PGCCDBS Guidelines for Workshops on Age Calibration

(update)

7 – 11 February 2011

Vienna, Austria.

Introduction

The main objective of an age reading workshop is to decrease the relative/absolute bias and to improve the precision (reduce CV) of age determinations (their reproducibility) between age readers of the different age reading laboratories. An exchange of calcified structures must be carried out first to indicate the errors in age reading before a recommendation for an age reading workshop can be made (see previous section).

Problems Indicated by the Exchange.

At a workshop an attempt should be made to solve the problems indicated by the exchange. The following possible problems in reading might exist:

- the age reading methods differ too much (as indicated by statistical tests);
- the precision in age reading is too low for certain age readers;
- there is a strong bias in the age readings of young and/or old fish;
- precision differs considerably for different preparation methods;
- inexperienced readers;
- other age reading problems.

It is very important to ensure that the workshop also addresses any issues relating to age reading as highlighted by the relevant assessment working group. The workshop coordinator should endeavour to get feedback from the assessment working group chair on what he/she feels are important outcomes that should be achieved from the upcoming workshop. It is recommended that the chair of the relevant assessment working group should be encouraged to contribute to the workshop as an end user of the data, either in person where possible or via webex etc.

Topics to Consider When Preparing for a Workshop

The following topics can be and all should be considered:

- The biology of the species;
- The results of previous exchanges and workshops;
- When and how the age reading technique was validated;
- The sample processing techniques used at the different age reading laboratories;
- If necessary, try to standardise the processing techniques of calcified structures;
- Agreement on age determination criteria;
- Discuss disagreements in age reading results from the sets of the calcified structures read during the exchange and at the workshop and try to agree on the age reading method;
- Determine at the end of the workshop the precision in age reading and the relative bias (if possible the absolute bias);
- Estimate improvement in age reading concerning precision and bias by comparing exchange set and the last set at the workshop;
- Make recommendations on how to improve the age reading quality;
- Indicate which calcified structures can be used for the "agreed collection" and (if possible) produce digitised images.

Other topics may be addressed based on the conclusions from the exchange.

PGCCDBS recommends that workshop coordinators use the following criteria for classifying age reading performance into 'good', 'medium' or 'bad'.

- **Bad ageing performance:** When the quality of the data is unknown or there are serious concerns about the reliability of the age data and/or its value to stock assessment WGs. Indicators may include poor agreement between age readers and age data that do not appear to agree with other methods of growth estimation for the stock/species. Causes may include difficulty in observing/interpreting calcified structure (CS) growth patterns, no protocol for preparation/age reading and the use of inappropriate CS or preparation methods.

- **Medium ageing performance:** The age data is sufficiently reliable to be used for stock assessment purposes but improvement is required. Indicators may include levels of agreement between age readers that are below a reference target value for the stock/species (e.g. VIIa cod - 90%, redfish - 40%), difficulty in interpreting aspects of CS growth patterns (e.g. disagreement over the location of the first annulus or otolith edge interpretation), protocols for age reading are used but may need revision and the use of less reliable preparation/observation methods.

- **Good ageing performance:** The age data is considered reliable. Indicators may include repeated high levels of agreement between age readers at successive exchanges or workshops. Causes may include calcified structure CS growth patterns that are easier to interpret, good protocols for preparation/age reading and the implementation of QA and/or QC procedures at individual institutes.

  Stock coordinators should be aware of levels of percentages of agreement and CV's amongst stock assessment readers. Age coordinators should recommend achievable percentage agreement and CVs based on the most recent exchange and workshops.

Workshop Participants

Everyone who participated in the exchange should also participate in the workshop, and vice versa; no one should participate in the workshop unless they also took part in the exchange.

Experimental Design in Age Reading Workshops

Workshops usually compare the performance of readers between the start and end of the workshop. These comparisons need to be planned from the start of the exchange and carried out using the principles of designed experiments. The most important ideas for experimental design are to compare like with like and to control for other variables that affect age reading ability. For example, do not provide otoliths for the exchange from one area then read otoliths from a different area at the end of the workshop.

