**CSIRO** PUBLISHING

www.publish.csiro.au/journals/mfr *Marine and Freshwater Research*, 2005, **56**, 773–782

# Quality issues in the use of otoliths for fish age estimation

*A. K. Morison*[A,E], *J. Burnett*[B], *W. J. McCurdy*[C] *and E. Moksness*[D]

[A]Marine and Freshwater Systems, Department of Primary Industries, PO Box 114, Queenscliff, Vic. 3225, Australia.
[B]National Marine Fisheries Service, Northeast Fisheries Science Center, 166 Water Street, Woods Hole, MA 02543, USA.
[C]Aquatic Systems Branch, Agriculture Food and Environmental Science Division, Department of Agriculture and Rural Development for Northern Ireland, Newforge Lane, Belfast BT9 SPX, Northern Ireland, UK.
[D]Institute of Marine Research Arendal, Flødevigen Marine Research Station, N-4817 HIS, Norway.
[E]Corresponding author. Email: sandy.morison@dpi.vic.gov.au

*Abstract.* Quality issues in fish age estimation, which historically have focused mainly on inadequacies in the validation process, are increasingly directed at ways to measure and control the errors or inconsistencies in the application of established and validated methods. The process of age estimation, as undertaken by human operators, involves a complex mix of pattern recognition and interpretation based on knowledge and experience. It is best characterised as a skill rather than an art. Such an approach promotes the use of well-recognised techniques designed to maintain and enhance skills that also assist in maintaining standards. The results of a questionnaire completed by representatives of over 50 ageing laboratories worldwide were used to assess current quality assurance and quality control practices. Results indicate a great diversity in attention to, and no clear consensus on desirable standards for, quality issues, including staff training, use of reference sets, reading protocols, and post-reading analyses. This is considered more likely to reflect variation in awareness of the importance of quality issues than variation in the need for quality assurance and quality control measures. Greater attention to a range of quality control processes is urged, particularly the more regular use of reference sets.

*Extra keywords*: bias, error, growth, mortality, precision, recruitment, stock assessment.

## Introduction

The use of otoliths to provide age estimates is their most common application. Their use for this purpose is subject to error from two major sources: (*1*) error inherent in the structure itself, and (*2*) error in the process of interpreting its incremental structure (Campana 2001). The first type of error can be estimated but not controlled; the second type of error can be both estimated and controlled. It is the measurement and control of this second type of error that is the subject of this paper.

Attention has regularly been drawn to the extent and impact of errors in age estimation (Beamish and McFarlane 1983*a*, 1995; Lai and Gunderson 1987; Tyler *et al*. 1989; Bradford 1991; Richards *et al*. 1992; Reeves 2003). Such errors will affect all of the age-based population parameters such as yield, growth, mortality and recruitment. And as fisheries assessment and management systems frequently depend on the correct estimation of such parameters, errors in the age estimates will be propagated through the assessment and management process. The result may be overly conservative management measures, but it may also be excessive fishing mortality and over-fished stocks. These errors are at least partly the result of a failure to rigorously apply appropriate quality assurance and quality control processes (QA/QC), and the costs have been substantial for some fisheries.

Errors in the estimation process can affect either accuracy (closeness to the true value) or precision (closeness of repeated estimates) (Kalish *et al*. 1995). The prevalence and impact of inaccurate age determinations on the accuracy of population dynamics studies cannot be overstated (Campana 2001), and the issue of accuracy, the purpose of age validation, has received deserved attention since the important paper by Beamish and McFarlane (1983*a*). The prevalence and impact of 'imprecise' age determinations has, however, only recently begun to be evaluated. Particularly, in laboratories providing ongoing estimates of age for fisheries management, there is a need to estimate and control the errors in the interpretation of incremental structure, including precision (Morison *et al*. 1998; Campana 2001). Even valid age estimation techniques may not always be correctly applied (Buckmeier 2002).

 10.1071/MF04217 1323-1650/05/050773

The massive annual effort applied to the estimation of fish age (well over 1 million fish worldwide in 1999; Campana and Thorrold 2001) highlights both the value attached to the information obtained, and the potential for poor QA/QC practices to lead to a huge waste of resources. But although there is a range of publications on the practical aspects of fish age estimation (Bedford 1983; Cowan *et al*. 1995), on the application of validation procedures (Beamish and McFarlane 1983*a*; Campana 2001), on measures of precision (Beamish and Fournier 1981; Chang 1982; Kimura and Lyons 1991), on estimation of required sample sizes (Gröger 2003), and on the analysis of age data and treatment of ageing errors (Kimura 1977; Beamish and McFarlane 1983*b*; Richards *et al*. 1992; Gröger 1999; Schaalje *et al*. 2002), guidance and advice on how to establish and maintain quality in the outputs from ageing facilities is more limited (but see Campana *et al*. 1995; Morison *et al*. 1998; Campana 2001). This paper aims to provide an overview of the current theory and practise of the application of QA/QC to fish age estimation.

## Materials and methods

The theory aspects are based on publications concerning fisheries applications and concerning similar processes in other analytical laboratories. An appreciation of the unique aspects of the fish age estimation process, however, is important in assessing the applicability of processes from other contexts. The overview of current practice of quality assurance and quality control in fish age estimation laboratories around the world is based mainly on answers to a questionnaire distributed through informal networks of professionals. The experiences of the authors and examples from the laboratories in which they have worked are also used where appropriate.

Although there is no widely accepted definition of the terms, those outlined in NATA (1995) are used here. Quality assurance (QA) is considered to cover those procedural matters that are generally applied laboratory-wide and are not sample specific (staff training, procedure manuals, calibration, instrument maintenance, audit and review). A test of a reader's performance on a reference set of otoliths is a relevant example. Quality control (QC) is used here to include the methods and processes that are used in checking of the results obtained for individual samples. The making of repeated age estimates to estimate precision is a relevant example.

The extensive and diverse literature on QA/QC was sampled by using the Internet and library searches for relevant publications. Those selected included some giving general overviews of QA/QC issues, some pertinent to QA/QC in analytical laboratories in general, and a small number dealing with the process of fish age estimation in particular.

