Question Q2: Summary document

This document combines pertinent comments and responses contained in various documents submitted to the Small pelagic Working Group during the course of 2020 that are associated with Question Q2.

QUESTION Q2: It has been asserted (see FISHERIES/2020/SEP/SWG-PEL/96rev) that the marked (and apparently relatively precisely estimated) change in the estimated survival rate at Robben (but not Dassen) island from the Kaplan-Meier estimates of chick survival rates after some 50 days exposure needs to be better understood before the associated results could be used with confidence as inputs to estimators of island closure effects – see the reasons given in FISHERIES/2020/AUG/SWG-PEL/82 (bottom of page 32) and FISHERIES/2020/AUG/SWG-PEL/84 (last paragraph on page 2). Is this assertion justified?

Overview of material included under 4 steps: 1 = assertion, 2 = response, 3 = response to response, 4 = further responses. Note interpretation of document purpose below is that of Janet Coetzee. Author DSB = Doug Butterworth, RS = Richard Sherley, MOB = Mike Bergh.

	Step	Author	Original source document and main reference documents	Pg
Exchange 1: Butterworth vs Sherley	1	DSB	**FISHERIES/2020/AUG/SWG-PEL/82 (**Asserts that certain aspects of the chick survival results reported in FISHERIES/2020/JUL/SWG-PEL/53REV require further investigation)	2
	2	RS	**FISHERIES/2020/SEP/SWG-PEL/85** (Response to several technical issues raised in respect of chick survival estimates in FISHERIES/2020/AUG/SWG-PEL/82)	4
	3	DSB	**FISHERIES/2020/SEP/SWG-PEL/96rev** (In response to FISHERIES/2020/SEP/SWG-PEL/85 notes opposite effects for chick survival and fledging success for Dassen island) **FISHERIES/2020/OCT/SWG-PEL/103** (notes unexpected features of the survival data analyses) **FISHERIES/2020/OCT/SWG-PEL/110** (Response to FISHERIES/2020/OCT/SWG-PEL/102, additionally notes the unjustified assumption of a common value for the impact of a closure effect in the baseline estimation using chick survival data for Dassen and Robben Islands **FISHERIES/2020/OCT/SWG-PEL/111** (Response to FISHERIES/2020/OCT/SWG-PEL/105rev, re-iterates points made above in respect of chick survival)	6
	4	RS	Provides correlation results to negate concerns about negative correlation effects between chick survival and fledging success are unfounded **FISHERIES/2020/SEP/SWG-PEL/87** (results using island specific effects for chick survival were basically unchanged) **FISHERIES/2020/NOV/SWG-PEL/117rev** (the inference about the effect on chick survival remains unchanged when chicks that were not monitored from hatching were excluded)	9
	5	DSB	In commenting on PEL/117REV asserts that the K-M results remain essentially unchanged, so excluding chicks not monitored from hatching has not resolved the issue. Re-iterates that more time is needed for further analyses of this (extended) dataset, before reliable conclusions about an island closure effect might be drawn from these data.	12
	6	RS	Provides biological reasons for why the K-M curves might differ at different colonies.	14
Exchange 2: Bergh vs Sherley	1	MOB	**FISHERIES/2020/AUG/SWG-PEL/84** (Comments on FISHERIES/2020/JUL/SWG-PEL/53REV and asserts that the estimates of chick survival are potentially biased.	17
	2	RS	**FISHERIES/2020/SEP/SWG-PEL/87** (Response to FISHERIES/2020/AUG/SWG-PEL/84) **FISHERIES/2020/SEP/SWG-PEL/85** (Response to FISHERIES/2020/AUG/SWG-PEL/82, suggests that any systematic bias is likely small)	18
	3	MOB	**FISHERIES/2020/OCT/SWG-PEL/106** (K-M curves exhibits evidence of substantial dependence of the survival rate on the time since observations per chick were initiated) **FISHERIES/2020/OCT/SWG-PEL/107** (Response to FISHERIES/2020/SEP/SWG-PEL/87, suggests a possible bias in chick survival estimates) **FISHERIES/2020/OCT/SWG-PEL/113** (Response to FISHERIES/2020/OCT/SWG-PEL/105rev, a proposal based on the results in FISHERIES/2020/JUL/SWG-PEL/53REV, questions raised about the chick survival analyses have not yet been answered)	19
	4	RS	**FISHERIES/2020/NOV/SWG-PEL/117rev** (the inference about the effect on chick survival remains unchanged when chicks that were not monitored from hatching were excluded)	21
	5	MOB	The robustness of the estimation method applied needs to be checked. More time is required to carry out further analyses on the chick survival data, once provided.	22
	6	RS	Notes agreement of addressing further analyses and provision of data after the review, if necessary.	23

EXCHANGE 1: Butterworth/Sherley

Step 1 – Butterworth assertion: Extracts (in red) from:

Document FISHERIES/2020/AUG/SWG-PEL/82 "A response to Sherley: FISHERIES/2020/JUL/SWG-PEL/53REV" by D. S. Butterworth

Note: For readers' ease, responses have, in the main, been inserted at appropriate points in the original document below in red, and in *italics* in the main text.

Re: Chick survival (from Appendix 4)

