International Journal of Innovative Research and Advanced Studies (IJIRAS)
Volume 6 Issue 3, March 2019

ISSN: 2394-4404

Validity Of Capture-Recapture Methods In Estimating Population Size Of Fake Drug Syndicates In Nigeria

Chinwuba E. Emeka

Department of Mathematics & Computer Science,
Ebonyi State University Abakaliki, Nigeria

***Abstract:** Capture-recapture method of analysis is a new phenomenon for Nigerian researchers, especially for researchers in the areas of public health, illicit and counterfeit drugs, and our attention drawn on its usage to estimate the population size of fake drug syndicates in Nigerian. How valid is the use of capture-recapture methods in estimating the population size of fake drug syndicates in Nigeria from NAFDAC records is the object of this study. NAFDAC, National Agency for Food and Drug Administration and Control, is an agency responsible for checkmating counterfeit drugs in Nigeria. Relying solely on capture-recapture methods of analysis without cross-validating its results may be misleading. This we believe could happen especially when the underlying assumptions validating the general use of capture-recapture (CR) methods are violated. Truncated models are therefore used to cross-validate its usage. The traditional CR analyzes data from two or more sources, while truncated model is used to analyze count or frequency data from a single or multiple sources. While some authors have used CR to estimate the population size of fake drug syndicates from NAFDAC records, we however apply truncated models to their data to validate the use of CR in estimate this difficult-to-reach population. Overall, they identified 542 cases of which 440 were from NAFDAC Onitsha zone, 270 were from at least 2 other zones of Lagos and Kano, and 136 were common to all the three zones. The sample coverage of CR estimate $\hat{N}$ is 560. We cross-validate the estimate on the same data with Zelterman, Chao and Binomial truncated models. Respectively they yielded 542, 577 and 559. These models showed an appreciable degree of unison with sample coverage of CR employed by other researchers. Falsified drugs were 336, while 206 were at least on expired, unregistered and banned drugs. Falsified, expired, unregistered and banned drugs were analyzed for internal validity. Their sample coverage of CR estimate yielded 346, 58, 96 and 60 respectively, while our corresponding Zelterman truncated model estimate yielded 336, 56, 92 and 58 respectively; Chao yielded 357, 60, 98 and 62 respectively; and Binomial yielded 350, 60, 98 and 60 respectively. Internal validity holds as no major discrepancy existed between the two methods, truncated models and the sample coverage. We therefore conclude that CR method is valid technique for the estimate of fake drug syndicates.*

***Keywords:** Fake drug, capture-recapture, sample coverage, truncated models*

I. INTRODUCTION

National Agency for Food and Drug Administration and Control (NAFDAC), an agency responsible for checkmating illicit and counterfeit drugs in Nigeria has allayed fears on the problem imposed on the Nigerian Healthcare system by fake drugs. As (Chinwendu, 2008) remarked; "the result of fake drug proliferation has led to treatment failures, organ dysfunction or damage, worsening of chronic disease conditions and the death of many Nigerians". And (Akunyili, 2004) added, "The situation became so bad that even when patients were treated with genuine drugs, there is no response due to resistance caused by previous intake of fake drugs".

The health problems associated with the consumption of fake and counterfeit drugs cannot be over stated (Osisiogu & Chinwuba, 2019). The authors stated that administering counterfeit drugs to patients has led to drug resistance, abuse or even death. Fake, adulterated and substandard drugs as (NAFDAC. 2007) reported, resulted to the death of over 150 children as a result of paracetamol syrup containing diethylene glycol in 2001 alone. While (Chinwendu, 2008) expressed fears that the problem of fake drugs was so severe that neighbouring countries such as Ghana and Serra, some time ago, officially banned the sale of drugs made in Nigeria; adding that the issue of fake drugs went to the extent that drugs were hawked even in commercial buses, open markets and streets.

In 1988, World Health Organization (WHO) Assembly urged countries all over the world to help combat the global health threat by counterfeit pharmaceuticals (Osisiogu & Chinwuba, 2019). The authors stated that WHO remarked that the prevalence of fake medicine is higher in countries with weak regulations, enforcement, and scarcity of supply of basic medicines, unregistered markets and unaffordable prices. While (Chinwendu, 2008) remarked that the high demand for medicines and low cost of production prompts counterfeiters to continue because adequate drug deterrent legislation is lacking. He said, “Around 70% of drugs in Nigeria are imported, and India is a major exporter of these drugs”, stating that “Some Nigerian importers connived with some Indian manufacturers to produce fake and substandard drugs at a cheap rate with less active ingredient and sold at a cheaper rate”. In order to checkmate the proliferation of fake drugs, NAFDAC therefore insisted that all drugs sold in Nigeria must have NAFDAC registration number; otherwise such drug is fake.

