Scope
Volume 13 Number 3 September 2023

Zero-Truncated Model Estimations of Fake Drug Syndicates Inonitsha Southeast, Nigeria

Ugochukwu Osisiogu[1], Theresa Efor[2], Chinwuba Emeka[3]

[1] Ebonyi State University, Department of Industrial Mathematics and Applied Statistics, Abakaliki, Ebonyi State, Nigeria.
[2]Ebonyi State University, Department of Industrial Mathematics and Applied Statistics, Abakaliki, Ebonyi State, Nigeria.
[3]Ebonyi State University, Department of Industrial Mathematics and Applied Statistics, Abakaliki, Ebonyi State, Nigeria.

Corresponding Author**: Theresa Efor**

Abstract

This paper examines the use of zero truncated models for the estimate of the population size of fake drug Syndicates in Onitsha, Southeast Nigeria. Various estimators under zero-truncated Poisson models and geometric models were used to estimate this kind of hidden population. The estimators are Maximum Likelihood, Turing, Chao's Lower Bound and Zelterman. The weighted estimators for the four candidate estimators under zero-truncated Poisson model gives the population size of fake drug Syndicates as 6579. As 1434 were observed, this means that only about 22% of these Syndicates are observed with 95% confidence interval of 19%-25% leaving about 78% unobserved still in the distribution chain of this drugs. Similarly, the weighted estimator for the four candidate estimators under zero-truncated geometric model is 12649 with only about 11% of the population being observed with confidence interval of 10%-13%, leaving about 89% still in the distribution chain. This is not good news for National Agency for Food and Drug Administration and Control (NAFDAC), the agency responsible for checkmating illicit and counterfeits drugs in Nigeria. The study also shows that of all the offense committed by fake drug Syndicates, falsification of genuine drugs was rampantly committed, followed by the selling of expired drugs, while the least committed offense was the selling of banned drugs

Key words: Fake drug Syndicates, zero-truncated Poisson models, zero-truncated geometric models, maximum Likelihood estimator, Turing estimator, Chao's Lower Bound estimator and Zelterman estimator.

1. Introduction

Onitsha, a commercial city of Anambra State, is in the Southeast of Nigeria. According to the National Population Commission (NPC) estimate, the city has a population of 1.5 million inhabitants in 2021. The National Bureau of Statistics (NBS) 2021 report, estimated annual volume of trade in Onitsha-Market to be in excess of three billion US dollars, with about 40% of this earnings coming from none banking transactions. This makes Onitsha market one of the highest Gross Domestic Product (GDP) in Nigeria and one of the biggest Markets in Africa, the report stated.

Being the commercial hub of Southeast of Nigeria, different goods and services are traded in Onitsha markets. Among them are Building Material Market; Ceramic Market; Foods and Vegetables Market;

Footwear Market; Abada Market; Main Market; Drug Markets and so forth. Following frequent demand for cheap drugs, fake and adulterated drugs freely entered into Onitsha drug market without control. The influx had nearly peaked in 2007 before the National Agency for Food and Drug Administration Control (NAFDAC)set up a Special Zonal Office in Onitsha to control the spread.

NAFDAC Onitsha Special Office thus embarked on aggressive surveillance and arrested many syndicates, sealed off many pharmaceutical stores found with fake or adulterated drugs; seizedtruck loads of fake or adulterated drugs and either destroyed or confiscated them. The Daily Champion Newspapers in its March, 2007 edition reported that NAFDAC Onitsha Zonal Office intercepted many truck loads of fake, banned and substandard drugs worth about 6.5 billion naira. In fact, the alarming situation resulted in closing down the whole drug markets in Onitsha by NAFDAC in 2007.

The Chairman of Kano State Taskforce on Counterfeit Drugs also added his voice to the return of fake drugs in Onitsha drug markets. He, therefore, called for immediate closure of the market to insulate the country from unwholesome drugs where counterfeit drugs were being shipped to Kano, (Guardian Newspapers May 2018). The chairman said: "From 2012 when the Taskforce was established till date, we have confiscated and destroyed fake drugs including codeine, tramadol and other out of prescription medication drugs from Onitsha drug markets"

The genuine traders in Onitsha drug markets even lamented on the return of drug cartels. Bridge Head Drug Market Traders Association raised the alarm in a press conference in 2020, saying that drug barons were trying to use the upcoming election in the market to return to the business by foisting one of their candidates as a leader. "If you know the history of our market", they said, you will know that the market was closed down many years ago by Prof. Dora Akunyili when she was Director General of NADFAC because the market became notorious for fake drugs (Daily Post Newspapers, March 15, 2020).These were some of the disturbing comments about the return of fake drugs in Onitsha Markets. .