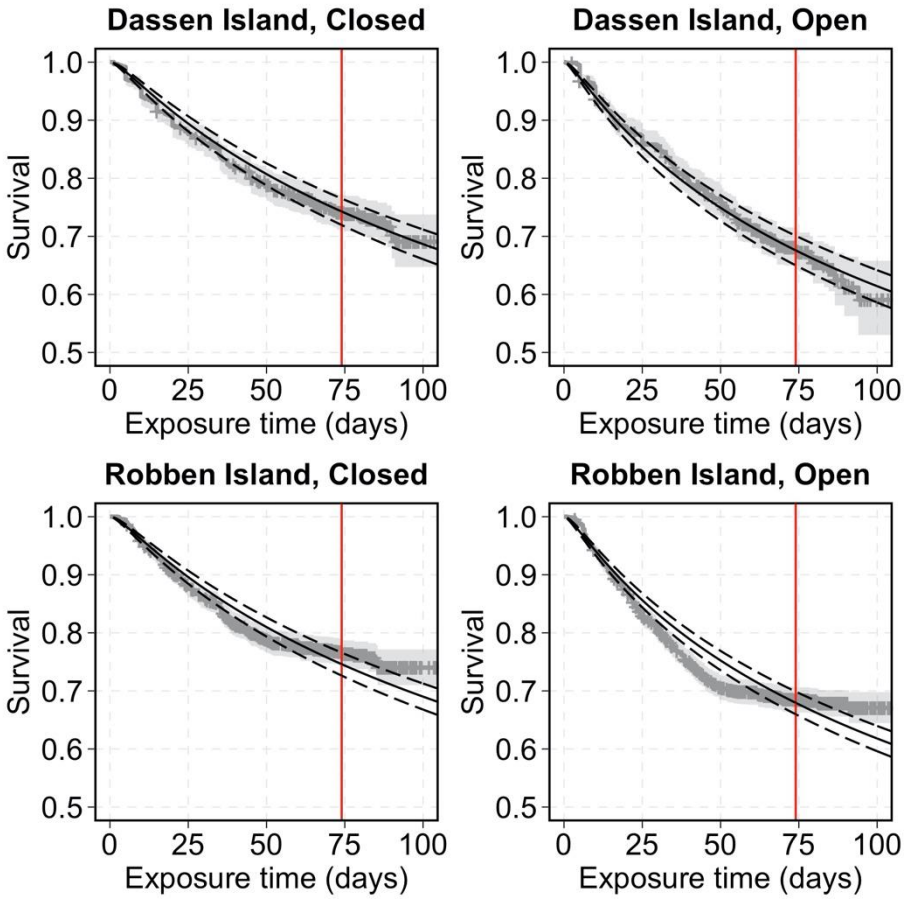

Figure A4.13. Model validation plots for Chick Survival at Dassen Island (top) and Robben Island (bottom) during years that were Closed (left) and Open (right) to fishing. Panels show the comparison of the non-parametric Kaplan-Meier (KM) estimate of survival (grey points, +) and its 95% confidence intervals (grey polygons) and the predicted survival rates (solid black curves) and 95% credible intervals (black dashed curves) based on a model with a log-normal hazard function and no shared frailty term. The vertical red line marks time = 74 days, the age at which the predicted chick survival is compared between islands and closure statuses in the results section of this document and elsewhere (Sherley et al. 2013, 2015, 2018, 2019). Crucially, the predictions from the log-normal model and the KM estimate (which is derived only from the observations) are not credibly different at 74 days, which indicates adequate model fit to predict chick survival at time = 74 days.

The use of an exponential model for survival rather than the "log-normal" model would be simpler and more readily interpreted, and seems attractive given indications (if I am understanding correctly from Sherley's

comments during the 30 July meeting) that the differences in results of interest are not large. The particular reason for this is that then the non-equivalence of exposure time and chick age (because of variable commencement of the age at which different chicks are first recorded) does not potentially confound results. But then the marked (and apparently relatively precisely estimated) change in the estimated survival rate at for Robben (but not Dassen) from the KM estimates after some 50 days exposure becomes a concern. To what extent then might these estimates of cumulative survival be confounded by different distributions of the chick age at which this monitoring commences? Some restrictions on the data used for these analyses, for example through elimination of data for chick for which monitoring is known to have started only at a fairly late stage, might be desirable. However, the matter should first be discussed to check whether some prior further diagnostic investigations might provide insight, before perhaps embarking on further onerous data extractions.

Step 2 – Sherley responses: Extracts (in blue) from:

Document FISHERIES/2020/SEP/SWG-PEL/85 "A response to Butterworth: FISHERIES/2020/AUG/SWG-PEL/82" by Richard B. Sherley

HEADLINE: A model assuming mortality was constant (independent of age) was already presented in 53REV.

Extract from R18, page 18 of FISHERIES/2020/SEP/SWG-PEL/85: "Results comparing the log-normal and exponential hazard functions were given already in an appendix of FISHERIES/2020/JUL/SWG-PEL/53REV. It makes no appreciable difference to the inference whether a log-normal or exponential hazard function is used".

HEADLINE: Model selection was used to compare the most common hazard functions used in birds.

Extract from R18, page 18 of FISHERIES/2020/SEP/SWG-PEL/85: "preliminary analysis using the flexsurvreg function from the 'flexsurv' library in R (Jackson 2016) that used a frequentist approach with null models and no frailty terms to compare exponential, Weibull, Gompertz and log-normal hazard functions supported the use of log-normal hazards (ΔAICc = 49.1 over the next best model)".

Extract from Appendix 2, page 15 of FISHERIES/2020/JUL/SWG-PEL/53REV: "The log-normal model gives a more parsimonious fit to the data based on model selection".

HEADLINE: A residual analysis indicated that if there was any systematic bias in the dataset, it was diminishing, not driving the closures effect.

Extract from R18, page 18 of FISHERIES/2020/SEP/SWG-PEL/85: "the mean error between the KM model and the [log-normal] LN model is less than 1.5% … the LN over-estimates survival more on average and at 74 days for the Open years than for the Closed years at Robben Island. This means that the meaningful closure effect detected at Robben Island is in spite of, not caused by, any bias that might exist in the dataset from the different chick ages at which this monitoring commences".

References

Jackson C. 2016. flexsurv: A Platform for Parametric Survival Modeling in R. Journal of Statistical Software 70: 1–33.

Sherley RB, Barham BJ, Barham PJ, Campbell KJ, Crawford RJM, de Blocq A, Grigg J, Le Guen C, Hagen C, Ludynia K, Makhado AB, McInnes A, Meyer A, Morris T, Pichegru L, Steinfurth A, Upfold L, van Onselen M, Visagie J, Weller F and Winker H. 2019. A Bayesian approach to understand the overall effect of purse-seine fishing closures around African penguin colonies. Department of Environment, Forestry and Fisheries Report: MARAM/IWS/2019/PENG/P4. pp. 1–25.

Sherley RB, Barham BJ, Barham PJ, Campbell KJ, Crawford RJM, Grigg J, Horswill C, McInnes A, Morris TL, Pichegru L, Steinfurth A, Weller F, Winker H and Votier SC. 2018. Bayesian inference reveals positive but subtle effects of experimental fishery closures on marine predator demographics. *Proceedings of the Royal Society B: Biological Sciences* 285: 20172443.

Sherley RB, Winker H, Altwegg R, van der Lingen CD, Votier SC and Crawford RJM 2015. Bottom-up effects of a no-take zone on endangered penguin demographics. *Biology Letters* 11: 20150237.

Sherley RB, Underhill LG, Barham BJ, Barham PJ, Coetzee JC, Crawford RJM, Dyer BM, Leshoro TM and Upfold L. 2013. Influence of local and regional prey availability on breeding performance of African penguins *Spheniscus demersus. Marine*

Step 3 – Butterworth response to responses (extracts from previous documents)

Note: On occasions, simply to provide the necessary context, extracts need to include more material than pertinent to Q1 in isolation. In such instance, the text concerned is shown in yellow highlight.

