

4-28-2020

A Simulation Study on Increasing Capture Periods in Bayesian Closed Population Capture-Recapture Models with Heterogeneity

Ross M. Gosky

Appalachian State University, goskyrm@appstate.edu

Joel Sanqui

Appalachian State University, sanquijat@appstate.edu

Follow this and additional works at: <https://digitalcommons.wayne.edu/jmasm>



Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Gosky, R. M., & Sanqui, J. (2019). A simulation study on increasing capture periods in Bayesian closed population capture-recapture models with heterogeneity. *Journal of Modern Applied Statistical Methods*, 18(1), eP2555. doi: 10.22237/jmasm/1556668920

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in *Journal of Modern Applied Statistical Methods* by an authorized editor of DigitalCommons@WayneState.

A Simulation Study on Increasing Capture Periods in Bayesian Closed Population Capture-Recapture Models with Heterogeneity

Cover Page Footnote

We wish to thank Ms. Vilma Sanqui for her assistance with various computational aspects of this paper.

A Simulation Study on Increasing Capture Periods in Bayesian Closed Population Capture-Recapture Models with Heterogeneity

Ross M. Gosky

Appalachian State University
Boone, NC

Joel Sanqui

Appalachian State University
Boone, NC

Capture-Recapture models are useful in estimating unknown population sizes. A common modeling challenge for closed population models involves modeling unequal animal catchability in each capture period, referred to as animal heterogeneity. Inference about population size N is dependent on the assumed distribution of animal capture probabilities in the population, and that different models can fit a data set equally well but provide contradictory inferences about N . Three common Bayesian Capture-Recapture heterogeneity models are studied with simulated data to study the prevalence of contradictory inferences in different population sizes with relatively low capture probabilities, specifically at different numbers of capture periods in the study.

Keywords: Capture-recapture, Bayesian estimation, heterogeneity, MCMC, WinBUGS

Introduction

Closed Population Capture Recapture Models with Heterogeneity Effects

Capture-Recapture studies are often performed on closed animal populations, where the population size N is assumed constant during the study. Likelihood models based upon the multinomial distribution can be used in these studies to make inferences about N . A thorough introduction of these models is given in Otis et al. (1978). These models allow animal capture probabilities to vary based on three types of effects: time effects, heterogeneity effects, and behavioral effects.

Specifically, heterogeneity effects refer to individual animals having unequal capture probabilities, with some animals being relatively easy to catch and others being relatively difficult to catch. Time effects refer to capture probabilities that vary by capture period. Behavioral effects refer to changes in an animal's capture probability after initial capture, with some animals becoming "trap-shy" or "trap-happy" (less likely, or more likely to be recaptured, respectively) after initial capture. Each type of effect can be present or absent in the model, and the subscripts used to describe the model indicate which effects are present. This leads to eight possible models, where the model with none of the three types of effects is denoted as M_0 . For example, a model with only heterogeneity effects is denoted as M_h , while Model M_{bh} refers to a model with both behavioral and heterogeneity effects present, but not time effects.

Heterogeneity models have been a focus of much continued research, with proposed approaches differing in specification of the assumed heterogeneity in the population. Pledger (2000) proposed using finite mixture models to fit heterogeneity effects in capture-recapture data and discussed the use of Akaike's Information Criterion (AIC) as a model selection tool. In these finite-mixture models, groups of animals are assumed to exist with different capture probabilities between groups, but equal capture probability within each group. Other frequentist M_h models include logistic-normal models (Coull & Agresti, 1999) and beta-binomial models (Dorazio & Royle, 2003). However, an important paper by Link (2003) showed that estimates of the population size N under M_h models depend heavily on the assumed distribution of capture probabilities in the population. He showed that different, reasonable models might fit the capture data equally well but result in very different inferences for N . Some work to resolve this issue has been done, and examples include Holzmann et al. (2006), Mao (2008), and Farcomeni and Tardella (2012). Nevertheless, for practitioners, the possibility of contradictory inferences about the population size in different heterogeneity models is both a theoretical and practical concern.

An alternative to frequentist approaches for inferences about N is the Bayesian approach. Several Bayesian M_h models for closed populations have been proposed. One example developed by Ghosh and Norris (2005) involves finite mixture models for M_{bh} , of which M_h is a special case. These models are similar to the mixture models of Pledger (2000). Additionally, Dorazio and Royle (2003) proposed a Bayesian version of the beta-binomial model, and King and Brooks (2008) proposed a Bayesian model-averaged estimation method across the eight common models in Otis et al. (1978). Bayesian statistical models such as these can be fit using WinBUGS or OpenBUGS software. Information about the WinBUGS

BAYESIAN HETEROGENEITY POPULATION MODELS

software is available in Lunn et al. (2000), and the software is freely available at <https://www.mrc-bsu.cam.ac.uk/software/bugs/>. We focus in this paper on Bayesian mixture models and the beta-binomial model. These approaches differ in their modeling of the capture probabilities in the population, with the mixture models modeling heterogeneity via a finite number of subgroups in the population, each with its own capture probability, and with the beta-binomial model allowing each animal to have its own capture probability, but with the entire capture probability distribution being modeled via the beta distribution. We chose these approaches for our study due to their common use in modeling heterogeneity among practitioners, that they represent different approaches to modeling heterogeneity between a latent class approach in the mixture model and a continuous probability distribution in the beta-binomial model, and the fact that they have been the focus of comparison in past research, such as in Dorazio and Royle (2003) and Pledger (2005).

Many simulation studies have compared the performance of different M_h models in estimating population size for simulated data sets under different data generating assumptions. For example, Pledger (2005) used a simulation to compare the performance of several M_h models, including the mixture models mentioned previously, and the beta-binomial models from Dorazio and Royle (2003). Her paper generally found that M_h models performed better than M_0 models when heterogeneity is present in estimating the population sizes in the presented simulations. When comparing different types of heterogeneity models to each other, in some cases, the performance of the beta-binomial models of Dorazio and Royle was superior to the mixture model approach, but in other cases, the opposite was true. Among the findings in this simulation, Pledger noted that a two-point mixture model underestimated N when the data generating distribution had a large amount of probability density near zero, and the beta-binomial tended to overestimate N when the data generating distribution had a high degree of skewness.

