Received: 27 April 2022 | Revised: 20 May 2022 | Accepted: 28 June 2022 | Published online: 30 June 2022

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2023, Vol. 2(3) 249–259
DOI: 10.47852/bonviewJCCE2202226

A Model for Estimating the Population Size of Disproportionate Two Sample Capture Recapture Methods

B. Z. Reuben[1,*], D. Jibasen[1], I. J. Dike[1] and S. S. Abdulkadir[1]

[1]*Modibbo Adama University, Nigeria*

Abstract: The size of a population $\widehat{N}$ in a dynamic setting can be estimated using closed population capture-recapture techniques. This entails drawing a sample from a sampling frame denoted by $n_{1.}$, mark and return into the population. Thereafter, another sample is drawn independently and denoted by $n_{.1}$ items or individuals selected in both samples are recorded and denoted by n_{11}. The sizes of samples drawn are not necessarily at bay. Supplementary Immunization Activities (SIAs) are vaccination campaigns conducted at National and Subnational scale to boost immunity against Vaccine Preventable Diseases (VPD) such as polio and Lot Quality Assurance Survey (LQAS) was introduced to assess the coverage of the immunization activities with the view to get quick idea of what the coverage is so as to decide whether to accept or reject the lot(s) on the basis of a predetermine number of acceptable defects (unvaccinated). Since the size of the first sample is inordinately different to the second sample in LQAS, a disproportionate capture-recapture (C-R) model was developed to address the disparity, estimate SIAs coverage and enhance precision of estimate classification.

Keywords: supplementary immunization activities, lot quality assurance survey, capture-recapture (C-R) models, disproportionality, global polio eradication initiative

1. Introduction

Two sample capture-recapture (C-R) techniques are used to evaluate and approximate the size of a population using individuals or items selected from two samples of almost equal magnitude. A special type of two sample C-R methods with disproportionality between the sizes of the two samples is obtained as exemplified in the biggest internationally organized public health project, the Global Polio eradication Initiative (GPEI) which is lunched in 1988 (Tebbens et al., 2010). The improvement in the quality of life from the paralysis and deaths that would be avoided and money that would have been used for procurement of vaccines deployed to other areas of lives was the anticipated relevance of the GPEI (Thompson et al., 2006). Sangrujee et al. (2004) estimated the gains accruable from post-polio certification and found out that the budgets on the vaccine per dose and targeted number factors of the global costs of post certification polio vaccination.

When National Immunization Plus Days are conducted, primary supervision is vital for problem identification in planning and reporting, observing particular areas of inadequate coverage, and forecasting the tendency of the spread of disease. The use of supervisory tools before, during and after National Immunization Plus Days (NIPD) is hampered because entrenching and adopting reliable techniques to generate data can be improved in the buildup and implementation of humongous national activities in addition to unreliable regular monitoring data which show appreciable level of coverage in almost all areas, including those with current cases of virus circulation. To mitigate this problem, LQAS was adopted because data collected is quickly and easily interpreted (Brown et al., 2014).

LQAS which was initially developed for industrial quality control, has been applied to health surveys. It is a quick sampling technique deployed to evaluate the viability of immunization coverage sequel to SIAs in a settlement that has been determined ahead of time using small sample size. It is particularly used in areas with risk or polio dominated area to execute corrective action such as mop-up in areas identified to be weak in coverage (Manual, 2012; Jutand & Salamon, 2000; Olives, 2011). With a small investment, this method enables programme directors to identify rapidly the health facilities with below standard services and therefore requiring special attention (Valadez, 1991).

As a statistically dependable tool for supervising polio immunization, LQAS has proven its relevance in determining campaign quality. Information supplied by LQAS in determining the quality of SIAs coverage has helped to distinguish the areas that seriously in need of intervention. More so, the capacity of

*Corresponding author: B. Z. Reuben, Modibbo Adama University, Nigeria. Email: zangaluka@gmail.com

GPEI to track the patterns of immunization quality overtime can be hinged on LQAS, an improvement on SIAs pre-implementation of LQAS (Brown et al., 2014).

Some of the limitations of CLQAS as enunciated in (Manual, 2012) are: (i) if the lot is too enormous and heterogenous in coverage, LQAS may not give the coverage of the entirety of the lot, thereby diffusing its reliability across the entire lot (e.g. ward); (ii) LQAS doesn't give us the point estimate of the coverage but classification of SIA coverage; and (iii) there is a tendency for misclassification due to the relatively small sample size and clustering approach. In fact, statistical error can be very high in lot where coverage varies greatly between clusters.

Capture-Recapture (C-R) techniques are used for assessing the size of a population based on ratio of tagged to untagged individual (Amstrup & Mcdonald, 2010). Mingoti and Caiaffa (2006), noted that C-R can be used to estimate the size of unknown finite population size. Pollock (1981), affirmed that the population under review can be sampled more than a time. At every occasion, every untagged individual caught is specially tagged; previously tagged individuals have their capture history recorded and returned into the population. Therefore, by the time the study is concluded, the researcher has the comprehensive history of each individual handled.

In using LQAS, subsamples of the population (lots) are either accepted or rejected based on the number of defects in a random sample (N) of a given lot. Should the number of defects is higher than decision value (d), the lot is rejected and remedial measures recommended in the lot; should the number of defects is equal or less than d, the lot is accepted (Pezzoli & Kim, 2013). Since the variableness in the percentage of children immunized among clusters within a lot has remarkable relevance on the remarkable impact on the coverage estimates, the probability of error is increased by high variability thereby weakening the strength of the "pass/fail" determination (Okayasu et al., 2014). To this end, overestimation of immunization coverage may leave populations at risk, whilst underestimation can lead to unnecessary catch-up campaigns (Alberti et al., 2008). Additionally, there were instances where the Wild Polio Virus (WPV) were recorded in lots where the LQAS coverage were estimated to have been high. As a result of the high estimate, program activities were relapsed and the virus was spreading. To this effect, reliance on LQAS coverage alone might not in all cases give a true reflection of the reality on ground, hence the need to incorporate estimated population size. Intrinsic heterogeneity in C-R techniques is reduced by stratification (Sutherland & Schwartz, 2005). Stratification in LQAS would address the effect of intrinsic heterogeneity in C-R techniques and the estimated population size would address the frailty of LQAS due to small sample size. Therefore, C-R techniques in alliance with CLQAS (where lots are classified into clusters) would provide a more precise insight of the coverage estimate, hence the need for this study.

This work is aimed at developing an effective model that could be used to estimate SIAs coverage with a view to enhance the precision of the estimate classification by developing a model for disproportionate two sample capture recapture.

2. Literature Review

2.1. Immunization coverage surveys

Siddiqi *et. al.* (2021) studied how powerful, vaccine data are generated via Electronic Immunization Registers (EIR) are used to supervise vaccination workers and ensuring that remedial measures are strategically targeted at communities identified as chronically missed. They suggested the importance of generating and using quality data for evidence-based decision making to overcome the obstacles inherent in immunization system in order to attain the Sustainable Development Goal (SDGs) of ensuring healthy lives and well-being for all persons at all ages, especially for newborn and children under the age of 5.