Arrest of syndicates of these drugs by NAFDAC, National Drug Law Enforcement Agency (NDLEA) and the Police were huge, confiscation and destruction of them were alarming, yet the influx continued to soar. This means there are many fake drug syndicates that escaped arrest by the NAFDAC. Hence the cell f_0 is empty in the NAFDAC records.

Frequency counts of observed cases in a single register or multiple registers with the aim to estimate the number of unobserved cases give rise to zero truncated models. This method has been used to estimate the hidden populations of illicit drug users and homeless persons in [1].Let us consider a population of size N and count variable Y taking values from the set of integers {0, 1, 2, 3…}. For example, in the study of drug syndicates Y might represent the number of times a syndicate is arrested. Also denote withf_0, f_1, f_2 ..., the frequency with which a 0, 1, 2, …, occurs in this population. Again consider a list where every syndicate arrested is included except Y = 0, meaning that the syndicate escape arrest. This list reflects a count variable truncated at zero which we denote by Y_0.Accordingly, the list will have observed frequenciesf_1, f_2, f_3 ..., but the frequency f_0 of zeros in the population is unknown.

Let n denotes the size of the observed zero-truncated counts with f_k being the frequency of observing exactly k counts.

In this paper, we use similar description of[2], that fake drug syndicates (in general terminology can be regarded as units) that are arrested only once are also called singleton, units that occur twice are called doubletons and units that occur thrice are called tripletons, and so forth. In Table I, there are 1250 singletons, whereas there are only 94 doubletons and 25 tripletons. This huge number of singletons might be easily explained as being caught once by the NAFDAC officials.

1.1. Objectives

The objectives of this paper are the following

- To use estimators namely Maximum Likelihood, Turing, Chao's Lower Bound, and Zelterman estimators under zero truncated Poisson distribution in estimating fake drug Syndicates in Onitsha
- To use estimators namely Maximum Likelihood, Turing, Chao's Lower Bound, and Zelterman estimators under zero truncated geometric distribution to estimate the population size of fake drug Syndicates in Onitsha
- To construct a weighted estimators for estimating the fake drug Syndicates in Onitsha

2. Materials and Methods

Capture-recapture methods have been proven to provide reliable estimates of hidden populations, including illegal populations in [3]. The method relies on a pattern found in the observed part of the population to make inference on the unobserved part. We start by reviewing some methodologies relevant to our study.

Table 1 represents the Frequency distribution (per count) of NAFDAC records on fake drug syndicates (FDSs) in Onitsha Markets from January 2011—December 2011 (figures extracted from A Quarterly Magazines of National Agency for Food and Drug Administration and Control (NAFDAC) Vol.2 No.2 2011

TABLE I

Kind of offense committed	Number of arrest per month												n
	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	
Sale of falsified drug	565	54	15	9	8	5	5	3	3	1	1	0	669
Sale of expired drug	288	20	6	3	3	2	1	1	0	1	0	0	325
Sale of unregistered drug	210	14	3	4	3	2	1	1	0	0	0	0	238
Sale of banned drug	187	6	1	2	2	2	1	1	0	0	0	0	202
Count of FDSs	1250	94	25	18	16	11	8	6	3	2	1	0	1434

Frequency distribution (per count) of NAFDAC

2.1 Horvitz-Thompson estimator: With capture-recapture experiment, the frequency counts of arrested syndicates are the variable of interest. NAFDAC records provides a count $Y_k > 0$ of how many times a syndicate k^{th} has been arrested, for $k = 1, 2, \ldots, n$ and $Y_k = 0$ for syndicates that escape arrest. If we let P_0 be the probability of a syndicate that escaped arrest, then $N(1- P_0)$ is the expected number of syndicates that has been arrested which can be estimated. This leads us to a simple equation used to estimate the population size N. Thus we have:

$$N = NP_0 + N(1 - P_0) = NP_0 + n. \qquad (2.1)$$

This simple equation (2.1) can be solved by estimating N to provide the Horvitz-Thomson estimator which we shall note as:

$$\hat{N}_{HTE} = \frac{n}{(1-P_0)} \qquad (2.2)$$

The variance of Horvitz-Thomson estimator ($\hat{N}_{HTE}$) as can be seen in[4]is given as

$$Var(\hat{N}_{HTE}) = \frac{n(1-P_0)P_0}{(1-P_0)^2} \qquad (2.3)$$

A common approach for deriving an estimator, P_0, is based upon counting repeated trials. According to [4], this is referred to as a capture–recapture experiment in continuous time (CRECT).The estimators we use in this study is built upon this Horvitz-Thompson estimator.

