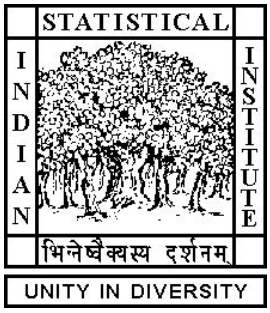
A New Capture-Recapture Model in Dual-record System

Technical Report No. ASU/2017/8
Dated: 26 May, 2017

Kiranmoy Chatterjee
Interdisciplinary Statistical Research Unit,
Indian Statistical Institute,
Kolkata 700108
kiranmoy07@gmail.com

Prajamitra Bhuyan
Applied Statistics Unit,
Indian Statistical Institute,
Kolkata 700108
bhuyan.prajamitra@gmail.com



A New Capture-Recapture Model in Dual-record System

Kiranmoy Chatterjee*
Prajamitra Bhuyan†

Abstract

Population size estimation based on two sample capture-recapture type experiment is an interesting problem in various fields including epidemiology, pubic health, ecology, population studies, etc. The Lincoln-Petersen estimate is popularly used under the assumption that capture and recapture status of each individual is independent. However, in many real life scenarios, there is some inherent dependency between capture and recapture attempts which is not well-studied in the literature for dual system or two sample capture-recapture method. In this article, we propose a novel model that successfully incorporates the possible causal dependency and provide corresponding estimation methodologies for the associated model parameters. Simulation results show superiority of the performance of the proposed method over existing competitors. The method is illustrated through analysis of some real data sets.

Key words: Behavioural dependency, Bivariate Bernoulli, Disease surveillance, Method of moments, Maximum likelihood, Post-stratification.

*Indian Statistical Institute, Kolkata, India & Bidhannagar College, Kolkata, India; E-mail: *kiranmoy07@gmail.com*

†Indian Statistical Institute, Kolkata, India.

1 Introduction

Estimation of the true size of a population is an interesting problem in different disciplines of epidemiological, medical, social and demographic study. In order to formulate policies for public heath related issues, federal agencies are generally interested to know the actual size of a diseased population (e.g. Encephalitis patients) or vital events (e.g. child mortality) occurring in a specified region. Any attempt to count all the individuals belonging to a population of interest is always subject to error and the degree of error depends on many factors, such as, population size, individual's capture probability, etc. Thus, it is common practice to gather information (lists of names and other identities) from two independent attempts. In order to draw inference, one needs to combine the data obtained from these two independent surveys and determine how many people are included in both the lists and how many are included only in one of the lists. This 2×2 cross-classified data structure is well known as dual system or dual-record system (DRS) (Wolter, 1986; Chatterjee and Mukherjee, 2016b). In DRS, counts for the three cells are available, however the last cell count remained unknown which makes the true population size N unknown. Since, the fourth cell is missing, this DRS often called an incomplete 2×2 data. The primary goal is to estimate the missing cell count, equivalently N, from the available data. This is somewhat close to the capture-recapture experiment, widely practiced in wild-life studies, with only one recapture attempt. In general, the entire population can be post-stratified into two mutually exclusive and exhaustive sub-populations based on demographic characteristics (e.g. age, sex, etc.), and it is also of great interest to estimate sub-population sizes (Bell, 1993; Wolter, 1990).

It is a common practice to assume causal independence between capture and recapture attempts, which helps to reduce the dimension of the unknown parameters associated with the model, and hence the model becomes estimable. The resulting estimate is popularly known as the Lincoln-Petersen estimate (Otis et al., 1978; Bohning and Heijden, 2009) in the broad literature of capture-recapture theory. Chatterjee and Mukherjee (2016b) discussed details of such model, known as M_t model (Otis et al., 1978), along with various likelihood-based estimates of N in DRS. However, M_t model often fails due to positive dependence among the lists, especially in the fields of public health and demography, which leads to underestimation of N (Hook and Regal, 1982; Chao et al., 2001). For

example, patients with positive result from a serum test for Hepatitis A Virus (HAV), are prone to visit hospital for further treatment. Therefore, the ascertainment of the serum sample and that of the hospital sample becomes dependent. In census-undercount study, Fay et al. (1988) and Bell (1993) observed such behavioral dependence among adult males but not for females in the Post Enumeration Programs conducted for evaluating the US Censuses in 1980 and 1990 respectively. In epidemiological or demographic surveillance study, usually a population possesses positive nature of dependence between the two lists which is known as *recapture proneness*. Similarly, there are some populations in which individuals possesses negative nature of dependency, known as *recapture aversion*, such as the population of drug abused, population of patients affected with HIV or any other diseases that bear social stigma.

Modeling of the capture-recapture data without the causal independence assumption is an important but challenging task in DRS. Nour (1982) proposed an estimate of total number of vital records assuming such positive dependence between DRS of vital registration systems. Wolter (1990) provided estimation for post-strata wise population sizes under two different models, assuming sex-ratio is known. In the first model, Wolter (1990) considered that the cross-product ratios in DRSs for male and female post-strata are same but unknown and, in the second one, causal independence is assumed for the female only. Isaki and Schultz (1986) also worked on the same problem for 1980 Post Enumeration Program and suggested an estimate based on demographic analysis. Later, Bell (1993) proposed some variations of the methods suggested by Wolter (1990) for the estimation of the cross-product ratios for both male and female populations. However, sex-ratio is calculated at the time of census for larger population (e.g. national level population). In many situations, it is not realistic to assume that the sex-ratio remains constant over time or holds true for the sub-population under consideration. Moreover, the availability of sex-ratio is very much limited across the various fields where the DRS type data structure is commonly used (e.g., epidemiological or disease surveillance data; *See* Section 6).

In this article, we propose a novel model to incorporate this inherent dependency between capture and recapture attempts in DRS and provide estimation methodologies for the population size N under two different modeling scenarios. Our work is motivated from two real datasets on public health: (i) *Encephalitis incidence* in England, 2006-2007 and (ii) *child mortality* in western Kenya, 2000-2001, where the existing methods proposed by

Wolter (1990) and Nour (1982) are not applicable. Our models and associated estimates exhibit superiority with respect to ease of interpretation and relative root mean squared error (RRMSE) over the existing competitors available in the literature (*See* Section 4, 5). We first describe the DRS and the associated data structure in Section 2. In Section 3, we propose a Bivariate Bernoulli model under the DRS. Next, in Section 4, we derive method of moments estimates (MMEs) and discuss maximum likelihood estimation of the model parameters. Comparison of the proposed estimators with its existing competitors is studied through simulation and two illustrative data analyses in Sections 5 and 6, respectively. Finally, we end with some concluding remarks in Section 7.

