Shortcomings in data access Procedure B.

WILLIAM K. de la MARE
Australian Antarctic Division, Channel Highway, Kingston, Tasmania. Australia, 7050.
Contact email: bill.delamare@aad.gov.au

ABSTRACT

In the light of experience in requesting access to data under Procedure B, it is proposed that the Procedure is revised to distinguish between cases where first use is being made of data and those where analyses of data have already been published. In the case of published analyses, the procedure should enable independent analysis and verification. The application of Procedure B for the conduct of periodic reviews of Special Permit research also needs to ensure that data are made available well in advance of the reviews and without requirements that might inhibit independent analysis and verification.

KEYWORDS: DATA AVAILABILITY, SPECIAL PERMIT REVIEWS, STATISTICAL ANALYSIS, INDEPENDENT VERIFICATION

INTRODUCTION

Access to data collected under Special Permit is subject to a protocol specified in IWC (2004) Annex T. There are two procedures under the protocol:

> *Procedure A is the process for obtaining access to data for analyses that are needed to provide the best management advice on catch limits (e.g. the RMP and AWMP).*
>
> *Procedure B is the process for obtaining access to data for analyses the Committee believes would be valuable in providing other advice to the Commission.*

Last year I requested, under Procedure B, access to data on body condition collected by the Japanese Whale Research Program under Special Permit in the Antarctic (JARPA). This request followed on from simulation studies (presented in de la Mare, 2011) that demonstrated that methods used in analyses of these data published in the open scientific literature ((Konishi *et al*., 2008 in Polar Biology) were statistically inappropriate. Subsequently, the Scientific Committee agreed that:

> *... that further analysis of the data was warranted to determine: (i) whether the models fitted so far captured all the main features of the data; and (ii) whether the estimate of trend (whose confidence limits using the best fitting model ranged from near zero to values that could be of appreciable biological significance) could be made more precise. Inter alia, revised analyses should consider the two sexes separately and consider latitudinal band as a random effect. The Committee recommends that further analyses are presented next year. To facilitate this it suggests that the authors of SC/63/O16 and of Konishi et al. as appropriate apply for access to the data under Procedure B of the Data Availability Agreement; it requests the data holders to consider such requests favourably.*

I submitted a request (given in Attachment A) under Procedure B towards the end of the Scientific Committee Meeting as hardcopy by hand and by email on June 11, 2011 to both the Chair of the Data Access Group (DAG) and to Takashi Hakamada of the Institute of Cetacean Research (ICR). After the exchange of some correspondence, the DAG approved my proposal on August 12, 2011 and forwarded it to the ICR. On September 29, 2011 I received from the DAG Chair the response from the ICR (dated August 26, 2011, Attachment B).

I decided I could not comply with all the conditions specified by the ICR. The ICR stipulations concerning co-authorship were more restrictive than Protocol B. They would have required me to include as co-authors several of the authors of the original paper and to negotiate and agree the substance and presentation of my analyses. This would not be consistent with independent further analyses. No process was indicated by the ICR as to what would happen if the ICR nominated co-authors and I did not agree on any parts of the analysis. I could only assume that the analyses could be vetoed and access to the data withdrawn. The ICR also stipulated that the title of any paper I produced should specify that the analyses proposed are focused on corroborating the analyses and results published in Konishi *et al*. (2008). The purpose of the data request was not to corroborate such analyses, but to conduct an independent analysis of the data. As noted above, the simulation study presented in de la Mare (2011) demonstrated that methods used in analyses of data by Konishi were statistically inappropriate.

The second problem was the denial of access to the data that were instrumental in choosing the dependent variables used in the body condition analyses. Konishi *et al*. (2008) selected the dependent variables mid-lateral blubber thickness, fat weight and half girth based on the pre-analysis of Konishi (2006) which selected these particular variables from a number of related measurements.

A full analysis needs to begin with all the relevant data, not just starting at the point where a number of data-selection decisions have already been made. My concern is further increased because the ICR claims that the reason for withholding these pre-analysis data was that they had only "limited quality control" (despite the results being published in Konishi, 2006). This statement suggested to me that it is even more important to re-examine the pre-analyses in case there is some underlying sensitivity to the choices made there that subsequently affect the overall result. Given that inappropriate statistical models were used in the subsequent analysis, I could not be confident that the pre-analyses were not also similarly flawed.

The third difficulty is related to the first. The ICR indicated an intention to make public announcements that an Australian Government scientist was collaborating with the ICR, presumably to imply that the Australian Government somehow accepted the validity of JARPA. This is patently not the case.

I did not withdraw my request for the data and I requested the DAG to forward it to the ICR without the acceptance of their proposed conditions. I was generally satisfied that the DAG did their best to facilitate my data request; the problems stem from the data access Procedure B.

