Mambilla Journal of Science and the Environment Vol. 10, Issue 1, Page No. 127-139

MAMBILLA JOURNAL OF SCIENCE AND THE ENVIRONMENT

journal homepage: https://oer.tsuniversity.edu.ng/index.php/mjse

HYBRID AND HUMAN-IN-THE-LOOP MINING OF NATURAL LANGUAGE–PROGRAMMING LANGUAGE PAIRS ACROSS LANGUAGES

[1*]Matthew, P.K., [1]Chama, F.A., [1]Emmanuel, R.

[1]Department of Mathematics and Statistics, Taraba State University, Jalingo, Nigeria

ARTICLE INFO

Article history:
Received 14 January 2026
Received in revised form 02 February 2026
Accepted 06 February, 2026

Keywords:
Capture-Recapture, Out-of-School Children, Population Size, Taraba.

ABSTRACT

This study investigated the street-child phenomenon and the applicability of capture-recapture (C-R) methods for estimating such hard-to-count populations. The method was applied to estimate the number of street children in Taraba State, Nigeria. Primary data were collected using a purposive sampling design, with five wards chosen as strata from Jalingo, Gassol and Wukari Local Government Areas. The data obtained were coded, cross-matched according to specific matching criteria and a degree of matching. Four C-R models were employed: the *Model with Non-Factor Interaction Effect* (M_o), *Petersen Model*(M_s), *Listability Model* (M_a) *and Behavioral Response Model* (M_b) model performed best in all the selected wards with lowest AIC (Kona ward, AIC = 5.1229, Barade ward, AIC = 5.9221, Sintali ward, AIC = 6.1743, Kachalla-Sembe ward, AIC = 5.5345 and Mayogwoi ward, AIC = 5.1400); for Gassol LGA, (M_s)model performed best in Sandirde with AIC=5.9322 while behavioral response (M_b)model was most efficient in Mutum biyu I, Mutum biyu II, Namnai and Tutare wards (AIC = 5.4511, 5.0165, 4.3838 and 4.9032, respectively); for Wukari LGA the Petersen model (M_s) model performed best in Avyi, Bantaje and Hospital wards (AIC=6.2922,6.1837,6.1930) respectively while, behavioral response model (M_b) model performed best in Kente and Puje wards(AIC=5.6010,5.4224), respectively. Accordingly, the estimated numbers of street children per ward using the best-fitting model were: for Jalingo LGA: Kona ward, N = 101, Barade ward, N = 160, Sintali ward, N = 185, Kachalla-Sembe ward, N = 128 and Mayogwoi ward, N = 102.The estimated total number of street children in Jalingo LGA is 676; for Gassol LGA: Sandirde ward, N = 161, Mutum biyu I ward, N = 122, Mutum biyu II ward, N = 95, Namnai ward, N = 66 and Tutare ward, N = 189. The estimated total number of street children in Gassol ward is 533; for Wukari LGA: Avyi ward, N = 198, Bantaje ward, N = 186, Hospital, N = 187, Kente ward, N = 133 and Puje ward, N = 120. The estimated total number of street children in Wukari LGA is 824.This study demonstrates that capture-recapture methods can effectively estimate the size of elusive or difficult-to-count populations.

1. Introduction

Estimating the population of street children across the world, in continents or even within nations, has been an uphill task over the last two and a half decades. Between 1999 and 2016, UNICEF estimated the numbers of street children or homeless children as running into tens of million across the world but quoted in each case a figure of 123 million (UNICEF 2016). Capture-recapture (C-R) sampling is generally used to estimate the number of hard-to-find populations, where naturally, making a complete census of an entire animal population is nearly impossible and therefore cannot be taken easily.

In a typical wildlife survey, animals are captured, marked, released, and allowed to mix back into the population. On a subsequent survey, captured animals are counted, and the number of already-marked individuals is recorded. This