2 Dual-record System (DRS)

The idea of DRS is similar to the capture-recapture sampling in wildlife management for estimation of the population size N. Laplace (1783) pioneered such sampling plans in order to estimate the population size of France from vital events like births, marriages and deaths. Let us consider a population U of size N. It is not feasible to capture all the N individuals by means of a population census. In order to estimate N, at least two such independent attempts are organized, providing two corresponding lists. The individuals captured in the first list (e.g. census) are matched one-by-one with the list of individuals captured from the second survey, also known as Post Enumeration Survey (PES). Let $p_{j1\cdot}$ and $p_{j\cdot 1}$ denote the capture probabilities of the jth individual in the first sample (List 1) and the second sample (List 2), respectively. Then estimate of N is obtained under different assumptions on the capture probabilities in both the lists. In this article, we consider the DRS with the following assumptions:

($S1$) Population is closed until the second sample is taken,

($S2$) Individuals are homogeneous with respect to their capture probabilities.

Assumption ($S2$) ensures that $p_{j1\cdot} = p_{1\cdot}$ in List 1 and $p_{j\cdot 1} = p_{\cdot 1}$ in List 2 for $j = 1, 2, \cdots, N$. The data structure is presented in Table 1, popularly known as the Dual-record system or shortly, DRS. The number of untapped individuals in both the surveys, denoted

as x_{00}, is unknown which makes the total population size N unknown. The probabilities attached to all the cells are also provided in Table 1 and these notation will be followed throughout this paper. As discussed before, casual independence among individuals is assumed between capture and recapture attempts, which is formally written as:

($S3$) Inclusion of each and every individual, belonging to U, in List 2 is *causally independent* to its inclusion in List 1 (i.e. $p_{11} = p_{1\cdot}p_{\cdot 1}$).

Now assuming ($S3$), estimate of N is given as $\hat{N}_{ind} = [x_{\cdot 1}x_{1\cdot}/x_{11}]$, which is popularly

Table 1: Dual-record-System (DRS): 2×2 data structure with cell probabilities mentioned in [ ] and $p_{\cdot\cdot}$=1

	List 2		
List 1	In	out	Total
In	$x_{11}[p_{11}]$	$x_{10}[p_{10}]$	$x_{1\cdot}[p_{1\cdot}]$
Out	$x_{01}[p_{01}]$	$x_{00}[p_{00}]$	$x_{0\cdot}[p_{0\cdot}]$
Total	$x_{\cdot 1}[p_{\cdot 1}]$	$x_{\cdot 0}[p_{\cdot 0}]$	$x_{\cdot\cdot} = N[p_{\cdot\cdot}]$

known as the Lincoln-Petersen estimate and widely used in different contexts for homogeneous human population (Bohning and Heijden, 2009). Also, this estimate is identical with the resulting estimate from M_t model (Wolter, 1986). Several authors criticized the causal independence assumption ($S3$) in the context of surveys and censuses for human populations (ChandraSekar and Deming, 1949). In many situations, the failure in capturing one individual in both the attempts may be due to some common causes. In some other cases, individuals may be less keen to be enlisted in List 2. This phenomena is grossly known as behavioral response variation (*See* Wolter, 1986).

3 Proposed Model

In this section, we first introduce a Bivariate Bernoulli model (BBM), which is useful in measuring the degree of association between two dichotomized quantitative characters. Although the problem can be generalized to a multivariate setup, in the present paper we focus our attention to the bivariate version only. This model will be used to incorporate

the inherent dependency between capture and recapture attempts in the DRS. Now, we will introduce the Bivariate Bernoulli model for Dual-record System (BBM-DRS).

Let us define a paired variable (Y, Z) such that Y_i and Z_i, denote, respectively, the List 1 and List 2 inclusion status of the i^{th} individual belonging to U. Suppose assume (Y_i, Z_i), for $i = 1, \ldots, N$, are *iid* bivariate random variables distributed as

$$(Y_i, Z_i) \sim \begin{cases} (X_1, X_2) & \text{with prob. } 1-\alpha, \\ (X_1, X_1) & \text{with prob. } \alpha, \end{cases} \tag{1}$$

where X_1 and X_2 are independently distributed Bernoulli random variables with parameters p_1 and p_2, respectively. We write $p_{yz} = Prob(Y = y, Z = z)$, for $y, z = \{0, 1\}$. Thus, based on the parameters involved in the above model (1), we have the following cell probabilities in the DRS (*See* Table 1):

$$\begin{aligned} p_{11} &= \alpha p_1 + (1-\alpha) p_1 p_2, \\ p_{10} &= (1-\alpha) p_1 (1-p_2), \\ p_{01} &= (1-\alpha)(1-p_1) p_2, \\ p_{00} &= \alpha(1-p_1) + (1-\alpha)(1-p_1)(1-p_2). \end{aligned}$$

Consequently, the marginal probabilities are

$$\begin{aligned} p_Y = p_{1\cdot} &= p_1, \text{ and} \\ p_Z = p_{\cdot 1} &= \alpha p_1 + (1-\alpha) p_2, \end{aligned}$$

with $Cov(Y, Z) = \alpha p_1 (1 - p_1)$. Note that the proposed Bivariate Bernoulli model incorporates positive dependence between capture status in Lists 1 and 2. In particular, when $\alpha = 0$ (i.e. there is no case of causal dependency), our proposed Bivariate Bernoulli model in (1) reduces to the M_t model.

***Remark* 1** *One can define the proposed BBM-DRS model in order to capture negative dependency (or, recapture aversion) by rewriting (1) as*

$$(Y_i, Z_i) \sim \begin{cases} (X_1, X_2) & \textit{with prob. } 1-\alpha, \\ (X_1, 1-X_1) & \textit{with prob. } \alpha. \end{cases}$$

***Remark* 2** *The parameters of BBM-DRS possess easy interpretations with practical significance. The dependence parameter α represents proportion of causally dependent individuals, and p_i is the capture probability of an causally independent individual in the ith List, for $i = 1, 2$.*

4 Estimation Methodologies

Let us assume that the population U of size N can be divided into two mutually exclusive and exhaustive sub-populations U_A and U_B with size N_A and N_B, respectively. In practice, one can easily consider post-stratification of the entire population into two mutually exclusive and exhaustive sub-populations (*See* Wolter, 1990; Eisele et al., 2003; Granerod et al., 2013). We also assume that for any individual, belonging to U_A, the capture status in either of the two lists is independent of the same of an individual belonging to U_B. In order to denote the cell counts and the associated probabilities for the 2×2 table obtained under the DRS for the sub-population U_k, we consider the same notation as mentioned in Table 1, with an additional suffix k (for example, List 1 capture probability for the sub-population U_k is denoted as $p_{1\cdot k}$), for $k = A, B$. Now we consider two different models and propose methodologies for estimation of the associated parameters including the population size $N(= N_A + N_B)$, the parameter of primary interest.