While the Nigerian government has been at war with fake drugs since at least the early 1980s, concern about fake drugs became especially intense in 2001, when NAFDAC started a toughest war against fake drugs. Seizure of fake drugs and arrests for fake drug possessions and sale skyrocketed. Media attention to fake drug increased and public increasingly saw it as one of the greatest problem facing Nigerian health sector. Concern about fake drug peaked in late 2005 and 2006, when NAFDAC established a special zone at the commercial city of Onitsha, where the inflow of fake drugs was high. But to us this is only part of the story; those who import the drugs, those who distribute them to various shops in Nigeria and those who market them on the streets, buses, and at the overhead bridges (we called syndicates) are also major contributors to the circulation of fake drugs. These syndicates are many and until something very serious is done to curtail its excess the fight against fake drug may be a fight in futility. The syndicates have no visibility of location, as they migrate from one place to another, especially in rural areas, looking for a place where they can sell their products without NAFDAC molestation whatsoever. Unfortunately, due to its illegality and criminalization it becomes really a big challenge to determine the population size of them by enumeration or any standard sampling technique. In other words, this hidden population requires a special technique to estimate its size. (Osisiogu & Chinwuba, 2019) looked at the dangers impose on the Nigerian health sector by these syndicates and proposed a technique known as capture-recapture analysis to estimate the population size of the syndicates of these fake drugs for the health officials and NAFDAC to visualize.

Capture-recapture (CR) method of analysis was used to estimate hidden or partial hidden populations, (McKegancy et al. 1992; Fisher et al. 1992; Fisher et al. 1994; Squires et al. 1995; Aaron et al. 2002). It was first used to estimate animal abundance (Amstrup, 2005) before it was recently applied to epidemiological studies (Post, 2013). In Europe, for instance, capture-recapture was the recommended method for estimating the population size of illicit drug users and in UK capture-recapture was used to monitor the effectiveness of drug policy (Jones et al. 2013). In non-academics, U.S government applied capture-recapture to control census undercount, (Nanan DJ and White F, 1997), while NASA used the technique to count the number of stars in the universe, and British Society of Statistics used this methodology to estimate the size of the World Wide Web (Fienberg and Stephen, 1998).

II. BACKGROUND OF STUDY

NAFDAC, the National Agency for Food and Drug Administration and Control is a Nigerian government agency responsible for regulating and controlling the manufacture, importation, exportation, advertisement, distribution, sale and use of food, drugs, cosmetics, medical devices, chemicals and pre-packed water (Osisiogu & Chinwuba, 2019). According to them, the Agency was created following the World Health Organization Assembly resolution in the 1988 that countries all over the world should help in combating the global health threat posed by counterfeit pharmaceutical; and amidst growing concerns about the problem of fake and poorly regulated drugs circulating in Nigeria markets. They also noted that the Agency was created in 1994 and had its offices in the six geopolitical zones and the 36 state of Nigeria. As remarked by (Chinwendu, 2008), India is a major drug exporter to Nigeria. According to Chinwendu, these drugs are imported into Nigeria, sold to the wholesalers and retailers who may or may not know if these drugs are fakes or not. (Osisiogu and Chinwuba, 2019) reported that surveys on prevalence rate of fake drug at Onitsha market alone stands at 30% as against 10% in other parts of the country. They noted however that Crude method (aggregated cases divided by the observed population) was used to estimate the prevalence. Source of data collection was based on NAFDAC records of fake drug syndicates. But they use the state-of-art method called capture-recapture to estimate the population size of syndicates of these fake drugs. In our own case however we use truncated models. (Bohning et al. 2004) noted that conventional capture-recapture technique involves two sources (e.g., hospital and police) or three sources (e.g., treatment centre, survey and family doctors). As earlier noted by (Osisiogu and Chinwuba, 2019), NAFDAC has different offices in 36 states plus Abuja, the Nation’s capital.

The choice of Onitsha, Lagos and Kano NAFDAC zonal offices, according to Osisiogu and Chinwuba was because these zones have heavy flow of fake drugs. They noted that this was to reduce the number of data sources which may result to increasing variation of estimates as (Van Hest et al, 2007) envisaged. An alternative to conventional capture-recapture is the truncated model. Truncated models were used on count data (Rob van Hest, 2007). We use unique identifier such as demographic information of the syndicates to identify who the syndicates are and how many times they have been apprehended (repeated entries), (Dankmar , 2004).

The aim of this research is to reexamine the data used by (Osisiogu and Chinwuba, 2019) to estimate fake drug syndicates with various truncated models for the purpose of cross-validating their use of CR to estimate population size of fake drug syndicates.

Truncated model is based on a single source data which makes it less dependent on matching entries from different sources (Dankmar, 2004). (Van Hest et al, 2007) used this idea to validate the use of capturer-recapture analysis in estimating infectious disease from different sources. If there is no discrepancy between the two approaches then the estimate of capture-recapture on fake drug syndicates from NAFDAC records by them is valid. But if there is disagreement, we can be sure that NAFDAC records alone is not sufficient enough to be used to estimate the population size of fake drug syndicates in Nigeria. This study will therefore help Nigerian health workers and epidemiologists adopt a specific methodology for use in estimating the population size of fake drugs syndicates in Nigeria.

