Antarctic Blue Whale surveys: augmenting via genetics for close-kin and ordinal age

Mark V. Bravington[1], Simon N. Jarman[2], Hans J. Skaug[3]

[1]: CSIRO Marine Lab., Hobart, Tas. 7000, Australia

[2]: Australian Antarctic Division, Kingston, Tas 7050, Australia

[3]: Department of Mathematics, P. O. Box 7800, 5020 Bergen, Norway

April 2014

Acknowledgments

Thanks to Bill de la Mare for suggesting this particular application of epigenetic age estimation.

Abstract

The biopsies from a proposed genetic mark-recapture study for Antarctic Blue Whales could also be used to identify parent-offspring pairs. These "recaptures" can be accommodated in an extended mark-recapture model, which would dramatically improve the precision of abundance estimates from the study. Also, with parent-offspring pairs, the model could be made robust against bias arising from unmodelled heterogeneity. Precision could be improved further using epigenetic age data to tell which animal is the parent and which the offspring.

Introduction

SORP has proposed a multi-year mark-recapture survey programme to estimate the abundance of Antarctic Blue Whales. Paper *IWC/65/SHxxx* examines the feasibility, in terms of the amount of survey effort and timespan involved. It concludes that reasonable CVs on 2020 abundance (25%-30%) could be obtained in a 12-year biopsy-based survey, using an open-population model in which birth and death rates are also estimated.

This paper shows how the precision of the abundance estimate, and of the birth and death rates, can be dramatically improved from exactly the same tissue samples at minimal extra cost, using two newly-developed approaches: Close-Kin Mark-Recapture (CKMR; Bravington et al., 2014), enhanced by relative age estimates inferred from methylation levels across the genome (Polanowski et al., 2014). Specifically, the genotype of each tissue sample is used to establish the animal's sex, and is then compared with the genotype of every other tissue sample to see whether (i) the two animals are the same, as in standard genetic mark-recapture, or (ii) the two are a Parent-Offspring Pair (POP), or (iii) neither of the above. If (ii), then further genetic analyses, including approximate ageing-by-methylation, can show which of the two is the Parent and which is the Offspring.

POPs will be much more common than individual recaptures, and provide a great deal of extra demographic information that can be incorporated easily into an extended mark-recapture framework. In addition, with POPs it is possible to make the abundance estimate robust against individual Heterogeneity of Capture Probability (HCP, a.k.a. "unmodelled heterogeneity") due to, say, shyness; this phenomenon is otherwise a well-known potential source of significant bias in MR abundance estimates, and one that is largely unresolvable when— as for ABW— multiple recaptures are unlikely. POPs also yield straightforward results about stock structure over a single-generation timescale, without the limitations of conventional population genetics or tagging.

In the rest of this paper, section 1 briefly introduces the survey approach, the background to genetic ageing and its feasibility for ABWs, and the way that close-kin genetics can be used in this particular MR setting. Section 2 shows likely precisions for a couple of possible survey designs and scenarios about the true abundance and demography, with and without various levels of information about POPs. Section 3 gives a very brief review of the different technologies that can be used for genotyping— a rapidly-changing field— and considers which are likely to be best for a fairly long-term study on ABW. Our overall conclusions are given in section 4.

1 Methods

1.1 Survey background

There are no reliable current estimates for the abundance of Antarctic blue whales, the species most heavily depleted by 20th century Antarctic whaling operations. The most recent estimates come from the IDCR/SOWER circumpolar abundance surveys, with an abundance of 2280 (CV=0.36) in 1998 (Branch et al., 2007). ABWs were not commonly encountered even in late-period SOWER (early 2000s) when abundance was presumably highest, and a purely visual-based survey would not generate enough samples for mark-recapture to work. Line-transect abundance estimates would similarly require very high effort levels and infeasible logistic rigidity about platforms and designs.

However, encounter rates for ABW can be boosted considerably by acoustic tracking, using sonobuoys to zoom in on calling males and nearby females from distances of many 10s of kilometres, as successfully demonstrated in a 2013 pilot study (Double et al., 2013). Peel et al. (2014) show that an acoustically-assisted mark-recapture program of genetic mark-recapture from biopsies, running at lower effort levels than SOWER but for at least a decade, should generate enough samples (typically 500-800 in most scenarios examined) and individual recaptures (typically 30-40) to allow a MR estimate of abundance with reasonable CV (typically 25-30%). That paper considers some of the issues that would need addressing (impact of spatial fidelity, etc.), concluding that there are no fundamental impediments provided that enough ship time is available. Nevertheless, the number of recaptures expected is still fairly low, the likely precision is not spectacular, and the inferences would still be somewhat vulnerable to unmodelled heterogeneity in capture probability. It is worth considering whether these aspects could be addressed by using the samples in other ways.

1.2 Genetic age estimation via methylation

Age-related changes in DNA methylation have been recorded for a range of mammalian genes expressed in skin tissue (Maegawa et al., 2010; Grönniger et al., 2010; Bocklandt et al., 2011; Polanowski et al., 2014). Specific cytosines have been found to become either more or less methylated with age in a process thought to be part of the regulated genetic program of development and senescence. Many of the methylation changes

recorded in human and mouse genes have a linear relationship with age (Maegawa et al., 2010; Grönniger et al., 2010; Bocklandt et al., 2011; Koch and Wagner, 2011) and methylation levels can be measured with reasonable precision by parallel sequencing methods (Grönniger et al., 2010; Polanowski et al., 2014). This aspect of the molecular ageing process is consequently useful for developing proxy assays for age. At least four assays for estimating human age have been developed based on DNA methylation at small numbers of cytosines that are known to have an especially strong relationship with age (Bocklandt et al., 2011; Koch and Wagner, 2011; Garagnani et al., 2012; Horvath, 2013). A humpback whale age estimation assay was recently developed that works with skin biopsies (Polanowski et al., 2014). The assay, which used pyrosequencing technology, was based on methylation changes in the humpback whale orthologues of genes known to have age-related methylation levels in humans and mice. The absolute ages of those humpback whales is known because of long-term photo-ID studies from birth, and the standard error of the epigenetic estimates was about 2.9 years; however, most of the variation (2.2 years) was due to between-assay variability on the same tissue samples. This suggests that the standard error could be reduced by generating larger numbers of pyrosequencing reads per assay, or by finding gene regions with stronger age-related changes in methylation. Estimates accurate to as little as 1 year might eventually be possible given sufficient determination and financial investment.