4.1 Model I

In this model, we consider the assumption $(S3)$ for the sub-population U_B, which implies $p_{11B} = p_{1\cdot B}p_{\cdot 1B}$. Therefore, the popular Lincoln-Petersen estimate of N_B is given as $\hat{N}_B = \left[\frac{x_{1\cdot B}x_{\cdot 1B}}{x_{11B}}\right]$. In order to incorporate the behavioural dependency present in the sub-population U_A, we consider BBM-DRS model as described in subsection 3, which consists of four parameters with $p_1 = p_{1A}$, $p_2 = p_{2A}$,$\alpha = \alpha_A$, and $N = N_A$. In addition to $(S3)$, we consider the following assumption:

$(S4)$ Initial (List 1) capture probabilities for the individuals belonging to both the sub-populations U_A and U_B are the same (i.e. $p_{1\cdot A} = p_{1\cdot B} = p_1, say$).

The assumption $(S4)$ ensures identifiability of the model parameters. Note that List 1 is prepared before List 2 and hence, List 2 capture probabilities for different sub-populations may differ due behavioral dependence, if exists. Also, it is quite reasonable to consider the same List 1 capture probability for different sub-populations when possibly there is no prejudice. Similar assumption has been considered by several authors in the past (Bell, 1993). Note that the model under consideration is similar to the Model 2 proposed by Wolter (1990), where the estimate of N_A is obtained from $\hat{N}_B$ using the available knowledge of the sex-ratio. As discussed before, the availability of reliable estimate of the sex-ratio remains a practical challenge (*See Section 6*). As mentioned before, N_B is estimated assuming causal independence, and hence, one needs to find the estimate of N_A in order to estimate the population size N. Since α_A can be interpreted as the proportion of behaviorally dependent individuals, its estimation may provide interesting insight of the capture-recapture mechanism.

First we consider method of moments estimation of the parameters associated with Model I. Note that the MME of N_B is same as the Lincoln-Petersen estimate $\hat{N}_B = \left[\frac{x_{1\cdot B}x_{\cdot 1B}}{x_{11B}}\right]$, and the MMEs of p_{1B} and p_{2B} are given as $\hat{p}_{1B} = \frac{x_{11B}}{x_{\cdot 1B}}$ and $\hat{p}_{2B} = \frac{x_{11B}}{x_{1\cdot B}}$, respectively. From the assumption $(S4)$, the estimate of p_{1A} is given as $\hat{p}_1 = \frac{x_{11B}}{x_{\cdot 1B}}$. Now, equating expected and observed number of cell counts from the 2×2 table obtained under the DRS for the sub-population U_A, we get

$$\begin{aligned} N_A p_{11A} &= x_{11A}, \\ N_A p_{10A} &= x_{10A} \text{ and} \\ N_A p_{01A} &= x_{01A}, \end{aligned} \tag{2}$$

involving three unknown parameters N_A, p_{2A}, and α_A. Solving these equations in (2), the MMEs of the model parameters are obtained as

$$\begin{aligned} \hat{N}_A &= \left[\frac{x_{1\cdot A}x_{\cdot 1B}}{x_{11B}}\right], \\ \hat{p}_{2A} &= \frac{x_{01A}x_{11B}}{x_{10A}x_{01B} + x_{01A}x_{11B}}, \\ \hat{\alpha}_A &= \min\left\{\max\left\{0, \frac{x_{\cdot 1A}}{x_{1\cdot A}} - \frac{x_{01A}x_{\cdot 1B}}{x_{01B}x_{1\cdot A}}\right\}, 1\right\}. \end{aligned}$$

The detailed derivation for finding the above mentioned MMEs and some asymptotic

results of the estimator $\hat{N}_A$ are provided in the *Appendix*.

Next, we obtain the maximum likelihood estimate (MLE) of $\theta = (N_A, N_B, \alpha_A, p_1, p_{2A}, p_{2B})$ based on the available data from the DRS under the assumptions of Model I. The likelihood function of θ is given by

$$\begin{aligned} L(\theta|\underline{\mathbf{x}}_A, \underline{\mathbf{x}}_B) \;\propto\; & \frac{N_A! N_B!}{(N_A - x_{0A})!(N_B - x_{0B})!}[\alpha_A p_1 + (1-\alpha_A)p_1 p_{2A}]^{x_{11A}} \\ & \times p_1^{(x_{10A}+x_{11B}+x_{10B})}(1-p_1)^{(x_{01A}+N_B-x_{11B}-x_{10B})} p_{2A}^{x_{01A}} \\ & \times p_{2B}^{(x_{11B}+x_{01B})}(1-p_{2A})^{x_{10A}}(1-p_{2B})^{(N_B-x_{11B}-x_{01B})}(1-\alpha_A)^{(x_{10A}+x_{01A})} \\ & \times [\alpha_A(1-p_1) + (1-\alpha_A)(1-p_1)(1-p_{2A})]^{(N_A-x_{0A})}, \end{aligned} \tag{3}$$

where $\underline{\mathbf{x}}_k = (x_{11k}, x_{10k}, x_{01k})$, $x_{0k} = x_{11k} + x_{10k} + x_{01k}$, for $k = A, B$. However, explicit solution for MLE of θ is not possible. The Newton-Raphson method can be used to maximize the log-likelihood in order to estimate θ, assuming N_A and N_B as continuous parameters. Alternatively, any standard software package equipped with general purpose optimization (e.g., *optim* in the package R) can be used. Note that the log-likelihood function involves $\ln(N_A!)$, which may create computational difficulty for large values of N_A. In order to avoid such issues we approximate $\ln(N_A!)$ as $N_A \ln(N_A) - N_A + \frac{1}{2}\ln(2\pi N_A)$ (Wells, 1986, p. 45).

