American Journal of Mathematics and Statistics 2012, 2(6): 199-205
DOI: 10.5923/j.ajms.20120206.06

Alternative Estimation Method for a Three-Stage Cluster Sampling in Finite Population

L. A. Nafiu[1,*], I. O. Oshungade[2], A. A. Adewara

[1]Department of Mathematics and Statistics, Federal University of Technology, Minna, +234, Nigeria
[2]Department of Statistics, University of Ilorin, Ilorin, +234, Nigeria

Abstract This research investigates the use of a three-stage cluster sampling design in estimating population total. We focus on a special design where certain number of visits is being considered for estimating the population size and a weighted factor of / is introduced. In particular, attempt was made at deriving new method for a three-stage sampling design. In this study, we compared the newly proposed estimator with some of the existing estimators in a three-stage sampling design. Eight (8) data sets were used to justify this paper. The first four (4) data sets were obtained from[1],[2],[3] and[4] respectively while the second four (4) data sets represent the number of diabetic patients in Niger state, Nigeria for the years 2005, 2006, 2007 and 2008 respectively. The computation was done with software developed in Microsoft Visual C++ programming language. All the estimates obtained show that our newly proposed three-stage cluster sampling design estimator performs better.

Keywords Sampling, Three-stage, Cluster Design, Estimator, Bias and Variance

1. Introduction

In a census, each unit (such as person, household or local government area) is enumerated, whereas in a sample survey, only a sample of units is enumerated and information provided by the sample is used to make estimates relating to all units[5] and[6]. In designing a study, it can be advantageous to sample units in more than one-stage. The criteria for selecting a unit at a given stage typically depend on attributes observed in the previous stages[7]. Multistage sampling is where the researcher divides the population into clusters, samples the clusters, and then resample, repeating the process until the ultimate sampling units are selected at the last of the hierarchical levels[8]. If, after selecting a sample of primary units, a sample of secondary units is selected from each of the selected primary units, the design is referred to as two-stage sampling. If in turn a sample of tertiary units is selected from each selected secondary unit, the design is three-stage sampling[9].

The aim of this paper is to model a new estimator for three-stage cluster sampling scheme which is to be compared with the other existing seven conventional estimators.

2. Methodology

* Corresponding author:
lanconserv@yahoo.com (L. A. Nafiu)
Published online at http://journal.sapub.org/ajms

Sub sampling has a great variety of applications[3] and the reason for multistage sampling is administrative convenience[10]. The process of sub sampling can be carried to a third stage by sampling the subunits instead of enumerating them completely[11]. Comparing multistage cluster sampling with simple random sampling, it was observes that multistage cluster sampling is better in terms of efficiency[12]. Multistage sampling makes fieldwork and supervision relatively easy[4]. Multistage sampling is more efficient than single stage cluster sampling[13] and references had been made to the use of three or more stages sampling[9].

Let N denote the number of primary units in the population and n the number of primary units in the sample. Let M_i be the number of secondary units in the primary unit. The total number of secondary units in the population is

$$M = \sum_{i=1}^{N} M_i \tag{1}$$

Let y_{ij} denote the value of the variable of interest of the jth secondary unit in the ith primary unit. The total of the y-values in the ith primary unit is

$$Y_i = \sum_{j=1}^{M_i} y_{ij} \tag{2}$$

Accordingly, the population total for over-all sample in a two-stage is given as

$$Y = \sum_{i=1}^{N}\sum_{j=1}^{M_i} y_{ij} \tag{3}$$

For a three-stage sampling, the population contains N first-stage units, each with M second-stage units, each of which has K third-stage units. The corresponding numbers for the sample are n, m and k respectively. Let y_{iju} be the value obtained for the uth third-stage units in the jth second-stage units drawn from the ith primary units. The relevant population total for over-all sample in a three-stage is given as follows:

$$Y=\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{u=1}^{K}y_{iju} \tag{4}$$

For any estimation $\hat{\Theta}_h$ in the hth cell based on completely arbitrary probabilities of selection, the total variance is then the sum of the variances for all strata. The symbol E is used for the operator of expectation, V for the variance, and $\hat{V}$ for the unbiased estimate of V. We may then write

$$V(\hat{\Theta}_h)=\underset{1}{V}(\underset{>1}{E}(\hat{\Theta}_h))+\underset{1}{E}(\underset{>1}{V}(\hat{\Theta}_h)) \tag{5}$$

The expression (5) may be written into three components as:

$$V(\hat{\Theta}_h)=\underset{1}{V}(\underset{2}{E}(\underset{>2}{E}(\hat{\Theta}_h)))+\underset{1}{E}(\underset{2}{V}(\underset{>2}{E}(\hat{\Theta}_h)))+\underset{1}{E}(\underset{2}{E}(\underset{>2}{V}(\hat{\Theta}_h))) \tag{6}$$

where ">1" is the symbol to represent all stages of sampling after the first[3].