* Corresponding author: +2347039321996
E-mail address: pkadaro@yahoo.com

process may be repeated over *k* surveys, and the resulting capture and recapture data are used to estimate total population size, i.e., the number of animals never captured. Such experiments are also known as mark-recapture, tag-recapture, or multiple-record systems. The simplest case involves only two samples: an initial capture sample and a later recapture sample. The two-sample case is often called a dual-record system, particularly in the context of census undercount estimation. When three or more samples are used, it is referred to as a multi-list record system. In human population studies (e.g, epidemiology, social sciences) for estimating the number of drug users or homeless persons, these are termed dual-system methods (for two lists) or multi-list record systems (for more than two lists). In applying C-R methods to lists is analogous to a closed-population C-R experiment, which rests on the following assumptions:
(i) The population is closed – no births, deaths, immigration, or emigration during the sampling period (demographic closure).
(ii) Every individual in the defined population has an equal probability of being observed (captured) in any sample.
(iii) Individuals identified in one source can be perfectly matched to another source without error.
(iv) The sources are independent of one another–appearing on one list does not affect the probability of appearing on another.
Violations of any assumption affect estimate quality. Assumptions (i), (ii), and (iv) are not particularly difficult to meet when studying street children, though (iv) can be relaxed by using log-linear models (Wesson *et al*., 2023).
Despite the country's legal and institutional framework, disbursement of funds, loans and educational policies, Nigeria continues to be home to the highest number of out-of- school children (OOSC) in the world (UNICEF, 2022a). The country's Basic Education Statistics established the number of out-of- school children aged 6 to 11 at 10 million, and an estimate of 10.2 million in 2022 (Nigeria Digest of Education Statistics, 2022). However, if the secondary school-aged children are included, the out-of- school population reaches an estimated 20 million (UNESCO Global Education Monitoring, 2023). The OOSC phenomenon is even more prevalent in northern Nigeria (Taraba inclusive), with a sharp 15% decrease in primary school enrolment in the northern geopolitical zones compared to the southern regions (UNESCO, 2022). The number of out-of-school children in Taraba State is unknown, making it challenging for policymakers and stakeholders to develop effective strategies to address this issue. Traditional methods of estimating the population of out-of-school children, such as surveys and censuses, may under-estimate the true number due to difficulties in reaching all segments of the population. The capture-recapture method offers a potential solution to this problem by providing a more accurate estimate of the population size. The objectives of the study are to: (i) Estimate the population size of out-of-school children in Taraba State using the capture-recapture method. (ii) Identify the factors associated with being out-of-school among children in Taraba State. (iii) Inform policies and programs aimed at increasing access to education for out-of-school children.

1.1 Out-of-school children: The situation in Nigeria

Available literature shows there is a positive correlation between human development (UNDP, 2016) and the availability of formal primary education (Farswan, 2023). According to the most recent report from Statista (2023), with a median of 18.7 years, Africa is renowned to have the most youthful population in the world. The United Nations (2019) report on world population prospects projected that by the middle of the century, Africa will be home to over a billion children. With the wide increase in the young population, Evans and Acosta (2020) maintain that the continent is critical for addressing the global education crisis—particularly out-of-school children. Within this broader context, Nigeria's situation warrants particular attention. Despite not having the highest proportional rate of OOSC, Nigeria's sheer number of 20.2 million out-of- school children highlight an urgent need for comprehensive study. According to Adegbami and Adesanmi (2018), understanding the Nigerian situation can reveal critical barriers to education that persist despite ongoing interventions.

1.1.1 Economy of household

Due to factors such as inflation, unemployment and high cost of living that plague the Nigerian economy (Tule *et al*., 2017), financial resources are acutely limited in many homes. This makes education deprioritized in favour of addressing immediate needs for survival in these households (Olonade *et al*., 2022). Adebowale (2018) insists that families and households in Nigeria engaging in child labor and early marriage is indicative of the pervasive socioeconomic crisis in the region. These challenges underline the importance of considering the economic state of households when assessing the prevalence of OOSC in Nigeria.

1.1.2 Conflict and displacement

Adeleke and Alabede (2022), in studying the spatial variation of OOSC in the different regions in Nigeria, conclude that the literature shows that northern geopolitical zones of Nigeria suffer the most from the problem of out-of- school children. An elaborate state-by-state study by UBEC (2018) showed 66% of all OOSC in Nigeria are in the North-East and North-West regions of the country.

1.2 Record Linkage in Capture-Recapture Data

Record linkage is necessary when a researcher wishes to combine data from two or more files that are believed to share common characteristics. This field has expanded across history, public health, and demography. There are two broad types: deterministic and probabilistic.

***1.2.1 Deterministic record linkage*:** This method uses a unique shared key (or keys). Records are considered matched if the linkage fields agree, and unmatched if they disagree. This method is common in research but assumes a known linking key exists. It produces only two mutually exclusive categories: "matched" and "unmatched" (Sayers *et al*., 2016). Asher *et al*. (2020) noted that deterministic linkage may be infeasible or only possible for a subset of records. Moreover, errors or missing values in a unique identifier can prevent some true matches from being linked. Deterministic linkage can lead either to too many invalid links (Type I error) or to true matches that remain unlinked (Type II error).

***1.2.2 Probabilistic record linkage*:** This method first requires that all matching fields be uniquely identifiable across files. After merging (joining), the agreement pattern between the two sets of keys is assessed. Probabilistic methods use one or more fields from each list and a probability model to determine the likelihood that two records refer to the same entity. For example, common names (e.g., Mohammed Ahmed, James Clement) yield a low likelihood that two records with those names belong to the same person; rare names (e.g., Kadaro Pascalis) greatly increase that likelihood. Combined with date of birth, a rare name can serve almost as a unique identifier (Asher *et al.*, 2020). Three underlying probabilistic models are summarized below.