Other more recent work in evaluating performance of capture-recapture M_h models has been performed by Grimm et al. (2014), who evaluated the performance of different capture-recapture models in estimating the size of an arboreal gecko population, using reference population sizes constructed from a set of primary sampling periods as the goal of their inference. In their analyses, the mixture models did not perform as well as other heterogeneity models. Such field studies provide useful conclusions about applying methods to real data and are complementary to simulation studies, because real data do not require data generating assumptions as in simulation studies. However, the sheer volume of analyses that can be done in

simulation studies along with reasonable data generating assumptions also provide significant value in comparing different methods.

Still, while simulation studies are often useful and insightful, one study typically cannot consider all possible combinations of data generating factors in complex situations. So, while such studies provide suggestions as to model performance, general conclusions often require further comparative studies to expand the number of cases studied.

The number of capture periods does not vary substantially in capture-recapture simulation studies. Increasing the number of capture periods in a study could be expected to improve the quality of inferential conclusions about N , both with accuracy of the estimation of N and decreased uncertainty about N . Increasing the number of capture periods is within the control of the researcher, at least to some extent, and for this reason it is of interest in this study. However, non-identifiability concerns about N suggest different models may produce interval estimates that are highly specific but reach different conclusions about the population size.

The goal of this study is to compare the performance of different Bayesian M_h models via simulations, particularly with different numbers of capture periods as a significant factor in the data generating process. The aim is to determine if different models converge toward contradictory inferences when the number of capture periods increases, where a contradictory inference would be defined as the non-overlap of 95% interval estimates of N between the two models. Assessing the relative bias of the different model estimations of N in the simulation is important, as well as coverage probability and average length of 95% interval estimates of N from the posterior distribution of each of the models we consider.

Bayesian Closed Population Heterogeneity Models

Bayesian statistical models involve both the likelihood function of the data given the model parameters and the joint prior distribution of the model parameters. For closed population models let k represent the number of capture periods, and for $i = 1, 2, \dots, N$ let p_i denote the probability of animal i being captured during a capture period in the study. This capture probability only varies by animal and not by capture period in model M_h . Denote the number of animals captured exactly m times as Z_m for $m = 0, 1, 2, \dots, k$. The values Z_1, Z_2, \dots, Z_k are observable, but Z_0 is unobserved, representing the number of animals not captured during the study. Define P_{mi} as the probability that animal i is captured exactly m times for $m = 1, 2, \dots, k$, and $i = 1, 2, \dots, N$. For animal i ,

BAYESIAN HETEROGENEITY POPULATION MODELS

$$P_{mi} = \binom{k}{m} p_i^m (1 - p_i)^{k-m}$$

under the assumption that time effects and behavioral effects are not present in the population. The mixture approach from Ghosh and Norris (2005) assumes a finite mixture distribution of capture probabilities in the population with potentially r different groups in the population, each with constant capture probability for every member of the group. In this approach, the capture probability in each group $w = 1, 2, \dots, r$ is denoted as θ_w and π_w denotes the probability an animal is in group w . Because every animal belongs to exactly one of the r groups in the population,

$$\sum_{w=1}^r \pi_w = 1.$$

The beta-binomial approach from Dorazio and Royle (2003) assumes that $p_i \stackrel{i.i.d.}{\sim} \text{Beta}(\alpha, \beta)$ and from this approach we find that

$$P_m = \binom{k}{m} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + m)\Gamma(k - m + \beta)}{\Gamma(\alpha + \beta + k)}$$

for $m = 1, 2, \dots, k$. Then, under either the mixture model approach or the beta-binomial approach, write the likelihood function as

$$\Pr[Z_1 = z_1, \dots, Z_k = z_k \mid N, P_1, \dots, P_k] = \frac{N!}{(N - S)! \prod_{l=1}^k Z_l!} \prod_{l=1}^k P_l^{Z_l} \left(1 - \sum_{l=1}^k P_l\right)^{N-S},$$

where $S = N - Z_0$ is the number of animals observed during at least one capture period, and $L = 2^k - 1$.

Prior distributions for the population size N is commonly chosen as $\Pr(N = n) \propto (1 / n^\delta)$ for $n = 1, 2, \dots, N_{\max}$ with $\delta > 0$ fixed at a specific value and N_{\max} fixed at a finite value based upon prior knowledge of the population size N . In a Bayesian model, N_{\max} could be chosen to be very large if little prior information is known about the population size to express uncertainty, whereas a smaller value would indicate prior information about the population size. A uniform prior

distribution for N is obtained with $\delta = 0$, and choices of $\delta = 0.5$ or $\delta = 1$ are still non-informative prior distributions.

For the mixture models, common choices for prior distributions are: $(\pi_1, \pi_2, \dots, \pi_r)$ have a joint prior distribution that is Dirichlet(χ), and the capture probabilities $\theta_w \stackrel{i.i.d.}{\sim} \text{Beta}(a, b)$ for $w = 1, 2, \dots, r$. For the beta-binomial models, a reasonable but noninformative prior distribution for each of α and β are $\alpha, \beta \stackrel{i.i.d.}{\sim} \text{Uniform}(u_1, u_2)$ where $u_2 > u_1 > 0$.

Simulation Design and Results

Simulation Study Purpose and Design

Link (2003) showed different heterogeneity models could fit capture-recapture data comparably well via the AIC criterion, but give different inferences about N . An example was given where this could occur even when k , the number of capture periods, was twenty, which is large. The question arises how often this problem of contradictory inferences would arise as a function of k . It would be necessary to determine if many different data sets would have this contradictory inference problem, with different numbers of capture periods as one factor in the simulation design, and how the true distribution of capture probabilities in simulated populations would affect the rates of contradictory inferences among different M_h models fit to the data. This leads to comparing performance of the different models under our simulation conditions and determining if any conclusions could be reached that might recommend one of the models over the others.