In light of these difficulties, the requirements of IWC's Data Access Procedure B need revisiting. Procedure B is weighted towards protecting rights of agencies collecting the data to make first use of it. Independent review of already published analyses should be distinguished from first use.

Procedure A takes into account (IWC, 2004) that:

> *if important management decisions are to be made, they should be based on a full scientific review of both data quality and analysis that can be independently verified.*

It is not clear why a lesser standard applies to Procedure B. The Konishi *et al*. (2008) analyses have been published in the open literature and have been cited as evidence of trends in the Antarctic marine environment even though the published analyses are statistically invalid. In this case there has been a failure of peer review, but such failures will occur from time to time because normal peer review does not involve independent verification of data quality and analysis. Independent verification is the scientifically appropriate process for the identification and correction of flaws in published analyses.

My proposal is:

- Once first use is made of data, Procedure B should allow for independent analysis and review.
- First use is deemed to have been satisfied once an analysis has been submitted to the Scientific Committee, or published in a scientific journal.
- The data used in a first use analysis should not be interpreted narrowly, but should include data used in pre-analyses or supporting analyses or those that are necessary for the proper understanding of the context and processes under which the data were collected.
- Once data have been provided they cannot subsequently be withdrawn, subject to the requirement that the recipient does not use the data for purposes other than independent verification.
- Analyses under these provisions would be eligible for publication in the same or similar journals as the first-use analysis.
- Prior to first-use existing requirements for encouraging a collaborative approach could remain.

Requests for data access after the first-use criterion has been met could still be facilitated through the Scientific Committee and DAG, to ensure that proposals for independent verification meet appropriate scientific standards. Any alternative analyses should be eligible for publication in the same or similar journal in which the first use of the data has been published (with appropriate acknowledgements). If the first-use data analyses were only presented to the Scientific Committee, this would require any re-analysis to also be confined to the Committee. Independent analysis and verification by definition should not require an agreement to collaborate. Although collaboration is desirable and should be facilitated, it should not be mandatory once first use has been made of the data, and in particular it should not be used to inhibit proper review and re-analysis of published studies.

The "Process for the Review of Special Permit Proposals and the Research Results from Existing and Completed Permits" (Annex P, IWC, 2009) states that data access should follow the recommended approach of Procedure B. It is not at all clear how this could work under the existing arrangements in terms of vetting proposals by the Scientific Committee and DAG and the requirement to offer co-authorship to researchers whose work is to be reviewed.

In the process of periodic reviews of whaling programs under Special Permit I propose that:

- All data should be made available to accredited members of the Scientific Committee at least 12 months in advance of the review without any requirements to submit research proposals to the Committee and the data holders.
- There should be no mandatory requirements for collaboration.
- The data to be included is not to be interpreted narrowly (as set out above).

- In the case of programs that have been rolled over, such as JARPA into JARPA II, a review needs to allow for independent verification of long term analyses that rely on data from all of the programs, not just their most recent incarnation.

The requirements for data access in the IWC are considerably more restrictive than those adopted by international agencies such as the World Meteorological Organisation and the Scientific Committee on Oceanographic Research. In both these agencies there are not even provisions for first use; all data are made publicly available without restriction immediately. In the USA and Australia data collected by public agencies are usually freely available under "creative commons" types of arrangements. My proposal does not seek to implement those standards, but the Scientific Committee should take into account that scientific practice is heading towards fewer restrictions on access to data.

REFERENCES

de la Mare, W. K. 2011. Are reported trends in Antarctic minke whale body condition reliable? Paper SC/63/O16 (revised) presented to the IWC Scientific Committee, June 2011, Tromsø, Norway. 25pp.

IWC, 2004. Report of the Data Availability Working Group. *J. Cetacean Res. Manage*. (Suppl.) 6:406-8.

IWC, 2009. Process for the Review of Special Permit Proposals and the Research Results from Existing and Completed Permits (Annex P) *J. Cetacean Res. Manage*. (*Suppl*.) 11:398-401.

Konishi, K. 2006. Characteristics of blubber distribution and body condition indicators for Antarctic minke whales (*Balaenoptera bonaerensis*). *Mammal Study* 31:15-22.

Konishi K, Tamura T, Zenitani R, Bando T, Kato H, Walloe L. 2008. Decline in energy storage in the Antarctic minke whale (*Balaenoptera bonaerensis*) in the Southern Ocean. *Polar Biol.* 31(12):1509-20.

ATTACHMENT A

Application under Procedure B of the Data Access Agreement

(a) *Title* **Review of analysis of trends in Southern Hemisphere minke whale body condition.**

(b) *Investigators:*

William de la Mare, Australian Antarctic Division, Channel Highway, Kingston, Tasmania, Australia, 7050.