### *Questionnaire*

The questionnaire was structured around questions covering current and, in some instances, preferred practices for five topics: background statistics on the laboratory, staff training, use of reference sets, reading protocols, and post-reading analyses. This was distributed by email using the networks of contacts of the authors. The main aim of the questionnaire was to gain information about the current practices of age estimation laboratories; but other aims were to provide laboratories with a benchmark of current practices against which they could assess their own, and to provoke some thought about QA/QC issues in all laboratories, whether or not the questionnaire was completed.

## Results

### *Questionnaire responses*

Responses were obtained from 53 laboratories from 23 countries. Responding laboratories employed between fewer than one full time staff member and over 30 staff, but two-thirds had five or fewer staff employed full time. The number of age estimates generated by these laboratories annually was similarly variable, but 45% of laboratories each generated more than 10 000 age estimates annually. Marine species were aged by over 80% of laboratories, but over 50% also examined at least some freshwater species. There was a similar mix between daily and annual age estimates being provided, although daily age estimates were provided for fewer species than annual age estimates. The total numbers of age estimates undertaken annually by all responding laboratories was between 630 000 and 1.1 million (respondents were asked to nominate a range). This total indicates that responding laboratories represent a significant proportion of the total worldwide annual effort on fish age estimation, which was estimated at over 1 million fish in 1999 (Campana and Thorrold 2001), and suggests that the responses provide a representative picture of currently employed practices.

One of the main findings is that there is a large amount of variation in the manner and extent of use of QA/QC practices among laboratories. Complete results of the questionnaire are available on request from the corresponding author, but some of the more significant conclusions from the questionnaire are as follows.

- Practical aspects of sample preparation have been well documented, but there has been less (and we would suggest insufficient) attention given to other aspects of the issue. Whereas 43% of respondents had technical protocols documented for every species, and 30% had the same level of documentation for their reading protocols, only 25% had quality control protocols documented for every species.
- Current practices would suggest that laboratories differ greatly in the amount of training that they consider necessary for their staff. The level of experience of the reader, and the complexity of the otolith macrostructure in the species being examined were expected to be the most important determinants of training needs. Inter-laboratory variation overshadowed these factors (Fig. 1) with, for example, some laboratories proposing that experienced staff members required in excess of 1000 samples for species with clear increments, and others that new staff members could be adequately trained with fewer than 100 samples for difficult species.

We examined whether larger laboratories required more training of their staff. Two situations were examined representing special cases. Case 1 was for experienced readers dealing with species with clear increments (the situation considered to require the least training). Case 2 was for

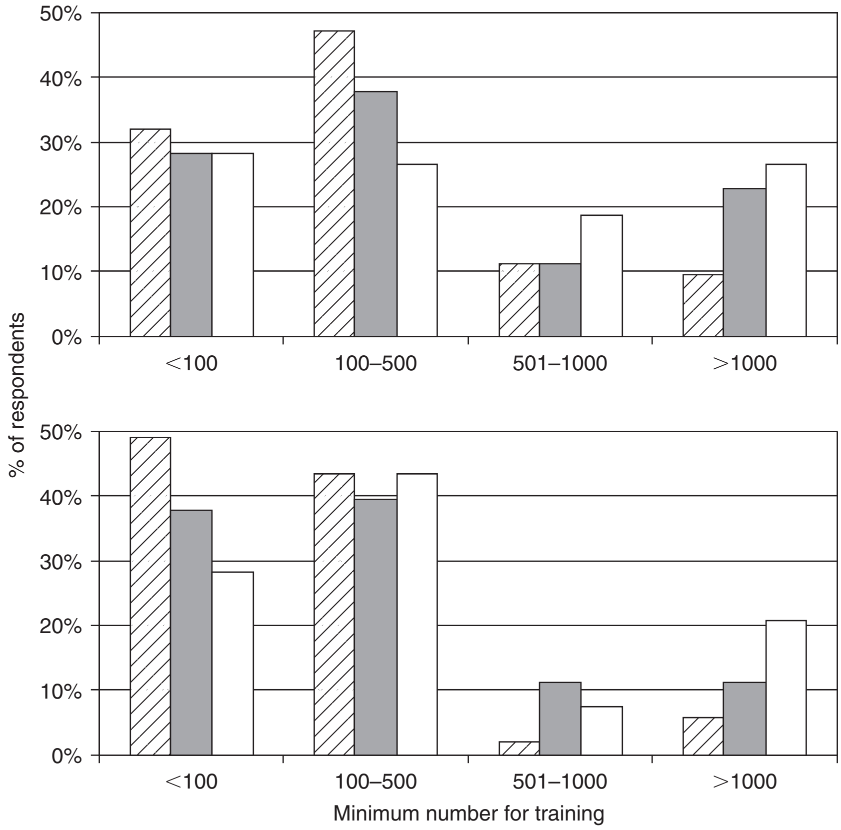


**Fig. 1.** Minimum number of otolith-responding laboratories considered necessary for a new reader (*a*) and an experienced reader (*b*) to examine in training to achieve consistency in interpretation, for species with clear increments (diagonal stripe), moderately difficult species (shaded bars), and difficult species (open bars), $n = 53$.

new readers dealing with species with difficult increments (the most difficult situation, requiring the highest level of training). A $\chi^2$-test was used to test whether the numbers required for training was independent of laboratory size. Laboratory size and numbers required for training were categorised with the groupings chosen to keep expected frequencies in each cell above five (Zar 1984). Laboratory size was categorised for both cases as either as small (fewer than 10 000 age estimates undertaken annually) or large (10 000 or more age estimates undertaken annually). Groupings for numbers required for training were less than 100 and 100 or more for Case 1, and less than 500 and 500 or more for Case 2. There was a significant relationship between laboratory size and the minimum number of samples considered necessary for training of staff for both Case 1 ($\chi^2 = 4.8$, d.f. $= 1$, $P < 0.05$) and Case 2 ($\chi^2 = 11.3$, d.f. $= 1$, $P < 0.01$).