With ABWs, as oppposed to humpback whales, absolute known ages will not be available. However, a more limited *age-ordering* assay for ABWs based on DNA methylation changes in skin cells is quite feasible. The purpose of this assay would merely be to determine, for each POP detected, which animal was the Parent (i.e. whichever one is older, in any given year). Since ABWs are believed to require at least 6 years to reach maturity, the age-gap in a POP must be at least 6 years, so that the imprecision in estimating comparative ages should be small compared to the signal. The assay could be calibrated from recaptured individuals through the duration of the study; around 30-40 such (re)capture-pairs are expected in the baseline scenario of Peel et al. (2014), and some are sure to span at least 10 years (recall that there are already many samples from the SOWER period). Differences in methylation level at a range of potential cytosines would then be measured in each recapture-pair following the same process as Polanowski et al. (2014), and compared to the number of years elapsed between the first and second captures. Those cytosines that consistently had the strongest changes in methylation relating to intra-individual change in age would be combined to make a multi-site predictor of ordinal age, i.e. relative to an unknown offset.

The cost of the assays in Polanowski et al. (2014) was around AUD 60, not much more than the cost of individual genotyping; if used as suggested here, only individual recaptures and POPs would need the assays, so the total cost would be very small. The role of DNA methylation and other epigenetic changes in the ageing process of mammals is being intensively researched by many laboratories worldwide, and the knowledge gained will make epigenetic age estimation assays increasingly easy to develop with time.

1.3 Close-kin genetics

The message of this section is that reliable identication of POPs, based solely on pairwise comparisons of genotypes from biopsies of large numbers of mostly unrelated animals, is entirely possible, and can nowadays be thought of as routine, in the same way as genetic mark-recapture of individuals. Of course, it is necessary to be very careful about the details, and the process is certainly more complicated than genetic mark-recapture, but it need not be expensive (see section 3) nor particularly error-prone. Readers already familiar with genetics, and readers who are completely uninterested in it, may wish to skip this section.

The principles of POP identification from genetic markers have been well-known for some time; Skaug et al. (2010) explains how to best to go about it using limited numbers of genetics markers. Recently, it has

become realistic to use larger numbers of dedicated genetic markers in studies specifically designed for POP-finding, which has made it affordable to control error rates down to arbitrarily low levels. The explanation given here is closely based on the first author's experience of recently completing a very large POP-based CKMR study of Southern Bluefin Tuna (Bravington et al., 2014), in which about 5000 adult and 8000 juvenile fish collected between 2006 and 2010 were genotyped at 20-27 carefully selected microsatellite loci. After careful QC screening, the adult and juvenile genotypes were compared pairwise to identify POPs, yielding about 45 in about 38,000,000 comparisons overall, with tight bounds on plausible errors in the number of POPs. The setup for ABWs would be different mainly because many fewer animals would be involved, and essentially all comparisons would be between adults, and the genetic technology has advanced enough so that the microsatellite route is probably not now the best option (see section 3). However, the principles are very similar, as described next.

Modern genotyping methods allow us to measure many different loci across an animal's genome[1]. Each measurement can take different values ("alleles"), depending on genetic variability in the population at that locus. The nature of an "allele" depends on what type of genetic marker is found at the locus: for example, for microsatellite markers the alleles are numeric values in the 10-100s, or a single genetic "letter" (G/A/T/C) for a SNP marker. The suite of alleles encountered in the population depends on the locus; most SNP loci are either G/A or C/T, for example, whereas useful microsatellites for POP work will typically have, say, 8-20 possible numeric values.

The loci used in POP identification are "diploid", meaning that each animal has two versions of the locus. These two might be different ("heterozygous") or the same ("homozygous"). One version is inherited from each parent. Therefore, a POP must have at least one allele in common at every locus. Of course, an unrelated pair may also just happen to have one or two matching alleles at any particular locus. However, the probability of this happening at *every* locus is very low if the number of loci examined is large enough. Therefore, the simplest principle for identifying POPs is "Mendelian exclusion": a pair is treated as POP if, and only if, the two animals have at least one allele in common at all loci.

It is evident that there are two types of potential error. The first is False Positives, whereby an unrelated pair just happen to have an allele in common at every locus compared. The second is False Negatives, where a true POP is rejected according to Mendelian exclusion because of either a mutation (rare) or a genotyping error (rarity depends on tissue quality, careful choice of markers, and general QC); note that just one error at one locus could lead to true POP being rejected under the most rigid version of exclusion.