Document FISHERIES/2020/JAN/SWG-PEL/96rev

Summary comments on analyses of the island closure experiment

D.S. Butterworth

Possible further steps needed in moving towards management recommendations

3) Response variables to be considered

In principle, chick survival is a valuable addition to the set of response variables for which data are available, as associated changes link directly to penguin population dynamics, and hence to estimates of related changes in population growth rates. However, the marked (and apparently relatively precisely estimated) change in the estimated survival rate at Robben (but not Dassen) island from the KM estimates of these rates after some 50 days exposure is a concern. A further concern is that although chick survival is a component of fledging success, for Dassen island these two response variables offer estimates of δ which have different signs and are (in simple terms) near statistically significantly different at the 10% level. Questions also arise about the reliability of the foraging-related variables given that the signs of point estimates of δ to which they give rise differ for the Western and the Eastern Cape islands concerned (see Figure 1). These are matters that require further discussion, as the reasons for the features above need to be better understood before the associated results can be used with confidence.

Document FISHERIES/2020/AUG/SWG-PEL/103

A proposal for a basis to consider future island closures, taking account especially of the current results from the island closure experiment

D. S. Butterworth

On the biological basis for a proposal

In principle, given the agreed decision criterion offered earlier by the Panel, the basis to interpret the results from the application of the 2016 Panel algorithm to the island closure experiment results in terms of whether they demonstrate a biologically meaningful effect of closure on the penguin population growth rate should be straightforward: in simple terms, check for point estimates of the closure impact parameter δ amongst the response variables (only those which are able to be linked directly to penguin demographics) that are less than -0.1.

Only two variables (see also Figure 1) meet this criterion – chick survival and fledging success (though these are available for Robben and Dassen islands only). Dassen meets the criterion for chick survival and Robben for fledging success (and is close to doing so for chick survival). However, as discussed in FISHERIES/2020/JAN/SWG- PEL/96rev, this inference is confounded by two other aspects related to these data. The first is a reverse result for δ for fledging success at Dassen island which is (in simple terms) in the opposite direction and is nearly statistically significantly different at the 10% level from that for the chick survival. The second concerns certain unexpected features of the

survival data analyses which indicate a need for these first to be more closely examined before they might be used with confidence as a basis from which to draw such inferences. Specifically, these relate to the substantial increase (which is indicated to be relatively precisely estimated) in the estimated survival at for Robben island (but not Dassen island) by the Kaplan-Meier method after some 50 days of exposure (FISHERIES/2020/JUL/SWG- PEL/53REV); this aspect of these results first needs to be explained, so as to be clear on whether or not its source results in confounding of estimates of cumulative chick survival.

Document FISHERIES/2020/AUG/SWG-PEL/110

A Response to FISHERIES/2020/JUL/SWG-PEL/102

D. S. Butterworth

Note: For readers' ease, responses have, in the main, been inserted at appropriate points in the original document below in red and in *italics.*

Dassen Island

- ***Both sets of analyses currently show that the closure of Dassen Island to fishing will benefit penguin chick survival.***

While in principle chick survival data have relatively high potential information content for the purposes of the closure experiment, present results need to be considered in the context of a number of caveats:

- *The substantial increase (which is indicated to be relatively precisely estimated) in the estimated survival at for Robben island (but not Dassen island) by the Kaplan-Meier method after some 50 days of exposure (PEL/53REV); this aspect of these results first needs to be explained, so as to be clear on whether or not its source results in confounding estimates of cumulative chick survival. The fact that exposure time does not relate directly to chick age may be contributing to such confounding.*
- *The assumption made in baseline individual data-based estimation using chick survival data for Dassen and Robben Islands of a common value for the impact of a closure effect is unjustified, given general indications of inter-island differences in this effect (see PEL/82 and also PEL/84).*

Document FISHERIES/2020/AUG/SWG-PEL/111

Response to FISHERIES/2020/JUL/SWG-PEL/105rev

D. S. Butterworth

Note: For readers' ease, responses have, in the main, been inserted at appropriate points in the original document below in red and in *italics.*

Island closure results

The Panel's recommendations for further work regarding selection of an appropriate random effects structure and model selection were addressed by Sherley (2020) *[PEL/53REV]* and presented and discussed at a Seabird Scientific Task Team meeting in May 2020. In these updated analyses, a combination of random effects structures were incorporated and the results of the best fitting models demonstrated positive island closure effects on African Penguins from three islands where appreciable purse-seine fishing effort took place during the experiment period: Dassen, Robben and St Croix islands. The results included positive effects of island closures on chick survival, a parameter that has a direct impact on the demographic process, on the only 2 islands where this parameter was measured throughout the experiment period, Robben and Dassen islands.

While in principle results based on chick survival rates are of particular value, they are confounded here first by a reverse result for fledging success at Dassen island which is (in simple terms) in the opposite direction and is nearly statistically significantly different at the 10% level from that for the chick survival. Secondly, there are certain unexpected features of the survival data analyses which indicate a need for these first to be more closely examined before they might be used with confidence as a basis from which to draw such inferences (see PEL/103).

Step 4 – Sherley further responses:

NOTE: Some comments above pertain to Question 1, so responses to those are not repeated here.

Extracts from previous documents in blue:

HEADLINE: The comments that chick survival at Dassen is unreliable because the fledging success results trends in the opposite direction are built on a spurious comparison.

Sherley (2020) [FISHERIES/2020/SEP/SWG-PEL/85]: The chick survival dataset… spans 2008–2015 for Dassen Island. The fledging success dataset… spans 1995–1999 and then 2008–2015 (with a gap from 2000 to 2007) at Dassen Island. First, it is difficult to be confident in directly comparing data from the 1990s with data collected from 2008 onwards in this context because there is strong evidence that the ecosystem, the availability of forage fish resources to fisherman and predators, and penguin population dynamics have changed markedly over this timeframe (e.g. van der Lingen et al. 2005, Roy et al. 2007, Robinson et al. 2013, Crawford et al. 2019). We cannot be sure that the trend in the opposite direction is not a consequence of these differences in the state of the ecosystem.

Second, FISHERIES/2020/JAN/SWG-PEL/09 indicates that the experiment would need to continue for more than 10 years before a biologically meaningful fishing effect is likely to be detected for fledging success at Dassen Island. In other words, the fledging success effect at Dassen Island isn't meaningfully different from zero. On the other hand, the chick survival dataset for Dassen Island already provides evidence of a biologically meaningful fishing effect. Thus, the two do not offer equally strong opposing evidence…

Third, the above ignores the fact that at Robben Island the fledging success dataset which also already enough power in FISHERIES/2020/JAN/SWG-PEL/09 analysis to provide evidence of a biologically meaningful fishing effect is in the same direction as the chick survival effects at both Robben Island and Dassen Island reported in FISHERIES/2020/JAN/SWGPEL/09 and FISHERIES/2020/JUL/SWG-PEL/53REV.

Fourth, and most importantly, if we actually do a like for like… comparison between chick survival and fledging success, we find they are positively correlated with one another…

Robben Island 2001 to 2015: Pearson's product-moment correlation, $r = 0.981$, $t_{13} = 18.37$, $p < 0.001$.