- Reference collections were not in use in one-third of laboratories. Large and small laboratories were equally likely to use reference sets. Where they were used, they mostly relied on previously aged material. This is most likely a reflection of the fact that known-age fish are not widely available.
- Reference collections were used more for training than QA/QC. Even among those laboratories that used reference collections for QC, a low proportion had their use prescribed as a standard procedure.
- Birth dates were not routinely set for a significant proportion of species being aged by laboratories. Birth dates are used in combination with the date of capture and an assessment of the state of completion of the outermost increment, to convert an increment count to an age estimate. Without an established birth date, there will almost always be some level of inconsistency in the assignment of ages to fish of the same year-class.
- Auxiliary data (e.g. fish size, date of capture etc.) were often available to readers, much of it not explicitly provided, but the types of data, and the manner of their use, was quite varied. It was mostly used during the reading process, where it is able to influence a reader's interpretation.
- Most laboratories use a similar range of measures to detect errors, including the comparison of estimates for reference sets, comparison of growth curves/parameters, inspection

of scatter plots of age *v.* otolith weight or fish length, examination of age composition, examination of age-length keys, examination of mean lengths-at-age, as well as other methods. Their frequency of use, however, is very variable and is frequently *ad hoc*. Without the regular application of error detection measures, the extent of errors in the age estimation process will not be known. Statistical correction of ageing errors was rarely undertaken, but measures of precision are more commonly incorporated into assessments.
- Percentage of agreement was still the most commonly used measure of precision (49% of respondents), despite its limitations and previous criticisms. The coefficient of variation (CV; 36% of respondents) and the index of average per cent error (30%) were also commonly used. There was little consensus on what constitutes an acceptable level of error for any measure.
- Re-reading of samples was practised by 74% of respondents, but only 30% of respondents undertook it always or regularly, and 25% did not undertake any re-reading. The numbers re-examined ranged from the very few to 100%. Secondary readers were also commonly used.

### *Examples of failures and successes of quality control*

One example of unexpected errors and one of the benefits of quality control are instructive. The example of errors concerns the effect of working with size-sorted samples on age estimates of eastern gemfish (*Rexea solandri*) (Morison, unpublished data). Otoliths of this species are usually examined whole with reflected light while immersed in water. Otolith samples had been sorted by fish length before age estimates were made, and this was known to the reader; however, the length of each individual fish was not known when examining the otoliths. Calibration checks of interpretation against previously aged material before examination of the sample indicated consistency in interpretation. Nevertheless, length–frequency distributions of the younger age classes exhibited an abnormally wide spread of size-at-age and multiple modes for both sexes (Fig. 2), indicating unacceptable ageing errors. Knowledge that the sample was sorted in order of fish length had apparently increased the likelihood that the reader would assign ages that were the same as the previously aged fish. Ambiguous increments were more likely to be interpreted to give the same ages, causing the observed errors. Re-ageing of the sample after the batch had been randomly mixed produced single modes and narrower distributions that were more consistent with those previously observed (Fig. 3). The reader was also less experienced than others in the laboratory, suggesting that inexperience may contribute to a greater susceptibility to such sources of error.

The example of the benefit of the implementation of QA/QC procedures is in the interpretation of ambiguous data. A time-series of age estimates for redfish, *Centroberyx affinis*, showed inter-annual variation in age composition that could not be readily explained by variation in mortality, recruitment or known stock structuring (Morison and Rowling 2001). Of the several hypotheses that potentially explained the observed data, the one that could be immediately rejected was one that attributed the variation in age composition to inconsistency in the age estimates; repeat readings of the sample across years indicated a clear consistency in the interpretation.

## Discussion

The process of fish age estimation is a fundamentally different process from those applied by most other types of analytical laboratories. The assigning of ages to otoliths or other structures is clearly not simply a counting of features. There is both a reading stage and an interpretation stage (Sych 1974), and the interpretation process includes a component that could be best described as pattern recognition. This strong cognitive component (see Troadec and Benzinou 2002) complicates other important aspects of the process including the description of how ages are arrived at, the training of new staff, and the comparisons of ages arrived at by different readers. The recognition of this subjective component is part of the reason why age estimation is commonly referred to as ‘more art than science’. This description has possibly contributed to an apparent reluctance to apply QA/QC principles to the process.

It is more useful and more accurate, however, to view the process of age estimation from calcified structures as a skill rather than an art; it is not a creative process, but a diagnostic and analytical one. It shares attributes of other skills:

- The elements of the process can be taught, learned, and practiced.
- The level of skill increases with practise and experience, and will decline with time without practise.
- Some practitioners are naturally far more skilled than others.
- All practitioners vary in the skill they display at different times.

There are several important consequences for quality control from this: even those skilled and experienced in fish age estimation will make errors at some time, constant use is important to maintain levels of skill, and skilled practitioners are an uncommon and valuable resource. The experience of the age reader has been shown to be one of the significant factors affecting the accuracy of the age estimates of fish larvae (Campana and Moksness 1991), demonstrating its importance to the age estimation process.

The recognition of this subjective component to the age estimation process has led to a range of attempts to automate the process, or at least to create an objective one. These include: statistical approaches, the substitution of proxies for age, and the use of increasingly powerful computing

