What is the best way to age Antartic minke whales?

Paul R. Wade

National Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA, Seattle, WA, USA, 98115

Accurate aging of Antarctic minke whales is crucial to some of the stated objectives of the NEWREP-A program. In particular, Objective I(ii) (*Improvement in the precision of biological and ecological parameters*) relies on accurate age determination to estimate parameters such as Age of Sexual Maturity (ASM), and estimates from the Statistical Catch-At-Age (SCAA) analysis, such as MSYR(1+). Furthermore, the outputs from the SCAA are listed as being crucial to Objective I (iv) (*Specifications of RMP/IST*). It goes without saying that estimates of the age of Antarctic minke whales is essential to an SCAA analysis.

Kitakado et al. (2013) investigated reading errors in aging Antarctic minke whales using ear plugs. They pointed out that both random age-reading error and age-reading bias can affect the results of the SCAA analyses. They noted that Reeves (2003) found that random errors led to 'smoothing' of annual recruitment estimates. Bias in age estimates is potentially a greater problem. Importantly, Kitakado et al. (2013) noted that a 'drift' of age-reading methods could have led population models to estimate spurious trends in recruitment for the Antarctic minke whale, citing Butterworth and Punt (2009). Kitakado et al. (2013) provide a method for adjusting for possible bias due to age-reading errors. Here I comment on the issue of bias and precision in estimating the age of minke whales from ear plugs, and on what is accomplished by the Kitakado et al. (2013) adjustment.

Precision

The NEWREP-A proposal states that alternative methods of aging such as DNA methylation methods (e.g., Polanowski et al. 2014) are not precise enough to be used in models such as the SCAA. I do not believe this statement can be supported by the facts. The reported standard error of the humpback whale age estimates in Polanowski et al. (2014) is 2.99. Note that this is a standard error from the *true* age, as this study used known age animals (photo-identified as calves). Therefore, this amounts to a mean squared error that incorporates both precision and bias. The standard deviation of the chemical method alone, using repeated analyses of the same DNA, was only 2.2 years (Polanowski et al. 2014).

Kitakada et al. (2013) report standard errors for the four Japanese readers under the important assumption that the ages provided by Lockyer are the truth. Obviously, this is an untestable assumption at this point. Those standard errors are not substantially smaller than the 2.99 reported by Polanowski et a. (2014), and in fact they generally increase with age, which is problematic as well (whereas in the DNA methylation technique the errors remain the same regardless of the age of the animal). In Figure 8 of Kitakado et al. (2013) (included below as Fig. 1), the standard errors of the age estimates for the four readers range approximately from 1-7, 1-3, 1-3, and 1.5-3,

depending upon the age of the whale. It is clear that these standard errors are not substantially smaller that 2.99, and for some age ranges are substantially larger.

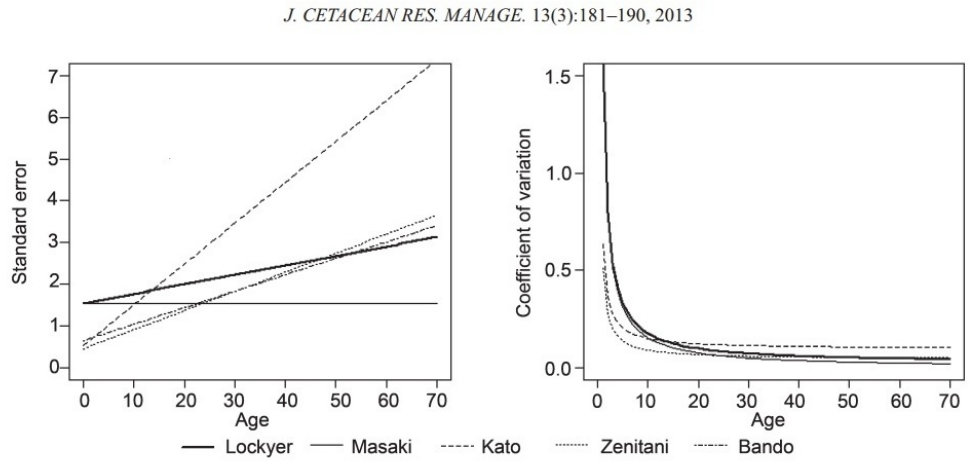

Fig. 8. Standard errors (left) and coefficients of variation (right) for the control and Japanese readers if the control reader (Lockyer) is assumed to be unbiased (Case 1).

