Exploration of a future Mark-Recapture study of Antarctic Blue Whales

DAVID PEEL[ab], MARK BRAVINGTON[ab], NATALIE KELLY[ab] and MICHAEL DOUBLE[b]

[a] *CSIRO Digital Productivity and Services/Wealth from Oceans National Research Flagship, Castray Esplanade, Hobart, Tasmania, 7000, Australia; Australian Marine Mammal Centre, Australian Antarctic Division, Department of the Environment, Channel Highway, Kingston, 7050, Australia*

[b] *Australian Marine Mammal Centre, Australian Antarctic Division, Department of the Environment, Channel Highway, Kingston, 7050, Australia*

ABSTRACT

To properly conserve and manage wild populations, it is important to have information on abundance and population dynamics. In the case of rare and cryptic species, especially in remote locations, surveys can be difficult and expensive, and run the risk of not producing sample sizes large enough to produce precise estimates. Therefore, it is crucial to conduct preliminary analysis to determine if the study will produce useable estimates. The focus of this paper is a proposed mark-recapture study of Antarctic blue whales (Balaenoptera musculus intermedia). Antarctic blue whales were hunted to near extinction up until the mid-1960s, when commercial exploitation of this species ended. Current abundance estimates are a decade old. Furthermore, at present, there are no formal circumpolar-level cetacean surveys operating in Antarctic waters and, specifically, there is no strategy to monitor the potential recovery of Antarctic blue whales. Hence the work in this paper was motivated by the need to inform decisions on strategies for future monitoring of Antarctic blue whale population. The paper describes a model to predict the precision and bias of estimates from a hypothetical survey programme. The analysis showed that mark-recapture is indeed a suitable method to provide a circumpolar abundance estimate of Antarctic blue whales, with precision of the abundance, at the mid-point of the programme, predicted to be between 0.2 and 0.3. However, this was only if passive acoustic tracking was utilised to increase the encounter rate. The analysis also provided guidance on general design for an Antarctic blue whale programme, showing it required a 12 year duration; although surveys did not necessarily need to be run every year if multiple vessels are available to clump effort. Mark-recapture is based on a number of assumptions; it was evident from the analysis that ongoing analysis and monitoring of the data would be required to check such assumptions hold (e.g., test for heterogeneity), and the modelling adjusted as needed.

KEYWORDS:

Mark-recapture, survey design, cetaceans, Antarctic blue whale, *Balaenoptera musculus intermedia*

INTRODUCTION

For endangered species, unbiased and precise estimates of their population size and dynamics are vital to inform conservation and management decisions (Cooke 1995). These population estimates are only useful if they have the precision to be effective; for example, to detect changes in population size or address other specific scientific questions (Reynolds, Thompson & Russell 2011). Therefore, it is paramount to conduct thorough planning studies and establish in advance if candidate survey approaches and designs are capable of producing estimates of suitable precision. This is an accepted, but not universally adhered to premise (Reynolds, Thompson & Russell 2011).

This paper is concerned specifically with estimating marine mammal abundance. Surveys to derive estimates of abundance of marine animals, particularly for species distributed across large and remote areas, are often logistically challenging and expensive (Williams & Thomas 2009). Therefore, there is an even greater need for planning, to help ensure the most efficient design to produce sufficient sample size and, hence, adequately precise parameter estimates.

Although there are a number of different survey methods used to derive abundance estimates (Schwarz & Seber 1999), the two most common and proven methods for cetaceans are distance sampling methods (Buckland *et al.* 2001; Borchers, Buckland & Zucchini 2002), particularly line transect surveys (Branch & Butterworth 2001; Hedley & Buckland 2004) and mark-recapture methods (using either photo-identification data (Hammond 1986) or genetic samples (Carroll *et al.* 2011; Rew *et al.* 2011; Constantine *et al.* 2012)). Both line transect and mark-recapture have advantages and disadvantages (Evans & Hammond 2004). This paper focuses on mark-recapture.

Mark-recapture approaches can be less labour intensive, less reliant on formal track locations and can yield other information such as life history and population structure, but can be prone to biases when heterogeneities in sampling probabilities exist. In the context of this paper, surveys are conducted and individuals are identified, via natural markings (e.g., genetic samples from biopsies and colouration and shape of the body, usually recorded in photographs (Hammond 1986)).

The motivation for this paper is to design a proposed study of Antarctic blue whales (Double *et al.* 2013). During the twentieth century, blue whale populations were hunted to near extinction (Branch *et al.* 2007b) and they still remain, after four decades of protection from harvesting, a massively deleted population. As such, the species is classified as critically endangered by the International Union for Conservation of Nature (Reilly 2008) and therefore is considered one of the most at risk baleen whale species in the Southern Ocean (Leaper & Miller 2011). Our understanding of the impact of the whaling era on the circumpolar population of Antarctic blue whales is predominantly based on two sources of information: catch data derived from the logbooks of whaling vessels (International Whaling Commission 2009) and from circumpolar sightings surveys for cetaceans. Circumpolar Antarctic sighting surveys, operating under the auspices of the International Whaling Commission (IWC), were initiated in 1978 as the International Decade of Cetacean Research (IDCR) and later Southern Ocean Whale Ecosystem Research (SOWER) initiatives (Branch & Butterworth 2001). SOWER (as the overall programme is henceforth referred to) provides, to date, the most comprehensive survey data for cetacean species in the Southern Ocean (Branch *et al.* 2007b). Estimates from SOWER show the Antarctic blue whale population to have been around 453 (CV = 0.4) 1981, 559 (CV = 0.47) around 1988 and 2,280 (CV = 0.36) in 1998 (Branch 2007). This may indicate the beginning of recovery of this species, but with such imprecise abundance estimates, it is difficult to draw strong conclusions.

Unfortunately, these abundance estimates are more than a decade out of date. Currently, there are no formal circumpolar-level Antarctic cetacean surveys and, specifically, there is no strategy to monitor the potential recovery of Antarctic blue whales. Hence the work in this paper was motivated by the need to inform decisions on future monitoring strategies.

Given these motivations, a survey programme has been proposed to provide information on the status of the Antarctic blue whale population. The programme will be an international collaboration with varied vessels, and is planned to span over a number of years/seasons.

The following sections will describe some of the challenges and possible solutions to a mark-recapture study of this kind. This is followed by an analysis to predict the precision of abundance estimates based on what is known about the population. Using this analysis we are able to test various programme designs and examine the effect on bias and precision if some of assumptions are invalid.

Although this paper has a strong focus on the specific question of an Antarctic blue whale mark-recapture survey, the general approach is applicable to any mark-recapture survey. In particular, in early planning when perhaps little is known and few data exist, this type of analysis can be informative.

PROPOSED STUDY AREA

The study area (see Fig. 1) corresponds to the summer feeding grounds of Antarctic blue whales, that is the circumpolar region between the ice-edge and 60°S (Branch *et al.* 2007a).

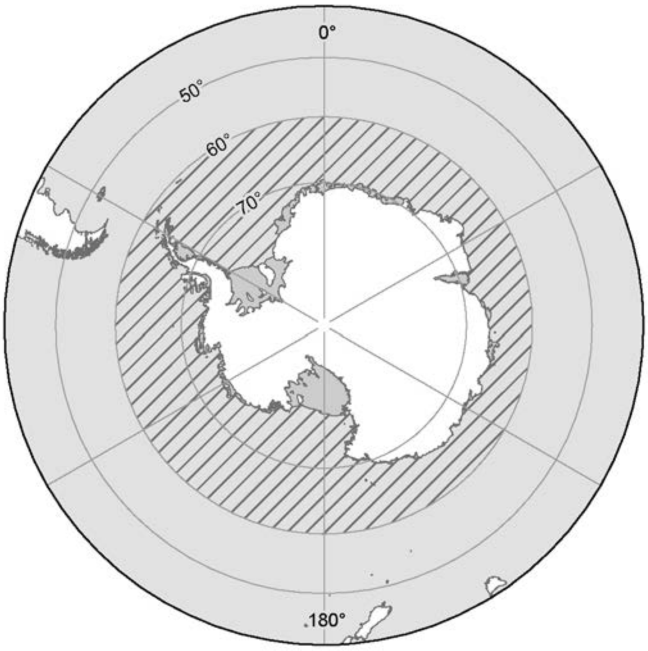

Figure 1: The study region we are considering (hashed area) covers the Antarctic blue whale feeding grounds i.e., the circumpolar area south of 60°S.

METHODS

Potential survey and analysis issues

There are a number of issues facing a mark-recapture study of Antarctic blue whales. The following is a description of some of the potential key issues, repercussions and possible solutions:

Low sample size

Collecting enough biopsies or photographs to obtain precise mark-recapture population estimates may be difficult, due to low densities and the difficulty of surveying in the Southern Ocean. Passive acoustics can detect vocalising blue whales at a distance of an order magnitude greater than visual observation. Therefore, it has been proposed to use passive acoustics, in the form of deployed sonobuoys, to detect and track down blue whales to increase the encounter rate.

