SC/67A/SCSP/02

An approach to quantifying potential improvements in management performance from scientific research programmes

Andre E Punt, Greg P Donovan

INTERNATIONAL WHALING COMMISSION

Papers submitted to the IWC are produced to advance discussions within that meeting; they may be preliminary or exploratory.
It is important that if you wish to cite this paper outside the context of an IWC meeting, you notify the author at least six weeks before it is cited to ensure that it has not been superseded or found to contain errors.

SC/67a/SPXX

An approach to quantifying potential improvements in management performance from scientific research programmes

ANDRÉ E. PUNT[1] AND GREG P. DONOVAN[2]

1 School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195-5020, USA
2 International Whaling Commission, The Red House, 135 Station Road, Impington, Cambridge CB4 9NP, UK
Contact e-mail: aepunt@uw.edu

ABSTRACT

An approach for using the RMP *Implementation Simulation Trial* framework to inform quantification of the management-related benefits of research programmes is outlined. This framework quantifies the benefits of scientific research in terms of the improvement in catches given a fixed level of risk. A simple example is provided for a case in which whaling occurs in coastal areas and there is uncertainty about productivity (as quantified using MSYR) and stock structure (one or two stocks)

INTRODUCTION

Management procedures such as the RMP are usually designed to meet catch and conservation objectives. These are expressed in different ways in different management organisations, but broadly cover:

(1) ensuring 'stocks' are not depleted to undesirably low levels or, if already depleted, allowing them to rebuild as fast as possible to user-defined 'optimal' levels – for the RMP, the IWC expresses this such that catches should not be allowed on stocks below 54%[1] of the estimated maximum number of whales that the environment can support ('carrying capacity').
(2) maximising catches over a specified time period – for the RMP this is the considered to be that the highest possible continuing catch is obtained from a management stock (typically 100-year simulations have been used); and
(3) keeping inter-interval variation in catches as small as possible – for the RMP this is expressed such that catch limits are as 'stable as possible'.

It is recognised that there is often a conflict in achieving both catch and conservation objectives and that some 'trade-off' is necessary. In such cases, the approach is often that a minimum standard for one of these is set - for the IWC it relates to 'conservation' such that catches are not allowed on stocks that are below 54% of the unexploited level.

The selection of a management option then involves selecting from those for which this performance criterion has been met against various trade-offs between other aspects of performance such as average catch, catch variation, etc. This approach to selecting a management option is known by some as 'satisficing' (Miller and Shelton, 2010; Punt, in press).

While there is always a trade-off between the three aims, it may be possible to achieve a better performance trade-off (e.g. the higher catches for the same level of risk) given an investment in research to reduce uncertainty. Bergh and Butterworth (1987) illustrate this well for the fishery for anchovy off South Africa, where increased investment in surveys could lead to substantially higher catches without increasing risk. Unfortunately, there has been relatively little work to quantitatively estimate the value of increased research in terms of performance statistics.

This paper first outlines the basic principles that could be used to evaluate the management-related benefits of research. It then provides an illustrative (and simplified) worked example and finally relates the principles and examples to the

[1] This is 10% below the level at which it is assumed that whales are most productive (the 'maximum sustainable yield level' (MSYL) that is assumed to be 60% of unexploited levels).

situation of evaluating Special Permit Whaling programmes that aim to 'improve management' (such as NEWREP-A and NEWREP-NP – see Government of Japan, 2015 and 2017).

METHODS

The general approach

Let us consider the situation in which the catch of animals in a geographical region is to be managed and a minimum conservation performance standard has been agreed.

The initial step is to develop a set of hypotheses for uncertainties related to the population dynamics (e.g. stock structure, productivity) and monitoring (e.g. abundance) that is represented in terms of a set of system models considered 'plausible' (*sensu Implementation Simulation Trials*). The system models can be projected forwards under various management options (known as 'variants' in an RMP context). The 'winning' management option will be that from the set that satisfies the minimum conservation standard and also achieves the best performance for a catch-based performance statistic (e.g. maximum long-term catch) over all of the system models. In effect, the selected management option will depend on its performance for whichever system model is 'most challenging' in terms of achieving the conservation standard (sometimes called the 'worst case plausible scenario').