***Remark* 3** *The above likelihood function (3) can be simplified using Stirling's approximation of* $\ln(N_A!) \approx N_A \ln(N_A) - N_A$ *(Whittaker and Robinson, 1967, p. 138-140), and obtain closed form expression of the MLEs. Interestingly, the MLEs for all the parameters are exactly equal to the respective MMEs.*

***Remark* 4** *If the ratio of the sub-population sizes (equivalent to sex-ratio for male-female stratification)* r *is known, one can easily incorporate such information in the likelihood function (3) taking* $N_B = r^{-1}N_A$.

4.2 Model II

In Model II, we relax the assumption $(S3)$ and the BBM-DRS model is considered for both the sub-populations U_A and U_B with parameters $p_1 = p_{1k}$, $p_2 = p_{2k}$, $\alpha = \alpha_k$, and $N = N_k$, for $k = A, B$. Similar to Model I, we consider the assumption $(S4)$ (i.e. $p_{1A} = p_{1B} = p_1$, say) and additionally we assume $\alpha_A = \alpha_B = \alpha_0$, say, which ensures

identifiability of Model II. Note that this model is similar to the Model 1 proposed by Wolter (1990). Note that the estimator of N in Wolter (1990) is given by $\hat{N}_B = \frac{Kx_{0B}-x_{0A}}{K-r}$, where $K = \frac{x_{11B}(x_{1\cdot A}-x_{11A})(x_{\cdot 1A}-x_{11A})}{x_{11A}(x_{1\cdot B}-x_{11B})(x_{\cdot 1B}-x_{11B})}$ and r denotes the available sex-ratio, which becomes infeasible when $K \leq r$.

We first consider the method of moments for estimating the parameters associated with the Model II. We equate the expected and observed cell counts from the 2×2 tables obtained under the DRS involving six parameters $N_A, N_B, p_1, p_{2A}, p_{2B}, \alpha_0$ and find the following MMEs as

$$\begin{aligned}
\hat{p}_{2A} &= \frac{x_{01B}(x_{1\cdot A}x_{10B} - x_{1\cdot B}x_{10A})}{x_{1\cdot B}(x_{01A}x_{10B} - x_{10A}x_{01B})}, \\
\hat{p}_{2B} &= \frac{x_{01A}(x_{1\cdot A}x_{10B} - x_{1\cdot B}x_{10A})}{x_{1\cdot A}(x_{01A}x_{10B} - x_{10A}x_{01B})}, \\
\hat{\alpha_0} &= 1 - \frac{x_{10A}}{x_{1\cdot A}}\frac{1}{1-\hat{p}_{2A}}, \\
\hat{p}_1 &= \frac{1}{1 + \frac{x_{01A}}{x_{10A}}\left(\frac{1}{\hat{p}_{2A}} - 1\right)}, \\
\hat{N}_A &= \frac{x_{1\cdot A}}{\hat{p}_1}, \\
\hat{N}_B &= \frac{x_{1\cdot B}}{\hat{p}_1}.
\end{aligned}$$

The derivation for finding the above mentioned MMEs is similar to that of Model I and hence skipped. In some cases $\hat{p}_{2A} > \frac{x_{01A}}{x_{01A}-x_{10A}}$, and hence, the estimates for p_1, N_A and N_B become negative, as in Wolter (1990). Such issues with MME has been discussed in the literature (*See* Bowman and Shenton (1998, p. 2092-2098) for more details). Therefore, it is not advisable to use MME and we propose to estimate N_A and N_B using the maximum likelihood method.

The likelihood function of $\theta = (N_A, N_B, \alpha_0, p_1, p_{2A}, p_{2B})$, under the assumptions of

Model II, based on data obtained from the DRS is given by

$$\begin{aligned} L(\theta|\underline{\mathbf{x}}_A, \underline{\mathbf{x}}_B) \;\propto\; & \frac{N_A! N_B!}{(N_A - x_{0A})!(N_B - x_{0B})!}[\alpha_0 p_1 + (1-\alpha_0) p_1 p_{2A}]^{x_{11A}} \\ & \times [\alpha_0 p_1 + (1-\alpha_0) p_1 p_{2B}]^{x_{11B}} p_1^{(x_{10A}+x_{10B})} (1-p_1)^{(x_{01A}+x_{01B})} \\ & \times p_{2A}^{x_{01A}} p_{2B}^{x_{01B}} (1-p_{2A})^{x_{10A}} (1-p_{2B})^{x_{10B}} (1-\alpha_0)^{(x_{10A}+x_{01A}+x_{10B}+x_{01B})} \\ & \times [\alpha_0(1-p_1) + (1-\alpha_0)(1-p_1)(1-p_{2A})]^{(N_A - x_{0A})} \\ & \times [\alpha_0(1-p_1) + (1-\alpha_0)(1-p_1)(1-p_{2B})]^{(N_B - x_{0B})}, \end{aligned} \tag{4}$$

where $\underline{\mathbf{x}}_k = (x_{11k}, x_{10k}, x_{01k})$, $x_{0k} = x_{11k} + x_{10k} + x_{01k}$, for $k = A, B$. As discussed before, explicit solution for MLE of θ is not possible, and hence, any standard software package equipped with general optimization (e.g., *optim* in the package R) can be used. Here also, for computation of the log-likelihood function, we consider the Stirling's approximation for $\ln(N_k!)$, for $k = A, B$. As remarked in Subsection 4.1, one can consider the same reparametrization $N_B = r^{-1} N_A$ in the likelihood function (4), if the ratio of the sub-population sizes r is known.

5 Simulation Study

In this section, the properties of the proposed estimators are investigated through simulation and compared with the works of Nour (1982), and Wolter (1990). First, we consider Model I and generate data for the sub-population U_B under the M_t model with six choices of capture probabilities $(p_{1\cdot B}, p_{\cdot 1B})$ as given by $(0.60, 0.80)$, $(0.60, 0.70)$, $(0.80, 0.55)$, $(0.80, 0.70)$, $(0.50, 0.75)$ and $(0.50, 0.60)$, denoted $P1, \ldots, P6$, respectively, with $N_B = 200$ and 1000. The sub-population U_A is also generated with size 240 and 1200, keeping the same six pairs of capture probabilities for $(p_{1\cdot A}, p_{\cdot 1A})$, with $\alpha_A = 0.4$ and 0.8. Since the Lincoln-Petersen estimator of N_B produces efficient results under the M_t model, our primary interest is to estimate N_A and α_A only. The mean, RRMSE, and 95% confidence interval of the estimates are obtained based on 5000 replication and presented in Table 2.