Dassen Island 2008 to 2015: Pearson's product-moment correlation, $r = 0.818$, $t_{6} = 3.48$, $p < 0.013$.

Robben Island 2008 to 2015: Pearson's product-moment correlation, $r = 0.93$, $t_{6} = 6.65$, $p < 0.001$.

Dassen Island 2008 to 2015: Pearson's product-moment correlation, $r = 0.817$, $t_{6} = 3.47$, $p < 0.013$.

Thus, Butterworth's concerns about negative correlation effects are unfounded.

HEADLINE: In still making comment in FISHERIES/2020/OCT/SWG-PEL/110 (presented at the October SWG-PEL meeting) to chick survival results that present a common value for the impact of a closure effect, Butterworth is disregarding that results using island specific effects were already presented (following his and Mike Bergh's request) in FISHERIES/2020/SEP/SWG-PEL/87 (presented at the September SWG-PEL meeting) and the results were basically unchanged.

Sherley (2020) [FISHERIES/2020/SEP/SWG-PEL/87], extract from R12, page 14: "in the interests of full transparency, I have refit M1 in FISHERIES/2020/JUL/SWG-PEL/53REV, but including the Island x Closure Interaction within the fixed effects structure. This model (let's call it M1.I) yields the following results (results for M1 in FISHERIES/2020/JUL/SWG-PEL/53REV in square parentheses; as a reminder the closure main effect (95% HPDI) was 0.38 (0.21–0.55 for M1)):

Robben Island Closure effect (95% HPDI) = 0.37 (0.16–0.57), percentage increase = 9.8% (4.1–15.7%) [9.9% (1.1–18.2%)], percentage of posterior > 0 = 99.9% [97.9%], percentage of posterior >10% = 46.8% [49.0%].

Dassen Island Closure effect (95% HPDI) = 0.41 (0.11–0.70), percentage increase = 11.4% (2.4–20.4%) [10.1% (0.2–19.7%)], percentage of posterior > 0 = 99.7% [97.7%], percentage of posterior >10% = 59.8% [51.0%].

So, whether model M1 is used (Island + Closure) or M1.I is used (Island x Closure), the overall effect size and the inference about whether these effects are biologically meaningful would be unchanged.

HEADLINE: Refitting the Western Cape chick survival model excluding the 8% of chicks not monitored from hatching does not change the inference about the effect of the island closures.

Sherley 2020 [FISHERIES/2020/NOV/SWG-PEL/117rev]: "Of the 3,219 African penguin chicks monitored for survival at Robben Island between 2008 and 2018, 357 (~11%) were not monitored from hatching. And of the 1,673 penguin chicks monitored for survival at Dassen Island over those same years, 31 (~2%) were not monitored from hatching".

"Refitting the Western Cape chick survival model to exclude the ~11% of chicks that were not monitored from hatching… the closure effect increased from a 9.8% improvement in chick survival during Closed years relative to Open years, to an 11.2% improvement at Robben Island (Table 1). The Dassen Island results remain unchanged with an 11.4% improvement in chick survival during Closed years relative to Open years (Table 1)… the probability that the closures to purse-seine fishing around Robben and Dassen Island improved penguin chick survival exceeded 99.5% at both islands. Moreover, the percentage effect size exceeded the 10% pre-agreed threshold for what constitutes a biologically meaningful effect at both islands (Table 1) and the inference about the effect of the island closures experiment on chick survival remains unchanged".

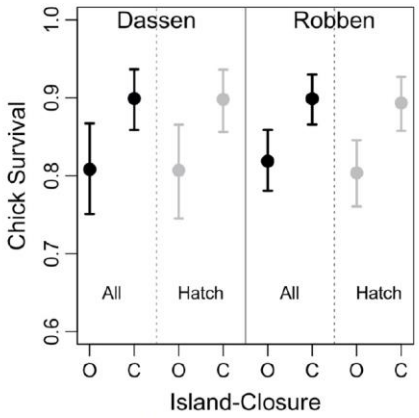

Figure 1: Model estimated mean (± 95% highest posterior density intervals, HPDI) chick survival between 2008 and 2018 at Dassen Island and Robben Island in years when a 20 km radius around each island was open (O) to or closed (C) to purse-seine fishing using either all (n = 4,892) African penguin chicks monitored in that time frame (All, black points and whiskers) or only the 92% of chicks (n = 4,504) monitored from hatching (Hatch, grey points and whiskers).

Table 1: Results from models tested to assess the impact of the fishing closures on African penguin chick survival at Robben and Dassen Islands. M1 presents the results from the best fitting model from the model selection presented in FISHERIES/2020/JUL/SWG-PEL/53REV. M1.I presents the results from an update of M1 to include island-specific closure effects as requested by two participants at the SWG-PEL, and first presented in Sherley (2020c; FISHERIES/2020/SEP/SWG-PEL/87). M1.I.H presents an update of M1.I using only the 92% of chicks (n = 4,504) that were monitored from the egg stage.

Model Number	Random effects structure	Island and Closure fixed effects structure	Robben Closure effect mean (95% HPDI)	Percentage difference (95% HPDI)	Probability of effect	Dassen Closure effect mean (95% HPDI)	Percentage difference (95% HPDI)	Probability of effect
M1	Island/Year/BirdID	Island + Closure	0.38 (0.21–0.55)	10.3% (5.4–15.2%)	100%	0.38 (0.21–0.55)	10.6% (5.2–16.2%)	100%
M1.I	Island/Year/BirdID	Island × Closure	0.37 (0.16–0.57)	9.8% (4.1–15.7%)	99.9%	0.41 (0.11–0.70)	11.4% (2.4–20.4%)	99.7%
M1.I.H	Island/Year/BirdID	Island × Closure	0.39 (0.18–0.62)	11.2% (4.6–18.0%)	100%	0.41 (0.10–0.69)	11.4% (2.2–20.9%)	99.6%

Notes: HPDI = highest posterior density intervals. Probability of effect = the percentage of the closure effect posterior estimates > 0. Model M1 used an additive Island and Closure fixed effects structure, so only one overall closure effect for the two islands is estimated. In this table, the same effect size is given for both islands, but only one estimate is made in the model (the percentage difference for each island individually can be extracted from the model posteriors as a derived parameter).

Step 5 – Butterworth further responses:

Implications of Document FISHERIES/2020/NOV/SWG-PEL/117REV

Document PEL/117REV became available for the first time only for Pelagic Working Group meeting on 16 November, which was after question Q2 (and the material on which comments in Step 1 and Step 3 had to be based) had already been finalised. Its contents have been raised for the first time in comments in Step 4, so this extra Step 5 has become necessary to be able to respond to the associated new material introduced in Step 4.

