

An Efficient Two Sample Capture-Recapture Model with High Recaptures.

Danjuma Jibasen¹ and Yusuf J. Adams²

This paper proposed an efficient two sample capture-recapture model (M_a) with high recaptures and compared it with the existing models like the model of no factor effect (M_o), behavioral response model (M_b) and the Petersen model (M_s), using simulated data. We found that the Petersen model provides a better estimate of the population size when the observations follow a hypergeometric distribution and the population is overestimated when recapture n_{11} is high. It was also found that the proposed model provides a better estimator of the population size than the existing ones when the recapture is high. This model is particularly useful in situations where individuals respond positively to capture. This model can be applied to the estimation of the population of a locality and can be used to check inflated or disputed census figures effectively.

Keywords: Captures-recapture, High recaptures, Fewer recaptures, Petersen model,

JEL Classification: C13, C15, C52, C63

1.0 Introduction

The idea of obtaining information about a population by marking or labeling some of its members can be traced back several centuries. However, two noteworthy applications to ecology were made by Petersen (1896) in (Seber, 1982a), using tagged plaice, and Lincoln (1930) in (Seber, 1982a) who used band returns to estimate the size of the North American waterfowl population. Since the 1930s, Capture-recapture (C-R) methods have been used by ecologists to study different animal populations that are difficult to enumerate and followed over time. After trapping, marking, and releasing individuals on one or more capture occasions, inference about the population is obtained by comparing numbers of marked and unmarked individuals captured at subsequent times (Bonner and Schwarz, 2006). The technique has been recently considered in relation to estimating the size of human population from several incomplete lists of the population, (Seber (1982a); IWGDMF, (1995a) and Xu *et al.* (2007)).

¹ Department of Statistics and Operations Research, Modibbo Adama University of Technology Yola djibasen1@yahoo.co.uk

² Statistics Department, Central Bank of Nigeria YJADAMS@cbn.gov.ng

The objectives are to estimate the number of individuals in the population that are never observed, and thereby obtain an estimate of the population size (N), as well as information on survival rates (Dorazio and Royle, 2003). When more than one recapture or more than two records systems are used, the usual descriptions are multiple-recapture, multiple-system and multiple record systems methods. This paper proposed a two sample capture-recapture model for estimating populations that experience high recaptures, this proposed model was compared with other existing two sample models such as the classical Lincoln-Petersen model (here called the Petersen model).

2.0 Literature Review

The Petersen model is the starting point for all capture-recapture methods and is the classical method for two sample C-R problems. It has been used extensively to estimate population size for animal and human populations; in the latter case it is known as dual-list method, Schwarz and Seber (1999). Chao and Tsay (1998) presented reviews of its use to estimate census undercounts. Rockwood and Whiting (1992) presented an example where the method is used to estimate the number of hunter-trips where the first sample is a self-completed questionnaire and the second sample is obtained from a telephone survey of license holders. Laska and Meisner (1993) illustrated the use of a modified Petersen type model where the first sample was replaced by a set of planted individuals who were added to the initial population. The second sample recorded the total number of planted individuals observed.

The assumptions and properties of the Petersen model are well known; (1) the population is closed to additions or deletions, (2) all animals are equally likely to be captured in each sample, and (3) marks are not lost and are not overlooked by the observer. Recent work has concentrated on developing variants of the Petersen model to account for violation of the assumptions. Gaskell and George (1972) presented a Bayesian modification of the Petersen estimator and showed that the Petersen performs poor with fewer recaptures. Skalski and Robson (1982) discussed a modification of the Petersen model where there is a series of permanent removals that compose the second. They proved that the Petersen performs wells compare to other method, unless the removals rate is very high. Wolter (1990) presented an extension of the Petersen model where there are two types of animals; this method allows for dependence between the two samples, he applied this extension to the census under count problem. Rajwani and Schwarz (1997) showed how to modify the

Petersen estimator to account for tags that were overlooked during the initial recovery sample. (Seber *et al.* 2000) developed a multinomial tag loss model for dual list method, based on the assumption that the two lists are independent.

The assumption of equal catchability causes problems in most applications. There may be inherent variability (heterogeneity) in capture probabilities of individuals due to sex, age, or other factors. There could also be positive response (trap- happy) or negative response (trap-shy).

The effect of heterogeneity, can sometimes be reduced by stratification, an example of this was given by Doscher and Woodward (1983), though Darroch (1961) was the first to work on stratification to remove the effect of heterogeneity. Other work on this can be seen in Plante (1990); IWGD MF (1995a); Plante *et al.* (1998) and Schwarz and Taylor (1997).