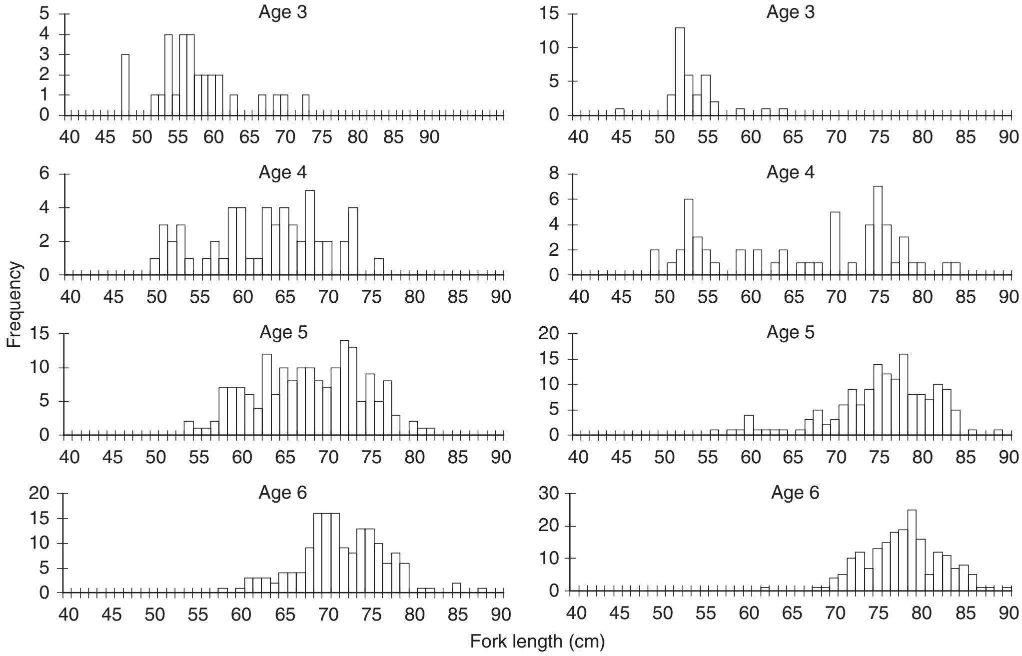


**Fig. 2.** Length–frequency distributions for male (*a*) and female (*b*) eastern gemfish (*Rexea solandri*) by age class, when age estimates were made from samples sorted in length order.

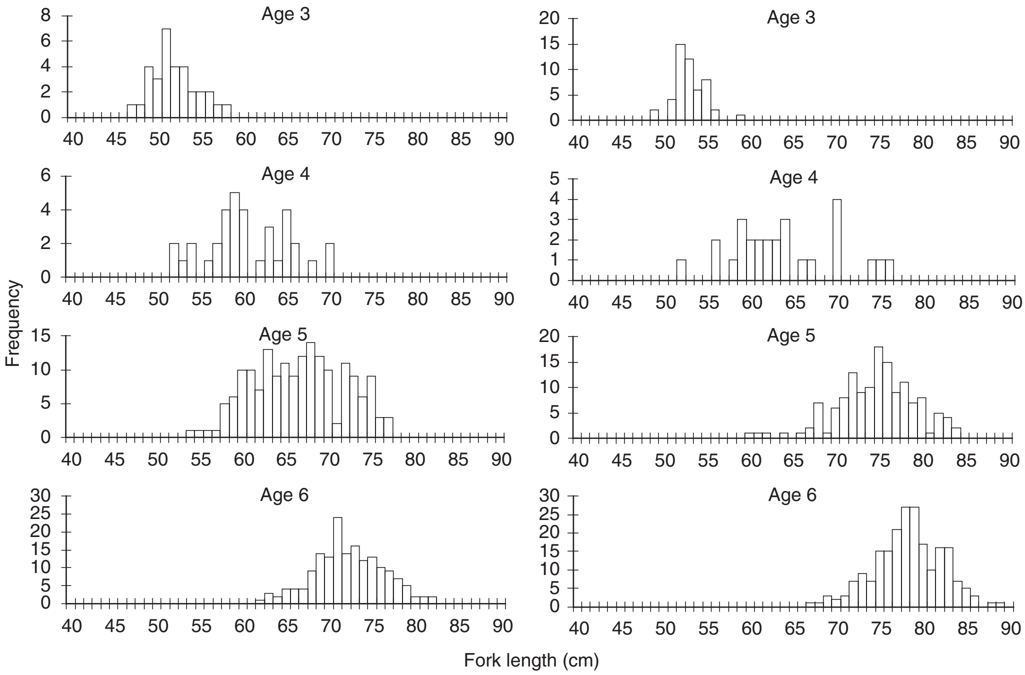


**Fig. 3.** Length–frequency distributions for male (*a*) and female (*b*) eastern gemfish (*Rexea solandri*) by age class, when age estimates were made from samples in random order.

capabilities and increasingly sophisticated image analysis software (Boehlert 1985; McGowan *et al*. 1987; Small and Hirschhorn 1987; Szedlmayer *et al*. 1991; Troadec 1991; Fletcher 1995; Macy 1995; Welleman and Storbeck 1995; Worthington *et al*. 1995; Cailliet *et al*. 1996; Lagardère and Troadec 1997; Morales-Nin *et al*. 1998; Morison *et al*. 1998; Robertson and Morison 1999, 2003; Pilling *et al*. 2003). It is a recognition of the complexity of the process that no age estimation laboratories have been able to replace their human readers.

The questionnaire showed that many laboratories are not using reference sets as a component of routine quality assurance. Whereas two-thirds of laboratories use reference sets, these are mostly used for training, and not as a regular part of a standard procedure. We believe that the development and regular use of reference collections is probably the most important single measure that laboratories should implement as part of a QA/QC system. They provide a standard, allowing assessments of the performance of readers through repeated age estimates on the same samples. As Campana (2001) points out, the reference collection approach has some strong advantages over other approaches, such as re-ageing of the previous year's samples. These include an ability to detect drift that may otherwise go undetected and, where the reference collection is based on accurately aged samples, the monitoring will assess accuracy as well as precision. The scarcity of known age material means that reference collections must frequently be based on samples for which the ages have been established by consensus or expert readers. Such reference collections allow consistency to be assessed but not bias.

The QC analyses of one kind or another should constitute at least 10% and possibly 20% of all analyses carried out (Campana 2001). Similar frequencies are suggested for other types of analytical laboratories (NATA 1995). In age estimation, where errors are potentially common, QC checks should be run often, and only reduced in frequency when experience shows it is justifiable. The regular undertaking of QC checks should also allow the progressive establishment of benchmarks for species and laboratories against which comparisons can be made between readers over time or between laboratories. If the use of a reference collection shows up regular errors, it would suggest that more attention is needed in the training (for inexperienced staff) or re-orientation (for more experienced staff) steps that often proceed the examination of a new batch of samples. As the detection of such errors will usually require the re-ageing of all structures from the point of the last quality control check, the use of more frequent checks would be prudent until the results are back within more acceptable error levels.