1.2.3 Fellegi-Sunter method

Multiple fields are compared for each pair of records. A decision is made whether each pair of fields agrees (flexibly defined e.g., year of birth within one year). Two probabilities are estimated:

m probability: the fields agree given that the two records represent the same individual.

u probability: the fields agree given that the two records represent different individuals.

An agreement weight of $\log_2(m/u)$ is assigned if the fields agree, and a non-agreement weight of $\log_2((1-m)/(1-u))$ (always negative) if they disagree. The weights are summed to produce an overall match weight. *m* probabilities are typically close to 1; *u* probabilities vary (e.g., for gender, about 50% agreement by chance). For large datasets, a blocking strategy (e.g., using ZIP code and year of birth) is used to avoid comparing every possible pair. Multiple blocking passes may be applied (sequential or overlapping).

1.2.4 Bayesian record linkage

Bayesian techniques rely on prior probabilities of match/non-match based on expert opinion or previous studies. These priors are combined with new data to obtain posterior probabilities for each record pair. Bayesian approaches extend the Fellegi-Sunter algorithm by assuming that the *m* and *u* probabilities follow a probability distribution (e.g., uniform or Beta). The mean values from the posterior distributions are then used to compute match weights and complete the linkage.

1.3 Empirical review on Capture-recapture

Yadav (2019), used a two-sample capture-recapture method for ascertainment of fatal and non-fatal injuries reported from 1st January to 31st December 2017, in India. The first capture was data of injuries extracted from First Information Reports registered by the police. The study concluded that both records of police and government health facilities underestimated fatal and non-fatal injuries with under-reporting more pronounced in police records.

Koivuniemi *et al*. (2019), evaluated the use of photo-identification (photo-ID) and mark-recapture techniques for estimating the population size of the endangered Saimaa ringed seal (*Phoca hispida saimensis)*. The results indicated high survival rates and site fidelity among the adult seals.

Kevin, Reena and Hladik (2017), used capture-recapture techniques to estimate the number of men who have sex with men and among female sex workers in 11 selected towns in Uganda. The study utilized conventional 2-source capture-recapture (CRC) to estimate the population of female sex workers' men who have sex with men.

Waller *et al*. (2017), used multiple linked dataset and capture-recapture techniques to estimate rates of dementia among women in Australia.

Post *et al*. (2019) conducted a study using a single source of data, such as police records, or combining data from multiple sources results in an undercount of gun-related injuries.

Kreshpaj *et al*. (2021) used capture-recapture methods to estimate the magnitude of under-reporting of occupational injuries (OIs) among precarious and non-precarious workers in Sweden in 2013. The capture-recapture methods were applied using the national OIs register and records from a labour market insurance company. The study concluded that, OIs under-reporting may represent unrecognized injuries that especially burden precariously employed workers as financial, health and social consequences shift from the employer to the employee.

Guure *et al*. (2021) conducted a study to estimate the overall population of female sex workers (FSW) in all the 16 regions of Ghana using different Population size estimation (PSE) methods. Mapping of venues and complete enumeration of seaters was conducted at the formative stage prior to the bio-behavioral survey (BBS).

2. Material and Methods

Subject to closure, the simplest form of capture-recapture experiment, which is often called the Lincoln Index is the Petersen model which has a long history. A sample of n_1 animals is caught, marked and released. Later a sample of n_2 animals is captured of which m of them have been marked (Seber, 1973). Intuitively, one can derive an estimator of the population size (N) based on the notion that the ratio of marked to total animals in the sample should reflect the same ratio in the population so that;

$$\frac{m}{n_1} \approx \frac{n_2}{N} \quad (1)$$

which gives the intuitive estimator of the population size ($\widehat{N}$)

$$\widehat{N} = \frac{n_1 n_2}{m} \quad (2)$$

where:

$\widehat{N}$ = Number of individuals in the population in the giving location

n_1 = Number of individuals marked on the first occasion
n_2 = Number of individuals captured on the second occasion
m = Number of marked individuals

The Lincoln-Peterson estimator is asymptotically unbiased as sample size approaches infinity, but it is biased at small sample sizes, an alternative less biased estimator of population size is given by the Chapman estimator;

$$\widehat{N_c} = \frac{(n_1+1)(n_2+1)}{m+1} - 1 \quad (3)$$

2.1 The Estimation of Street Children in Taraba State

This descriptive study is an analytical study of street children (Almajiri's) in a group of urban communities which is still undergoing urbanization just like the study area Taraba State. A cluster or strata sample of street children (Almajiri's) as defined by the United Nations was used and considered while taking the data from three (3) LGAs (Jalingo, Gassol and Wukari), one (1) in each of the three (3) senatorial districts of the state. Five (5) wards were selected from each of the three (3) LGAs, using a purposive sampling plan. The Five (5) wards from each of the three (3) LGAs served as a trapping session or survey occasion or even as strata. A structured questionnaire was designed and used in the data collection; the questionnaire consists of information as; Names, Age, Village/LGA and State the child come from, Street Name, Name of Tsangaya/School, Time of the capture and recapture. Research assistants were independently hired to capture the data per ward. Data collection lasted one week; Monday was the first capture occasion and Wednesday the second. Recaptures were identified by cross-examining the two lists.

