

The Zero Inflated Negative Binomial - Shanker Distribution and Its Application to HIV Exposed Infant Data

Stella Kibika, Collins Odhiambo*, Elphas Okango

Strathmore University, Nairobi, Kenya

Abstract Motivated by HIV exposed infants (HEI) sero-conversion data, we provide an extension of Zero-inflated Negative Binomial (ZINB) distribution to Zero Inflated Negative Binomial - Shanker (ZINB-SH) distribution. In this setting the ZINB-SH, distribution provides an alternative to the Poisson-Shanker distribution in particular, when data exhibits over dispersion brought by excess zeros. The HIV Exposed infant data is characterized by both structured and non-structured zeroes which makes the feature ideal in this context. We describe the properties of ZINB-SH distribution and estimate its parameters. Extensive simulations were conducted and the results in terms of goodness-of-fit, compared to the standard Negative Binomial, Shanker, Zero-Inflated Negative Binomial and Negative Binomial –Shanker distributions. The ZINB-SH distribution is competitive under different settings of simulation and does well as sample size increases. To validate the distribution, we apply real typical HIV-Infant exposed data.

Keywords Shanker Distribution, Zero Inflated Negative Binomial, Generalized Linear Model, HIV Exposed Infants, Mother-to-Child-Transmission

1. Introduction

Zero inflated models have been developed to analyze count data that exhibits many zeros. The skewed nature of the resulting distribution makes it difficult to transform the data to a normal distribution. Count models such as Poisson and negative binomial are preferred. In cases where these models are not able to handle the number of zeros, they are extended to zero-inflated models. The zero inflated Poisson (ZIP) and zero inflated negative binomial (ZINB) are the most popular models and have been used many literature [1,2 and 3]. Hurdle models are an alternative way of modelling zero inflated data and has also been utilized in several literature [4]. The Poisson hurdle model was introduced by John Mullahy [3]. The hurdle models also model the zeros and the non-zero values (zero-truncated) separately. The difference between the zero inflated models and the hurdle models is that the hurdle models do not distinguish between the structured and random zeros. All the zeros are assumed to be structured [3]. Failure to take into account both structure and unstructured zeros may introduce bias into the final

results if there exists random zeros in the data [1,2 and 3].

In cases where the data exhibits over dispersion, the ZINB is preferred to the Poisson. The zero inflated models introduce a second link function, the logit link, which allows the zeros to be distinguished as either structured or unstructured. There also exists hurdle models that assume all the zeros are structured. It truncates the non-zero part of the data. The hurdle models do not distinguish the zeros and may be applicable when the researcher is sure that all zeros arising from the data are structured for example when all the observations exist in a controlled space. An example would be looking at the occurrences of a disease when the entire population under investigation has been vaccinated for the particular disease. The zero inflated models will outperform hurdle models whenever the zeros in the data arise both structurally and randomly [5].

In order to create more flexible models, there are some mixed models such as zero inflated negative binomial-Crack (ZINB-CR) [6], zero inflated negative binomial-Sushila (ZINB-S) [7] and zero inflated negative binomial-Generalized exponential (ZINB-GE) [8] that have been developed. These models have been shown to perform better than the ZINB with fewer parameters. They are all parametric models that involve estimation of parameters.

The disadvantage of the extended models is the complexity that is introduced with mixing two distributions. Estimation of parameters becomes more complex and requires algorithms such as the Newton Raphson and

* Corresponding author:

codhiambo@strathmore.edu (Collins Odhiambo)

Published online at <http://journal.sapub.org/ijps>

Copyright © 2020 The Author(s). Published by Scientific & Academic Publishing

This work is licensed under the Creative Commons Attribution International

License (CC BY). <http://creativecommons.org/licenses/by/4.0/>

Expectation-Maximization (EM) algorithms. To achieve convergence, statistical software is used to make the process easier. There is also no standard procedure to tell when a dataset is exhibiting zero inflation [9].

