INTERNATIONAL JOURNAL OF DEVELOPMENT MATHEMATICS

ISSN: 3026-8656 (Print) | 3026-8699 (Online)

journal homepage: https://ijdm.org.ng/index.php/Journals

Estimation of Probabilities for Open Population Capture-Recapture Using Generalized Estimating Equations

Pascalis K. Matthew[a]*, Danjuma Jibasen[b] and Ikeme J. Dike[b]

[a]*Department of Mathematics and Statistics, Taraba State University, Jalingo, Nigeria.*

[b]*Department of Statistics, Modibbo Adama University, Yola, Nigeria.*

ARTICLE INFO

Article history:
Received 10 June 2025
Received in revised form 10 August 2025
Accepted 20 August 2025

Keywords:
Capture-recapture, Open Population, Capture probabilities, GEE

MSC 2020 Subject classification:
62J12, 62J99, 62P10, 62P12

ABSTRACT

This study was designed to estimate the capture probabilities for open population capture-recapture experiment using Generalized Estimating Equation (GEE) approach accounting for heterogeneity, behavioural effects and environmental effects. An open population capture-recapture data was simulated using three different trapping occasions(m=4,5,6) with population sizes (N=500,1000).The results show that,M_{bhe} has the best performance in estimating the capture probability because of the smallest value of RMSE as compare to M_h,M_{bh} across the different simulations. The study concludes that; environmental covariates plays important role when estimating the capture probabilities of open population capture-recapture experiments.

1. Introduction

Capture-recapture is a research methodology in which multiple instances of observation are gathered on the identical individuals over a period of time. This technique is employed to estimate the size of a population and other associated parameters by examining the proportions of marked and unmarked individuals (Wesson *et al*., 2023). When the study population comprises a finite number of N elements, with N being unknown, capture-recapture methods can be utilized to estimate the size of N (Koivuniemi *et al*., 2019). The population under investigation is sampled on two or more occasions, during which unmarked animals are uniquely marked (typically using numbered leg bands in avian studies); previously marked animals are recorded upon recapture, and subsequently, all animals are released back into the population (Wesson *et al*., 2017). At the conclusion of the study, the researcher possesses the complete capture history of each handled animal. Capture-recapture methods are widely employed in animal population estimation, as well as in other disciplines such as quality control and epidemiology (Akanda & Alpizar-Jara, 2017). Numerous models and their corresponding inferential procedures have been proposed in the literature on capture-recapture to estimate animal population size based on capture-recapture data (Chowdhury et al., 2019). There exist two categories of capture-recapture models: (a) models that assume a constant population throughout the duration of the study (closed population models), and (b) models that allow for the addition of individuals due to births, immigration, or reintroduction, as well as the removal of individuals from the population due to deaths, emigration, or other causes (open population models). Open population models can also be classified based on whether the recaptures pertain to observations of live animals or deceased animals. The capture-recapture data collected on the same individuals across successive capture occasions represent binary longitudinal or repeated measurements data (Akanda & Alpizar-Jara, 2014). The occurrence of these repeated observations often exhibits a correlation that persists over time. This dependency or correlation structure can be induced by incorporating individual heterogeneity. Failing to account for this dependency may result in estimates that are biased (Chao *et al*., 2021). The Generalized Estimating Equations (GEE) approach serves as a potent statistical tool that expands the application of capture-recapture models to open populations, thereby enabling more precise population estimates. The GEE approach addresses the limitations of closed population models by providing a framework for estimating population size while considering the demographic changes. Worthington *et al*. (2018) devised a multi-state model to estimate the size of a closed population based on capture-recapture studies. The

* Corresponding author. Tel.: +2347039321996
E-mail address: pkadaro@yahoo.com (Pascalis K Matthew)
https://doi.org/10.62054/ijdm/0203.16