3. Proposed Three-Stage Cluster Sampling Design

To estimate the population size at different hospitals using three-stage sampling, the unbiased estimator of population total can be derived as follows. In a three-stage sampling without replacement design supported by[3],[4] and[14]; a sample of primary units is selected, then a sample of secondary units is chosen from each of the selected primary units and finally, a sample of tertiary units is chosen from each selected secondary unit. For instance, the state consists of N number of local government areas out of which a simple random sampling of n number of local government areas is selected. Each local government area consists of M_i number of cities out of which a simple random sampling without replacement of m_i number of cities is selected. Finally, from the selected sample of city containing K_{ij} number of hospitals, k_{ij} number of hospitals is selected at random without replacement and the number of diabetic patients in this hospital is collected.

Then;

$$y=\sum_{i=1}^{n}\sum_{j=1}^{m_i}y_{ij} \tag{7}$$

Again, let n be the number of primary units (local government areas) sampled without replacement, m_i be the number of secondary units (cities) selected without replacement from the ith sampled primary unit (local government area) and k_{ij} be the number of tertiary units (hospitals) selected from the jth secondary unit (city) in the ith primary unit (local government area). An unbiased estimator of the population total at jth secondary unit in the ith primary unit in the sample is:

$$\hat{y}_{ij}=\frac{1}{\gamma_{ij}}\sum_{l=1}^{k_{ij}}\frac{N_i}{n_i^2}y_{ijl}$$
$$=\frac{K_{ij}}{k_{ij}}\sum_{l=1}^{k_{ij}}\frac{N_i}{n_i^2}y_{ijl} \tag{8}$$

where $\gamma_{ij}=\frac{k_{ij}}{K_{ij}}$ is the known sampling fraction for tertiary units in the jth secondary unit of the ith primary unit.

Also, let y_{ijl} denote the number of individuals (tertiary units) in the sample from the jth secondary unit of the ith primary unit who engage in the treatment of diabetes. An unbiased estimator of the population total in the ith primary unit in the sample is:

$$\hat{y}_i=\frac{M_i}{m_i}\sum_{j=1}^{m_i}\hat{y}_{ij} \tag{9}$$

Finally, an unbiased estimator of the population total of the diabetic patients undergoing treatment in all the hospitals at the jth secondary unit (city) in the ith primary unit (local government area) is:

$$\hat{Y}_{3NPE}=\frac{N}{n}\sum_{i=1}^{n}\hat{y}_i$$
$$=\frac{N}{n}\sum_{i=1}^{n}\{\frac{M_i}{m_i}\sum_{j=1}^{m_i}(\frac{K_{ij}}{k_{ij}}\sum_{l=1}^{k_{ij}}\frac{N_i}{n_i^2}y_{ijl})\} \tag{10}$$

4. Theorems

4.1. Theorem 1: $\hat{Y}_{3NPE}$ is Unbiased for the Population total Y

Proof:

We know that expectation of $\hat{y}_{ij}$ given by equation (8) conditional on samples s_1 and s_2 of primary units and secondary units respectively equals y_{ij} of engaging in the variable of interest in each primary unit and each secondary unit [15]. That is;

$$E(\hat{y}_{ij}|s_1,s_2)=y_{ij} \tag{11}$$

Also, the expectation of $\hat{y}_i$ given by equation (9) conditional on sample s_1 of primary units equals y_i of engaging in the variable of interest in each primary unit.

That is;

$$E(\hat{y}_i|s_1)=y_i \tag{12}$$

To obtain the expected value of $\hat{Y}_{3NPE}$ given by equation (10) over all possible samples of primary units.

Then, the expectation of $\hat{Y}_{3NPE}$ is :

$$E(\hat{Y}_{3NPE})=E_1[E_2\{E_3(\hat{Y}_{3NPE}|s_1,s_2)|s_1\}] \tag{13}$$

where s_1 and s_2 denote the samples of primary units and secondary units respectively.

$$E(\hat{Y}_{3NPE})=E[E\left\{E\left(\frac{N}{n}\sum_{i=1}^{n}\hat{y}_{ij}\middle|s_1,s_2\right)\middle|s_1\right\}]$$
$$=E[E\{\frac{N}{n}\sum_{i=1}^{n}\hat{y}_i|s_1\}]$$
$$=E\{\frac{N}{n}\sum_{i=1}^{n}y_i\}$$

$$= \frac{N}{n}\frac{n}{N}\sum_{i=1}^{N} y_i$$

$$= Y \qquad (14)$$

This implies that the proposed estimator $\hat{Y}_{3NPE}$ is unbiased.

