

## Addendum

# Independent Expert Peer Review of the Close Kin Mark Recapture Assessment for School Shark

### Review panel:

Professor Colin Simpfendorfer (Chair)

Professor Sean Cox

Dr Kevin Stokes

Dr Robin Waples

This addendum was prepared by the Panel Chair following a meeting between Mr Patrick Cordue and two members of the review panel. The addendum represents the views of the panel only, and summarises the Panel's recollection of the meeting.

Addendum provide: 7 June 2021

### Background

Following the submission of our review of the School Shark CKMR assessment in February 2021, AFMA arranged for the panel to meet with Patrick Cordue who had previously reviewed the assessment for the fishing industry. The review panel had extensively used Mr Cordue's report in its consideration, but agreed to the requested meeting to ensure that all aspects of assessment were fully considered. Two members of the review panel (Prof Colin Simpfendorfer, Dr Robin Waples) met with Mr Cordue via Microsoft Teams; a representative of AFMA was also in attendance but did not participate in the meeting.

The meeting agenda as proposed by Mr Cordue was:

1. *The data that could have been collected if a careful survey and experimental design had been undertaken (5 minutes)*
2. *The inability to distinguish between FSPs and POPs genetically (10 minutes)*
3. *Ageing error and the problems of using a method that conditions on age when the nature and extent of the ageing error is not known (20 minutes)*
4. *Technical errors that were made in the assessment and on which the panel made no comment:*
  - a. *Neglecting to include nibblings pairs (10 minutes)*
  - b. *Incorrect calculation of the probabilities of MHSPs and PHSPs (at least according to the documentation) (5 minutes)*
  - c. *Incorrect calculation of the likelihood (10 minutes)*
5. *The use of a CK model that does not need to condition on age (thus avoiding the problems with ageing error) (5 minutes)*

Following the meeting Mr Cordue provided a written explanation of item 4b, which is provided at Appendix A

## Issues raised

### 1. The data that could have been collected if a careful survey and experimental design had been undertaken

**Mr Cordue's input:** Shared thoughts on the survey design used in the sampling of CKMR data. Thinks it should have posed different stock structures to test and form the sampling strategy; suggests they should have collected a lot more information (e.g. age frequencies, size structure, etc..) that would have helped with interpretation of the data.

**Panel conclusion:** The post-hoc assessment of data collected is of limited value to the reievw of the assessment, but can be useful in consideration of future assessments. Aspects of this point on stock structure were covered in the original review report. As indicated in that report, further work on stock structure and the interpretation of the results would be beneficial to the assessment. The panel also suggests that prior to any future assessment that SharkRAG and researchers discuss assumptions to be used and what they mean for the sampling strategy to ensure there is a shared understanding.

### 2. The inability to distinguish between FSPs and POPs genetically

**Mr Cordue's input:** Was concerned that CSIRO was not able to genetically distinguish between POPs and FSPs. The assessment avoided this issue by excluding animals older than the age at maturity (ie no POPs would be possible), but Mr Cordue argued that the high level of ageing error meant that this was not 100% effective.

**Panel conclusion:** In CSIRO's response to this question for the original Cordue review, they identified that null alleles made the distinguishing of FSPs and POPs difficult and hence they restricted the age range to deal with this issue. The panel did not examine the genetic data in detail but generally agreed with the CSIRO approach in the current assessment. However, this topic may need to be considered further in future assessments that includes a longer data set or if individuals above the age at maturity are included. While the issue of ageing error and whether POPs would have been fully excluded in the current assessment, the panel covered issues related to ageing error in its original report. Future assessments will significantly benefit from reductions in the uncertainty in ages and should be a focus of future work.

### 3. Ageing error and the problems of using a method that conditions on age when the nature and extent of the ageing error is not known

**Mr Cordue's input:** Expressed concerns about the consequences of the ageing error. He was of the opinion that they should have included higher CV in the calculations; and that error may have been biased (e.g. as a result of maturity not being knife-edged as assumed). Suggested that sensitivity tests would have been useful in understanding the consequences.

**Panel conclusion:** The panels original review report contains extensive comment on the issue of ageing error, and we have nothing further to add to this topic.

4. **Technical errors that were made in the assessment and on which the panel made no comment:**

a. **Neglecting to include nibblings pairs**

**Mr Cordue's input:** Suggests ageing uncertainty means that nibblings should be included.

**Panel conclusion:** CSIRO have indicated that if required future assessments would include these possible outcomes. The panel also considered that the low likelihood of repeat mating means that this issue is of limited importance to the assessment. Future assessments should include nibblings where they are a possibility, but this would have had minimal effect on the current assessment.

b. **Incorrect calculation of the probabilities of MHSPs and PHSPs (at least according to the documentation)**

**Mr Cordue's input:** Suggested that need to remove p(same father) for MHSP; worried about the inclusion of the nuisance parameters.