2.2Poisson model with zero-truncation

A typical capture-recapture method is interested in finding appropriate models for the count variables. In the study of animal abundance, the probability of capturing of animals usually followed a binomial distribution. If there are many trapping occasions with little catch,binomial distribution need to be approximated by Poisson which work well in rare a success. Let Y follows a Poisson distribution with parameter Λ so that:

$$P(Y = k) = P_0(k|\Lambda) = \frac{\exp(-\Lambda)\Lambda^k}{k!}, \quad k = 0, 1, 2, 3, \ldots \qquad (2.4)$$

The zero-truncated Poisson distribution is defined as a probability function conditional on y > 0, such that:

$$P(Y_+ = k) = P_{0+}(k|\Lambda) = \frac{\exp(-\Lambda)\Lambda^k}{1-\exp(-\Lambda)k!}, \quad k = 0, 1, 2, 3, \ldots \qquad (2.5)$$

If we let n to be the number of syndicates arrested, and f_0 the frequencies of syndicates not arrested, then population of syndicates N shall be:

$$\text{N} = n + \hat{f}_0 (2.6)$$

2.3Geometric model with truncation: Geometric distribution on the other hand arises as a result of a mixtureof Poisson parameter with exponential distribution according to [5]. The geometric distribution has a major interesting property that turns out to be useful for the truncated process:

Let $(1-p)^k\, p$ be the geometric for $k = 0,1, \ldots$. The zero-truncated geometric is of the form

$(1-p)^{k-1}p; k = 1,2,3, \ldots$

Assuming that each syndicate in the population has equal chance of being arrested, it then follows that:

$$P_k(Y > 0) = P(Y > 0) = 1 - P(Y = 0) \qquad (2.6)$$

Hence, therefore, the population size of syndicates can be estimated by means of the Horvitz-Thompson estimator:

$$\widehat{N} = \sum_{k=1}^{n} \frac{1}{1-P(Y=0)} = \frac{n}{\exp(-\Lambda)} \qquad (2.7)$$

3. Estimators of Poisson and Geometric Models

3.1 Maximum likelihood estimator ($\widehat{N}_{MLE}$): Authors in **[6]** defined Maximum likelihood method as a traditional technique used to derive estimators. If we let $K_1, K_{2,\ldots,K_n}$be a random sample with probability density function$f(k;\theta)$, the likelihood function is defined as:

$$L(\theta) = \prod_{i=1}^{n} f(k;\theta) \qquad (3.0)$$

We can then obtain the maximum likelihood estimator (MLE) for unknown parameter θ by maximizing the function,$L(\theta)$, by differentiating it with respect toθ and equating to zero. For instance, if we let Y_i be the number of times that i^{th}syndicate was arrested over the surveillance period say,$K = 1,2,3, \ldots, k$. The count data K is modeled by the zero-truncated Poisson distribution with probability function:

$$P_0^+(k;\lambda) = \frac{\exp(-\lambda)\lambda^k}{k!(1-\exp(-\lambda))};\ \lambda > 0,\ \ k = 1, 2, 3, \ldots \qquad (3.1)$$

If we let f_k denotes the frequencies of syndicates arrested k times over the period of the surveillance where $k = 1,2, \ldots m$ and $\sum_{k=1}^{m} f_k = \mathrm{n}$. Then, the likelihood function for this zero-truncated count density shall be:

$$L(\lambda) = \prod_{k=1}^{m}\left(\frac{P_0(k,\ \lambda)}{1-\exp(-\lambda)}\right)^{f_k}\ k = 1,2, \ldots, m \qquad (3.2)$$

With the information in (3.2), the log-likelihood function shall be:

$$l(\lambda) = -n\lambda + \log\lambda \sum_{\mathrm{k}=1}^{\mathrm{m}} \mathrm{k} f_k - \sum_{\mathrm{k}=1}^{\mathrm{m}} f_k \log(\mathrm{k}!) - \mathrm{n}\log(1-\exp(-\lambda)) \quad (3.3)$$

When we take derivative of $l(\lambda)$ with respect to λ we shall have:

$$\frac{\partial l}{\partial \lambda} = -n + \frac{1}{\lambda}\sum_{k=1}^{m} k f_k - \frac{n\,exp(-\lambda)}{1-\exp(-\lambda)} = 0 \qquad (3.4)$$

$$\frac{1}{n}\sum_{k=1}^{m} k f_k = \frac{\lambda}{1-\exp(-\lambda)}$$

$$\bar{v} = \frac{\hat{\lambda}}{1-\exp(-\hat{\lambda})} \text{ or } \hat{\lambda} = \bar{v}(1-\exp(-\hat{\lambda})$$

ByapplyingTaylor's series approximation, the value of $\hat{\lambda}$ can be approximated as

$$\hat{\lambda} = 2\left(\frac{\bar{v}-1}{\bar{v}}\right)$$

The likelihood function (3.3) can be maximized with algorithm between the E-Step and M-step.