Proposed research programmes that aim to improve management (i.e. increase catches whilst maintaining conservation performance) should thus show that the 'most challenging' system model is implausible and in effect therefore identify a new 'most challenging' system model. Any proposed research programme should thus estimate the likelihood that it will be able to reject the most challenging system model; the resultant improvement in management can be quantified by examining the catch performance between the original and new 'most challenging' system model.

This approach will then allow decision-makers to evaluate whether the improvement in performance is worth the cost of the research programme (note that such a programme may have several elements).

This is in effect the rationale behind the IWC's 'acceptable with research' option (IWC, 2012).

An illustrative simple example

Let us assume that a country has a whaling operation in its coastal waters, but it is known from surveys that whales are also found offshore. This country aims to manage its whaling operation using a harvest control rule in which the catch limit is set as a proportion of an estimate of the most recent estimate of total abundance (coastal and offshore combined). The country has adopted the performance standard that the probability of any 'management stock' impacted by whaling operating being below 0.6K is not more than 0.05 over all 'medium-high plausibility' hypotheses, where K is carrying capacity.

It takes time to process survey results so the catch limit for year t (CL_t) will be set as $\beta\hat{N}_{t^*}$ where β is a constant rate and $\hat{N}_{t^*}$ is the most recent estimate of abundance (note that in this simple example, unlike the RMP and AWMP, only the most recent estimate of abundance is used). In the context of the principles above, the set of management options is the set of possible values for β. The data currently available to any management options are:

(1) an estimate of abundance of 10,000 whales (3,000 in the coastal area (called Area A) and 7,000 in the offshore area (called Area B) with a CV of 0.5); and
(2) historical catches which have been 100 whales a year, all from Area I.

Surveys are planned to be conducted for coastal and offshore waters (areas A and B) every six years with survey CVs of 0.5 in both areas.

The scientists in the country identified three key areas of uncertainty relevant to determining β and thus establishing the highest possible safe catch limits while meeting the conservation performance standard*:*

(a) survey imprecision;

(b) uncertainty about productivity (the maximum sustainable yield rate, MSYR, expressed in terms of the total number of animals); and

(c) uncertainty about stock structure (i.e. how the total abundance and catches are split by management stock).

They summarised their uncertainty using the four system models (Trials) in Table 1, with a focus on MSYR and stock structure. Survey uncertainty is considered in the context of (a) 'conditioning' (i.e. ensuring that the system models are consistent with the data) and (b) projections under different levels of improvement in the survey CVs (See Appendix A for details of the specifications of the trials). The value of β (and hence catch) was selected based upon meeting the agreed performance standard. In order to examine the management implications of each system model, trials were run and output statistics examined.

Table 1

The trials used to select β. Note that the lowest plausible MSYR is 0.01.

Trials	**MSYR**	**Stock structure**
1	0.01	One stock
2	0.03	One stock
3	0.01	Two stocks
4	0.03	Two stocks

As one would expect, the 'most challenging' system model was that expressed in Trial 3 with two stocks and a low MSYR. To meet the country's minimum performance standard, β is estimated to be 0.00096, which leads to a catch limit of 9 whales (I.e. substantially less than the current catch of 100 whales). In contrast, for the least challenging trial (Trial 2 with one stock and high MSYR), the catch limit would have been 218 (i.e. β =0.02276). Thus, in the extreme, if it could be shown that Trials 1, 3 and 4 were implausible, the 'improvement' in management would result in an almost 25 times increase in catch level.

Given this potential benefit, the country wished to explore whether it is possible by increased research to (i) 'improve' management (i.e. obtain higher catches for the same level of risk), and (ii) if so, determine the likely extent of any such improvement to be able to examine the cost-benefit of undertaking that research. An initial canvassing of scientists led to four initial potential scientific objectives:

(a) estimating MSYR;

(b) estimating the maturity ogive (often used in fish stock assessments e.g. ICES, 2008);

(c) exploring stock structure to try to refine the plausible hypotheses; and

(d) improving the CVs of the abundance estimates.