simulation results demonstrate that the estimation of population size may become biased if the movement between states is not taken into account. Mamadou *et al*. (2019) put forth a novel algorithm for estimating the parameters of a robust design with a substantial number of capture occasions and a simple parametric bootstrap variance estimator. Kevin *et al*. (2017) employed capture-recapture techniques to estimate the number of men who engage in sexual activities with men and female sex workers in 11 selected towns in Uganda. The study utilized the conventional 2-source capture-recapture (CRC) method to estimate the population of female sex workers and men who engage in sexual activities with men. You *et al*. (2021) introduced a modern approach to estimate population size based on capture-recapture designs of k-samples. The study encompasses models that assume a single constraint (identification assumption) on the k-dimensional distribution, ensuring the identification of the target quantity and the unrestricted nature of the statistical model. The study presents solutions for linear and non-linear constraints commonly assumed to identify capture-recapture models, including the absence of k-way interaction in linear and log-linear models, as well as independence or conditional independence. Kreshpaj *et al*. (2021) employed capture-recapture methods to estimate the extent of under-reporting of occupational injuries (OIs) among precarious and non-precarious workers in Sweden in 2013. The study empirically demonstrated that under-reporting of OIs is 50% higher among precariously employed workers in Sweden. Capture-recapture studies were conducted to estimate the population of female sex workers in Ghana (Guure *et al*., 2021), in Bhutan (Khandu *et al*., 2021), in Juba (Okiria *et al*., 2019), and in Hai Phong province (Nguyen et al., 2021). Wesson *et al.* (2023) utilized capture-recapture methods to estimate the size of hidden populations by leveraging the proportion of overlap in the population across independent lists. The study employed simulations to model scenarios where capture-recapture methods yield biased or unbiased estimates and compared various model selection criteria. The study concluded that conventional capture-recapture model selection fails when dealing with sparse cell counts. This simplistic approach to model selection should be replaced with the implementation of multiple different models in order to triangulate the truth in real-world applications. Huang and Pan (2021) proposed an innovative method that simultaneously models the mean and within-subject correlation coefficients for longitudinal binary data, while taking into account the constraints imposed by upper bounds. A joint GEE technique was formulated for the specific purpose, and the resulting correlation coefficients were demonstrated to satisfy the imposed constraints. The parameter estimators were proven to possess asymptotic normality, and simulation studies were conducted under various scenarios. These studies indicate that the suggested joint GEE method performs remarkably well even when the working covariance structures are not accurately specified. Zhuoran *et al*. (2023) employed GEE in order to examine the relationship between exposures and hearing loss due to its robustness against the misspecification of the correlation matrix. The study proposed to model the correlation coefficients and employ second-order GEE for estimating the correlation parameters. Through simulation studies, it was observed that the proposed method achieves a higher level of efficiency for the coefficients of the covariates related to within-cluster variation compared to the coefficients of cluster-level covariates. Asep *et al*. (2020) utilized GEE as one of the longitudinal data regression methods for composite index data on children's welfare (CWCI) in the province of West Java. The research concluded that this approach is significant as an alternative method for analyzing longitudinal data when a correlation issue arises in the observed data. Nasrin *et al*. (2021) put forth a method for analyzing spatially clustered real data based on GEE, which was originally designed for analyzing longitudinal data. A simulation study was conducted, and the proposed GEE approach yielded superior results in terms of parameter estimation and capturing spatial processes compared to the widely used conditional auto-regressive model. Goldenberg *et al*. (2023) employed bivariate and multivariate logistic regression with GEE to model the association between unstable housing exposure and evictions with intimate partner violence (IPV) and workplace violence among a community-based longitudinal cohort of cisgender and transgender women sex workers in Vancouver, Canada, from 2010 to 2019. The study concludes that women sex workers face a considerable burden of unstable housing and evictions, which are connected to higher odds of experiencing intimate partner and workplace violence. Thus, there is an urgent need for improved access to safe, women-centered, and non-discriminatory housing. The applications of the GEE approach by Zhang (2012), Akanda and Alpizar-Jara (2014a, 2017), and Chao *et al*. (2021) were all focused on closed population capture-recapture models. The objective of this paper is to propose a GEE approach for estimating the captured-probability of an open population capture-recapture data while considering individual heterogeneity, behavioral effects, and environmental covariates.

2. Model Formulation

2.1. Linking GEE to Capture-Recapture

Let N be the total number of individuals in the population of study, and let m be the possible number of capture occasions, $(m \geq 2)$.Let Y_{ij} the indicator of the i^{th} individual being caught in the j^{th} occasion, that is, $Y_{ij} = 1$ if i^{th} individual is captured in the j^{th} occasion, and $Y_{ij} = 0$, otherwise. Let T_i be the number of samples that the i^{th} individual belongs to, then let $T_i = \sum_{j=1}^{m} Y_{ij}$ .Let $Y_{ij} = \left(y_{i1}, \dots, y_{im}\right)^T$ be the $m \times 1$ random vector to record the capture history of the i^{th} individual for the m occasions. Covariates z_i and x_{ij} are associated with the i^{th} individual and the j^{th} occasion, for $i = 1, \dots, N$, $j = 1, \dots, m$. Let z_i be an individual measurable covariate, such as age, sex or weight, etc. Let x_{ij} be a measurable occasion related covariate for the i^{th} individual at j^{th} occasion, and let

$$X_i = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_i & z_i & \cdots & z_i \\ x_{i1} & 0 & \cdots & 0 \\ 0 & x_{i2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & x_{im} \end{bmatrix}^T$$

be the covariance matrix.

2.1.1 Individual heterogeneity covariates model (M_h)

Zhang (2012) proposed the probability P_{ij} that the i^{th} individual is captured on j^{th} occasion is given by

$$P_{ij} = \Pr(Y_{ij} = 1 \mid z_i, x_j) = h(\beta_0 + \beta_1 z_i + \beta_{j+1} x_{ij}) \tag{1}$$

for $i = 1, \dots, N$, $j = 1, \dots, m$, where $h(u) = \frac{e^u}{1+e^u} = (1 + (e^{-u}))^{-1}$ is the logistic function. However if $x_{i_1 j} = x_{i_2 j} = x_j$ for $i_1 \neq i_2$, this condition give rise to model (2) below

$$P_{ij} = \Pr(Y_{ij} = 1 \mid z_i, x_j) = h(\beta_0 + \beta_1 z_i + \beta_{j+1} x_j) \tag{2}$$

for $i = 1, \dots, N$, $j = 1, \dots, m$.