Hence, the variance of the newly proposed estimator $\hat{Y}_{3NPE}$ of the population total is derived as follows:

In line with[3] and[14], we use

$$V(\hat{Y}_{3NPE}) = V\{E(\hat{Y}_{3NPE}|s_1,s_2)|s_1\} + E\{V[(\hat{Y}_{3NPE}|s_1,s_2)|s_1]\} =$$

$$V\{E(\hat{Y}_{3NPE})\} + E\{V[E(\hat{Y}_{3NPE})]\} + E\{E[V(\hat{Y}_{3NPE})]\} \quad (15)$$

Because of the simple random sampling of primary units and secondary units without replacement at the first stage and second stage respectively, the first term to the right of the equality in equation (15) is:

$$V\{E(\hat{Y}_{3NPE}|s_1,s_2)|s_1\} = V\{\tfrac{N}{n}\textstyle\sum_{i=1}^{n} y_i\} = \frac{N(N-n)}{n}\sigma_1^2 \quad (16)$$

The second term to the right of the equality in equation (15) is:

$$V\{(\hat{Y}_{3NPE}|s_1,s_2)|s_1\} = V\{\frac{N}{n}\sum_{i=1}^{n}\hat{y}_i|s_1\}$$

$$= \left(\frac{N}{n}\right)\sum_{i=1}^{n} V(\hat{y}_i|s_1) + (\frac{N}{n})^2\sum_{i=1}^{n} V(\hat{Y}_{3NPE})$$

$$= \left(\frac{N}{n}\right)\sum_{i=1}^{n}\sigma_{ij}^2 + (\frac{N}{n})^2\sum_{i=1}^{n}\sigma_i^2$$

$$E[V\{(\hat{Y}_{3NPE}|s_1,s_2)|s_1\}]$$

$$= \frac{N}{n}\sum_{i=1}^{N}\frac{M_i}{m_i}\sum_{j=1}^{M_i} K_{ij}(K_{ij}-k_{ij})\frac{\sigma_{ij}^2}{k_{ij}}$$

$$+ (\frac{N}{n})^2\frac{n}{N}\sum_{i=1}^{N} M_i(M_i-m_i)\frac{\sigma_i^2}{m_i}$$

$$= \tfrac{N}{n}\textstyle\sum_{i=1}^{N} M_i(M_i-m_i)\frac{\sigma_i^2}{m_i} + \frac{N}{n}\sum_{i=1}^{N}\frac{M_i}{m_i}\sum_{j=1}^{M_i} K_{ij}(K_{ij} -$$

$$kij)\sigma ij2kij \qquad (17)$$

Equations (16) and (17) give;

$$V(\hat{Y}_{3NPE}) = N(N-n)\frac{\sigma_1^2}{n} + \tfrac{N}{n}\textstyle\sum_{i=1}^{N} M_i(M_i-m_i)\frac{\sigma_i^2}{m_i} +$$

$$Nni{=}1NMimij{=}1MiKijKij{-}kij\sigma ij2kij \quad (18)$$

where

$$\sigma_1^2 = \frac{\sum_{i=1}^{N}(Y_i-\bar{Y})^2}{N-1} \qquad (19)$$

$$\sigma_i^2 = \frac{\sum_{j=1}^{M_i}(Y_{ij}-\frac{Y_i}{M_i})^2}{M_i-1} \qquad (20)$$

$$\sigma_{ij}^2 = \frac{K_{ij}^2}{k_{ij}^2}\sum_{l=1}^{k_{ij}}(\frac{N_i^2}{n_i^4}-\frac{N_i}{n_i^2})(y_{ijl}-\bar{y}_{ij})^2 \qquad (21)$$

We note that the first term to the right of the equality in equation (18) is the variance that would be obtained if every tertiary unit in a selected secondary unit and every secondary unit in a selected primary unit was observed, that is, if y_i's were known for $i = 1,2,\cdots,n$. The second term contains variance that would be obtained if every tertiary unit in a selected secondary unit was observed, that is, if y_{ij}'s were known for $i = 1,2,\cdots,n$ and $j = 1,2,\cdots,m_i$. The third term contains variance due to estimating the y_{ij}'s from a subsample of tertiary units within the selected secondary units. An unbiased estimator of the variance of $\hat{Y}_{3NPE}$ given in equation (18) is obtained by replacing the population variances with the sample variances as follows:

$$\hat{V}(\hat{Y}_{3NPE}) = N(N-n)\frac{s_1^2}{n} + \tfrac{N}{n}\textstyle\sum_{i=1}^{n} M_i(M_i-m_i)\frac{s_i^2}{m_i} +$$

$$Nni{=}1nMimij{=}1miKijKij{-}kijsij2kij \quad (22)$$

where

$$s_1^2 = \frac{\sum_{i=1}^{n}(y_i-\frac{\hat{Y}_{3NPE}}{n})^2}{n-1} \qquad (23)$$

$$s_i^2 = \frac{\sum_{j=1}^{m_i}(y_{ij}-\frac{y_i}{m_i})^2}{m_i-1} \qquad (24)$$

$$s_{ij}^2 = \frac{K_{ij}^2}{k_{ij}^2}\sum_{l=1}^{k_{ij}}(\frac{N_i^2}{n_i^4}-\frac{N_i}{n_i^2})(y_{ijl}-\bar{y}_{ij})^2 \qquad (25)$$