The scientists recognised that estimating MSYR (objective a) without error is not possible - thus the objective would be to reduce the uncertainty over the current estimates - in terms of improved management, the key parameter is the lower confidence interval. The scientists believed that it may be possible to do this using data obtained from lethal samples and they estimated the level of improvement that may be potentially obtained for a range of sample sizes (catch levels) under certain assumptions (Table 2, Program I). These values will be used in simulation test to quantify the improved management that may be obtained by the research programme for this objective.

While recognising the scientific interest in estimating the maturity ogive, upon review the scientists decided not to pursue it since knowing the maturity ogive would not affect the results of the trials under the proposed management regime (Appendix A). However, it was noted that data to allow its calculation could be collected in conjunction with the lethal sampling required to try to estimate MSYR and understand stock structure and agreed that estimating the maturity ogive could be included as an ancillary objective.

With respect to objective (c), the scientists recognised that whilst eliminating a hypothesis of two stocks in a region is almost impossible, obtaining improved estimates of dispersal rates may in principle either confirm the need to retain a two-stock hypothesis or render it of minimal consequence for management (see Table 2 for the relationship between sample size and the lower limit of dispersal rate. They recognised that tissues samples for analysing the genetics data could be collected using biopsy samples or from lethal sampling.

To evaluate the research programme, there are two main approaches:

(1) implement one of Programs I, II or III and hence treat each of objectives (a), (c), and (d).
(2) create a research program that includes elements of Programs I, II and III and try to address objectives (a), (c) and (d) simultaneously.

Table 2
Impact of different scientific programme types (estimation of MSYR or examination of stock structure uncertainty) as a function of sample size.

Sample Size	Program I Lower confidence limit for MSYR (if MSYR=0.03)	Program II Lower limit of dispersal rate (if there is two stocks)	Program III Sampling CV
0	-	-	0.5
A	0.005	0.01	0.4
B	0.015	0.05	0.3
C	0.020	0.1	0.2
D	0.025	0.2	0.15

Table 3 lists sixteen potential research programmes. Cases I, II and III explore research programmes that address each objective in turn (case I: objective (a) via Program I; case II: objective (c) via Program II; case III via Program III). Case IV explores research programs that address objectives (a), (b) and (c) simultaneously. Results are shown for the status-quo situation in which no research programme is implemented (but surveys continue). Note that the research programs are categorized according to the sample size (A, B, C and D).

Table 3
Alternative "worst case" trials as a function of outcomes for scientific investments.

Approach	MSYR	Stock structure	Survey CV
Status quota	0.01	Two stocks	0.5
Case I-A	0.005	Two stocks	0.5
Case I-B	0.015	Two stocks	0.5
Case I-C	0.020	Two stocks	0.5
Case I-D	0.025	Two stocks	0.5
Case II-A	0.01	Two stocks with dispersal of 0.01 each year	0.5
Case II-B	0.01	Two stocks with dispersal of 0.05 each year	0.5
Case II-C	0.01	Two stocks with dispersal of 0.1 each year	0.5
Case II-D	0.01	Two stocks with dispersal of 0.2 each year	0.5
Case III-A	0.01	Two stocks	0.4
Case III-B	0.01	Two stocks	0.3
Case III-C	0.01	Two stocks	0.2
Case III-D	0.01	Two stocks	0.15
Case IV-A	0.01	Two stocks with dispersal of 0.01 each year	0.5
Case IV-B	0.015	Two stocks with dispersal of 0.05 each year	0.5
Case IV-C	0.020	Two stocks with dispersal of 0.1 each year	0.5
Case IV-D	0.025	Two stocks with dispersal of 0.2 each year	0.5

In terms of quantifying the effects of the research programme, then the given the results from the research program, Trial 3 would be replaced by a trial that (a) reflects the lower 95% confidence interval for MSYR (Cases I-A-D), (b) a trial with two stocks but dispersion between them (Cases II-A-D), (c) a trial with two stocks and low value for MSYR but more precise survey estimates of abundance (Cases III-A-D), or (d) a trial with a higher value for MSYR and dispersal between the two stocks (Cases IV-A-D).

To quantitatively compare the 16 alternative approaches (and the status quo), projections are conducted where each of trials 1-4 are conducted in which Trial 3 is replaced by the alternative trial that reflects the outcome of each approach. Performance is measured in terms of the value for β (which is equivalent to the catch limit for the next year given an abundance estimate of 10,000), and the average catch over 100 years.