2.2 Method of Data Analysis

Degree of matching of capture-recapture data

Here multiple blocking passes was used;

1. Block A: Required that, all the Name of the child and name of school (Tsangaya)
2. Block B: Required that, all the Name of the child, state of origin
3. Block C: Required that, all the Name of the child and the age

An individual was accepted to be recapture if his identity matched any of the above blockings. In line with the suggestion given by IWGDWF (1995), the analysis was done based on the population of study. The population was estimated separately for each stratum and the estimates was added together to get an estimate of total population size. The strata consist of political wards from each of the three (3) LGAs, namely:

Jalingo LGA

1. Kona ward strata;
2. Barade ward strata;
3. Sintali ward strata;
4. Kachalla-Sembe ward strata;
5. Mayogwoi ward strata;

Gassol LGA

1. Sandirde ward strata;
2. Mutum biyu I ward strata;
3. Namnai ward strata;
4. Tutare ward strata;
5. Mutum biyu II ward strata;

Wukari LGA

1. Avyi ward strata;
2. Bantaje ward strata;
3. Hospital ward strata;
4. Kente ward strata;
5. Puje ward strata;

2.2.1 Two sample capture-recapture model employed

Four capture-recapture models were employed, the best fit was selected using, Akaike"s Information Criterion(J) (AIC_J) proposed by Jibasen (2011),

$$AIC_J = -\beta \sum\sum log_e \left(\frac{n_{i.} * n_{.j}}{\hat{N}^2}\right) + 2(C_0 + C_1 - 2)$$

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, $\hat{N}$ is estimated based on the given model, $n_{i.}$ and $n_{.j}$ are the marginals.

2.2.2 Behavioral Response Model (M_b)

This model applies when individuals become "elusive" (shy away) after being captured once which is analogous to avoiding re-arrest or running away from rehabilitation. Street children are naturally elusive, though incentives might make them "trap-happy". The estimator is:

$$\hat{N}_b = \frac{n_{1.}^{\ 2}}{n_{1.} - (n_{.1} - n_{11})} = \frac{n_{1.}^{\ 2}}{n_{.1} - n_{11}} \tag{4}$$

According Jibasen and Adams, (2013), this model is equivalent to the two-sample removal estimator given by Seber and Le Cren (1967).

2.2.3 Listability Model

The listability model as proposed by Jibasen (2011) for the elusive population is given by

$$\hat{N}_s = \frac{r}{\hat{p}_k} \tag{5}$$

Where, $\hat{p}k = \frac{r}{N}$,

$$r = n_1 + n_2 - n_{11}$$

It has been shown by Jibasen (2011)

$$\hat{p}_k = \frac{n}{sr} \tag{6}$$

Where; r = the number of different individuals listed (caught),

$\hat{p}$ = estimated capture probabilities

n = number of index on both lists or trapping sessions

s = number of systems

2.2.4 Non-factor Model (M_o)

This model assumes no time variation, no individual heterogeneity, and no behavioral response. All individuals are equally likely to appear on any list at any time. The estimator is:

$$\hat{N}_0 = {n^2}/{4n_{11}}, \tag{7}$$

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

See Jibasen and Adams, (2013) for details

2.2.5 The Petersen Model Estimator (M_s)

This model assumes all individuals have the same probability of being listed within a given system, but probabilities may vary between systems. The estimator is:

$$\hat{N}_s = \frac{n_{.1}\, n_{.1}}{n_{.11}} \tag{8}$$

3. Results and Discussion

3.1 Jalingo LGA

The following was the combined capture recaptured data collected on street children in the five (5) selected political wards out of the sixteen (16) wards of Jalingo LGA in Table 1. This presents ($n_{1.}$) = Number of Street Children Captured, ($n_{.1}$) = Number of Street Children on the second capture and n_{11} = Number of Street Children that Matched (recaptured) over the survey session in Kona, Barade, Sintali, Kachalla-Sembe and Mayogwoi wards, respectively.

Table 1: Data Presentation

Iteration	**The Selected Political Wards in Jalingo LGA**	**Number of Street Children Captured ($n_{1.}$)**	**Number of Street Children on the second capture ($n_{.1}$)**	**Number of Street Children Matched or Recaptured (n_{11})**
1	Kona Ward	56	50	19
2	Barade Ward	83	68	25
3	Sintali Ward	98	92	40
4	Kachalla-Sembe Ward	75	68	24
5	Mayogwoi Ward	67	62	18
	TOTAL	**379**	**340**	**126**

Table 1 presents data captured; $n_{1.}$ = (56, 83, 98, 75, 67), second capture ($n_{.1}$) = (50, 68, 92, 68, 62) and number of children who were found in both lists (recaptures), n_{11} = (19, 25, 40, 24, 18) in Kona, Barade, Sintali, Kachalla-sembe and Mayogwoi wards, respectively.