As is stated in Q2 itself, the primary reason for posing this question arises from the patterns in survival with exposure time evidenced by the Kaplan-Meier (K-M) estimates, including their marked differences for Robben and Dassen island. One suggestion made earlier was that chicks not monitored from immediately following hatching be excluded from the analyses, as the non-equivalence of exposure time and chick age that pertained for some chicks might be the reason underlying these patterns, which in turn could bias estimates of the island closure effect.

PEL/117REV reports on the results that follow from this suggested exclusion. Importantly they show that the **patterns** previously evident in the K-M results **remain essentially unchanged**, so that the suggestion to exclude chicks monitored only from some time after hatching does **not** provide resolution of the original concern arising from the patterns in these K-M estimates.

The updated results have provided stronger indication for changes with age in chick survival, but what is the explanation for these – certainly daily mortality could depend on age, but why in as extreme a form as indicated? At Robben island, apparently the daily chick mortality drops enormously after the chicks reach an age of about 50 days, but at Dassen island there is hardly any change. This is “strange”, and warrants more thorough investigation, in particular as to whether it is a real effect or perhaps rather some anomaly possibly introduced by certain confounding effects in the analyses; these is turn could lead to bias in the estimates of the island closure effect. This pattern itself also suggests that estimation be based on a model parametrised to allow for the possibility of patterns similar to those indicated, which is not the case for analyses to date; this would again be to attempt to avoid potential estimation bias.

Furthermore, and perhaps more so than for other response variables, chick survival may be dependent on time-of-year, as one might expect it to be higher during the peak of the breeding season than when near either the start or the end. This in turn suggests standardisation for time-of-year in the analyses, as it is possible that different sampling patterns from year to year in terms of this time could be biasing results.

These survival data, because of their more direct relationship to penguin demographics, are of particular importance to the process of estimation of the island closure effect, and consequently need especially thorough evaluation before conclusions might be drawn from them. This follows given the consequences for both penguins and the fishing industry that may result from decisions based on those results.

However, as indicated in the Appendix, these data have only quite recently become available for scrutiny and analysis by PWG participants in terms of the associated agreed process. The most recent data as used in PEL/117REV have yet to be made available, and these would be required in fuller form to allow for the standardisation adjustments motivated above. Accordingly, notwithstanding PEL/117REV, the assertion of Q2 remains justified – hence more time is needed for further analyses of this (extended) dataset, once provided to and available through DEFF in terms of agreed processes; this is before reliable conclusions about an island closure effect might be drawn from these data.

Appendix

Data Procedures and Timelines in the PWG for Analyses of the Island Closure Experiment Results and Chick Survival data

- When analyses of island closure data sets for consideration in the PWG commenced almost a decade ago, certain procedures were agreed. These were that any data to be used for such analyses must be available to all PWG participants, with these data to be lodged with (then) DAFF – a publication embargo for a limited period might apply to such data should the data provider so request, in terms of the Code of Conduct signed by all participants in such Working Groups.

- Coetzee has taken responsibility for curating such data submitted, and typically annually provides a "data-document" listing the data so available. Originally only annually aggregated data were requested for such submission, but later some data providers came to provide individual data as well.

- There is no mention of chick survival data sets in these documents up to and including 2017.

- In late 2019 efforts were made to update these data sets as held by what was then DEFF. This led to Coetzee's updated data document FISHERIES/2019/NOV/SWG-PEL/23. For the first time this included reference to chick survival data, though only individual data were available and not annual means. The data provider had advised then that reliable values for annual means were not available at that time.

- A particular reason for updating these data sets at that time was to provide inputs to update the standard approach (based on annually aggregated data) for the estimation of and power analysis for the island closure effect parameter δ. Because annually aggregated values for chick survival were not available, a method was developed to estimate these from the individual data then available (OLSPS: FISHERIES/2020/JAN/SWG-PEL/06). Those results were used to provide the aggregated analyses for chick survival in the update document FISHERIES/2020/JAN/SWG-PEL/09rev.

- Annually aggregated estimates of chick survival were first provided to Coetzee by the data provider at the end of July 2020, and subsequently formally advised to be available in data document FISHERIES/2020/SEP/SWG-PEL/100.

- Suggestions were first made in September (see e.g. FISHERIES/2020/AUG/SWG-PEL/82) that it would be desirable to adjust these data to restrict to chicks for which monitoring commenced at hatching (so that exposure time and age became equal). However, advice given at the September PWG meeting when this was first discussed was that this was not possible in the short term as it would be an onerous data extraction task needing substantial time.

- At the 16 November PWG meeting, FISHERIES/2020/NOV/SWG-PEL/117 was tabled and discussed. This reported on analyses of the chick survival data restricted to records of chicks for which monitoring commenced at hatching. These data have yet to be provided to the DEFF data repository.

Step 6 – Sherley further responses:

There is no evidence or even rationale presented here for why the shape of the mortality function at Robben Island would lead to a systematic bias in the closure effect estimate, when the apparent "anomaly" is present at Robben Island in both 'Open' and 'Closed' years. All that has been offered is a straw man argument that suggest there may be some 'anomaly possibly introduced by certain confounding effects in the analyses', for example suggesting that time-of-year at which nests were monitored could now be biasing results. However, monitoring of penguin nests has been carried out consistently during all years of the closures experiment, as outlined in peer-reviewed papers using these data (e.g. Sherley et al. 2018).

Of course daily mortality depends on age. This is perfectly normal biology, is well documented for African penguins in the peer-reviewed literature (e.g. Seddon and van Heezik 1991) and is why the lognormal represents an improvement over the exponential curve (see Appendix 2 of SWG-PEL/REV53). Both parents provision the chicks, usually in shifts until about 40 days post-hatch after which chicks may be left unguarded as both parents forage to meet their growing energy needs (Cooper 1977). Using a cohort of 104 chicks that all fledged at Robben Island in 2004, Bouwhuis et al. (2007) found that the chicks' daily metabolizable energy requirement peaked at 53 days, but 90% of this value was reached at age 36 days and it remained fairly constant until fledging. The parents cannot leave the chicks alone to both forage when the chicks are less than about 40 days old as they are susceptible to exposure and gull predation. But, after the peak energy requirement at 53 days, both parents can go to sea to forage. Thus, if chicks make it to 53 days, their probability of starvation is much lower from then on. Moreover, penguin chicks usually begin to fledge from about 60 days old, but can take up to 130 days depending on food availability (Cooper 1980; Williams and Cooper 1984; Seddon and van Heezik 1991, Sherley et al. 2013). In the Robben Island dataset we have some that fledge from what we estimate to be about 55 days (note we do not know age exactly because our nest checks are about 5 days apart on average). And these data are censored, so it should be remembered that as chicks fledge and so disappear from the study, they show up in the KM plot as crosses (mortalities) when in fact they are not mortalities in the model. At Dassen, penguin chick growth rates are generally slower in a given year than at Robben Island (Sherley 2010, Sherley 2012b), thus it is not at all strange to see later fledging dates represented in the KM curve for Dassen. That these curves differ for different colonies is not "strange"; all kinds of prevailing conditions differ at different colonies.