**Panel conclusion:** The panel did not consider this to be of importance to the assessment.

c. **Incorrect calculation of the likelihood**

**Mr Cordue's input:** Suggested that the model should use a likelihood based on Multinomial rather than Binomial trials as per his original review of the assessment. A short paper which outline his concerns is provided in Appendix A.

**Panel conclusion:** The panel concluded that there were substantive errors in Mr Cordue's document that render it as non-issue for the assessment. Our reasoning is below:

**Mr Cordue:** *Intuitively it makes no sense that a categorical trial can be split into an equivalent set of independent Bernoulli trials.*

**Panel:** This is absolutely not true. The standard procedure to generate a random multinomial (categorical) vector is to loop through a sequence of Bernoulli (or Binomial for  $N > 1$ ) draws. The procedure is outlined in various Numerical Recipes textbooks.

**Mr Cordue:** *Simple examples show that this cannot generally be done. For example, consider the roll of a fair dice. In a single roll there are six possible outcomes and each outcome has a probability of  $1/6$ . Suppose when we roll our dice that we get a 1. The likelihood function is simply  $1/6$ . If we follow the approach of Dr. Bravington, where we loop through the alternative outcomes treating each as an independent Bernoulli trial, the likelihood function is very different from  $1/6$ . We first consider whether we have a 1 or not. As we do, this gives the first contribution to the likelihood of  $1/6$ . We then move on to consider whether we have a 2 or not. We don't so the next contribution to the likelihood is  $5/6$ . Repeating the process for 3, 4, 5, and 6 we get a total likelihood of  $1/6 \times (5/6)$ .*

*Panel:* This is not correct. The quantity  $p(1-p)$  is the Bernoulli variance not the likelihood. The real Bernoulli probability is:

$$p(x | q) = q^x (1-q)^{(1-x)}$$

where  $p$  is the true probability (e.g., of face=1, which is  $1/6$ ), and  $x=1$  if success (face=1) and  $x=0$  if failure (face  $\neq 1$ ). So, according to his example, we rolled a fair die and got a 1. So, the string of categorical data (or independent Bernoulli's) are 1,0,0,0,0,0 for faces 1, 2,...6. This gives a likelihood of:

$$p(x=1 | q) = 1/6$$

not  $(1/6)^*(5/6)$  because looping over all possible faces we would have:

$$p(x=1 | q) = (1/6)^1 (1/6)^0 (1/6)^0 (1/6)^0 (1/6)^0 (1/6)^0 = 1/6$$

Writing out the probabilities shows that the categorical likelihood is equivalent to a sequence of Bernoulli's and hence the CKMR methods used are appropriate.

#### 5. The use of a CK model that does not need to condition on age (thus avoiding the problems with ageing error)

*Mr Cordue's input:* Suggested that the ageing error requires that the assessment not be conditioned on age; he developed a different approach and got a result that was 2x the level of mature school sharks.

*Panel conclusion:* The issue of conditional probability was dealt with in the panel's original review of the assessment. The report concluded that conditional probability was the appropriate approach for the assessment. The alternative approach proposed by Mr Cordue (i.e. not conditioned on age) requires an equilibrium model, something that panel did not consider appropriate for this species. That report also identified the issues of ageing uncertainty and the need for it to be reduced in future assessments.

## Conclusions

The input of Patrick Cordue into the panel's review for the school shark CKMR assessment added to the robustness of the exercise. A number of the issues raised had already been dealt with in the panels original review report, and so were not considered further. Some issues may be of consequence to future assessments (e.g. FSPs vs POPs, niblings) and should be considered when a new assessment is undertaken. And some issues were considered to not be of relevance or to contain errors, and so were not dealt with further.

A number of Mr Cordue's concerns would be dealt with by reducing ageing uncertainty. As such the panel wishes to reiterate that further work on ageing of individuals, and especially in reducing uncertainty about ages, would be important for future assessments. The panel also recommends that a planning session precede the next assessment to identify the uncertainties to be addressed, the methods for doing that, the data to be collected and criteria for adopting the results.

## Likelihood function for close kin data

Patrick Cordue

3 April 2021

In my review of the CKMR assessment of school shark I noted that an incorrect likelihood function had been used and its use probably biased the estimates low by a factor of 2 or more. I noted that the CK data consisted of a series of multinomial trials but the likelihood function used corresponded to a series of Bernoulli trials. In Dr. Bravington's response to my review report he wrote that his use of the likelihood for Bernoulli trials would make no difference to the results. He did not provide any evidence for his claim.

It is better to use the terminology categorical distribution in this discussion rather than a "single multinomial trial". There is a subtle distinction as the multinomial likelihood refers to the vector of successes in each category but it is more natural to consider the likelihood for a series of independent categorical random variables. The log likelihoods only differ by a constant.