Table 2: Analysis for Kona ward

Models	$n_{1.}$	$n_{.1}$	n_{11}	N	AIC
Behavioral Response Model (M_b)	56	50	19	101	5.1229
Listability Model (M_a)	56	50	19	142	5.7148
No Factor Effect (M_o)	56	50	19	147	5.7749
Petersen (M_s)	56	50	19	146	5.7631

Table 2 shows the outcome of the comparison between the Behavioral model, No Factor-effect model, Listability Model and Petersen Model for the data collected from Kona ward. The result shows that the Behavioral Model appears to be more efficient with the smallest AIC value of 5.1229.

Table 3: Analysis for Barade Ward

Models	$n_{1.}$	$n_{.1}$	n_{11}	N	AIC
Behavioral Response Model (M_b)	83	68	25	160	5.9221
Listability Model (M_a)	83	68	25	210	6.3945
No Factor Effect (M_o)	83	68	25	228	6.5374
Petersen (M_s)	83	68	25	227	6.5297

Table 3 showed the comparison between Behavioral, Listability Model and Petersen Model for the data collected from Barade ward. The result indicates that Behavioral Response Model appears to be more efficient with the smallest AIC value of 5.9221.

Table 4: Analysis for Sintali ward

Models	$n_{1.}$	$n_{.1}$	n_{11}	N	AIC
Behavioral Response Model (M_b)	98	92	40	185	6.1743
Listability Model (M_a)	98	92	40	237	6.6046
No Factor Effect (M_o)	98	92	40	226	6.5221
Petersen (M_s)	98	92	40	225	6.5144

Table 4 above shows the comparison between Behavioral, Listability Model and Petersen Model for the data collected in Sintali ward. The result also indicates that Behavioral model appears to be more efficient with the smallest AIC value of 6.1743.

Table 5: Analysis for Kachalla-Sembe ward

Models	$n_{1.}$	$n_{.1}$	n_{11}	N	AIC
Behavioral Response Model (M_b)	75	68	24	128	5.5345
Listability Model (M_a)	75	68	24	198	6.2923
No Factor Effect (M_o)	75	68	24	213	6.4192
Petersen (M_s)	75	68	24	212	6.4110

Table 5 shows the comparison between Behavioral, Listability Model and Petersen Model for the data collected from Kachalla-Sembe ward. The result also indicates that Behavioral Response Model appears to be more efficient with the smallest AIC value of 5.5345.

Table 6: Analysis for Mayogwoi Ward

Models	$n_{1.}$	$n_{.1}$	n_{11}	N	AIC
Behavioral Response Model (M_b)	67	62	18	102	5.1400
Listability Model (M_a)	67	62	18	96	5.0347
No Factor Effect (M_o)	67	62	18	231	6.5601
Petersen (M_s)	67	62	18	230	6.5525

Table 6 shows the comparison between Behavioral, Listability Model and Petersen Model for the data collected from Mayogwoi ward. The result also indicates that Behavioral Response Model appears to be more efficient with the smallest AIC value of 5.1400.

3.2 Gassol LGA

The following was the combined capture recaptured data collected on street children in the five (5) selected political wards out of the sixteen (14) wards of Gassol LGA in Table 7. This presents ($n_{1.}$) = Number of Street Children Captured, ($n_{.1}$) = Number of Street Children on the second capture and n_{11} = Number of Street Children that Matched (recaptured) over the survey session in Sandirde, Mutum biyu I, Mutum biyu II, Namnai and Tutare wards, respectively.

Table 7: Data Presentation

Iteration	**The Selected Political Wards in Gassol LGA**	**Number of Street Children Captured ($n_{1.}$)**	**Number of Street Children on the second capture ($n_{.1}$)**	**Number of Street Children Matched or Recaptured (n_{11})**
1	Sandirde Ward	75	65	30
2	Mutum biyu I Ward	68	63	25
3	Mutum biyu II Ward	55	52	20
4	Namnai Ward	43	42	14
5	Tutare Ward	50	46	18
	TOTAL	**291**	**268**	**107**

Table 7 presents data captured; $n_{1.}$ = (75, 68, 55, 43, 50), second capture ($n_{.1}$) = (65, 63, 52, 42, 46) and number of children who were found in both lists (recaptures), n_{11} = (30, 25, 20, 14, 18) in Sandirde, Mutum biyu I, Mutum biyu II, Namnai and Tutare wards, respectively.

Table 8: Analysis for Sandirde ward

Models	$n_{1.}$	$n_{.1}$	n_{11}	N	AIC
Behavioral Response Model (M_b)	75	65	30	161	5.9329
Listability Model (M_a)	75	65	30	173	6.0578
No Factor Effect (M_o)	75	65	30	163	5.9544
Petersen (M_s)	75	65	30	161	5.9322

Table 8 shows the outcome of the comparison between the Behavioral model, No Factor-effect model, Listability Model and Petersen Model for the data collected from Sandirde ward. The result shows that the Petersen model appears to be more efficient with the smallest AIC value of 5.9322.