Figure 1. Plots of standard error and coefficient of variation of the age estimates by Japanese readers, This is Figure 8 from Kitakado et al. (2013).

Bias

Kitakado et a. (2013) use a model to correct the four Japanese reader's age estimates to the standard provided by Lockyer. However, there is one fundamentally important assumption in the correction – that of what constitutes the "true" age. As they state:

> *"It should be noted that the analyses on which this paper are based are predicated on Lockyer's age estimates. It cannot necessarily be assumed that Lockyer provides unbiased estimates of true age. Overall, the results suggest that the age-reading errors for Lockyer and the four Japanese readers differ."* (Kitakado et al. 2013).

In other words, Lockyer's ages were different from the four Japanese readers but it is unknown which, if either, is the truth. At least one set of readings, and perhaps both, are biased. There is no way that true age can be ground-truthed from ear plug data alone. It is entirely possible that the four Japanese reader's estimates are less biased, and the method corrects the estimates further from the truth, and makes the final age estimates more biased.

The minimum bias in one (unknown) set of age estimates is not small. For an age 1 whale the bias ranges from ~1 to 3 years, for an age 20 whale the bias also ranges from ~ 1 to 3 years, and for an age 40 whale the bias ranges from ~ 3 to 7 years (Fig 2). The true biases would actually be larger if all the readers (including the control reader) were biased in the same direction. The model correction in Kitakado et al. (2013) may make the ages less biased (if the control reader was unbiased), or it may make the more biased (if the control reader themselves is biased). At a minimum, it would seem that the SCAA should be run under both assumptions.

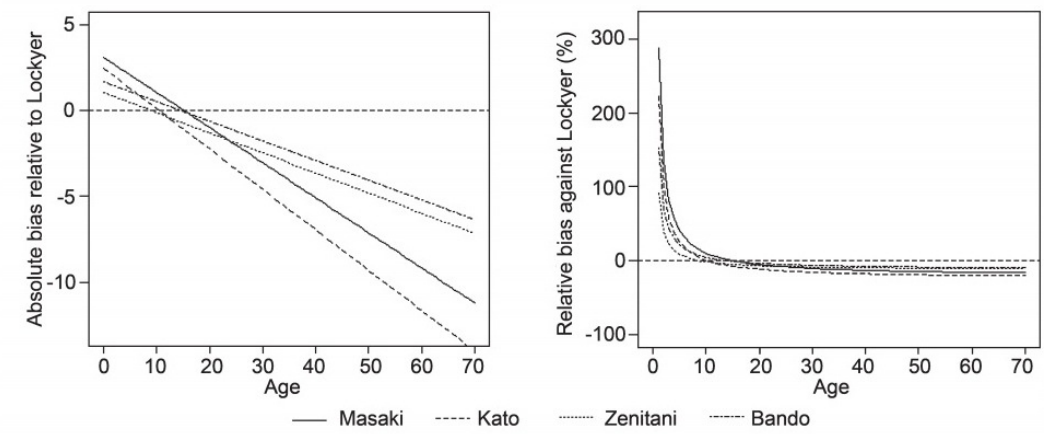

Fig. 7. Absolute (left) and relative (right) biases for the Japanese readers relative to the control reader (Lockyer), who is assumed to be unbiased (Case 1).

Figure 2. Absolute biases (on the left) for the Japanese readers relative to the control reader (Lockyer), who is assumed to be unbiased. This is Figure 7 from Kitakado et al. (2013).

Consistency and repeatability

Lockyer's reading of the ear plugs is held up as the gold standard for aging baleen whale ear plugs. However, note that in her first reading she was able to age 228 out of 250 ear plugs (91.2%), whereas one week later, she was only able to age 216 out of the 250 *same* ear plugs (86.4%). In other words, the age of 12 of the 250 (4.8%) ear plugs could be determined in one reading, but could not be determined during a second reading only one week earlier, by the same reader. This is not intended to be a criticism of Lockyer's ability; it is clear that the reading of baleen whale ear plugs is an extremely difficult task, and the nature of the task makes it an exercise that it is not fully repeatable. 4.8% seems an unacceptability high percentage of plugs to move from being readable to being undetermined, and raises the question as to whether this is a wholly reliable method of aging whales. Furthermore, the four Japanese readers determined the age of 100%, 100%, 100%, and 99% of the same ear plugs. Given that Lockyer was used as the control reader, it seems surprising that the control reader was not comfortable determining the age of 9% or 14% of the ear plugs, but the 4 Japanese readers were comfortable determining 100% of them. Any correction of the Japanese readers by using Lockyer's ages would not be able to use those 9% or 14% of the samples in the correction. Given that Lockyer felt she could not age these samples, these were clearly the most difficult samples, and likely to have the greatest reading error. Therefore, the estimates of aging error in Kitakado et al. (2013) will likely be biased low (under-estimated) because they do not include those difficult samples. This again points to the fact that the DNA methylation estimates are not, in fact, less precise than the ear plug estimates.