Empirical tests would also be useful in other aspects of the development and use of reference sets such as the minimum size for a reference set needed to avoid a significant proportion of the material becoming recognised by readers. This will probably depend on the frequency of use and the range of ages, at least. The frequency with which new material should be added to avoid such learning and to keep the reference set representative also remains to be established. Furthermore, tests of performance on reference samples when known to be part of a reference set, compared with when the same samples were read among other material, would help guide the best way to introduce reference material to readers. Reference material may not be able to be introduced without a reader's knowledge where, for example, many otolith sections are mounted on slides that have unique identifiers. Temporary re-labelling of slides may overcome this problem but an additional cost of the coding and de-coding process will be incurred.

The questionnaire also identified a common practice of age estimates being generated without the use of a standard birth date for each species. A protocol for interpreting the outermost increment is necessary for consistent age estimation. There are good examples in the literature of such protocols (Francis *et al*. 1992; Fowler and Short 1998; Morales-Nin and Panfili 2002) and they should be applied more often. The absence of an assigned birth date may not be important in only a few situations. One is where samples are all collected at a time of year well away from the time of increment formation, so the interpretation of the outermost increment would be consistent. Another is where ages are being assigned to very old animals (such as the orange roughy, *Hoplosteuthus antarctica*), so that any inconsistency in interpretation of the outermost increment is likely to be less than the variation inherent in the counting of the large number of other increments.

The analysis of results of the questionnaire indicate that the staff of laboratories that undertake more age estimates consider that more samples are needed for training both new and experienced staff than do staff of smaller laboratories. If the larger laboratories are considered those with the better perspective on this issue, and not merely those with higher training standards, then this finding is instructive and suggests that many laboratories underestimate the level of training that should be required of staff.

The best use of auxiliary data during age estimation is also an area requiring further investigation. It is clear that a large variety of such data are frequently available during age estimation, much of it implicit rather than explicitly provided. Such data have the potential to both reduce error and introduce bias (Morison *et al*. 1998). The relative strength of these opposing effects may not be immediately obvious in many cases. It would be valuable to explore the effects of providing or removing some data types on the accuracy and precision of the age estimates obtained to ensure that their use improves the quality of the results and does not degrade it. For example, knowledge of the year of recapture and of the expected strong year classes improved the accuracy of age estimates for snapper (*Pagrus auratus*) (Francis *et al*. 1992) and walleye pollock (*Theragra chalcogramma*; Kimura *et al*. 1992) respectively.

The questionnaire showed that percentage of agreement remains the most commonly used statistical tool for comparing repeated age estimates. The age-bias plot, in conjunction with a CV estimate (Campana *et al*. 1995) or the functionally similar index of an average per cent error (Beamish and Fournier 1981), however, are far superior measures to detect bias and measure precision. Quality assurance will be improved if these alternative measures are more commonly used.

The rarity with which any form of statistical correction of ageing errors is used in laboratories is noteworthy. The questionnaire did not ask why these methods were not used, but our experience suggests that a combination of an ignorance of their existence, a lack of readily available and user-friendly software to implement the methods, and an absence of any demand for them from users of the data, have all been impediments.

### *QA/QC for age estimation laboratories*

The example of a failure of quality control for eastern gemfish demonstrates how errors can arise as a result of a reader's (flawed) expectations. Many factors may influence the judgement call of a reader when they are assessing alternative interpretations, and significant errors can arise from quite unexpected sources. Human frailty and subjectivity will be sources of error until a fully objective method of age estimation is perfected. This is not likely in the short term. Systems are needed that will detect such errors and quantify their extent, regardless of how they arise.

The decision as to what constitutes an acceptable error is itself not necessarily a straightforward one. The users of the data should be consulted in this process to help identify the boundaries of acceptable performance. This is almost certain to vary with the species and with the intended use of the data. It is also likely to depend on other sources of error that may be contributing to the analyses. For example, improving the precision of age estimates that are being incorporated into some form of population model may be unnecessary if the model is known to be far more sensitive to errors in other contributing datasets. Sensitivity tests for stock assessment models can explore the impact of ageing error on model outputs (e.g. Punt *et al*. 2001). Many fisheries modelling programmes now also incorporate ageing error matrices as one of the data inputs, and so this explicitly accounts for the errors without having to correct them.

Well-designed databases can contribute significantly to the implementation of effective QA/QC systems in an age estimation laboratory. Several levels of data should be recorded, including details specific to the reader (such as results from training or recalibration), details of the batch being processed (such as species, area and date of collection, collection method etc.), details of the preparation and viewing method used, and the date on which readings were made. In addition, consideration should be given to recording increment radii, the presence of false increments, and the edge type, even if the adjustment from increment counts to age estimates is made at the time of reading. Ideally, any subsequent changes made to original data should be traceable. A well-designed system should allow for the effective recording and management of multiple readings, both for reference sets and also for the re-reading for estimates of precision. The more detail that is recorded, the more likely it is that any future questions can be answered, potentially saving time in terms of re-reading and helping identify the cause and cure for any systematic errors.

The storage of images of preparations is now widely practised in fish ageing facilities and has many potential QA/QC benefits. Readily available software also allows for the storage of layers of related information (such as the position of counted increments) that is of relevance for QA/QC. Databases for the storage of images, however, should also include details on a suite of additional parameters (e.g. magnification, image format, colour depth), not usually recorded in databases of age estimates (Troadec 1998, 2000).

### *General issues in the implementation of QA/QC processes*

The rigorous application of robust QA/QC protocols will not only give confidence to readers in the quality of their work, but should also provide proof to any independent assessment. The important principle is that some procedures should be set in place, used, and the results checked as to whether they have delivered the results being sought. If experience shows that the QC checks have failed to detect a problem, the reason for the failure should be investigated and corrective action taken. This may include increasing the frequency of QC checks. Procedures should be reviewed and revised as necessary and the process continued. The establishment of this feedback loop is a critical aspect of the whole system and is often described as a Plan/Do/Check/Act cycle (Allison 1995).

