bluegills from southern lakes was significantly less than that for fish from northern lakes (least-squares means; $F = 7.73$; $P = 0.01$).

Age had a significant effect on precision of estimation for both otoliths ($r^2 = 0.75$; df = 5; $P = 0.01$) and scales ($r^2 = 0.97$; df = 4; $P < 0.01$). However, as age increased, the effect was more pronounced for scales than for otoliths (ANCOVA; slope, $F_{1,10} = 24.21$; $P < 0.01$; intercept, $F_{1,10} = 20.61$; $P < 0.01$; Figure 4). With otoliths, percent agreement was initially high, declining gradually after age 3 years. With scales, percent agreement declined after age 2. In addition, the mean age of disagreement was significantly greater than the

TABLE 1.—Mean coefficient of variation (CV = 100 × SE/mean; ±SE) and percent agreement (±SE) for scales and otoliths collected from May through July in 12 Illinois bluegill populations with different latitudes (north or south) and population size structures (stunted or quality). Statistics are reported for bluegills of different sex (male or female) and maturity (mature or immature). Significant differences within structures due to treatment effects are indicated by asterisks (three-way analysis of variance; * $P < 0.10$; ** $P < 0.05$).

	CV		Percent agreement	
Group	Otoliths	Scales	Otoliths	Scales
All lakes	2.62 (0.56)	7.14 (0.74)	90.5 (1.75)	73.0 (2.96)
Latitude				
North	2.82 (0.84)	5.84 (1.03)*	89.7 (2.66)	78.9 (3.52)**
South	2.40 (0.80)	8.41 (0.83)*	91.2 (2.50)	67.1 (3.51)**
Size structure				
Quality	2.06 (0.85)	7.73 (0.97)	92.7 (2.97)	71.1 (3.41)
Stunted	3.16 (0.71)	6.52 (1.15)	88.2 (1.64)	74.9 (5.06)
Sex				
Male	3.35 (0.78)	7.73 (1.05)	90.9 (2.17)	73.8 (3.49)
Female	2.32 (0.55)	8.61 (0.95)	89.3 (2.89)	66.5 (4.42)
Maturity				
Mature	2.55 (0.49)	9.49 (0.85)**	87.0 (2.40)*	58.2 (4.63)**
Immature	3.22 (0.94)	6.32 (1.19)**	94.1 (2.10)*	87.3 (1.70)**

TABLE 2.—Mean age (±SE) of agreement and disagreement between readers for bluegill scales, otoliths, and both structures combined. Bluegills were collected from May through July in 12 Illinois reservoirs of different latitudes and population size structures. Mean ages were tested for differences within each structure and combined with analysis of variance ($P < 0.05$).

Structure	Mean age		F	P
	Agree	Disagree		
Scales	2.12 (0.10)	3.32 (0.19)	31.11	<0.001
Otoliths	2.40 (0.12)	3.71 (0.38)	12.57	0.002
Combined	2.26 (0.08)	3.49 (0.20)	34.58	<0.001

mean age of agreement for scales, otoliths, and both structures combined ($F_{1,41} = 34.58$; $P < 0.01$; Table 2).

The percent agreement and CV of ages estimated from scales and otoliths were not significantly different between male and female bluegills (percent agreement, $F_{1,91} = 0.94$; $P = 0.34$; CV, $F_{1,91} = 0.90$; $P = 0.34$; Table 1). Maturity affected the precision of age estimates from scales more than those from otoliths (structure × maturity interaction; percent agreement, $F_{1,91} = 11.19$; $P < 0.01$; CV, $F_{1,91} = 4.26$; $P = 0.04$). For scales, ages estimated from mature fish had significantly higher CVs (least-squares means; $F = 8.60$; $P < 0.01$) and lower percent agreement ($F = 41.34$; $P < 0.01$) than did immature fish (Table 1). The effects of maturity were similar for both male and female bluegills (sex × maturation interaction; percent agreement, $F_{1,91} = 1.62$; $P = 0.21$; CV, $F_{1,91} = 2.12$; $P = 0.15$).

Discussion

Our results suggest that otoliths are more precise than scales for estimating bluegill age. This agrees with studies of other fishes, including yellow perch *Perca flavescens* (Robillard and Marsden 1996; Niewinski and Ferreri 1999), lake trout *Salvelinus namaycush* (Sharp and Bernard 1988), striped bass *Morone saxatilis* (Welch et al. 1993), and white crappie *Pomoxis annularis* (Boxrucker 1986; Hammers and Miranda 1991). In contrast, scales have been recommended over otoliths for black crappie *P. nigromaculatus* and Pacific herring *Clupea pallasi* because the results are equally precise, are easier to obtain, and do not require killing the fish (Chilton and Stocker 1987; Kruse et al. 1993). Our results suggest the high precision of age estimates based on scales in those studies may result from fish being collected in northern latitudes. Even though otoliths provided a more precise estimate of age than did scales in our study, the precision of the use of each structure was affected by population-specific variation.