Simulation Study Design

The factors in our simulation design were population size N (100, 500, 1000), theoretical average capture probability (5%, 10%), data generating model for capture probabilities (2 point mixture, 3 point mixture, beta, logistic), and the fitted model for the data (Bayesian 2 point mixture model, Bayesian 3 point mixture model, Bayesian beta-binomial model), and finally the number of capture periods ($k = 6, 8, 10, 12, 14, 16$). Excluding the fitted-model factor, there were 144 different factor combinations. Twenty-five unique data sets were randomly generated at each of these 144 factor combinations, and then fit each of these 3600 data sets under the three Bayesian Models, leading to 10800 data analyses. The data-generating

BAYESIAN HETEROGENEITY POPULATION MODELS

model for the simulation is referred to as DGM, and the theoretical average capture probability as average p_{ij} .

We chose relatively small average capture probability rates, because situations with high heterogeneity and low average capture probabilities are known to be more difficult for accurate estimation of population size (Pledger, 2005). This permits cases where, despite a large number of capture periods in the simulated data, a reasonable proportion of the population remained uncaptured, so that differences between inferences from the different models might be clearer.

SAS version 9.1 was used to generate the capture probabilities and capture histories for the different data sets. The capture probabilities were generated for a specific population as follows. For data in the two-point mixture model, we generated $\pi^* \sim \text{Uniform}(0, 1)$ and then computed $\theta_1, \theta_2 \sim F(\mu + \kappa Z_i)$, where F is the Logistic distribution function, $\mu = F^{-1}(0.05)$ or $F^{-1}(0.10)$ depending on whether the average capture probability was high or low, $Z_i \stackrel{i.i.d.}{\sim} N(0,1)$, and $\kappa = 0.75$.

For each simulated individual in the population in the two-point mixture DGM, we generated a Bernoulli random variable with π^* probability of success, which classified the animal as being part of group 1 with capture probability θ_1 , or as being part of group 2 with capture probability θ_2 . For simulated data in the three-point mixture DGM, generate $\pi_1^* \sim \text{Uniform}(0, 1)$, and then $\pi_2^* | \pi_1^* \sim \text{Uniform}(0, 1 - \pi_1^*)$, and finally $\pi_3^* = 1 - \pi_1^* - \pi_2^*$. Capture probabilities for the three groups were generated similarly to the two-point mixture simulation.

For simulated data in the beta DGM, we generated $\beta \sim \text{Uniform}(0, 80)$. Then, we generated $\alpha | \beta = \beta(19 + 1.5 Z_i)^{-1}$ when the average capture probability was low, and generated $\alpha | \beta = \beta(9 + 1.5 Z_i)^{-1}$ when the average capture probability was high. When α was very small, defined as less than 0.1, we set $\alpha = 0.1$ to keep α from being too small, which would result in a very limited number of simulated captures.

For simulated data in the logistic DGM, for each simulated population we generated $\mu = F^{-1}(0.05) + 0.5Z_p$ for simulated populations with low capture probabilities, and $\mu = F^{-1}(0.10) + 0.5Z_p$ for simulated populations with high capture probabilities, where $Z_p \sim N(0, 1)$ and is chosen once for the simulated population. Then, the individual animal capture probabilities were generated as $p_i \sim F(\mu + 0.5 Z_i)$, where again $Z_i \stackrel{i.i.d.}{\sim} N(0,1)$.

For fitting the specific Bayesian models in WinBUGS, we chose hyperparameters as $\delta = 0.5$ for the prior distribution of N , and a conditional upper bound on N as $S + 2500$. For the Bayesian Mixture models for both $r = 2$ and $r = 3$, we chose $a = b = 0.5$ for hyperparameters for capture probabilities θ_w and we chose

$\chi = (0.5, 0.5, 0.5)$ as the hyperparameters for the Dirichlet prior distribution for π_i ($i = 1, 2$ for $r = 2$; $i = 1, 2, 3$ for $r = 3$).

For fitting the Beta-binomial models in WinBUGS, we chose $u_1 = 0.1$ and $u_2 = 80$ as hyperparameters for the prior distribution of α, β . Three Markov Chain Monte Carlo (MCMC) chains were simulated with initial values for the parameters dispersed among the three chains. This choice follows the recommendations for multiple chains to be used at dispersed starting values of the parameters, as discussed in Gelman et al. (2014), for example. Then, 10000 observations were sampled from the posterior distribution using each of the three chains but discarded the first 2500 samples from each chain to allow the convergence of the MCMC chains to a stable distribution. The posterior distribution estimates were based on 22500 total samples. The Brooks-Gelman-Rubin statistic in WinBUGS was used (Brooks & Gelman, 1998) to check the convergence of the posterior distributions of the model parameters.

Simulation Study Results

For the simulation described in the previous section, use the median of the posterior distribution of N as a point estimate of N due to the right-skewed nature of most of the posterior densities of N in the simulation. Consider the effects of the different factors on each of the following outcome measures: (i) the relative bias of the posterior median of N , (ii) coverage probability of the 95% posterior intervals for N in this simulation, (iii) length of 95% equal-tail posterior intervals for N , and (iv) the percentage of data sets for which the 95% posterior intervals for N for the three Bayesian models overlap. The first three outcome measures assess each model fit to each data set separately, and can be interpreted as assessing the bias, reliability, and precision of each model for a data set. The fourth outcome measure is intended as a comparison of the models to each other across one data set, to see the rate at which contradictory inferences occur.

Summaries of the simulation results for the effects of the design factors on outcome measures (i) to (iii) are presented in Figures 1 to 3. Each figure is an interaction plot to represent the mean value of an outcome measure as a function of each pair of the simulation design factors.

In each of these figures, the mean outcome measure for the simulated data sets is plotted against each pair of factors in the simulation design. For example, the upper left graph in Figure 1 shows the mean relative bias of the posterior median of N for data sets generated at each of the three levels of N , and for each of the three models fit to the data. The other graphs within Figure 1 present similar information

BAYESIAN HETEROGENEITY POPULATION MODELS

about the mean relative bias for other combinations of simulation design factors. From the plots in Figure 1, we can see that the relative bias is positive for $N = 100$ but decreases with N for all three models. The beta-binomial model fits are somewhat positively biased for all three population sizes, while the 3-point mixture model's relative bias approaches 0% at $N = 1000$, and the 2-point mixture model is slightly negatively biased at $N = 1000$.

