Our goal is to evaluate potential survey designs for a CKMR survey of walrus, i.e. to predict the variance of some quantity-of-interest that would be obtained if the data was collected according to that design. In many other contexts we would need to create a series of simulations for each design, estimate parameters (and the quantities of real interest, which may be functions of the raw parameters) for each simulation, then directly calculate the variance of those estimates. Happily, this cumbersome process can be circumvented almost entirely with the formulation of CKMR+IMR that we use, because of three key properties:

- it is based on a pseudo-log-likelihood that sums over pairwise comparisons between samples;
- the outcome of each comparison is discrete yes-or-no: do these two samples the target kinship, or not?
- individual samples have multiple covariates (e.g. year of sampling, age), but the range of possible values for each covariate is limited.

Compared to general mark-recapture frameworks, this leads to some remarkably simple formulae for computations. In particular, because the pairwise comparisons are almost mutually independent (see below), the parameter covariance matrix can be predicted, without any simulated data, just from two types of information:

- the number of pairwise comparisons with each particular combination of covariates, basically the product of the numbers of relevant samples (which constitute "the design");
- a fairly-easy-to-compute function of the kinship probabilities, which determines how much statistical information about the parameters will be obtained, on average, from a single pairwise comparison of samples with those covariates.

We present a slightly simplified explanation here, dealing first with the latter. Let y_{ijk} be the kinship outcome for samples i and j and target kinship k: $y_{ijk} = 1$ if their actual kinship $K_{ij} = k$, or 0 if $K_{ij} \neq k$; and let y be the set of all data (all kinship outcomes). Also define $p_{ijk}(\theta) = \mathbb{P}[K_{ij} = k | z_i, z_j, \theta]$ to be the kinship probability for parameter values θ, computed from a formula such as (**??**); we just write p_{ijk} where there is no ambiguity about θ. Each comparison has a very low probability of "success" ($y_{ijk} = 1$), on the order of the reciprocal of adult abundance, and is well approximated by a Poisson distribution with mean p_{ijk}. The pseudo-log-likelihood Λ is thus[1]:

$$\Lambda(\theta; y) = \text{const} + \sum_{i<j;k\in\mathcal{K}} \{-p_{ijk} + y_{ijk} \log p_{ijk}\} \tag{1}$$

With a real dataset y, we would estimate $\hat{\theta}(y)$ by maximizing (1), and then infer the variance of $\hat{\theta}(y)$ from the inverse of the Hessian (second derivative matrix) of (1)[2]. For design purposes, we work instead with $H(\theta_0)$, the expected value over datasets Y of $d^2\Lambda(\theta_0; Y)/d\theta^2$ at the true parameter value θ_0. Since Λ itself consists of a sum of terms over individual comparisons, the same is true of the second derivative and its expectation, which we can write say as $H(\theta_0) = \sum_{i<j;k\in\mathcal{K}} h_{ijk}(\theta_0)$. Some algebra (ref**) shows that the single-comparison expected Hessian is given by

$$h_{ijk}(\theta_0) = 4 d_{ijk}(\theta_0)\, d_{ijk}(\theta_0)^\top \tag{2}$$
$$\text{where } d_{ijk}(\theta) = \frac{d\sqrt{p_{ijk}}}{d\theta}$$

The term $d_{ijk}(\theta_0)$ is the derivative vector of the square-root of the kinship-k probability for covariate values z_i and z_j with respect to the elements of the parameter vector θ, at the value $\theta = \theta_0$. This can be obtained efficiently for all (i, j, k) by numerical differentation of the probabilities calculated by the CKMR model, using some reasonable guess about θ_0; the whole process takes just a minute or two for our walrus example.

The remaining requirement for design, is to group similar comparisons, i.e. across all pairs with identical covariate values. Let z_i denote all the covariate values for sample i that are needed to compute p_{ijk} (note that this

[1] In practice, certain types of comparison are usually excluded *a priori* based purely on their covariate values and the target kinship (e.g. second-order kin born a long time apart), but this does not alter the basic argument.