From Table 2, it is observed that both the proposed estimators (MME and MLE) of N_A performs far better in terms of RRMSE than other existing estimator proposed by Nour (1982). The RRMSE of the MME appears to be better compared to that of the

MLE. As discussed before, the estimator proposed by Wolter (1990) is not applicable for the cases where sex-ratio is not available. Assuming the sex-ratio of the simulation model to be known, the method of Wolter (1990) (Model 2) has been used along with the proposed estimators for a comparison study. Our method seems to perform better compared to that of Wolter (1990) in terms of RRMSE (not reported here).

Next, we generate data from Model II considering the same parameter choices $(P_1, \ldots, P_6)$ for the capture probabilities $(p_{1\cdot k}, p_{\cdot 1k})$ for $k = A, B$, with common dependence parameter $\alpha_0 = 0.4$ and 0.8, and the same values of N_A and N_B considered earlier. The results are presented in Table 3. As discussed before, the proposed MME and the estimator proposed by Wolter (1990) (Model 1) are often found to be negative; hence, these estimators have not been considered for this simulation study. It is clear from the results presented in Table 3 that the performance of the proposed MLE is significantly better than that of Nour (1982). As expected, the RRMSE of the MLE decreases as the population size N increases.

6 Applications

In this section, we first analyze a data set on Encephalitis (infectious and noninfectious) incidence in England during November 2006 to October 2007 (Granerod et al., 2013), presented in the top panel of Table 4. This particular data was collected adhering to an encephalitis code in any of the 20 diagnostic fields, and segregated into two strata, Children (< 18 years) and Adult (≥ 18 years). A patient detected with encephalitis by a hospital clinician was likely to be recorded in Hospital Episode Statistics (HES) and also included in the the Public Health England (PHE) study. Thus, these two sources are likely to be positively dependent (Granerod et al., 2013, p. 1461), and the M_t (or equivalently, Lincoln-Petersen) model possibly underestimated the true number of cases. Note that the estimator proposed by Nour (1982) cannot be applied for both the strata as its underlying condition ($x_{11}^2 > x_{10}x_{01}$) is not satisfied. Also, the estimators proposed by Wolter (1990) can not be applied as the ratio of adult and child patients (equivalent to sex-ratio for male-female stratification) is not available here.

As remarked in Subsection 4.1, the MMEs are approximately equal to the MLEs under Model I, and hence we only consider maximum likelihood estimation for our data

analysis. For analyzing the data under Model I, we consider both the cases separately where capture recapture status for Children and Adult are independent. Comparing the relative standard error (r.s.e) values, we find that our proposed estimator under Model I performs better with independent assumption for Children than that for Adult and the corresponding results are reported in the top panel of Table 5. Estimate of the dependence parameter indicates that 5% of the Adult encephalitis patients are causally dependent. Under Model II, the estimated number of patients is larger compared to that of under Model I. The estimated proportion of causally dependent patients for both Adult and Children are 3% under Model II. It is interesting to note that the relative standard error (r.s.e.), based on 1000 bootstarp resamples, of the MLE under Model II (Model I) is substantially smaller compared to those of the MLE under Model I (M_t Model).

Now we consider another dual system dataset from Wagai and Yala divisions in western Kenya on child mortality, named as *Gem* in the article by Eisele et al. (2003), presented in the bottom panel of Table 4. This study is on the completeness and differential ascertainment of vital events related to child health among male and female children (less than five years old) registered in demographic surveillance system (DSS) based on two-sample capture-recapture experiment. Here also, both the methods, proposed by Wolter (1990) and Nour (1982), are not applicable. Analyzing the data, we find that performance of our proposed estimator $\hat{N}_k$s under Model I, for k = Male, Female, performs better with the assumption that capture recapture status for Female are independent. The results are presented in the bottom panel of Table 5. It is seen that the estimates for female deaths based on Model I and M_t model are very close, however the r.s.e is for Model I is smaller compared to that of M_t model. Estimate of the dependence parameter indicates that 7% of male child are causally dependent under Model I. Under Model II, the MLEs are marginally lower compared to those of under Model I. In this case the relative standard error (r.s.e.) of the estimates under Model I(Model II) is smaller compared to those under Model II (M_t model).

7 Concluding Remarks

This article deals with a very interesting problem when causal independence assumption in DRS, which is common in practice in the fields of public health and demography, is

not valid. We introduce a model, called Bivariate Bernoulli model, that successfully accounted for the possible dependence between capture and recapture attempts. Though the proposed model discusses positive correlation or recapture proneness, one can rewrite the model easily in order to model negative dependence equivalently, recapture aversion (See Ramark 1). Our proposed model seems to have an edge in terms of ease of interpretation and has much wider domain of applicability. In case, the ratio of the subpopulation sizes (e.g., sex-ratio for male-female stratification) is also known, estimates based on our proposed models incorporating the behavioural aspect of the capture-recapture mechanism may be preferred. This also allows analysis of any additional information like, for example, the sex-ratio to make more efficient inference. Although the primary objective is to obtain estimate of the population size N, the estimates of the other model parameters, especially $\hat{\alpha}$, give specific insights into the capture-recapture mechanism. The BBM can also be extended for multiple list or multiple capture-recapture problems which is commonly encountered in the study of wildlife population. It is also an interesting problem to develop testing procedure in order to test the significance of bahavioural dependence between two sources in DRS, which will be taken up in future work.

Appendix

Derivation for MME under Model I:

We get from (2), the following equation in terms of p_{2A}, α_A, and N_A:

$$\begin{aligned} N_A\alpha_A\hat{p}_1 + (1-\alpha_A)\, N_A\hat{p}_1 p_{2A} &= x_{11A}, && (5)\\ N_A\hat{p}_1\,(1-p_{2A})\,(1-\alpha_A) &= x_{10A} && (6)\\ N_A p_{2A}\,(1-\hat{p}_1)\,(1-\alpha_A) &= x_{01A}, && (7) \end{aligned}$$

where $\hat{p}_1 = \hat{p}_{1\cdot A} = \frac{x_{11B}}{x_{\cdot 1B}}$. Now, by adding (5) and (6), we get the MME of N_A as

$$\hat{N}_A = \frac{x_{1\cdot A}x_{\cdot 1B}}{x_{11B}}.$$

Again, by adding the equations (5)-(7), we have

$$N_A\hat{p}_1 + N_A p_{2A}\,(1-\alpha_A)\,(1-\hat{p}_1) = x_{0A} \qquad (8)$$

and by subtracting (7) from (6), we get

$$N_A\,(1-\alpha_A)\,(\hat{p}_1 - p_{2A}) \;=\; (x_{10A} - x_{01A})\,.$$

Now, using the estimates $\hat{N}_A$ and $\hat{p}_1$ in (8), we get

$$p_{2A}\,(1-\alpha_A) \;=\; \frac{x_{01A}x_{11B}}{x_{01B}x_{1\cdot A}}. \tag{9}$$