The implementation of QA/QC procedures has important social and organisational dimensions, as well as the practical ones. It involves attitudinal and behavioural changes as much as enforcing of standards. Good change-management processes and skills will help to achieve success including commitment from staff to the process, flexibility, and communication of the importance to stakeholders. The complexities of this process will increase as you move up the scale from the individual, to the laboratory, to the organisation, to the national, and to the international arena. The European programme called TACADAR (Towards Accreditation and Certification of Age Determination of Aquatic Resources; http://www.efan.no/tacadar) and its predecessor, the European Fish Aging Network, are good examples of the critical need to devote substantial amounts of time, planning, and resources to these aspects of the process.

Practical aspects of the implementation will also vary. Good documentation or recording of processes and practices is important for training, tracking the processing of samples, and demonstrating to others how QA/QC standards are set and maintained. It can be particularly important when it comes to backtracking to identify when subsequently detected reading errors may have started to become evident, and for securing procedural improvements to prevent the recurrence of such errors. In addition, for training and inter-laboratory checks, the exchange of data, samples, or images is valuable and may be sufficient, but structured workshops may also be needed to demonstrate practical skills or help resolve difficult issues of interpretation.

Expectations may need to be managed. The implementation of QA/QC procedures could potentially decrease a laboratory's performance in the short-term. As an example, in one age estimation laboratory, the time needed to develop reference collections initially caused a 30% decline in the number of samples processed. This is a significant impact and suggests that, unless the need is urgent, substantial change to practices may need to be implemented gradually. Such a short-term cost is, however, likely to be recovered by the long-term gain from having good QA/QC procedures in place.

There are likely to be other trade-offs in the implementation of QA/QC procedures, apart from the obvious one of diminishing returns. For example, measures that are designed to reduce error rates, such as the removal of outliers, the rejection of samples when full agreement between multiple readers is not achieved, and the rejection of samples with low readability, also have the potential to cause bias in the final dataset. The effects of the application of such filters should be tested rather than assuming that they will improve data quality.

Laboratories could benefit from developing insight into any differences in perception between themselves and their clients about the quality and priority of product services (Nicholson 1995). Effort may be wasted by improving aspects of work that are not considered important (overkill), or more seriously, by giving low priority to aspects on which clients are highly focused (a critical weakness). For example, spending time resolving reader discrepancies may be less important to stock assessment scientists than having the data more promptly.

All quality programmes should be assessed on the basis of common sense, logic and measured achievement (Hardie 1997). It should be realised that quality remains a means to an end and not the end itself. The QA/QC processes should aim to increase not only the skills of the persons involved, but also the awareness of the problems involved in QA that may affect the quality (McCurdy *et al.* 1999). A QA/QC system, however, cannot substitute for critical thinking, intellectual honesty, and professional skill (Adams 2002).

## Conclusions

Laboratories involved in fish age estimation are diverse with respect to numbers of staff, numbers of age estimates produced, and the extent to which they have good QA/QC systems established. It is likely that they are also quite varied in the extent to which they produce information that is free from significant errors or uncertainties. The need for change to current practices will obviously depend on the extent to which laboratories already have good systems in place. But we consider the reported variation in work practices that is evident from the responses to the questionnaire to reflect variation in awareness of the importance of quality issues rather than variation in the need for quality assurance and quality control measures.

We endorse the approach recommended by Campana (2001) for those seeking to establish a QA/QC protocol:

(1) The early development of a reference collection, preferably consisting of known age or consensus-aged structures.
(2) The periodic ageing of a randomly-drawn, blind-labelled subsample of the reference collection, intermixed with a subsample of structures recently aged as part of routine ageing.
(3) The use of age-bias graphs and CV as tools to evaluate the results of the monitoring.

A more empirical approach to quality issues, including the publication of findings, would be of general benefit to age estimation laboratories. The explicit examination of the effects of a range of different procedural options on the types and extent of errors produced would be very useful. There are some very instructive examples in this regard (Campana and Moksness 1991; Kimura *et al.* 1992), but much more could be learned.

Sound guidelines are now available for the implementation of QA/QC into fish age laboratories, although in some areas, the implications of alternative operating protocols have yet to be fully tested. Although there are still many laboratories around the world that are yet to enact these guidelines, it is clear that substantial progress is being made. There are laboratories able to demonstrate the local practical benefits of the implementation of such guidelines, and international networks have been created that serve as valuable models of what can be achieved on the broader scale. The challenge is for more laboratories to follow their lead. Improved fish age estimation will contribute to a reduction in the uncertainty or improvement in the precision of fisheries assessments, and ultimately to the sustainability of fisheries.

## Acknowledgments

Thanks are owing to those who distributed and completed questionnaires, to Gavin Begg and members of the Steering Committee of the Third International Symposium on Fish Otolith Research and Application for the invitation to contribute a keynote address on this topic, and to past and present staff of the Central Ageing Facility, Queenscliff, Australia for sharing their experiences, knowledge, and insights.

## References

Adams, T. M. (2002). 'Guide for the Estimation of Measurement Uncertainty in Testing.' American Association for Laboratory Testing, Frederick, MD. Available online at http://www.a2la.org/guidance/est_mu_testing.pdf, verified June 2005.

Allison, M. J. (1995). TQM comes to the laboratory. In 'Proceedings of the Sixth International Symposium on the Harmonization of the Role of Laboratory Quality Assurance in Relation to Total Quality Management'. (Ed. M. Parkany.) pp. 86–92. (The Royal Society of Chemistry: Melbourne.)

Beamish, R. J., and Fournier, D. A. (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 McFarlane, G. A. (1983*a*). The forgotten requirement for age validation in fisheries biology. *Transactions of the American Fisheries Society* **112**, 735–743. doi:10.1577/1548-8659(1983)112<735:TFRFAV>2.0.CO;2

Beamish, R. J., and McFarlane, G. A. (1983*b*). Current trends in age determination methodology. In 'Age and Growth of Fish'. (Eds R. C. Summerfelt and G. E. Hall.) pp. 15–42. (Iowa State University: Ames, IA.)

Beamish, R., and McFarlane, G. A. (1995). A discussion of the importance of aging errors, and an application to walleye pollock: the worldś largest fishery. In 'Recent Developments In Fish Otolith Research'. (Eds D. H. Secor, J. M. Dean and S. E. Campana.) pp. 545–565. (University of South Carolina Press: Columbia, SC.)