Because bluegills exhibit sex-specific differences in growth and maturation schedules (Ehlinger 1997), we hypothesized sex could affect the precision of the estimated ages. However, CV and percent agreement did not differ significantly between male and female bluegills. Conversely, scales of mature bluegills were significantly less precise for estimating ages than were those of immature bluegills. Similarly, Sharp and Bernard (1988) found that ages of only immature lake trout could be precisely estimated with scales. The poor precision of estimates based on scales from mature bluegills probably results from the decrease in growth after maturation (Roff 1983), which makes determining annuli difficult.

Decreased precision of estimating ages from scales at lower latitudes may be the result of differences in annulus formation between northern and southern populations. High precision of age estimates based on scales collected in other northern regions (e.g., South Dakota) has been attributed to distinct winters, which produce easily distinguishable annuli (Kruse et al. 1993). Bluegills in our southern populations experienced longer, more variable growing seasons, thereby causing false and nondescript annuli, which reduced the ability to estimate accurately and precisely the ages of these fish from their scales. In contrast, latitude did not affect the precision of age estimates based on otoliths.

Two reasons may explain why we could not find a relationship between population size structure and precision with either scales or otoliths. First, there may have been a relationship that was undetectable. This seems unlikely, however, given the large sample sizes we obtained from the study lakes. Second, and we believe more important, precision is dictated by the ability to detect annuli on each structure. As long as annuli are distinct and no false annuli are present to confound age estimation, precision should be high regardless of the proximity of annuli to each other (influenced by growth rate) or the size of the structures themselves (influenced by population size structure).

Percent agreement has been shown to decrease with age for other species such as yellow perch (Niewinski and Ferreri 1999) and haddock *Melanogrammus aeglefinus* (Kohler and Clark 1958). Our results agree with these investigations and demonstrate that increasing age in bluegills decreases precision of age estimates for both scales and otoliths, probably because of the physiology of annuli formation on bony structures. The decrease in precision is particularly pronounced with scales. As fish age and mature, decreasing growth rates result in annuli coming closer together and compressing near the distal margin of the structure. Detecting annuli on older fish becomes more difficult because distinguishing the margin of the structure from the annuli is harder. In addition, otoliths become thicker as fish age. As a result, annuli become more difficult to distinguish in the increasingly opaque structure. Previous researchers have suggested sectioning bluegill otoliths after age 5 to distinguish crowding annuli (Hales and Belk 1992). Although we do not necessarily concur with a fixed age for sectioning otoliths (the actual age for sectioning is likely to be population dependent), we agree that at some point most older bluegill otoliths will need to be sectioned for accurate and precise estimates of age.

Statistical Analyses

We found no single measure that was best suited for estimating age precisely based on bony structures of bluegills. However, our analyses do demonstrate the value of using multiple statistical methods to identify sources of variation between readers and structures. By far the easiest statistic to calculate and interpret is percent agreement; this measure, however, is less conservative than CV. Despite debate in the literature regarding use of percent agreement, we agree with Hoenig et al. (1995) that it is a valuable, informative starting point. Percent agreement identifies important areas of disagreement between readers for each structure. The use of additional statistics is dictated by the goals of the investigation. To examine the effect of age on precision of estimates and to detect reader bias, age bias plots are useful (Campana et al. 1995). Assuming no bias exists between or among readers, regression analysis and ANOVA can provide test statistics to identify ages at which precision declines to less than acceptable. If multiple populations are to be compared, especially of different species, the age-dependent CV statistic is useful. In combination, these statistics can be used to identify sources of variation among readers, structures, and populations.

Management Recommendations

We recommend use of otoliths for precise and accurate age estimation for bluegills from most populations and of most ages. However, use of otoliths requires killing the fish, and losses of adult fish could be a problem for some bluegill populations. For younger fish (i.e., up to about age 3, depending on the maturation schedule of the population), scales can provide age estimates with satisfactory precision. In addition, ages of fish from populations subject to relatively short, distinct growing seasons can be more precisely estimated with scales than those with longer, indistinct growing seasons (i.e., lower latitudes). Because precision depends on these population-specific parameters, we suggest a preliminary examination of fish throughout the size range of the population to determine which ages can be precisely estimated with scales and those that will require otoliths. For comparisons of multiple populations, our results suggest that latitude, as well as age range and maturity of individuals within the populations, will have a greater effect on precision than will population size structure.