The safeguard against both types of error is basically the same: *use a lot of loci, and only bother with good-quality comparisons*. It is straightforward statistically to work out the expected number (and likely upper limit) of False Positives, based on the expected number of pairwise comparisons, the number of loci likely to be used in each comparison, and the allele frequency distributions at the loci, which can estimated from preliminary genotyping of a modest number of animals (40, say). By increasing the number of loci, and keeping a tight tolerance on which pairwise comparisons will be deemed usable (i.e. where enough loci are genotyped successfully in both animals), the FP rate can be reduced to arbitrarily low levels, say to less than 1% of a plausible expected number of *true* POPs (see below). By *further* increasing the number of loci, it is moreover possible to control the expected number of unrelated comparisons that exclude not only at zero but also at just one (or two, or some fixed small number) loci. If this number is kept well below 1, then there is a high probability that any pairs which *do* turn out to exclude at just one (or two, etc.) loci are actually False Negatives; the number of such pairs found can then be used to bound the likely proportion of FNs.

[1] For now, consider only the diploid parts of the nuclear genome, which is after all most of it.

Since it is hard to predict genotyping error rates in advance (especially for detecting close relatives, where the same error might in fact be likely in both relatives), the key is to plan for enough "clear blue water", in terms of number of excluding loci, between true POPs (where there should be zero excluding loci, but might occasionally be one or two because of errors) and the "luckiest" unrelated pairs (ie those that by chance happen to have alleles in common at the most loci). In practice, for the SBT study we were able to use a subset of comparisons to set an upper bound on error rates, which could be extrapolated to the remaining comparisons. We estimated that the overall number of errors around our 45 POPs was unlikely to exceed 1 FP or 3 FNs; so far, subsequent validation using extra markers has failed to find any FNs or FPs.

1.3.1 Is the POP a PO or an OP?

The genotyping process identifies POPs reliably, but does not tell us which animal was the P and which the O. Epigenetic ageing should be able to do this, but there is also a simpler method which can distinguish the order for the subset of POPs where the two animals are different sexes, by comparing mtDNA haplotypes[2]. (Sex can readily be determined during the genotyping process.) Offspring always inherit their mtDNA from their mother not their father, so, if the two animals in a mixed-sex POP have different mtDNA haplotypes, then the male must be the O and the female the P. Further, provided there is a wide spread of mtDNA haplotypes in the population, most mixed-sex POPs that have the same haplotype are likely to be mother-son pairs. Some father-daughter pairs will also share a haplotype by chance; the proportion is predictable, and a mixture-distribution approach can be used to handle uncertainty about the nature of the relationship (see next subsection).

For ABW surveyed in AMR mode, about 75% of biopsies are likely to be male, so about 38% of POPs will be mixed-sex. In the examples of section 2, the mtDNA-based ordering (i.e. without using epigenetic ordinal age) is assumed to be accurate on all 38%, though in reality the information content would be slightly lower because of chance haplotype-sharing in father-daughter pairs.

1.4 Demographic equations

Remark 1. The material in this section is drawn from a draft MS by Bravington and Skaug, which sets out the general theory of CKMR for sampling adult populations. For brevity, details of the algebra and most of the actual formulae are omitted here. Many of the results are special cases of more general formulae in Keyfitz and Caswell (2005), particularly chapters 5 and 15.

The demographic framework here is for an open population sampled only as adults, with constant but unknown birth and death rates, no immigration or emigration, and fixed, known age-at-maturity (assumed to be 6 for now). The population is assumed to be in exponential growth and at quasi-equilibrium, i.e. to have been steadily increasing for long enough to have reached a stable age structure (which is *not* the age structure of a stable population that has *stopped* increasing; see below). We assume a true sex ratio of 50:50, and that the demographics of both sexes are the same. This framework is consistent with *IWC/65/SHxxx*, with the extra assumptions of quasi-equilibrium and known age-at-maturity. As in that paper, we have not made explicit provision for spatial population structure, but see section 4 for further discussion.

In what follows, β is the "single-sex adult birth rate", i.e. the actual birth rate of (say) female juveniles multiplied by the proportion that survive through to maturity at age α. The death rate of adults is δ. It can be readily shown that, in quasi-equilibrium, the adult age distribution is exponential with slope $\beta' \triangleq \beta \exp(-\rho\alpha)$,

[2] Pointed out long ago to the first author by Scott Baker– thanks!

where the population growth rate ρ satisfies the nonlinear equation $\rho = \beta e^{-\rho\alpha} - \delta$ so that $N_t = N_0 e^{\rho t}$, where N is adult abundance in year t referenced to a nominal "year zero". In a completely stable population, the age structure would of course have slope δ.

The approach to recaptures is fundamentally the same regardless of the Type of the recapture: Self, Parent, or Offspring. For each Type, given that the first capture is in some year, the key quantity is the expected number of living possible-recaptures of that Type y years later. For Self-recaptures, the animal was clearly alive at 0 but might have died since, so the expected number of possible-recaptures is obviously

$$\mathbb{E}\left[\text{Living Self @ } y\right] \exp\left(-\delta y\right)$$

The analogous formula for "recapturing a Parent" (of given sex, say female) turns out to be

$$\mathbb{E}\left[\text{Living mothers @ } y\right] = \frac{\beta'}{\beta' + \delta} \exp\left(-\delta\left(y + \alpha\right)\right) \tag{1.1}$$