Abbott et al. (2021) showed how Measurement and Improvement (M&I) strategy has helped to mitigate the variableness across Immunization Information System and strengthen immunization data in Immunization Information System (IIS) which is more comprehensive, reliable and can be used with certain degree of certainty that is of particular relevance in actualizing Sustainable Development Goal targeted at enhancing healthy lives and elevating well-being for all age groups via robust immunization system.

Pincipal target of immunization programmes is to assuage the frequency of occurrence of vaccine preventable of diseases (VPDs) by reaching high levels of routine immunization coverage with viable vaccines (Uwaibi & Omozuwa, 2020). As one of the countries accounting for 62% of under and unvaccinated children worldwide, Nigeria needs to strengthen its immunization system (Olaniyan et al., 2021). The major attention of the health-related SDGs number three is universal health coverage (UHC), encompassing access to secured, robust, excellent, and affordable essential medicines and vaccines. However, the problems to realizing UHC are enormous, particularly with increases reliance on the health sector whose budget is either stagnant or plummeting (Chopra et al., 2020).

2.2. Underlying assumptions

The assumptions required for incorporating the number of units selected in both samples and the number of units selected in just one sample to evaluate the number of units not selected in both samples, therefore, providing estimate of the entire population magnitude, N, can be itemized in a number of ways, but the substratum is unpacked as following (International Working Group for Disease Monitoring and Forecasting, 1995):

i. Closure: the population under study is closed that is, the population is unaffected by change in birth, death or migration during the study period.
ii. Perfect matching: subjects captured in one sampling unit can be precisely paired to another sampling unit with no variation (no unpair, no loss of tag, etc.).
iii. Homogeneity: within each source, all subjects have equal chance of being selected (that is the "catchability" is equal for all subjects).
iv. Independence: the two sources are independent, that is, the likelihood of a subject being selected in one sampling unit is independent of subject selected from other sampling units whether the unit was captured in the other source.

2.3. Capture recapture techniques

To improve the Petersen estimates when heterogeneity is deemed to affect the estimates, Sekar and Deming (1949) employed stratification to evaluate the rate of birth and death using two lists. Pollock (1976) explicitly highlighted a step by step approach to building models and the importance of assumptions in the building of such models. He used the trap response models to accentuate the problem of non-identifiability of the parameter N. He found that N could not be estimated unless

the often-unrealistic assumption that the probability of capture of the unmarked animal is constant for both samples.

Manning and Goldberg (2010) designed a method to build spatial explicit capture-recapture selection histories from sites of untagged species for evaluating population magnitude with conventional C-R techniques. They applied the technique to data from point coordinate capture-recapture sampling method for more species with the probability of detecting error.

Focusing on the model formulation rather than on the estimation methods (which include inverse prediction, maximum likelihood and Bayesian methods) in a non-technical way, omitting much of the algebraic detail, Borchers (2012) reviewed capture recapture models that do include an explicit spatial component. He observed in an attempt to synthesize these models, that starting with circular plot survey models and moving through conventional distance sampling models, with and without measurement errors, through mark-recapture distance sampling (MRDS) model; concluded that spatial explicit capture-recapture (SCR) models can be viewed as an endpoint of a series of spatial sampling models.

Jibasen et al. (2012) presented a robust capture-recapture model for estimating the size of elusive epidemiologic events. They compared a proposed estimator $\widehat{N}_c$ the Petersen estimator $\widehat{N}_s$ and another estimator $\widehat{N}_0$ using the Akaike Information Criterion (AIC) and the Mean Absolute Deviation (MAD) through simulation studies. The study shows that both AIC and MAD revealed that $\widehat{N}_c$ is a better and robust estimator. The research further discovered that $\widehat{N}_s$ under estimates the total elusive population N, $\widehat{N}_0$ over estimates N while $\widehat{N}_c$ was always consistent and performs better than the other two and hence, recommended that the proposed estimator $\widehat{N}_c$ be used for estimating dual system elusive events.

Jibasen and Adams (2013) proposed an efficient two sample capture-recapture model (M_a) with high recaptures and compared it with the existing models such as the model of no factor effect (M_o), behavioral response model (M_b) and the Petersen model (M_s), using simulated data. They found that the proposed model provides a better estimator of the population size than the existing ones when the recapture is high, that is, in situations where individuals respond positively to capture, and also found that the Petersen model provides a better estimate of the population size when the observations follow a hyper-geometric distribution.

Sekar and Deming (1949), Ahlo (1990), Chao et al. (2008); Chao et al. (2001), and Royle and Converse (2014) use stratification to address the biasness in Petersen estimator under population heterogeneity. Manning and Goldberg (2010); and Borchers (2012) reviewed C-R models that include explicit spatial components for estimating population size. Chao et al. (2001), Chao et al. (2008); Pollock (1976), and Clavel et al. (2008) presented an intuitive interpretation for independence between capture sample and recapture sample while "trap happy" and "trap shy" were buttressed with explicit exposition on elusive event by Jibasen *et al.* (2012).

Apart from Okayasu *et al.* (2014) who conducted a pilot evaluation in four LGAs in Nigeria with an expanded LQAS sample size 16 clusters instead of the standard 6 clusters of 10 subjects each and found out that improvement in precision was deemed insufficient to warrant the effort, most literatures reviewed on LQAS were more emphatic on its application rather than its formation. They also noted that since variability in the proportion of children vaccinated among clusters within a lot has a remarkable impact on the coverage estimates, the probability of error is increased by high variability thereby compromising the robustness of the "pass/fail" determination. This may lead to overestimation of vaccination coverage which may leave populations at risk or underestimation which can lead to unnecessary catch-up campaigns (Alberti et al., 2008).

This work proposes C-R models that takes into consideration the disproportionality between the first and the second sample sizes in two sample capture recaptures.

3. Methodology

3.1. Model direction

This work focused on two-sample C-R model where the first sample typically is the enumeration of the target population (i.e. children under the age of 5 years that were immunized during SIAs) while the second sample is a small fraction of the first. During SIAs, vaccination teams move from house-to-house, immunizing children under the age of 5 years and finger marking them as indication that they have been immunized. Two days after the immunization campaign, independent surveyors are deployed to take sample of 60 eligible children from selected lots (wards) and coverage based on the principle of LQAS is reached by considering children that have been finger marked vis-à-vis those not finger marked. This is a topology of two-sample C-R technique. While the usual two-sample C-R methods considered two independent samples of almost the same size, this research is looking at a situation where the two samples sizes are greatly disproportionate.