III. MATERIALS AND METHOD OF DATA ANALYSIS

A. MATERIALS FOR DATA ANALYSIS

(Osisiogu and Chinwuba, 2019) extracted fake drug syndicates from NAFDAC records from January 2015 to December, 2015. In its newsletters (Consumer Safety Bulletins, Quarterly Magazines, websites) and in its press releases, NAFDAC published all the counterfeit drugs confiscated in various drug markets in Nigeria. The report also contained all violations and kind of offenses committed and arrest made during the period. In the literature we noted that the Agency was created in 1994 and it has offices in the six geopolitical zones and the 36 state of Nigeria. However, since the inflow of fake drugs are usually heavy at the commercial cities of Onitsha in Eastern Nigeria, Lagos in Western and Kano Northern Nigeria, (Osisiogu and Chinwuba, 2019) collapsed the multiple offices into three. (Rob van Hest, 2007) pointed out that, it is neither practical to have as many data sources as possible because of budgetary constraint, and too increasing number of source causes decreasing overlap which may result increase in variation of estimates, and cells in the multi-way contingency table may even contain zero cases.

Some of the syndicates after granting a bail by a court of law may migrate to other zones to commit the same offense as (Osisiogu and Chinwuba, 2019) noted. They also noted that some of them were rearrested the second or even the third times. They therefore constructed capture-recapture variables by counting the number of times a syndicate was arrested. The subscript Z_{111} showed the number of syndicates arrested in Onitsha, Lagos and Kano, as they noted. While Z_{110} means the number of syndicates arrested in Onitsha and Lagos but not in Kano. The number of fake drug syndicates not arrested in any of the three zones is indicated by Z_{000}. As shown in Table 1, they use the nature of offenses committed to construct capture history of fake drug syndicates. But because of the type of offenses committed, suspects arrested were stratified into four categories namely falsified, expired, unregistered and banned drugs as can be seen in Table 2. For data in Table 1, they counted 84 syndicates arrested in the Onitsha zone only, 44 syndicates arrested in Lagos only and 8 in Kano only. 130 syndicates were both arrested in Onitsha and Lagos but not in Kano. Similarly we interpret other records.

Onitsha zone (list 1)	Lagos zone (list 2)	Kano zone (list 3)	No. of arrest (Z_{ijk})
1	1	1	Z_{111}= 136
1	1	0	Z_{110} = 130
1	0	1	Z_{101} = 90
1	0	0	Z_{100} = 84
0	1	1	Z_{011} = 50
0	1	0	Z_{010} = 44
0	0	1	Z_{001}= 8
0	0	0	Z_{000} = ?

Source: (Osisiogu& Chinwuba, 2019).1 represents arrest, 0 represents no arrest

Table 1: Observed cases of fake drug syndicates

a. SOURCES OF CASE AND RECORD-LINKAGE

Three sources were used to identity syndicates arrested from January-December, 2015(Osisiogu & Chinwuba, 2019).The first source was the arrest made by the officials of NAFDAC Onitsha special zone. The second source was the arrest made by staff of NAFDAC Lagos zone and third was Kano zone. The raids were made independently. Cases in various lists were merged and after correction for duplicate entries with the aid of 'excel software', the records of syndicate arrest were matched by a deterministic linkage procedure using identifiers such as full name of the syndicate, proximity of dates, address, geographical and demographic information. Overall, we identified 542 cases of arrest as presented in Fig 1 of which 440 were from Onitsha, 270 were through at least 2 other zones and 136 were common to all the three zones. Arrests were also stratified to distinguish offenses as presented in Table 2. Arrest made on falsified drugs was 336 while 206 arrests were made on other offenses combined.

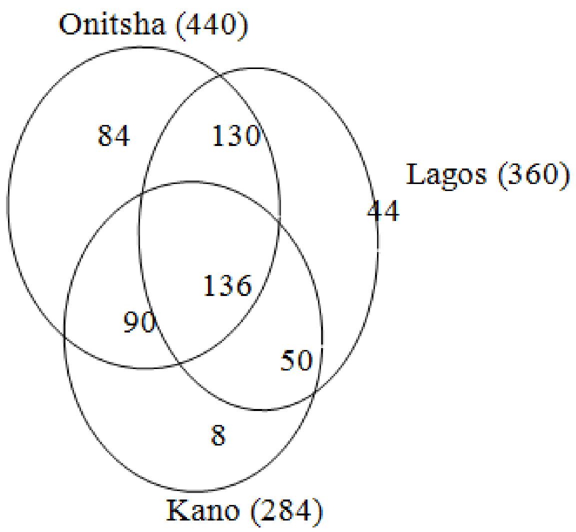

Figure 1: Vann diagram showing distribution of fake drug syndicates arrested in the investigated sources