Table 9: Analysis for Mutum Biyu I Ward

Models	$n_{1.}$	$n_{.1}$	n_{11}	N	AIC
Behavioral Response Model (M_b)	68	63	25	122	5.4511
Listability Model (M_a)	68	63	25	172	6.0477
No Factor Effect (M_o)	68	63	25	171	6.0376
Petersen (M_s)	68	63	25	172	6.0477

Table 9 showed the comparison between Behavioral, Listability Model and Petersen Model for the data collected from Mutum Biyu I ward. The result indicates that Behavioral Response Model appears to be more efficient with the smallest AIC value of 5.4511.

Table 10: Analysis for Mutum Biyu II ward

Models	$n_{1.}$	$n_{.1}$	n_{11}	N	AIC
Behavioral Response Model (M_b)	55	52	20	95	5.0165
Listability Model (M_a)	55	52	20	141	5.7025
No Factor Effect (M_o)	55	52	20	143	5.7270
Petersen (M_s)	55	52	20	151	5.8215

Table 10 above shows the comparison between Behavioral, Listability Model and Petersen Model for the data collected in Mutum Biyu II ward. The result also indicates that Behavioral model appears to be more efficient with the smallest AIC value of 5.0165.

Table 11: Analysis for Namnai ward

Models	$n_{1.}$	$n_{.1}$	n_{11}	N	AIC
Behavioral Response Model (M_b)	43	42	14	66	4.3838
Listability Model (M_a)	43	42	14	119	5.4078
No Factor Effect (M_o)	43	42	14	129	5.5480
Petersen (M_s)	43	42	14	129	5.5480

Table 11 shows the comparison between Behavioral, Listability Model and Petersen Model for the data collected from Namnai ward. The result also indicates that Behavioral Response Model appears to be more efficient with the smallest AIC value of 4.3838.

Table 12: Analysis for Tutare Ward

Models	$n_{1.}$	$n_{.1}$	n_{11}	N	AIC
Behavioral Response Model (M_b)	50	46	18	89	4.9032
Listability Model (M_a)	50	46	18	127	5.5208
No Factor Effect (M_o)	50	46	18	128	5.5345
Petersen (M_s)	50	46	18	127	5.5208

Table 12 shows the comparison between Behavioral, Listability Model and Petersen Model for the data collected from Tutare ward. The result also indicates that Behavioral Response Model appears to be more efficient with the smallest AIC value of 4.9032.

3.3 Wukari LGA

The following was the combined capture recaptured data collected on street children in the five (5) selected political wards out of the sixteen (16) wards of Jalingo LGA in Table 13. This presents ($n_{1.}$) = Number of Street Children Captured, ($n_{.1}$) = Number of Street Children on the second capture and n_{11} = Number of Street Children that Matched (recaptured) over the survey session in Avyi, Bantaje, Hospital, Kente and Puje wards, respectively.

Table 13: Data Presentation

Iteration	The Selected Political Wards in Wukari LGA	Number of Street Children Captured ($n_{1.}$)	Number of Street Children on the second capture ($n_{.1}$)	Number of Street Children Matched or Recaptured (n_{11})
1	Avyi Ward	91	61	28
2	Bantaje Ward	100	80	43
3	Hospital Ward	120	78	50
4	Kente Ward	79	70	23
5	Puje Ward	87	84	21
	TOTAL	**379**	**340**	**126**

Table 13 presents data captured; $n_{1.}$ = (91, 100, 120, 79, 87), second capture ($n_{.1}$) = (61, 80, 78, 70, 84) and number of children who were found in both lists (recaptures), n_{11} = (28, 43, 50, 23, 21) in Avyi, Bantaje, Hospital, Kente and Puje wards, respectively.

Table 14: Analysis for Avyi ward

Models	$n_{1.}$	$n_{.1}$	n_{11}	N	AIC
Behavioral Response Model (M_b)	91	61	28	250	6.6974
Listability Model (M_a)	91	61	28	202	6.3526
No Factor Effect (M_o)	91	61	28	206	6.3611
Petersen (M_s)	91	61	28	198	6.2922

Table 14 shows the outcome of the comparison between the Behavioral model, No Factor-effect model, Listability Model and Petersen Model for the data collected from Avyi ward. The result shows that the Petersen model appears to be more efficient with the smallest AIC value of 6.2922.

Table 15: Analysis for Bantaje Ward

Models	$n_{1.}$	$n_{.1}$	n_{11}	N	AIC
Behavioral Response Model (M_b)	100	80	43	270	6.8311
Listability Model (M_a)	100	80	43	206	6.3611
No Factor Effect (M_o)	100	80	43	188	6.2023
Petersen (M_s)	100	80	43	186	6.1837

Table 15 showed the comparison between Behavioral, Listability Model and Petersen Model for the data collected from Bantaje ward. The result indicates that Petersen Model appears to be more efficient with the smallest AIC value of 6.1837.