3.2. The proposed model (M_p)

Disproportionate two sample Capture Recapture, was derived from the general (unrestricted) two sample capture recapture model. The general two-sample C-R model is given as:

$$P(n_{.1}, n_{1.}, n_{11}) = \binom{N}{n_{1.}}\binom{n_{\cdot 1}}{n_{11}}\binom{N-n_{1.}}{n_{.1}-n_{11}} P_{1.}^{n_{1.}}(1-P_{1.})^{N-n_{1.}} *$$

$$C^{n_{11}}(1-C)^{n_{1.}-n_{11}} P_{.1}^{n_{.1}-n_{11}}(1-P_{.1})^{N-n_{1.}-n_{.1}+n_{11}} \quad (1)$$

Where,
$n_{1.}$ = number of captures in the first sample
$n_{.1}$ = number of captures in the second sample
n_{11} = number of captures in both sample
$P_{1.}$ = Captures probability in the first sample
$P_{.1}$ = Captures probability in the second sample
C = Capture probability in both samples

Also note that robustness when conditions are altered by relapsing and constraining at least a parameter results in the following: when,

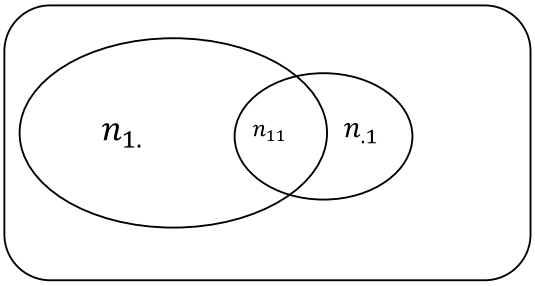
i. $P_{1.} = C$, while $P_{.1}$ is unaffected, Petersen Model M_s ensue.
ii. $P_{.1} = C$, while $P_{1.}$ is unaffected, Effective Model for High Recaptures M_a ensue.
iii. $P_{.1} = P_{1.}$, while C is unaffected, Behavioral Model M_b ensue.
iv. $P_{.1} = P_{1.} = C$, No Effect Model (Restricted Model) M_o ensue.

In SIAs, the number of children immunized and the number of children sampled during LQAS represent the first and second sample respectively.

Since the number of children sampled during LQAS is a small proportion of number of children immunized during House to House campaign, it is important that a suitable CR model be developed to address the issue of gross disproportionality between the first

and the second sample as shown in Figure 1. $f(n_{.1})$ as depicted in equation (5) is introduced as a replacement to the second sample distribution into the general model to curb the parity between the two samples.

Figure 1
Venn diagram showing disproportionate two sample capture recapture

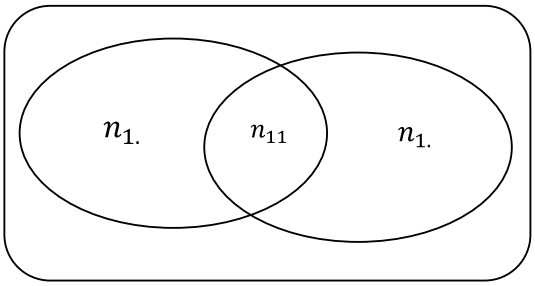


3.3. Assumptions of the proposed C-R models

Assumptions for the proposed dipropionate capture recapture model are as follows:

i. Let $n_{1.}$ and $n_{.1}$ be the observed sample sizes of the first and second samples respectively as depicted in Figure 2.

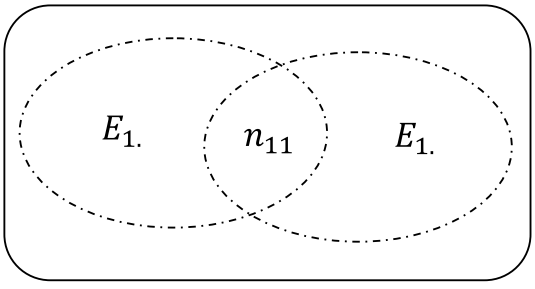
Figure 2
Venn diagram showing first sample and second sample observed



ii. Let $E_{1.}$ and $E_{.1}$ be the expected sample sizes of the first and second samples respectively.

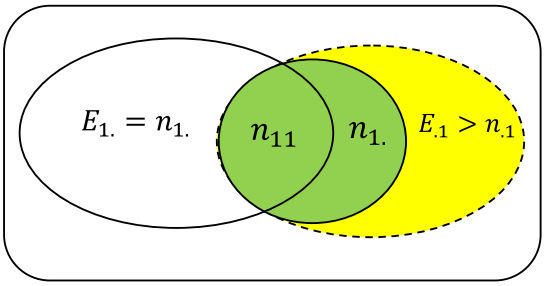
iii. The expected sample sizes of both samples are approximately the equal ($E_{1.} = E_{1.}$) as shown in Figure 3.

Figure 3
Venn diagram showing expected sample sizes of the first and second samples



The expected and observed sample sizes of the first sample are approximately the same ($E_{1.} = n_{1.}$) as demonstrated in Figure 4.

Figure 4
Venn diagram showing equal observed and expected sizes of the first sample



iv. The subset and superset of the second sample have the same interception with the first sample.

v. Let:

Combination of the observed subset

$$Cm_{sub1} = \binom{n_{1.}}{n_{.1}} \tag{2}$$

Combination of the unobserved subset

$$Cm_{sub2} = \binom{N - 2n_{1.}}{n_{1.} - n_{.1} - n_{11}} \tag{3}$$

Combination of the expected superset

$$Cm_{sup} = \binom{N - n_{1.}}{n_{1.} - n_{11}} \tag{4}$$

Let $P_{sup} = \binom{N - n_{1.}}{n_{.1} - n_{11}} P_{.1}^{n_{.1} - n_{11}} (1 - P_{.1})^{N - n_{1.} - n_{.1} + n_{11}}$ be the probability of the expected superset of the second sample from whence the undercount could be adjusted. By merging equations (2), (3) and (4), we get a hypergeometric probability, denoted by $f(n_{.1})$, thus:

$$f(n_{.1}) = \frac{\binom{n_{1.}}{n_{.1}}\binom{N - 2n_{1.}}{n_{1.} - n_{.1} - n_{11}}}{\binom{N - n_{1.}}{n_{1.} - n_{11}}} \tag{5}$$

Therefore, equation (1) becomes:

$$P(n_{.1}, n_{1.}, n_{11}) = \binom{N}{n_{1.}} P_{1.}^{n_{1.}} (1 - P_{1.})^{N - n_{1.}} \binom{n_{.1}}{n_{11}} C^{n_{11}} (1 - C)^{n_{1.} - n_{11}} * \frac{\binom{n_{1.}}{n_{.1}}\binom{N - 2n_{1.}}{n_{1.} - n_{.1} - n_{11}}}{\binom{N - n_{1.}}{n_{1.} - n_{11}}} \tag{6}$$

As the binomial distribution of the second sample which is written thus:

$$\binom{N-n_{1.}}{n_{.1}-n_{11}}P_{.1}^{n_{.1}-n_{11}}(1-P_{.1})^{N-n_{1.}-n_{.1}+n_{11}}$$

becomes a hypergeometric distribution:

$$\frac{\binom{n_{1.}}{n_{.1}}\binom{N-2n_{1.}}{n_{1.}-n_{.1}-n_{11}}}{\binom{N-n_{1.}}{n_{1.}-n_{11}}}$$