Kind of Offense	Place of Arrest List1 (100)	List2 (010)	List3 (001)	List1&2 (110)	List1&3 (101)	List2&3 (011)	All list (111)	No. of Arrest
Falsified	56	25	3	81	59	28	84	336
Expired	9	4	1	13	10	5	14	56
Unregtd	11	10	2	21	13	12	23	92
Banned	58	5	2	15	8	5	15	58

List1=Onitsha zone; List2 =Lagos zone &List3=Kano zone

Table 2: No. of fake drug syndicates stratified by kind of offense

B. METHODS OF DATA ANALYSIS

"There exists a general belief that one knows something only when it has been counted" (Douglas, 1967). Numbers are increasingly involved in understanding and evaluating intersection of the social construction of drug problems (Himmelstein, 2013). (Osisiogu & Chinwuba, 2019) extracted fake drug syndicates as contained in NAFDAC News Magazines, NAFDAC Consumer Safety Bulletins, NAFDAC Public Alert notices, NAFDAC press releases and newsletters, and NAFDAC websites and so forth, and used CR on the data generated to estimate the population size of fake drug syndicates.

Assuming that the true population size of fake drug syndicates is N as (Osisiogu and Chinwuba, 2019) noted which may be indexed 1, 2,…., N; and suppose the observed cases of fake drug syndicates arrested is M. Then N-M is the number of syndicates not arrested. These individuals have capture history Z_{000} as shown in Table 1. Let $Z_{s1,\, s2,\ldots,st}$ be the number of syndicates with records s_1, s_2, …, s_t, where $s_j = 0$ denotes absence in sample zone j and $s_j = 1$ denotes presence in sample zone j. For t = 3, there would be seven observed cells of arrested cases namely Z_{001}, Z_{010}, Z_{011}, Z_{100}, Z_{101}, Z_{110} and Z_{111}, where Z_{001} is the number of syndicates arrested in Kano zone only, Z_{110} is the number of syndicates arrested in Onitsha and Lagos zones but not in Kano. A similar interpretation follows other capture histories. Thus syndicates not arrested (missed) has cell Z_{000} = N-M. This is equivalent to predicating no arrest in all the three zones (i.e., Z_{000} = N-M). When we add over a sample zone, the subscript corresponding to that zone is replaced by a "+" sign (Chao et al, 2001). For example $Z_{+11} = Z_{011} + Z_{111}$ and $Z_{++1} = Z_{001} + Z_{011} + Z_{111}$ and Z_{+++} = N. Also $Z_{1+1} = Z_{101} + Z_{111}$ and $Z_{11+} = Z_{110} + Z_{111}$. If we let n_j j = 1, 2, ---,t be the number of individuals arrested in sample zone$j$. For t =3, we have $n_3 = Z_{++1} = Z_{001} + Z_{011} + Z_{101} + Z_{111}$ and $n_2 = Z_{+1+} = Z_{010} + Z_{110} + Z_{011} + Z_{111}$ while $n_1 = Z_{1++} = Z_{100} + Z_{110} + Z_{101} + Z_{111}$

a. *SAMPLE COVERAGE APPROACH*

Sample coverage approach of capture-recapture was developed by (Chao et al, 2001) to estimate N. The concept of sample coverage was originally proposed by (Turing and Good, 1953) but was purified by (Chao et al, 2001). The basic idea is that the sample coverage can as well estimate the presence of two types of dependencies. Thus an estimate of population size can be derived via the relationship between population size and the sample coverage. Estimators of sample coverage as seen in (Chao A., Tsay PK, 1998) valid for this study are:

$$\hat{C} = 1 - \frac{1}{3}\left(\frac{Z_{100}}{n_1} + \frac{Z_{010}}{n_2} + \frac{Z_{001}}{n_3}\right) \qquad (1)$$

which is the average (over three sample zones) of the fraction of cases found more than once. Z_{100}, Z_{010} and Z_{001} are the number of individuals arrested only in one sample zone, and thus have no information about overlap; while n_1, n_2 and n_3 are the number of identified cases of arrest in each sample zone. They are called independent sources. The following estimators seen in (Chao et al, 2001) are applied:

$$D = \frac{1}{3}\left[(M - Z_{100}) + (M - Z_{010}) + (M - Z_{01})\right] \quad (2)$$

where $(Z_{100} + Z_{010} + Z_{001})/3$ represents the average of the non-overlapped cases and M denotes the total number of identified cases of arrest. Thus D can be interpreted as the average of the overlapped cases of arrest. When arrest in the three sample zones is independent, a sample population size estimator is derived as

$$\hat{N}_0 = D / \hat{C} \qquad (3)$$

(Chao et al, 2001) noted that when dependence exists among the zones and the overlap information is large enough then

$$\hat{N} = \left(\frac{Z_{+11} + Z_{1+1} + Z_{11+}}{3\hat{C}}\right) \Big/ \left\{ \begin{array}{c} 1 - \frac{1}{3\hat{C}}\left(\frac{(Z_{1+0} + Z_{+10})Z_{11+}}{n_1 n_2}\right) + \\ \frac{(Z_{10+} + Z_{+01})Z_{1+1}}{n_1 n_3} + \frac{(Z_{0+1} + Z_{01+})Z_{+11}}{n_2 n_3} \end{array} \right\} \quad (4)$$