4.2. Theorem 2: $\hat{V}(\hat{Y}_{3NPE})$ is Unbiased for $V(\hat{Y}_{3NPE})$

Proof:

We note that

$$s_1^2 = \frac{1}{n-1}(\textstyle\sum_{i=1}^{n} y_i^2 - \frac{n\hat{Y}_{3NPE}^2}{N^2}) \qquad (26)$$

Next, we note that

$$E(\sum_{i=1}^{n} y_i^2) = E\{E(\sum_{i=1}^{n} y_i^2|s_1)\}$$

$$= E(\sum_{i=1}^{n}[V(y_i|s_1) + \{E(y_i|s_1)\}^2])$$

$$= E(\sum_{i=1}^{n}\sigma_i^2 + \sum_{i=1}^{n} Y_i^2)$$

$$= E(\sum_{i=1}^{N} z_i\sigma_i^2 + \sum_{i=1}^{N} z_i Y_i^2)$$

$$= \tfrac{n}{N}(\textstyle\sum_{i=1}^{N}\sigma_i^2 + \sum_{i=1}^{N} Y_i^2) \qquad (27)$$

In addition;

$$E(\hat{Y}_{3NPE}^2) = V(\hat{Y}_{3NPE}) + \{E(\hat{Y}_{3NPE})\}^2$$

$$= \frac{N(N-n)}{n}\sigma_1^2 + \tfrac{N}{n}\textstyle\sum_{i=1}^{N}\sigma_i^2 + Y^2 \qquad (28)$$

Combining equations (27) and (28), we have:

$$E(s_1^2) =$$

$$\frac{n}{N(n-1)}(\textstyle\sum_{i=1}^{N}\sigma_i^2 + \sum_{i=1}^{N} Y_i^2) - \frac{n}{N^2(n-1)}(\frac{N(N-n)}{n}\sigma_1^2 +$$

$$Nni{=}1N\sigma i2{+}Y2)$$

$$= \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\sigma_i^2 + \frac{(N-1)n}{(n-1)N}\left\{\frac{1}{N-1}\left(\sum_{i=1}^{N} Y_i^2 - \frac{Y^2}{N}\right)\right\} + \frac{(N-n)}{N(n-1)}\sigma_1^2 \quad (29)$$

Using the fact that:

$$\sigma_1^2 = \frac{1}{N-1}\left(\sum_{i=1}^{N} Y_i^2 - \frac{Y^2}{N}\right)$$

Equation (29) becomes;

$$E(s_1^2) = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 + \frac{(N-1)n}{(n-1)N}\sigma_1^2 - \frac{(N-n)}{N(n-1)}\sigma_1^2$$

$$= \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\sigma_i^2 + \frac{N(N-n)}{n}\sigma_1^2 \qquad (30)$$

Next, we note that;

$$E(s_i^2) = E\{E(\sum_{i=1}^{n} s_i^2)\}$$
$$= E\{E(\sum_{i=1}^{n} s_i^2)\}$$
$$= E\{E(\sum_{i=1}^{n} s_i^2|s_1)\}$$
$$= E\{\sum_{i=1}^{n} \frac{M_i^2}{m_i^2}\sigma_i^2)$$
$$= E\{\sum_{i=1}^{N} z_i \frac{M_i^2}{m_i^2}\sigma_i^2)$$
$$= \frac{n}{N}\{\frac{N^2}{n^2}\sum_{i=1}^{N} M_i(M_i - m_i)\frac{\sigma_i^2}{m_i} - \frac{1}{n}\sum_{i=1}^{N} M_i(M_i - m_i)\frac{\sigma_i^2}{m_i}\}$$
$$= \frac{N}{n}\sum_{i=1}^{N} M_i(M_i - m_i)\frac{\sigma_i^2}{m_i} - \frac{1}{N}\sum_{i=1}^{N} M_i(M_i - m_i)\frac{\sigma_i^2}{m_i}$$

Therefore;

$$E(s_i^2) = \frac{N}{n}\sum_{i=1}^{N} M_i(M_i - m_i)\frac{\sigma_i^2}{m_i} - \frac{1}{N}\sum_{i=1}^{N} M_i(M_i - mi\sigma i2mi \tag{31}$$