Negative response to the initial capture leads to under estimation of the actual population size, this problem can be handle using removal method as in fisheries and with improved sampling effort (Seber, 1982b). Jibasen (2011) developed a two sample C-R model for estimating Elusiveness (negative response) in epidemiologic events.

On the other hand, positive response leads to under estimation of the population size, this problem receive less attention in C-R literatures (Otis *et al.* (1978) and Pollock *et al.* (1990)).

In this work therefore, we develop a two sample C-R model for positive response (that is, high recaptures) for demographic and epidemiologic problems. Positive response in demographic and epidemiologic events is a situation where people are encourage to respond to vaccinations, treatments, payment of taxes, registration of vital statistics or even enumeration where people are vigorously encourage to participate. This method can be used to check inflated or even disputed census figures, where the main census will form one list (sample 1) and a second pre or post census will form the second list (sample 2).

C-R models are classified based on factors responsible for catchability. Key of these models are; heterogeneity model M_h , time variation M_t and behavioral response effect M_b . These factors; “heterogeneity”, denoted by ‘ h ’, “time variation”, denoted by ‘ t ’ and “behavioral response”, denoted ‘ b ’ further gave

rise to a number of other models; M_{th} , M_{tb} , M_{hb} , and M_{thb} (Otis *et al.* (1978) and Pollock *et al.* (1990)). Others include; modeling the effects of covariates on the individual by Yip *et al.* (1996), generalized linear models introduced by Fienberg (1972) and developed by Cormack (1989), mixture models (Norris and Pollock, 1996 and Pledger, 2000), models incorporating tag loss (Seber and Felton, 1981; Seber, 1982b and Seber *et al.* 2000), likelihood based approach introduced by Huggins (1991) and so on. Model M_o is also considered for situations where variation of any kind is not suspected. Here, models M_o , M_b , and M_t (refers to as M_s) were compared with the proposed model M_a .

3.0 Methodology

The usual capture-recapture census comprised of s -samples, list, or rosters. Those individuals included in the i^{th} sample typically are divided into two groups: those who have been identified, tagged, or marked in preceding samples, and those who have not. The only information provided by traditional capture-recapture data regarding an individual in the i^{th} sample is whether or not he was in some previous sample (Bishop *et al.* 1975).

In this work, we focused on situations where there is a set of two samples (or systems), lists or rosters, which may or may not be ordered sequentially in time. The individuals are assumed to be uniquely labeled or identified so that we can determine whether or not an individual is present or not in each of the two-systems considered as capture occasions.

The observed individuals are categorized into 2^2 cross-tabulation; the cross-tabulation has one missing cell corresponding to absence in all 2-systems. A typical assumption of C-R models is that the 2-samples are independent; this may not be the case with lists, since patients or suspected convicts may be referred to another system for treatment or counseling or for legal action as the case may be.

3.1 Notations and Definition of terms

The following notations are defined in the language familiar in the C-R literatures:

N = total population size,

s = number of (capture occasions) systems; s for systems replacing t for time.

$p_{1.}$ = capture probability of any individual on the 1st capturing system
(a measure of performance of the 1st capturing system)

$p_{.1}$ = capture probability of any individual on the 2nd capture system
(a measure of performance of the 2nd capturing system)

$n_{1.}$ = number of individuals caught in the 1st capture system (Number on list 1)

$n_{.1}$ = number of individuals caught in the 2nd capture system (Number on list 2)

n = total number of individuals on both lists.

r = number of different individuals on both lists (that is, without replacement)

\hat{N}_o = Estimated population using model M_o

N_s = Estimated population using model M_s

\hat{N}_a = Estimated population using model M_a

\hat{N}_b = Estimated population using model M_b

Catchability – according to Cormack (1966), catchability is the probability with which an animal in the catchable population places itself in a position where the experimenter is able to catch it.

“Listability” – the chance that an individual in the “listable” population will be on a list.

3.2.1 Modelling Listability

The basic C-R data can be conveniently expressed in matrix form as;

$$A = \begin{bmatrix} n_{11} & n_{12} & \cdots & n_{1k} \\ n_{21} & n_{22} & \cdots & n_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ n_{N1} & n_{N2} & \cdots & n_{Nk} \end{bmatrix}$$

where,

$$A = (n_{ij}), \text{ for } i = 1, 2, \dots, N; j = 1, 2, \dots, k$$

$$n_{ij} = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ individual is on the } j^{\text{th}} \text{ list (system)} \\ 0 & \text{otherwise} \end{cases}$$

The A matrix is a simple way to record the individuals in C-R experiments. Row i gives the listing results of individual i , while column j gives the results of the j^{th} recording system.

Matrix \mathbf{A} may not be observed in its entirety because some individuals may never be seen or captured or listed. Therefore those rows of \mathbf{A} are all zeros, but one cannot tell how many such rows there are.