Spatial heterogeneity

One of the underlying assumptions for unbiased mark-recapture estimation is equal probability of capture across all animals. One of the major potential threats to this assumption is spatial population structure. For example, if the population is made up of sub-groups distinct in space, and disproportionate sampling results in different sub-groups having different probabilities of capture. Unfortunately, little is known of Antarctic blue whale population structure. Analyses of Discovery mark data (see Branch *et al.* 2007b) and of genetic analyses from biopsy samples (Sremba *et al.* 2012) are somewhat inconclusive in regards to support for the existence of population structure. A historic Discovery mark series, reported in (Branch *et al.* 2007b), indicate that whilst the majority of within-season recaptures suggest animals did not move more than 60° of longitude between marking and slaughter, the series also found that as the number of seasons between these events increased, so, too, did longitudinal movement, with some individuals killed nearly 180° of longitude away from their marking. With a study of contemporary individual marking data from biopsy samples, (Sremba *et al.* 2012) do report on evidence of population structure in Antarctic blue whales. However, due to a small sample size, (Sremba *et al.* 2012) were forced to apply *a priori* assumptions to the clustering of the genetic samples (i.e., clustering was based on the location of samples within predefined management areas), which may influence, to some degree, the results of the test of genetic differentiation between these areas.

As the future survey progresses, more biopsy samples, and subsequent genetic analyses, and basic information on animal movement from recaptures, should help us strengthen conclusions about any population structure in Antarctic blue whales.

Another potential form of heterogeneity is if there is clustering of animals, for example, into family units. Again, as genetic data is collected, the samples could be compared to establish relatedness and compared in terms of sampling time, and location, to test for non-randomness.

Sexual capture heterogeneity

Heterogeneity in capture probabilities could also be introduced with the use of acoustic tracking of animals as there is a distinct bias towards finding vocalising males. The amount of sampling bias is not obvious, since whales are usually detected and found at the group level rather than individual animals. Hence, non-vocalising females will be sampled if they are within a group containing a vocalising male. A recent pilot study (Double 2013) using passive acoustics to track Antarctic blue whales resulted in 13 male and 4 female biopsies, i.e., a 76% male sampling bias. If sexing of subsequent biopsy samples is available, heterogeneity in marking of the genders can be handled by the model, i.e. as described in Appendix A. If sex is not known (e.g. if using photographic-ID) or there is another unknown mechanism for differential capture probabilities, then a more complex model would be required (e.g., Cubaynes *et al.* 2010).

Multiple catalogues

For Antarctic blue whales, distinct patterns in mottled pigmentation along the back and variable dorsal fin shapes and sizes aid in identifying individuals (Sears & Perrin 2008). Such pigmentation pattern can be distinctive for both sides of an animal, therefore, unless both sides are photographed; two distinct mark catalogues would arise. Along with genetic biopsy data, this potentially results in three mark-recapture catalogues. There are a number of models proposed to handle multiple catalogues (Bonner and Holmberg (2013) and McClintock *et al.* (2013)). However, for simplicity, in this paper we do not consider the photographic catalogues, and just use the catalogue of genetic biopsy data in the calculations. It would be expected that if photograph catalogues are included in future analysis, it should result in greater precision due to the larger sample size.

Proposed survey

The proposed programme would involve one or more vessels conducting annual surveys during the Antarctic austral summer, beginning in 2015. Acoustic sonobuoys would be used to detect and track down whale groups, as well as visual observers. Once found, whales would be photographed and biopsies taken.

No formal track or design would be specified, but rather for each 'vessel-year' (the unit of effort a single vessel could reasonably complete in a given season) a survey region would be defined (most likely in terms of a longitudinal range) within the study area (Fig.1). Within this region, the vessel would search based on acoustic detections, as well as being informed by historical density data and models, weather forecasts, logistics, and sea ice coverage. It should be noted that a mark-recapture approach can easily incorporate other more opportunistic data from less formal surveys. In this paper, however, we focus on these dedicated surveys only.

For a well-mixed unstructured population, mark-recapture methods do not require completely uniform spatial coverage. Therefore, the survey can take advantage of the putative higher densities of Antarctic blue whales near the summer sea ice edge, relative to those areas further north (Branch *et al.* 2007b), and focus on regions in the Southern stratum (0-200 km from the ice edge). However, it would be prudent to distribute effort longitudinally around the Antarctic summer sea ice-edge, to combat problems from any potential population structure.

Estimating precision

We now describe the calculations to predict the precision of the abundance estimates arising from our conceptual mark-recapture programme.

Mark-recapture model

We adopted a simple open mark-recapture model of both sexes in the adult population, on which to base variance prediction. Since little is known about the spatial heterogeneity and movement of Antarctic blue whales we assumed no spatial heterogeneity. The model is based on a Poisson distribution of recaptures (Cormack 1989) and does not consider within-year recaptures (see Appendix A for full details). From this model: population size (N_0) in the specified start year, mortality (M) and logistic growth rate (r) and sex ratio (λ) can be estimated by maximizing the likelihood. Alternatively, if these are known, they can be fixed.

Precision

To estimate precision we simulated realised populations based on various potential population scenarios and fitted the mark-recapture model, with the following steps:

1. *Simulate a presence history matrix*
   Based on exponential annual population growth over the planned survey period, and a given mortality, growth rate and population size in the start year, we created a presence history matrix containing the fates of all animals over the sampling period of interest (for each year every whale is allocated a '1' if alive and a '0' if dead or not born yet).
2. *Estimate sample sizes*
   To provide an indicative mark-recapture survey sample size, we used an individual-based simulation study (as described in Peel *et al.* In Press). Specifically, for a given whale density, whales were randomly placed in a survey region and an acoustic-assisted mark-recapture survey was simulated based on search protocols/rules and assumed acoustic/biological properties (see Supplementary material A for full details).
3. *Generate a capture history*
   A mark-recapture sampling framework was applied to the simulated presence history, to create a capture history. As per our simple mark-recapture model it was assumed that animals in the

population are homogenous in terms of probability of being encountered, and sampling was random. Alternative assumptions were implemented as part of the sensitivity analysis described later.

4. *Fitting Mark-Recapture model*

To the simulated capture histories we fitted the mark-recapture model and calculated the estimates of N_0, M, r and λ along with their variances, using the inverse Hessian of the negative log likelihood. If we are only concerned with estimating the variance, and not interested in assessing bias, we could forego the full fitting process and simply examine the Hessian of the likelihood surface at the true parameters and the variances, averaged over a small number of simulated populations. Otherwise, to examine bias, we completed the full process.

In theory we could calculate the variances from a single simulated population. However, we found it prudent to simulate multiple replicated populations and take the median variance or CV.

Historically, Antarctic whale abundance is reported as a circumpolar decadal average so we re-parameterised the likelihood in terms of N_{mid} and r, where N_{mid} is the abundance in the middle year of the multi-year programme.

For comparison, we also included in our analysis a typical visual-based mark-recapture survey (i.e., without passive acoustic assistance) where the vessel searches within the southern stratum in a zig-zag pattern stopping and marking any animals that are found visually.

A summary of the main parameters used in the simulation and sources is given in Table 1. Unless stated otherwise, these are the parameters used to simulate populations; we will refer to these as the 'base case' (in particular N_0 = 2280, r = 6.4%, M = 4%, λ = 0.473 and a 12 year survey programme, with a single vessel surveying annually).

It is worth examining the mechanisms that dictate precision. In mark-recapture, the precision of the abundance estimates is related to the number of recaptures, which is driven by the true underlying population size and its relationship to sample size and probability of recapture. We therefore also found it valuable to examine the predicted sample sizes and recaptures in the analysis.

Table 1: 'Base case' parameters for the simulations/analysis

Parameter	**Value**	**Comment/Reference**
Population size in 1998 (N_0)	2280 (1350,3450)	International Whaling Commission (2009)
Population growth rate (r)	6.4% (2.4,8.4)	International Whaling Commission (2009)
Mortality (M)	4% (1-7%)	Branch, Matsuoka and Miyashita (2004)
Population sex ratio (λ)	0.473 females	Branch (2004) (5,637 females out of 11,942 foetuses = 0.472) and 87,098 females out of 184,280 adults = 0.473)
Expected group size	1.59	Branch (2007)
Within-season re-sight rate	16%	Olson (2010) and Olson *et al.* (2013)
Expected Biopsy success	0.60	
Density of groups in 1998	2.269185×10^4 groups per km^2	Based on a SOWER encounter rate in the area 200 km to the summer ice edge
Typical survey duration	37.5 days	Based on pilot study (Double 2013)
Existing historical catalogues (*To avoid the complexity of modelling multiple catalogues, in this exploratory analysis we consider genetic data only.*)	166 individuals years: 1991-2009 17 individuals in 2013	Attard *et al.* (2012); Sremba *et al.* (2012) Double (2013)

Heterogeneity and Spatial structure

As discussed earlier, the potential for heterogeneity in capture probability due to spatial population structure is a concern, as it will introduce bias in the estimates. However, the hope is that as data is collected, any spatial structure could be detected and modelled from the locations of the recaptures (with respect to the marks). To better quantify this idea, we applied our model to various simulated data generated with inherent spatial structure, and estimated bias; we then applied a simple χ^2 test to try to detect the spatial structure.

The framework (see Fig. 2) to generate this spatial structure was to define two hypothetical survey areas (‘East’ and ‘West’) with uneven sampling effort, E_E and E_W, respectively. Next, we created two types of whales: those that prefer the Eastern survey area and those that prefer the Western area. In any given year, any particular whale has a defined probability, P_H, of being in its home area, H, and (1- P_H) of being in its non-preferred home area. This behaviour and unequal sampling effort between areas results in uneven capture probabilities for the two sub-populations. We examined a gradual range from no site fidelity (P_H = 0.5) to strong site fidelity (P_H = 1.0) and survey effort ratio East to West (E_E:E_W) ranging from 50:50% to 90:0%.

