**WGBIOP GUIDELINES FOR AGE READING REFERNCE COLLECTION**

Reference collections of fish age interpretations are an essential tool for assessing consistency in age interpretations of any hard structure (otoliths, scales, vertebrae or ….). These reference collections may serve as tool for training purposes and as a benchmark for assessing age interpretation consistency between age readers and across time. The requirements for addressing these two objectives differ. Guidelines for TRAINING and CONTROL reference collections are therefore treated separately.

1. **TRAINING REFERENCE COLLECTION (TRC)**

OBJECTIVES

Training of new readers and self-testing of experienced readers, specifically:

* *Training of new readers*: A collection that enables new readers to learn to use age interpretation
* *Confirmation of correct age*: The quality assessment aims to verify that age interpretation, on average, is consistent with the actual age in cases where “known age” is available.
* *Consistency within readers*: Self-assessing age interpretation consistency within individual age readers.

DESIGN OF TRAINING COLLECTION

The TRAINING reference collection should consist of samples that have been discussed by (advanced) age readers during exchanges or workshops, from which consensus age samples are selected. The samples discussed should be selected following the outlines for SmartDots events, assuming that samples cover all areas, full age range, all quarters and both sexes of a given species.

*1. Step: Consensus agreement*

* The first step of setting up a TRC is to define what “consensus agreement” is for a given species.
* Consensus agreement for a sample was defined to be when a minimum of 3 advanced readers provide the same age interpretation, with no more than 2 from one institute, agree (EFAN Report 3-2000, Guidelines and Tools for Age Reading Comparisons). But the applicability of this recommendation may not be applicable to all species/stock etc.

*2. Step: Agreement threshold*

Select an agreement threshold, above which age interpretation can be considered as consensus ages.

* This threshold will depend on the species, the age range and level of age interpretation difficulty.
* Example: For long-lived species a PA > 80% (from advanced readers) could be considered as being consensus age, while the same level would not be high enough for short-lived species.

*3. Step: TRAINING reference collection setup*

Define the setup of the samples added to the TRC

* Should TRC be static or dynamic?
* Two possible scenarios:
  + TRC is static → samples from new SmartDots events are NOT included. This type of TRC may not represent the visual appearance of recent samples when the visual appearance and growth are changing as response to i.e. climate change
  + TRC is dynamic → samples from new SmartDots events ARE included in the TRC. This implies that the number of samples in the TRC is increasing with each event, unless a) samples are remove, or b) a SmartDot function/module is developed where specific strata can be selected.

SAMPLE SELECTION FOR TRAINING AND SELF-MONITORING

The development of a useful approach to TRC sample selection depends a) on what type of TRC is selected (static or dynamic), and b) on whether a SmartDots module for extracting specific samples will be developed.

Sample selection strategy should consider:

* Random selection of consensus age samples from the entire TRC sample database – again, this depends on whether the TRC will be static or dynamic
* If the TRC is dynamic, we need a tool for selecting images with consensus ages stratified from specific strata (year of event, age range, area, etc.).
* If SmartDots module will not be available in the near future, we need to consider other tools for obtaining a suitable selection of images for training.

1. **CONTROL REFERENCE COLLECTION (CRC)**

OBJECTIVES

Assess changes in interpretation over time, specifically

* *Confirmation of correct age*: The quality assessment aims to verify that age interpretation, on average, is consistent with the actual age in cases where “known age” is available.
* *Long-term consistency*: Measurements should ensure that age interpretation (and accuracy if “known age” samples are available) do not change over time, i.e. ensures consistency in both short and long term.
* *Consistency between readers*: Quality assessment to ensure that age interpretation between different age readers is comparable in cases where a change in age reader is occurring.
* *Consistency between interpretation methods*: Assessment of the impact of new or updated interpretation methods over time, to ensure that i.e. a new interpretation method will on average result in the same divergence (if the updated method results in such one) is consistent back in time.

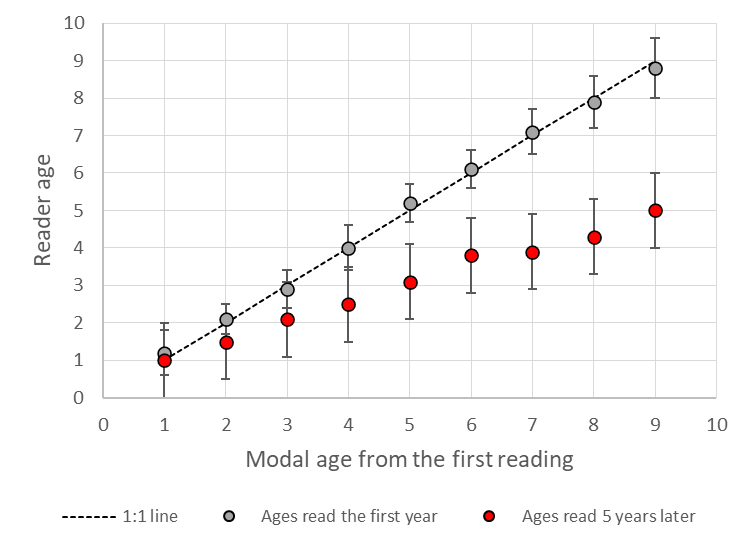


Figure 1 Graphical illustration of how a control reference collection can be used to asses changes in age interpretation over time for 1) within individual readers over time, where interpretations of the first reading correspond to the modal age at the time (grey dots), while age interpretations 5 years later are consistently lower (red dots), or 2) between a previous reader (grey dots) and a new reader (red dots), or 3) if interpretation methods have changed in the time between the original (grey dots) and the new (red dots) interpretation methods.

DESIGN OF CONTROL COLLECTION

The CONTROL reference collection should consist of samples representative of the data that is submitted for stock assessment and should cover all areas, full age range, all quarters and both sexes of a given species – and as many years as possible.

*1. Step: Design of sample selection*

* All samples from all previous and new SmartDots exchanges should be considered as being part of the CRC sample database
* CRC should consist of samples representative of what is submitted for stock assessment
* All AQ scores should be included, even AQ3s, to monitor if otolith interpretation is perceived the same way over time or between old and new readers
* *2. Step: Design of sample selection*

*2. Step: Assessment of historic reference collections*

The quality of images and data from historic collections (pre-SmartDots), should where possible, be included in the SmartDots database. An assessment of this quality will need to include whether:

* Available images and associated data are of sufficient quality
* If only physical samples are available, whether these are in sufficiently good shape to allow imaging
* An assessment whether including historic collections would provide useful information for control of age interpretation consistency – or be useful for obtaining Age Error Matrices

SAMPLE SELECTION FOR CONTROL ASSESSMENT

* Samples for temporal consistency in age interpretation should be extracted randomly from combined CRC, with a balanced number of samples distributed across:
  + time series
  + modal ages
  + areas
  + AQ scores
  + sex
* Sample numbers to be selected for consistency control:
  + Campana (2001) suggested that a sample size of 100 from the CRC in addition to 100 samples from the most recent exchange should be sufficient to insure reasonable statistical power for statistical quality control testing
  + However, the number of samples required to obtain a suitable consistency control depends on the age range, geographic range, consistency in age interpretation etc. of a given species and stock
* Development of sample selection tool: Would be nice with a tool to simulate what sample size would be required with a given set of parameters specified above