2.1.2. Individual heterogeneity and behavioural covariates model (M_{bh})

Akanda and Alpizar-Jara (2014) Models introduced v_{ij} into the model (2) which is an indicator that, the i^{th} individual is captured prior to j^{th} occasion, and which depends on occasions as well as the individual i.e

$$v_{ij} = \begin{cases} 1, & caught \\ 0, & otherwise \end{cases} \tag{3}$$

The probability P_{ij} that the i^{th} individual is captured on j^{th} occasion is given by

$$P_{ij} = \Pr(Y_{ij} = 1 \mid z_i, x_j, v_{ij}) = h(\beta_0 + \beta_1 z_i + \beta_{j+1} x_j + \beta_{bh} v_{ij}) \tag{4}$$

for $i = 1, \dots, N$, $j = 1, \dots, m$.

2.1.3. Individual heterogeneity, behavioural and environmental covariates model (M_{bhe} ***)***

Matthew *et al*. (2023) introduced f_{ij} into the (4) which denote the environmental covariates such as mean temperature, humidity or rainfall recorded for i^{th} individual at j^{th} occasions.

Therefore, the probability P_{ij} that the i^{th} individual is captured on j^{th} occasion is given by

$$P_{ij} = \Pr(Y_{ij} = 1 | z_i, x_j, v_{ij}, f_{ij}) = h(\beta_0 + \beta_1 z_i + \beta_{j+1} x_j + \beta_{bh} v_{ij} + \beta_{ev} f_{ij}) \tag{5}$$

for $i = 1,..., N$, $j = 1,..., m$. In this case the covariance matrix X_i now becomes

$$X_i = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_i & z_i & \cdots & z_i \\ x_1 & x_2 & \cdots & x_m \\ v_{i1} & v_{i2} & \cdots & v_{im} \\ f_{i1} & f_{i2} & \cdots & f_{im} \end{bmatrix}^T \tag{6}$$

The assumptions of the model are:

i. Capture probabilities vary according to environmental characteristics.
ii. Capture probabilities vary according to individual heterogeneity and environment.
iii. probabilities of capture vary according to an individual's prior capture history and the environmental covariates.

The mean vector of Y_i is specified as

$$\mu_i = \begin{bmatrix} E(Y_{i1}) \\ E(Y_{i2}) \\ \vdots \\ E(Y_{im}) \end{bmatrix} = \begin{bmatrix} P_{i1} \\ P_{i2} \\ \vdots \\ P_{im} \end{bmatrix} = P_i \tag{7}$$

$P_{ij} = \mu_{ij} = \Pr(Y_{ij} = 1 / z_i, x_{ij}, f_{ij}, v_{ij})$; for $i = 1,..., N$; $j = 1,..., m$. The probability of not capturing the i^{th} individual on the j^{th} occasion is given by $(1 - P_{ij})$, and the variance of Y_{ij} is $P_{ij}(1 - P_{ij}) = \mu_{ij}(1 - \mu_{ij})$,

Therefore, the $m \times m$ variance-covariance matrix of Y_i is given by

$$Var(Y_i) = \begin{bmatrix} Var(Y_{i1}) & Cov(Y_{i1}, Y_{i2}) & \cdots & Cov(Y_{i1}, Y_{im}) \\ Cov(Y_{i1}, Y_{i2}) & Var(Y_{i2}) & \cdots & Cov(Y_{i2}, Y_{im}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Y_{i1}, Y_{im}) & Cov(Y_{i2}, Y_{im}) & \cdots & Var(Y_{im}) \end{bmatrix} \tag{8}$$

Let $R_i(\alpha)$ denote the working correlation matrix, which is an approximation of the actual correlation matrix of Y_i. The working correlation matrix can be expressed in the form:

$$V_i = A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}} \tag{9}$$

Where, $A_i = diag\left[Var(Y_{i1}),..., Var(Y_{im})\right]$ is a $m \times m$ diagonal matrix with

$$A_i = \begin{bmatrix} Var(Y_{i1}) & 0 & \cdots & 0 \\ 0 & Var(Y_{i2}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Var(Y_{im}) \end{bmatrix}$$

where $\beta = (\beta_0, \beta_1, \beta_2)^T$ is the vector of parameters and $m \times m$ working correlation structure among $Y_{i1},...,Y_{im}$ is $R_i(\alpha) = Corr(Y_i)$, which describes the average dependency of individuals being captured from occasion to occasion, then the $R_i(\alpha)$ has the form: $R_i(\alpha) = I$, where I is a $m \times m$ identity matrix.