Using maximum likelihood estimation method, equation (6) yields maximum likelihood estimator (MLE) as:

$$\hat{P}_{1.}=\frac{n_{1.}}{\widehat{N}} \tag{7}$$

$$\hat{C}=\frac{n_{11}}{n_{1.}} \tag{8}$$

$$\frac{L(N)}{L(N-1)}=1 \Rightarrow \frac{\widehat{N}}{\widehat{N}-n_{1.}}*\frac{\widehat{N}-2n_{1.}}{\widehat{N}-3n_{1.}+n_{.1}+n_{11}}*\frac{\widehat{N}-2n_{1.}+n_{11}}{\widehat{N}-n_{1.}}*(1-\hat{P}_{1.})=1 \tag{9}$$

Substituting equation (8) in equation (9) gives us the appropriate Maximum Likelihood estimator of N, thus:

$$\widehat{N}_p=\frac{n_{1.}(n_{1.}+n_{.1}-n_{11})}{n_{.1}} \tag{10}$$

Using Delta Method as expounded by Jibasen (2011), where the variance $Vf(x)$ of a function of x is estimated as:

$$Varf(x)=\left(\frac{\partial f}{\partial x}\right)_E^2 var(x)$$

Such that $(\,)_E$ represent replacement of the expected value for x in the differentiation of the bracket while $var(x)$ represents the variance of x.

Therefore, the variance of M_P is given by:

$$var\left(\widehat{N}_P\right)=var\left(\frac{n_{1.}(n_{1.}+n_{.1}-n_{11})}{n_{.1}}\right)$$

$$var\left(\widehat{N}_P\right)=var\left(\frac{(n_{1.}^2+n_{1.}n_{.1})}{n_{.1}}-\frac{n_{1.}n_{11}}{n_{.1}}\right)$$

$$var\left(\widehat{N}_P\right)=var\left(-\frac{n_{1.}n_{11}}{n_{.1}}\right)$$

$$var\left(\widehat{N}_P\right)=\left(-\frac{n_{1.}}{n_{.1}}\right)^2 var(n_{11})$$

$$var\left(\widehat{N}_P\right)=\frac{n_{1.}^2}{n_{.1}^2}var\left(\frac{\partial f(n_{11})}{\partial n_{11}}\right)_E^2 var(n_{11})$$

$$var\left(\widehat{N}_P\right)=\frac{n_{1.}^2}{n_{.1}^2}\left(\widehat{N}_P p_{1;}q_{1.}\right)$$

$$var\left(\widehat{N}_P\right)=\frac{n_{1.}^2}{n_{.1}^2}\widehat{N}_P\left(\frac{n_{1.}}{\widehat{N}_P}\right)\left(\frac{\widehat{N}_P-n_{1.}}{\widehat{N}_P}\right)$$

$$var\left(\widehat{N}_P\right)=\frac{n_{1.}^3}{n_{.1}^2}\left(\frac{\widehat{N}_P-n_{1.}}{\widehat{N}_P}\right) \tag{11}$$

3.4. Model selection criteria

Comparing the proposed model with existing models, Mean Absolute Deviation (MAD) and Akaike Information Criteria (AIC) were used.

Mean Absolute Deviation (MAD) in case of simulation and is given as:

$$MAD=\frac{\left|N-\widehat{N}\right|}{n} \tag{12}$$

Unrealistically simple assumptions are made which lead to high bias, poor prediction, and missed opportunities for insight when choosing a model with too few parameters. Such models lack the flexibility to explain the sample or the population well. A model with too many parameters can fit the observed data very well, but be too closely tailored to it. Such models may generalize poorly. Penalized-likelihood information criteria, such as Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC), are widely used for model selection (Dziak et al., 2020).

The AIC is computed as follows:

$$AIC=-2logL(\theta^{\wedge})+2k \tag{13}$$

where
θ = the set (vector) of model parameters
$L(\theta^{\wedge})$= the likelihood of the candidate model given the data when evaluated at the maximum likelihood estimate of θ
k = the number of estimated parameters in the candidate model

There is no problem of subjectively specifying an arbitrary significance level to test the models, and comparisons are not restricted to two models which are nested or hierarchically ordered. It is easy to calculate AIC once the maximum likelihood estimators of the parameters of a model are determined. A model with a minimum value of AIC is chosen to be the best fitting model among several competing models (Takane & Bozdogan, 1987).

The BIC is computed as follows:

$$BIC=-2logL(\theta^{\wedge})+klog(n) \tag{14}$$

where
θ = the set (vector) of model parameters
$L(\theta^{\wedge})$= the likelihood of the candidate model given the data when evaluated at the maximum likelihood estimate of θ
k = the number of estimated parameters in the candidate model
n = the sample sizes

When n should be used in the context of mark-recapture is ambiguous. While some are advocating that n is the total number of recorded individuals in the population, others are of the opinion that it should instead be the number of releases, excluding those released from the last sample. AIC is preferable to avoid such inconsistency (Burnham et al., 2011).

Since we are dealing with disproportionality between the first and the second samples whose sizes differ greatly, attempt to consistently estimate the dimension model which requires that the sample size is very large, in model selection we focused more on the distance rather than on dimension of the true model. To this end, the suggestion by Anderson and Burnham (1999) which recommends the use of criteria that are based on Kullback-Leibler information in biological sciences was adopted. AIC is better in situations when a false negative finding would be considered more misleading than a false positive, and BIC is better where false positive is as misleading as, or more misleading than a false negative (Acquah, 2010). In capture recapture model selection, AIC performs slightly better than the BIC methods, which tend to select simpler models (Hook, & Regal, 2000). The usually preferred model selection method in capture recapture studies is the AIC (Zwane et al., 2004).

The Akaike Information Criteria (AIC) used in this work was proposed by Sanni and Jolayemi as cited in Jibasen and Adams (2013) as:

$$AIC = -\beta \sum_{i_0}^{c_0} \sum_{i_1}^{c_1} n(i_0, i_1) log_e \frac{n(i_0, .)n(., i_1)}{\widehat{N}^2} + 2(c_0 + c_1 - 2) \tag{15}$$

Where, c_0 and c_1 are the dimensions of the contingency table, β is equal to 2 just like it is in the classical, or identified as Abs$(N - \widehat{N})$ in case of simulation, n is the number of observations, N is the hypothesized (in case of simulation) and $\widehat{N}$ is the estimated population size. Taking into account that population size estimate may be very sensitive if certain cells are null or very sparse, using log-linear capture-recapture methods Hooks and Regal (1997) suggested that the use of AIC in model selection appears to be preferrable over its BIC counterpart.