The data for C-R experiments are samples, this imposes the need for a probabilistic treatment of the data to derive correct estimation and inference procedures, and these are presented below.

3.2.2 The Generalized Two sample Capture-recapture Model

The joint probability density function (pdf) that described the probability of individuals belonging to a capture-recapture category with no restriction on the capture probabilities is given below,

$$p(n_{1.}, n_{.1}, n_{11}) = \binom{N}{n_{1.}} p_{1.}^{n_{1.}} (1 - p_{1.})^{N - n_{1.}} \times \binom{n_{1.}}{n_{11}} c^{n_{11}} (1 - c)^{n_{1.} - n_{11}} \times \binom{N - n_{1.}}{n_{.1} - n_{11}} p_{.1}^{n_{.1} - n_{11}} (1 - p_{.1})^{N - r} \quad (1)$$

The first term refers to those on the first list only: the second term describes those on both lists and the third term refers to those on the second list only: c denotes the probability of individuals that are listed on both systems and $r = n_{1.} + n_{.1} - n_{11}$. The probability $p_{1.}$ is the probability of those on the first system's list and $p_{.1}$ is the probability of those on the second.

Estimation of N under each model involves only simple statistics computed from the matrix \mathbf{A} . This is in accordance with Otis *et al.* (1978), who recommends that estimates be based on sufficient statistics. Individuals on the lists are not employed in the modeling, but various sums derived from the matrix \mathbf{A} . These are statistics such as $n_{1.}$, $n_{.1}$, n_{11} , n and r . For any model there exists a set of simple statistics called its minimal sufficient statistics, MSS.

Using maximum likelihood method, the general model (1) yields maximum likelihood estimator as follows;

$$\hat{p}_{1.} = \frac{n_{1.}}{\hat{N}}$$

$$c = \frac{n_{11}}{n_{1.}} \quad (2)$$

$$p_{.1} = \frac{n_{1.} - n_{11}}{\hat{N} - n_{1.}}$$

The Maximum Likelihood Estimator (MLE) of N does not exist. This is because there are four parameters ($N, p_{1.}, p_{.1}$ and c) and only three statistics ($n_{1.}, n_{.1}$ and n_{11}).

3.2.3 Model with no Factor Effect (M_0)

M_0 assumes a constant probability of capture for individuals on both systems, that is, these factors; time variation, individual heterogeneity and behavioural response to capture do not affect captures. In other words, all the individuals are equally likely to be on any of the lists. The systems themselves do not affect listing (capture) probabilities. The assumption is such that individuals listed were not influenced by the listing systems or previous listing history or heterogeneity (individual differences).

Hence, $p_{1.} = p_{.1} = c = p$. This implies, the capture and recapture probability of individuals is the same for both systems and for all individuals. The joint probability distribution of the set of possible listing histories $\{n_\omega\}$ is,

$$P(n_\omega) = p(n_{1.}, n_{.1}, n_{11}) = \binom{N}{n_{1.}} \binom{N - n_{1.}}{n_{.1} - n_{11}} \binom{n_{1.}}{n_{11}} p^{n_{1.} + n_{.1}} (1 - p)^{2N - n_{1.} - n_{.1}} \quad (3)$$

The parameters to be estimated here are N and p , while the Minimum Sufficient Statistics (MSS) are $n_{1.}, n_{.1}$ and n_{11} . Using maximum likelihood method,

$$\hat{p} = n / 2\hat{N}_0, \quad \hat{N}_0 = n^2 / 4n_{11}, \quad (4)$$

$$\text{with variance, } \text{Var}(\hat{N}_0) = \frac{\hat{N}_0 q^2}{p^2}, \quad (5)$$

where $n = n_{1.} + n_{.1}$

3.2.4 The Petersen Model M_S

In this case, all individuals in a recording system are assumed to have equal probability of being listed. These probabilities vary from system to system. In

other words, individuals on a particular system, exhibit the same behaviour. This is equivalent to the assumption of the Petersen method, that is, sample independence. Those listed twice (n_{11}) are assumed to have the same behaviour with those in the second system ($n_{.1}$), hence $c = p_{.1}$.

The parameters to be estimated include $p_{1.}, p_{.1}$ and N . The set of MSS are $n_{1.}, n_{.1}$ and n_{11} .