Thus, Y_{ij} is conditional on the captured individuals n (i.e., $T_i \geq 1$) as in (Zhang, 2012; Akanda & Alpizar-Jara, 2014; Matthew et al., 2023) . If $Y_{i1}, Y_{i2},...,Y_{im}$ are independent, given that the i^{th} individual is captured at least in one occasion, the probability that, the i^{th} individual is captured on j^{th} occasion is given by

$$P(Y_{ij} = 1 | T_i \geq 1) = \frac{P_{ij}}{1 - \prod_{j=1}^{m} (1 - P_{ij})} \tag{10}$$

By adopting the notation in Matthew *et al*. (2023), Let D_i be the matrix of derivatives $\frac{du_i}{d\beta^T}$. Since $z_i, x_{ij}, v_{ij}, f_{ij}$ are observable covariates for each captured individual, then the vector of parameters $\beta = (\beta_0, \beta_1, \beta_2, \beta_{bh}, \beta_{ev})^T$ for model (5) can be estimated by solving the following generalized estimation equations

$$h(\beta) = \sum_{i-1}^{N} D_i^T V_i^{-1} (Y_i - \mu_i) = 0 \tag{11}$$

Therefore, (11) is a transformed GEE for capture-recapture models. Thus given the following GEE model

1. **Model for Individual Heterogeneity Covariates (M_h)**

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_0 + \beta_{ih} z_i ; (i = 1,..., n; j = 1,..., m) \tag{12}$$

This model assumes that the probabilities of capture vary according to individual covariates.

Let $p_{ij} = h(\beta_0 + \beta_{bh} v_{ij} + \beta_{ih} z_i)$,then,

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} (Y_{ij} - \mu_{ij}) + \sum_{i=1}^{n} \sum_{j=1}^{m} (Y_{ij} - \mu_{ij}) z_i = 0 \tag{13}$$

2. **Model for behavioural and heterogeneity covariates M_{bh}**

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_0 + \beta_{bh} v_{ij} + \beta_{ih} z_i ; (i = 1,..., n; j = 1,..., m) \tag{14}$$

This model assumes that the probabilities of capture vary according to an individual's heterogeneity covariates and prior capture history.

Let $p_{ij} = h(\beta_0 + \beta_{bh} v_{ij} + \beta_{ih} z_i)$,then,

$$=\sum_{i=1}^{n}\sum_{j=1}^{m}\left(Y_{ij}-\mu_{ij}\right)+\sum_{i=1}^{n}\sum_{j=1}^{m}\left(Y_{ij}-\mu_{ij}\right)z_i+\sum_{i=1}^{n}\sum_{j=1}^{m}\left(Y_{ij}-\mu_{ij}\right)v_{ij}=0 \quad (15)$$

3. **Model for behavioural, heterogeneity and environmental covariates (M_{bhe})**

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right)=\beta_0+\beta_{bh}v_{ij}+\beta_{ih}z_i+\beta_{ev}f_{ij};(i=1,\ldots,n;\ j=1,\ldots,m) \quad (16)$$

This model assumes that the probabilities of capture vary according to an individual's heterogeneity covariates, prior capture history and the environmental covariates.

Let $p_{ij} = h(\beta_0 + \beta_{bh} v_{ij} + \beta_{ih} z_i + \beta_{ev} f_{ij})$,then,

$$=\sum_{i=1}^{n}\sum_{j=1}^{m}\left(Y_{ij}-\mu_{ij}\right)+\sum_{i=1}^{n}\sum_{j=1}^{m}\left(Y_{ij}-\mu_{ij}\right)z_i+\sum_{i=1}^{n}\sum_{j=1}^{m}\left(Y_{ij}-\mu_{ij}\right)v_{ij}+\sum_{i=1}^{n}\sum_{j=1}^{m}\left(Y_{ij}-\mu_{ij}\right)f_{ij}=0 \quad (17)$$

2.2 Fitting Generalized Estimating Equation Models

The following steps was used in fitting the proposed GEE models as suggested (Wentao,2019):

1. **Linear predictor and best link function:** A link transformation function is specified that allows the response variable to be expressed as a function of parameter estimate β of an additive model, with the model expected value of the marginal response for the population $\mu_i = \mathrm{E}(Y_i)$ to a linear combination of the covariates. The choice of link function depends primarily on the distribution specified.
2. **Distribution of the response variable:** The GEE allows the specifications of the distribution from the exponential family of distributions, including Normal, inverse Normal, Poisson, Binomial, Gamma distribution and negative Binomial. Prior knowledge of the distribution of the response variable is important.
3. **Structure of the correlation within the response variable:** Although GEE is robust when it comes to misspecification of the correlation structure, it is important to take precaution when it comes to specifying the structure. The working correlation employed are as follows: (a) Independent Structure(b)Exchangeable Structure (c) Autoregressive Structure (d) Unstructured Structure

2.3 Model Selection of Generalized Estimating Equations

A modification of AIC known as the Quasi-Likelihood Information Criterion (QIC) was introduced by Pan (2001). The QIC is constructed on the basis of independence of the working correlation assumption. The QIC can be used to select the appropriate working correlation structure that is nearest to the true model from a set of potential models in GEE analyses.