The ZINB model is a model that is used to address the issue of over dispersion in count data with excess zeros. We seek to extend the ZINB probability distribution function (PDF) to allow for greater flexibility by introducing more randomness. Structured zeros occur due to the presence of a group that is not at risk of exhibiting the phenomena under study. Structured zeros are inevitable while unstructured zeros occur by chance. Ridout [4] gives an example of the distinction between the zeros using horticulture data where, in counting disease lesions on plants, a plant may have no lesions either because it is resistant to the disease (structured), or simply because no disease spores have landed on it (unstructured). In this study, the focus will be on mothers who undergo prenatal and postnatal care in facilities that are equipped to prevent transmission (structured) and mothers who do not receive prenatal and postnatal care or visit facilities that are not equipped to prevent transmission (unstructured). Distinguishing structured and unstructured zeros is important here because the two groups have different probabilities of mother to child transmission. The zero inflated models are applied to such kind of data, with more zeros than the probability distribution expects, to account for the excess zeros. Creating more flexible models is important for the bias-variance trade off and creating models that have better predictive ability. The mixed models mentioned in the introduction have been proven to perform better than the original form of ZINB.

Majority of sero-conversion among HIV Exposed Infants (HEI) occurs during the pregnancy, delivery and breastfeeding process [16]. The government and World Health Organization have put measures in place for Prevention of Mother to Child Transmission (PMTCT). These interventions have been introduced as a result of the high infant mortality rate recorded due to HIV in the period 1970-1990 where it also emerged that Mother to Child Transmission was a key factor [10]. The interventions put in place have resulted in a decrease in the number of MTCT from 29.7% in 2015 to 11.5% in 2017 with PMTCT coverage of 77% [11]. The difference in the quality of health services offered across the country leads to suboptimal procedures for PMTCT in certain facilities leading to random zeros in data collected for HIV sero-conversion [12].

In many studies relying on count data where the counts may exhibit more zeros than the common count models can handle, zero inflated models can be used to model the outcomes. This is useful especially when some of the zeros recorded are not occurring randomly. For example, if an intervention is put in place to prevent a phenomenon from occurring, the data collected will contain zeros occurring randomly and those that are a result of the intervention.

The negative binomial distribution is useful since it

accounts for over dispersion that may be present in the data. However, it is not able to cater for excess zeros; therefore the ZINB distribution is applied. The ZINB applies weights to the structured and random zeros. It gives a weight π to the structured zeros and $(1-\pi)$ to the random zeros and other count values greater than zero.

Given a random variable y ,

$$f(y) = \begin{cases} \pi + (1-\pi) \left(1 + \frac{\theta}{r}\right)^{-r}, & y=0 \\ (1-\pi) \frac{\Gamma(y+r)}{y! \Gamma(r)} \left(1 + \frac{\theta}{r}\right)^{-r} \left(1 + \frac{r}{\theta}\right)^{-y}, & y>0 \end{cases} \quad (1)$$

where, π =proportion of structured zeros

θ =probability of success

r =dispersion parameter.

This paper seeks to extend the ZINB to allow more flexibility to the model. The ZINB-SH will allow a parameter of the NB to be random and follow its own distribution. Mixture models have been used to increase the flexibility and robustness of probability distributions. This work aims to determine whether a mixture of the NB and Sushila distributions will provide greater flexibility and robustness when fitting zero inflated models. Fitting more than one model to a given dataset is common to establish the best model for a given situation.

The rest of the sections are organized in the following manner: in section two we look at the methodological approach when modelling ZINB-SH distribution. We also theoretically, derive parameter estimates. Simulation criteria and data description is also provided in section 2. In section three we provide results for both simulated data and real data HEI application. In section four we elaborate on interpretation of the results under discussions.

2. Methodology

2.1. Research Design

This study will create a mixture distribution for the ZINB and Shanker distributions. The model will then be applied on a real dataset of HIV exposed infants. The new distribution is made of the negative binomial and Shanker distributions and then adjusted for zero inflation.

The negative binomial Distribution

The probability distribution function is given by:

$$f(y) = \binom{m+y-1}{y} \theta^m (1-\theta)^y, \quad y=0,1,2,\dots, m > 0, 0 < p < 1 \quad (2)$$

where m = total number of trials

θ =probability of success.

The first two moments about the origin are:

$$E(y) = \frac{m(1-\theta)}{\theta} \quad (3)$$

$$E(y^2) = \frac{m(1-\theta)[1+m(1-\theta)]}{\theta^2} \quad (4)$$

Shanker Distribution

The Shanker distribution is proposed by Shanker [13] for modelling life data. It is a one parameter distribution which is a mixture of exponential (θ) and gamma (2, θ) distributions. This mixture gives the final probability of the Shanker [13] as:

$$f(y; a) = \frac{a^2}{a^2+1} (a + y)e^{-ya}, y > 0, a > 0 \tag{5}$$

where a =exponential rate.