The joint probability distribution is,

$$p(n_{1.}, n_{.1}, n_{11}) = \binom{N}{n_{1.}} p_{1.}^{n_{1.}} (1 - p_{1.})^{N - n_{1.}} \binom{N - n_{1.}}{n_{.1} - n_{11}} \binom{n_{1.}}{n_{11}} p_{.1}^{n_{11}} (1 - p_{.1})^{N - n_{11}} \quad (6)$$

The maximum likelihood estimators for $p_{1.}, p_{.1}$ and N are

$$\begin{aligned} \hat{p}_{1.} &= n_{1.} / \hat{N}_S, \\ \hat{p}_{.1} &= n_{.1} / \hat{N}_S, \\ \hat{N}_S &= \frac{n_{1.} n_{.1}}{n_{11}}, \end{aligned} \quad (7)$$

with variance,

$$Var(\hat{N}_S) = \frac{n_{1.} n_{.1} n_{12} n_{21}}{n_{11}^3} = \hat{N}_S \left(\frac{n_{12} n_{21}}{n_{11}^2} \right) \quad (8)$$

\hat{N}_S is exactly the Petersen estimator which the hypergeometric distribution is assumed. This reveals that, the assumption of independence of the samples is equivalent to assuming that capture probabilities of individuals depends on the systems. Thus $p_{1.}$ and $p_{.1}$ measure the relative effectiveness of the systems.

3.2.5 M_b : Behavioral Response Model (M_b)

Behavioral response can be likened to situation where individuals shy away ('elusify') from a listing system, once listed by that system. This is similar to when individuals shy away or avoid arrest and re-arrest (or even run away from rehabilitation). This model is based on the behaviour of individuals to the capturing systems. Here an individual is assumed to change behavior after rehabilitation or treatment. The joint probability density is

$$p(n_{1.}, n_{.1}, n_{11}) = \binom{N}{n_{1.}} p_{1.}^{n_{1.}} (1 - p_{1.})^{N - n_{1.}} \times$$

$$\binom{n_{1.}}{n_{11}} c^{n_{11}} (1-c)^{n_{1.}-n_{11}} \binom{N-n_{1.}}{n_{1.}-n_{11}} p_{.1}^{n_{1.}-n_{11}} (1-p_{.1})^{N-r} \quad (9)$$

N is not estimable just as in (1). For N to be estimable, we need to make further assumptions. These assumptions, in form of restrictions, yield the following results;

Restricting: $p_{1.}, p_{.1} = c$ will yield M_s ,

$p_{.1}, p_{1.} = c$, yields M_a ,

$c, p_{1.} = p_{.1} = p$ – This is considered here.

If $c > p$, it means more individuals are recaptured by the second system. If $c < p$, less individuals are recaptured. The joint pdf now becomes,

$$p(n_{1.}, n_{1.}, n_{11}) = \binom{N}{n_{1.}} \binom{N-n_{1.}}{n_{1.}-n_{11}} p^{n_{1.}+n_{1.}-n_{11}} \times \\ (1-p)^{2N-2n_{1.}-n_{1.}+n_{11}} \binom{n_{1.}}{n_{11}} c^{n_{11}} (1-c)^{n_{1.}-n_{11}} \quad (10)$$

The MLE yields,

$$\hat{p} = \frac{n_{1.}+n_{1.}-n_{11}}{2N-n_{1.}} = \frac{r}{2N-n_{1.}} \quad (11)$$

$$\hat{c} = n_{11}/n_{1.}$$

$$\hat{N}_b = \frac{n_{1.}^2}{n_{1.}-(n_{1.}-n_{11})} = \frac{n_{1.}^2}{n_{1.}-n_{21}} \quad (12)$$

This is equivalent to the two sample removal estimator given by Seber and Le Cren (1967), in Seber and Whale (1970); the variance is given as;

$$Var(\hat{N}_b) = \frac{N_b q^2 (1+q)}{p^3} \quad (13)$$

3.3 The Proposed Model (M_a)

This model treats situations where the capture probability of individuals on the first system and those recaptured is the same, but the capture probability of individuals on the second system has a different probability. The assumption is that those caught by the first system are likely to be recaptured; that is, to say those captured for the first time are encouraged to show up again. The joint probability distribution with $p_{1.} = c = p$, $p_{.1}$ is;

$$p(n_{1.}, n_{.1}, n_{11}) = \binom{N}{n_{1.}} \binom{n_{1.}}{n_{11}} p^{n_{1.}+n_{11}} (1-p)^{N-n_{11}} \binom{N-n_{1.}}{n_{.1}-n_{11}} p_{.1}^{n_{.1}+n_{11}} (1-p_{.1})^{N-r} \quad (14)$$

where p denotes the probability of individuals that are captured by system 1 and those recaptured by system 2, $p_{.1}$ denotes the capture probability of individuals on records of system 2 only and $r = n_{1.} + n_{.1} - n_{11}$. The maximum likelihood estimators of N_a , p , and $p_{.1}$ are,

$$\hat{p} = \frac{n_{1.}+n_{11}}{\hat{N}+n_{1.}},$$

$$\hat{p}_{.1} = \frac{n_{.1}-n_{11}}{\hat{N}-n_{1.}}, \quad (15)$$

$$\hat{N}_a = \frac{n_{11}(n_{.1}-n_{11})+(n_{1.}+n_{11})n_{1.}}{2n_{11}}$$

and

$$var(\hat{N}_a) = \frac{n_{1.}^4 q_1^2}{4N^3 p^6} + \frac{Np^2 q^2}{4} \quad (16)$$

This model can be suitable for processes where ‘baits’ are used to attract individual animals. In epidemiology and public health surveys, this is similar to situations where patients are encouraged via some inducements e.g. free drugs, treatments, vaccinations, mosquito nets etc.