They also noted that for relatively low sample coverage data where information about the syndicates is not sufficient enough to accurately estimate the population size of fake drug syndicates, the following estimators may apply:

$$\hat{N}_1 = \frac{D}{\hat{C}} + \frac{1}{3\hat{C}} \left[\begin{array}{c} (Z_{1+0} + Z_{+10})\left(\frac{DZ_{11+}}{\hat{C} n_1 n_2} - 1\right) + (Z_{10+} + Z_{+01}) \\ \left(\frac{DZ_{1+1}}{\hat{C} n_1 n_2} - 1\right) \\ + (Z_{01+} + Z_{0+1})\left(\frac{DZ_{+11}}{\hat{C} n_2 n_2} - 1\right) \end{array} \right] \quad (5)$$

Simulation studies by (Chao et al, 1996) have suggested that the estimated sample coverage should be at least 55 percent to adequately estimate any population.

We prefer sample coverage to log-linear model because of two inherent advantages over log-linear; i.e., no model selection or model comparison is needed and no further difficulty arises when the number of lists increases overlap information.

C. TRUNCATED MODEL

In epidemiological studies, violation underlying capture-recapture assumptions is unavoidable. This and other limitations call for cross-validation. Alternative models related to capture-recapture analysis have been suggested by (Van Hest et al, 2007). As stated in the literature, truncated models are used to cross-validate the use of CR. If a suspect is arrested and later released, there is likelihood he may be rearrested if he goes back to the same illegal business again. The number of such arrests helped estimate the total number of syndicates. Consider for example a population size of fake drug syndicates to be N and suppose a count variable Y taking values in the set of integers {0, 1, 2, 3, ---} is the number of suspects arrested. Also denote for example f_0, f_1, f_2, ---- the frequency with which a 0, 1, 2--- occurs in the population.

Since a syndicate is only observed if he is arrested, y = 0 will not be observed in the list. Hence the list reflects a count variable truncated at zero which we shall denote by Y. Accordingly, the list has observed frequencies f_1, f_2, ---, but the frequency f_0 of zeros in the population is unknown. Since we do not know f_0, we form the zero-truncated Poisson distribution by defining a conditional probability function $y > 0$, as

$$P(y_i / y_i > 0, \lambda) = \frac{P(y_i / \lambda)}{P(y_i > 0 / \lambda)} = \frac{\exp(-\lambda)\lambda^{y_i}}{y_i!(1-\exp(-\lambda))} \quad (6)$$

with $p(y_i > 0/\lambda)\ 1-\exp(-\lambda), \quad i = 1, \cdots, N_{obs}$.

(Van der Heijden et al, 2003) noted that if an estimator for λ is $\hat{\lambda}$, then the probability of an individual not arrested (unobserved) shall be $\hat{p}_0 = \exp(-\hat{\lambda})$. Thus, the number of unobserved individual denoted by $\hat{f}_0$ can then be calculated as

$$\hat{f}_0 = [\hat{p}_0/(1-\hat{p}_0)]N_{obs} \quad (7)$$

where N_{obs} is the observed number of individual in the population. The estimated population size $\hat{N}$ is then obtained by

$$\hat{N} = \hat{f}_0 + N_{obs} \quad (8)$$

However, Zelterman and Chao have derived a simpler estimator for truncated model. In most cases, λ is known but if it is not it can be estimated. They said, with maximum likelihood under the assumption that Poisson distribution is homogeneous, λ can be estimated. However, instead of estimating λ under the assumption of a homogeneous Poisson distribution (Zelterman, 1988) argued that the Poisson assumption might not be valid over the entire range of possible values for Y but it might be valid for small ranges of Y such as from y to y_{+1}, so that it would be meaningful to use only the frequencies f_i and f_{i+1} in estimating λ. Since for any i both the truncated and un-truncated Poisson distribution have the property that

$P_0(i+1|\lambda)/P_0(i|\lambda) = \lambda/(i+1)$ and

$P_{0+}(i+1|\lambda)/P_{0+}(i|N) = \lambda/(i+1)$

respectively, λ can be derived as

$$\hat{\lambda} = \frac{(i+1)P_0(i+1|\lambda)}{P_0(i|\lambda)} = \frac{(i+1)P_{0+}(i+1|\lambda)}{P_{0+}(i|\lambda)} \quad (9)$$