3.5. Simulation studies

The hypothesized population size N as well as the first sample size $n_{1.}$ and the second sample size $n_{.1}$ which represent house to house immunization during immunization plus days and LQAS coverages respectively were used to simulate the size of capture in both samples using the hypergeometric setting (Jibasen et al., 2012), thus:

$$P(n_{11}) = \begin{cases} \dfrac{\binom{n_{1.}}{n_{11}}\binom{n_{.1}}{n_{01}}}{\binom{N}{n_{.1}}}, & n_{11} = \max(0, n_{.1} - n_{1,}) to \min(n_{.1}, n_{1,}) \\ 0, & otherwise \end{cases} \tag{16}$$

Simulation scheme was repeated ten times (it could be more) for every hypothesized first, second and population sizes and population sizes for N_o, N_s, N_b, N_a (see Jibasen & Adams, 2013) and N_p were estimated.

$$\widehat{N}_0 = \frac{n^2}{4n_{11}} \tag{17}$$

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

$$\widehat{N}_s = \frac{n_{1.}n_{.1}}{n_{11}} \tag{18}$$

$$\widehat{N}_b = \frac{n_{1.}^2}{n_{1.} - (n_{.1} - n_{11})} \tag{19}$$

$$\widehat{N}_a = \frac{n_{11}(n_{.1} - n_{11}) + (n_{1.} + n_{11})n_{1.}}{2n_{11}} \tag{20}$$

4. Results and Discussion

Tables 1–10 show comparison between the proposed disproportional C-R model and some existing C-R models using simulated data, AIC and MAD. Results of simulated data for different hypothesized values of N, $n_{1.}$ *and* $n_{.1}$ are presented in this session, the simulated data were used to compute estimated population size, AIC values and MAD using the five models, these are M_o, M_s, M_b, M_a, *and* M_p which are No factor effect model, Petersen model, Behavioral model, High recapture model and the proposed model. Each iteration is a complete set of simulation as depicted in Tables 1–10.

Table 1
Ten simulated data sets: N=100, $n_{1.}$ =50 and $n_{.1}$=30

Simulation		Estimated population					AIC				
Iteration	n_{11}	$\widehat{N}_o$	$\widehat{N}_s$	$\widehat{N}_b$	$\widehat{N}_a$	$\widehat{N}_p$	M_o	M_s	M_b	M_a	M_p
1	20	32,000	75	63	37,000	100	85,711	11	11	101,288	4
2	18	28,800	83	66	32,544	103	68,314	9	10	78,653	5
3	20	32,000	75	63	37,000	100	85,711	11	11	101,288	4
4	19	30,400	79	64	34,761	102	76,752	10	11	89,562	5
5	21	33,600	71	61	39,260	98	95,197	11	11	113,839	5
6	18	28,800	83	66	32,544	103	68,314	9	10	78,653	5
7	19	30,400	79	64	34,761	102	76,752	10	11	89,562	5
8	17	27,200	88	68	30,354	105	60,396	7	10	68,552	6
9	21	33,600	71	61	39,260	98	95,197	11	11	113,839	5
10	20	32,000	75	63	37,000	100	85,711	11	11	101,288	4
MAD		30,780	22	36	35,348	2					

Table 1 shows that the proposed model M_p is better than any other model in AIC, MAD and its estimation of the population size.

Table 2
Ten simulated data sets: N=100, $n_{1.}$=60 and $n_{.1}$=45

Simulation		Estimated population					AIC				
Iteration	n_{11}	$\widehat{N}_o$	$\widehat{N}_s$	$\widehat{N}_b$	$\widehat{N}_a$	$\widehat{N}_p$	M_o	M_s	M_b	M_a	M_p
1	30	82,688	90	80	87,750	100	365,321	7	9	390,839	4
2	26	71,663	104	88	73,502	105	269,021	5	7	276,904	6
3	27	74,419	100	86	77,031	104	291,643	4	8	303,327	5
4	27	74,419	100	86	77,031	104	291,643	4	8	303,327	5
5	30	82,688	90	80	87,750	100	365,321	7	9	390,839	4
6	27	74,419	100	86	77,031	104	291,643	4	8	303,327	5
7	24	66,150	113	92	66,528	108	226,661	9	6	228,140	7
8	22	60,638	123	97	59,686	111	188,115	13	5	184,744	8
9	27	74,419	100	86	77,031	104	291,643	4	8	303,327	5
10	28	77,175	96	84	80,584	103	315,230	5	8	331,121	5
MAD		73,768	6	14	76,292	4					

Table 2 clearly shows that while M_o and M_a models performing poorly, M_p performed excellently well with the lowest AIC and MAD values. And much better estimation of the population size.

Table 3
Ten simulated data sets: N=200, $n_{1.}$=90 and $n_{.1}$=60

Simulation		Estimated population					AIC				
Iteration	n_{11}	$\widehat{N}_o$	$\widehat{N}_s$	$\widehat{N}_b$	$\widehat{N}_a$	$\widehat{N}_p$	M_o	M_s	M_b	M_a	M_p
1	41	230,625	132	114	257,665	164	1,521,317	37	35	1,723,244	28
2	46	258,750	117	107	296,332	156	1,942,546	40	36	2,261,856	34
3	36	202,500	150	123	219,672	171	1,153,806	30	32	1,264,607	22
4	46	258,750	117	107	296,332	156	1,942,546	40	36	2,261,856	34
5	40	225,000	135	116	250,000	165	1,443,540	36	35	1,625,131	27
6	34	191,250	159	127	204,748	174	1,021,681	26	31	1,103,351	19
7	44	247,500	123	109	280,808	159	1,767,565	39	36	2,036,810	32
8	36	202,500	150	123	219,672	171	1,153,806	30	32	1,264,607	22
9	44	247,500	123	109	280,808	159	1,767,565	39	36	2,036,810	32
10	46	258,750	117	107	296,332	156	1,942,546	40	36	2,261,856	34
MAD		232,113	68	86	260,037	37					

Table 3 clearly shows that M_p performed better than all the model under consideration following closely by M_s model.

Table 4
Ten simulated data sets: N=300, $n_{1.}$=90 and $n_{.1}$=35

Simulation		Estimated population					AIC				
Iteration	n_{11}	$\widehat{N}_o$	$\widehat{N}_s$	$\widehat{N}_b$	$\widehat{N}_a$	$\widehat{N}_p$	M_o	M_s	M_b	M_a	M_p
1	29	113,281	109	96	157,818	247	498684	77	67	725,554	50
2	27	105,469	117	99	145,071	252	428129	76	65	614,267	43
3	30	117,188	105	95	164,250	244	536093	77	68	785,151	53
4	26	101,563	121	100	138,762	255	394973	75	63	562,537	40
5	31	121,094	102	94	170,717	242	574931	76	69	847,405	57
6	30	117,188	105	95	164,250	244	536093	77	68	785,151	53
7	28	109,375	113	98	151,424	249	462698	77	66	668,599	46
8	27	105,469	117	99	145,071	252	428129	76	65	614,267	43
9	20	78,125	158	108	102,000	270	225338	63	54	305,313	23
10	24	93,750	131	103	126,288	260	332872	73	61	466,789	34
MAD		105,950	182	201	146,265	49					

Table 4 vividly portray M_p is the best model in both AIC and MAD values as well as estimates of the population.