3.4 Model Selection Criteria

In comparing the existing models and the proposed model, two selection criteria were used, these are; the Akaike Information Criterion (AIC) and the Mean Absolute Deviation (MAD).

3.4.1 Akaike’s Information Criterion (AIC)

The Akaike’s Information Criterion (AIC) used in this work was proposed by Sanni and Jolayemi (2009) and modified by Jibasen (2011) as;

$$AIC = -\beta \sum \sum \log_e \left(\frac{n_{i_0} n_{i_1}}{\hat{N}^2} \right) + 2(c_0 + c_1 - 2) \quad (17)$$

Where, c_0 and c_1 are the dimensions of the contingency table, β is equal to 2 as in the classical, or can be defined as $Abs(N - \hat{N})$ in case of simulation.

3.4.2 Mean Absolute Deviation (MAD)

Mean Absolute Deviation (MAD) was also used to assess the appropriate model alongside the AIC. This is given as

$$MAD = \frac{1}{n} |(N - \hat{N})|, \quad (18)$$

where n is the number of cases under consideration, N is the population size and \hat{N} is the estimated population size.

The results are presented in Table 1.

Table 1: Ten Simulated data sets: $N=90$, $n_{1.} = 50$ and $n_{.1} = 10$

S/No.	n_{11}	\hat{N}_o	AIC	\hat{N}_s	AIC	\hat{N}_a	AIC	\hat{N}_b	AIC
1	7	129	7.141	71	5.366	205	14.96	53	7.50
2	7	129	7.141	71	5.366	205	14.96	53	7.50
3	6	150	9.137	83	4.490	235	18.50	54	7.21
4	8	113	5.760	63	6.112	182	12.44	52	7.89
5	8	113	5.760	63	6.112	182	12.44	52	7.89
6	8	113	5.760	63	6.112	182	12.44	52	7.89
7	10	90	4.000	50	-	150	9.14	50	-
8	7	129	7.141	71	5.366	205	14.96	53	7.50
9	8	113	5.760	63	6.112	182	12.44	52	7.89
10	5	180	12.197	100	4.759	278	23.77	56	6.98
MAD		44		28		138		47	

Source: Jibasen (2011)

4.0 Comparison of Models Using Simulated Data

In this section, we present results of simulated data for different values of N , $n_{1.}$ and $n_{.1}$. The simulated data were used to compute estimated population size and other measures using the four models, namely M_o , M_s , M_a , and M_b . Note that each serial number is a complete set of simulation and estimation. Simulation was done using program R version 2.9.0. The simulation was based on the hypergeometric settings, this is because of the finite number of possibilities in the distribution, where simulated variates approximately satisfied the expected value of the hypergeometric distribution. That is,

$$E(n_{11}) = \frac{n_1 \cdot n_1}{N} \tag{19}$$

Table 1 shows that the Petersen M_s performed better, but breaks down when $n_{11} = n_1$, that is, when those on the list of the second system are proper subset of the first list. M_a and M_o over estimate while M_b underestimates the population size N . This shows that M_a performs very poorly with low recaptures n_{11} .

Table 2: Ten Simulated data sets: $N= 90$, $n_1 = 50$ and $n_1 = 40$

S/No.	n_{11}	\hat{N}_o	AIC	\hat{N}_s	AIC	\hat{N}_a	AIC	\hat{N}_b	AIC
1	27	75	4.868	74	4.927	78	4.698	68	5.38
2	28	72	5.040	71	5.100	76	4.829	66	5.52
3	26	78	4.692	77	4.750	80	4.563	69	5.24
4	25	81	4.509	80	4.567	83	4.423	71	5.10
5	27	75	4.868	74	4.927	78	4.698	68	5.38
6	26	78	4.692	77	4.750	80	4.563	69	5.24
7	25	81	4.509	80	4.567	83	4.423	71	5.10
8	28	72	5.040	71	5.100	76	4.829	66	5.52
9	21	96	4.360	95	4.293	94	4.225	81	4.53
10	24	84	4.316	83	4.375	85	4.276	74	4.96
MAD		15		16		12		25	

Table 2 shows that model M_a is a better model followed by models M_o and M_s . Observe that the recaptures n_{11} are high, all others perform poorly.