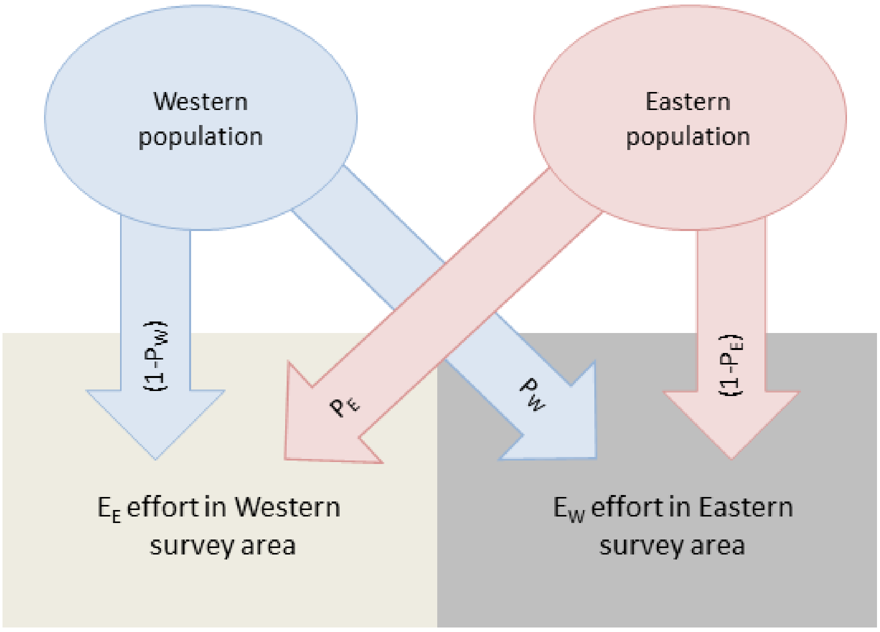

Figure 2: Simple simulation model to generate data with spatial structure and hence unequal capture probabilities. Consisting of two sub-populations of whales (East and West) that demonstrate site-fidelity depending on the parameter P_E and P_W.

RESULTS

Model Choice

The mark-recapture model used to calculate precision contains four parameters: N_0, M, r and λ. There is existing information on the sex ratio (λ) of Antarctic blue whale and, to a much lesser extent, mortality (M).

There is extensive historical data on sex ratio (λ) of Antarctic blue whales Branch (2004), hence the existing estimate (Table 1) should be reasonably precise. Whereas, for mortality (M), there is some uncertainty (Branch, Matsuoka and Miyashita (2004) report mortality of between 1 and 7%). Given the uncertainty about mortality, we chose to adopt the model that fixes λ and fits N_0, r and M.

To examine the effect of estimating λ and M in the model, versus using the existing estimates, we generated 1000 base case simulations and calculated the precision of $\widehat{N}_{mid}$ for the various models arising from fitting, or fixing λ and M (see Fig. 3a).

In the unlikely event that the historical sex ratio parameter value is incorrect, some bias would be introduced into the estimates. To determine how pronounced this bias would be, we generated data with other sex ratios and fitted the model with our historical value fixed (see Fig. 3b). It is evident that for small errors in sex ratio, the bias would be acceptable.

In general we simulated data only from populations with mortality equal to 4%. However, as this value is far from certain, we tested how assumed true mortality affected the simulated precision of $\widehat{N}_{mid}$ by fitting the model over a range of true mortalities (Fig. 3c). In all cases, the precision was still within acceptable limits.

Precision

Given the uncertainty in the population parameters N_0 and r, we also examined populations with a range of N_0 between the 90% confidence intervals (1350-3450) derived by Branch (2007) and r between 0 and 10%. For each (N_0, r) parameter combination, 100 populations were generated and the precision estimated for the base case (i.e., 12 year survey programme, with a single vessel survey annually).

As expected, acoustic-assisted mark-recapture provided a significantly larger sample size than visual mark-recapture, for all combinations of N_0 and r (Fig. 4a, 4b and 4c). We found that as r or N_0 increase, the expected number of recaptures in mark-recapture decreases for acoustic-assisted mark-recapture (Fig. 4d and 4f) and increases slightly for visual mark-recapture (4e and 4f). As expected from the disparate sample sizes, acoustic assisted mark-recapture provided many more recaptures than the visual-only method (Fig. 4d, 4e and 4f).

The sample size and the number of resultant recaptures determine the final precision (Fig. 4g and 4h). The predicted CV of visual mark-recapture surveys decreased with increasing r and N_0 (Fig. 4h). The CV of acoustic-assisted mark-recapture on the other hand increased with increasing N_0 or r (Fig. 4g). It is interesting to directly compare how differently the precision each method responds to changes in r (Fig. 4i).

In terms of a future Antarctic blue whale survey programme, if we look at the results at the base case (i.e. a 12 year annual survey with the agreed IWC Antarctic blue estimates of N_0, r), the predicted CVs for acoustic-assisted mark-recapture are around 0.28; and the CVs for visual mark-recapture were much larger at 0.50 (the base case parameters are denoted by small squares in Fig. 4).

With regard to estimating the growth rate (r) and mortality (M), it was found that the estimates were too imprecise to be useful. For example, a 95% confidence interval on a population rate of increase estimated in the base case was (-1.9%, 12.5%). Similarly, the estimates of mortality were not precise enough for further consideration.

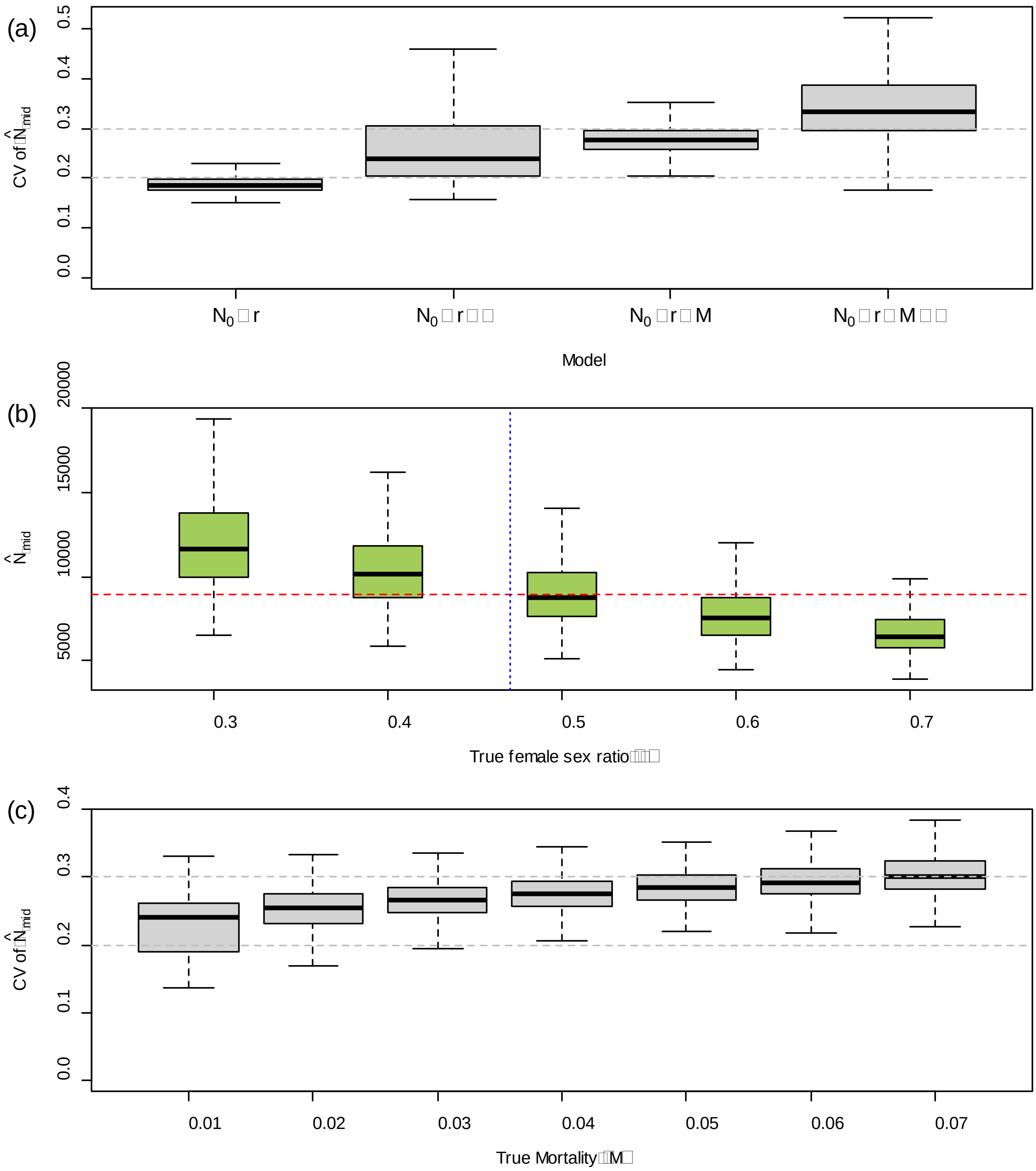

Figure 3: (a) The precision of $\widehat{N}_{mid}$ results for various model options, in terms of which parameters are estimated e.g., (N_0,r,M,λ) denotes N_0, growth r, mortality M, and sex ratio λ were estimated, and (N_0,r) denotes only N_0 and r were estimated. (b) A plot of $\widehat{N}_{mid}$ to examine bias arrising from mispecify the sex ratio in the model. The dashed line denotes the true value of N_{mid}. (c) Precision of $\widehat{N}_{mid}$ from data based on simulated populations with various true mortality rates. All results in these plots based on 1000 simulated populations with N_0 = 2280 and rate of increase = 6.4%.