Since $N_A\hat{p}_1 = x_{1\cdot A}$, (5) implies

$$\alpha_A + p_{2A}\,(1-\alpha_A) \;=\; \frac{x_{11A}}{x_{1\cdot A}}. \tag{10}$$

Subtracting (9) from (10), the MME of α_A is obtained as

$$\hat{\alpha}_A \;=\; \frac{x_{\cdot 1A}}{x_{1\cdot A}} - \frac{x_{01A}x_{\cdot 1B}}{x_{01B}x_{1\cdot A}}. \tag{11}$$

Using $\hat{\alpha}$ in (9), MME of p_{2A} is given as

$$\hat{p}_{2A} \;=\; \frac{x_{01A}x_{11B}}{x_{10A}x_{01B} + x_{01A}x_{11B}}.$$

In order to ensure that MME of α_A lies in $[0,1]$, we modify (11) and consider

$$\hat{\alpha}_A = \min\left\{\max\left\{0, \frac{x_{\cdot 1A}}{x_{1\cdot A}} - \frac{x_{01A}x_{\cdot 1B}}{x_{01B}x_{1\cdot A}}\right\}, 1\right\}.$$

□

Theorem 1 *For large N_A, the mean and variance of the estimator $\hat{N}_A$ are given by*

$$(i)E(\hat{N}_A) \approx N_A + r\frac{p_{01B}}{p_1 p_{\cdot 1B}^2}, \qquad (ii)V(\hat{N}_A) \approx N_A(1-p_1) + r\frac{p_{01B}(1+p_1)}{p_1^2 p_{\cdot 1B}^2},$$

respectively, where $r = N_A/N_B$.

Proof: Since capture status of individual belongs to U_A is completely independent to that of U_B, we have

$$E(\hat{N}_A) \;=\; E(x_{1\cdot A})E\left(\frac{x_{\cdot 1B}}{x_{11B}}\right), \tag{12}$$

$$V(\hat{N}_A) \;=\; E(x_{1\cdot A})V\left(\frac{x_{\cdot 1B}}{x_{11B}}\right) + V(x_{1\cdot A})E\left(\frac{x_{\cdot 1B}}{x_{11B}}\right). \tag{13}$$

Now $E(x_{1\cdot A}) = N_A p_{1\cdot A}$, and $E\left(\frac{x_{\cdot 1B}}{x_{11B}}\right) \approx \frac{p_{\cdot 1B}}{p_{11B}} + \frac{p_{01B}}{N_B p_{11B}^2}$, for large N_B (Chatterjee and Mukherjee, 2016a, Lemma 1, p.1061). Considering assumptions $(S3)$ for the sub-population U_B and $(S4)$, we have, from (12),

$$E(\hat{N}_A) \approx N_A + r\frac{p_{01B}}{p_1 p_{\cdot 1B}^2},$$

where $r = N_A/N_B$. Now, $V(x_{1\cdot A}) = N_A p_{1\cdot A}(1 - p_{1\cdot A})$.

Again, from the Lemma 1 (Chatterjee and Mukherjee, 2016a, p.1061), we have

$$\begin{aligned} V\left(\frac{x_{\cdot 1B}}{x_{11B}}\right) &= V\left(\frac{x_{01B}}{x_{11B}}\right) \\ &\approx \frac{E^2(x_{01B})}{E^2(x_{11B})}\left[\frac{V(x_{11B})}{E^2(x_{11B})} + \frac{V(x_{01B})}{E^2(x_{01B})} - 2\frac{Cov(x_{01B}, x_{11B})}{E(x_{01B})E(x_{11B})}\right], \end{aligned} \tag{14}$$

where $E(x_{01B}) = N_B p_{01B}$, $E(x_{11B}) = N_B p_{11B}$, $V(x_{01B}) = N_B p_{01B}(1 - p_{01B})$, $V(x_{11B}) = N_B p_{11B}(1 - p_{11B})$, and $Cov(x_{01B}, x_{11B}) = -N_B p_{01B} p_{11B}$.

Simplifying (14) we get

$$V\left(\frac{x_{\cdot 1B}}{x_{11B}}\right) \approx \frac{p_{\cdot 1B} p_{01B}}{N_B p_{11B}^3}. \tag{15}$$

Under the same assumptions $(S3)$ and $(S4)$, from (13) and (15) we obtain

$$\begin{aligned} V(\hat{N}_A) &\approx N_A p_{1\cdot A}\frac{p_{\cdot 1B} p_{01B}}{N_B p_{11B}^3} + N_A p_{1\cdot A}(1 - p_{1\cdot A})\left[\frac{p_{\cdot 1B}}{p_{11B}} + \frac{p_{01B}}{N_B p_{11B}^2}\right] \\ &= N_A(1 - p_1) + r\frac{p_{01B}(1 + p_1)}{p_1^2 p_{\cdot 1B}^2}. \end{aligned}$$

□

Acknowledgments

The authors are thankful to Prof. Anup Dewanji and Prof. Murari Mitra for many helpful comments and suggestions.
Conflict of Interest: None declared.

References

Bell, W. R. (1993). Using information from demographic analysis in post-enumeration survey (pes) estimation. *Journal of the American Statistical Association*, 88:1106–1118.

Bohning, D. and Heijden, P. V. D. (2009). Recent developments in life and social science applications of capturerecapture methods. *Advanced Statistical Analysis*, 93:1–3.

Bowman, K. O. and Shenton, L. R. (1998). *Encyclopedia of Statistical Sciences*. John Wiley & Sons.

ChandraSekar, C. and Deming, W. E. (1949). On a method of estimating birth and death rates and the extent of registration. *Journal of the American Statistical Association*, 44:101–115.

Chao, A., Tsay, P. K., Lin, S-H. Shau, W.-Y., and Chao, D.-Y. (2001). Tutorial in biostatistics: The applications of capture-recapture models to epidemiological data. *Statistics in Medicine*, 20:3123–3157.