RESULTS AND DISCUSSION

Illustrative Application

Expected value of perfect information

Table 4 lists the short-term catch and the average catch over 100 years for Trials 1-4 when the management option is based on achieving the performance standard for Trial 3. The short-term catch is always 9 (10,000 multiplied by 0.00096 and rounded down), but the average catch over 100 years is larger for Trials 1, 2, and 4, but not by a substantial amount. This is partially because the harvest control does not try to "learn" about the productivity of the stock.

Table 4
Impact of basing management on achieving the performance standard for Trial 3 and the basis for calculating the expected value of perfect information. The values in parenthesis are differences from the case where there is no knowledge which trial is correct.

	Trial 1	Trial 2	Trial 3	Trial 4
β =0.00096				
Short-term catch	9	9	9	9
Average catch (100 years)	1087	1063	1061	1085
Risk	0	0	0.05	0
β selected by trial				
β	0.00928	0.02276	0.00096	0.00386
Short-term catch	92 (83)	227 (218)	9 (0)	38 (29)
Average catch (100 years)	8401 (7314)	19335 (18272)	1061 (0)	4163 (3078)
Risk	0.05	0.05	0.05	0.05

The expected value of perfect information is the difference in short-term and average catch over 100 years if β is selected for each Trial in turn and the short-term catch and the average catch over 100 years when β is set to 0.00096. These differences reflect the maximum improvement in catch possible if the set of trials reflect possible 'states of nature'. The improvements in both short-term and average catch over 100 years are substantial, particularly if Trial 2 can be shown to be the true state of nature. Note that the optimal values for β are not 0.01 and 0.03 for Trials 1 and 2 because the performance standard reflects a probability of 0.95 of being above 0.6*K* and also because the estimates of abundance are imprecise and only available every six years.

Alternative research projections

Table 5 lists the values for β and the change in average catch over 100 years compared to the status-quo. The value for β is based on the new 'worst case plausible scenario' given the outcomes from the research (Table 3). The results for case I-A are identical to those for the status quota (change in short-term catch of zero) because this approach would be expected to lead to a lower bound for MSYR that is lower than the current minimum value of 0.01 suggesting this approach would be scientific and financially unjustifiable.

Table 5
Change in short-term catch and average catch over 100 years for each of the 16 research programs

Research programme	**β**	**Short-term**	**Trial 1**	**Trial 2**	**Trial 3**	**Trial 4**
Case I-A	0.00096	0	0	0	0	0
Case I-B	0.00185	9	977	962	974	976
Case I-C	0.00262	17	1796	1784	1807	1803
Case I-D	0.00328	23	2470	2473	2500	2487
Case II-A	0.00239	14	1517	1536	1507	1554
Case II-B	0.00461	37	3663	3847	3674	3853
Case II-C	0.00596	50	4841	5205	4858	5195
Case II-D	0.00736	64	5958	6575	5975	6549
Case III-A	0.00097	0	10	10	9	10
Case III-B	0.00099	0	19	19	19	20
Case III-C	0.00100	1	36	36	35	36
Case III-D	0.00100	0	32	32	31	33
Case IV-A	0.00239	14	1517	1536	1507	1554
Case IV-B	0.00535	44	4508	4597	4521	4598
Case IV-C	0.00811	72	7133	7279	7138	7249
Case IV-D	0.01252	116	10990	11199	10975	11137

Several features are immediately apparent from Table 5 (for now treating the research programs for each level of sampling as equally "costly"). In particular, improving the survey CV has a much smaller impact on both short-term and long-term catches than research to increase the lowest plausible value for MSYR and understanding about stock structure. Improving understanding about stock structure leads to greater improvements in catch than increasing the lowest plausible value for MSYR for the same sample size (Case II vs Case III), while conducting a research program that simultaneously increases the lowest plausible value for MSYR and estimates dispersal not surprisingly leads to a highest catches. Somewhat surprisingly, the results for Case IV are not simply to sum of those for Cases I and II. Unsurprising, greater benefits accrue from higher sample sizes.