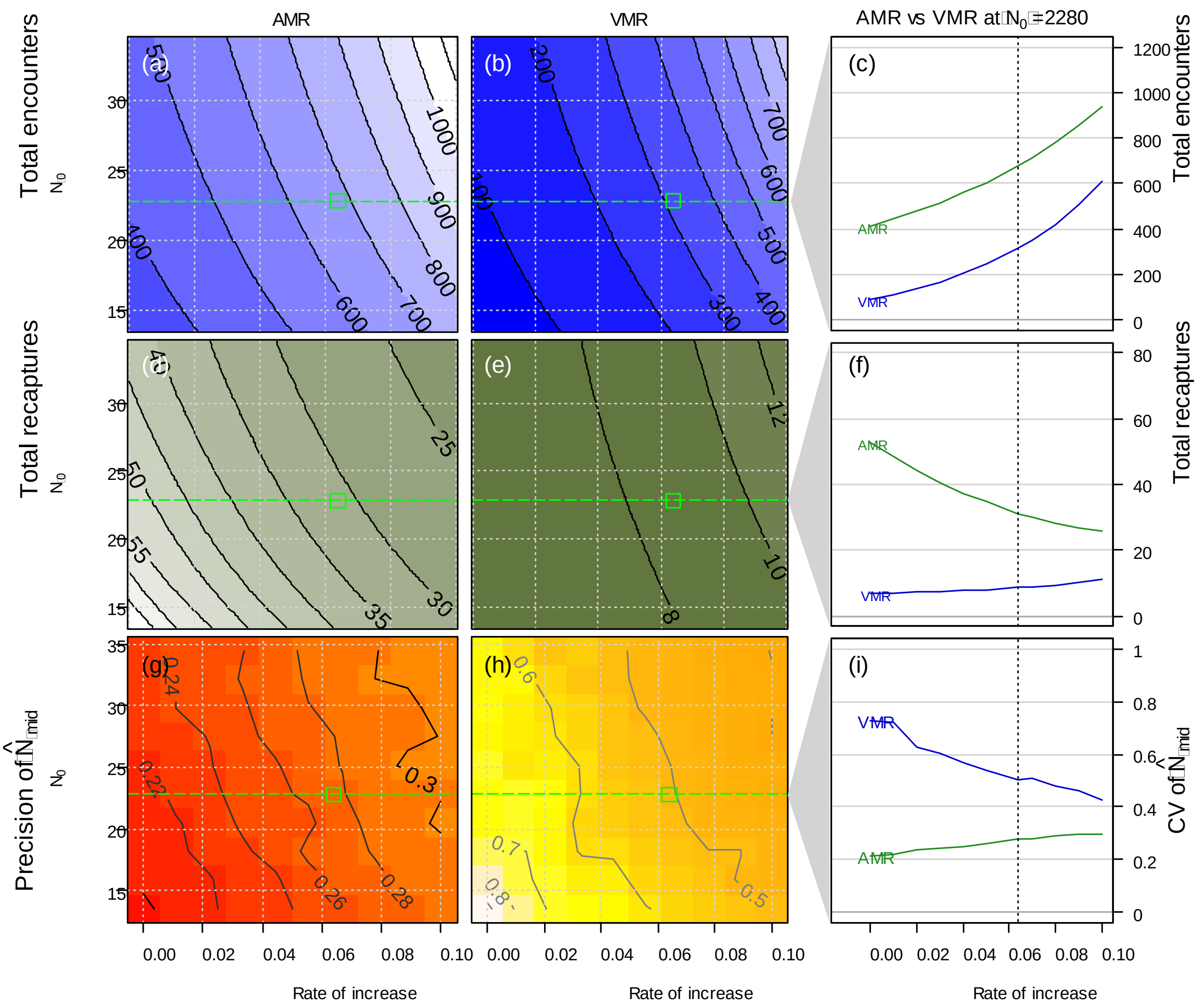

Figure 4: Predictions from the simulation study for a 12 year single vessel every season acoustic assisted mark-recapture survey (AMR) and a visual-only mark-recapture survey (VMR). For populations generated with a range of N_0 and r (the current best estimate is indicated by the square). Plots (a) and (b) shows the total number of encounters (d) and (e) the total number of recaptures and (g) and (h) the precision (CV) of the mid-survey circumpolar abundance (N_{mid}). Plots (c), (f) and (i) show the comparison between AMR and VMR for a range of r and N_0 fixed at 2280 (corresponding to the dashed vertical line in the other contour plots). The results for the IWC agreed estimates of N_0 and r are indicated by the small squares in the contour plots, and the dotted vertical line in plots (c), (f) and (i).

Survey effort design

In terms of survey design the results presented so far have been for a single vessel conducting a 12 year annual survey programme. To investigate other options we calculated the precision of $\widehat{N}_{mid}$, at the base case, arising from survey programme length programmes running between 2 and 12 years, and with 1 to 6 vessels each season (Fig. 5a). As expected, the longer the survey programme, the greater the precision of the abundance estimate. These results can be used to gain a rough indication of the choice between programme length and number of vessels to obtain a target CV (0.2 – 0.3 denoted in the Fig. 5a). However, it should be noted that, since mid-programme abundance is being considered, the different length programmes are providing abundance estimates for different years (see top axis of Fig. 5a).

In practice, regular, long-term annual surveying can be difficult to maintain. Therefore, we also investigated more complex programme designs involving irregular sampling regimes. We applied a range of irregular sampling strategies (Table 2), each 12 years in length, with a total effort of 12 vessel seasons and estimated the precision of N_{mid} (Fig. 5b). It was found that as long as the programme consisted of a number of surveys (e.g., four surveys each with three vessels) within the 12 years, precision was fairly robust to how effort was distributed (i.e., which years the surveys were conducted).

Table 2: Summary of effort regimes examined.

Scenario	Description
Annual	Even annual sampling (i.e., one vessel surveying ever year for 12 years)
Bookend	Placing all effort in the first and last year of the programme (i.e., six vessels year one, and six vessels year 12)
Triennial	Surveying every three years (i.e., four vessels every three years)
Quadrennial	Surveying every four years (i.e., three vessels every four years)
Increasing	Surveying every four years but gradually increasing effort over the life of the programme
Decreasing	Surveying every four years but gradually decreasing effort over the programme
Late	Surveying only in the last four years of the programme (i.e. three vessels over four years)
Big-Start	Placing all the effort in the first year and matching to the existing catalogue (i.e., 12 vessels in the final year)
Big-Finale	Placing all the effort in the final year and matching to the existing catalogue (i.e., 12 vessels in the final year)