An estimator for λ is obtained by replacing $P_{0+}(i|\lambda)$ by the empirical frequency f_i:

$$\hat{\lambda}_i = (i+1) f_{i+1}/f_i \quad (10)$$

If $i = 1$, $\hat{\lambda}_1 = 2f_2/f_1$, and this estimator is often considered for two reasons; $\hat{\lambda}_1$is using frequencies in the vicinity of f_0 which is the target of prediction and in many application studies for estimating f_1 and f_2. As Zelterman noted, this estimator is unaffected by changes in the data for counts larger than 2, which contributes largely to its robustness. Various truncated model that will apply are:

Truncated binomial model:

$$est(N) = obs(N) + (f_1)^2/3f_2 \quad (11)$$

Truncated Poisson mixture (Zelterman) model:

$$est(N) = obs(N)/[1-\exp(-2f_2/f_1)] \quad (12)$$

Truncated Poisson heterogeneity (Chao) model:

$$est(N) = obs(N)+(f_1)^2/2f_2 \quad (13)$$

Out of many possible methods, we have chosen the above combinations of truncated models because according to (Hook & Regal, 1982; Hook & Regal, 1995) they are alternative to capture-recapture methods.

The ratio between the number of syndicate arrested once (f_1) and twice (f_2) plays an important role in the use of truncated models. When '1' represents arrest and '0' no arrest and the three linked zones are used, frequency count f_1 is the sum of the cells 100, 010, and 001 in the 2x2x2 contingency table and frequency count f_2 corresponds to the sum of the cells 110, 101, and 011. Similarly, syndicates arrested in all the three zones, f_3 are donated as 111. We use the f_1/f_2 ratio to examine a possible relationship between this ratio and the performance of the truncated models. In Table 4 were sample coverage estimates while truncated model estimates were in Table 5. Comparison of the two approaches was in Table 6.

IV. RESULTS OF DATA ANALYSES

Table 3 below depicts the results of sample coverage as $\hat{C}$ = 88.6 per cent and D = 496.667 (see eqns.1 &2). An estimate without possible dependency is $\hat{N}_0$ = 561 (see eqn.3). Estimate when the source is dependent and sample coverage is adequate is $\hat{N}$=560 (see eqn.4) while the estimate when the source is dependent and sample coverage is inadequate is $\hat{N}_1$ is 547 (see eq. 5).

	M	D	$\hat{C}$	est	cil	ciu
$\hat{N}_0$	542	496.667	0.886	561	515	608
$\hat{N}$	542	496.667	0.886	560	515	608
$\hat{N}_1$	542	496.667	0.886	547	502	595

sample coverage =88.6% which is adequate (55% or more is adequate)

Table 3: Unstratified estimate of fake drug Syndicates

DEFINITIONS: We adopt similar notations used by (Chao et al, 2001)

M: number of individuals arrested in at least one sample zone

D: the average number of individuals arrested in at least one sample zone

$\hat{C}$: sample average estimate see eq. (1)

est: population size estimate of fake drug syndicates

se: estimated standard error of the population size estimation of fake drug syndicates

cil: 95% confidence interval lower limit

ciu: 95% confidence interval upper limit

$\hat{N}_0$: population size estimate of fake drug syndicates for independent sample zone see eq. (3)

$\hat{N}$: population size estimate for sufficiently high sample coverage cases see eq. (4)

$\hat{N}_1$: one-step population size estimate for low sample coverage cases see eq. (5)

ISSN: 2394-4404

Type of arrest	If source is dependent and sample coverage is adequate, estimates ($\hat{N}$) in eq.4 are:	If source is dependent but sample coverage is inadequate, estimates ($\hat{N}_1$) in eq.5 are:
Falsified	346 (95%CI = 310-384)	338 (95%CI = 303–376)
Expired	58 (95%CI = 44-75)	57 (95%CI = 43 – 74)
Unregist	96 (95%CI =78-117)	93 (95%CI = 75 – 114)
Banned	60 (95%CI = 46-77)	60 (95%CI = 75 – 114)
*Fake	560(95%CI = 515-608)	547(95%CI = 515 – 608)

**Fake in general terms means of falsified, expired, unregistered and banned drugs; Estimates (N) = estimated population*

Table 4: Estimate of Fake drugs Syndicates by Sample-Coverage Method

Type of Arrest	Obs (N)	Zelterman model est(N) eq.12	% Obs (N)/ est(N)	Chao model est(N) eq.13	% Obs(N) /est(N)	Binomial model est(N) see eq.11	% Obs(N)/ est(N)
Falsified	336	336	100	357	94.1	350	96.0
Expired	56	56	100	60	93.3	60	93.3
Unregistd	92	92	100	98	93.9	98	93.9
Banned	58	58	100	62	93.5	60	96.7
Fake	42	542	100	577	93.9	559	97.0