Shanker also provides the moments for the distribution which are given as:

$$E(y) = \frac{a^2+2}{a(a^2+1)} \tag{6}$$

$$E(y^2) = \frac{2(a^2+3)}{a^2(a^2+1)} \tag{7}$$

$$\text{Var}(y) = \frac{a^4+4a^2+2}{a^2(a^2+1)^2} \tag{8}$$

The Negative Binomial-Shanker distribution

This is a compound distribution developed by Thhaloganyang [14]. It is a mixture of Negative Binomial and Shanker. Thhaloganyang [14] ends up with the probability distribution function of the NB-SH as:

$$f(y; r, \theta) = \binom{m+y-1}{y} \sum_{r=0}^y \binom{y}{r} (-1)^r \left(1 + \frac{\theta(m+r)}{\theta^2+1}\right) \left(1 + \frac{m+r}{\theta}\right)^{-2} \tag{9}$$

where, θ =probability of success
 r =dispersion parameter.

This distribution is a special case of the generalized negative binomial-Shanker (GNB-SH) with the parameter $\beta=1$. The distribution assumes $p=\exp(\lambda)$, λ follows a Shanker (θ) distribution.

The following are the properties of the NB-SH as shown by Thhaloganyang [14]:

$$\text{var}(y) = (1-\pi) \frac{m(m+1)\theta^2}{\theta^2+1} (\rho-2\omega+\delta) - \left[\frac{m\theta^2}{\theta^2+1} (\omega-\delta) \right]^2 + \pi(1-\pi) \frac{m(m+1)\theta^2}{\theta^2+1} (\rho-2\omega+\delta) \tag{17}$$

The parameters of the distribution were estimated using the maximum likelihood estimation method. This method differentiates the product of the probability distribution function with respect to each of the parameters. Numerical methods are used to solve the final equations.

Estimation of Parameters for ZINB-SH

The likelihood function:

$$L(m, \pi, \theta, r; y) = \prod_{i=1}^n \left[\left[I_{y_i=0} \left(\pi + (1-\pi) \left(1 + \frac{\theta m}{\theta^2+1} \right) \left(1 + \frac{m}{\theta} \right)^{-2} \right) \right] + \left[I_{y_i>0} (1-\pi) \binom{m+y_i-1}{y_i} \sum_{r=0}^{y_i} \binom{y_i}{r} (-1)^r \left(1 + \frac{\theta(m+r)}{\theta^2+1} \right) \left(1 + \frac{m+r}{\theta} \right)^{-2} \right] \right] \tag{18}$$

The log-likelihood function:

$$\binom{m+y_i-1}{y_i} = \frac{\Gamma(m+y_i)}{\Gamma(m+1)\Gamma(y_i)} \tag{19}$$

$$\log L = \sum_{i=1}^n \left[\log \left[I_{y_i=0} \left(\pi + (1-\pi) \left(1 + \frac{\theta m}{\theta^2+1} \right) \left(1 + \frac{m}{\theta} \right)^{-2} \right) \right] + \left[I_{y_i>0} \log(1-\pi) + \log \left(\Gamma(m+y_i) \right) - \log \Gamma(m+1) - \log \Gamma y_i + \log \sum_{r=0}^{y_i} \binom{y_i}{r} (-1)^r + \log \left(1 + \frac{\theta(m+r)}{\theta^2+1} \right) + \log \left(1 + \frac{m+r}{\theta} \right)^{-2} \right] \right] \tag{20}$$

$$E(y) = \frac{m\theta^2}{\theta^2+1} (\omega-\delta) \tag{10}$$

$$E(y^2) = \frac{m(m+1)\theta^2}{\theta^2+1} (\rho-2\omega+\delta) \tag{11}$$

where,

$$\delta = \frac{\theta^2+1}{\theta^2}, \omega = \frac{\theta(\theta-1)+1}{(\theta-1)^2}, \rho = \left(\frac{\theta-1}{\theta-2} \right)^2 \tag{12}$$

We create a ZINB-SH distribution; this will be achieved by using the method used by Lambert [15] which distinguishes the structured and random zeros. The model is a mixture of Bernoulli and Negative binomial-Shanker. The random zeros will follow the NB-SH distribution in equation 9. The model for zero inflation is as below:

$$f(y) = \begin{cases} \pi + (1-\pi) \left(1 + \frac{\theta m}{\theta^2+1} \right) \left(1 + \frac{m}{\theta} \right)^{-2}, & y=0 \\ (1-\pi) \binom{m+y-1}{y} \sum_{r=0}^y \binom{y}{r} (-1)^r \left(1 + \frac{\theta(m+r)}{\theta^2+1} \right) \left(1 + \frac{m+r}{\theta} \right)^{-2}, & y>0 \end{cases} \tag{13}$$

where, π =proportion of structured zeros.