Also;

$$E(s_{ij}^2) = E\{E[E(s_{ij}^2|s_1, s_2)]\}$$
$$= E\{E[\frac{K_{ij}^2}{k_{ij}^2}\sum_{l=1}^{k_{ij}} E(y_{ijl}|s_1)]\}$$
$$= E\{\frac{M_i^2}{m_i^2}E[\sum_{j=1}^{m_i}\frac{K_{ij}^2}{k_{ij}^2}\sum_{l=1}^{k_{ij}} E(y_{ijl}|s_1)]\}$$
$$= E\{\frac{M_i^2}{m_i^2}E[\sum_{j=1}^{M_i}\frac{K_{ij}^2}{k_{ij}^2}\sigma_{ij}^2]\}$$
$$= E\{\sum_{i=1}^{N}\frac{M_i^2}{m_i^2}E[\sum_{j=1}^{M_i}\frac{K_{ij}^2}{k_{ij}^2}\sigma_{ij}^2]\}$$
$$= E\{\sum_{i=1}^{N} z_i\frac{M_i^2}{m_i^2}\sum_{j=1}^{M_i} K_{ij}(K_{ij} - k_{ij})\frac{\sigma_{ij}^2}{k_{ij}}\}$$
$$= \frac{n}{N}\{\frac{N^2}{n^2}\sum_{i=1}^{N}\frac{M_i}{m_i}\sum_{i=1}^{M_i} K_{ij}(K_{ij} - k_{ij})\frac{\sigma_{ij}^2}{k_{ij}}\}$$
$$= \frac{N}{n}\sum_{i=1}^{N}\frac{M_i}{m_i}\sum_{i=1}^{M_i} K_{ij}(K_{ij} - k_{ij})\frac{\sigma_{ij}^2}{k_{ij}}$$

Therefore;

$$E(s_{ij}^2) = \frac{N}{n}\sum_{i=1}^{N}\frac{M_i}{m_i}\sum_{i=1}^{M_i} K_{ij}(K_{ij} - k_{ij})\frac{\sigma_{ij}^2}{k_{ij}} \tag{32}$$

Combining equations (30), (31) and (32), we have;

$$E\{\hat{V}(\hat{Y}_{3NPE})\} = N(N-n)\frac{\sigma_1^2}{n} + \frac{N}{n}\sum_{i=1}^{N} M_i(M_i - m_i)\frac{\sigma_i^2}{m_i} + \frac{N}{n}\sum_{i=1}^{N}\frac{M_i}{m_i}\sum_{j=1}^{M_i} K_{ij}(K_{ij} - k_{ij})\frac{\sigma_{ij}^2}{k_{ij}}$$
$$= V(\hat{Y}_{3NPE}) \tag{33}$$

That is;

$$E\{\hat{V}(\hat{Y}_{3NPE})\} = V(\hat{Y}_{3NPE})$$

Hence, $\hat{V}(\hat{Y}_{3NPE})$ is an unbiased sample estimator of the proposed estimator ($\hat{Y}_{3NPE}$) in three-stage cluster sampling design.

This estimator, $\hat{Y}_{3NPE}$, is then compared with these seven conventional three stage cluster sampling design estimators:

i. $\hat{Y}_{HH} = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{Y}_{i..}}{P_i}$ (34)

ii. $\hat{Y}_{HHG} = M_0\frac{\sum_{j=}^{m_i}\sum_{i=1}^{n}\hat{y}_{ij}}{\sum_{i=1}^{n} M_i}$ (35)

iii. $\hat{Y}_{HT} = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{Y}_{i..}}{\pi_i}$ (36)

iv. $\hat{Y}_{RHC} = \sum_{g=1}^{n}\frac{Z_g}{z_g}M_g\bar{Y}_{g..}$ (37)

v. $\hat{Y}_C = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{u=1}^{k_{ij}} y_{iju}$ (38)

vi. $\hat{Y}_T = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{u=1}^{k_{ij}} y_{iju}$ (39)

and

vii. $\hat{Y}_O = \frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{j=1}^{m_i}\frac{K_{ij}}{k_{ij}}\sum_{u=1}^{k_{ij}} y_{iju}$ (40)

5. Data Used for this Study

There are eight (8) categories of data sets used in this paper. The first four (4) data sets were obtained from[1],[2],[3] and[4] respectively. The second four (4) data sets used are of secondary type and were collected from[16] and[17] where we constructed a sampling frame from all diabetic patients with chronic eye disease (Glaucoma and Retinopathy) in the twenty-five (25) Local Government Areas of the state between 2005 and 2008.

6. Results

The estimates obtained with the aid of software developed using Visual Basic C++ Programming Language[18] are given in tables 1 - 12 for the illustrated and the real-life data respectively.