Chatterjee, K. and Mukherjee, D. (2016a). An improved estimator of omission rate for census count: with particular reference to india. *Communication in Statistics: Theory and Methods*, 45:1047–1162.

Chatterjee, K. and Mukherjee, D. (2016b). An improved integrated likelihood population size estimation in dual-record system. *Statistics & Probability Letters*, 110:146–154.

Eisele, T. P., Lindblade, K. A., Rosen, D. H., Odhiambo, F., Vulule, J. M., and Slutsker, L. (2003). Evaluating the completeness of demographic surveillance of children less than five years old in western kenya: A capture-recapture approach. *The American Journal of Tropical Medicine and Hygiene*, 69:92–97.

Fay, R. E., Passel, J. S., and Robinson, J. G. (1988). The coverage of population in the 1980 census. *Evaluation and Research Report PHC8O-E4, US. Department of Commerce, Bureau of the Census*.

Granerod, J., Cousens, S., Davies, N. W. S., Crowcroft, N. S., and Thomas, S. L. (2013). New estimates of incidence of encephalitis in england. *Emerging Infectious Diseases*, 19:1455–1462.

Hook, E. B. and Regal, R. R. (1982). Validity of bernoulli census, log-linear, and truncated binomial models for correction for underestimates in prevalence studies. *American Journal of Epidemiology*, 116:168–176.

Isaki, C. T. and Schultz, L. K. (1986). Dual-system estimation using demographic analysis data. *Journal of Official Statistics*, 2:169–179.

Laplace, P. S. (1783). Sur les naissances, les mariages et les morts. *Histoire de L'Academie Royale des Sciences, Paris*, pages 693–702.

Nour, E. S. (1982). On the estimation of the total number of vital events with data from dual collection systems. *Journal of the Royal Statistical Society, Series A*, 145:106–116.

Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs: A Publication of Wildlife Society*, (62):3–135.

Wells, D. (1986). *The Penguin Dictionary of Curious and Interesting Numbers*. Penguin Books.

Whittaker, E. T. and Robinson, G. (1967). *The Calculus of Observations: An Introduction to Numerical Analysis*. Dover Publications.

Wolter, K. M. (1986). Some coverage error models for census data. *Journal of the American Statistical Association*, 81:338–346.

Wolter, K. M. (1990). Capture-recapture estimation in the presence of a known sex ratio. *Biometrics*, 46:157–162.

Table 2: Summary results of the estimates $\hat{N}_A$ and $\hat{\alpha}_A$ under Model I

			MME		MLE		Nour
$(p_{1\cdot A}, p_{\cdot 1A})$	N_A	α_A	$\hat{N}_A$ [RRMSE] C.I. for N_A	$\hat{\alpha}_A$	$\hat{N}_A$ [RRMSE] C.I. for N_A	$\hat{\alpha}_A$	$\hat{N}_A[RRMSE]$ C.I. for N_A
P1	240	0.4	241 [0.084] (204, 284)	0.388	242 [0.083] (206, 285)	0.397	202 [0.161] (188, 217)
		0.8	241 [0.083] (203, 284)	0.796	242 [0.086] (202, 287)	0.801	161 [0.333] (146, 175)
	1200	0.4	1201 [0.037] (1115, 1293)	0.398	1236 [0.066] (1122, 1402)	0.426	1010 [0.159] (977, 1042)
		0.8	1200 [0.038] (1114, 1294)	0.799	1244 [0.074] (1126, 1415)	0.809	803 [0.331] (770, 835)
P2	240	0.4	241 [0.088] (203, 287)	0.388	242 [0.087] (205, 287)	0.393	201 [0.167] (185, 217)
		0.8	241 [0.089] (202, 285)	0.796	242 [0.089] (202, 286)	0.796	159 [0.339] (144, 174)
	1200	0.4	1202 [0.039] (1117, 1298)	0.399	1213 [0.049] (1117, 1348)	0.407	1004 [0.164] (968, 1039)
		0.8	1202 [0.039] (1110, 1296)	0.799	1225 [0.058] (1114, 1384)	0.805	796 [0.337] (763, 828)
P3	240	0.4	240 [0.058] (216, 269)	0.380	241 [0.058] (217, 270)	0.388	219 [0.090] (207, 232)
		0.8	240 [0.058] (215, 270)	0.794	242 [0.062] (215, 271)	0.798	198 [0.175] (186, 210)
	1200	0.4	1201 [0.026] (1141, 1263)	0.397	1209 [0.031] (1146, 1285)	0.408	1096 [0.088] (1068, 1124)
		0.8	1200 [0.026] (1142, 1262)	0.799	1227 [0.057] (1149, 1407)	0.804	991 [0.174] (965, 1017)
P4	240	0.4	241 [0.054] (217, 268)	0.382	241 [0.053] (219, 267)	0.397	221 [0.083] (210, 232)
		0.8	240 [0.053] (217, 267)	0.795	242 [0.053] (216, 270)	0.791	200 [0.170] (188, 211)
	1200	0.4	1201 [0.024] (1146, 1259)	0.397	1204 [0.025] (1148, 1264)	0.402	1104 [0.081] (1079, 1129)
		0.8	1200 [0.024] (1145, 1258)	0.799	1221 [0.049] (1148, 1372)	0.807	998 [0.169] (972, 1023)
P5	240	0.4	241 [0.106] (198, 296)	0.391	243 [0.105] (200, 297)	0.397	191 [0.206] (174, 208)
		0.8	242 [0.107] (196, 297)	0.797	247 [0.114] (201, 309)	0.801	140 [0.418] (124, 155)
	1200	0.4	1201 [0.047] (1094, 1319)	0.398	1274 [0.076] (1163, 1375)	0.445	957 [0.203] (920, 997)
		0.8	1202 [0.047] (1095, 1319)	0.799	1229 [0.061] (1109, 1366)	0.805	700 [0.417] (666, 734)
P6	240	0.4	242 [0.115] (195, 301)	0.381	243 [0.115] (196, 302)	0.396	187 [0.226] (168, 206)
		0.8	242 [0.114] (194, 302)	0.797	243 [0.115] (194, 303)	0.798	137 [0.430] (122, 152)
	1200	0.4	1202 [0.050] (1090, 1327)	0.399	1213 [0.054] (1095, 1345)	0.404	934 [0.222] (893, 975)
		0.8	1203 [0.050] (1092, 1325)	0.799	1224 [0.068] (1104, 1458)	0.803	685 [0.430] (651, 719)