Table 3: Ten Simulated data sets: $N= 100$, $n_1 = 60$ and $n_1 = 40$

S/No.	n_{11}	\hat{N}_o	AIC	\hat{N}_s	AIC	\hat{N}_a	AIC	\hat{N}_b	AIC
1	24	104	4.238	100	4.000	113	4.745	82	5.10
2	21	119	5.100	114	4.820	125	5.470	88	4.71
3	23	109	4.497	104	4.248	117	4.965	84	4.97
4	23	109	4.497	104	4.248	117	4.965	84	4.97
5	29	86	4.812	83	5.034	98	4.139	73	5.75
6	27	93	4.427	89	4.647	103	4.181	77	5.48
7	27	93	4.427	89	4.647	103	4.181	77	5.48
8	21	119	5.100	114	4.820	125	5.470	88	4.71
9	18	139	6.325	133	5.971	141	6.462	95	4.30
10	31	81	5.179	77	5.415	93	4.429	71	6.04
MAD		15		13		15		18	

Table 3 shows that the Petersen is better followed by M_o and M_a . This is because of low recaptures.

Table 4: Ten Simulated data sets: $N= 200$, $n_{1.} = 100$ and $n_{.1} = 50$

S/No.	n_{11}	\hat{N}_o	AIC	\hat{N}_s	AIC	\hat{N}_a	AIC	\hat{N}_b	AIC
1	37	152	6.880	135	8.027	192	4.508	115	10.09
2	32	176	5.452	156	6.619	215	4.950	122	9.20
3	31	181	5.116	161	6.313	221	5.304	123	9.04
4	34	165	6.065	147	7.197	205	4.311	119	9.54
5	35	161	6.348	143	7.476	200	4.022	118	9.71
6	39	144	7.383	128	8.593	184	4.982	112	10.52
7	36	156	6.619	139	7.751	196	4.251	116	9.90
8	27	208	4.515	185	4.894	247	7.030	130	8.45
9	34	165	6.065	147	7.197	205	4.311	119	9.54
10	32	176	5.452	156	6.619	215	4.950	122	9.20
MAD	33			50		14		80	

Model M_a proves a better model followed by model M_o .

Table 5: Ten Simulated data sets: $N= 200$, $n_{1.} = 100$ and $n_{.1} = 60$

S/No.	n_{11}	\hat{N}_o	AIC	\hat{N}_s	AIC	\hat{N}_a	AIC	\hat{N}_b	AIC
1	41	156	6.555	146	7.166	181	5.078	123	8.99
2	45	142	7.441	133	8.094	169	5.817	118	9.68
3	39	164	6.080	154	6.692	189	4.659	127	8.68
4	39	164	6.080	154	6.692	189	4.659	127	8.68
5	40	160	6.321	150	6.931	185	4.873	125	8.83
6	33	194	4.355	182	5.057	215	4.899	137	7.81
7	35	183	4.997	171	5.654	205	4.317	133	8.09
8	42	152	6.782	143	7.398	178	5.273	122	9.15
9	39	164	6.080	154	6.692	189	4.659	127	8.68
10	44	145	7.224	136	7.860	172	5.642	119	9.49
MAD		37		48		17		74	

With increase in the number of individuals on the list of the second system $n_{.1}$ model M_a performed well. M_o still followed.

Table 6: Ten Simulated data sets: $N= 200$, $n_{1.} = 100$ and $n_{.1} = 70$

S/No.	n_{11}	\hat{N}_o	AIC	\hat{N}_s	AIC	\hat{N}_a	AIC	\hat{N}_b	AIC
1	52	139	7.648	135	7.980	155	6.574	122	9.19
2	45	161	6.249	156	6.549	174	5.496	133	8.08
3	47	154	6.661	149	6.963	168	5.823	130	8.38
4	47	154	6.661	149	6.963	168	5.823	130	8.38
5	44	164	6.036	159	6.336	177	5.324	135	7.94
6	51	142	7.452	137	7.774	158	6.429	123	9.01
7	42	172	5.586	167	5.893	183	4.962	139	7.65
8	42	172	5.586	167	5.893	183	4.962	139	7.65
9	50	145	7.256	140	7.571	160	6.282	125	8.85
10	56	129	8.458	125	8.850	146	7.136	116	9.94
MAD		47		52		33		71	

Table 6 shows that M_a performs better with higher recaptures and the classical Petersen estimator M_s under estimated N .