Obs(N) = observed number of cases, est(N) = estimated number of case

Table 5: Estimates of fake drug syndicates by truncated models and percentage

Type of analysis	Fake	Type of offense: Falsified	Expired	Unregistered	Banned
Sample coverage method from eqn. 1& 2	560 (547)	346 (338)	58 (57)	96 (93)	60 (60)
Truncated binomial model est(N) = obs(N) + $(f_1)^2/3f_2$	559	350	60	98	60
Truncated Poisson mixture model est(N) = obs(N)/[1-exp(-$2f_2/f_1$)]	542	336	56	92	58
Poisson heterogeneity model(Chao) est(N) = obs(N) + $(f_1)^2/2f_2$	577	357	60	98	62
f_1/f_2	0.5	0.5	0.5	0.5	0.54

Obs(N) = observed number of cases, est(N) = estimated number of case; () = estimate by eqn. 2

Table 6: Comparison of estimates by sample coverage method and Poisson mixture model (Zelterman), truncated Poisson heterogeneity model (Chao) and truncated Binomial model

V. DISCUSSION

A. DISCUSSION

Assumptions underlying the use of capture-recapture in epidemiological studies include; (1) the population must be closed, (2) the capture sources must be independent, (3) all members of the population must have equal chance of being in the list and the capture history of each member must be accurate.

The closure assumption can only be reached to a reasonable extent, especially when the study period is short. In this study, NAFDAC records of fake drug syndicates were extracted from January to December, 2015. The second assumption which is source independent is also difficult to satisfy, but the source dependency is relaxed when sample coverage approach is employed. The third assumption that all members of the population shall have equal chance of being in the list is easy to satisfy. This is because all the syndicates faced the same risk of being arrested. The fourth assumption is that the capture history of each member shall be accurate; i.e., all true matches only are identified. Fake drug syndicates are matched by proximity of dates, address, name of the syndicates, and geographical information. Furthermore, we stratified them as falsified, unregistered, banned and expired drugs for internal validity.

Though the assumptions of CR to large extent are satisfied we still use truncated models to verify its validity in case some of the assumptions are violated. Overall, we identified 542 cases of which 440 were from Onitsha zone, 270 were from at least 2 other zones of Lagos and Kano and 136 were common to all the three zones (See fig 1). The sample coverage of CR estimate $\hat{N}$ is 560 (See Tables 3&6). We cross-validate the estimate with Zelterman truncated Poisson mixture model, Chao truncated Poisson heterogeneity model and Binomial truncated models. Respectively they yielded 542, 577 and 559 (Also see Table 5 & 6). These models showed an appreciable degree of unison with sample coverage. Falsified drugs were 336 while 206 were at least on expired, unregistered and banned drugs. Falsified, expired, unregistered and banned drugs were analyzed for internal validity. Their sample coverage estimate yielded 346, 58, 96 and 60 respectively (see Table 4), while corresponding Zelterman estimate yielded 336, 56, 92 and 58 respectively; Chao yielded 357, 60, 98 and 62 respectively; and Binomial yielded 350, 60, 98 and 60 respectively (see Table 5). The internal validity also holds as no major discrepancy existed between truncated models and the sample coverage, f_1/f_2 ratio is also same in the estimators, thus showing perfect agreement.

VI. CONCLUSION

The population size of fake drug syndicates by capture-recapture analysis requires adequate data base register. In capture-recapture analysis small variations in the quality of data and record-linkage can lead to highly variable outcomes. NAFDAC News magazine, newsletters etc., may not contain all the fake drug syndicates and not including them may lead to underestimation. Truncated models were then used as a heuristic tool to identify possible failure in capture-recapture analysis models. Since there is no major discrepancies between CR analysis and truncated models, capture-recapture stands as a good statistical stool to estimate the size of fake drug syndicates in Nigeria.

ACKNOWLEDGMENTS

The author is deeply grateful to Professor U.A Osisiogu for proofreading the article and Professor Walford Chukwu for going through the analyses. U.A Osisiogu (Professor of Mathematics) was a lecturer at the Department of Mathematics & Computer Science, Ebonyi State University, Abakaliki, Nigeria. While W. Chukwu (Professor of Statistics) was lecturer at the Department of Statistics, University of Nigeria, Nsukka, Nigeria.

REFERENCES

[1] Aaron DJ, Chang Y.F, Markovic N, LaPorte RE. Estimating the lesbian population: a capture-recapture approach. Journal of epidemiology community health, 2003; 52, 207-209.

[2] Akunyili DN. Fake drug and counterfeit drugs in the health sector. Annals of Ibadan Postgraduate Medicine 2004; Vol. 2 No. 2.

[3] Amstrup S.C., McDonald T.L., Manly B.F.J. Handbook of Capture-recapture Analysis. Princeton Univ. Press; 2005

[4] Bohning D., Suppawattanadodee B., Kusolvisitkul W., Viwatwongkasem C. Estimating the number of drug users in Bangkok 2001: Capture-recapture approach using repeated entries in one list. European Journal of Epidemiology 2004; 19: 1075-1083.