This positive bias is consistent with the results of Pledger (2005) because for most of our data sets, average capture probability is relatively small and the skewness of capture probabilities in some data sets is large, which led to a positive relative bias in estimation of population size in the simulation results described in that paper. The top graph in the second column plots the average biases via both the data generating model and the fitted model, which allows for assessments when the fitted model matches (and does not match) the data generating assumptions.

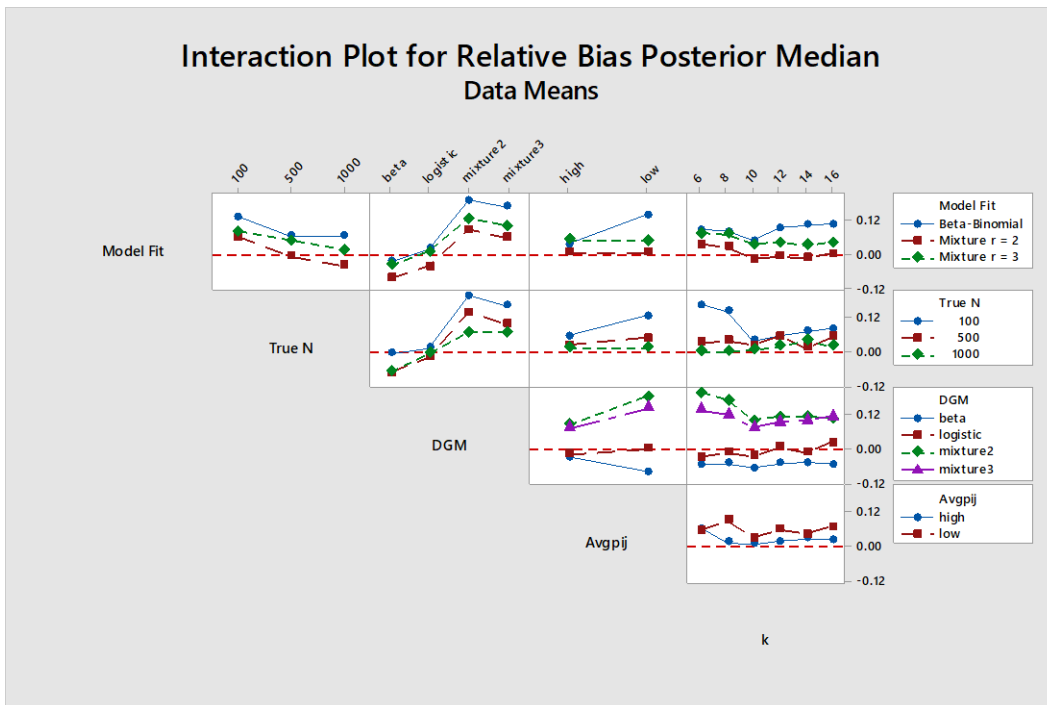


Figure 1. Effects of simulation design factors on relative bias of the posterior median

For capture probabilities generated under a continuous probability distribution, the relative bias is modestly negative overall but fairly close to zero for the beta-binomial and three-point mixture model fits. For data generated under the mixture assumptions, all models show positive relative bias. The mixture models have biases that are closer to zero but still are positive. The third column shows that when average capture probability in the simulated data set is low, the average relative bias is positive and higher overall, but this effect varies in each of the plots. For example, the relative bias increases for the beta-binomial model fit for low average capture probabilities but is consistent for the mixture models. The plots in the fourth column show the effect of k , the number of capture periods. In terms of relative bias, there is not a significant difference at the different levels of k overall. A few noteworthy points are that the largest relative biases occur when k is small and the simulated population size is also small, and when k is small and the capture probabilities are generated from a mixture distribution. The largest relative biases occur (on average) when N is 100, and when k is 6 or 8. When k is 6, the bias is high regardless of whether the average capture probability is low or high, and when k is 8, the bias is higher for the low capture probability data sets in the simulation. These positive biases persist, but decline somewhat, as k increases in the study for these smaller populations, suggesting that the benefit of increasing k in practice may be more significant for smaller populations with smaller capture probabilities.

In [Figure 2](#), which measures the coverage probabilities in the simulation, a vertical line is placed in each graph at 95% for reference. From this figure, the coverage probabilities drop in the simulation as the true N increases from 100 to 1000, although the coverage probability remains consistent and strong in the three-point mixture model. The second column in the figure shows that the lowest coverage probability occurs when the beta-binomial model is fit to data generated under a two-point or three-point mixture model. Overall, the coverage probability of the three-point mixture model is consistently strong and slightly above the 95% nominal level. In the third column in [Figure 2](#), the coverage probabilities remain fairly consistent for data with relatively high or low capture probabilities with slightly higher coverage probabilities at low capture probability levels. Finally, in the fourth column of [Figure 2](#), the coverage probabilities drop (in aggregate) as the number of capture periods increase from six to sixteen. However, this decline does not occur for the three-point mixture model, and it is less notable when the true N is smaller (100 or 500) compared with the larger value of 1000. The decline in coverage probabilities as k increases is fairly consistent across the different data generating assumptions, and average capture probability levels.