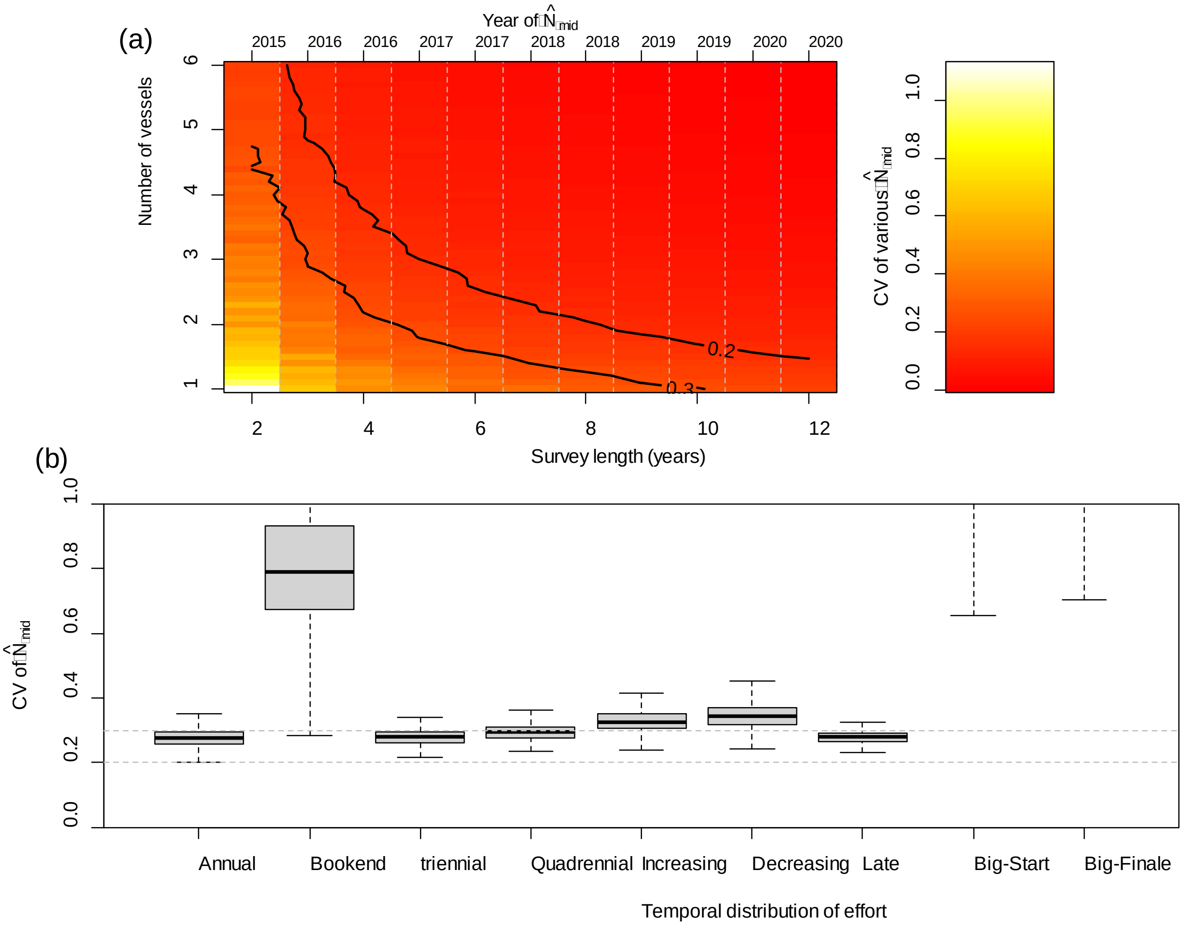

Figure 5: The predicted effect on the CV of $\widehat{N}_{mid}$ for various other programme design scenarios (a) from the length of the study, assuming even annual sampling (b) for the nine survey regimes described in Table 2 (based on 1000 simulated populations with base case parameters).

Spatial structure testing

To examine the issue of spatial heterogeneity, we ran the simulation with spatial structure as described in the methods section (Fig. 6). Specifically, we looked for the situation where there exists spatial structure that is undetectable but is strong enough to cause significant bias. The bias (Fig. 6a) is most pronounced when animals show strong site fidelity (P_H = 0.8 – 1.0) combined with uneven survey effort (E_E = 80% – 90%). Spatial structure (i.e., detected with a χ^2 *p*-value < 0.05) is found only when survey effort mix is greater than 80% in one region, and this threshold increases as site fidelity decreases (Fig. 6c). So the 'danger' situation (i.e., bias of $\widehat{N}_{mid}$ > 15% and where the χ^2 test does not detect the heterogeneity) only occurs for survey effort mix > 85% and site fidelity > 75% (Fig. 6b). So, in this simple test case, as long as the survey programme can guarantee reasonable even effort coverage of sub-populations, there would be no cause for concerned. In Antarctic blue whales this would most likely be achieved by assuring there is reasonable longitudinal coverage proportional to whale density.

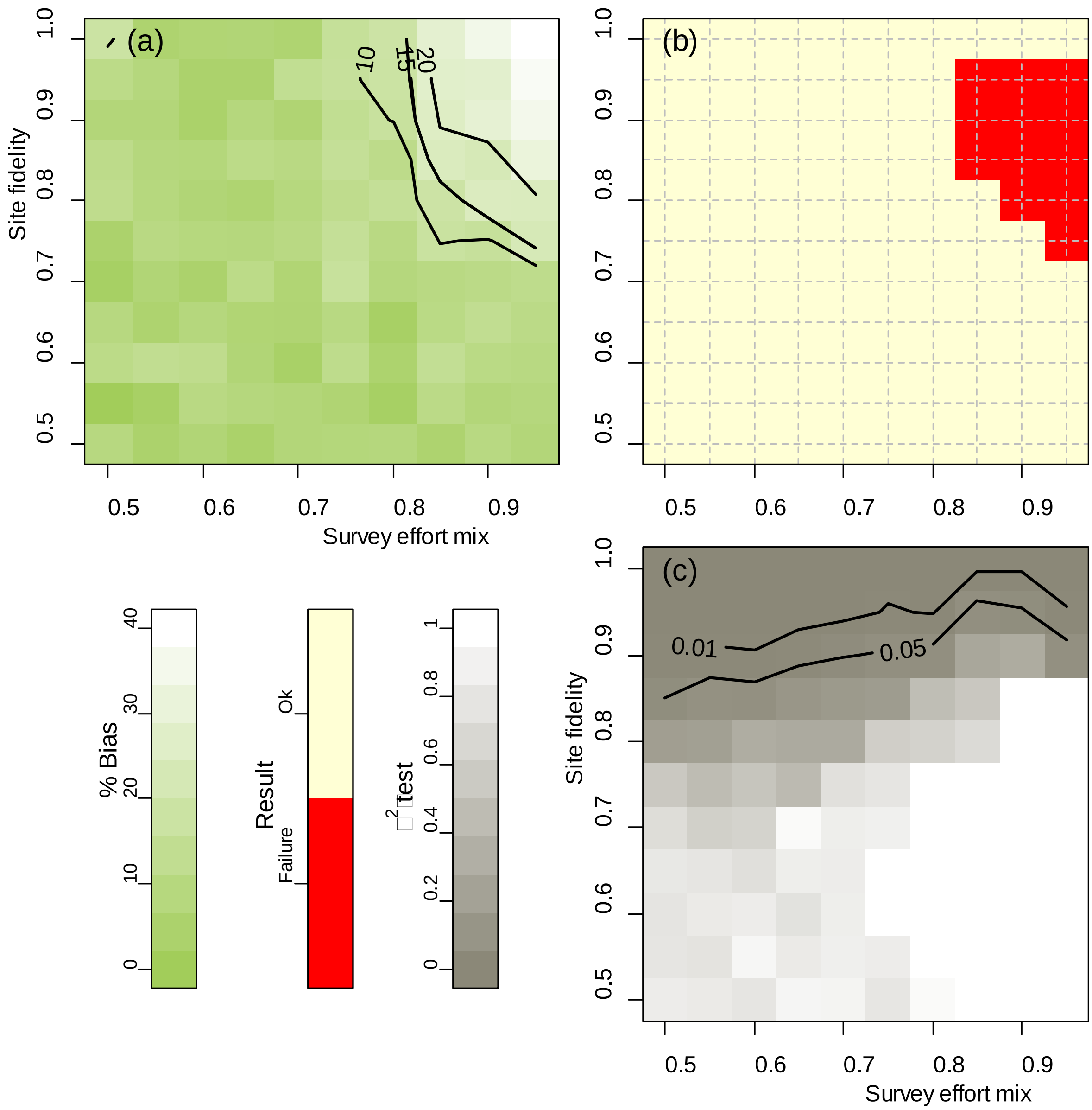

Figure 6: Plot of simulated spatial heterogeneity parameter space with (a) bias of $\widehat{N}_{mid}$ (c) the result of the χ^2 test and (b) areas where the bias of the estimate $\widehat{N}_{mid}$ > 15% and the χ^2 test does not detect the heterogeneity.

DISCUSSION

This analysis established that a mark-recapture approach would be a good candidate to estimate circumpolar Antarctic blue whale abundance. For the expected Antarctic blue whale population trajectory of N_0 = 2280 and r = 6.4% (International Whaling Commission 2009), the predicted precision of 0.28 from an acoustic-assisted mark-recapture programme was below our acceptable target of 0.3. However, it was clearly demonstrated that for mark-recapture to be feasible, surveys would require acoustic-assistance to track animals, given that visual mark-recapture did not achieve reasonable CV under any of our scenarios.

In terms of survey duration, a minimum programme length of 12 years was indicated. However, upon investigation of different effort distributions over the duration of the survey programme (see Fig. 5) it was found surveys do not necessarily need to be conducted every year (see Fig. 5b), but rather can be ran at other intervals. The main requirement is that there a minimum number of surveys, rather than all effort being concentrated into 1 or 2 years. This arises as mark-recaptures precision is dependent on the number of recaptures, which is proportional to sample size (or effort), but also the number of temporal 'sampling events', i.e., surveys; as each survey adds another set of data to do a mark-to-recapture comparison. Schweder and Sadykova (2009) looked at a similar question of distributing temporal effort and found benefit in applying substantial effort at the end of the survey programme.

A virtue of the mark-recapture framework is it can produce an estimate of the population rate or increase (r). However, the estimate was found to be extremely imprecise in this study. Upon inspection it appears the population trajectories estimated from mark-recapture in this application are highly variable, pivoting around the mid-survey abundance. Hence, estimates of N_0 and r are imprecise, but the mid-survey year abundance estimate $\widehat{N}_{mid}$ is precise. Therefore, caution is required when interpreting the point estimates of N_0, r and M from any future survey programme. If a population growth estimate is needed, a population model including previous line transect estimates may be more appropriate.

Characteristics of Precision

The precision of an abundance estimate from an acoustic-assisted mark-recapture survey was not as sensitive as visual-based mark-recapture to our assumed population parameters (N_0 and r), but did show some loss of precision for population-level parameter combinations that resulted in larger overall population sizes. Whereas, visual-only mark-recapture steadily gained precision for larger population parameters. This is explained by examining the number of recaptures (Fig. 5d and 5e) from each method, which directly effects precision in mark-recapture. The number of recaptures depends on the probability of recapture and the sample size, which both in turn depend on the population size. The probability of recapture is inversely proportional to N, whereas sample size is proportional to N. Therefore, the number of recaptures (and hence precision) will be determined by the overall effect of these two counteracting drivers. Hence, the effect of larger r and N_0, will be based on the balance between the resulting decrease in probability of recapture and increase in sample size. It is clear that for population parameters determining larger population sizes, the number of recaptures in acoustic assisted mark-recapture decreases, and visual-only recaptures increase. This difference in behaviour is due to the effect of tracking and marking time in this study: as r increases, the gain in sample size is hindered by the need to mark ever increasing numbers of animals. So, the detrimental effects from the reduction in the probability of recapture outweigh the improvement of larger sample size and, hence, the precision decreases.