[5] Chao A., Tsay PK. A sample coverage approach to multiple-system estimation with application to census undercount. Journal of the American Statistical Association 1998; 93: 283-293.

[6] Chao A., Tsay PK., Shau WY., Chao DY. Population size estimation for capture-recapture models with applications to epidemiological data. Proceedings of Biometrics Section, American Statistical Association .,1996; 108-117.

[7] Chao A., Tsay P.K., Sheng-Hsiang Lin., Wen-Yi Shau and Day-Yu Chao. Tutorial in Biostatistics. The applications of capture-recapture models to epidemiological data. Statist. Med.2001; 20:3123-3157.

[8] Chinwendu. The fight against fake drugs by NAFDAC in Nigeria 44th International Course in Health Development (ICHD) September 24, 2007-September 12, 2008. Royal Tropical Institute Development Policy & Practical/ Vrije Uniersiteit Amsterdam. 2008

[9] Dankmar B., Busaba S., Wilai K., Chukiat V.. Estimating the number of drug users in Bangkok 2001: A capture-recapture approach using repeated entries in one list. European Journal of epidemiology. 2004; 19: 1075-1083.

[10] Douglas, J.D. The social meaning of suicide. Princeton, NJ: Princeton University Press; 1967

[11] Fienberg and Stephen E. Innovative methodology for capture-recapture studies. Workshop notebook from the North Eastern Biometrics Conference, Pittsburg, PA 1998: March 31.

[12] Fisher N., Turner SW., and Pugh R. Estimating numbers of homeless and homeless mentally ill people in North East Westminster by using capture-recapture analysis. British medical journal1994; 308, 27-30.

[13] Himmelstein J L. Hidden choices and missing

[14] numbers: Selecting drug numbers for media attention.

[15] Journal of drug Issues 2013; 43(4) 437-449.

[16] Hook B.E., Regal R.R.Validity of Bernoulli census, log-linear and truncated binomial models for correcting underestimates in prevalence studies. American journal of epidemiology 1982; 116: 168-176.

[17] Hook B.E., Regal R.R. Capture-recapture methods in epidemiology: methods and limitation. Epidemiologic Review 1995; 17: 243-263.

[18] Jones H.E., Hickman M., Welton J.N., De Angelis. D., Harris Ross., Ades A.E. Recapture or Pre-capture? Fallibility of Standard Capture-recapture Methods in the Presence of Referral between Sources; American Journal of Epidemiology 2014; Vol. 179 No 11 pp 1383-1393.

[19] McKegancy NM., Barnard A, Leyland.. Female street-working prostitution and HIV infection I Glasgow. British Medical Journal1992; 305, 801-4.

[20] NAFDAC. Factors encouraging counterfeiting of drugs. nafdacnigeria.org/about.html. 2005. NAFDAC.2007 http://www.informationng.com/2013/06/nafdac-parades-4-over-importation-of-fake-drugs-worth-n59bn.html.

[21] Nanan DJ and White F. Capture-recapture: Reconnaissance of a Demographic Technique in Epidemiology. Chronic Diseases in Canada , 1997; 18, No. 4: 144-148.

[22] Osisiogu U.A and Chinwuba E.E (2019):Estimating Population size fake drug syndicates in Nigeria: A Capture-Recapture Approach by Sampling Coverage Methods. Zambrut International Journal of Biology, Physics & Mathematics

[23] Post L.A., Zhang H., Barboza G.E., Conner T. Simulation Demonstrate Feasibility of Capture-recapture. Michigan State University; 2013

[24] Rob van Hest. Capture-recapture methods in surveillance of tuberculosis and other infectious diseases. 2007

[25] Squires NF., Beeching NJ., Schlecht BJM and Ruben SM. An estimate of prevalence of drug misuse in Liverpool and a spatial- analysis of known addiction. Journal of Health Medicine 1995; 17, 103-109.

[26] Turing and Good. The population frequencies of species and the estimation of population parameters. Biometrika 1953.

[27] Van der Heijden PG., Cruyff MJ., Van Houwelingen, H. Estimating the size of a criminal population from police registrations using truncated Poisson regression model. Stat Neerl ., 2003; 57: 289-304.

[28] Van Hest NAH., Grant AD., Smit F., Story A., Richardus JH. Estimating infectious diseases incidence: Validity of capture-recapture analysis and truncated models for incomplete datacount. EpidemiolInfect;Publishedonline; 11March, 2007; doi: 10, 1017/S0950268807008254.

[29] WHO. Good governance for medicine. Curbing corruption in medicines regulation and supply. 2007 http://www.who.int/medicines/policy/goodgoverance/home/en/index.hmt.

[30] Zelterman D. Robust estimation in truncated discrete distributions with applications to capture-recapture experiment. J. Statistic Plan Inf 1988; 18; 225-37.