Table 1. Estimated Population Totals using Illustrated Data

Estimator	Cases			
	I	II	III	IV
$\hat{Y}_{HH}$	400	98,983	36	13,898
$\hat{Y}_{HHG}$	440	111,310	22	14,000
$\hat{Y}_{HT}$	479	139,527	27	14,791
$\hat{Y}_{RHC}$	437	98,830	31	15,214
$\hat{Y}_{C}$	385	131,675	39	13,963
$\hat{Y}_{T}$	456	102,635	34	14,576
$\hat{Y}_{O}$	397	141,194	24	13,950
$\hat{Y}_{3NPE}$	492	99,136	42	15,016

Table 2. Estimated Population Totals using Real Life Data

Estimator	POP1	POP2	POP3	POP4
$\hat{Y}_{HH}$	26,022	26,541	28,428	29,356
$\hat{Y}_{HHG}$	24,355	25,019	25,162	28,610
$\hat{Y}_{HT}$	25,514	27,197	28,731	29,096
$\hat{Y}_{RHC}$	26,043	26,428	27,301	27,451
$\hat{Y}_{C}$	24,420	25,321	26,851	28,365
$\hat{Y}_{T}$	25,804	27,197	27,609	29,472
$\hat{Y}_{O}$	27,204	28,124	28,631	29,251
$\hat{Y}_{3NPE}$	26,151	26,625	27,511	28,090

Table 3. Biases of Estimated Population Totals using Illustrated Data

Estimator	Cases I	Cases II	Cases III	Cases IV
$\hat{Y}_{HH}$	29	516	2	446
$\hat{Y}_{HHG}$	19	337	8	540
$\hat{Y}_{HT}$	29	427	3	133
$\hat{Y}_{RHC}$	15	364	2	461
$\hat{Y}_{C}$	25	309	1	265
$\hat{Y}_{T}$	33	463	9	369
$\hat{Y}_{O}$	13	261	3	476
$\hat{Y}_{3NPE}$	11	219	1	112

Table 4. Biases of Estimated Population Totals using Real Life Data

Estimator	POP1	POP2	POP3	POP4
$\hat{Y}_{HH}$	196	136	136	125
$\hat{Y}_{HHG}$	147	176	127	168
$\hat{Y}_{HT}$	122	117	114	128
$\hat{Y}_{RHC}$	155	153	134	118
$\hat{Y}_{C}$	127	146	151	120
$\hat{Y}_{T}$	164	156	124	195
$\hat{Y}_{O}$	137	137	137	158
$\hat{Y}_{3NPE}$	112	104	103	107

Table 5. Variances of the Estimated Population Totals using Illustrated Data

Estimator	Case I	Case II	Case III	Case IV
$\hat{V}(\hat{Y}_{HH})$	584.3718	163,955.9954	3.0753	2,012.5838
$\hat{V}(\hat{Y}_{HHG})$	498.0637	146,797.7757	2.9402	1,958.2391
$\hat{V}(\hat{Y}_{HT})$	444.7838	131,942.6350	1.6489	1,861.7476
$\hat{V}(\hat{Y}_{RHC})$	425.8846	109,764.4333	1.6226	1,844.8793
$\hat{V}(\hat{Y}_{C})$	418.4156	107,030.5504	1.6201	1,842.4048
$\hat{V}(\hat{Y}_{T})$	315.7266	96,800.8501	1.5117	1,834.4763
$\hat{V}(\hat{Y}_{O})$	296.8283	94,624.7532	1.1674	1,810.5586
$\hat{V}(\hat{Y}_{3NPE})$	274.6806	91,963.5764	1.1105	1,619.5559

Table 6. Variances of the Estimated Population Totals using Real Life Data

Estimator	POP1	POP2	POP3	POP4
$\hat{V}(\hat{Y}_{HH})$	17,464	16,636	15,360	15,146
$\hat{V}(\hat{Y}_{HHG})$	15,714	16,286	14,139	15,100
$\hat{V}(\hat{Y}_{HT})$	13,419	14,626	11,493	13,315
$\hat{V}(\hat{Y}_{RHC})$	11,684	11,788	11,398	12,396
$\hat{V}(\hat{Y}_{C})$	9,985	11,532	10,938	12,239
$\hat{V}(\hat{Y}_{T})$	9,749	11,441	10,532	12,069
$\hat{V}(\hat{Y}_{O})$	9,568	10,507	10,330	11,087
$\hat{V}(\hat{Y}_{3NPE})$	9,118	9,726	9,884	10,612

Table 7. Standard Error for Estimated Population Total using Illustrated Data

Estimator	Case I	Case II	Case III	Case IV
$\hat{Y}_{HH}$	24.1738	404.9148	1.7537	44.8618
$\hat{Y}_{HHG}$	22.3173	383.1420	1.7147	44.2520
$\hat{Y}_{HT}$	21.0899	363.2391	1.2841	43.1480
$\hat{Y}_{RHC}$	20.6370	331.3072	1.2738	42.9521
$\hat{Y}_{C}$	20.4552	327.1552	1.2728	42.9232
$\hat{Y}_{T}$	17.7687	311.1283	1.2295	42.8308
$\hat{Y}_{O}$	17.2287	307.6114	1.0805	42.5507
$\hat{Y}_{3NPE}$	16.5785	302.2597	1.0538	40.2437