Table 7: Ten Simulated data sets: $N= 200$, $n_{1.} = 90$ and $n_{.1} = 70$

S/No.	n_{11}	\hat{N}_o	AIC	\hat{N}_s	AIC	\hat{N}_a	AIC	\hat{N}_b	AIC
1	46	139	7.485	137	7.628	145	7.115	123	8.70
2	43	149	6.887	147	7.026	153	6.661	129	8.22
3	42	152	6.679	150	6.818	155	6.503	131	8.07
4	50	128	8.265	126	8.422	136	7.692	116	9.37
5	51	125	8.463	124	8.626	134	7.834	114	9.55
6	51	125	8.463	124	8.626	134	7.834	114	9.55
7	51	125	8.463	124	8.626	134	7.834	114	9.55
8	53	121	8.872	119	9.049	130	8.120	111	9.94
9	46	139	7.485	137	7.628	145	7.115	123	8.70
10	53	121	8.872	119	9.049	130	8.120	111	9.94
MAD	67			70		60		82	

All the models performed averagely, though M_a is better. This is because the number of individuals on the second system $n_{.1}$ is high, as well as the recaptures n_{11} .

Table 8: Ten Simulated data sets: $N= 300$, $n_{1.} = 150$ and $n_{.1} = 70$

S/No.	n_{11}	\hat{N}_o	AIC	\hat{N}_s	AIC	\hat{N}_a	AIC	\hat{N}_b	AIC
1	40	303	4.156	263	6.279	371	8.704	188	11.39
2	42	288	4.734	250	7.029	357	7.706	184	11.70
3	41	295	4.302	256	6.663	364	8.191	186	11.54
4	44	275	5.528	239	7.717	344	6.814	181	12.02
5	51	237	7.801	206	9.864	305	4.317	172	13.29
6	52	233	8.083	202	10.160	300	4.022	170	13.50
7	49	247	7.213	214	9.271	315	4.949	174	12.90
8	53	228	8.357	198	10.46	296	4.262	169	13.72
9	44	275	5.528	239	7.717	344	6.814	181	12.02
10	53	228	8.357	198	10.459	296	4.262	169	13.72
MAD	40			74		31		122	

Table 8 shows that with increase in recapture n_{11} , all the models performed well, with M_a still being the ‘best’.

5.0 Findings and Conclusion

The capture-recapture models considered in this work are those concerned with two systems, otherwise known as dual systems estimation in epidemiology and demography. M_a was introduced, and it proved better compared to the other three models when there is a good sampling effort, that is, high n_{11} . The classical Petersen method M_s which has been proved by Gaskell and George (1972) and collaborated by Jibasen (2011) to be poor with poorer sampling effort, was also found to be poor with high recaptures n_{11} . Since the expected value of the recaptures (n_{11}) in Petersen model is the same

as the expected value of the hypergeometric distribution, that is, equation (19). The Petersen performs better when recaptures closely satisfy (19) see Table 3 M_b generally under estimates N , where its estimates were approximately equivalent to the number of different individuals listed, r .

The simulation results showed that model M_a is appropriate for a process with a good recaptures, that is, high n_{11} . This model can be applied to animal population where the experiment has good sampling effort. It is particularly useful in public health research where the patients (victims) are encouraged to show up for treatment or rehabilitation.

Reference

- Bishop, Y.M.M, Fienberg, S.E. and Holland, P.W. (1975). Discrete Multivariate Analysis: Theory and Practice. Cambridge: The MIT Press.
- Bonner, S. J. and Schwarz, C.J. (2006). An extension of the Cormack-Jolly-Seber model for continuous covariates with application to *Microtus pennsylvanicus*. *Biometrics* **62**, 142-149
- Chao, A and Tsay, P.K. (1998). A sample coverage approach to multiple-system estimation with application to census undercount. *J. Amer. Statist. Assoc.* **93** 283-293
- Cormack, R. M. (1989). Log-linear model for capture-recapture models. *Biometrics* **48**, 395-413
- Cormack, R. M. (1966). A test of equal catchability. *Biometrics* **28**, 337-343
- Darroch, J. N. (1961). The Two-Sample Capture-recapture Census When Tagging and Sampling are Stratified *Biometrika*.
- Dorazio, R.M. and Royle, J. A. (2003). Mixture models for estimating the size of a closed population when capture rates among individuals. *Biometrics* **59**, 351-364.
- Doscher, M. L. and Woodard, J. A. (1983). Estimating the size of subpopulation of Heroin Users: Application of Log-linear Models to Capture-Recapture Sampling. *Int. J Addict* **18**, 167-182