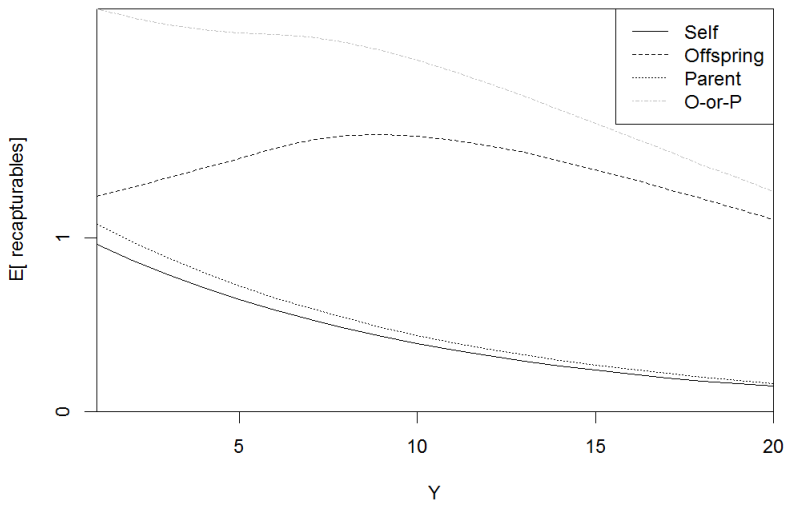
and for "recapturing an Offspring" (again, of given sex) is

$$\mathbb{E}\left[\text{Living daughters @ } y\right] = \frac{\beta}{\beta' + \delta} \exp\left(\beta\left(y - \alpha\right)\right) \tag{1.2}$$

provided that $y \leqslant \alpha$, or a slightly more complicated form if not. These quantities (aggregated to include both male and female kin) of Self and P-then-O and O-then-P, and the last two combined as PO/OP, are shown in Figure 1.1, as a function of y and for base case parameters (Table 1). A couple of features stand out:

1. There are about the same number of surviving parents as there are "selfs"
2. There are substantially *more* surviving offspring
3. The rising and falling limbs of the "offspring" curve (which in fact have different slopes) are easily distinguished, but the subtleties of shapre are somewhat masked in the combined "O or P" curve

Figure 1.1: Expected "recapturables"



These formulae show the expected numbers of *potential* recaptures of the given Type. The actual *probability* that a given comparison will in fact yield a recapture of that Type, is equal to the potential recaptures divided by the number of other animals that might be caught, in effect the adult population size[3]. For a comparison between a pair of animals caught at t_1 and t_2, the probability formulae look like this (NB only one Type shown here):

$$\begin{aligned}&\mathbb{P}\left[\text{Comp.result=O-then-P}|t_1,t_2\right]\\&=\frac{1}{N_0\exp\left(\rho t_2\right)}\frac{\beta'}{\beta'+\delta}\exp\left(-\delta\left(\left(t_2-t_1\right)+\alpha\right)\right)\end{aligned}$$

1.4.1 An interesting thing

One interesting feature of eqns (1.1) and (1.2) is that, for fixed y and α, the expected number of living POPs depends on the ratio of β and δ, i.e. on the population's rate of growth; there are more close relatives per capita in a growing population than in a shrinking population of the same (current) size. This differs from conventional "Self" mark-recapture, where clearly You are alive in the year You are first caught regardless of what the population is doing. The dependence on rates means that CKMR-only mark-recapture (i.e. without Self-recapture, as would be the case if sampling was lethal) exhibits some degree of confounding between abundance and trend. In other words, estimates of abundance are partly correlated with estimates of birth and death rates in CKMR. This is eventually resolved given a sufficient length of study, even without Self-recapture, but the use of Self-recapture data considerably improves the ability to quickly separate the effects of abundance, birth, and death. The same general phenomenon was noted by Skaug (2001), who deals with a generalized CKMR beyond the POP case, but without Self-recapture. In the case of ABWs, the abundance is expected to be increasing, which makes CKMR even more effective (because it generates more recaptures, and thus has better precision) relative to pure Self-recapture than it would be in a static population.

1.4.2 Constructing the likelihood and predicting the uncertainty

Depending on whether epigenetic age and/or mtDNA are used, we may not be able to order a POP, i.e. to tell whether it is actually PO or OP where the letter order matches the capture order. We therefore need to consider the possibility of a fourth *observed* Type, PO/OP. Given that a POP is actually PO, the probability that it will be *recorded* as PO rather than PO/OP depends on what additional genetic analyses are performed: 1 if epigenetic ageing is used, $2q\left(1-q\right)$ if mtDNA is used and the sampled sex ratio is q, or 0 otherwise.

Because recaptures of any kind will be rare (under 1% of ABW samples, at least in the first decade), it is reasonable to consider the four Types of recapture as independent events, [4]. Also, because the sampling covers such a small fraction of the population, it is statistically reasonable to treat all comparisons as independent[5]. For any one comparison-and-Type i where the probability of "success" is p_i say, the log-likelihood of the outcome (success or failure) is very well approximated by a Poisson distribution with mean p_i (true only because $p_i \ll 1$). Thanks to independence, all the individual log-probabilities can be summed to give the overall log-likelihood,

[3]The population is assumed large enough that the potential recaptures are a negligible proportion.

[4]Logically, this is preposterous, since e.g. an animal cannot be its own offspring, but statistically it's OK.

[5]Again, this is not *strictly* correct. For example, once a Mother has been found, then the probability in subsequent comparisons that she will match other sampled offspring rises slightly, because she is a posteriori likely to be older than an average female. However, the per-comparison probabilities are so small anyway— less than 1/1000— that the changes are irrelevant.

which can be maximized to find the MLE $\hat{\theta}$, where θ inclues all the demographic parameters that govern the p_i's: N_0, β, etc. The variance of $\hat{\theta}$ can be estimated via the usual asymptotic statistical arguments as H^{-1}, where

$$H = \mathbb{E}\left[ss^\top\right] = \sum_i \mathbb{E}\left[s_i s_i^\top\right] \tag{1.3}$$

and s is the derivative of the log-likelihood evaluated at the true parameter values θ_0. Thanks to the Poisson property, and a *soupçon* of algebra, this all turns out to have a particularly simple form:

$$\mathbb{E}\left[s_i s_i^\top\right] = p_i^{-1}\left(\frac{dp_i}{d\theta}\frac{dp_i}{d\theta}^\top\right) \tag{1.4}$$

For survey design purposes, we can simply plug in a guesstimate of θ_0, predict the number of comparisons by year and sex (based on the demographics and the particular survey design being evaluated), evaluate equations (1.4) and (1.3), and invert H to see what the uncertainty is likely to be when the survey is complete.