Acknowledgments

We thank R. Anderson, J. Boase, M. Desjardins, and K. Eisenbarth for help with field collection and T. Edison, M. Engel, H. Leonard, and T. Mason for additional assistance in the laboratory. This manuscript benefited from comments by P. Braaten, J. Davis, J. Hart, K. Ostrand, and two anonymous reviewers. This study was funded in part by the Illinois Department of Natural Resources (IDNR) through Federal Aid in Sport Fish Restoration, project F-128-R. S. Stuewe and L. Dunham coordinated activities with the IDNR.

References

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.

Beamish, R. J., and G. A. McFarlane. 1987. Current trends in age determination methodology. Pages 15–42 *in* R. C. Summerfelt and G. E. Hall, editors. The age and growth of fish. The Iowa State University Press, Ames.

Boxrucker, J. 1986. A comparison of the otolith and scale methods for aging white crappies in Oklahoma. North American Journal of Fisheries Management 6:122–125.

Campana, S. E., M. C. Annand, and J. L. 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.

Chilton, D. E., and M. Stocker. 1987. A comparison of otolith and scale methods for aging Pacific herring. North American Journal of Fisheries Management 7:202–206.

Claussen, J. C. 1991. Annual variation in the reproductive activity of a bluegill population: effect of clutch size and temperature. Master's thesis. University of Toronto, Toronto.

DeVries, D. R., and R. V. Frie. 1996. Determination of age and growth. Pages 483–512 *in* B. R. Murphy and D. W. Willis, editors. Fisheries techniques, 2nd edition. American Fisheries Society, Bethesda, Maryland.

Ehlinger, T. J. 1997. Male reproductive competition and sex-specific growth patterns in bluegill. North American Journal of Fisheries Management 17: 508–515.

Hales, L. S., Jr., and M. C. Belk. 1992. Validation of otolith annuli of bluegills in a southeastern thermal reservoir. Transactions of the American Fisheries Society 121:823–830.

Hammers, B. E., and L. E. Miranda. 1991. Comparison of methods for estimating age, growth, and related population characteristics of white crappies. North American Journal of Fisheries Management 11: 492–498.

Hoenig, J. M., M. J. Morgan, and C. A. Brown. 1995. Analysing differences between two age determination methods by tests of symmetry. Canadian Journal of Fisheries and Aquatic Sciences 52:364–368.

Jennings, M. J., J. E. Claussen, and D. P. Philipp. 1997. Effect of population size structure on reproductive investment of male bluegill. North American Journal of Fisheries Management 17:516–524.

Kohler, A. C., and J. R. Clark. 1958. Haddock scale–otolith comparisons. Journal of the Fisheries Research Board of Canada 15:1239–1246.

Kruse, C. G., C. S. Guy, and D. W. Willis. 1993. Comparison of otolith and scale age characteristics for black crappies collected from South Dakota waters. North American Journal of Fisheries Management 13:856–858.

Maceina, M. J., and R. K. Betsill. 1987. Verification and use of whole otoliths to age white crappie. Pages 267–278 *in* R. C. Summerfelt and G. E. Hall, editors. The age and growth of fish. The Iowa State University Press, Ames,.

Niewinski, B. C., and C. P. Ferreri. 1999. A comparison of three structures for estimating the age of yellow perch. North American Journal of Fisheries Management 19:872–877.

Regier, H. A. 1962. Validation of the scale method for estimating age and growth of bluegills. Transactions of the American Fisheries Society 91:362–374.

Robillard, S. R., and J. E. Marsden. 1996. Comparison of otolith and scale ages for yellow perch from Lake Michigan. Journal of Great Lakes Research 22:429–435.

Roff, D. A. 1983. An allocation model of growth and reproduction in fish. Canadian Journal of Fisheries and Aquatic Sciences 40:1395–1404.

SAS Institute. 1989. SAS user's guide: statistics. SAS Institute, Cary, North Carolina.

Schramm, H. L., Jr. 1989. Formation of annuli in otoliths of bluegills. Transactions of the American Fisheries Society 118:546–555.

Sharp, D., and D. R. Bernard. 1988. Precision of estimated ages of lake trout from five calcified structures. North American Journal of Fisheries Management 8:367–372.

Welch, T. J., M. J. Van Den Avyle, R. K. Betsill, and E. M. Driebe. 1993. Precision and relative accuracy of striped bass age estimates from otoliths, scales, and anal fin rays and spines. North American Journal of Fisheries Management 13:616–620.