It should be noted that in all the variance calculations in this paper, uncertainty in the sex-ratio value was not propagated through to the final CVs. However, given the sample sizes that the sex-ratio value is based on, and the consistency between foetal and adult estimates (see Table 1), it seems reasonable to assume sex-ratio uncertainty would have minimal contribution to the final variances.

As expected, when we investigated the effect of true population mortality (Fig. 3), higher mortality decreased precision. This occurs because with higher mortality, more animals die during the programme and are not available for subsequent recapture. However, in all cases, the precision was still within our target CV range. Obviously, we have considered a range of true mortality values for the base case (i.e., orthogonally, and in isolation to the range of possible parameters N_0 and r) and with a combination of high N_0, r, and M, the precision would be worse than we report. However, this situation is unlikely, especially given the relationships between parameters, e.g., given marine mammal birth rates, both extremely high population growth and high mortality occurring is incongruent.

Assumptions

As well as the standard assumptions behind mark-recapture methods, any model, or analysis, is based on assumptions that may not be consistent with reality. Generally, the assumptions made in analyses within this paper are reasonable given the current knowledge of Antarctic blue whales.

For example:

We assumed that the proportion of pygmy blue whales (*B. musculus brevicauda*) in higher latitudes is small enough to be ignored. Recent results indicate there may be slightly more pygmy blue whales, or blue-pygmy whale hybrids, in the Southern Ocean than previously assumed (Attard *et al.* 2012). Unlike line transect, where this situation could bias abundance estimates, with a mark-recapture, the main repercussion would be wasted survey time tracking and marking the wrong species. Subsequent genetic analyses should identify the correct species, resulting in no bias.

No consideration was given to variations in the longitudinal distribution of Antarctic blue whales. Distribution of sightings during SOWER surveys suggests some patchiness in densities longitudinal distribution. This information could be used to improve a design for mark-recapture, where targeting areas of putative higher densities could boost the sample size and recapture rate.

The individual-based simulation to predict mark-recapture sample size is based on many assumptions (see Supplementary material). The assumptions used in the simulation generally erred on the conservative side. The 2013 pilot study had slightly fewer encounters than would be expected according to the mean encounter rate predicted by the model (Fig. S2); although this realised encounter rate was well within the range of simulated values. However, the number of final genotypes obtained in the 2013 pilot study was much lower than predicted by the simulation. This could possibly be because the simulation assumed that for every whale found, an attempt to biopsy was made, and in the pilot study this was not the case. In terms of the variance calculations for the future programme, it is reasonable to assume rates of biopsy attempts will be higher in the main programme than the pilot study. If this is not the case, then the fact that in any analysis of actual data, a multiple catalogue model will be used (i.e., photographic catalogues will be included), so samples sizes will increase to be closer to what was predicted.

Overall, it was apparent that ongoing analysis and monitoring of the data would be required to check assumptions (e.g., test for heterogeneity) and the modelling adjusted as needed.

Bias and Spatial Heterogeneity in Capture Probabilities

In terms of spatial heterogeneity (Fig. 6), it was found that, as expected, strong spatial population structure will produce bias, but that it should be reasonably straightforward to detect by examining the spatial location of marks and recaptures. At the other end of the spectrum, we found more subtle spatial structure was difficult to detect, but the effect on bias will be minimal. Hence, in both cases, spatial heterogeneity is manageable. However, there is potential area of concern between these two extremes, where spatial structure is pronounced enough to incur significant bias, but not enough to be detectable. We attempted to quantify this issue in a simple simulated case. It was found that the problematic area did exist, but only when effort was distinctly un-even in its distribution across subpopulations. This indicates that any survey programme should attempt to have reasonable coverage across any putative sub-populations, proportional to the sub-population size. Obviously, the simulation was a simple artificial example, but it provides some guidance to how this issue may present itself and be handled.

Other Considerations

Although precision is a fundamental aspect of judging the performance of a survey method, there are other satellite considerations that may influence attractiveness of a particular method. Mark-recapture is relatively easy and simple to implement; does not require defined track lines (beyond considerations for ensuring homogeneity in capture probabilities), and can focus instead on higher density regions; and, when using genetics in particular, can aid in understanding individual movements and population structure. Mark-recapture is also robust to sampling platform, which is relevant given the proposal to use vessels from many nations and other platforms of opportunity (Williams, Hedley & Hammond 2006). Finally, continually adding to biopsy and photograph-identification databases produces a legacy, where, due to the longevity of blue whales, these samples can, over the coming decades, contribute to newer abundance estimates.

Management and Conservation implications

The obvious implication from this work for management and conservation is the finding that for a future monitoring programme, mark-recapture is a suitable method to estimate circumpolar Antarctic blue whale abundance. Furthermore, a number of other aspects of the results in this paper will inform management/conservation of Antarctic blue whales and an associated monitoring program: most importantly, the repercussions of certain survey planning decisions on precision (e.g., survey programme length, number of vessels, temporal and spatial distribution of effort, and the need prioritise collection of biopsy data over photograph-ID). In particular, given the difficultly of securing survey resources continually for 12 years, the results related to surveying with a number of vessels every few years is relevant and useful.

The exercise of developing a mark-recapture model for the simulation also gave insight into some of the issues that need to be considered in future implementation, e.g., since data on the sex of each sample makes the handling of the sex sampling bias in the analysis much simpler, it is strongly recommended that where possible all animals should be biopsied.

Conclusion

Given the expense and effort required to survey wild populations, pre-survey studies such as presented in this paper are important for planning and, more fundamentally, validating that the planned programme can theoretically meet its aims and answer the questions being asked. Overall, this work provides a powerful tool for early stage planning and decision making of Antarctic blue whale monitoring. Beyond blue whales, and cetaceans in general, we believe this type of analysis and study is warranted in the early stages of many potential mark-recapture studies.

ACKNOWLEDGEMENTS

We would like to thank the crew and scientists on the SORP Antarctic Blue Whale Project 2012 pilot study and the 2013 Antarctic voyage for the information and data that informed this work. Bran Miller, Bill de la Mare and Victoria Wadley (AAD) for valuable advice and comments on this work. Trevor Branch (University of Washington) for answering questions our many SOWER and blue whale questions along the way. Angela

Sremba and Scott Baker (Oregan State University) for answering questions about existing genetic catalogues. David Smith (AAD) for GIS expertise in producing the study area map. Spatial data was provided by the Australian Antarctic Data Centre and © Scientific Committee on Antarctic Research 1993-2012.

REFERENCES

Attard, C.R.M., Beheregaray, L.B., Jenner, K.C.S., Gill, P.C., Jenner, M.-N., Morrice, M.G., Robertson, K.M. & Möller, L.M. (2012) Hybridization of Southern Hemisphere blue whale subspecies and a sympatric area off Antarctica: impacts of whaling or climate change? *Molecular Ecology,* **21,** 5715-5727.

Bonner, S.J. & Holmberg, J. (2013) Mark-Recapture with Multiple, Non-Invasive Marks. *Biometrics,* **69,** 766-775.

Borchers, D.L., Buckland, S.T. & Zucchini, W. (2002) *Estimating Animal Abundance: Closed Populations*. Springer-Verlag, Great Britain.

Branch, T.A. (2007) Abundance of Antarctic blue whales south of 60ºS from three complete circumpolar sets of surveys. *Journal of cetacean research and management,* **9,** 253-262.

Branch, T.A. & Butterworth, D.S. (2001) Estimates of abundance 60°S for cetacean species sighted frequently on the 1978/79 to 1997/98 IWC/IDCR-SOWER sighting surveys. *Journal of cetacean research and management,* **3,** 251-270.

Branch, T.A., Matsuoka, K. & Miyashita, T. (2004) Evidence for increases in Antarctic blue whales based on Bayesian modelling. *Marine Mammal Science,* **20,** 726-754.

Branch, T.A., Matsuoka, K., and Miyashita, T. (2004) Evidence for increases in Antarctic Blue whales based on Bayesian modelling. *Marine Mammal Science,* **20,** 726-754.

Branch, T.A., Stafford, K.M., Palacios, D.M., Allison, C., Bannister , J.L., Burton, C.L.K., Cabrera, E., Carlson, C.A., Galletti vernazzani, B., Gill, P.C., Hucke, R., Jenner, K.C.S., Jenner, M.-N.M., Matsuoka, K., Mikhalev, Y.A., Miyashita, T., Morrice, M.G., Nishiwaki, S., Sturrock, V.J., Tormosov, D., Anderson , R.C., Baker, A.N., Best, P.B., Borsa, P., Brownell JR, R.L., Childerhouse, S., Findlay, K.P., Gerrodette, T., Ilangakoon, A.D., Joergensen, M., Kahn, B., Ljungblad, D.K., Maughan, B., Mccauley, R.D., Mckay, S., Norris, T.F., Group, O.w.a.D.R., Rankin, S., Samaran, F., Thiele, D., Van Waerebeek, K., and & Warneke, R.M. (2007a) Past and present distribution, densities and movements of blue whales *Balaenoptera musculus* in the Southern Hemisphere and northern Indian Ocean. *Mammal Review,* **37,** 116-175.