Sherley (2010) examined the influence of hatching date on chick survival at Robben Island over nine breeding seasons using a trigonomic model, which found that chick survival peaked in June (see Figure 2.5 A); but that chick survival was within ~10% of the June value between April and August. In other words, for hatching date to be generating the observed closure effect difference at Robben Island (of ~11%), we would have to have only monitored nests outside of the peak breeding season (i.e only at the ends of the breeding season) during 'Closed' years and only during the peak of the breeding season during 'Open' years. In other words, what Doug is suggesting here is that we have falsely reported our methods in the peer-reviewed literature. Moreover, where the best month for chick survival varies from year to year, this is likely a function of differing food availability from year to year (see Sherley et al. 2013). Food availability is accounted for in the models in SWG-PEL/REV53 via sardine and anchovy biomass estimates and additional year to year variation is accounted for by using Year in the random effect structure.

Statements above to the effect that the chick survival data and methods have only recently been made available to the SWG-PEL are incorrect and misleading. As Doug was reminded at the last SWG-PEL meeting, chick survival data (including annual aggregated values) and analyses for Robben Island were first presented to the SWG-PEL in 2012 (Sherley 2012a, see Table 1 extracted below). Disaggregated data for Robben 2001 to

2013 – exactly the data needed to produce the annually aggregated vales in OLSPS (2020; FISHERIES/2020/JAN/SWG-PEL/06) – have been available to anyone in the world on Dryad since June 2016: https://doi.org/10.5061/dryad.t446r. And data for 2008–2015 for both Robben and Dassen have been available on Dryad since December 2017 https://doi.org/10.5061/dryad.d4977. It is just the case that Doug decided to ignore the chick survival data in favour of fledging success until recently (while we are on the subject of the fledging success data, any issue with the chick survival dataset will also apply equally to the fledging success dataset and any analysis done using it, as the two metrics are derived from the same raw data).

Doug's statement about the procedures for data sharing that were agreed "almost a decade ago" is also misleading. It is not correct that "when analyses of island closure data sets for consideration in the PWG commenced almost a decade ago" that "a publication embargo for a limited period might apply to such data should the data provider so request, in terms of the Code of Conduct signed by all participants in such Working Groups". In fact this the Code of Conduct had to be reviewed and the capacity for a publication embargo added later to protect the interests of the penguin biologists passing data to DEFF because Doug and colleagues attempted to publish in the peer-reviewed literature an analyses based on unpublished datasets submitted to the SWG-PEL without the knowledge, input or permission of the data owners. At the time, this seriously risked the chances of the PhD of one data contributor being awarded and resulted in the paper eventually being withdrawn and an internal ethics review process at the University of Cape Town. The only reason for any delay (and I would argue there has been next to none of that) in any of the penguin biologists passing datasets to DEFF has been while we waited for confirmation that the Code of Conduct of the SWG-PEL would, in fact, protect first publication rights.

Table 1. Measures of breeding success, egg survival and chick survival from African Penguin nests monitored on Robben Island from 2001 to 2012.

Year	1H0F	1H1F	2H0F	2H1F	2H2F	1H nests	2H nests	Total nests	Hatched	Fledged	Fledges/ hatchling 1H	Fledges/ hatchling 2H	Fledges/nest	Fledglings/ hatchling	Egg survival	Chick survival	Overall survival	95% CI – Overall survival
2001	3	0	2	1	4	3	7	10	17	9	0	0.64	0.90	0.53	0.23	0.51	0.12	0.06–0.23
2002	7	0	3	1	2	7	6	13	19	5	0	043	0.38	0.26	0.38	0.34	0.13	0.07–0.23
2003	13	6	28	5	7	19	40	59	99	25	0.32	0.24	0.42	0.25	0.66	0.27	0.18	0.12–0.27
2004	21	9	36	4	17	30	57	87	144	47	0.30	0.33	0.54	0.33	0.55	0.30	0.17	0.12–0.24
2005	11	8	7	3	16	19	26	45	71	43	0.42	0.67	0.96	0.61	0.52	0.60	0.31	0.22–0.44
2006	17	5	20	10	15	22	45	67	112	45	0.23	0.44	0.67	0.40	0.50	0.43	0.21	0.15–0.30
2007	4	16	19	6	38	20	63	83	146	98	0.80	0.65	1.18	0.67	0.58	0.68	0.39	0.32–0.49
2008	33	42	46	16	98	75	160	235	395	254	0.56	0.66	1.08	0.64	0.63	0.65	0.41	0.35–0.47
2009	18	28	32	23	40	46	95	141	236	131	0.61	0.54	0.93	0.56	0.50	0.56	0.28	0.23–0.34
2010	22	17	19	18	49	39	86	125	211	133	0.44	0.67	1.06	0.63	0.61	0.60	0.36	0.30–0.44
2011	27	23	30	16	63	50	109	159	268	165	0.46	0.65	1.04	0.62	0.57	0.65	0.37	0.31–0.45
2012	16	30	18	8	51	46	77	123	200	140	0.65	0.71	1.14	0.70	0.55	0.70	0.39	0.32–0.46
Overall*	–	–	–	–	–	–	–	**1147**	**1918**	**962**	**0.43**	**0.65**	**0.95**	**0.58**	**0.55**	**0.57**	**0.31**	**0.29–0.34**

* The overall value is either the total value[1], the median value or (for survival estimates) the overall survival estimate from the model in the absence of Year as an explanatory variable as appropriate to the column. The data in the table exclude Treasure oiled birds which are known to have had significantly lower breeding success. Breeding success (e.g. not survival) estimates are restricted to nests where observations started with chicks, rather than eggs, i.e. this is only measuring CHICK-REARING SUCCESS. Double clutches are included and treated as separate attempts. Egg and chick survival probabilities are derived from all breeding attempts monitored, including those monitored during incubation.

References:

Bouwhuis, S., Visser, G. H., and Underhill, L. G. 2007. Energy budget of African penguin Spheniscus demersus chicks. In Kirkman, S. P., editor, Final Report of the BCLME (Benguela Current Large Marine Ecosystem) Project on Top Predators as Biological Indicators of Ecosystem Change in the BCLME, pages 125–127. Avian Demography Unit, Cape Town.

Cooper, J. 1977. Energetic requirements for growth of the jackass penguin. Zoologica Africana, 12:207–213.

Cooper, J. 1980. Breeding biology of the jackass penguin with special reference to its conservation. In Johnson, D. N., editor, Proceedings of the 4th Pan-African Ornithological Congress. Southern African Ornithological Society, Johannesburg.