Bedford, B. C. (1983). A method of preparing sections of large numbers of otoliths embedded in black polyester resin. *Journal du Conseil, Conseil International pour l'Exploration de la Mer* **41**, 4–12.

Blumenthal, T. K. (1995). Motivation and 'marketing' of analytical laboratory for TQM: pride of place. In 'Proceedings of the Sixth International Symposium on the Harmonization of the Role of Laboratory Quality Assurance in Relation to Total Quality Management'. (Ed. M. Parkany.) pp. 1–7. (The Royal Society for Chemistry: Melbourne.)

Boehlert, G. W. (1985). Using objective criteria and multiple regression models for age determination in fishes. *Fishery Bulletin* **83**, 103–117.

Bradford, M. J. (1991). Effects of ageing errors on recruitment time series estimated from sequential population analysis. *Canadian Journal of Fisheries and Aquatic Sciences* **48**, 555–558.

Buckmeier, D. L. (2002). Assessment of reader accuracy and recommendations to reduce subjectivity in age estimation. *Fisheries* **27**, 10–14. doi:10.1577/1548-8446(2002)027<0010:AORAAR>2.0.CO;2

Cailliet, G. M., Botsford, L. W., Brittnacher, J. G., Ford, G., Matsubayashi, M., King, A., Watters, D. L., and Kope, R. G. (1996). Development of a computer-aided age determination system – evaluation based on otoliths of bank rockfish off California. *Transactions of the American Fisheries Society* **125**, 874–888. doi:10.1577/1548-8659(1996)125<0874:DOACAA>2.3.CO;2

Campana, S. E. (2001). Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *Journal of Fish Biology* **59**, 197–242. doi:10.1111/J.1095-8649.2001.TB00127.X

Campana, S. E., and Moksness, E. (1991). Accuracy and precision of age and hatch date estimates from otolith microstructure examination. *ICES Journal of Marine Science* **48**, 303–316.

Campana, S. E., and Thorrold, S. R. (2001). Otoliths, increments, and elements: keys to a comprehensive understanding of fish populations? *Canadian Journal of Fisheries and Aquatic Sciences* **58**, 30–38. doi:10.1139/CJFAS-58-1-30

Campana, S. E., Annand, M. C., and McMillan, J. I. (1995). Graphical and statistical methods for determining the consistency of age determinations. *Transactions of the American Fisheries Society* **124**, 131–138. doi:10.1577/1548-8659(1995)124<0131:GASMFD>2.3.CO;2

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.

Cowan, J. H., Shipp, R. L., Bailey, H. K., and Haywick, D. W. (1995). Procedure for rapid processing of large otoliths. *Transactions of the American Fisheries Society* **124**, 280–282. doi:10.1577/1548-8659(1995)124<0280:NPFRPO>2.3.CO;2

Fletcher, W. J. (1995). Application of the otolith weight–age relationship for the pilchard, *Sardinops sagax neopilchardus. Canadian Journal of Fisheries and Aquatic Sciences* **52**, 657–664.

Fowler, A. J., and Short, D. A. (1998). Validation of age determination from otoliths of the King George whiting *Sillaginodes punctata* (Perciformes). *Marine Biology* **130**, 577–587. doi:10.1007/S002270050279

Francis, R. I. C. C., Paul, L. J., and Mulligan, K. P. (1992). Ageing of adult snapper (*Pagrus auratus*) from otolith annual ring counts: validation by tagging and oxytetracycline injection. *Australian Journal of Marine and Freshwater Research* **43**, 1069–1089.

Gröger, J. (1999). A theoretical note on the interpersonal correction of age readings by means of calibration techniques. *Archive of Fishery Marine Research* **47**, 77–101.

Gröger, J. (2003). Theoretical and practical considerations with respect to sample size in statistical quality control of otolith readings. *Archive of Fishery & Marine Research* **50**, 161–173.

Hardie, N. (1997). 'The Consequences of Quality: A Critical Look at the Quality Movement.' (Consensus Books: Sydney.)

Kalish, J. M., Beamish, R. J., Brothers, E. B., Casselman, J. M., Francis, R. I. C. C., *et al.* (1995). Glossary for otolith studies. In 'Recent Developments in Fish Otolith Research'. (Eds D. A. Secor, J. M. Dean and S. E. Campana.) pp. 723–729. (University of South Carolina Press: Columbia, SC.)

Kimura, D. K. (1977). Statistical assessment of the age-length key. *Journal of the Fisheries Research Board of Canada* **34**, 317–324.

Kimura, D. K., and Lyons, J. J. (1991). Between reader bias and variability in the age-determination process. *Fishery Bulletin* **89**, 53–60.

Kimura, D. K., Lyons, J. J., MacLellan, S. E., and Goetz, B. J. (1992). Effects of year-class strength on age determination. *Australian Journal of Marine and Freshwater Research* **43**, 1221–1228.

Lagardère, F., and Troadec, H. (1997). Age estimation in common sole *Solea solea* larvae – validation of daily increments and evaluation of a pattern recognition technique. *Marine Ecology Progress Series* **155**, 223–237.

Lai, H. L., and Gunderson, D. R. (1987). Effects of ageing errors on estimates of growth, mortality and yield per recruit for walleye pollock (*Theragra chalcogramma*). *Fisheries Research* **5**, 287–302. doi:10.1016/0165-7836(87)90048-8

Macy, W. K., III (1995). The application of digital image processing to the aging of long-finned squid, *Loligo pealei*, using the statolith. In 'Recent Developments In Fish Otolith Research'. (Eds D. H. Secor, J. M. Dean and S. E. Campana.) pp. 283–302. (University of South Carolina Press: Colombia, SC.)

McCurdy, W. J., Hammer, C., and Appelberg, M. (1999). EFAN in the context of quality assurance and quality control. Working Document for the EFAN Annual Meeting 1999, Crete. EFAN Report 5 – 1999. European Fish Ageing Network, Institute of Marine Research, Bergen, Norway.

McGowan, M. F., Prince, E. D., and Lee, D. W. (1987). An inexpensive microcomputer-based system for making rapid and precise counts and measurements of zonations in video displayed skeletal structures of fish. In 'Age and Growth of Fish'. (Eds R. C. Summerfelt and G. E. Hall.) pp. 385–395. (The Iowa State Univeristy Press: Ames, IA.)