(i) Expectation (E-Step): The expected value of unobserved case f_0 given that the observed variable is known, the current estimates of likelihood parameter are derived as follows:

$$\hat{f}_0 = E(f_0, f_1, f_2 \ldots.., f_m;\ \lambda) \quad = p_0 N \qquad (3.5)$$
$$= \exp(-\lambda)(\mathrm{n} + \hat{f}_0\ )$$

Hence

$$\hat{f}_0 = \frac{n\,p_0}{1-p_0} = \frac{n\exp(-\lambda)}{1-\exp(-\lambda)} \qquad (3.6)$$

(ii) Maximization (M-Step): In this step, the unobserved, complete data likelihood function is maximized by using observed cases (n) and unobserved cases (f_0) that is imputed from initial value in first iteration and from $\hat{f}_0$ from E-Step for next iteration. The estimate of λ in M-Step is

$$\hat{\lambda}_{MLE} = \frac{1}{n+f_0}(0f_0 + 1f_1 + 2f_2 + \cdots + mf_m) \qquad (3.7)$$

Where, n is the total number of observed cases and on the condition that $f_0 = \hat{f}_0$. The EM-algorithm requires iterating between E-step and M-step until $\hat{\lambda}_{MLE}$ and $\hat{f}_0$converges. The initial value is very important to start the procedure, so it should be selected carefully. In [5], the authors suggested that the value may beset to sample mean.

(iii) Maximization (Poisson Approach): As a result of replacing $\hat{\lambda}_{MLE}$in Horvitz-Thompson approach (2.2), the population size estimator with regard to maximum likelihood under Poisson shall be:

$$\hat{\mathrm{N}}_{MLE-p} = \frac{n}{1-\exp(\hat{\lambda}_{MLE})} \qquad (3.8)$$

The variance of (3.7) was estimated as

$$\widehat{Var}(\hat{N}_{MLE-P})=\frac{\hat{N}_{MLE-P}}{\left(exp\left(\frac{\sum kf_k}{\hat{N}_{MLE-P}}\right)-\frac{\sum kf_k}{\hat{N}_{MLE-P}}\right)} \quad (3.9)$$

For variance estimate (3.9), the reader isreferred to [7]), [8], [9] and [10])for more understanding.

(iv). Maximization (Geometric Approach): Here, we consider maximum likelihood estimation under geometric model. We assume that count data K is modeled by a geometric distribution with probability function $p_k = (1-p)^k$p; $k = 0, 2, \ldots$and the zero-truncated geometric likelihood is of the form

$$L(p)=\prod_{k=1}^{m}((1-p)^{k-1}p^{f_k}).$$

The log-likelihood function is

$$logL(p) = \log(1-p)\sum_{k=1}^{m} f_k(k-1) + \log p \sum_{k=1}^{m} f_k \quad (3.10)$$

To find the maximum likelihood estimator (MLE) of unknown parameter p we differentiate (3.10) with respect to p and set it to zero, and we shall have:

$$\frac{\partial l}{\partial \mathrm{p}} = -\frac{\sum_{k=1}^{m} f_k(k-1)}{1-p} + \frac{\sum_{k=1}^{m} f_k}{p} = 0$$
$$\hat{p} = \frac{n}{S}$$

Hence under the assumption of zero-truncated geometric model the population size estimator with the maximum likelihood approach shall be:

$$\hat{N}_{MLE-G} = \frac{n}{1-n/S} \quad (3.11)$$

Wheres $=\sum_{k=1}^{m} kf_k$. The variance estimation of the MLE-G in (3.11) can be estimated as

$$\widehat{Var}(\hat{N}_{MLE-G}) = \frac{S^2n^2}{(S-n)^2} \quad (3.12)$$

For more details of (3.12), the reader is referred to [11].