OLSPS Marine. 2020. A simple restructuring of penguin chick survival data to estimate survivorship at Dassen and Robben Islands. FISHERIES/2020/JAN/SWG-PEL/06.

Seddon, P. J. and van Heezik, Y. 1991. Effects of hatching order, sibling asymmetries, and nest site on survival analysis of jackass penguin chicks. Auk, 108:548–555.

Sherley RB. 2010. Factors influencing the demography of Endangered seabirds at Robben Island, South Africa: Implications and approaches for management and conservation. PhD Thesis, University of Bristol, U.K.

Sherley RB. 2012a. Update on the breeding productivity of African Penguins (*Spheniscus demersus*) on Robben Island. FISHERIES/2012/NOV/SWG-PEL/ICTT/06.

Sherley 2012b. Report on the analysis of growth in African Penguin chicks for the Island Closure Feasibility Study in South Africa, 2012. FISHERIES/2012/NOV/SWG-PEL/ICTT/03.

Sherley RB, Underhill LG, Barham BJ, Barham PJ, Coetzee JC, Crawford RJM, Dyer BM, Leshoro TM and Upfold L. 2013. Influence of local and regional prey availability on breeding performance of African penguins *Spheniscus demersus*. Marine Ecology Progress Series 473: 291–301

Sherley RB, Barham BJ, Barham PJ, Campbell KJ, Crawford RJM, Grigg J, Horswill C, McInnes A, Morris TL, Pichegru L, Steinfurth A, Weller F, Winker H and Votier SC. 2018. Bayesian inference reveals positive but subtle effects of experimental fishery closures on marine predator demographics. Proceedings of the Royal Society B: Biological Sciences 285: 20172443.

Williams, A. J. and Cooper, J. 1984. Aspects of the breeding biology of the jackass penguin, *Spheniscus demersus*. In Ledger, J., editor, Proceedings of the 5th Pan-African Ornithological Congress. Southern African Ornithological Society, Johannesburg.

EXCHANGE 2: BERGH/SHERLEY

Step 1 – Bergh assertion: Extracts (in red) from:

<u>Document FISHERIES/2020/AUG/SWG-PEL/84</u> "Comments on Revisiting the key results in MARAM/IWS/2019/PENG/P4 in light of the 2019 Panel recommendations (FISHERIES/2020/JUL/SWG-PEL/53REV) by Richard B. Sherley" by Mike Bergh

"The Kaplan Meier results suggest that at Robben Island chick survivorship is dependent on time, or chick age (see Figure A4.13 of FISHERIES/2020/JUL/SWG-PEL/***53REV***). The potential that this has biased the closure effect estimate because of the selection of chicks at different ages/times for estimating chick survivorship needs to be fully explored. This has not been done in FISHERIES/2020/JUL/SWG-PEL/***53REV***. "

Step 2 – Sherley responses: Extracts (in blue) from:

Document FISHERIES/2020/SEP/SWG-PEL/87 "A reply to Bergh: FISHERIES/2020/AUG/SWG-PEL/84" by Richard B. Sherley

And

Document FISHERIES/2020/SEP/SWG-PEL/85 "A response to Butterworth: FISHERIES/2020/AUG/SWG-PEL/82" by Richard B. Sherley

Headline: A residual analysis indicated that if there was any systematic bias in the dataset, it was diminishing, not driving the closures effect.

Extract from R7, page 7 of FISHERIES/2020/SEP/SWG-PEL/87: "Please see Response 18 in Sherley (2020e), FISHERIES/2020/SEP/SWG-PEL/85. In principle, this could be explored further, but it is far from clear that this is necessary".

Extract from R18, page 18 of FISHERIES/2020/SEP/SWG-PEL/85: "the mean error between the KM model and the [log-normal] LN model is less than 1.5%... meaning that on average, the estimates from the LN are within 1.5% of the KM … And, at Robben Island, where the LN fits the least well of the two islands, the deviation is more positive during the Open years (1.35% mean error and −0.46% at 74 days), than during the Closed years (0.26% mean error and −1.91% at 74 days)… This means that the meaningful closure effect detected at Robben Island is in spite of, not caused by, any bias that might exist in the dataset from the different chick ages at which this monitoring commences".

Step 3 – Bergh responses to responses (Extracts from previous documents)

(Question relating to features in Kaplan Meier plots)

Document FISHERIES/2020/OCT/SWG-PEL/106

Proposals for Experimental Design Decisions and Island Closure Decisions in relation to the island closure experiment.
By
Mike Bergh
6 October 2020

"The above comments notwithstanding, a significant issue has arisen for the chick survival estimates at Dassen Island and Robben Island. These are that the underlying data exhibits evidence of substantial dependence of the survival rate on the time since observations per chick were initiated, as shown in Fig A4.13 of FISHERIES/2020/JUL/SWG-PEL/53REV. Further work is imperative to dispel concerns that this may have substantially biased estimates of the related closure effect and its variance."

Document FISHERIES/2020/OCT/SWG-PEL/107

Comments on FISHERIES/2020/SEP/SWG-PEL/87

by

Mike Bergh
20 September 2020

"FISHERIES/2020/SEP/SWG-PEL/87 responds, using numbering R1 – R14, to the 14 points made in FISHERIES/2020/AUG/SWG-PEL/84 on FISHERIES/2020/JUL/SWG-PEL/53REV. The following further comments on points R1-R14 in FISHERIES/2020/SEP/SWG-PEL/87 use numbering R1-R14 as well (text in italics taken directly out of FISHERIES/2020/SEP/SWG-PEL/87):**"**

"R7. Kaplan Meier, Possible Bias in Chick Survival Estimates. The R7 argument in FISHERIES/2020/SEP/SWG-PEL/87 suggests that the chick survival estimates may be biased. This is a reason to carry out further work to correct for this bias, by, for example and as suggested in FISHERIES/2020/SEP/SWG-PEL/87, excluding from analyses chicks that were not monitored from their hatching date."

Document FISHERIES/2020/OCT/SWG-PEL/113

Comments on FISHERIES/2020/SEP/SWG-PEL/105REV, "Recommendations for island closures around African Penguin colonies".

by

Mike Bergh
19 October 2020

There have also been important technical questions raised about the chick survival analyses which have not yet been answered, viz. there is apparently some time dependence of survival which requires urgent clarification and investigation, since the implication may be that there are biases in the chick survival estimates. These technical questions have implications for the results in both **FISHERIES/2020/JUL/SWG-PEL/53REV** and **FISHERIES/2020/SEP/SWG-PEL/97REV.**

Step 4 – Responses: Extracts from previous documents in blue:

HEADLINE: Refitting the Western Cape chick survival model excluding the 8% of chicks not monitored from hatching does not change the inference about the effect of the island closures.