BAYESIAN HETEROGENEITY POPULATION MODELS

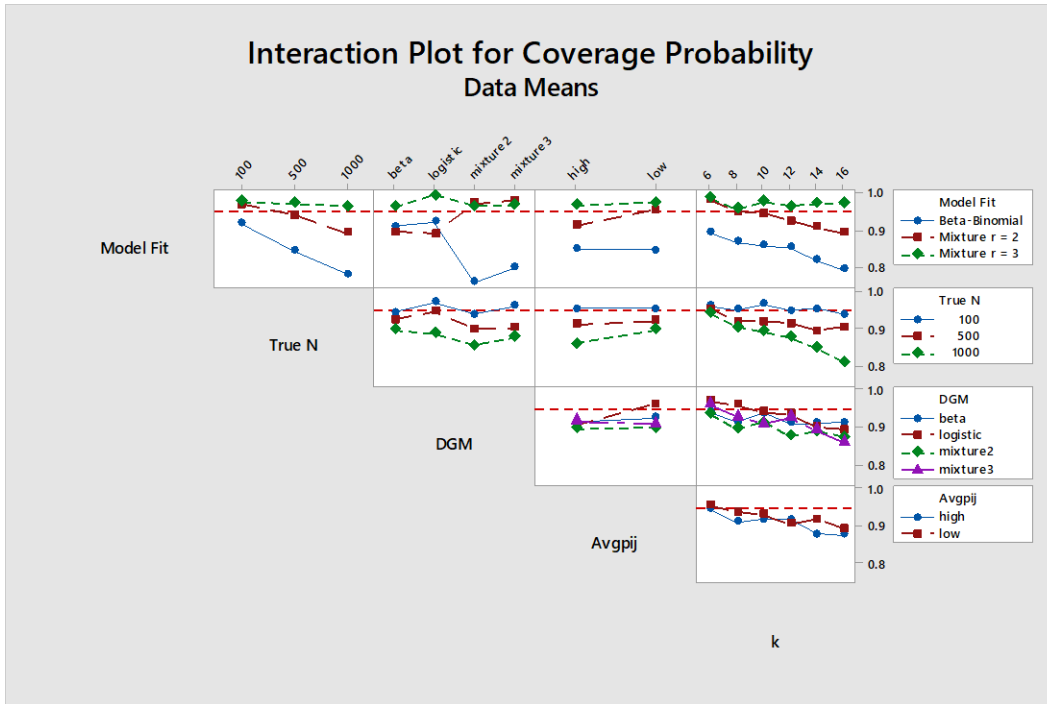


Figure 2. Effects of simulation design factors on the coverage probability of 95% equal-tailed posterior intervals for N

In Figure 3 the outcome measure is the length of the 95% equal tail posterior intervals for N in the simulation. From this figure, initially the average interval length increases with the true population size in the simulation, although this observation is not terribly surprising, as one might expect interval lengths to increase for larger population sizes. In the second column of Figure 3, overall the mean lengths of the posterior intervals are consistent for the different data generating assumptions. The average length is consistently shorter for the beta-binomial model than for the mixture models. The third column in Figure 3 shows that the length of the intervals is somewhat longer on average for data generated with lower capture probabilities than for higher capture probabilities. In the fourth column average length of the posterior interval decreases with k across the other simulation design factors.

Taken together, Figures 1, 2, and 3 provide the following general conclusions from this simulation study. Collectively the three-point mixture model performed the best in our simulations. There is some positive bias in the posterior median of N for this model, but it decreases with population size. Secondly, the coverage

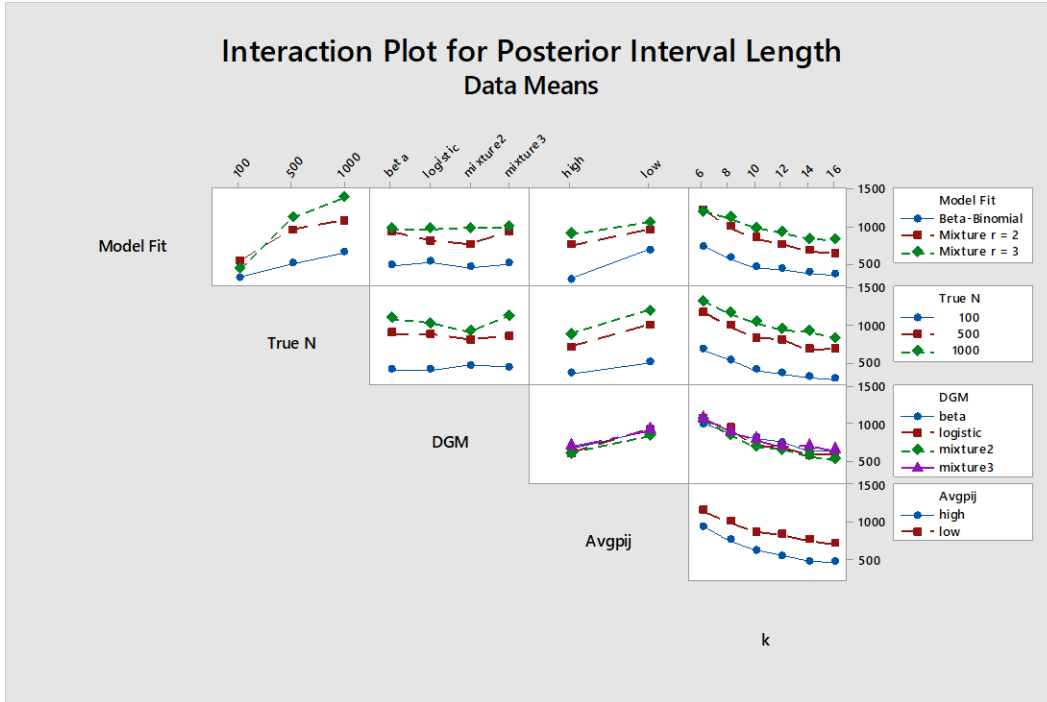


Figure 3. Effects of simulation design factors on length of 95% equal-tailed posterior intervals for N

probability of this model remained above the nominal 95% level even as k increased and as the average length of the 95% posterior interval decreased. The two-point mixture model performed relatively well and had smaller relative bias than the three-point mixture model for data generated under a mixture-distribution. However, the coverage probability of the 95% posterior interval for this model decreased more notably below the nominal 95% level as k increased. The beta-binomial model did not perform as well as the mixture models in our simulation, having higher relative bias than the other models and poorer than reported coverage probabilities for the 95% posterior intervals for N . Specifically, the beta-binomial model did not fit data generated under mixture distribution assumptions well, as can be seen in Figures 2 and 3.

From Figures 1 through 3, some general conclusions can be made about the effects of increased k in our simulation. In our simulation study, larger values of k did not result in significant changes in the relative bias of the posterior median, with the only notable reductions in the relative bias occurring when the True N in the simulation is 100, or for data generated under the mixture distribution assumptions.

BAYESIAN HETEROGENEITY POPULATION MODELS

The coverage probability declined with increasing k in general, especially for the beta-binomial model, and when the True N was 1000. As noted previously, the coverage probability of the posterior interval stayed consistent with the three-point mixture model as k increased in the simulation. The decrease in the average length of the 95% posterior interval as k increased in the simulation study, and that this occurs consistently across the different design factors in the study.