Table 16: Analysis for Hospital ward

Models	$n_{1.}$	$n_{.1}$	n_{11}	N	AIC
Behavioral Response Model (M_b)	120	78	50	514	7.9495
Listability Model (M_a)	120	78	50	221	6.4832
No Factor Effect (M_o)	120	78	50	196	6.2747
Petersen (M_s)	120	78	50	187	6.1930

Table 16 above shows the comparison between Behavioral, Listability Model and Petersen Model for the data collected in Hospital ward. The result also indicates that Petersen model appears to be more efficient with the smallest AIC value of 6.1930.

Table 17: Analysis for Kente ward

Models	$n_{1.}$	$n_{.1}$	n_{11}	N	AIC
Behavioral Response Model (M_b)	79	70	23	133	5.6010
Listability Model (M_a)	79	70	23	213	6.4192
No Factor Effect (M_o)	79	70	23	241	6.6337
Petersen (M_s)	79	70	23	240	6.6265

Table 17 shows the comparison between Behavioral, Listability Model and Petersen Model for the data collected from Kente ward. The result also indicates that Behavioral Response Model appears to be more efficient with the smallest AIC value of 5.6010.

Table 18: Analysis for Puje Ward

Models	$n_{1.}$	$n_{.1}$	n_{11}	N	AIC
Behavioral Response Model (M_b)	87	84	21	120	5.4224
Listability Model (M_a)	87	84	21	263	6.7855
No Factor Effect (M_o)	87	84	21	348	7.2720
Petersen (M_s)	87	84	21	348	7.2720

Table 18 shows the comparison between Behavioral, Listability Model and Petersen Model for the data collected from Kente ward. The result also indicates that Behavioral Response Model appears to be more efficient with the smallest AIC value of 5.4224.

3.4 The Estimated Summary according to each LGA

3.4.1 Jalingo LGA

Table 19 presents the summary of the estimated N obtained in each ward using the models with the least AIC, this is presented in Table 19, where for Kona ward, $N = 101$, Barade ward, $N = 160$, Sintali ward, $N = 185$, Kachalla-Sembe ward, $N = 128$ and Mayogwoi ward, $N = 102$.The estimated total number of street children in Jalingo LGA is 676.

Table 19: Summary of results based on ward

Strata	**Adopted Models**	**Estimated *N* per Strata**
Kona Ward	Behavioral (M_b)	101
Barade Ward	Behavioral (M_b)	160
Sintali Ward	Behavioral (M_b)	185
Kachalla-Sembe Ward	Behavioral (M_b)	128
Mayogwoi Ward	Behavioral (M_b)	102
	TOTAL	**676**

From table 19, it shows that Behavioral Model (M_b) performs better than other models in all the selected wards of Jalingo LGA.

3.4.2 Gassol LGA

This presents the summary of the estimated N obtained in each ward using the models with the least AIC, this is presented in Table 20, where for Sandirde ward, $N = 161$, Mutum biyu I ward, $N = 122$, Mutum biyu II ward, $N = 95$, Namnai ward, $N = 66$ and Tutare ward, $N = 189$. The estimated total number of street children in Gassol ward is 533.

Table 20: Summary of results based on the strata

Strata	**Adopted Models**	**Estimated *N* per Strata**
Sandirde Ward	Petersen (M_s)	161
Mutum biyu I Ward	Behavioral (M_b)	122
Mutum biyu II Ward	Behavioral (M_b)	95
Namnai Ward	Behavioral (M_b)	66
Tutare Ward	Behavioral (M_b)	89
	TOTAL	**533**

From table 20, it shows that Petersen Model (M_s) performs better than other models in Sandirde ward, while Behavioral Model (M_b) performs better than other models in Mutum biyu I, Mutum biyu II, Namnai and Tutare wards of Gassol LGA respectively.

3.4.3 Wukari LGA

This presents the summary of the estimated N obtained in each strata using the models with the least AIC, this is presented in Table 21, where for Avyi ward, $N = 198$, Bantaje ward, $N = 186$, Hospital, $N = 187$, Kente ward, $N = 133$ and Puje ward, $N = 120$. The estimated total number of street children in Wukari LGA is 824.

Table 21: Summary of results based on the strata

Strata	Adopted Models	Estimated *N* per Strata
Avyi Ward	Petersen (M_s)	198
Bantaje Ward	Petersen (M_s)	186
Hospital Ward	Petersen (M_s)	187
Kente Ward	Behavioral (M_b)	133
Puje Ward	Behavioral (M_b)	120
	TOTAL	**824**

From table 21, it shows that Petersen Model (M_s) performs better than other models in Avyi, Bantaje and Hospital wards of Wukari LGA, while Behavioral Model (M_b)performs better than other models in Kente and Puje wards of Wukari LGA respectively.