Branch, T.A., Stafford, K.M., Palacios, D.M., Allison, C., Bannister, J.L., Burton, C.L.K., Cabrera, E., Carlson, C.A., Vernazzani, B.G., Gill, P.C., Hucke-Gaete, R., Jenner, K.C.S., Jenner, M.N.M., Matsuoka, K., Mikhalev, Y.A., Miyashita, T., Morrice, M.G., Nishiwaki, S., Sturrock, V.J., Tormosov, D., Anderson, R.C., Baker, A.N., Best, P.B., Borsa, P., Brownell, R.L., Childerhouse, S., Findlay, K.P., Gerrodette, T., Ilangakoon, A.D., Joergensen, M., Kahn, B., Ljungblad, D.K., Maughan, B., McCauley, R.D., McKay, S., Norris, T.F., Whale, O., Rankin, S., Samaran, F., Thiele, D., Van Waerebeek, K. & Warneke, R.M. (2007b) Past and present distribution, densities and movements of blue whales *Balaenoptera musculus* in the Southern Hemisphere and northern Indian Ocean. *Mammal Review,* **37,** 116-175.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. & Thomas, L. (2001) *Introduction to Distance Sampling*. Oxford University Press, Great Britain.

Carroll, E.L., Patenaude, N.J., Childerhouse, S.J., Kraus, S.D., Fewster, R.M. & Baker, C.S. (2011) Abundance of the New Zealand subantarctic southern right whale population estimated from photo-identification and genotype mark-recapture. *Marine Biology,* **158,** 2565-2575.

Constantine, R., Jackson, J.A., Steel, D., Baker, C.S., Brooks, L., Burns, D., Clapham, P., Hauser, N., Madon, B., Mattila, D., Oremus, M., Poole, M., Robbins, J., Thompson, K. & Garrigue, C. (2012) Abundance of humpback whales in Oceania using photo-identification and microsatellite genotyping. *Marine Ecology Progress Series,* **453,** 249-261.

Cooke, J.G. (1995) The international whaling commission's revised management procedure as an example of a new approach to fishery management. *Developments in Marine Biology,* **4,** 647-657.

Cormack, R.M. (1989) Log-Linear Models for Capture-Recapture. *Biometrics,* **45,** 395-413.

Cubaynes, S., Pradel, R., Choquet, R., Duchamp, C., Gaillard, J.M., Lebreton, J.D., Marboutin, E., Miquel, C., Reboulet, A.M., Poillot, C., Taberlet, P. & Gimenez, O. (2010) Importance of Accounting for Detection Heterogeneity When Estimating Abundance: the Case of French Wolves. *Conservation Biology,* **24,** 621-626.
Double, M.C., Miller, B.S., Kelly, N., Leaper, R., Andrews-Goff, V., Schmitt, N., Wadley, V. & Lindsay, M. (2013) Antarctic Blue Whale Voyage 2013: Science Plan.
Double, M.C.B., J; Miller, B; Olson, P; Andrews-Goff, V; Leaper, R; Ensor, P; et al. (2013) Cruise report of the 2013 Antarctic blue whale voyage of the Southern Ocean Research Partnership pp. 1-16. Jeju Island, Republic of Korea, SC/65a/SH21.
Evans, P.G.H. & Hammond, P.S. (2004) Monitoring cetaceans in European waters. *Mammal Review,* **34,** 131-156.
Hammond, P.S. (1986) Estimating the size of naturally marked whale populations using capture-recapture techniques. pp. 253-282.
Hedley, S.L. & Buckland, S.T. (2004) Spatial models for line transect sampling. *Journal of Agricultural Biological and Environmental Statistics,* **9,** 181-199.
International Whaling Commission (2009) Report of the Scientific Committee. Annex H. Report of the sub-committee on other Southern Hemisphere whale stocks. *Journal of Cetacean Research and Management. (Suppl.),* **11,** 220-247.
Leaper, R. & Miller, C. (2011) Management of Antarctic baleen whales amid past exploitation, current threats and complex marine ecosystems. *Antarctic Science,* **23,** 503-529.
McClintock, B.T., Conn, P.B., Alonso, R.S. & Crooks, K.R. (2013) Integrated modeling of bilateral photo-identification data in mark-recapture analyses. *Ecology,* **94,** 1464-1471.
Olson, P.A. (2010) Blue whale photo-identification from IWC IDCR/SOWER cruises 1987/1988 to 2008/2009. pp. 6 pp.
Olson, P.A., Ensor, P., Schmitt, N., Olavarria, C. & Double, M.C. (2013) Photo-identification of Antarctic blue whales during the SORP Antarctic Blue Whale Voyage 2013. *Paper SC/65a/SH11 presented to the IWC Scientific Committee*.
Peel, D., Miller, B.S., Kelly, N., Dawson, S., Slooten, E. & Double, M.C. (In Press) A simulation study of acoustic-assisted tracking of whales for mark-recapture surveys *PLoSOne*.
Reilly, S.B., Bannister, J.L., Best, P.B., Brown, M., Brownell Jr., R.L., Butterworth, D.S., Clapham, P.J., Cooke, J., Donovan, G.P., Urbán, J. and Zerbini, A.N. (2008) Balaenoptera musculus ssp. intermedia. *IUCN 2010. IUCN Red List of Threatened Species. Version 2010.2*.
Rew, M.B., Robbins, J., Mattila, D., Palsboll, P.J. & Berube, M. (2011) How many genetic markers to tag an individual? An empirical assessment of false matching rates among close relatives. *Ecological Applications,* **21,** 877-887.
Reynolds, J.H., Thompson, W.L. & Russell, B. (2011) Planning for success: Identifying effective and efficient survey designs for monitoring. *Biological Conservation,* **144,** 1278-1284.
Schwarz, C.J. & Seber, G.A.F. (1999) Estimating animal abundance: Review III. *Statistical Science,* **14,** 427-456.
Schweder, T. & Sadykova, D. (2009) Information is gained increasingly fast in capture-recapture surveys-bowheads assessed by photographic surveys, and minke whales by genetic marking. *Paper SC/61/SH21 presented to the IWC Scientific Committee*
Sears, R. & Perrin, W.F. (2008) Blue whale (Balaenoptera musculus). *Encyclopedia of Marine Mammals*, pp. pp 120-124.
Sremba, A.L., Hancock-Hanser, B., Branch, T.A., LeDuc, R.L. & Baker, C.S. (2012) Circumpolar diversity and geographic differentiation of mtNDA in the critically endangered Antarctic blue whale (Balaenoptera musculus intermedia). *Plos One,* **7**.
Williams, R., Hedley, S.L. & Hammond, P.S. (2006) Modeling distribution and abundance of Antarctic baleen whales using ships of opportunity. *Ecology and Society,* **11**.
Williams, R. & Thomas, L. (2009) Cost-effective abundance estimation of rare animals: Testing performance of small-boat surveys for killer whales in British Columbia. *Biological Conservation,* **142,** 1542-1547.

APPENDIX A: DETAILS OF MARK-RECAPTURE MODEL

We now describe the simple mark-recapture model on which we based our variance prediction. The model assumes a Poisson distribution of recaptures (Cormack 1989) and does not consider within-year recaptures. The probability that any individual animal is caught in year *t* is given by,

$$p_t = \frac{n_t}{N_t}$$

where n_t is the number of animals captured each year *t* and N_t is the population size. Given that the size of the Antarctic blue whale population remains at a small proportion of its original size, it is reasonable to assume any increases in abundance would still be exponential (i.e. density independent),

$$N_t = N_0 \cdot (1 + r)^t$$

where *r* is the annual rate of increase and N_0 is the abundance at the beginning of the study period (i.e., $t = 0$).

The expected number of recaptures in year t_2 of animals marked in year t_1 is given by

$$\hat{m}_{t_1,t_2} = p_{t_1} p_{t_2} N_{t_1} e^{-M(t_2 - t_1)}$$

where *M* is a specified mortality rate. So the likelihood contribution for the number of recaptures between years t_1 and t_2, assuming a Poisson distribution, is

$$L = \frac{\left(\hat{m}_{t_1,t_2}\right)^{m_{t_1,t_2}} e^{-\hat{m}_{t_1,t_2}}}{m_{t_1,t_2}}$$

And hence the log-likelihood is then,

$$-\log_e L \propto \sum_{t_1=} \sum_{t_2=} \left[-m_{t_1,t_2} \log\left(\hat{m}_{t_1,t_2}\right) + \hat{m}_{t_1,t_2} \right]$$

From this model N_0 , *r* and *M* can be estimated by maximizing the likelihood. The variances of the estimates can be calculated numerically from the inverse Hessian of the negative log likelihood.

As discussed in the main text, if sex typing of samples is available, the sex-based sampling bias can easily be accommodated in the model. To do this we treat the data from the male and female animals separately and introduce a population sex ratio parameter, λ, that can be estimated (or fixed if known), such that

$$N_0^{(Female)} = \lambda N_0$$
$$N_0^{(Male)} = (1 - \lambda) N_0$$

Then the N_t in the likelihood calculations are as before, e.g.,

$$N_t^{(Female)} = N_0^{(Female)} \cdot (1 + r)^t$$

This solves the issue of unequal sex sampling and allows both data to contribute to the estimation of the common parameters N_0, *M* and *r*.