An additional concern for practitioners is the possibility of contradictory inferences in these models, which was a concern introduced by Link (2003). In terms of the simulation, a data set produces a contradictory inference if the 95% equal-tail posterior intervals from two or more of the models do not overlap each other. When different models produce contradictory inferences in these situations, Link also showed that model selection criteria, such as Akaike's Information Criterion could provide equal measures of fit in some of these cases. When this occurs, the assumption of capture probability distribution in the population becomes important, but in practice, it is not verifiable. However, if these situations are somewhat rare in practice, then they are less of a concern to practitioners.

The rate at which these contradictory inferences occurred was examined. A high rate of occurrence would suggest concern for practitioners, forced to choose between models giving different conclusions, while a low rate of occurrence would suggest that the different models generally reach similar conclusions.

Table 1 shows the percentage of simulated data sets that have one, two, and three overlapping posterior intervals from each of the three fitted models (beta-binomial, and each of the two mixture models). A non-contradictory inference occurs when all three intervals overlap, and a contradictory inference occurs when fewer than three of the intervals overlap. In the simulation study, at least one overlap occurred for every data set. The percentages in Table 1 are calculated by pairwise comparison of the posterior intervals for the three models for each data set, and the results are grouped separately by each simulation design factor (k , True N , DGM, and Average Capture Probability) respectively. From Table 1, as the number of capture periods k increases, the rate of contradictory inferences rises modestly although in the vast majority of cases the three intervals overlap. Some additional investigation of these occurrences showed that in all cases but one, the posterior intervals from the two mixture models overlap, and the contradictory inference for the data set arises from one or both of the posterior intervals from the mixture models not overlapping with the interval from the beta-binomial model. The intervals from the mixture models overlap in virtually all of the simulated data sets in the study.

In Table 1, the percentage of contradictory inferences is higher when N is larger in the simulated data set, and for higher theoretical capture probabilities in the simulated data set (10% theoretical capture probability rate). These increased rates largely correspond with the reduced mean length of the posterior intervals we saw in Figure 3 in these cases. Also in Table 1, the rate of contradictory inferences is higher for data generated under mixture distribution assumptions. Collectively the percentages of cases where all three intervals overlapped is high (at least 94% of cases in our simulation in all conditions). This is encouraging but notice a tendency toward contradictory inferences as the number of capture periods increases, implying that the different models converging to different answers in these cases is a real possibility.

For all four of these explanatory factors, differences in the rates of contradictory inferences are statistically significant when analyzed via a frequentist binary logistic regression model.

Table 1. Percentages of overlap among 95% equal-tailed posterior intervals for N for each of the three Bayesian models

		Overlapping intervals (compared pairwise)			
K	Zero overlaps	One overlap	Two overlaps	Three overlaps	
6 captures	0.00%	0.17%	0.17%	99.67%	
8 captures	0.00%	0.00%	0.83%	99.17%	
10 captures	0.00%	0.17%	1.33%	98.50%	
12 captures	0.00%	0.83%	1.33%	97.83%	
14 captures	0.00%	0.17%	3.00%	96.83%	
16 captures	0.00%	0.83%	4.00%	95.17%	

True N	Zero overlaps	One overlap	Two overlaps	Three overlaps
100	0.00%	0.00%	0.08%	99.92%
500	0.00%	0.00%	1.00%	99.00%
1000	0.00%	1.83%	4.25%	94.67%

Average p_{ij}	Zero overlaps	One overlap	Two overlaps	Three overlaps
Low	0.00%	0.22%	1.11%	98.67%
High	0.00%	0.50%	2.44%	97.06%

DGM	Zero overlaps	One overlap	Two overlaps	Three overlaps
Beta	0.00%	0.89%	3.56%	95.56%
Logistic	0.00%	0.11%	2.44%	97.45%
Mixture $r = 2$	0.00%	0.22%	0.56%	99.22%
Mixture $r = 3$	0.00%	0.22%	0.56%	99.22%

Real Data Analysis

To illustrate the results of increased capture periods for real data sets, two real data sets were analyzed using each of the three models. The first data set is a deer mice data set as discussed in Otis et al. (1978). The second is a data set from Karanth et al. (2004), who used photographic and capture-recapture methods to estimate the density of a tiger population in Central India.

The deer mice data had 38 mice captured over six nightly capture periods. Assume closure of the population for this analysis. This data set was analyzed using each of the three models by first analyzing only the first two capture periods, then analyzing the first three capture periods, and so on, sequentially until all six capture periods were analyzed. This demonstrates how the estimates of N and the 95% equal-tailed posterior intervals for N change for each model as k increases.

Presented in Table 2 are the results of the analysis of the deer mice data. The posterior medians of N are consistently a bit higher for the mixture models than for the beta-binomial model, but at the end of all six captures the point estimates are close for all three models. Given that 38 mice were eventually captured in this study, the posterior median is below the true population size for all models when $k = 2$, and for the beta-binomial model up through $k = 5$ captures. The 95% equal-tailed posterior interval for N is below the true population size for the beta-binomial model when $k = 2$.

Table 2. Posterior median and 95% posterior intervals (in brackets) of population size for deer mice data

k	Fitted model		
	Mixture $r = 2$	Mixture $r = 3$	Beta-binomial
2 captures	31 (24, 657)	34 (24, 345)	27 (24, 36)
3 captures	43 (30, 855)	46 (31, 434)	33 (30, 41)
4 captures	39 (33, 497)	42 (33, 271)	34 (33, 40)
5 captures	38 (36, 90)	40 (36, 133)	37 (36, 50)
6 captures	43 (39, 195)	46 (39, 215)	43 (39, 82)

Another notable difference is that the length of the 95% equal-tailed posterior intervals for the mixture models being consistently larger than that of the beta-binomial model. The length of the posterior intervals for the mixture models decreases as the number of capture periods increases but remains larger than that of the beta-binomial model. These results are consistent with the results of our simulation study in that the length of the posterior intervals for the beta-binomial model was the smallest, but it differs in that the posterior median of N for the beta-binomial model is below that of the mixture models. The beta-binomial model often had a narrower 95% posterior interval than the mixture models, but also had lower coverage probability than the other models. The intervals in this analysis are consistent with this finding in that the posterior intervals are narrower for the beta-binomial model. It cannot be evaluated whether the intervals are accurate because the true population size is not known.