(c) *Objectives and rationale of the study* as specified by the Scientific Committee along with the appropriate reference to the report(s) of the Scientific Committee. This will include the reasons why the proposed analyses are important and how they fit into previous work.

Objectives:

To examine whether:

1. the models fitted to the data so far captured all the main features of the data; and
2. the estimate of trend (whose confidence intervals range from near zero to values that would be of appreciable biological significance) could be made more precise.

Rationale:

Analyses of trends in Antarctic minke whale condition have been reported in the open literature (Konishi *et al.* 2008). A simulation study (de la Mare, SC/63/O16, 2011) has indicated that aspects of those analyses may have led to statistical bias in the results or overestimated their statistical precision.

The Working Group on Ecosystem Modelling in 2011 agreed that further analyses of the data were warranted to determine:

> *(i) whether the models fitted to the data so far captured all the main features of the data; and*
>
> *(ii) whether the estimate of trend (whose confidence intervals using the preferred model in Appendix 2 range from near zero to values that would be of appreciable biological significance) could be made more precise.*
>
> *The Group requested, inter alia, results from analysing the two sexes separately and the inclusion of slopes by latitudinal band as a random effect.*
>
> *The Group recommended that the authors of SC/63/O16 and of Konishi et al. apply for access to the data under Procedure B of the Data Availability Agreement, and requested that the data holders consider these requests favourably so that further analyses can be reviewed by the Group next year.*

The proposed work is in accordance with the recommendations of the Working Group endorsed by the Scientific Committee.

(d) *Data to be used* will include a general description of all data to be used as well as data held by ICR. For the ICR-held data, the precise requirements will be given, including the level of disaggregation.

The complete JARPA data on minke whale body condition, including all the variates used in the analyses of Konishi *et al.* (2008), are requested for each whale taken:

Body length
Body weight (when measured)
Age
Diatom adhesion level
All blubber thicknesses at the sites L1, L2, L3, L4 (and D3 when measured)
Half girth (at the three sites evg, um, am)
Blubber weight
Foetus length
Date of capture (year, month, day)
Position of capture (Latitude and longitude)
Sex
Sexual maturity
Pregnant

For each whale, the data on the morphometric and condition indicators specified in Table 1 of Konishi, 2006 are also requested.

The data need to be fully disaggregated as data records for each whale sampled. Where data are missing or not measured for a particular whale, the missing data-field should be left blank. A data description will be needed to link the names of data fields in the records for each whale with names of variables used in the Konishi (2006) and Konishi *et al.* (2008) analyses.

(e) *Description of the methods* likely to be used. The level of detail must be in accordance with the level of novelty of the proposed methods and the particular research questions they will address. References to similar analyses should be included where available.

Standard statistical methods including the usual methods of exploratory data analysis, linear, non-linear and mixed effects model regression will be used in conjunction with bootstrapping methods for the validation of confidence intervals. Model selection will be based on Akaike's Information Criterion.

(f) *Schedule of the work:* this should include estimated times for the various analyses to be carried out and an indication of which investigators will collaborate on individual components. If the project is a long-term project, annual progress reports will be required by ICR and the Scientific Committee.

The analyses should be completed within the time scheduled under Data Agreement, Procedure B. The results of analysis will be reported at SC64.

(g) *Condition for data access*

We agree with all the conditions established in the Scientific Committee data access protocol for Procedure B.

(h) *Output of the research*: this will follow the rules for publication agreed at the Scientific Committee meeting and given below. ICR may consider requests for less stringent conditions (e.g. presentations at non-IWC scientific meetings, publications, etc.). Such requests should be detailed here.

The analyses would only be presented to the Scientific Committee, with the required annotation, according to the Data Agreement, Procedure B. Additional permission

may be sought in due course to publish the updated analyses in Polar Biology or the Journal of Cetacean Research and Management, or in a similar peer reviewed journal.

References

Konishi, K. 2006. Characteristics of blubber distribution and body condition indicators for Antarctic minke whales (*Balaenoptera bonaerensis*). *Mammal Study* 31:15-22.

Konishi K, Tamura T, Zenitani R, Bando T, Kato H, Walloe L. 2008. Decline in energy storage in the Antarctic minke whale (*Balaenoptera bonaerensis*) in the Southern Ocean. *Polar Biol*. 31(12):1509-20

ATTACHMENT B – Letter to DAG from ICR

Tokyo 26 August 2011

Dear Toshi and DAG,

Thank you for your mail and research proposal from Dr. de la Mare. To be honest I am surprised to see an Australian Government scientist asking for data obtained by lethal-technique in the Antarctic. As you know Australia has strongly criticized JARPA and JARPAII programs in the Antarctic.