4. Conclusion

Street children are a common and underestimated phenomenon in Taraba State particularly in the selected LGAs, sharing characteristics with street children elsewhere in Nigeria (especially the northern part of Nigeria) and in other developing countries. Capture-recapture appears to be a suitable method for studying such elusive populations.

Acknowledgement
The authors wish to thank TETfund for sponsoring this research.

References

Adebowale, S. A. (2018). Dynamics of child marriage and marital timing in Nigeria: A retrogression or progression? Health Care for Women International, 39(9), 975–993. https:// doi. org/ 10. 1080/ 07399 332. 2018. 1490742

Adegbami, A., & Adesanmi, O. (2018). Nigeria's educational system and sustainable development goals' attainment: A wild goose chase. *Journal of Contemporary Asia*, *5*(1), 217–233.

Adeleke, R., & Alabede, O. (2022). Geographical determinants and hotspots of out-of- School children in Nigeria. *Open Education Studies*, *4*(1), 345–355.

Asher, J., Resnick, D., Brite, B., Brackbill, R., and Cone, J. (2020). An Introduction to Probabilistic Record Linkage with a Focus on Linkage Processing for WTC Registries. *Int. J. Environ. Res. Public Health* 2020, *17*(18), 6937; https://doi.org/10.3390/ijerph17186937

Evans, D., & Acosta, A. (2020). Education in Africa: What are we learning? *Journal of African Economies*, *30*(2),13–54.

Farswan, D. S. (2023). Role of education and culture in social development. *Journal of Social Review and Development*, *2*(1), 13–16.

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.

International Working Group for Disease Monitoring and Forecasting (IWGDMF) (1995). 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, Nigeria.

Jibasen, D. and Adams, Y. J. (2013). An efficient two sample capture-recapture model with high recaptures. *CBN Journal of Applied Statistics*, ISSN 2476-8472, The Central Bank of Nigeria, Abuja, 4(2); 141-158.

Kevin, A., Reena, H. D., & 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.*

Koivuniemi, M., Kurkilahti, M., Niemi, M., Auttila, M., & 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.

Nigeria Digest of Education Statistics (2022). Federal Ministry of Education.

Olonade, O. Y., George, T. O., Imhonopi, D., Egharevba, M. E., & Kasa, A. G. (2022). Youths' socio-economic well-being in Southwest Nigeria: What role can empowerment/poverty reduction programmes play? *Cogent Social Science*, *8*(1), 2115694.

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.

Post, L.A., Balsen, Z., Spano, R., & Vaca, F.E. (2019). Bolstering gun injury surveillance accuracy using capture–recapture methods. *J Behav Med*, 42(4),674–80.

Sayers, A. Ben-Shlomo, Y., Blom, A. W., and Steele F. (2016). Probabilistic record linkage. *International Journal of Epidemiology*, 45(3). 954–964 doi: 10.1093/ije/dyv322 Advance Access Publication

Statista Search Department. (2023, September 22nd). *Median age in Africa 2023, by country*.

Tule, K. M., Egbuna, E. N., Dada, E., Ebuh, G. U., & McMillan, D. (2017). A dynamic fragmentation of the misery index in Nigeria. *Cogent Economics & Finance*, *5*(1), 1–13.

UBEC Personnel Audit and Digest of Basic Education Statistics in Nigeria. (2018). *National personnel report(NPA)*.

UNICEF (2016). Situation of Women and Children in Nigeria,

UNDP. (2016). Goal 4 |Department of Economic and Social Affairs.

UNESCO. (2018). Combining data on Out-of-school children, completion and learning to offer a more comprehensive view on SDG 4 (Information Paper No. 61). UNESCO Institute for Statistics.

UNESCO Institute for Statistics (UIS). (2022). Out-of-schoolchildren, adolescents, and youth.

UNESCO Institute for Statistics, Global Education Monitoring Report Team & World Bank. (2023). Education Finance Watch.

UNESCO Institute of Statistics. (2022). 244M children won't start the new school year (UNESCO)–World.

UNESCO. (2022). New estimation confirms out-of-school population is growing in sub-Saharan Africa. UNESCO Digital Library Policy Paper.

UNESCO. (2023). Out-of-school numbers are growing in sub-Saharan Africa Global Education Monitoring Report.

UNESCO. (2021). Setting commitments: National SDG 4 benchmarks to transform education Global Education Monitoring Report.

UNICEF. (2022a). Educational opportunities for out of school children (OOSC).

Yadav, S.S. (2019). How safe are industries in India? Ascertaining industrial injuries in Dadra and Nagar Haveli, India by capture-recapture method. *Indian Journal of Occupational and Environmental Medicine*, 23(1), 15-20.

Wesson, P., Jewell, N.P., McFarland, W.M.D., & 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.