Table 3. Posterior median and 95% posterior intervals (in brackets) of population size for tiger data

k	Fitted model		
	Mixture $r = 2$	Mixture $r = 3$	Beta-binomial
4 captures	11 (6, 142)	11 (6, 95)	8 (6, 47)
5 captures	13 (7, 220)	13 (7, 125)	9 (7, 40)
6 captures	20 (9, 327)	20 (9, 191)	16 (9, 103)
7 captures	20 (9, 282)	20 (9, 179)	16 (9, 124)
8 captures	19 (9, 252)	19 (9, 169)	17 (9, 140)
9 captures	19 (9, 237)	19 (9, 158)	18 (9, 153)
10 captures	24 (11, 274)	23 (11, 186)	23 (11, 215)
11 captures	24 (11, 254)	23 (11, 178)	25 (11, 226)
12 captures	18 (10, 140)	18 (10, 108)	19 (10, 133)
13 captures	24 (13, 262)	24 (13, 183)	25 (13, 168)
14 captures	24 (13, 247)	24 (13, 177)	25 (13, 175)
15 captures	26 (13, 297)	26 (13, 205)	23 (13, 144)

BAYESIAN HETEROGENEITY POPULATION MODELS

The tiger data set had eleven distinct tigers photographed during fifteen capture periods. For this data, Karanth et al. (2004) determined that model M_h was the most plausible model due to a combination of goodness of fit tests and knowledge of the spatial patterns of tiger movement and other factors. For this data set, we proceed similarly to the deer mice data by analyzing each model at each capture period using only the data obtained up through that period. Table 3 provides the results of this analysis. For this data set, the first possible value of k for analysis was $k = 4$, which was the first capture period with a recapture. There were no tigers sighted in capture periods, 7, 8, 9, 11, and 14, which explains the stability of the posterior medians at these points of the sequential analysis. In comparing the different models, the posterior median of N of each of the models is generally similar at each value of k . The length of the 95% equal-tailed posterior interval for N for the beta-binomial model is shorter than for the mixture models. This data set is again consistent with the results of our simulation study with regard to the length of the posterior interval for N , but in this example, the posterior medians of all three models are again close to each other, which differs slightly from the results of our simulation study.

In both of the real data analyses, increasing k leads to an initial increase in the posterior median of N to presumably more accurate levels, as we know the posterior medians of N for the smallest k value in each data set are below the true population size. These posterior medians become more stable as k increases further, which differs somewhat from our simulation results, where the relative biases did not change much with increasing k . This may be attributable to the fact that the smallest k in our simulations was $k = 6$, and the analyses of these real data sets started at $k = 2$ and 4, respectively.

These population sizes are both likely smaller than in our simulations, although in many cases the 95% posterior interval does contain larger values that overlap with our simulated population sizes. Across both data sets, at each level of k , the posterior medians of N of each model are similar. The difference between the models is seen primarily in the width of the 95% posterior intervals for N . From a practical perspective, interpreting these results depends on the believability of the upper bound of the posterior interval, as a wide posterior interval gives less precision in the population size estimate. In cases like these, though, a practitioner may have a stronger prior opinion about the population size than we did in this analysis, and this opinion can be incorporated into the prior distribution for N , which could shorten the posterior interval of each of the models. The posterior distributions for N in each of these models is right-skewed, so the interval length

can be shortened by lowering the probability level of the interval or by choosing an interval other than an equal tailed interval.

Discussion

Capture-recapture modeling when heterogeneity is present in the population presents a significant challenge for population size estimation. Practitioners face challenges not only in deciding among the eight closed population models described in Otis et al. (1978) but also challenges deciding between different heterogeneity models. These challenges can be alleviated somewhat if a practitioner has information or insight about the distribution of capture probabilities in the population based upon knowledge of the population being studied. However, the possibility of different heterogeneity models arriving at different conclusions is a concern for practitioners as well.

The simulation results should provide some useful guidance in these considerations, but as with all simulation studies, only a small number of factors (and levels of these factors) are considered relative to all possible factors. For example, the focus was on relatively smaller capture probabilities in our simulation design. For each data set the parameters governing the capture probabilities were also randomly generated, and so some data sets had more heterogeneity than others due to this approach.

The three-point mixture model generally performed the best. Despite some positive relative bias at all population sizes, the model showed strong coverage probability of its 95% equal-tailed posterior interval for N along with decreasing average length of that interval as the number of capture periods k increased. This performance occurred across data sets generated under both finite mixture assumptions and continuous assumptions about the capture probabilities in the population, which may alleviate some concern from a practitioner who is unsure about the type of heterogeneity present in the population.

The beta-binomial model did not perform as well, with results that were more precise than the other models (with smallest average posterior interval length) offset by lower than desirable coverage probability of the posterior interval for N as k increased. Regarding even at the largest number of capture periods in the simulation ($k = 16$), the relative bias for the beta-binomial model was relatively large on average for data generated under mixture distribution assumptions. This occurs to some extent due to more notable right skewness in the simulation results for the posterior median of the beta-binomial model than in other models, as the overestimation from this model was more extreme in magnitude than for the other

BAYESIAN HETEROGENEITY POPULATION MODELS

models. The two-point mixture model performed reasonably well, having average relative bias a bit closer to zero than even the three-point mixture model. However, the coverage probability of the posterior interval for N declined more notably below the 95% nominal level as k increased.

The other simulation design factors also affected the simulation results. Population size was a factor, as the relative biases of the posterior median of N were highest when N was 100 in the simulations and when k was 6 or 8, which were the smallest two levels of k in our simulation. For larger population sizes in our simulation, when N was 1000, the relative biases were relatively small, but the coverage probability of the posterior intervals declined noticeably below the nominal 95% level as k increased. The data generated under mixture distribution assumptions caused some positive bias in the posterior median of N for all the models, and the beta-binomial model had notably low coverage probability in the 95% posterior interval for N for data generated under mixture assumptions. Data sets generated under the lower capture probability assumptions had somewhat higher positive bias than the higher capture probability assumptions, along with also having slightly higher coverage probability and length of the posterior interval.