[2] Assuming sparse sampling...

may vary for different k, given different roles that the sample may be playing: for brevity we omit the k here). Now let $m(z)$ denote the number of samples with covariate values z. The number of comparisons between samples that have covariates z_1 and samples that have covariates z_2 is $m(z_1)\, m(z_2)$. The grouped version of the expected Hessian can be written as

$$H(m_{\mathcal{Z}};\theta_0) = \sum_{z_i<z_2\in\mathcal{Z};k\in\mathcal{K}} m(z_1)\, m(z_2)\, h(z_1,z_2,k) \tag{3}$$

where $h(z_1,z_2,k)$ is the single-comparison expected Hessian for two samples with covariates z_1 and z_2 respectively[3]. The set $\mathcal{Z}$ comprises all possible combinations of covariates, and $m_{\mathcal{Z}}$ is the corresponding breakdown of total sample size by covariate combinations (e.g. year and age and sex). Once H has been computed, it can be inverted to give the average predicted variance $V(m_{\mathcal{Z}};\theta_0)$ of a parameter estimate. CVs or standard errors of any quantity-of-interest $g(\theta)$ that can be obtained from θ, can then be approximated by the Delta method:

$$\mathbb{V}[g(\theta);m_{\mathcal{Z}},\theta_0] \approx \left[\left.\frac{dg}{d\theta}\right|_{\theta_0}\right] V(m_{\mathcal{Z}},\theta_0) \left[\left.\frac{dg}{d\theta}\right|_{\theta_0}\right]^{\top} \tag{4}$$

While a "design" must, by definition, include some specification of sample sizes, it may not specify the full breakdown of samples into specific z-categories. For example, the plan might be to sample 1000 adult walruses per year, but the age composition cannot be controlled directly. However, we still need to know that detailed breakdown $m_{\mathcal{Z}}$ in order to apply the above steps, so some extra extra assumptions and calculations might be required. For example, our population-dynamics model does not explicitly represent the adult age composition within the population, let alone within the samples; probability formulae such as (**??**) are *conditioned* on sample age, but make no prediction about how many samples of each age there will be. It would be possible to calculate expected sample sizes based on quasi-stable age compositions and unselective-sampling assumptions (assumptions that are in fact implicit for the self-recapture probability (**??**)), but somewhat laborious. Instead, since we are simulating sampled datasets in any case, the simulated sample composition can be used directly for $m_{\mathcal{Z}}$.

The use of the pseudo-log-likelihood Hessian to approximate the inverse variance is not strictly justified in a mathematical sense, because the pairwise comparisons are not fully mutually independent. The proposed walrus sample size (about 15,000 in total) is so large relative to adult abundance (about 70,000 females, although in effect somewhat more because of turnover during the years modeled) that roughly 10% of samples are recaptured multiple times, as self and/or as kin. This means that a comparable proportion of pairwise comparisons have predictable outcomes based on the results of other comparisons, which breaks independence. Thus the "sparse sampling" assumption of **?**, which underlies the use of the Hessian, is not strictly justified; this does not lead to bias in point estimates, but the Hessian-based approximation is likely to underestimate the true variance somewhat. Accordingly, we have made some simple adjustments to "effective sample size" based on summaries of the simulated datasets, as explained in the Appendix. This should be quite adequate for design purposes— since, in any case, all our variance estimates have to be based on uncertain assumptions about true parameter values— but a more detailed treatment may be worthwhile when it comes to analysing the real data.

0.1 Something for the Appendix, perhaps: Adjustments for non-sparse sampling

From experience, any attempt at a comprehensive treatment of non-independence in CKMR is complicated, to say the least. In this paper, we restrict attention to some obvious aspects for walrus that are easy to address. We consider the comparisons in stages: first SelfPs, then MOPs, then XmHSPs. We adjust set the effective sample size for each stage based on recaptures from the preceding stages in one simulated dataset, as follows:

- Sample sizes are initially taken from the simulated dataset (thus allowing detailed breakdown of sample size by age, year, etc). All available samples are used for SelfP comparisons.
- If an individual is self-recaptured, only its final capture will be used in MOP and XmHSP comparisons (i.e. duly adjusting the sample sizes sample sizes for MOPs and XmHSPs, as well as the number of MOPs etc found if that individual is involved).

[3]The ordering "$z_1 < z_2$" is arbitrary, included just to avoid double-counting. Sometimes it makes sense to also do comparisons with $z_1 = z_2$, in which case an extra factor of $1/2$ is required.

- Any Offspring o identified in a MOP, will be excluded from XmHSP comparisons (since o's sibship with any other sample i can be deduced from the MOP results, based on whether i is also an offspring of o's Mother).

This deals with the implications of one type of kinship for the others, but does not deal with multiple recaptures within a kinship class (e.g. an individual who is sampled 3 times; given that sample A matches sample B, and B matches C, it is redundant to compare A with C). There are simple ways to handle that with real datasets, as long as age is known fairly accurately, . Taking self-recapture as an example, we can compare the first capture of an individual to other samples from successively later dates, stopping immediately after a second capture (i.e. a self-recapture) if there is one. Thereafter, the first sample is not used in self-comparisons against later samples, but the second sample (the recapture) should still be used until and unless it too is recaptured, and so on. For XmHSPs, where we are really recapturing the shared Mother rather than the samples themselves, we can proceed in an analogous fashion, dropping the firstborn halfsib from subsequent comparisons once its first halfsib has been found, and so on.