Table 5
Ten simulated data sets: N=300, $n_{1.}$=150 and $n_{.1}$=90

Simulation		Estimated population					AIC				
Iteration	n_{11}	$\widehat{N}_o$	$\widehat{N}_s$	$\widehat{N}_b$	$\widehat{N}_a$	$\widehat{N}_p$	M_o	M_s	M_b	M_a	M_p
1	65	936,000	208	180	1,100,938	292	10,940,576	74	72	13,101,316	14
2	60	864,000	225	188	999,000	300	9,238,940	63	69	10,857,017	4
3	60	864,000	225	188	999,000	300	9,238,940	63	69	10,857,017	4
4	63	907,200	214	183	1,060,007	295	10,241,851	70	71	12,175,385	10
5	56	806,400	241	194	918,512	307	7,985,655	52	65	9,230,167	11
6	53	763,200	255	199	858,892	312	7,108,291	42	62	8,107,397	16
7	61	878,400	221	186	1,019,280	298	9,567,232	66	69	11,287,086	6
8	65	936,000	208	180	1,100,938	292	10,940,576	74	72	13,101,316	14
9	60	864,000	225	188	999,000	300	9,238,940	63	69	10,857,017	4
10	54	777,600	250	197	878,688	310	7,394,805	45	63	8,472,442	15
MAD		859,380	73	112	993,125	5					

Table 5 depicts M_p to be the best among models under consideration.

Table 6
Ten simulated data sets: N=300, $n_{1.}$=150 and $n_{.1}$=120

Simulation		Estimated population					AIC				
Iteration	n_{11}	$\widehat{N}_o$	$\widehat{N}_s$	$\widehat{N}_b$	$\widehat{N}_a$	$\widehat{N}_p$	M_o	M_s	M_b	M_a	M_p
1	76	1,385,100	237	212	1,415,272	243	19,453,915	59	65	19,924,124	56
2	82	1,494,450	220	201	1,554,556	235	22,833,390	69	70	23,852,432	64
3	88	1,603,800	205	191	1,694,704	228	26,496,764	75	72	28,163,305	71
4	85	1,549,125	212	196	1,624,563	231	24,629,471	72	71	25,960,373	68
5	81	1,476,225	222	203	1,531,265	236	22,250,481	68	69	23,171,024	62
6	79	1,439,775	228	206	1,484,766	239	21,108,286	64	68	21,840,191	60
7	70	1,275,750	257	225	1,277,500	250	16,356,341	43	58	16,381,234	48
8	74	1,348,650	243	216	1,369,148	245	18,390,171	54	63	18,700,303	53
9	91	1,658,475	198	186	1,764,900	224	28,435,498	76	72	30,460,281	75
10	87	1,585,575	207	192	1,671,314	229	25,866,406	74	71	27,418,493	70
MAD		1,481,393	77	97	1,538,499	64					

M_p shows a much better estimate than any other model in Table 6.

Table 7
Ten simulated data sets: N=500, $n_{1.}$=150 and $n_{.1}$=50

Simulation		Estimated population					AIC				
Iteration	n_{11}	$\widehat{N}_o$	$\widehat{N}_s$	$\widehat{N}_b$	$\widehat{N}_a$	$\widehat{N}_p$	M_o	M_s	M_b	M_a	M_p
1	40	400,000	188	161	578,000	480	2,696,754	197	171	4,068,376	31
2	35	350,000	214	167	494,813	495	2,031,665	185	156	2,993,276	10
3	43	430,000	174	157	628,897	471	3,143,421	199	179	4,804,559	46
4	37	370,000	203	164	527,824	489	2,285,859	191	162	3,400,839	18
5	37	370,000	203	164	527,824	489	2,285,859	191	162	3,400,839	18
6	41	410,000	183	160	594,890	477	2,841,655	198	174	4,306,081	36
7	35	350,000	214	167	494,813	495	2,031,665	185	156	2,993,276	10
8	35	350,000	214	167	494,813	495	2,031,665	185	156	2,993,276	10
9	38	380,000	197	163	544,464	486	2,418,866	193	165	3,615,805	22
10	43	430,000	174	157	628,897	471	3,143,421	199	179	4,804,559	46
MAD		383,500	304	337	551,023	15					

In Table 7, M_p has the smallest MAD and AIC values as well as most reasonable estimate of the population size.

Table 8
Ten simulated data sets: N=500, $n_{1.}$=150 and $n_{.1}$=100

Simulation		Estimated population					AIC				
Iteration	n_{11}	$\widehat{N}_o$	$\widehat{N}_s$	$\widehat{N}_b$	$\widehat{N}_a$	$\widehat{N}_p$	M_o	M_s	M_b	M_a	M_p
1	77	1,203,125	195	177	1,379,109	260	17,024,961	222	187	19,806,099	282
2	79	1,234,375	190	174	1,422,356	257	17,971,039	218	185	21,027,366	288
3	75	1,171,875	200	180	1,335,938	263	16,105,613	224	188	18,623,831	275
4	79	1,234,375	190	174	1,422,356	257	17,971,039	218	185	21,027,366	288
5	78	1,218,750	192	176	1,400,724	258	17,494,654	220	186	20,411,869	285
6	69	1,078,125	217	189	1,207,121	272	13,507,287	227	190	15,312,340	255
7	78	1,218,750	192	176	1,400,724	258	17,494,654	220	186	20,411,869	285
8	72	1,125,000	208	184	1,271,376	267	14,776,567	227	190	16,923,877	265
9	76	1,187,500	197	179	1,357,512	261	16,561,949	223	188	19,210,079	279
10	77	1,203,125	195	177	1,379,109	260	17,024,961	222	187	19,806,099	282
MAD		1,187,000	302	321	1,357,132	239					

Though having the best MAD value in Table 8, M_p is not the most efficient in terms of AIC values, M_b is.

Table 9
Ten simulated data sets: N=500, $n_{1.}$=200 and $n_{.1}$=100

Simulation		Estimated population					AIC				
Iteration	n_{11}	$\widehat{N}_o$	$\widehat{N}_s$	$\widehat{N}_b$	$\widehat{N}_a$	$\widehat{N}_p$	M_o	M_s	M_b	M_a	M_p
1	73	1,642,500	274	231	2,064,842	454	22,438,886	222	197	28,900,119	82
2	73	1,642,500	274	231	2,064,842	454	22,438,886	222	197	28,900,119	82
3	71	1,597,500	282	234	1,997,195	458	21,163,022	217	194	27,092,801	74
4	74	1,665,000	270	230	2,098,788	452	23,091,473	224	198	29,828,401	87
5	76	1,710,000	263	227	2,166,912	448	24,426,016	227	200	31,734,357	95
6	76	1,710,000	263	227	2,166,912	448	24,426,016	227	200	31,734,357	95
7	70	1,575,000	286	235	1,963,500	460	20,539,720	215	192	26,213,697	70
8	70	1,575,000	286	235	1,963,500	460	20,539,720	215	192	26,213,697	70
9	61	1,372,500	328	248	1,664,660	478	15,366,526	181	177	19,030,554	37
10	74	1,665,000	270	230	2,098,788	452	23,091,473	224	198	29,828,401	87
MAD		1,615,000	220	267	2,024,494	44					

M_p has the best AIC as well as the best MAD values in Table 9. The population size appears to be well estimated.