Table 8. Standard Error for Estimated Population Total using Real Life Data

Estimator	POP1	POP2	POP3	POP4
$\hat{Y}_{HH}$	132.1502	128.9710	130.6399	123.0702
$\hat{Y}_{HHG}$	125.3573	127.6163	125.3413	122.8823
$\hat{Y}_{HT}$	115.8417	120.9390	113.0051	115.3810
$\hat{Y}_{RHC}$	108.0937	108.5724	112.5340	111.3351
$\hat{Y}_{C}$	99.9287	107.3872	110.2437	110.6305
$\hat{Y}_{T}$	98.7379	106.9641	108.1752	109.8593
$\hat{Y}_{O}$	97.8150	102.5056	107.1344	105.2964
$\hat{Y}_{3NPE}$	95.4893	98.6229	104.7944	103.0137

Table 9. 95% Confident Intervals for Estimated Population using Illustrated Data

Estimator	Case I	Case II	Case III	Case IV
$\hat{Y}_{HH}$	(353,447)	(98189,99777)	(33,39)	(13810,13986)
$\hat{Y}_{HHG}$	(396,484)	(110559,112061)	(19,25)	(13913,14087)
$\hat{Y}_{HT}$	(438,520)	(138815,140239)	(24,30)	(14706,14876)
$\hat{Y}_{RHC}$	(397,477)	(98181,99479)	(29,33)	(15130,15298)
$\hat{Y}_{C}$	(345,425)	(131034,132316)	(37,41)	(13879,14047)
$\hat{Y}_{T}$	(421,491)	(102025,103245)	(32,36)	(14492,14660)
$\hat{Y}_{O}$	(363,431)	(140591,141797)	(22,26)	(13867,14033)
$\hat{Y}_{3NPE}$	(460,524)	(98542,99730)	(40,44)	(14939,15095)

Table 10. 95% Confident Intervals for Estimated Population Totals using Real Life Data

Estimator	Population 1	Population 2	Population 3	Population 4
$\hat{Y}_{HH}$	(25760,26290)	(26290,26790)	(28170,28680)	(29110,29600)
$\hat{Y}_{HHG}$	(24110,24600)	(24770,25270)	(24920,25410)	(28370,28850)
$\hat{Y}_{HT}$	(25290,25740)	(26960,27430)	(28510,28950)	(26870,27320)
$\hat{Y}_{RHC}$	(25830,26250)	(26220,26640)	(27080,27520)	(27230,27670)
$\hat{Y}_{C}$	(24220,24620)	(25110,25530)	(26630,27070)	(28150,28580)
$\hat{Y}_{T}$	(25610,26000)	(26990,27410)	(27400,27820)	(29260,29690)
$\hat{Y}_{O}$	(27010,27400)	(27920,28320)	(28420,28840)	(29040,29460)
$\hat{Y}_{3NPE}$	(25960,26340)	(26410,26800)	(27310,27720)	(27890,28290)

Table 11. Coefficient of Variation for Estimated Population Totals using Illustrated Data

Estimator	Case I	Case II	Case III	Case IV
$\hat{Y}_{HH}$	6.04%	0.41%	4.87%	0.32%
$\hat{Y}_{HHG}$	5.07%	0.34%	7.79%	0.32%
$\hat{Y}_{HT}$	4.40%	0.26%	4.76%	0.29%
$\hat{Y}_{RHC}$	4.72%	0.34%	4.11%	0.28%
$\hat{Y}_{C}$	5.31%	0.25%	3.26%	0.31%
$\hat{Y}_{T}$	3.90%	0.30%	3.62%	0.29%
$\hat{Y}_{O}$	4.34%	0.22%	4.50%	0.31%
$\hat{Y}_{3NPE}$	3.98%	0.31%	2.51%	0.27%

Table 12. Coefficient of Variation for Estimated Population Totals using Real Life Data

Estimator	Populations			
	1	2	3	4
$\hat{Y}_{HH}$	0.44%	0.43%	0.42%	0.42%
$\hat{Y}_{HHG}$	0.51%	0.51%	0.44%	0.43%
$\hat{Y}_{HT}$	0.45%	0.44%	0.37%	0.34%
$\hat{Y}_{RHC}$	0.42%	0.41%	0.37%	0.41%
$\hat{Y}_{C}$	0.37%	0.42%	0.38%	0.33%
$\hat{Y}_{T}$	0.38%	0.33%	0.36%	0.37%
$\hat{Y}_{O}$	0.36%	0.36%	0.36%	0.36%
$\hat{Y}_{3NPE}$	0.34%	0.33%	0.35%	0.37%