Then the simplified version of QIC given by Pan (2001) as cited by Wang et al. (2015) is denoted by

$$QIC = -2Q(\hat{\mu}; I) + 2p \quad (18)$$

Therefore, QIC is used to select an optimal correlation structure. The model with the smallest QIC will be chosen as the best parsimonious model with the best correlation structure (Pan, 2001).

2.4 Capture-Probability Estimator and the Standard Error

Let p_{ij} $(i=1,\ldots,n;\ j=1,\ldots,m)$ represent the probability of capture for the i^{th} individual during the j^{th} capture occasion. The open-population capture–recapture models employed here relate a differentiable function of the p_{ij} to a linear function of external covariates (Matthew *et al.*, 2023). These covariates include sex, weight, behavioural effect and mean temperature. Individual capture probabilities were modeled with a logistic regression, so that,

$$P_{ij}=\frac{e^{\beta_0+\beta_{sex}sex_i+\beta_{wt}wt_i}}{1+e^{\beta_0+\beta_{sex}sex_i+\beta_{wt}wt_i}} \tag{19}$$

for M_h covariates.

$$P_{ij}=\frac{e^{\beta_0+\beta_{sex}sex_i+\beta_{wt}wt_i+\beta_{bh}v_{ij}}}{1+e^{\beta_0+\beta_{sex}sex_i+\beta_{wt}wt_i+\beta_{bh}v_{ij}}} \tag{20}$$

for M_{bh} covariates and

$$P_{ij}=\frac{e^{\beta_0+\beta_{sex}sex_i+\beta_{wt}wt_i+\beta_{bh}v_{ij}+\beta_{bh}f_{ij}}}{1+e^{\beta_0+\beta_{sex}sex_i+\beta_{wt}wt_i+\beta_{bh}v_{ij}+\beta_{bh}f_{ij}}} \tag{21}$$

for M_{bhe} covariates.

The estimates for the parameters β_i are obtained by the method of maximum likelihood. Given a set of observed captured histories, I_{ij} , where $I_{ij}=1$ if the i^{th} individual was seen at the j^{th} occasion and $I_{ij}=0$ otherwise.

The estimated capture probability and standard error for the i^{th} individual at j^{th} occasion is labelled as $\hat{p}_{ij}$ and $\hat{\pi}_{\hat{p}_{ij}}$ respectively. So that, $E\left[\hat{p}_{ij}\right]=p_{ij}$ and $E\left[\hat{\pi}_{\hat{p}_{ij}}\right]=\pi_{\hat{p}_{ij}}$.

3. Results

3.1 Simulation 1

In order to illustrate the GEE approach on open population capture-recapture data, a non-spatial open population data was simulated using R with the following parameters; $n=500$, $\phi=0.7$, $\lambda=1.1$, $p=0.3$, and $m=4$ trapping occasions was used to simulate the recapture history. The individual discrete covariate, sex, was simulated based on the Binomial distribution $BIN(500,0.5)$. The individual continuous covariate, weight, environmental covariate, temperature, were simulated based on the Normal distribution $N(15,5)$ and $N(30,10)$ respectively. The indicator variable, behavioural effect was generated from the captured history. Table 1 shows the Capture-Probability estimates for open population Capture-Recapture models when *m=4, N=500*.

Table 1: Capture-Probability estimates for open population Capture-Recapture models *m=4, N=500*

Model	$\hat{P}_{ij}$	$\hat{\pi}_{\hat{p}_{ij}}$	*RMSE*
M_h	0.261	10.900	99.3244
M_{bh}	0.250	14.086	99.3287
M_{bhe}	0.279	12.991	99.2484

3.2 Simulation 2

In order to illustrate the GEE approach on open population capture-recapture data, a non-spatial open population data was simulated using R with the following parameters; $n=500$, $\phi=0.7$, $\lambda=1.1$, $p=0.3$, and $m=5$ trapping occasions was used to simulate the recapture history. The individual discrete covariate, sex, was simulated based on the Binomial distribution $BIN(500,0.5)$. The individual continuous covariate, weight, environmental covariate, temperature, were simulated based on the Normal distribution $N(15,5)$ and $N(30,10)$ respectively. The indicator variable, behavioural effect was generated from the captured history. Table 2 shows the Capture-Probability Estimates for open population Capture-Recapture models when *m=5, N=500*

Table 2: Capture-Probability Estimates for open population Capture-Recapture models *m=5, N=500*

Model	$\hat{P}_{ij}$	$\hat{\pi}_{\hat{p}_{ij}}$	*RMSE*
M_h	0.245	12.2181	99.2453
M_{bh}	0.341	11.7584	99.2081
M_{bhe}	0.352	11.7320	99.2060

3.3 Simulation 3

In order to illustrate the GEE approach on open population capture-recapture data, a non-spatial open population data was simulated using R with the following parameters; $n = 500$, $\phi =$ 0.7, $\lambda =$ 1.1, $p =$ 0.3, and $m = 6$ trapping occasions was used to simulate the recapture history. The individual discrete covariate, sex, was simulated based on the Binomial distribution $BIN(500,0.5)$. The individual continuous covariate, weight, environmental covariate, temperature, were simulated based on the Normal distribution $N(15,5)$ and $N(30,10)$ respectively. The indicator variable, behavioural effect was generated from the captured history. Table 3 shows the Capture-Probability Estimates for open population Capture-Recapture models when *m=6, N=500*.