θ =probability of success.

r =dispersion parameter.

m =number of observations.

To get the properties of the ZINB-SH, some general rules on finding mean and variance of zero inflated models are used.

For a zero inflated model with random variable(y),

$$E(y) = (1-\pi)E(y|z=0) \tag{14}$$

$$\text{var}(y) = (1-\pi)\text{var}(y|z=0) + \pi(1-\pi)E(y|z=0)^2 \tag{15}$$

The mean of ZINB-SH is therefore,

$$E(y) = (1-\pi) \frac{m\theta^2}{\theta^2+1} (\omega-\delta) \tag{16}$$

Partial derivatives:

$$\frac{\partial \log L}{\partial \pi} = \sum_{i=1}^n \left[\left[I_{y_i=0} \left(\frac{1}{\pi + (1-\pi) \left(1 + \frac{\theta m}{\theta^2 + 1} \right) \left(1 + \frac{m}{\theta} \right)^{-2}} \right) \right] + \left[I_{y_i>0} \frac{1}{1-\pi} \right] \right] = 0 \tag{21}$$

$$\frac{\partial \log L}{\partial \theta} = \sum_{i=1}^n \left[\left[I_{y_i=0} \left(1 + \frac{1+\theta m}{\theta^2+1} \right)^{-1} \frac{\theta m(2\theta) - (\theta^2+1)m}{\theta m(\theta^2+1)} - 2 \left(\left(1 + \frac{m}{\theta} \right)^2 \theta \right) \right] + \left[I_{y_i>0} \left(\left(1 + \frac{\theta(m+r)}{\theta^2+1} \right)^{-1} \right) \left(\frac{(m+r)}{\theta(m+r)(\theta^2+1)} \right) - 2 \left(1 + \frac{m+r}{\theta} \right)^2 \left(-\frac{1}{\theta} \right)^{-1} \right] \right] = 0 \tag{22}$$

$$\frac{\partial \log L}{\partial r} = \sum_{i=1}^n \left[I_{y_i>0} \left(1 + \frac{\theta(m+r)}{\theta^2+1} \right)^{-1} \theta + \left(1 + \frac{m+r}{\theta} \right)^2 \right] \tag{23}$$

$$\frac{\partial \log L}{\partial m} = \sum_{i=1}^n \left[\left[I_{y_i=0} \left(1 + \frac{\theta m}{\theta^2+1} \right)^{-1} (\theta) - 2 \left(1 + \frac{m}{\theta} \right)^2 \right] + \left[I_{y_i>0} \frac{1}{\Gamma(m+y_i)} \frac{\partial}{\partial m} \Gamma(m+y_i) - \frac{1}{\Gamma(m+1)} \frac{\partial}{\partial m} \Gamma(m+1) + \left(1 + \frac{\theta(m+r)}{\theta^2+1} \right)^{-1} (\theta) - 2 \left(1 + \frac{m+r}{\theta} \right)^2 \right] \right] \tag{24}$$

2.2. Data

The data used will be secondary data from publicly available information. It will be from three high burden areas, Kisumu, Nairobi and Mombasa. This data exhibits zero inflation because of the measures that have been put in place to reduce the rate of Mother to Child Transmission (MTCT). This increases the chance of children being born HIV-free and the chance of recording a zero. Due to the intervention, the results will contain structured zeros. All HEI sero-conversion who were registered in the EID programme in Kisumu, Nairobi and Mombasa between January 2014 and December 2018 were included in the study. From study sampling frame, a total of 494 samples were collected from HEI visiting 60 health facilities across the three cities in Kenya and obtained PCR testing together with the results. HEI with missing age or greater than 2 years old were excluded from analysis.

Statistical analysis

Data was transferred from the Microsoft excel windows 12 to R Studio [16] is for analysis. The analysis involves generating new random variables and getting density curves. The density curves will show the manner in which a given distribution fits the data, especially the zeros. We obtain a chi-square goodness of fit test to compare the distributions.