Morales-Nin, B., and Panfili, J. (2002). Chapter IIIA. Age estimation. In 'Manual of Fish Sclerochronology'. (Eds J. Panfili, H. d. Pontual,

H. Troadec and P. J. Wright.) pp. 89–110. (IFREMER-IRD coedition: Brest, France.)

Morales-Nin, B., Lombarte, A., and Japon, B. (1998). Approaches to otolith age determination: image signal treatment and age attribution. *Scienta Marina* **62**, 247–256.

Morison, A. K., and Rowling, K. R. (2001). Age, growth and mortality of redfish *Centroberyx affinis. Marine and Freshwater Research* **52**, 637–649. doi:10.1071/MF00016

Morison, A. K., Robertson, S. G., and Smith, D. C. (1998). An integrated system for production fish aging: image analysis and quality assurance. *North American Journal of Fisheries Management* **18**, 587–598. doi:10.1577/1548-8675(1998)018<0587:AISFPF>2.0.CO;2

NATA (1995). Guidelines for quality control in the analytical laboratory, Australia Technical Note 23. National Association of Testing Authorities, Sydney.

Nicholson, A. (1995). Customer satisfaction: a reality in analytical laboratories? In 'Proceedings of the Sixth International Symposium on the Harmonization of the Role of Laboratory Quality Assurance in Relation to Total Quality Management'. (Ed. M. Parkany.) pp. 232–240. (The Royal Society for Chemistry: Melbourne.)

Pilling, G. M., Grandcourt, E. M., and Kirkwood, G. P. (2003). The utility of otolith weight as a predictor of age in the emperor *Lethrinus mahsena* and other tropical fish species. *Fisheries Research* **60**, 493–506. doi:10.1016/S0165-7836(02)00087-5

Punt, A. E., Smith, D. C., Thomson, R. B., Haddon, M., He, X., and Lyle, J. M. (2001). Stock assessment of the blue grenadier *Macruronus novaezelandiae* resource of south-eastern Australia. *Marine and Freshwater Research* **52**, 701–717. doi:10.1071/MF99136

Reeves, S. A. (2003). A simulation study of the implications of age-reading errors for stock assessment and management advice. *ICES Journal of Marine Science* **60**, 314–328. doi:10.1016/S1054-3139(03)00011-0

Richards, L. J., Schnute, J. T., Kronlund, A. R., and Beamish, R. J. (1992). Statistical models for the analysis of ageing error. *Canadian Journal of Fisheries and Aquatic Sciences* **49**, 1801–1815.

Robertson, S. G., and Morison, A. K. (1999). A trial of artificial neural networks for automatically estimating the age of fish. *Marine and Freshwater Research* **50**, 73–82.

Robertson, S. G., and Morison, A. K. (2003). Chapter 19. Age estimation of fish using a probabilistic neural network. In 'Ecological Informatics: Understanding Ecology by Biologically-Inspired Computation'. (Ed. F. Recknagel.) pp. 369–382. (Springer-Verlag: New York.)

Schaalje, G. B., Shaw, J. L., and Belk, M. C. (2002). Using nonlinear hierarchical models for analyzing annulus-based size-at-age data. *Canadian Journal of Fisheries and Aquatic Sciences* **59**, 1524–1532. doi:10.1139/F02-120

Small, G. J., and Hirschhorn, G. (1987). Computer-assisted age and growth pattern recognition of fish scales using a digitizing tablet. In 'Age and Growth of Fish'. (Eds R. C. Summerfelt and G. E. Hall.) pp. 397–410. (The Iowa State University Press: Ames, IA.)

Sych, R. (1974). The sources of errors in ageing fish and considerations of the proofs of reliability. In 'The Ageing of Fish'. (Ed. T. B. Bagenal.) pp. 78–86. (Unwin Brothers Ltd: London.)

Szedlmayer, S. T., Szedlmayer, M. M., and Sieracki, M. E. (1991). Automated enumeration by computer digitization of age-0 weakfish *Cynoscion regalis* scale annuli. *Fishery Bulletin* **89**, 337–340.

Troadec, H. (1991). Frequency demodulation on otolith numerical images for the automation of fish age estimation. *Aquatic Living Resources* **4**, 207–219.

Troadec, H. (Ed.) (1998). Establishment of a digital image reference database. EFAN Report 3 – 1998. European Fish Ageing Network, Institute of Marine Research, Bergen, Norway.

Troadec, H. (Ed.) (2000). Use of digital tools for exchanging annotated images of calcified structures. EFAN Report 8 – 2000. European Fish Ageing Network, Institute of Marine Research, Bergen, Norway.

Troadec, H., and Benzinou, A. (2002). Chapter VI. Computer-assisted age estimation. In 'Manual of Fish Sclerochronology'. (Eds J. Panfili, H. d. Pontual, H. Troadec and P. J. Wright.) pp. 199–241. (IFREMER-IRD coedition: Brest, France.)

Tyler, A. V., Beamish, R. J., and McFarlane, G. A. (1989). Implications of age determination errors to yield estimates. *Canadian Special Publication of Fisheries and Aquatic Science* **108**, 27–35.

Welleman, H. C., and Storbeck, F. (1995). Automatic ageing of plaice (*Pleuronectes platessa* L.) otoliths by means of image analysis. In 'Recent Developments in Fish Otolith Research'. (Eds D. H. Secor, J. M. Dean and S. E. Campana.) pp. 271–282. (University of South Carolina Press: Columbia, SC.)

Worthington, D. G., Doherty, P. J., and Fowler, A. J. (1995). Variation in the relationship between otolith weight and age – implications for the estimation of age of two tropical damselfish (*Pomacentrus moluccensis* and *P. wardi*). *Canadian Journal of Fisheries and Aquatic Sciences* **52**, 233–242.

Zar, J. H. (1984). 'Biostatistical Analysis.' 2nd edn. (Prentice-Hall: Englewood Cliffs, NJ.)

Manuscript received 17 August 2004; revised 21 March 2005; and accepted 6 April 2005.