3.2 Turing estimator ($\hat{N}_T$): According to[5], Turing estimation is formulated to estimate the number of classes or species of animals which is defined as the sum of probabilities of observed classes. For the author, the estimator can be used to estimate the total number of population. Let f_k be the frequency of individuals detected exactly k times, k = 0, 1, 2, .., m where m is the largest observed count. The total number of observed cases in the sample is $n = \sum_{k=1}^{m} f_k$ and the total number of captured cases can be defined as

$$S = 1f_1 + 2f_2 + \cdots + mf_m$$

3.3. Turing Estimation under Poisson: Under Poisson,let p_k denotes the probability that a syndicate has been arrested exactly k times. Assume that K has homogeneous Poisson distribution with parameter Λ so that $p_0 = \exp(-\Lambda)$ and $p_1 = \Lambda\exp(-\Lambda)$,we can write:

$$p_0 = \exp(-\Lambda) = \frac{\exp(-\Lambda)\Lambda}{\Lambda} = \frac{p_1}{E(K)} \quad (3.13)$$

The estimator of $p_0$can be calculated from observed frequency as follows:

$$\hat{p}_0 = \frac{f_1/N}{S/N} = \frac{f_1}{S} \quad (3.14)$$

If we plug $\hat{p}_0$ into Horvitz-Thompson estimator, Turing estimator for estimating the population size is given by:

$$\hat{N}_{T\_P} = \frac{n}{1-f_1/S} \quad (3.15)$$

The variance for Turing estimator can be estimated as

$$\widehat{var}(\hat{N}_{T\_P}) = \frac{nf_1/S}{(1-f_1/S)^2} + \frac{n^2}{(1-f_1/S)^4}\left(\frac{f_1(1-\frac{f_1}{N})}{S^2} + \frac{f_1^2}{S^3}\right) \quad (3.16)$$

3.4. Turing Estimation under Geometric: Let K have a marginal probability mass function following the geometric distribution with parameter p where $p_0 = p$; $p_1 = (1-p)p$ and $E(K) = (1-p)/p$ so that

$$\frac{P_1}{E(K)} = \frac{p(1-P)}{(1-p)/p} = p^2$$

$$\sqrt{\frac{P_1}{E(K)}} = p = p_0$$

The estimate of p_0 can be calculated from the observed frequency as follows:

$$\hat{p}_0^* = \sqrt{f_1/S} \quad (3.17)$$

Therefore, the extension of Turing estimator for estimating the population size under geometric model is given by:

$$\hat{N}_{T\_G} = \frac{n}{1-\sqrt{f_1/S}} \quad (3.18)$$

The variance of $\hat{N}_{T\_G}$ can be derived as

$$\widehat{var}(\hat{N}_{T\_G}) = \frac{n\sqrt{f_1/S}}{\left(1-\sqrt{f_1/S}\right)^2} + n^2\left(\frac{S+f_1}{4S^2\left(1-\sqrt{f_1/S}\right)^4}\right) \quad (3.19)$$

3.5 Chao's lower bound estimator ($\hat{N}_C$): Estimators we have so far discussed are developed under homogenous Poisson mode, but in practice it is rarely met. Therefore, it is more suitable to incorporate heterogeneity, and in doing that we assume that the target population may be composed of a variety of subgroups. Chao, in [12] provided a lower bound estimator for the population size N under the heterogeneous Poisson population.

3.6 Chao's lower bound estimator Poisson, $\hat{N}_{C\_P}$: Assuming that capture probability followed a Poisson mixture, Chao proved that the lower bound for the estimate of the number of unobserved shall be:

$$\hat{f}_0 = \frac{f_1^2}{2f_2} \quad (3.20)$$

Where, the inequality $\hat{f}_0 \leq f_0$and its expected value are asymptotical. Finally,by adding the estimator $\hat{f}_0$ to the observed cases n Chao's lower bound estimator shall be:

$$\hat{N}_{C\_P} = n + \frac{f_1^2}{2f_2} \quad (3.21)$$

The approximate variance formula for estimator in (3.21) which was provided by[11]is given as:

$$\widehat{var}(\hat{N}_{C\_P}) = \left(\frac{1}{4}\right)\frac{f_1^4}{f_2^3} + \frac{f_1^3}{f_2^2} + \left(\frac{1}{2}\right)\frac{f_1^2}{f_2} \quad (3.22)$$

3.7 **Chao's lower bound estimator geometric ($\hat{N}_{C\_G}$):** Suppose Chao's lower bound estimator under geometric heterogeneity is also considered,the estimate shall be:

$$\hat{N}_{C\_G} = n + \frac{f_1^2}{f_2} \quad (3.23)$$

while its variance is

$$\widehat{var}(\hat{N}_{C\_G}) = \frac{f_1^4}{f_2^3} + \frac{4f_1^3}{f_2^2} + \frac{f_1^2}{f_2} \quad (3.24)$$

3.8 Zelterman's estimator ($\hat{N}_Z$): Because Poisson assumption is frequently violated, Zelterman (1988), argued that homogeneity Poisson probability may be valid for small ranges of Y such as from k to k+1. For example, singleton $f_1$and doubleton f_2 follows a homogeneous Poisson distribution, whereas other counts might be arbitrarily distributed according to [5] Thus, the neighbouring frequencies f_k and f_{k+1} can be used to estimate a parameter λ by considering Poisson distributions of truncated and untruncated as shown below:

$\frac{P_0(k+1|\lambda)}{P_0(k|\lambda)} = \frac{\lambda}{k+1}$ and $\frac{P_{0+}(k+1|\lambda)}{P_{0+}(k|\lambda)} = \frac{\lambda}{k+1}$ respectively to estimate the λ.