None of the research programmes in table 5 are able to achieve the same performance as Trial in Table 4 because it is never possible to increase the lowest plausible value for MSYR to 0.03 and to completely eliminate the two-stock hypothesis

Relevance to RMP and Special Permit Whaling

The illustrative example is simplification of the situation for the RMP and Scientific Permit Whaling for several ways. In particular, the management options for RMP are not different values for the slope of a control rule, but are instead specifications for RMP variants. However, the basic approach is still analogous. Instead, of selecting the value of β to achieve a given probability of being about 0.6K in 100 years, the RMP process involves selecting the RMP variant that has the highest average catch given it achieves the performance expected under the rules developed by the Scientific Committee (IWC, 2012).

The operating models considered in the illustrative example are simple to condition and run – this is not the case for the *Implementation Simulation Trials* used in *Implementations* and *Implementation Reviews* (e.g. those for western North Pacific minke whales; IWC 2014; and the North Atlantic fin whales; IWC 2017). Nevertheless, the availability of *Implementation Simulation Trials* provides a way assess both the impact of catches on stocks as well as the ability of research to reduce the range of uncertainties considered in trials.

The example application does not outline how Table 2 was created, i.e. how the lowest plausible value for MSYR and the extent of dispersal are related to sample sizes (and the sampling design of any future research program). Here the *Implementation Simulation Trials* could be used. This would involve projecting the trials (except for the 'worst case plausible scenario' trial) forward under the planned removals (with the removals assigned to location and times during the season given the sampling plan), generating the data that would be collected during the research activities (taking due account of, for example, overdisperson and sampling error for age, etc.), and applying the planned analysis methods. In the context of estimation of MSYR, methods such as statistical catch-at-age analysis (e.g. Punt *et al*., 2014) could be applied, while standard methods exist to compute dispersal rates when there is a single stock but it is (incorrectly) assumed there is only stock (e.g. Pastene *et al*. 2016). The analyses conducted for NEWREP-A (Government of Japan, 2016) provide an example of the basic calculations, albeit not using *Implementation Simulation Trials*.

The calculations outlined in this document are non-trivial and would require considerable planning and analysis to be ready for inclusion in a research proposal. However, they would require the proposers of research to explicitly consider their objectives, to document how the sampling and analytical work will be used to refine uncertainties, and hence to provide information about likely benefits of future research. The approach of this document pre-supposes the existence of *Implementation Simulation Trials*, which will not always be the case (e.g. for sei whales in the North Pacific). However, the basic approaches adopted by the Scientific Committee could be used to develop a set of trials that approximates those that are likely to be developed during an *Implementation* process.

The results from the calculations outlined in this document relate directly to several of the expectations of the 'Annex P' process used to evaluate proposals of Scientific Permit programs. In particular, the calculations would need to take account of the objectives, be based on a fully-specified sampling program, and provide results that directly address the benefits of research to improving management.

REFERENCES

Bergh, M.O. and D.S. Butterworth. 1987. Towards rational harvesting of the South African anchovy considering survey imprecision and recruitment variability. *S. Afr. J. Mar. Sci.* 5: 937-51.

Government of Japan. 2015. Proposed Research Plan for New Scientific Whale Research Program in the Antarctic Ocean (NEWREP-A). Document submitted to the February 2016 review of NEWREP-A

Government of Japan. 2016. Results of the analystical work on NEWREP-A recommendations. IWC Document SC/66b/SP10. 23pp.

Government of Japan. 2017. Proposed Research Plan for New Scientific Whale Research Program in the western North Pacific (NEWREP-NP). IWC Document SC/J17/JR01. 163pp.

ICES. 2008. Report of the Workshop on Maturity Ogive Estimation for Stock Assessment (WKMOG), 3- June 2008 , Lisbon, Portugal. ICES CM2008/ACOM:33. 72 pp.

International Whaling Commission (IWC) 2012. Requirements and Guidelines for Implementations under the Revised Management Procedure (RMP). *J. Cetacean Res. Manage.* (*Supp*l.), 13:497–505.

International Whaling Commission (IWC). 2014. Report of Sub-Committee on the Revised Management Procedure. Annex D to Report of the Scientific Committee. *J. Cetacean Res. Manage.* (*Supplement*) 15: 87-188.