It is difficult to follow that approach in a design context, because the number of *possible* triads is extremely large, even though in practice only a limited number will be seen in any real dataset. Instead, we make simple overall adjustments, as follows:

- For SelfPs, we start by tabulating the number of triple captures, etc. If we were using the above scheme for an individual caught 3 times, we would want to end up counting only 2 recaptures, rather than the 3 that arise from all pairwise matches. Generally, for an individual caught N times, we would only count $N-1$ recaptures rather than $N\left(N-1\right)/2$ from all pairings of its samples. Let α be the proportional reduction in the number of SelfP pairs that would result. The number of comparisons should be reduced accordingly, and we apply the same reduction α across-the-board to *all* SelfP comparison categories. (About 2% of walrus self-recaptures in different years are expected to be 3rd or more captures.)

- An exactly analogous approach works for XmHSPs (ignoring yet more complications from false-negatives). (About 6% of walrus XmHSPs are expected to be part of triads.)

This should give a reasonably unbiased adjustment to ensure sample sizes in the design are comparable to the adjusted sample sizes we would get with real data, although it is not quite right because some types of comparison (i.e. pairs of covariates) are liable to be more/less susceptible to multiple recaptures than others. That nuance really does not seem important for design purposes. A more practical problem is that the adjustments are dependent on results from a single realization of sampling, and in particular on the number of triple-captures, which is fairly small and thus subject to some variability. Thus, as usual in statistics, bias correction entails some increase in uncertainty.

It is not necessary to adjust MOP counts in the same way, because the outcomes of comparing one potential-offspring to all adults for MOPship are almost independent[4] of the results for any other potential-offspring; mothers are allowed to have multiple offspring.

All the phenomena above would, if uncorrected, tend to lead to underestimates of variance. However, there is another related phenomenon in non-sparse-sampling which tends to have the *opposite* effect, basically of finite-population-correction. Consider a simplified setting where all adults have the same covariates and only one juvenile, O, is sampled. In the MOP comparisons between O and all m_{ad} adult samples, at most 1 of those adult samples can be O's Mother, whereas the Poisson approximation in theory allows for 2 or more, albeit with low probability. It can be shown that the Fisher information (for a model where the only parameter is N_{ad}, the number of adult females) based on the Poisson approximation is lower than the true Fisher information based on the yes-no outcome to all m_{ad} comparisons at once, by a factor $\left(N_{\mathrm{ad}}-m_{\mathrm{ad}}\right)/N_{\mathrm{ad}}$ in this simple case; that is, the pseudo-log-likelihood Hessian here leads to an *overestimate* of true variance. While the same qualitative effect presumably extends to more complex settings with different covariates, it is by no means obvious how to extend the calculations, and we propose generally ignoring this overestimation of variance. On the whole, it is usually a worse mistake to be over-confident than to be under-confident in an estimate; model-based variance estimates tend to be biased low anyway, through ignoring structural oversimplifications in the model; and with CKMR variance can usually be reduced anyway by modest increases in sample size. For walrus in particular, the total proposed adult sample size is under 10% of the population, so the effect on standard deviations would presumably not exceed 5%.

[4]Or almost completely independent: if the two potential-offspring are born in the same or successive years, then walrus biology precludes them from both being offspring of the same mother. Another complication arises with non-lethal adult sampling; some MOP pairs will have offspring born after non-lethal sampling of the mother, in which case the mother's known lifespan can be allowed for in comparisons with other samples. We ignore all this.

0.1.1 Some code I don't want to delete yet...

```
I ran these snippets inside add_data(), using my debugger

table( table( c( MOPs, XmHSPs, selfPs))
# 1 2 3
# 1524 120 4
length( unique( samples$Me))
# 14385
length( unique( samples$Me)) - length( unique( c( MOPs, XmHSPs, selfPs)))
# 13187 this should go with "0" in the first line

# So, quite a few: 10% of samples are in pairs, 10% of those are in triplets!
# Where are most of these triplets coming from?
table( table( c( XmHSPs)))
# 1   2
# 444 27
# ... so it's mostly not from the XmHSPs

table( metab) # SelfPs before trimming the xtuples
#    1    2    3
# 9286  212    3
# ... not them either

table( table( MOPs[,1]))
#   1   2
# 419  13
# ... so it's mostly not from multiple offspring of one mother...

table( table( MOPs[,2]))
# 1
# 445
# ... good; no-one has >1 mother!

# Thus, the bulk of the 124 must be "interference" across kinships
# which is probably worth fixing, using the first lot of simple steps above.
```