1.4.3 Robustness against "unmodelled heterogeneity"

It is conceivable that individual ABWs differ in their determination to avoid being sampled, in a way that is consistent over time but not linked to an obvious covariate such as sex or location that could be "modelled away". This kind of unmodelled heterogeneity is well-known to cause sometimes serious bias in conventional MR estimates of abundance, though much less so in estimates of birth and death rates (and not at all if the rates are uncorrelated with behaviour). Unfortunately, the phenomenon is also undetectable (and unfixable) unless there are numerous individuals with multiple-capture histories, which will certainly not be the case for the design proposed here (though it might be partly feasible using platform-of-opportunity photos, e.g. from cruise ships).

However, even if HCP for ABWs really is large enough to cause significant bias, it seems much less likely to be significantly *heritable*. We can choose therefore to rely on only the CK-recaptures for abundance estimation, while continuing to use all the data including Self-recaptures to estimate the demographic parameters. This is a powerful combination, because of the partial confounding between the rate and abundance parameters for the CK data alone (section 1.4.1), which the Self-recaptures help to resolve.

Statistically, a simple solution is to introduce a second, artificial, abundance parameter N_0', which affects the recapture probabilities for the Self-recaptures only in exactly the same way that N_0 affects the probabilities for the CK recaptures. All four parameters $(N_0, \beta, \zeta, N_0')$ are used in estimation and in forming H, but the row and column for N_0' are discarded *after* inverting H. Of course, there is a bias-variance trade-off; the simpler model without N_0 can use more recaptures to estimate abundance, but at the risk of bias due to HCP. The difference between $\hat{N}_0$ and $\hat{N}_0'$ in the more complex model is indicative of the magnitude of HCP so, if the difference turns out to be negligible, it might be reasonable to claim that the assumptions of the simpler model seem to be valid, and therefore to switch *post hoc* to the simpler and more precise model.

2 Results

For brevity, we concentrate on just one set of results, for the acoustically-assisted base case in Peel et al. (2014) (Table 1), with notes as follows:

- In all cases, the demographic model estimates birth and death rates (or, equivalently, Rate of Increase and death rate) rather than treating either as known; age at maturity was assumed known.
- Within-year comparisons are excluded (since repeat captures of an individual within a year are fairly common, and would otherwise be a source of heterogeneity bias), although it might in fact be OK to use within-year CK recaptures but not self-recaptures.
- The data includes the 166 SOWER/IDCR samples (excluding within-year repeats) but does not use the actual results of those comparisons (2 cross-year recaptures).
- The reference year of 2020 was chosen as a round number close to the mid-point of the study, which is roughly where the CV of N_{refyear} seems to be lowest.

Table 1: Base case scenario for results in this paper

Parameter	Value
N_{1998}	2280
age at maturity	6
Rate of increase	0.06
Adult mortality	0.04
Population sex ratio	50% male
Survey start	2015
Duration	12 years
Annual effort	35 days
Biopsy sex ratio	75% male
Reference year for CVs	2020

Results are shown in Table 2; the messages, which are pretty obvious, are spelt out later in section 4.

Table 2: Predicted CV %s for base case design and various CK options

	Self	CK no-ord	CK partial	CK full
Simple	27.0	23.2	18.9	17.4
Robust	-	29.0	24.4	22.8

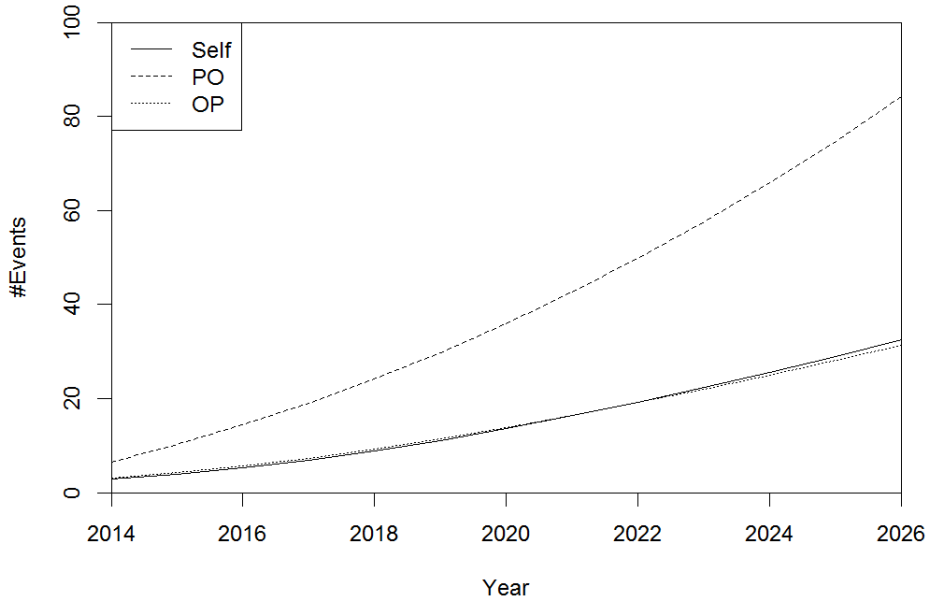
"Self" means conventional individual mark-recapture only. The CK options relate to how many of the POPs can be ordered (i.e. to know which is the P and which is the O). "Partial" assumes all mixed-sex POPs can be ordered via mtDNA; "Full" assumes all pairs can be ordered via epigenetic ageing; "no-ord" assumes none can be. The upper row of CVs is for the basic (CK)MR model, and the lower row for the HCP-robust version of section 1.4.3.