Intuitively it makes no sense that a categorical trial can be split into an equivalent set of independent Bernoulli trials. Simple examples show that this cannot generally be done. For example, consider the roll of a fair dice. In a single roll there are six possible outcomes and each outcome has a probability of  $1/6$ . Suppose when we roll our dice that we get a 1. The likelihood function is simply  $1/6$ . If we follow the approach of Dr. Bravington, where we loop through the alternative outcomes treating each as an independent Bernoulli trial, the likelihood function is very different from  $1/6$ . We first consider whether we have a 1 or not. As we do, this gives the first contribution to the likelihood of  $1/6$ . We then move on to consider whether we have a 2 or not. We don't so the next contribution to the likelihood is  $5/6$ . Repeating the process for 3, 4, 5, and 6 we get a total likelihood of  $1/6 \times (5/6)^5$ .

The CK situation is different from rolling a fair dice as for any pair of randomly selected sharks (be they conditioned on age/sex or not) the probability of obtaining any of the CK relationships (other than no-relationship) is close to 0. As a consequence,  $1 - p_{rel}$  is close to 1 (where  $p_{rel}$  is the probability that the random pair of sharks have a relationship of "rel". It may be that with the large number of pairs involved and the very low probabilities of CK relationships that the multinomial likelihood and the composite Bernoulli likelihood are almost equivalent.

Let us consider a simple example before we look at a more general case.

Suppose the CK relationships are numbered from 1 to 4 and that no-relationship is denoted by 0. Denote the corresponding probabilities as  $p_i$  for  $i=0...4$ . On a single trial (i.e., for a single pair of sharks) assume that we obtained no-relationship (i.e.,  $x=0$ ).

The categorical likelihood is simply  $p_0 = 1 - (p_1+p_2+p_3+p_4)$ .

The composite Bernoulli likelihood is derived by looping through the outcomes 1 to 4 and considering whether there was a "success" or a "failure" given the observed value of 0. In each case there was a failure, so the likelihood is  $(1 - p_1)(1 - p_2)(1 - p_3)(1 - p_4)$ .

The two likelihoods look very different as functions of the probabilities. However, they are approximately equal because they are identical up to first order terms of the  $p_i$  (and the  $p_i$  are very small).

To move to a general case, denote the number of successes for each close kin relationship as  $s_i$  and let  $n = \sum(s_i)$ . Use  $s_0$  for the number of failures.

The categorical likelihood is:

$$\prod_{i=0}^4 p_i^{s_i} = \left( \prod_{i=1}^4 p_i^{s_i} \right) \left( 1 - \sum_{i=1}^4 p_i \right)^{n - \sum_{i=1}^4 s_i}$$

The composite Bernoulli likelihood is:

$$\left( \prod_{i=1}^4 p_i^{s_i} \right) \left( \prod_{i=1}^4 (1 - p_i)^{n - s_i} \right)$$

The ratio of the composite Bernoulli to categorical is:

$$\frac{\prod_{i=1}^4 (1 - p_i)^{n - s_i}}{\left( 1 - \sum_{i=1}^4 p_i \right)^{n - \sum_{i=1}^4 s_i}}$$

This is obviously not a constant but it is a difficult expression to interpret. However, we can use some approximations to get an idea of what sort of scale the ratio is likely to take and how it might vary over the parameter space.

The ratio of the composite Bernoulli to categorical is approximately:

$$\exp \left( - \sum_{i=1}^4 p_i \sum_{j \neq i} s_j \right) \approx 1 - \sum_{i=1}^4 p_i \sum_{j \neq i} s_j$$

which was obtained using the exponential approximation:

$$(1 + b)^a \approx \exp(ab)$$

which applies for  $b < 1$  and  $ab$  large (and also the approximation  $\exp(x) \approx 1 + x$  for  $x \ll 1$ ).

For school shark the ratio is probably close to 1 but needs to be checked as it is not clear how large the probabilities are. The sums of successes are between 40 and 80.

If the probabilities are  $\sim 10^{-5}$  then there is no concern. If the probabilities are  $\sim 10^{-4}$  then it may be okay or it may not.

The big problem arises when there is a large sampling effort for a depleted population. The probabilities become larger and the number of successes could be relatively large.

For example if the sum of successes are  $\sim 1000$  each and the probabilities are  $\sim 10^{-4}$  then the ratio will be about 0.6 (and it will of course vary across the likelihood space and therefore the incorrect likelihood may deliver estimates which are very different from the maximum likelihood estimates).

The approximate formula is very close to the exact values when the probabilities are very low ( $\sim 10^{-5}$ ) but it does start to break down as they increase ( $\sim 10^{-4}$ ). It turns out that the ratio can be smaller or greater than 1.

In any case, it makes no sense to calculate the wrong likelihood and hope that the error isn't too large when it is straightforward to calculate the correct likelihood. This is not a situation where there can be any debate about the nature of the data (categorical) and the correct likelihood to use for maximum likelihood estimation. If there is some reason to use an *ad hoc* estimator then the reasons need to be stated and a justification supplied.