Before considering the possibility of providing data to Dr. de la Mare through the DAG it will be necessary that the research proposal be modified to fit the guidelines in the protocol for data access of our institute (this is a requirement under Procedure B). Our suggestions are as following:

(a) Title:
The title should be modified to specify that the analyses proposed are focused to corroborate the analyses and results published in Konishi *et al*. (2008). This was the original motivation for de la Mare to prepare document SC/63/O16. Then reference to this published paper should be made in the title.

(b) Investigators:
The ICR protocol stated that a list of co-authors should be provided and that this 'should include at least one scientist from ICR'. This means that this research should be conducted in collaboration with ICR scientists. Collaboration involves consultation among the co-authors for the different analyses and consultation/agreement among the authors for the conclusions of the analyses and output of the study (see item (g) below). This is a basic and common standard for research conducted under collaboration and has been followed in previous research collaboration under Procedure B.

From ICR we would like to propose the names of Kenji Konishi and Takashi Hakamada. We strongly recommend the inclusion of Lars Walloe as a co-author due to his involvement and familiarity with the topic of the research.

(c) Objectives and rationale of the study:
The ICR protocol stated that the explanation on objectives and rationale should 'include the reasons why the proposed analyses are important and how they fit into previous work'.

The objective and rationale in the research proposal are very general. The author lists up several recommendations for additional analyses derived from the Tromso IWC SC meeting, but he failed to put these analyses in the context of advancing the knowledge of Antarctic minke whale. How these analyses will contribute to understand better the ecology of Antarctic minke whale? Or, how these analyses will contribute to the assessment of this species in the Antarctic? (In addition to what we already known from the work of Konishi *et al*. (2008) and others).

The analyses conducted by Konishi *et al*. (2008) followed several recommendations offered by IWC SC members during the JARPA review meetings and IWC SC annual meetings. Furthermore new analyses were conducted at the Tromso IWC SC meeting that confirmed the results of Konishi *et al*. (2008). The research proposal by de la Mare does not take into account the previous SC discussions nor the analyses/results conducted at the Tromso IWC SC meeting. Therefore he fails to explain 'why the proposed analyses are important and how they fit into previous work'.

(d) Data to be used:
The research proposal fails to explain the 'precise requirement' of the data requested in the context of the analyses suggested by the IWC SC. The following text is an extract from the IWC SC report:

> 'Given the potential importance of body condition indices to the Scientific Committee's work, the Working Group agreed that further analysis of the data was warranted to determine:
> (i) whether the models fitted so far captured all the main features of the data; and
> (ii) whether the estimate of trend (whose confidence limits using the best fitting model in Appendix 2 range from near zero to values that could be of appreciable biological significance) could be made more precise.
> The Group requested, *inter alia*, results from analysing the two sexes separately and the inclusion of slopes by latitudinal band as a random effect.

> The Group recommended that the authors of SC/63/O16 and of Konishi *et al.*, as appropriate, apply for access to the data under Procedure B of the Data Availability Agreement, and requested that the data holders consider these requests favourably, so that further analyses can be reviewed by the Group next year'.

Thus the basic objective of the research proposal is to check the calculations and interpretation made in Konishi *et al*. (2008). For that objective only the primary data used by those authors in their analyses are necessary. Consequently we will provide data of blubber thickness and girth only for the sites used in that paper. We will provide data for males and females, but only for mature males and pregnant females, as used by Konishi *et al*. (2008). All the other data requested will be provided (BL, BW, age, diatoms, blubber weight (when measured), foetus length, date of capture, position of capture, sex).

We will not provide additional data because the quality control of data other than those used in Konishi *et al*. (2008) is limited.

The IWC SC has recognized the importance of investigating Antarctic minke whale conditions. The research proposal needs to explain why information on whale condition requires on data obtained by JARPA instead of those obtained by alternative research programs such as the SORP.

(e) Description of the methods:
The ICR protocol stated that 'the level of detail must be in accordance with the level of novelty of the proposed methods and the particular research questions they will address. Reference to similar analyses should be included where available'.

The research proposal fails to make reference to similar analyses. Furthermore the proposal by de la Mare includes methods other than the linear mixed effect model agreed in the EM Group. Methods should be those recommended by the EM Group

(g) Output of the research:
As explained earlier the study should be carried out in collaboration with ICR scientists. Therefore decisions on paper to be written, oral presentations, etc. should be agreed between the first authors and the co-authors. This is a normal practice in research under collaboration.

In case of any agreement, de la Mare must sign up a form agreeing to the terms of this research proposal as well the conditions for data access under procedure B.

Once completed any agreement ICR reserves the right to make public the request and use of data obtained by the lethal component of JARPA by an Australia Government scientist.

Sincerely,

Yoshihiro Fujise, Ph D
Director
Institute of Cetacean Research