We have not attempted a full sensitivity analysis, but in limited experiments the qualitative results across *methods* were consistent. With respect to biological parameters, across the range of values considered in Peel et al. (2014), all CVs are fairly insensitive (no more than a couple of percentage points) to different values of N_{1998} or rate of increase. One parameter that does have an impact, is age-at-maturity α: if the true α is actually 10 not 6, then the Simple results do not change much (not at all for Self, of course), but the Robust results are appreciably worsened (36%, 30%, 28% respectively when all other parameters are as in Table 1). Nevertheless, such CVs are still not too bad, and in any case the main impact of α is on how long it takes to get an answer, not on how much ship time is required; with $\alpha = 10$, then the CVs can be restored to those of Table 2 by extending the survey period to 18 years without increasing the total amount of effort. (It is actually possible to estimate α just from CKMR data, even without any age information, but we have not yet investigated the feasibility for ABWs.)

This illustrates a general point: survey duration is more important for CKMR studies (in terms of estimating N) than it is for Self-MR studies. In a self-MR study, lengthening the study while keeping total effort fixed tends to worsen the CV, because each animal has more chance of dying before being increased. For the CKMR components of the likelihood, though, it is essential to have decent estimates of demographic parameters because they, and not just N, affect the expected number of recaptures. Demographic parameters correspond to slopes in Figure 1.1, and slopes are easier to estimate when the X-range— ie the study duration— is longer. For another example, reducing the survey duration to 6 years while doubling the effort-per-year has little effect on CVs for the Simple model, but worsens the CVs appreciably for the Robust model (37%, 32%, 31% respectively).

Figure 2.1 shows how the expected number of recaptures of each type grows through the study (parameters as in Table 1). (It is a coincidence that, with the particular parameters used, the numbers of Self and OP recaptures are so similar.) The numbers for Self are very similar to those in Peel et al. (2014), which uses the same model but implemented with completely different computer code.

Figure 2.1: How "recaptures" accumulate through the study



3 Options for genotyping

This paper is not really about genetic techniques, but since CKMR has rather different requirements from individual MR, and since the available technology continues to change very fast, it is worth making a few points about what options are likely to be best for a study that will have to be long-term to work. Although "one can always change one's mind and re-genotype the tissues later", there are substantial development costs associated with some of the technologies, and it makes sense for IWC and SORP to discuss these issues right away.

The general strategy is that all samples would be genotyped at most of the markers (often, some loci cannot be genotyped in some individuals because of DNA quality issues), and at least the recaptures but not necessarily all the samples would undergo further genetic analysis (e.g. of mtDNA). The per-sample cost— at 2014

prices— would be somewhere between AUD 20 and AUD 50 depending on the genetic approach chosen and on who is doing the work; costs will fall over time, more so for some technologies than others. Even the most optimistic scenario in Peel et al. (2014) only generates about 1000 samples in the course of the study so, regardless of which technology is used, the running costs will be a minuscule proportion of the cost of sample collection from Antarctic voyages. Additional genotyping— for sex markers, mtDNA haplotype, and/or epigenetic methylation— would only be strictly necessary for samples in POPs or self-recaptures. These form a small proportion of the total, and so the extra expense of these additional analyses would be small: so small, in fact, that it might just be worth running the additional analyses on all the samples anyway to see if any extra information can be squeezed out.

The material here is an informal summary based on extensive recent experience at CSIRO, partly on the tuna study mentioned above and partly on several other ongoing and proposed close-kin projects.

Microsatellites have historically been popular for individual mark-recapture, and continue to be so today. For many species, candidate microsatellites *for conventional population genetics* are already available, and if unsatisfactory in their original state they can often be tweaked to perform better. However, microsatellites that are good for population genetics cannot be too variable otherwise the noise is guaranteed to mask any signal, and therefore are *not* ideal for individual mark-recapture, and are even worse for CKMR; a very large number would be needed for finding POPs, and the individual genotyping costs (which are roughly proportional to the number of microsatellites used) would be unnecessarily high. In studies that are forced to rely exclusively on pre-existing genotyping data, then trying to disentangle POPs from inadequate (for that purpose) numbers of not-particularly-variable microsatellites as in Skaug et al., 2010 can be a retrospective necessity— but it is a route to vigorously avoid in any new, dedicated study where ship time will be by far the limiting cost. For that reason, using microsatellites for ABW would require the deliberate development of panels of hypervariable microsatellites, which also need to be very reliable for scoring because the exclusion criterion for POPhood is sensitive to errors. Rough calculations along the lines done for SBT (taking into account the different population sizes) indicate that perhaps 20 such microsatellites would be needed, which (based again on SBT experience) might entail trialling closer to 100 to find loci with enough variability (i.e. estimating allele frequencies) and scoring reliability. The staff time required to do this would be considerable.

Further, "scoring" microsatellites is a notoriously difficult process to make consistent across laboratories, and even across personnel and equipment within institutions. For a study which is likely to last a decade and to involve multiple international partners, it seems advisable to expect changes in personnel, equipment, commercial organizations, and research institutions— types of change to which microsatellites are not particularly robust.

Targetted SNP assays, where a number of SNP loci are specially chosen for inclusion in a dedicated "assay" on a "chip" (a.k.a. cube, many other names...) which processes panels of about 50 or about 100 SNPs at a time for batches of typically around 100 animals at a time. For individual ID, a single panel of ~ 50 SNPs is quite adequate, but for reliably finding POPs, we calculate that close to around 200 SNPs (2 or 4 panels) would be needed— so the running costs would likely be four times higher. There are a number of different proprietory companies, with considerably varying requirements in infrastructure costs and running cost; whether all the companies will still be in business in 2025 is another question. On the plus side, there is no concern about subjectivity or repeatability when targetted SNP assays rather than microsatellites are used.