International Whaling Commission (IWC). 2017. Report of Sub-Committee on the Revised Management Procedure. Annex D to Report of the Scientific Committee. *J. Cetacean Res. Manage.* (*Supplement*) 18: 00-00.

Miller, D.C. and P.A. Shelton. 2010. “Satisficing” and trade-offs: evaluating rebuilding strategies for Greenland halibut off the east coast of Canada. *ICES J. Mar. Sci.* 67: 1896–1902.

Pastene, L.A., Goto, M., Taguchi, M. and T. Kitakado. 2016. Updated genetic analyses based on mtDNA and microsatellite DNA suggest possible stock differentiation of Bryde’s whales between management sub-areas 1 and 2 in the North Pacific. IWC Document SC/F16/JR44 (17pp).

Punt, A.E. In press. Strategic management decision-making in a complex world: quantifying, understanding, and using trade-offs. *ICES J. Mar. Sci.* 00: 00-00.

Punt, A.E., Bando, T., Hakamada, T. and Kitakado, T. 2014. Assessment of Antarctic Minke Whales using Statistical Catch-at-age Analysis. *J. Cetacean Res. Manage.* 14: 93–116.

Appendix A: The Operating Model

A.1 Single-stock operating model

The population dynamics are governed by the equation:

$$N_{y+1} = N_y + rN_y(1-(N_y / K)^z) - C_y \tag{A.1}$$

where N_y is the number of animals at the start of year *y*, *r* is the intrinsic rate of growth, *z* is the degree of compensation, *K* is carrying capacity, and C_y is the catch during year *y*. The population size in year -30, N_{-30}, is assumed to be equal *K*. The catches between years -30 and 0 are all 100.

The numbers of animals in areas 1 and 2 ($\tilde{N}_y^1$ and $\tilde{N}_y^2$ respectively) are given by χN_y and $(1-\chi)N_y$ respectively, where χ is the mixing proportion.

A.2 Two-stock operating model with mixing

The population dynamics are governed by the equations:

$$\bar{N}_{y+1}^1 = \tilde{N}_y^1 + r\tilde{N}_y^1(1-(\tilde{N}_y^1 / K^1)^z) - C_y \tag{A.2a}$$

$$\bar{N}_{y+1}^2 = \tilde{N}_y^2 + r\tilde{N}_y^2(1-(\tilde{N}_y^2 / K^2)^z) \tag{A.2b}$$

$$\tilde{N}_{y+1}^1 = (1-\beta^{1-2})\bar{N}_{y+1}^1 + \beta^{2-1}\bar{N}_{y+1}^2 \tag{A.2c}$$

$$\tilde{N}_{y+1}^2 = (1-\beta^{2-1})\bar{N}_{y+1}^2 + \beta^{1-2}\bar{N}_{y+1}^1 \tag{A.2d}$$

where $K^{1,2}$ are carrying capacities for stocks 1 and 2, β^{1-2} is the annual dispersal rate from stock 1 to stock 2, and β^{2-1} is the annual dispersal rate from stock 2 to stock 1 ($\beta^{2-1} = \beta^{1-2}K^1 / K^2$). $\tilde{N}_{-30}^1 = K^1$ and $\tilde{N}_{-30}^2 = K^2$.

A.3 Conditioning

The values for *r*, *z* and β^{1-2} (two stock operating model only) are pre-specified. The values for *r* and *z* are selected so MSYR and MSYL equal pre-specified values (0.01 or 0.03 for MSYR and 0.6 for MSYL). The values for the remaining parameters of the operating model (*K* and χ for the single-stock model, and K^1 and K^2 for two-stock operating model) are selected to match the generated numbers by area at the start of year 1. The latter numbers are generated as log-normal variables with means of 3,000 (area 1) and 7,000 (area 2) and a CV of 0.5

A.4 Data generation

An estimate of abundance is generated every 6th year, according to:

$$\hat{N}_y = \tilde{N}_y^1 e^{\varepsilon_y^1 - \sigma_A^2/2} + \tilde{N}_y^2 e^{\varepsilon_y^2 - \sigma_A^2/2} \tag{A.3}$$

where σ_A is the standard error of the logarithm of abundance by area.