Table 3: Capture-Probability Estimates for open population Capture-Recapture models *m=6, N=500*

Model	$\hat{P}_{ij}$	$\hat{\pi}_{\hat{p}_{ij}}$	*RMSE*
M_h	0.285	13.544	99.2960
M_{bh}	0.294	12.569	99.2874
M_{bhe}	0.230	12.536	99.2156

3.4 Simulation 4

In order to illustrate the GEE approach on open population capture-recapture data, a non-spatial open population data was simulated using R with the following parameters; $n = 1000$, $\phi =$ 0.8, $\lambda =$ 1.2, $p =$ 0.4, and $m = 4$ trapping occasions was used to simulate the recapture history. The individual discrete covariate, sex, was simulated based on the Binomial distribution $BIN(1000,0.5)$. The individual continuous covariate, weight, environmental covariate, temperature, were simulated based on the Normal distribution $N(15,5)$ and $N(30,10)$ respectively. The indicator variable, behavioural effect was generated from the captured history. Table 4 shows the Capture-Probability estimates for open population Capture-Recapture models when *m=4, N=1000*

Table 4: Capture-Probability estimates for open population Capture-Recapture models *m=4, N=1000*

Model	$\hat{P}_{ij}$	$\hat{\pi}_{\hat{p}_{ij}}$	*RMSE*
M_h	0.285	13.544	99.2960
M_{bh}	0.320	17.893	99.2874
M_{bhe}	0.335	17.432	99.2156

3.5 Simulation 5

In order to illustrate the GEE approach on open population capture-recapture data, a non-spatial open population data was simulated using R with the following parameters; $n = 1000$, $\phi = 0.8$, $\lambda = 1.2$, $p = 0.4$, and $m = 5$ trapping occasions was used to simulate the recapture history. The individual discrete covariate, sex, was simulated based on the Binomial distribution $BIN(1000, 0.5)$. The individual continuous covariate, weight, environmental covariate, temperature, were simulated based on the Normal distribution $N(15,5)$ and $N(30,10)$ respectively. The indicator variable, behavioural effect was generated from the captured history. Table 5 shows the Capture-Probability estimates for open population Capture-Recapture models *m=5, N=1000*.

Table 5: Capture-Probability estimates for open population Capture-Recapture models *m=5, N=1000*

Model	$\hat{P}_{ij}$	$\hat{\pi}_{\hat{p}_{ij}}$	*RMSE*
M_h	0.309	18.432	199.3017
M_{bh}	0.328	17.625	199.2573
M_{bhe}	0.358	17.608	199.2252

3.6 Simulation 6

In order to illustrate the GEE approach on open population capture-recapture data, a non-spatial open population data was simulated using R with the following parameters; $n = 1000$, $\phi = 0.8$, $\lambda = 1.2$, $p = 0.4$, and $m = 6$ trapping occasions was used to simulate the recapture history. The individual discrete covariate, sex, was simulated based on the Binomial distribution $BIN(1000, 0.5)$. The individual continuous covariate, weight, environmental covariate, temperature, were simulated based on the Normal distribution $N(15,5)$ and $N(30,10)$ respectively. The indicator variable, behavioural effect was generated from the captured history. Table 6 shows the Capture-Probability estimates for open population Capture-Recapture models when *m=6, N=1000*

Table 6: Capture-Probability estimates for open population Capture-Recapture models *m=6, N=1000*

Model	$\hat{P}_{ij}$	$\hat{\pi}_{\hat{p}_{ij}}$	*RMSE*
M_h	0.222	17.243	1193.2409
M_{bh}	0.251	17.246	1193.2222
M_{bhe}	0.263	17.206	1193.2053