7. Discussion of Results

The estimation methods given in equation (10) was applied to four different illustrated data (Cases I – IV) and four real life data (Populations 1 - 4). The population totals obtained for illustrated data are given in table 1 while the population totals obtained for real life data are given in table 2. Table 3 give the biases of the estimated population totals for illustrated data for our own estimator as 11, 219, 1, and 112 for cases I – IV respectively while table 4 gives that of the four life data sets as 112, 104, 103, and 107 respectively. This implies that our own estimator has the least biases using both data sets. Table 5 shows the variances obtained using illustrated data for our own estimator as 274.6806, 91963.5764, 1.1105 and 1619.5559 for cases I – IV respectively while table 6 shows that of life data sets as 9118.2037, 9726.4809, 9883.6215 and 10611.8216 respectively meaning that our own estimator has the least variances using both data sets. Table 7 shows the obtained standard errors for the estimated population totals using illustrated data for our own estimator as 16.5785, 302.2597, 1.0538 and 40.2437 for cases I – IV respectively while table 8 shows that of life data sets as 95.4893, 98.6229, 104.7944 and 103.0137 respectively meaning that our own estimator have the least standard errors using both data sets.

The confidence intervals of the estimated populations in table 1 are given in table 9 for α=, 5%. The confidence intervals of the estimated populations in table 2 are given in table 10 for α =5 % which shows that all the estimated population totals fall within the computed intervals as expected. For our own estimator, table 11 gives the coefficients of variations for the estimated population totals using illustrated data as 3.98%, 0.31%, 2.51% and 0.27% for cases I – IV respectively while table 12 gives that of life data sets as 0.34%, 0.33%, 0.35% and 0.37% respectively which means that our newly proposed three stage cluster estimator has the least coefficient of variation, hence it is preferred.

8. Conclusions

The alternative estimation method of population allows the use of certain number of visits to the venues (hospitals) within the clusters (cities) and a more precise (minimum mean square error) estimate was obtained and the estimates presented indicate that substantial reduction in the variances was obtained through the use of newly proposed estimator. We also observed that irrespective of the data considered, the variance of newly proposed estimator is always less than those of already existing estimators in three-stage cluster sampling designs. The newly proposed estimator ($\hat{Y}_{3NPE}$) is preferred to the already existing estimators considered in this study and is therefore recommended.

REFERENCES

[1] Horvitz, D. G. and D. J. Thompson, "A Generalization of Sampling without Replacement from a Finite Universe", Journal of American Statistical Association, vol. 47 663-685, 1952.

[2] Raj, D., The Design of Sample Survey, McGraw-Hill, Inc., New York, USA, 1972.

[3] Cochran, W.G., Sampling Techniques, Third Edition, John Wiley and Sons , New York, 1977.

[4] Okafor, F., Sample Survey Theory with Applications, Afro-Orbis Publications, Nigeria, 2002.

[5] Kish, L., Survey Sampling, John Wiley and Sons, New York, 1967.

[6] Adams, J. L., S. L. Wickstrom, M. J. Burgess, P. P. Lee, and J. J. Escarce, "Sampling Patients within Physician Practices and Health Plans: Multistage Cluster Samples in Health Services Research", The Global Journal for Improving Health Care Delivery and Policy, vol. 38: 1625-1640, 2003.

[7] Kuk, A., "Estimation of Distribution Functions and Medians under Sampling with Unequal Probabilities", Biometrika, vol. 75, 97-103, 1988.

[8] Goldstein, H., Multilevel Statistical Models, Halstead Press, New York, 1995.

[9] Tate, J.E. and M.G. Hudgens, "Estimating Population Size with Two-stage and Three-stage sampling designs", American Journal of Epidemiology,vol. 165 no.11, 1314-1320, 2007.

[10] Kalton, G., Introduction to Survey Sampling, Sage Publications, Thousand Oaks, C.A., 1983.

[11] Henry, G. T., *Practical* Sampling, Thousand Oaks, C.A.: Sage Publications, 1990.

[12] Fink, A., How To Sample In Survey, C.A. Sage Publications, Thousand Oaks, 2002.

[13] Kendall, M. G. and A. S. Stuart, The Advanced Theory of Statistics, Hafner Publishing Company, New York, 1968.

[14] Thompson, S. K., Sampling, John Wiley and Sons, New York, 1992.

[15] Nafiu, L. A., An Alternate Estimation Method for Multistage Cluster Sampling in Finite Population, Unpublished Ph.D Thesis, University of Ilorin, Ilorin, Nigeria, 2012.

[16] Niger State, Niger State Statistical Year Book, Niger Press Printing & Publishing, Nigeria, 2009.

[17] National Bureau of Statistics, Directory of Health Establishments in Nigeria, Abuja Printing Press, Nigeria, 2007.

[18] Hubbard, J. R., Programming with C++, Second Edition. Schaum's Outlines, Tata McGraw-Hill Publishing Company Limited, New Delhi, 2000.