Thus we have:

$$\lambda = \frac{(k+1)P_{0+}(k+1|\lambda)}{P_{0+}(k|\lambda)} = \frac{(k+1)P_{0+}(k+1|\lambda)}{P_{0+}(k|\lambda)} \quad (3.25)$$

The estimator for λ is obtained by replacing $P_{0+}(k|\lambda)$ with the empirical frequencyf_k so that we shall have:

$$\hat{\lambda}_k = \frac{(k+1)f_{k+1}}{f_k}. \quad (3.26)$$

Thus, if we let k = 1, we find that $\hat{\lambda}_1 = 2f_2/f_1$.The authors in [14]noted that the estimator (3.26) is often used for two reasons:

i. $\hat{\lambda}_1$is using frequencies in the vicinity of $f_0$which is the target of prediction,
ii. In many application studies for estimating f_0 the majority of counts fall into f_1 and f_2

For these two reasons, the estimator is not affected by changes in the data for counts larger than 2, which contributed largely to its robustness.

(i) Zelterman estimator based on Poisson ($\hat{N}_{Z\_P}$): Ifwe recall that if $\hat{\lambda}_1 = 2f_2/f_1$, thenZelterman estimator of Poisson distribution shall be:

$$\hat{N}_{Z\_P} = \frac{n}{1-exp(-\lambda)} = \frac{n}{1-\exp(-\frac{2f_2}{f_1})} \quad (3.27)$$

Bohning (2006) worked out the variance of (3.26) to be:

$$\widehat{var}\ (\hat{N}_{Z\_P}) = n(\frac{\exp(-2f_2/f_1)}{(1-\exp(-2f_2/f_1))^2})\left[1 + n(\frac{\exp(-2f_2/f_1)}{(1-\exp(-2f_2/f_1))^2})\left(\frac{2f_2}{f_1}\right)^2\left(\frac{1}{f_1}+\frac{1}{f_2}\right)\right] \quad (3.28)$$

(ii) Zelterman estimator based on geometric distribution ($\hat{N}_{Z\_G}$): In a similar way,Zelterman estimator under geometric distribution is given as

$$\hat{N}_{Z\_G} = \frac{nf_1}{f_2} \quad (3.29)$$

In [1], the variance of (3.27) is worked to be: $\widehat{var}(\widehat{N}_{Z\_G}) = \frac{nf_f(f_1-f_2)}{f_2^2} + n^2\left(\frac{f_1}{f_2^2} + \frac{f_1^2}{f_2^3}\right)$

4. Analysis and Discussions

Having reviewed some of the literatures relevant to our study,Table II shows the results of the population size estimates of fake drug syndicates with the four reviewed estimators under Poisson and geometric models.

The data on Table II depicts that both MLE and Turing estimators have negative biases, but with small variances. While Zelterman's estimator has small bias but with extra ordinary variance, Chao's estimator is negatively biased a little bit but with large variance though not to be compared with other two variances abovementioned. In [4]the authors suggested that in such situation, there is need to combine the positive aspects of the estimators by construction of a weighted estimator, whichwas constructed as:

$N_W = \left(w_1\widehat{N}_{MLE} + w_2\widehat{N}_T + w_3\widehat{N}_C + w_4\widehat{N}_Z\right)/(w_1 + w_2 + w_3 + w_4)$, where $\widehat{N}_{MLE}$, $\widehat{N}_T$, $\widehat{N}_C$ and $\widehat{N}_Z$ are the estimators of MLE, Turning, Chao and Zelterman respectively. Since the true variances are unknown it wise to use equal weights as follows: $N_W = \frac{1}{4}(\widehat{N}_{MLE} + \widehat{N}_T + \widehat{N}_C + \widehat{N}_Z)$. Table III, thus, compared the four estimators with the weighted estimator.