- Fienberg, S. E. (1972). The Multiple recapture Census for Closed Populations and incomplete 2^k Contingency Tables. *Biometrika* **59**, 591-603
- Gaskell, T. J. and George, B. J. (1972). A Bayesian Modification of the Lincoln Index. *The Journal of Applied Ecology* **9**, 337-384
- Huggins, R. (1991). Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics*. **47**, 725-732.
- International Working Group for Disease Monitoring and Forecasting (IWGDMF) (1995a). Capture-Recapture and Multiple-Record Systems Estimation I: History and Theoretical Development. *Am. J. Epidemiol.* **142**, 1047-58.
- Jibasen, D. (2011). Capture-recapture Type Models for Estimating the Size of An Elusive Population. A PhD Thesis submitted to the University of Ilorin, Ilorin, Nigeria.
- Laska, E. M. and Meisner, N. (1993). A plant-capture method for estimating the size of a population from a single sample. *Biometrics* **49** 209-220
- Lincoln, F. C. (1930). Calculating Waterfowl Abundance on the Basis of Banking Returns. Circular of the U.S. Department of Agriculture, 118, 1-4
- Norris, J. L. and Pollock, K.H. (1996). Nonparametric MLE under two capture-recapture models with heterogeneity. *Biometrics* **52**, 639-649.
- Otis, D.L., Burnham, K.P., White, G.C., and Anderson, D.R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monograph* **62**, 1-135
- Petersen, C. G. J. (1896). The Yearly Immigration of Young Plaice into the Limfjord from the German Sea. Report of the Danish Biological Station to the Ministry of Fisheries. 6, 1-48.
- Plante, N. (1990). Estimation de la taille d'une population animale a l'aide d un model de Capture-recapture avec Stratification. M. Sc. Dissertation, Univ. Laval, Canada.

- Plante, N, Rivest, L. P., Tremblay, G. (1998). Stratified Capture-recapture of the Size of a Closed Population. *Biometrics*, **54** 47-60.
- Pledger, S. (2000). Unified maximum likelihood for closed capture-recapture models using mixtures. *Biometrics* **56**, 434-442.
- Pollock, K.H., Nichols, J.D., Brownie, C. and Hines, J.E. (1990). Statistical inference from capture-recapture experiment. *Wildlife Monographs* **107**, 1-97
- Rajwani, K. and Schwarz, C.J. (1997). Adjusting for Missing tags in Salmon Escapement Surveys. *Canadian Journal of Fisheries and Aquatic Sciences* **54**, 259-264.
- Rockwood, S.V. and Whiting, R.M. (1992). Estimating waterfowl hunter-trips using a capture-recapture technique. *Wildlife Society Bulletin* **20**, 15-20
- Sanni, O.O.M. and Jolayemi, E.T. (2009). On Selection of an Optimal Model using AIC Approach. *Journal of Mathematical Sciences* **20**, 161-170
- Schwarz, C.J. and Taylor, C.G. (1997). The Use of Stratified-Petersen Estimator in Fisheries Management: Estimating the number of Pink Salmon (*Oncorhynchus gorbuscha*) Spawners in the Fraser River. *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 281-297.
- Schwarz, C.J. and Seber, G.A.F. (1999). Estimating animal abundance: *Statistical Science* **14**, 427-456.
- Seber, G. A. F. and Le Cren, E. D. (1967). Estimating Population Parameters from Catches Large Relative to the Population. *J. Animal Ecol.* **36**, 631-643
- Seber, G.A.F. and Whale, J.F. (1970). The Removal Method for Two and Three Samples. *Biometrics* **26**, 393-400
- Seber, G.A.F. and Felton, R. (1981). Tag loss and the Petersen Mark-recapture Experiment *Biometrika* **48**, 211-219
- Seber, G.A.F. (1982a). Capture-recapture methods: Encyclopedia of Statistical Sciences Vol.1 367-374. Edited by Samuel, K. and Norman, L.J. New York: John Wiley and sons.

-
- Seber, G.A.F. (1982b). *Estimating Animal Abundance and related Parameters*, 2nd edition. London: Charles Griffin & Sons.
- Seber, G.A.F, Huakau J.T and Simmons D. (2000). Capture-Recapture, Epidemiology, and List Mismatches: Two Lists. *Biometrics* **56**, 1227-1232
- Skalski, J. R. and Robson, D. S. (1982). A Mark and Removal Field Procedure for Estimating Animal Abundance. *J. Wildlife Manag.* **46**:741-751.
- Wolter, K. M. (1990). Capture-Recapture Estimation in the Presence of a Known Sex Ratio. *Biometrics* **46**, 157-162
- Xu, Y., Liu L., You, N., Pan, H. and Yip, P. (2007) Estimating population size for a continuous time frailty model with covariates in a capture-recapture study. *Biometrics*. **63**. 917-921.
- Yip, P., Huggins, R.M. and Lin, D.Y. (1996). Inference for capture-recapture experiments in continuous time with variable capture rates. *Biometrika* **83**, 477-483.