4. Discussion

Table 1 shows the estimated capture probability $\hat{P}_{ij}$ =0.261 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =10.900 and *RMSE* =99.3244 when M_h is considered. The estimated captured-probability $\hat{P}_{ij}$ =0.250 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =14.086 and *RMSE* =99.3287 when M_{bh} is considered. The estimated captured-probability $\hat{P}_{ij}$ =0.279 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =12.991 and *RMSE* =99.2484 when M_{bhe} covariates model is considered. In comparison among M_h, M_{bh} ,and M_{bhe} Table 1 reveals that, the M_{bhe} performs better in estimating the captured-probability because of the smallest value of *RMSE* =99.2484. Table 2 shows the estimated captured-probability $\hat{P}_{ij}$ =0.245 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =12.2181 and *RMSE* =99.2453 when M_h is considered. The estimated captured-probability $\hat{P}_{ij}$ =0.341 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =11.7584 and *RMSE* =99.2081 when M_{bh} is considered. The estimated captured-probability $\hat{P}_{ij}$ =0.352 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =11.7320 and *RMSE* =99.2060 when M_{bhe} covariates model is considered. In comparison among M_h, M_{bh} ,and M_{bhe} Table 2 reveals that, the M_{bhe} performs better in estimating the captured-probability because of the smallest value of *RMSE* =99.2060. Table 3 shows the estimated captured-probability $\hat{P}_{ij}$ =0.285 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =13.544 and *RMSE* =99.2960 when M_h is considered .The estimated captured-probability $\hat{P}_{ij}$ =0.294 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =12.569 and *RMSE* =99.2874 when M_{bh} is considered. The estimated captured-probability $\hat{P}_{ij}$ =0.230 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =12.536 and *RMSE* =99.2156 when M_{bhe} is considered. In comparison among M_h , M_{bh} ,and M_{bhe} Table 3 reveals that, the M_{bhe} performs better in estimating the captured-probability because of the smallest value of *RMSE* =99.2156. Table 4 shows the estimated captured-probability $\hat{P}_{ij}$ =0.285 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =13.544 and *RMSE* =99.2960 when M_h is considered .The estimated captured-probability $\hat{P}_{ij}$ =0.320 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =17.893 and *RMSE* =99.2874 when M_{bh} is considered. The estimated captured-probability $\hat{P}_{ij}$ =0.335 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =17.432 and *RMSE* =99.2156 when M_{bhe} is considered. In comparison among M_h , M_{bh} ,and M_{bhe} Table 4 reveals that, the M_{bhe} performs better in estimating the captured-probability because of the smallest value of *RMSE* =99.2156. Table 5 shows the estimated captured-probability $\hat{P}_{ij}$ =0.309 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =18.432 and *RMSE* =199.3017 when M_h is considered .The estimated captured-probability $\hat{P}_{ij}$ =0.328 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =17.625 and *RMSE* =199.2573 when M_{bh} is considered. The estimated captured-probability $\hat{P}_{ij}$ =0.358 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =17.608 and *RMSE* =199.2252 when M_{bhe} is considered. In comparison among M_h, M_{bh} ,and M_{bhe} Table 5 reveals that, the M_{bhe} performs better in estimating the captured-probability because of the smallest value of *RMSE* =199.2252. Table 6 shows the estimated captured-probability $\hat{P}_{ij}$ =0.222 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =17.243 and *RMSE* =1193.2409 when M_h is considered .The estimated captured-probability $\hat{P}_{ij}$ =0.251 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =17.246 and *RMSE* =1193.2222 when M_{bh} is considered. The estimated captured-probability $\hat{P}_{ij}$ =0.63 with a standard error $\hat{\pi}_{\hat{p}_{ij}}$ =17.206 and *RMSE* =1193.2053 when M_{bhe} is considered. In comparison among M_h , M_{bh} ,and M_{bhe} Table 6 reveals that, the M_{bhe} performs better in estimating the captured-probability because of the smallest value of *RMSE* =1193.2053.

5. Conclusion

In this study, a Generalized Estimating Equation (GEE) approach accounting for heterogeneity, behavioural effects and environmental effects to estimate the capture-probabilities for open population capture-recapture experiment was employed. An open population capture-recapture data was simulated using three different trapping occasions $(m = 4,5,6)$ with population sizes $(N = 500,1000)$ .The results of the simulations, show that, the comparison among M_h , M_{bh} ,and M_{bhe} shows that, the M_{bhe} performs better in estimating the capture probability because of the smallest value of *RMSE* across the different simulations. The study concludes that; environmental covariates plays important role when estimating the captured-probabilities of open population capture-recapture experiments.

References

Akanda, M.A.S., and Alpizar-Jara, R. (2014a). A generalized estimating equations approach for capture-recapture closed population models. *Environmental and Ecological Statistics*, 21(4), 667–688.

Akanda, M.A.S., and Alpizar-Jara, R. (2017). A generalized Estimating Equations Approach to model heterogeneity and Time dependence in Capture Recapture Studies. *European Journal of Ecology*, 3(1), 9–17.

Asep, S. A., Inge, W., and Hilda, N. (2020). Analysis of Longitudinal Regression Model Using the Generalized Estimating Equation (GEE) for the Child Welfare Composite Index (CWCI) in West Java. *Advances in Social Science, Education and Humanities Research, volume 550.* Proceedings of the 1st International Conference on Mathematics and Mathematics Education (ICMMED), 1-8.

Chowdhury, R.I., and Islam, M.A. (2017). *Generalized Estimating Equation. In: Analysis of Repeated Measures Data.* Springer, Singapore.

Chao, L., Ye, S., Qian, X., Stephen, L.R., Hui, H, and Yongtao, G. (2021). Mean corrected generalized estimating equations for longitudinal binary outcomes with report bias *Statistical Methods in Medical Research*. 31(2) 315–333.

Goldenberg, S.M., Buglioni, N., Krusi, A., Frost, E., Moreheart, S., Braschel, M., and Shannon K. (2023). Housing Instability and Evictions Linked to Elevated Intimate Partner and Workplace Violence Among Women Sex Workers in Vancouver, Canada: Findings of a Prospective, Community-Based Cohort, 2010–2019. *American journal of public health*. 113(4), 442-452.