It is important to avoid running the before and after comparisons on exactly the same set of otoliths. This is necessary if there are small numbers of otoliths but otherwise is undesirable as improvements seen in agreement may be from remembering specific cases and not apply in general. The procedure for generating two sets of otoliths for comparison of exchange and workshop results should be: Define the relevant strata

and assign otoliths by strata randomly to either the first or second set. The two sets do not have to be the same size. When the first set is for the exchange and the second set for the end of the workshop it is sensible to make the second set smaller. If the age workshop coordinator can specify changes in reading bias or CV that are biologically meaningful to detect then sample size calculations can be carried out to help decide how big the data sets should be.

The 'Tool for Age Reading Comparisons' was developed by Eltink *et al.* in 2000, has proved an invaluable contribution to Quality Control for fish age calibration. Eltink *et al.* (2000) advised that the precision errors in age readings are best described by the coefficient of variation CV by age group (CV = st. dev/mean age recorded). Although CV is often the preferred statistical tool for this task, the index of average percentage error (APE) is also commonly used. (Kimura, D. K., and Anderl, D.M. 2005; Morison *et al.* 2005). The dangers of the percent agreement statistic have long been recognised (Beamish and Fournier 1981; Chang 1982; Campana 2001), yet despite this Campana *et al.* (1995) reported that roughly 35% of 21 randomly sampled age comparison papers published between 1985 and 1995, used only percent agreement. More recently Morison *et al.* (2005) reported that responses to a questionnaire to assess current QA and QC practices that was completed by representatives of over 50 fish ageing laboratories worldwide, indicated that percentage agreement was still the most commonly used measure of precision (40% of respondents) despite its limitations and criticisms. Nevertheless, in order to ensure comparability between studies on different species, the CV and/or APE has to be reported as obligatory precision estimate.

Improvements to the original spreadsheet tool have been developed at CEFAS, UK. Eltink compared a number of results in the "work table of the bias test" of the original spreadsheet and the new spreadsheet, which calculates the results of the bias test in the overview table and so far has not found any discrepancies (Eltink *pers. com.*). Eltink advises that the new spreadsheet is much faster than the original one. The downside is that the new spreadsheet is limited in the number of otoliths as well as in the number of age-readers. The original spreadsheet did not have these restrictions. Eltink concludes that the new spreadsheet cannot replace the original one at this stage, but can be used within these restrictions.

Generic ToRs for ageing workshops

a) Provide information on participating laboratory procedures

- Sampling and storing of calcified structures.
- Equipment and preparation of calcified structures
- Documentation on processes and protocols (QA)
- How age determination are being checked within laboratories (QC):
  - availability of reference collections
  - results of age reading comparisons between readers
  - percentage of samples re-read
- Estimate (relative) accuracy and precision

b) Resolve interpretation differences between readers and laboratories.

Disagreements on the interpretation of annual increments can exist between experienced readers. Usually these differences are resolved when the readers discuss the otoliths jointly (note: annotated images largely simplify this process). However, this is not always the case and then follow-up actions must be formulated.

c) Create or update an ageing manual

There should be a standardised ageing manual for each species in a unified format that is internationally agreed upon by all experienced age readers. This manual focuses on the interpretation of the structures (e.g. date of birth, interpretation of rings and edges, period of opaque and translucent ring formation). The manuals on preparation of calcified structures are usually created and updated on the national level.

d) Collate agreed age reference collection.

The output of every workshop should be an agreed age reference collection. Preferably the agreed interpretation should be annotated (as a separate raster layer – see previous section) in the images. These sets of images could then be made available online to train new age readers or to have as a reference set for experienced readers. If establishing a digital collection on a website is not possible, then information about location of the reference collection and contact person should be available on the website.

e) Formulate follow-up actions

See the guidelines in the following section

f) Formulate species (and stock specific) target and threshold statistics

As tool for the evaluation of the quality of age readings we recommend that target and threshold statistics are formulated for each species and stock. The statistics refer to the percentage agreement, the CV and the bias. The target value is the value you would like to achieve and know is possible based on exchange and workshop results. The threshold value is the minimum value required before a reader is qualified to supply data to working groups and can if necessary be derived by discussion between expert readers. Usually, a CV of 5% is set as a threshold for sufficient data quality (Campana 2001).