Table 10
Ten simulated data sets: N=1000, $n_{1.}$=300 and $n_{.1}$=100

Simulation		Estimated population					AIC				
Iteration	n_{11}	$\widehat{N}_o$	$\widehat{N}_s$	$\widehat{N}_b$	$\widehat{N}_a$	$\widehat{N}_p$	M_o	M_s	M_b	M_a	M_p
1	74	2,960,000	405	328	4,222,588	978	42,681,835	752	639	63,113,564	60
2	74	2,960,000	405	328	4,222,588	978	42,681,835	752	639	63,113,564	60
3	80	3,200,000	375	321	4,624,000	960	50,284,113	776	675	75,390,367	114
4	84	3,360,000	357	317	4,894,848	948	55,714,388	785	696	84,265,633	152
5	77	3,080,000	390	325	4,422,534	969	46,401,717	766	657	69,097,653	86
6	81	3,240,000	370	320	4,691,480	957	51,614,454	779	680	77,557,022	123
7	76	3,040,000	395	326	4,355,712	972	45,143,727	761	651	67,068,813	77
8	79	3,160,000	380	323	4,556,681	963	48,971,889	773	669	73,258,322	104
9	75	3,000,000	400	327	4,289,063	975	43,903,774	757	645	65,074,155	69
10	78	3,120,000	385	324	4,489,524	966	47,677,764	769	663	71,160,786	95
MAD		3,111,000	614	676	4,475,902	33					

Table 10 shows the efficacy of M_p, both in AIC and MAD values as well as the estimates of the population size.

4.1. Discussion of the simulation study

M_p model performed better than any other model under consideration in terms of AIC and MAD values as well as closely approximating the corresponding hypothetical population sizes in Tables 1–10. The consistency of the MAD values and the closeness of estimates to the hypothetical population sizes show M_p to be more reliable in estimating disproportionate two-sample capture-recapture population size. This reveals that the proposed model is more efficient than any of the other four models as the AIC values appear to be much smaller when the ratio of the first sample and second sample is directly proportional to the ratio of the hypothesized population to the first sample.

5. Conclusion

Heterogeneity of clusters within a lot and relatively small sample sizes are the albatross associated with LQAS and making it susceptible to Type I or/and Type II errors given the large size of the population targeted for vaccination, hence the need to come up with a complementary tool which would not incur additional cost but to address the pitfalls in LQAS. CRC techniques addresses the issues of heterogeneity and takes into account the relatively small sample size by incorporating the SIAs coverage side by side with the LQAS coverage thereby mitigating the error accruable from heterogeneity and small sample size associated with LQAS.

There are number of CRC techniques used for estimating population size of a close population. Some of these techniques include: Petersen model, No Effect model, Behavioral mode, Efficient model for high recapture to mention but a few. The application of any of the mentioned models is encumber on satisfying the assumptions associated with each of them. SIAs immunization records and LQAS are a typology of C-R method with peculiarity to the disproportionality between the first sample (SIAs house-to-house) and the second sample (LQAS). Consequently, the need to develop a model that addresses this disproportionality.

A disproportionate C-R model was proposed and was compared against some existing C-R models using simulated data, AIC and MAD. The results showed that the AIC and MAD of the proposed model were the smallest compared to No factor, Petersen, Behavioral and Effective model for high recaptures when the ratio of the estimated population size to the first sample size is approximately equal to the ratio of the first sample to the second sample sizes in forestalling the disparity between first sample and second sample sizes in two-sample C-R associated to SIAs data.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

References

Abbott, E. K., Coyle, R., Dayton, A., & Kurilo, M. B. (2021). Measurement and improvement as a model to strengthen immunization information systems and overcome data gaps. *International Journal of Medical Informatics*, *148*, 104412. https://doi.org/10.1016/j.ijmedinf.2021.104412

Acquah, H. D. G. (2010). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Journal of Development and Agricultural Economics*, *2*, 001–006.

Ahlo, J. M. (1990). Logistic regression in capture-recapture models. *Biometrics*, *46*, 623–635. https://doi.org/10.2307/2532083

Alberti, K. P., Guthmann, J. P., Fermon, F., Nargaye, K. D., & Grais, R. F. (2008). Use of lot quality assurance sampling (LQAS) to estimate vaccination coverage helps guide future vaccination efforts. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *102*(3), 251–254. http://doi.org/10.1016/j.trstmh.2007.10.015

Amstrup, B. S. C., & Mcdonald, T. (2010). Capture-recapture estimation and polar bears. *Polar Beers International*, 1–4.

Anderson, D. R., & Burnham, K. P. (1999). Understanding information criteria for selection among capture-recapture or ring recovery models. *Bird Study, 46*(1), 14–21. https://doi.org/10.1080/00063659909477227

Borchers, D. (2012). A non-technical overview of spatially explicit capture-recapture models. *Journal of Ornithology*, *152*(2), 435–444. https://doi.org/10.1007/s10336-010-0583-z

Brown, A. E., Okayasu, H., Nzioki, M. M., Wadood, M. Z., Chabot-Couture, G., Quddus, A., … & Sutter, R. W. (2014). Lot quality assurance sampling to monitor supplemental immunization activity quality: An essential tool for improving performance in polio endemic countries. *Journal of Infectious Diseases*, *210*(1), 333–340. http://doi.org/10.1093/infdis/jit816

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*, 23–35. https://doi.org/10.1007/s00265-010-1029-6

Chao, A., Pan, H., & Chiang, S. (2008). The Petersen – Lincoln estimator and its extension to estimate the size of a shared population. *Biometrical Journal*, *50*(6), 957–970. http://doi.org/10.1002/bimj.200810482

Chao, A., Yip, P. S. F., Lee, S. M., & Chu, W. (2001). Population size estimation based on estimating functions for closed capture – Recapture models. *Journal of Statistical Planning and Inference*, *92*(1-2), 213–232. http://doi.org/10.1016/S0378-3758(00)00151-8

Chopra, M., Bhutta, Z., Blanc, D. C., Checchi, F., Gupta, A., Lemango, E. T., … & Victora, C. G. (2020). Addressing the persistent inequities in immunization coverage. *Bulletin of the World Health Organization*, *98*(2), 146. https://doi.org/10.2471/BLT.19.241620

Clavel, J., Robert, A., Devictor, V., & Julliard, R. (2008). Abundance estimation with a transient model under the robust design. *The Journal of Wildlife Management*, *72*(5), 1203–1210. http://doi.org/10.2193/2006-328

Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermiin, L. S. (2020). Sensitivity and specificity of information criteria. *Briefings in Bioinformatics*, *21*(5), 553–565. https://doi.org/10.1093/bib/bbz016

Hook, E. B., & Regal, R. R. (1997). Validity of methods for model selection, weighting for model uncertainty, and small sample adjustment in capture-recapture estimation. *American Journal of Epidemiology*, *145*(12), 1138–1144. https://doi.org/10.1093/oxfordjournals.aje.a009077

Hook, E. B., & Regal, R. R. (2000). Accuracy of alternative approaches to capture-recapture estimates of disease frequency: Internal validity analysis of data from five sources. *American Journal of Epidemiology*, *152*(8), 771–779. https://doi.org/10.1093/AJE/152.8.771

International Working Group for Disease Monitoring and Forecasting (1995). Capture-recapture and multiple-record system estimation i: History and theoritical development. *American Journal of Epidemiology*, *142*(10), 1047–1058. https://doi.org/10.1093/oxfordjournals.aje.a117558

Jibasen, D. (2011). *Capture-recapture type models for estimatingthe size of an elusive population*. Doctoral dissertation, University of Ilorin.