Guure, C., Dery, S., Afagbedzi, S., Tun, W., Weir, S.S., Quaye, S., et al. (2021). National and subnational size estimation of female sex workers in Ghana 2020: comparing 3-source capture-recapture with other approaches. *PLoS ONE*, 16(9),1–16.

Huang, Y., and Pan, J. (2021). Joint generalized estimating equations for longitudinal binary data. *Computational Statistics & Data Analysis*, ***155***, 107110. https://doi.org/10.1016/j.csda.2020.107110

Kevin, A., Reena, H. D., and Hladik, W. (2017). Capture-Recapture Among Men Who Have Sex with Men and Among Female Sex Workers in 11 Towns in Uganda. *Creative Commons Open Access Journal.*

Khandu, L., Tobgay, T., Kinley, K., Choida, N., Tashi, T., Wangdi, T., et al. (2021). Characteristics and population size estimation of female sex workers in Bhutan. *Sex Transm Dis*,48(10),754–60.

Koivuniemi, M., Kurkilahti, M., Niemi, M., Auttila, M., and Kunnasranta, M. (2019). A mark–recapture approach for estimating population size of the endangered ringed seal (*Phoca hispida saimensis*). *PLoS ONE, 14*(3), 64-82.

Kreshpaj, B., Bodin, T., Wegman, D.H., Matilla, S.N., Burstrom, B., Kjellberg, K., et al. (2021). Under-reporting of non-fatal occupational injuries among precarious and non-precarious workers in Sweden. *Occup Environ Med*, 79(1), 3-9.

Mamadou, Y.O.I., Louis,.P.R., and Greg, R. (2019). Capture-Recapture Methods for Data on the Activation of Applications on Mobile Phones. *Journal of the American Statistical Association*. 114(525), 105-114.

Matthew, P.K., Jibasen, D., and Dike, I.J. (2023). Estimating the Population Parameters for Open Population Capture-Recapture Data Using Generalized Estimating Equations Approach. American University of Nigeria 1st International Conference Proceedings November 2023. 1(1);498-516.

McDonald, T. L., and Amstrup, S. C. (2001). Estimation of population size using open capture-recapture models.

Journal of Agricultural, Biological, and Environmental Statistics, 6(2), 206-220.

Nasrin, L., Mohammad, S.A., and Syed, S. H. (2021). A Generalized Estimating Equations Approach for Modeling Spatially Clustered Data. *Austrian Journal of Statistics*: 50(36),36-52.

Nguyen, L.T., Patel, S., Nguyen, N.T., Gia, H.H., Raymond, H.F., Hoang, V.T.H., et al. (2021). Population size estimation of female sex workers in Hai Phong, Vietnam:use of three source capture–recapture method. *J Epidemiol Glob Health*,**11**(2),194–199.

Okiria, A.G., Bolo, A., Achut, V., Arkangelo, G.C., Michael, A.T.I., Katoro, J.S., et.al. (2019). Novel approaches for estimating female sex worker population size in conflict affected South Sudan. JMIR *Public Heal Surveill*,5(1),1–8.

Pan, W. (2001). Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics,* 57(1), 120-125.

Wang, Y., Orla, M., Wang, Z., Bhatnagar, S.R., Schultz, J., and Moodie, E., M. (2015). The Perils of Quasi-Likelihood Information Criteria, *Stat Journal*, 4(2), 246-254.

Wentao, G.(2019). Bootstrap-Adjusted Quasi-Likelihood Information Criteria for Mixed Model Selection. PHD Dissertation, Bowling Green State University.

Wesson, P., Reingold, A., and Mc Farland, W. (2017). Theoretical and empirical comparisons of methods to estimate the size of hard-to-reach populations: a systematic review. *AIDS Behav* 45(3),415-440.

Wesson, P., Jewell, N.P., Willi McFarland., M.D and Glymour., M.M. (2023). Evaluating tools for capture-recapture model selection to estimate the size of hidden populations: It works in practice, but does it work in theory? *Annals of Epidemiology*, 77(4), 24–30.

Worthington, H., Rachel S., McCrea., Ruth, K., and Griffiths, R.A. (2018). Estimation of Population Size When Capture Probability Depends on Individual States. *Journal of Agricultural, Biological, and Environmental Statistics*, 24(1), 154–172.

You, Y., Laan M Van Der, Collender, P., Cheng, Q., Hubbard, A., Jewell, N.P., Hu, Z.T., Mejia, R., and Remais, J.(2021). Estimation of population size based on capture-recapture designs and evaluation of the estimation reliability. ArXiv.2105.05373v1[maths.ST]

Zhang, S. (2012). A GEE approach for estimating size of hard-to-reach population by using capture-recapture data. *Journal of Theoretical and Applied Statistics*. 46(2), 175–183.

Zhuoran W.E., Hanbing, Z., Sharon, C., Gary, C., and Molin, W. (2023). Generalized Estimating Equations for Hearing Loss Data with Specified Correlation Structures *Stat.ME*