In contrast, the chemical techniques associated with DNA methylation do not seem subject to the same failure rate. Polanowski et al. (2014) did not report any samples that they failed to be able to age. In fact, they note that cytosine methylation is reasonably stable and has been successfully purified from ancient DNA as old as 60,000 years. They suggest this means that the DNA methylation technique would be

successful even on degraded DNA such as faecal samples or partially decomposed samples.

The aging of Antarctic minke whales using ear plugs has been shown to vary substantially between readers, with substantial random error as well as substantial bias. Additionally, there appears to be a lot of variability within a single reader. This does raise the possibility of 'drift' in the aging of ear plugs over time, which could be very problematic for the SCAA analyses. The experiment of adjusting the age estimates from the four Japanese readers to the age estimates of a control reader was well intended, and in theory should prevent 'drift', even if it cannot reliably correct for bias. However, there is no guarantee that the control reader themselves will not 'drift' through time. Presumably this correction would need to be done again in the future to ensure no drift has occurred, but if the control reader drifts as well, this could be masked. In contrast, the DNA methylation technique should be completely immune to any possible 'drift' through time, or differences between different technicians.

Conclusions

The epigenetic method of DNA methylation has some sampling error, so this will contribute to the possibility of the smoothing of annual recruitment estimates, but the age estimates from ear plugs have the same issue on the same order of magnitude. Therefore, the assertion that DNA methylation techniques are not precise enough to be useful in the SCAA analysis is clearly wrong. Given that it is currently impossible to identify known age Antarctic minke whales, it is impossible to determine the absolute bias of either DNA methylation or ear plug methods. However, one clear advantage of DNA methylation is that it is much more repeatable, and will not have errors and bias introduced based on which technician runs the samples. In contrast, the results of Kitakado et al. (2013) show that there is substantial error and bias between different ear plug readers, even down to the level of categorizing whether an ear plug can be reliably read. Moreover, DNA methylation should be able to be used on essentially all samples, whereas some ear plugs cannot be read. Both of these issues point to some strong advantages of the DNA methylation technique. Drift is the major issue, as it could lead to spurious trends (Kitakado et al.).

Moreover, Punt et al. (2013) confirm that the results of the statistical catch-at-age analysis for Antarctic minke whales are sensitive to whether age-reading error is ignored or accounted for. If you do not know which direction to correct the bias, then one cannot be certain that the results of the SCAA have been correctly adjusted, so the results of the SCAA will be subject to uncertainty about whether the results have been biased by biased ages or not.

I note that in Appendix 4 it is proposed to test the DNA methylation technique in Antarctic minke whales. I encourage this research, but note that there is no need for the NEWREP-A program to accomplish this. DNA methylation could be applied to existing ear plug and DNA samples from JARPA/JARPAII to provide a relationship for predicted age that could then be used for future DNA samples. Therefore, there is no need for lethal sampling in order to establish an aging technique for Antarctic minke whales from DNA methylation techniques.

References

Kitakado, T., Lockyer, C., and Punt, A.E. 2013. A statistical model for quantifying age-reading errors and its application to the Antarctic minke whales. J. Cetacean Res. Manage. 13(3):181–190.

Polanowski, A.M., Robbins, J., Chandler, D. and Jarman, S.N. 2014. Epigenetic estimation of age in humpback whales. Mol. Ecol. Resour. 14:976-987.

Punt, A.E., Bando, T., Hakamada, T. and Kishiro, T. 2013. Assessment of Southern Hemisphere Minke Whales using Statistical Catch-at-age Analysis Paper SC/65a/IA1 presented to the IWC Scientific Committee, June 2013 (unpublished). 48pp.