Increasing k did not significantly impact the relative bias, but it did result in increased precision (reduced length) but decreased accuracy (reduced coverage probability) for the 95% equal-tailed posterior interval for N in general, albeit not for the three-point mixture model, as mentioned before. This tendency for increasing k to be associated with increased precision but decreased accuracy also holds across the different population sizes, for the different data generating models, and for the different capture probability levels (low, high).

Increasing k was associated with an increasing but not alarming rate of contradictory inferences between models. This introduces some caution for a practitioner, but the outcomes described earlier should suggest that a researcher can increase k , within the scope of a study, and find that even when a specific heterogeneity model is not known ahead of time, the study should lead to useful conclusions about the population size N . However, the results concur with those of Link (2003) in that the estimates of N from the various models do not necessarily converge to the same answer as k increases.

One of the main factors noted in the coverage probability performance of these models in the simulation study was that for larger N and for larger k , the coverage probability of the beta-binomial model was lower than the reported 95% level when fitting data generated under mixture model assumptions. The coverage probabilities for the beta-binomial model were much closer to the nominal 95% rate for data generated with capture probabilities on a continuous distribution. This can

be seen in Figure 2 and suggests model mis-specification of this type became more prominent in larger populations and with more capture periods. This suggests for larger k and N , practitioners should exercise some caution in the conclusions of a beta-binomial model unless there is reason to believe the capture probabilities in the population are described by a continuous distribution rather than a finite mixture distribution.

When comparing the simulation results with the real data analyses, the posterior medians of all three models were close to each other as k increased, but the lengths of the posterior intervals were considerably wider for the two mixture models compared with the beta-binomial model. This is consistent with the simulation results, although the presumed population sizes for these studies are likely smaller than those in the simulation study.

A practical recommendation for a practitioner is to consider whether precision (a shorter posterior interval for N) or accuracy of the posterior median of N is more important in the context of the study. If smaller relative bias is more important in terms of the conclusions of the study, then increasing k to approximately 10 improved the average relative bias across all the models when N was smaller (100), but a smaller number of capture periods were sufficient when N was 500 or 1000. The mixture models were somewhat more robust in accurate estimation of N when fit to data generated under continuous capture probabilities than the beta-binomial was when fitting data with capture probabilities from the mixture distributions. However, if precision of the posterior interval for N is most important, then the beta-binomial model was more precise on average. This precision comes with a cost (higher average relative bias, and lower coverage probability in our simulation results), but if these costs are of lower relative concern, then increasing k and using a beta-binomial model could be chosen reasonably by a practitioner.

This simulation is somewhat preliminary, because more models can certainly be fit to these data, such as logistic-normal models, or the model-averaging approach proposed by King and Brooks (2008). Also, more factors may be varied within this simulation structure, as additional factor levels of r , k , N , and average p_{ij} can be studied. Furthermore, model selection continues to be a pertinent question for research and practitioners, and a simulation that focuses on both estimation and model selection may be of interest to further illuminate the problems of contradictory inferences in heterogeneity models.

References

- Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434-455. doi: 10.1080/10618600.1998.10474787
- Coull, B. A., & Agresti, A. (1999). The use of mixed logit models to reflect heterogeneity in capture-recapture studies. *Biometrics*, 55(1), 294-301. doi: 10.1111/j.0006-341x.1999.00294.x
- Dorazio, R. M., & Royle, J. A. (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics*, 59(2), 351-364. doi: 10.1111/1541-0420.00042
- Farcomeni, A., & Tardella, L. (2012). Identifiability and inferential issues in capture-recapture experiments with heterogeneous detection probabilities. *Electronic Journal of Statistics*, 6, 2602-2626. doi: 10.1214/12-ejs758
- Gelman, A., Carlin, J. B., Stern, H. S., Duoson, D. B., Ventari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd edition). Boca Raton, FL: CRC Press. doi: 10.1201/b16018
- Ghosh, S. K., & Norris, J. L. (2005). Bayesian capture-recapture analysis and model selection allowing for heterogeneity and behavioral effects. *Journal of Agricultural Biological and Environmental Statistics*, 10(1), 35-49. doi: 10.1198/108571105x28651
- Grimm, A., Gruber, B., & Henle, K. (2014). Reliability of different mark-recapture methods for population size estimation tested against reference population sizes constructed from field data. *PLoS One*, 9(6), e98840. doi: 10.1371/journal.pone.0098840
- Holzmann, H., Munk, A., & Zucchini, W. (2006). On identifiability in capture-recapture models. *Biometrics*, 62(3), 934-936. doi: 10.1111/j.1541-0420.2006.00637_1.x
- Karanth, K. U., Chundawat, S. R., Nichols, J. D., & Kumar, N. S. (2004). Estimation of tiger densities in the tropical dry forests of panna, central india, using photographic capture-recapture sampling. *Animal Conservation*, 7(3), 285-290. doi: 10.1017/s1367943004001477
- King, R., & Brooks, S. P. (2008). On the Bayesian estimation of a closed population size in the presence of heterogeneity and model uncertainty. *Biometrics*, 64(3), 816-824. doi: 10.1111/j.1541-0420.2007.00938.x

Link, W. A. (2003). Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics*, 59(4), 1123-1130. doi: 10.1111/j.0006-341x.2003.00129.x

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10(4), 325-337. doi: 10.1023/a:1008929526011

Mao, C. (2008). On the nonidentifiability of population sizes. *Biometrics*, 64(3), 977-979. doi: 10.1111/j.1541-0420.2008.01078.x

Otis, D. L., Burnham, K. P., White, G. C., & Anderson, D. R. (1978). *Statistical inference from capture data on closed animal populations*. Washington, D.C.: Wildlife Society.

Pledger, S. (2000). Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics*, 56(2), 434-442. doi: 10.1111/j.0006-341x.2000.00434.x

Pledger, S. (2005). The performance of mixture models in heterogeneous closed population capture-recapture. *Biometrics*, 61(3), 868-873. doi: 10.1111/j.1541-020x.2005.00411_1.x