Jibasen, D., & Adams, Y. J. (2013). An efficient two sample capture-recapture model with high recaptures. *CBN Journal of Applied Statistics*, *4*(2), 4, 141–158.

Jibasen, D., Yahya, W. B., & Jolayemi, E. T. (2012). Capture-recapture estimation for elusive events with two lists. *Mathematical Theory and Modeling*, *2*(8), 1–10.

Jutand, M., & Salamon, R. (2000). Lot quality assurance sampling: Methods and applications in public health. *Revue D'epidemiologie et de Sante Publique, 48*(4), 401–408..

Manning, J. A., & Goldberg, C. S. (2010). Estimating population size using capture-recapture encounter histories created from point-coordinate locations of animals. *Methods in Ecology and Evolution*, *1*(4), 389–397. http://doi.org/10.1111/j.2041-210X.2010.00041.x

Manual, F. (2012). Version edited for the global polio eradication initiative (GPEI). Retrieved from http://www.swachhtakipehel.com/images/knowledgehub/LQAS-assessing-vaccination-coverage-levels.pdf

Mingoti, S. A., & Caiaffa, W. T. (2006). A capture-recapture technique to estimate the size of the injecting drug user population attending syringe exchange programs: AjUDE-Brasil II Project Método de captura-recaptura para estimar o tamanho da população de usuários de drogas injetáveis atendid. *Cadernos de Saúde Pública, Rio de Janeiro*, *22*, 783–789. https://doi.org/10.1590/S0102-311X2006000400017

Okayasu, H., Brown, A. E., Nzioki, M. M., Gasasira, A. N., Takane, M., Mkanda, P., … & Sutter, R. W. (2014). Cluster lot quality assurance sampling: Effect of increasing the number of clusters on classification precision and operational feasibility. *Journal of Infectious Diseases*, *210*(1), 341–346. http://doi.org/10.1093/infdis/jiu162

Olaniyan, A., Isiguzo, C., & Hawk, M. (2021). The socioecological model as a framework for exploring factors influencing childhood immunization uptake in Lagos state, Nigeria. *BMC Public Health*, *21*(1), 867. https://doi.org/10.1186/s12889-021-10922-6

Olives, C. S. (2011). *Improving LQAS for monitoring and evaluation of health programs in resource-poor settings*. USA: Harvard University Press.

Pezzoli, L., & Kim, S. H. (2013). Monitoring health interventions–who's afraid of LQAS? *Global Health Action*, *6*(1), 21921. http://doi.org/10.3402/gha.v6i0.21921

Pollock, K. H. (1976). Building models of capture-recapture experiments. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *25*(4), 253–259. https://doi.org/10.2307/2988083

Pollock, K. H. (1981). Capture-recapture models: A review of current methods, assumptions and experimental design. *Studies in Avian Biology*, *71*, 426–435.

Royle, J. A., & Converse, S. J. (2014). Hierarchical spatial capture-recapture models: Modelling population density in stratified populations. *Methods in Ecology and Evolution*, *5*(1), 37–43. http://doi.org/10.1111/2041-210X.12135

Sangrujee, N., Cáceres, V. M., & Cochi, S. L. (2004). Cost analysis of post-polio certification immunization policies. *Bulletin of the World Health Organization, 82*(1), 9–15.

Sekar, C. C., & Deming, W. E. (1949). On a method of estimating birth rates and extent of registration. *Journal of the American Statistical Association*, *44*(245), 101–115. https://doi.org/10.1198/0003130042935

Siddiqi, D. A., Abdullah, S., Dharma, V. K., Shah, M. T., Akhter, M. A., Habib, A., … & Chandir, S. (2021). Using a low-cost, real-time electronic immunization registry in Pakistan to demonstrate utility of data for immunization programs and evidence-based decision making to achieve SDG-3: Insights from analysis of Big Data on vaccines. *International Journal of Medical Informatics*, *149*, 104413. https://doi.org/10.1016/j.ijmedinf.2021.104413

Sutherland, J., & Schwarz, C. J. (2005). Multi-list methods using incomplete lists in closed populations. *Biometrics*, *61*(1), 134–140. https://doi.org/10.1111/j.0006-341X.2005.021126.x

Takane, Y., & Bozdogan, H. (1987). Akaike information criterion (Aic)-introduction. *Psychometrika, 52*(3), 315–315.

Tebbens, R. J. D., Pallansch, M. A., Cochi, S. L., Wassilak, S. G., Linkins, J., Sutter, R. W., … & Thompson, K. M. (2010). Economic analysis of the global polio eradication initiative. *Vaccine*, *29*(2), 334–343. https://doi.org/10.1016/j.vaccine.2010.10.026

Thompson, K. M., Duintjer Tebbens, R. J., Pallansch, M. A., Kew, O. M., Sutter, R. W., Aylward, R. B., … & Cochi, S. L. (2006). Development and consideration of global policies for managing the future risks of poliovirus outbreaks: Insights and lessons learned through modeling. *Risk Analysis*, *26*(6), 1571–1580. https://doi.org/10.1111/j.1539-6924.2006.00841.x

Uwaibi, N. E., & Omozuwa, S. (2020). Childhood routine immunization coverage in children less than 5 years in Southern Nigeria: A descriptive cross-sectional survey. *Tropical Journal of Medicine and Dental Practice, 1,* 7–15. https://doi.org/10.47227/tjmdp/v1i1.2

Valadez, J. J. (1991). Assessment of decentralized health services using LQAS. *Community Participation in Research: Proceedings of a Workshop Held in Nairobi, Kenya*, 65–72. http://ovidsp.ovid.com/ovidweb.

Zwane, E. N., van der Pal-de Bruin, K., & van der Heijden, P. G. (2004). The multiple-record systems estimator when registrations refer to different but overlapping populations. *Statistics in Medicine*, *23*(4), 2267–2281. https://doi.org/10.1002/sim.1818

How to Cite: Reuben, B. Z., Jibasen, D., Dike, I. J., & Abdulkadir, S. S. (2023). A Model for Estimating the Population Size of Disproportionate Two Sample Capture Recapture Methods. *Journal of Computational and Cognitive Engineering* 2(3), 249–259, https://doi.org/10.47852/bonviewJCCE2202226

Nomenclature: AIC: Akaike's Information Criterion; BIC: Bayesian Information Criterion C-R: Capture Recapture; CLQAS: Cluster Lot Quality Assurance Sampling; GPEI: Global Polio Eradication Initiative; IIS: Immunization Information System; IPDs: Immunization Plus Days; IWGDMF: International Working Group for Disease Monitoring and Forecasting; LQAS: Lot Quality Assurance Sampling; MAD: Mean Absolute Deviation; MLE: Maximum Likelihood Estimator; NIPDs: National Immunization Plus Days; SDG: Sustainable Development Goals; SIAs: Supplementary Immunization Activities; UHC: Universal Health Coverage; VPD: Vaccine Preventable Diseases; WPV: Wild Polio Virus.