Sherley 2020 [FISHERIES/2020/NOV/SWG-PEL/117rev]: "Of the 3,219 African penguin chicks monitored for survival at Robben Island between 2008 and 2018, 357 (~11%) were not monitored from hatching. And of the 1,673 penguin chicks monitored for survival at Dassen Island over those same years, 31 (~2%) were not monitored from hatching".

"Refitting the Western Cape chick survival model to exclude the ~11% of chicks that were not monitored from hatching… the closure effect increased from a 9.8% improvement in chick survival during Closed years relative to Open years, to an 11.2% improvement at Robben Island (Table 1). The Dassen Island results remain unchanged with an 11.4% improvement in chick survival during Closed years relative to Open years (Table 1)… the probability that the closures to purse-seine fishing around Robben and Dassen Island improved penguin chick survival exceeded 99.5% at both islands. Moreover, the percentage effect size exceeded the 10% pre-agreed threshold for what constitutes a biologically meaningful effect at both islands (Table 1) and the inference about the effect of the island closures experiment on chick survival remains unchanged".

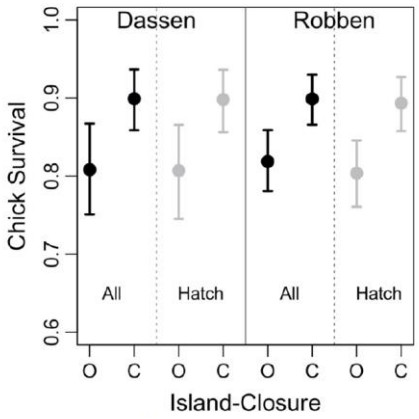

Figure 1: Model estimated mean (± 95% highest posterior density intervals, HPDI) chick survival between 2008 and 2018 at Dassen Island and Robben Island in years when a 20 km radius around each island was open (O) to or closed (C) to purse-seine fishing using either all (n = 4,892) African penguin chicks monitored in that time frame (All, black points and whiskers) or only the 92% of chicks (n = 4,504) monitored from hatching (Hatch, grey points and whiskers).

Table 1: Results from models tested to assess the impact of the fishing closures on African penguin chick survival at Robben and Dassen Islands. M1 presents the results from the best fitting model from the model selection presented in FISHERIES/2020/JUL/SWG-PEL/53REV. M1.I presents the results from an update of M1 to include island-specific closure effects as requested by two participants at the SWG-PEL, and first presented in Sherley (2020c; FISHERIES/2020/SEP/SWG-PEL/87). M1.I.H presents an update of M1.I using only the 92% of chicks (n = 4,504) that were monitored from the egg stage.

Model Number	Random effects structure	Island and Closure fixed effects structure	Robben Closure effect mean (95% HPDI)	Percentage difference (95% HPDI)	Probability of effect	Dassen Closure effect mean (95% HPDI)	Percentage difference (95% HPDI)	Probability of effect
M1	Island/Year/BirdID	Island + Closure	0.38 (0.21–0.55)	10.3% (5.4–15.2%)	100%	0.38 (0.21–0.55)	10.6% (5.2–16.2%)	100%
M1.I	Island/Year/BirdID	Island × Closure	0.37 (0.16–0.57)	9.8% (4.1–15.7%)	99.9%	0.41 (0.11–0.70)	11.4% (2.4–20.4%)	99.7%
M1.I.H	Island/Year/BirdID	Island × Closure	0.39 (0.18–0.62)	11.2% (4.6–18.0%)	100%	0.41 (0.10–0.69)	11.4% (2.2–20.9%)	99.6%

Notes: HPDI = highest posterior density intervals. Probability of effect = the percentage of the closure effect posterior estimates > 0. Model M1 used an additive Island and Closure fixed effects structure, so only one overall closure effect for the two islands is estimated. In this table, the same effect size is given for both islands, but only one estimate is made in the model (the percentage difference for each island individually can be extracted from the model posteriors as a derived parameter).

Step 5 – Bergh further responses:

I am in agreement with the Step 5 commentary by Butterworth on Q2 with respect to issues regarding the Kaplan-Meier estimate graphs. As regards the Appendix to those comments, much of the history of the availability of the chick survival date **prior to 2019** is not in my personal knowledge and experience. However, having consulted with persons who were closely involved with this I am satisfied that it is an accurate record of past events regarding the provision of chick survival data.

This Appendix (in Step 5 of Q2, by Butterworth) is correct that the individual chick survival data were first made available in terms of the agreed data protocol for the island closure experiment in late 2019, and that the aggregated data were first made available in mid-2020.

The data used in PEL/117REV have not yet been made available to all participants in the PWG, either via the agreed data protocol, or in any other way. As a result, it has not been possible for this author (or any other) to potentially replicate the calculations described in PEL/117REV and to check the robustness of the estimation method applied. This would require provision of a dataset that omits from the survival dataset those chicks that had not been observed as of their hatching date.

It is therefore clear, as stated in Step 5 of Q2 by Butterworth, that more time is required to carry out further analyses on the chick survival data, which could commence when those revised data are made available. Such analyses are likely to require access to more than just the dataset underpinning PEL/117REV – as just one example, it may well be necessary to include date of hatching as a variable in these analyses.

Step 6 – Sherley further responses:

As Mike admits that the history of the availability of the chick survival data is not in his personal knowledge, perhaps he should have refrained from comment here. The panel are directed to my response to Doug on this point.

The dataset underpinning PEL/117REV were clearly offered to the SWG-PEL at the last meeting I attended. They haven't yet been provided – and neither have the raw data underpinning them – because we agreed by email to first see whether the panel felt further analysis was necessary or not. Thus it surprises me to see above what essentially amounts to a suggestion that I have not acted with full transparency. I hope I have misunderstood the sentiment.

Below are extracts from our email conversation:

From Richard Sherley to Mike Bergh (copied Janet Coetzee and Doug Butterworth), sent 16 November 2020 at 16:45 (UK time):

"I think we should simply stop now and wait for the panel input on whether this is an issue or not. I am not prepared to spend more time going down rabbit hole after rabbit hole, going back to the data and providing more and more when you can't actually explain how your suggesitons [sic] would lead to this supposed bias. It isn't enough just to keep saying "we need to account for x because it could be causing a bias" because we could go on down that road *ad infinitum*.

So let's just put this to the panel as intended and if they think knowing the date of hatching will make a difference, then we can revisit this

Cheers,

Richard".

From Mike Bergh to Richard Sherley (copied Janet Coetzee and Doug Butterworth), received 16 November 2020 at 17:56 (UK time):

"Hi Richard

I agree with the suggestion in your last para. Let the panel decide whether the Kaplan Meier graphs raise questions that your analysis has not cleared up. If further data and analyses are necessary this can be addressed after the review.

Rgds

Mike"