Table II

Estimator	Observed	Poisson			Geometric		
		Estimated	SE	CI	Estimated	SE	CI
MLE	1434	2108	61.35	(1988, 2228)	5925	276.05	(5384-6466)
Turning	1434	4218	88.83	(4044, 4392)	7547	672.00	(6230-8864)
Chao	1434	9745627.01		(8516, 10974)	18056	1955.81	(14223-21889)
Zelterman	1434	10243906.70		(8466, 12020)	19069	2048.21	(15055-23083)

Results of estimate of fake drug syndicates in Onitsha(see Table I) by four different estimators

Table III

Observed	Model	Estimator				
		MLE SE	Turning SE	Chao SE	Zelterman SE	Weighted SE
1434	Poisson	2108 61.35	4218 88.83	9745 627.01	10243 906.70	6579 421.00
1434	Geometric	5925 276.05	7547 672.00	18056 1955.81	19069 2048.21	12649 1238.02

Result of MLE, Turing, Chao and Zelterman estimators and the weighted estimator

In this study, we also categorized fake syndicates according to the offense they committed. The offenses are falsifying of genuine drugs, selling of expired drugs, selling of unregistered drugs and selling of banned drugs. Table IV displays the kind of offense a syndicate committed, the observed number of the syndicates arrested by the NAFDAC and the estimate of the population size of the syndicates by the four different estimators which we explored under Poisson and geometric models. We also constructed the weighted estimator of various offenses committed by the syndicates to compare it with the four estimators. The results of it are shown in Table V.

Table IV

Offense	Estimator	Observed	Poisson Estimated	SE (CI)	Geometric Estimated	SE(CI)
Falsifying of drug	MLE	669	2230	105.6 (2023-2437)	1338	52.0 (1236-1440)
	Turning	669	1115	42.0 (1033-1197)	1673	82.0 (1591-1755)
	Chao	669	3625	476.0 (2692-4558)	6581	949.0 (4721-8441)
	Zeterman	669	3717	482.0(2772-4662)	7000	1027.0 (4987-9013)
Selling of expired drug	MLE	325	878	85.0 (711-1045)	1548	170.0 (1215-1881)
	Turning	325	1083	135.0 (818-1348)	2031	411.0 (1225-2837)
	Chao	325	4472	526.0 (3441-5503)	4472	1050.0 (2310-6634)
	Zeterman	325	2708	613.0 (1507-3909)	4680	1110.9 (2503-6857)
Selling of unregistered drug	MLE	238	680	75.3 (532-828)	1081	132.4 (822-1340)
	Turning	238	768	146.1 (482-1054)	1400	254.9 (900-1900)
	Chao	238	1813	475.4 (881-2745)	3388	949.2 (1528-5248)
	Zeterman	238	1983	535.8 (933-3033)	3570	1010.4 (1590-5550)
Selling of banned Drug	MLE	202	697	99.8 (501-893)	1122	165.3 (798-1446)
	Turning	202	842	243.4 (365-1319)	1554	514.3 (546-2562)
	Chao	202	3116	1264.9 (637-5595)	6030	2528.6 (1074-10986)
	Zeterman	202	3367	1332.3 (756-5978)	6296	2647.2 (1107-11485)

Results of estimate of fake drug syndicates in Onitsha Market by kind of offense committed (See Table I) by the four estimators

Table X

Offense	Observed	Model	MLE SE	Turing SE	Chao SE	Zelterman SE	Weighted SE
falsified drug	666 666	Poisson Geometric	2230 (105.6) 1338 (52.0)	1115 (42.0) 1673 (82.0)	3625 (476.0) 6581(949.0)	3717(482.0) 7000(1027.0)	2672(276.4) 4148(527.5)
Expired drug	325 325	Poisson Geometric	878(85.0) 1548 (170.0)	1083(135.0) 2031(411.0)	4472(526.0) 4472(1050.0)	2708(613.0) 4680(1110.9)	2286(340.0) 3183(685.5)
Unregistered drug	238 238	Poisson Geometric	680 (75.3) 1081 (132.4)	768 (146.1) 1400(254.0)	1813(475.4) 3388(949.2)	1983(535.8) 3570 (1010.4)	1311(308.2) 2360(586.5)
Banned Drug	202 202	Poisson Geometric	697 (99.8) 1122 (165.3)	842(243.4) 1554(514.3)	3116(1264.9) 6030(2528.6)	3367(1332.3) 6296(2647.2)	2006(735) 3751(1463.9)

Results comparing MLE, Turing, Chao and Zelterman estimators of the kind offense committed and the weighted estimator

4.1 Discussions of Results

Chao's lower bound, Turing and maximum likelihood estimators are some of the most qualified estimators used to estimate hidden population size in capture-recapture experiments. Turing and maximum likelihood estimations are developed under the Poisson homogeneity assumption whereas Chao's lower bound is developed allowing heterogeneity.