Table 3: Summary results of the estimates $\hat{N}_A$ and $\hat{N}_B$, and $\hat{\alpha}_0$ under Model II

			MLE			Nour	
$(p_{1.k},\ p_{.1k})$	(N_A, N_B)	α_0	$\hat{N}_A$ [RRMSE] C.I. for N_A	$\hat{N}_B$ [RRMSE] C.I. for N_B	$\hat{\alpha_0}$	$\hat{N}_A[RRMSE]$ C.I. for N_A	$\hat{N}_B[RRMSE]$ C.I. for N_B
P1	(240, 200)	0.4	241 [0.028]	201 [0.038]	0.407	201 [0.167]	168 [0.162]
			(231, 254)	(189, 216)		(185, 217)	(155, 182)
		0.8	244 [0.050]	203 [0.057]	0.805	159 [0.338]	134 [0.333]
			(228, 271)	(185, 227)		(144, 173)	(120, 146)
	(1200, 1000)	0.4	1201 [0.009]	1236 [0.014]	0.399	1004 [0.159]	842 [0.158]
			(1178, 1223)	(973, 1031)		(970, 1036)	(812, 871)
		0.8	1201 [0.012]	1001 [0.018]	0.801	796 [0.337]	669 [0.331]
			(1174, 1231)	(965, 1036)		(763, 830)	(640, 698)
P2	(240, 200)	0.4	242 [0.088]	201 [0.037]	0.406	202 [0.163]	167 [0.167]
			(231, 255)	(188, 215)		(186, 216)	(153, 181)
		0.8	243 [0.044]	203 [0.054]	0.803	160 [0.335]	133 [0.338]
			(227, 263)	(186, 226)		(145, 174)	(119, 146)
	(1200, 1000)	0.4	1200 [0.010]	1001 [0.014]	0.401	1007 [0.161]	837 [0.164]
			(1176, 1226)	(974, 1031)		(973, 1040)	(805, 868)
		0.8	1201 [0.013]	1000 [0.018]	0.800	800 [0.334]	663 [0.337]
			(1171, 1231)	(1114, 1384)		(766, 831)	(633, 692)
P3	(240, 200)	0.4	243 [0.042]	203 [0.050]	0.415	220 [0.088]	183 [0.092]
			(234, 269)	(192, 224)		(208, 232)	(171, 194)
		0.8	248 [0.074]	207 [0.077]	0.809	198 [0.175]	197 [0.182]
			(233, 296)	(191, 247)		(187, 210)	(154, 176)
	(1200, 1000)	0.4	1201 [0.007]	1001 [0.009]	0.401	1099 [0.084]	1019 [0.151]
			(1186, 1217)	(983, 1020)		(1072, 1126)	(987, 937)
		0.8	1201 [0.007]	1001 [0.011]	0.799	994 [0.172]	826 [0.174]
			(1183, 1220)	(979, 1024)		(967, 1019)	(802, 850)
P4	(240, 200)	0.4	243 [0.034]	202 [0.037]	0.417	221 [0.085]	184 [0.084]
			(234, 262)	(193, 218)		(209, 232)	(174, 195)
		0.8	248 [0.075]	207 [0.078]	0.813	199 [0.172]	166 [0.170]
			(233, 297)	(191, 247)		(187, 210)	(155, 177)
	(1200, 1000)	0.4	1201 [0.006]	1001 [0.009]	0.400	1102 [0.082]	920 [0.081]
			(1187, 1216)	(983, 1019)		(1076, 1128)	(898, 942)
		0.8	1201 [0.008]	1001 [0.011]	0.801	996 [0.170]	832 [0.169]
			(1183, 1221)	(978, 1024)		(971, 1022)	(808, 855)
P5	(240, 200)	0.4	242 [0.034]	202 [0.046]	0.406	190 [0.211]	160 [0.206]
			(228, 257)	(186, 220)		(173, 208)	(144, 175)
		0.8	244 [0.060]	204 [0.069]	0.804	139 [0.422]	117 [0.418]
			(266, 273)	(182, 230)		(124, 155)	(102, 131)
	(1200, 1000)	0.4	1201 [0.012]	1001 [0.018]	0.400	952 [0.208]	798 [0.203]
			(1173, 1231)	(965, 1037)		(911, 992)	(763, 835)
		0.8	1201 [0.015]	1000 [0.022]	0.799	695 [0.421]	583 [0.418]
			(1163, 1239)	(956, 1047)		(661, 730)	(551, 614)
P6	(240, 200)	0.4	241 [0.033]	202 [0.045]	0.404	187 [0.217]	156 [0.224]
			(227, 257)	(184, 220)		(171, 207)	(139, 173)
		0.8	245 [0.059]	204 [0.071]	0.805	138 [0.427]	114 [0.432]
			(184, 220)	(171, 207)		(139, 173)	(100, 128)
	(1200, 1000)	0.4	1200 [0.013]	1001 [0.019]	0.401	943 [0.215]	779 [0.203]
			(1168, 1232)	(965, 1037)		(903, 985)	(741, 816)
		0.8	1200 [0.015]	1001 [0.019]	0.800	689 [0.426]	570 [0.435]
			(966, 1040)	(966, 1040)		(666, 734)	(644, 714)

Table 4: Data sets used in illustration of the proposed methods

Dataset	Stratum	x_{11}	x_{10}	x_{01}	Total
Encephalitis	Adult	39	290	39	368
	Children	20	78	15	113
Children Death	Male	30	153	8	191
	Female	15	173	7	195

Table 5: Summary results of real data analysis under Models I and II

Dataset	Stratum (k)		Model I	Model II	M_t Model
Encephalitis	Adult	$\hat{N}_k$[r.s.e.]	660 [0.077]	739 [0.012]	658 [0.212]
		C.I.	(563, 760)	(731, 748)	(463, 988)
		$\hat{\alpha_k}$	0.052	0.031	-
	Children	$\hat{N}_k$[r.s.e.]	197 [0.104]	213[0.072]	171[0.317]
		C.I.	(160, 241)	(160, 241)	(101, 314)
Children Death	Male	$\hat{N}_k$[r.s.e.]	268 [0.054]	250 [0.092]	231[0.244]
		C.I.	(244, 303)	(204, 302)	(151, 362)
		$\hat{\alpha_k}$	0.070	0.006	-
	Female	$\hat{N}_k$[r.s.e.]	276 [0.052]	262 [0.097]	275 [0.424]
		C.I.	(250, 306)	(212, 324)	(145, 552)