Guidelines for follow-up actions

Dissemination of the results

Dissemination of the results is in principle the responsibility of the coordinator of the exchange and/or workshop. The full report of the workshop should be made available on the internet, and placed (in pdf-format) in the PGCCDBS document repository (http://www.ices.dk/reports/acfm/pgccdbs/PGCCDBSdocrepository.asp). An extended summary of all workshops and exchanges should be submitted to PGCCDBS and to the relevant working group/PGCCDBS liaison person, and the stock coordinator. This extended summary should provide sufficient information to enable the working group to judge whether or not the quality of the ageing data (by country) is sufficient to include the data in a quantitative stock assessment.

The extended summary should contain the following information:

1) Description of sets of calcified structures included in the exchange and/or workshop:
2) The number of calcified structures in each set
3) Composition (age and/or length structure, area)
4) Preparation methods
5) Images available?

6) Description of participants (numbers per country etc.)

7) Number of readers, laboratories and countries

8) Expertise level of each reader (trainee, intermediate, experienced)

9) Which readers provide ageing data to the WG's

10) Which laboratories provide ageing data to the WG's but are not represented in calibration

11) Accuracy and precision estimates

    i) Percentage agreement, CV and bias by age group

    ii) Only readers providing data to WG's

    iii) Readers combined

    iv) By reader (anonymous, but lab/country stated)

    v) If relevant, by stratum (spatial and/or temporal differentiation

12) Summarise currently existing ageing problems, either detected in exchange or not solved in workshop.

13) Evaluation of quality of age data provided to WG

    i) Preferably a quantitative evaluation (i.e. in relation to target and threshold statistics)

    ii) If not possible then a qualitative evaluation

14) A list of the expert groups to be informed.

Specific follow-up actions

If ageing problems are not solved within the ageing workshop, then the participants must formulate clear follow-up actions which will lead to solving the ageing problems. If there are no distinct ageing problems, but the workshop thinks the general ageing quality can be improved by follow-up actions than these should be formulated clearly. The workshop should point out who is responsible for coordinating and carrying out the follow-up actions and in what time frame. The required follow-up can differ depending on the species and the problem occurring. To aid the workshop coordinator some possible follow-up actions are listed here:

- Validation exercises must always be encouraged. A continuous comparison of age readings does not always solve the problem (an example to be learned from: the bias in hake ageing).
- In some species in which the contrast between the structures is poorly visible it may be advisable to improve preparation methods.
- If one or a few readers are disagreeing with the majority of experienced readers, then small scale regional exchanges and/or meetings can be organised.
- If interpretation problems of the first annuli are occurring, then back-calculated growth can provide an indication on the correct interpretation. If samples of '0'-group fish are available throughout the 1st year of life, the period of annual translucent zone may be determined by marginal incremental analysis.
- If age reading protocols are not available for all participants this should be remedied.
- When new age reading criteria are established and agreed at a workshop, all readers should be asked to implement the agreed criteria directly after

the workshop, using a small set of images or age reading material. This serves the dual purpose of ensuring that the agreed ageing criteria are adopted by all and also provides a format for testing the new criteria.

Workshop Checklist

1.	Inform the PG Liaison person/stock coordinator/ or chair of the relevant AWG of the pending workshop and look for feedback.	**[ ]**
2.	Establish list of participants from the exchange and direct them to the European Age Readers Forum (EARF).	[ ]
3.	Using the EARF, the workshop coordinator should agree a date, and location for the WK and any other house – keeping issue around the organisation of the WK.	[ ]
4.	Follow PGCCDBS Guidelines regarding the design and generic TOR's for the WK.	[ ]
5.	Conduct Workshop.	[ ]
6.	Complete analysis – follow PGCCDBS Guidelines on this.	[ ]
7.	Present analysis for age readers contributing to Stock Assessment.	[ ]
8.	Present analysis for all age readers in the annex of the report.	[ ]
9.	Circulate the draft report of the Workshop to all participants.	[ ]
10.	Forward the final report from the workshop to ICES, the AWG/stock coordinator and the Chair of PGCCDBS.	[ ]
11.	Provide an extended abstract to the PGCCDBS.	[ ]
12.	Upload images, or a link to where a set of agreed ages, resulting from the Workshop, can be found to the EARF.	[ ]