Currently, running a single 50-ish SNP panel is probably the cheapest option of the three for individual

MR, but not necessarily for CKMR. There is also substantial development cost in finding suitable SNP panels, perhaps comparable to what is required for microsatellites; to keep the total to 200, it is necessary to choose SNPs with the minor allele frequency fairly close to 0.5.

RAD-tagging, an example of Next-Generation Sequencing (there are several other semi-interchangeable semi-acronyms), is the newest applicable technology. It offers 1000s of SNPs per individual, with almost zero development cost. Per-genotype running costs are currently slightly higher than for microsatellites and targetted assays (maybe twice the cost), but the costs of the sequencing technologies that underlie this approach have fallen massively in recent years and continue to do so; cost reductions in microsatellites and targetted assays have been much lower. RAD-tagging is heavily "outsourced", especially compared to microsatellites which can largely be done "in-house". At least in principle, the technology is robust to being transferred to a different provider. In future, it is not inconceivable that RAD-tagging might be combined directly with the methyl-sensitive assays used for epigenetic age estimation, thus simplifying procedures and reducing running costs further.

At CSIRO, we have opted to use RAD-tagging on all new projects, except perhaps where individual ID is the only objective and sample sizes are very large (10,000s) in which case a targetted assay may be better. So far, our experience with RAD-tagging has been very positive— perhaps because we have found a particularly good and co-operative service provider.

4 Conclusions

There is nowadays no serious im pediment to reliably finding Parent-Offspring Pairs in large collections of biopsy samples, via Close-Kin Genetics. This offers the possibility of substantially improving a biopsy-based ABW Mark-Recapture study by using Close-Kin Mark-Recapture, at very little marginal cost. The basic strategy of an ABW MR study is discussed in Peel et al. (2014). As far as the CKMR aspects go, it is clear from the results in Table 2 that

1. CKMR offers a large improvement in CV over conventional Self-MR, using exactly the same samples. At the levels of sampling effort already considered, and under a typical scenario, it brings the achievable CVs well out of the just-acceptable zone around 30% with Self-MR, down to the pretty-good zone under 20%. If effort levels per year end up being lower than considered, then MR (of any kind) will still produce a precise answer in the end, but will take longer– and CKMR will considerably shorten the time required.

2. Further large improvements in CV are possible if at least some POPs can be ordered (to tell which animal is the Parent), which is possible for some but not all POPs via mtDNA analysis. The gains are biggest if all POPs can be ordered; this would require epigenetic ageing, but there is no strong reason to think this wouldn't work for ABWs, especially given the limited precision required for POP-ordering purposes.

3. With CKMR, the abundance estimate can be robustified against unmodelled individual heterogeneity, at some cost to CV but still with an acceptable precision under the base case scenario and survey design.

All this is possible essentially because CKMR will generate 4-5 times as many "recaptures" as conventional Self-MR; the ratio is particularly favourable for ABW because the population is (presumably) growing rather than stable or falling. Some of the extra information is required to estimate rate parameters rather than directly

for estimating abundance, but even so, in terms of abundance alone, the precision of CKMR for ABW would equal that of a pure Self-MR survey at about 240% of the original effort.

In addition, CKMR provides clear information about stock structure— or its absence— on "management timescales"[6], ie around one generation. Of course, stock structure information is a major selling point for Self-MR anyway (compared to, say, a pure Line-Transect survey) because it can show how much individual *adults* (the main biopsy target, in this case) move around. However, Self-MR cannot reveal any heritable signal in the unsampled parts of the population (e.g. if juveniles move around extensively but adults stay put), which is fundamental to stock structure. CKMR will do that, and also will work in situations where conventional population genetics cannot detect a clear signal, because the CKMR signal is not blurred by migration and does not require long time-scales of (partial) separation for allele frequency divergence (Palsbøll et al. (2010)).

Although POPs do require slightly more work genetically than Self-MR, there are no serious technical issues if the right technology is used. The marginal cost of doing the extra genetics is really negligible compared with the overhead of sending ships to Antarctica, and the cost of marker development is nowadays also very low if (but only if) the right technology is used. The assays required for epigenetic ageing are fairly cheap now (and will get cheaper), and do not need to be applied to every sample (only to POPs, if the unambitious approach of this paper). The development cost for the epigenetic ageing is not trivial, but again pales into insignifance when set against the extra ship time that would be required to achieve equivalent CVs without it.

This paper has been kept fairly simple in order to make its main points. Some of the points not covered include:

- we have not considered what to do with biopsies of immatures— presumably, there will not be that many, and it will be easy enough to tell, and with a little effort they could be incorporated into an augmented demographic model anyway.
- we have assumed that all mixed-sex pairs can be ordered via mtDNA. In practice, the number of mtDNA haplotypes is finite, and there will be some *a posteriori* uncertainty about the true orders of those pairs that have the same mtDNA haplotype: they will mostly be Mother/Son, but not exclusively. This would somewhat reduce the power of the "CK partial" option, but in any case we expect that epigenetic ageing will work well enough so that the "partial" option becomes irrelevant compared to the "CK full" option, which has the best CVs.
- we have assumed that the spatial sampling design will be sufficient to detect, and then allow for, stock structure, if the latter is important enough to matter. Presumably, though, the conclusion of Peel et al. (2014) would still apply, and even more strongly given the larger sample sizes of CKMR: that spatial coverage need only be "reasonable", rather than perfectly balanced, in order to largely eliminate the "black ice zone" where structure is important enough to bias abundance estimates but not large enough to be clearly evident in the data.