Coming back to the estimation of fake drug syndicates, data on Table II shows that 1434 syndicates were observed. Using zero-truncated Poisson model the estimators MLE, Turing, Chao and Zelterman yielded

2108, 4218, 9745, and 10243 respectively. But because the estimators have different results we consolidated their weights into one. The weighted estimator for Poisson gives the population size of fake syndicates as 6579. As 1434 were observed, this suggest to mean that only about 22% of the population of fake drug syndicates in Onitsha is observed with 95% confidence interval of 19% –25%. In similar way, the weighted estimator for geometric is 12649 with only about 11% of the population being observed with **95%**confidence interval of 10% –13%. We also looked at the most widely committed offense of fake drug syndicates. Data on Table IV shows that falsifying of genuine drug is more frequently committed offense, followed by selling of expired drugs.

6. Conclusion:

In conclusion, only 22% and 11% of the population of fake drug syndicates was observed under Poisson and geometric respectively is not good news for NAFDAC. It shows that the agency is not winning the war against fake drug circulation in Nigeria, especially in Onitsha drug markets. Consequently, NAFDAC officials should therefore redouble their efforts in the fight against the scourge, at least to reduce it to the barest minimum. More importantly, proper and regular record keeping of syndicate's arrest by NAFDAC is advocated, at least for the agency to have empirical facts and statistical evidence to assess their performance, not by mere newspaper publications and press releases. Unfortunately, since after the NAFDAC report on fake drug syndicates (FDSs) from January 2011—December 2011 (which was extracted from NAFDAC Quarterly Magazine Vol.2 No.2 2011) no other report was released by the agency. In United Kingdom, offense of drunk-driving (DD) is recorded by the Driver and Vehicle Licensing Agency (DVLA), In [2]; and in Bangkok, Office of the Narcotic Control Board (ONCB) has record of methamphetamine in [5]; this is to mention only but a few. These databases enabled the agencies to access their performance, we therefore suggest to NAFDAC to follow suit immediately.

Funding

The funding of this work was done by the three authors.

Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Ugochukwu. A. Osisiogu, Theresa. E. Efor and Chinwuba. E. Emeka. The first draft of the manuscript was written by Chinwuba. E. Emeka and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

1. Smite F, Reinking D, Reijerse M. (2002). Estimating the number of people eligible for healthservice. Evalprog plan , 25:101-105
2. Bohning and Peter G.M van der Heijden. (2019). The Identity of the zero-truncated, one- inflated likelihood and the zero-one-truncated likelihood for general count densities with an application to drink-driving in Britain. *Annals of Applied Statistics Vol. 13. No 2, 1198-1211*
3. Martin B., Owen G., and Karine, D. (2010). Estimation of the size of the illicit methamphetamine and MDMA markets in Canada. *A discussion paper on potential methods and sources*
4. DankmarBohning (2008). A simple variance formula for population size estimators by conditioning. *Statistical Methodology (5) 410-423*

5. PanichaKaskasamku (2018). Capture-recapture Estimation and Modeling for one-Infected count data. PhD thesis, University of Southampton

6. Casella, G. and Berger, R (2008). Statistical Inference, *2nd Edition California: Brooks/Cole Publishing Company*

7. Chao, A., and Lee, S. (1992). Estimating the Number of Classes via Sample Coverage. *Journal of the American Statistical Association, 87: 210-217*

8. Meng, X. L (1997). The EM algorithm and medical studies: A historical link. *Statistical methods in medical research, 6: 3-23*

9. Bohning, D., Dietz, E., Kuhnert, R., and Schon, D. (2005). Mixture models for capture-recapture count data. *Statistical Methods and Applications, 14:29-43.*

10. Viwatwongkasam, C., Kuhnert, R., and Satitvipawee, P. (2008). A comparison of population size estimators under truncated model with or without allowance for contamination. Biometrical Journal, 50(6):1006-1021

11. Niwitpong, S., Bohning, B., van der Heijden, P., and Holling, H. (2013). Capture-recapture Estimation based upon the geometric distribution allowing for heterogeneity. Metrika, 76:495-519

12. Chao, A. (1987). Estimating the Population Size for Capture-recapture Data with Unequal Catchability. *Biometrics, 43(4): 783-791*

13. DankmarBohning and Peter G.M.H. (2009). A Covariate Adjustment for Zero-truncated approaches to Estimating the Size of Hidden and Elusive Populations. *The Annals of Applied Statistics. Vol. 3 No. 2. 599-610*

14. Anan, O. (2016). Capture-recapture Modeling for Zero-truncated Count Data Allowing for Heterogeneity. *PhD thesis, University of Southampton*