SUPPLEMENTARY MATERIAL A: MARK-RECAPTURE SAMPLE SIZE PREDICTION

To provide an indicative mark-recapture survey sample size we used the R package WATS (Whale Acoustic Tracking Simulation) available from:

sourceforge.net/projects/watspackage/

and described in {Peel, In Press #6119}. The package simulates acoustic-assisted tracking of whales using an individual-based simulation framework. Specifically, for a given whale density, whales were randomly placed in a survey region and an acoustic-assisted mark-recapture survey was simulated based on search protocols/rules and assumed acoustic properties.

When applying the simulation to model a future Antarctic survey many of the parameters were unknown. So information was used from a pilot study {Double, 2013 #6490} as well as information from previous blue whale studies (see Table 2 in {Peel, In Press #6119).

The simulation was run at various assumed population densities (See Fig. S2) with each point based on 1000 replicated populations/surveys. The reported encounter rates are in terms of groups per hour, over the full survey period (i.e., this includes night and bad weather). It was found that an acoustic-assisted survey would be expected to produce approximately a 1.7-3.0 fold increase in encounter rate, over a traditional line transect survey, at predicted Antarctic blue whale densities. This gain arises from the increased range of acoustic methods over visual-detection, as well as the benefit from the ability to track overnight and in adverse weather conditions, countered by the reduced on-effort time in mark-recapture due to time spent biopsying animals.

For the mark-recapture variance calculations in the main paper, we interpolated the encounter data (in Fig. S1) to provide an encounter rate for any given whale density and then multiply by adjustments for expected group size, within-season resights, biopsy success rate and voyage length to get sample sizes (Fig. S2).

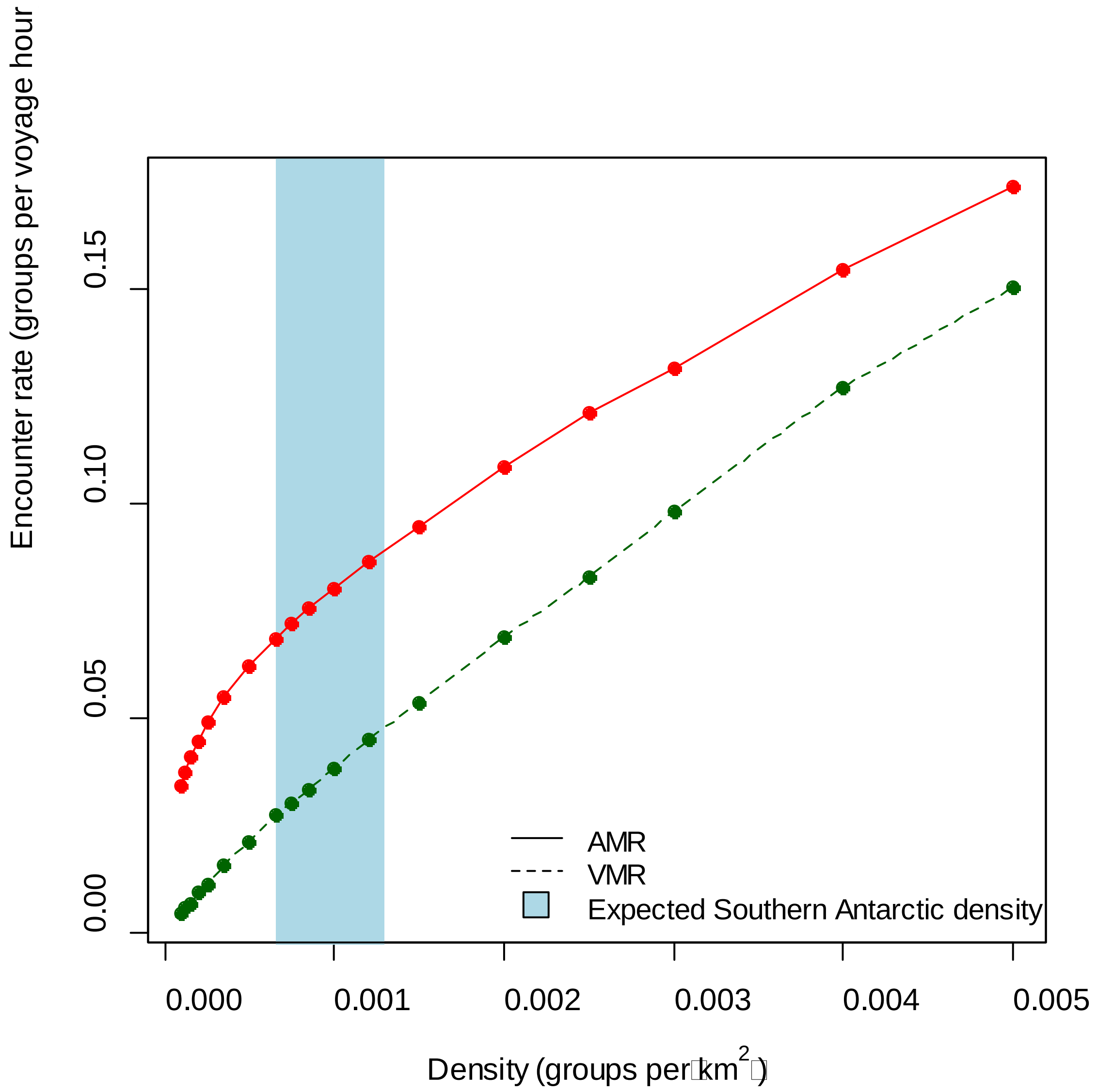

Fig. S1 Simulated whale group encounter rates for a given whale density.

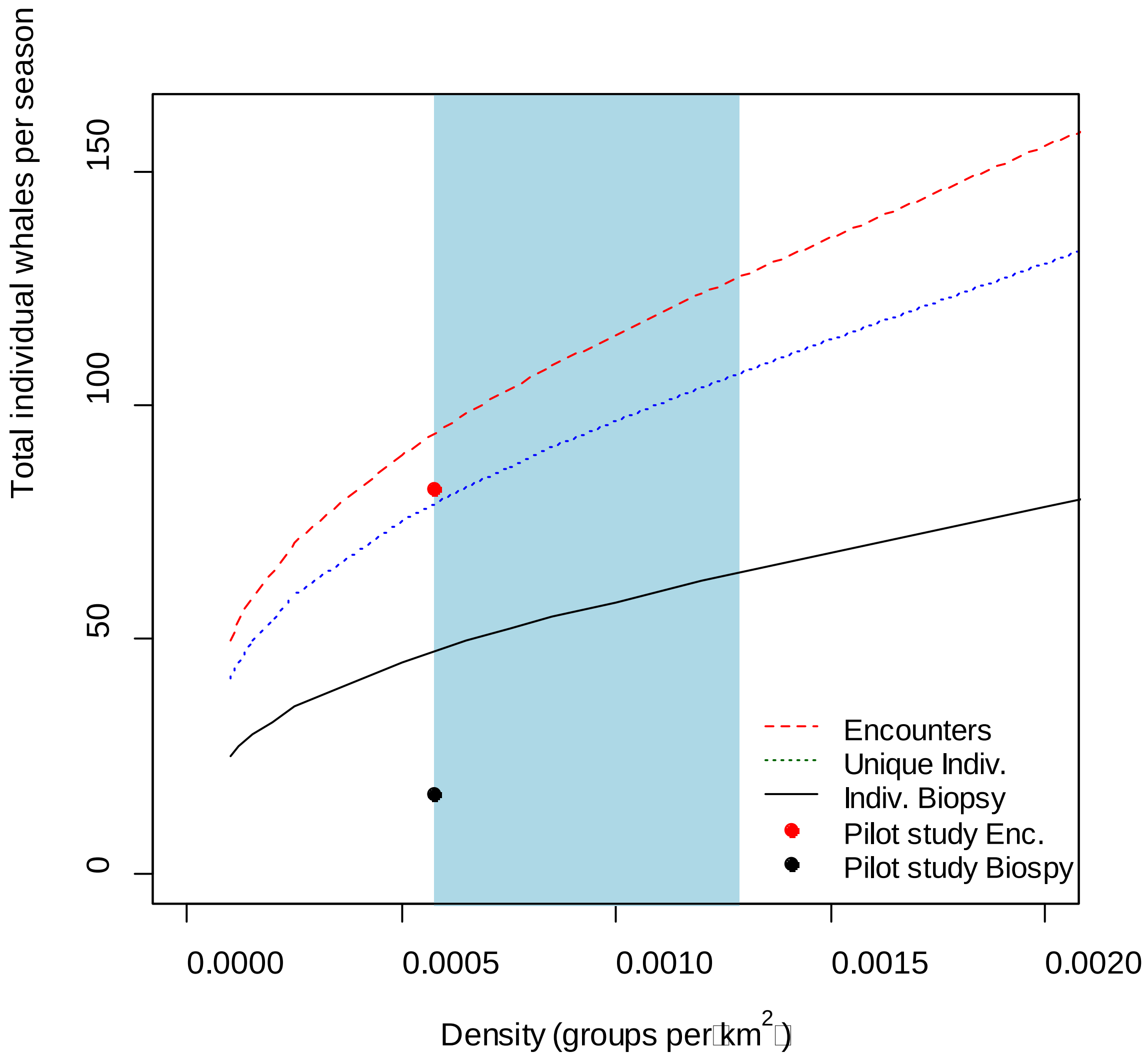

Fig. S2 Simulated total sample size for 37.5 days of effort for a given whale density.