There are not many downsides. One point is that CKMR obviously requires biopsies, whereas self-MR on its own *might* be possible from photo-ID alone. It *might* be the case that ABW biopsies are harder to collect (longer times-to-closure, lower probability of success) than photo-ID quality photos, so that a biopsy-based survey *might* lead to fewer samples than a pure photo-ID survey. But even this is not clear, and has to be balanced against several points (see also Peel et al. (2014)):

[6] In the case of ABWs, unlike for exploited species, it is perhaps not as clear what a "management timescales" might be, since hopefully no-one has any "management" planned.

1. There is a both a left-side and a right-side photo-ID catalogue for ABW, and it is not always possible to get both left- and right-side photos, so that in effect some recaptures will be lost due to not recognizing the same whale;

2. Photo-ID is rarely a perfect process, and blue whales are not the easiest species to get rock-solid IDs for;

3. By far the main time-to-closure issue in a acoustic-assisted survey of ABW (acoustic assistance being essential for getting even a remotely useful number of samples within a bearable time frame) is getting close enough to the whales to begin the sampling attempt— any increased closing time in that final phase is probably unimportant. Failure-to-close is obviously more serious, but ABW biopsy failure rates were low, at least in SOWER.

Although the practical impact of these points is not clear, in all it nevertheless seems most unlikely that a photo-ID based study could come even remotely close to getting the same precision as a CKMR study, since doing so would require 240% more (same-side) encounters. And, perhaps crucially, the photo-ID-focussed studies cannot necessarily be made robust against HCP. Of course, photo-ID can and should be built into protocols as a useful adjunct to biopsy samples, as discussed in Peel et al. (2014), but it would be wishful thinking to imagine that a survey where biopsies are "optional extras" is likely to generate enough POPs to allow the extra statistical power of CKMR to have much effect.

We have been deliberately unambitious in considering the scope of epigenetic ageing for ABW; the approach proposed here is in no way stretching the technology to its limits. Although direct calibration will not be possible with ABWs in the way that is has been for NW Atlantic humpbacks, it seems plausible that much extra further information about demographic processes (birth and death rates) could be extracted from comparing distributions of "epigenetic age" (related to true age by an unknown, but monotonic, relationship) with demographic predictions from the CKMR and Self-MR models. That would further improve the precision of CKMR abundance estimates, because the latter are correlated with demographic parameters. It will take several years before there are enough individual recaptures to start the development of epigenetic age estimators for ABW, and in the meantime there will certainly be advances in sequencing technology and in the general understanding of methylation rates across mammalian genomes. Hence, the prospects for using epigenetic age markers are likely to be even better towards the end of a multi-year ABW mark-recapture program than they are now.

In short, there seems to be a compelling case for making biopsy-focussed POP-based CKMR, and augmentation via epigenetic age data, into central planks of an ABW MR study.

References

Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S, and Vilain E (2011). “Epigenetic predictor of age”. *PloS one* 6.6, e14821.

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

Bravington M, Grewe P, and Davies C (2014). *Fishery-independent estimate of spawning biomass of Southern Bluefin Tuna through identification of close-kin using genetic markers*. FRDC Report 2007/034. CSIRO, Australia.

Double M, Miller B, Olson P, Andrews-Goff V, Leaper R, and Ensor Pea (2013). *Cruise report of the 2013 Antarctic blue whale voyage of the Southern Ocean Research Partnership*. Scientific Committee paper IWC/65a/SH21. International Whaling Commission.

Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, Di Blasio AM, Gentilini D, Vitale G, Collino S, et al. (2012). "Methylation of ELOVL2 gene as a new epigenetic marker of age". *Aging cell* 11.6, pp. 1132–1134.

Grönniger E, Weber B, Heil O, Peters N, Stäb F, Wenck H, Korn B, Winnefeld M, and Lyko F (2010). "Aging and chronic sun exposure cause distinct epigenetic changes in human skin". *PLoS genetics* 6.5, e1000971.

Horvath S (2013). "DNA methylation age of human tissues and cell types". *Genome biology* 14.10, R115.

Keyfitz N and Caswell H (2005). *Applied mathematical demography*. Springer.

Koch CM and Wagner W (2011). "Epigenetic-aging-signature to determine age in different tissues". *Aging (Albany NY)* 3.10, p. 1018.

Maegawa S, Hinkal G, Kim HS, Shen L, Zhang L, Zhang J, Zhang N, Liang S, Donehower LA, and Issa JPJ (2010). "Widespread and tissue specific age-related DNA methylation changes in mice". *Genome research* 20.3, pp. 332–340.

Palsbøll P, Peery Z, and Berube M (2010). "Detecting populations in the 'ambiguous' zone: kinship-based estimation of population structure at low genetic divergence". *Molecular ecology resources* 10.5, pp. 797–805.

Peel D, Bravington M, Kelly N, and Double M (2014). *Examination of a Future Mark-Recapture study of Antarctic Blue Whales*. Scientific Committee paper IWC/65/SHxxx. International Whaling Commission.

Polanowski AM, Robbins J, Chandler D, and Jarman SN (2014). "Epigenetic estimation of age in humpback whales". *Molecular Ecology Resources*.

Skaug H (2001). "Allele-Sharing Methods for Estimation of Population Size". *Biometrics* 57, pp. 750–756.

Skaug H, Berube M, and Palsbøll P (2010). "Detecting dyads of related individuals in large collections of DNA-profiles by controlling the false discovery rate". *Molecular ecology resources* 10.4, pp. 693–700.