Chi-square test for goodness of fit test:

$$\chi^2_{\alpha,d} = \frac{(O-E)^2}{E} \tag{25}$$

where, O=observed Value
 E=expected Value
 α = significance level
 d = degrees of freedom

Ethics

Data is secondary and readily available from National AIDS and STI Control Programme (NAS COP) website. No patient identification information is included in NAS COP database.

2.3. Simulations

Simulation will be achieved by use of common simulation methods such as the acceptance rejection region method to fit a new probability distribution. The method is an iterative method and statistical software which will be used to run the algorithm. The goal will be to generate random numbers following the ZINB-SH (m, θ, π, and r). The results will be used to generate plots to visualize the shape of the distribution. The steps used will be:

- i. Generate U from the Uniform (0, 1) distribution.
- ii. Let λ come from the Shanker Distribution (θ).
- iii. Generate Y from the NB (m, p) distribution. p=exp(-λ)
- iv. Generate U* from the Uniform (0, 1) distribution.
- v. if U* > χ² set X=Y, otherwise X=0

3. Results

A simulation was carried out based on four distributions. Negative Binomial (NB), Negative Binomial – Shanker (NB-SH), Zero Inflated Negative Binomial (ZINB) and Zero Inflated Negative Binomial Shanker (ZINB-SH). The simulation is based on the following parameters estimated using maximum likelihood method in R statistical software:

Table 1. Parameter Estimation

	NB	SH	NB-SH	ZINB	ZINB-SH
Θ	0.2432		3.5572	0.8396	0.7647
Exponential rate		2.3429			
M				479	479
Π				0.4141	0.4142
R (dispersion)				0.5574	0.5574

The simulations produce the density plots below:

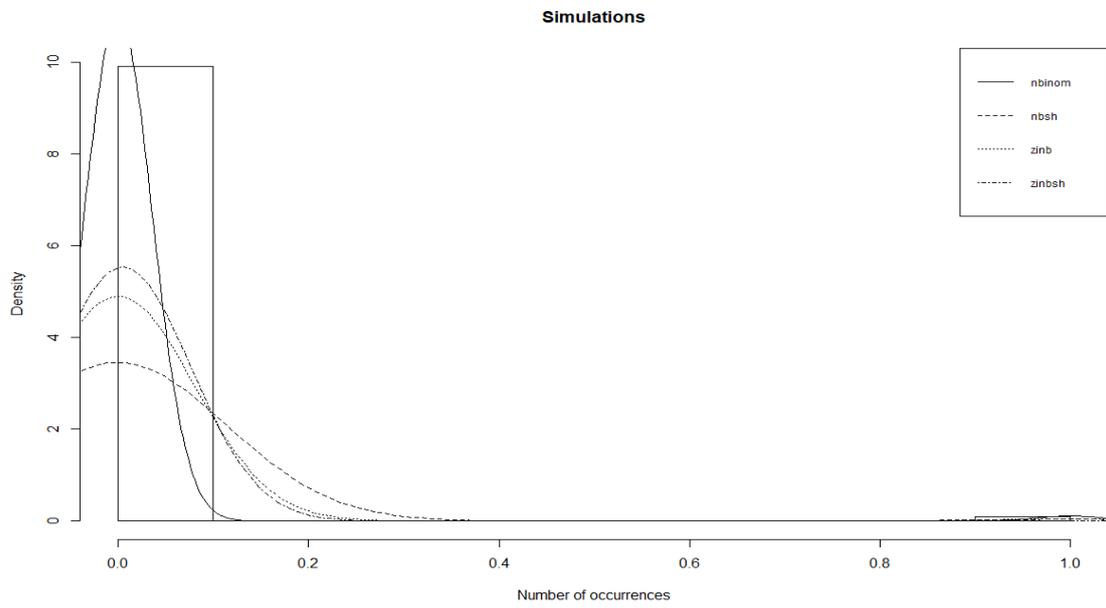


Figure 1. Density plots from simulations

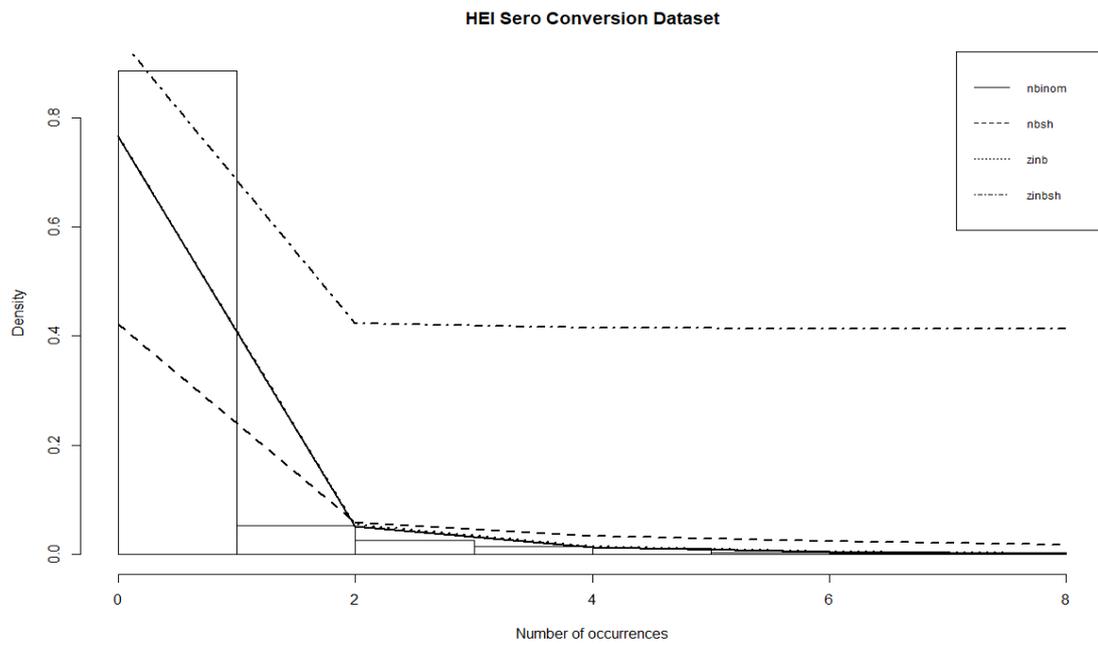


Figure 2. Density plots from a HEI-Sero conversion data-set

Table 2. Observed and expected count values for ZINB and ZINB-SH

Number of counts	Observed	ZINB	ZINB-SH	NB	NB-SH
		Expected	Expected	Expected	Expected
0	378	380	381	381	454
1	59	61	47	47	29
2	26	27	23	23	8
3	13	10	20	20	3
4	7	7	10	10	0
>4	11	9	13	13	0
χ^2 goodness of fit test (p-value)		0.09	0.78	0.78	0.05

The density plots show how the different models fit the data with zero inflation, the ZINB, ZINB-SH and NB-SH distributions partition the zeros. A portion of the zero values that may be considered random appear on the density plots. The NB distribution attempts to fit all the zero under the density curve since all the zeros are considered random.

The goodness of fit statistics show that there is no significant difference between the observed and expected values of the distributions. This implies that the models can be implemented in different scenarios and the best model chosen based on criteria such as Akaike Information Criterion (AIC).

4. Discussion

Interventions that targets PMTCT remain important HIV management considerations and are intended to mitigate the risk of HIV transmission from HIV infected mothers to their children. Kenya, in particular, has to a large extent, made significant progress in reducing MTCT rate. Literature has shown, for effective reduction in transmission, interventions for HEI begin well before delivery [17]. The HEI care and treatment, focuses on reducing the risk of infection with PreP, monitoring for signs and symptoms of HIV sero-conversion, and adhering to PCR testing schedule and starting treatment immediately. The UNAIDS report describe a 26% decline in incident HIV infections between 2009 and 2015 in the Global Plan priority countries in Sub-Saharan Africa [18,21]. This study is motivated by typical HEI sero-conversion data. In this setting the ZINB-SH, distribution provides an alternative to the Poisson-Shanker distribution in particular, when data exhibits over dispersion brought by excess zeros. The HIV Exposed infant data is characterized by both structured and non-structured zeroes which makes the feature ideal in this context. HEI sero-conversion data is collected routinely by ministry of health (MoH) in Kenya. Naturally, implementation of PMTCT is heterogeneous and results to structure-zero among positive HEI (situation where PMTCT is implemented optimally) and random zero among positive HEI (situation where PMTCT is implemented sub-optimally).

Failure to accommodate structured and random zero-inflation may result in false inference [14]. With this backdrop, the conventional zero-inflated distributions that do not consider structured and random zeros in data may give misleading results. Several rigorous and non-rigorous count data analysis approaches with zero inflation have been proposed by different researchers. Nekesa [12] did a comparison of four zero inflated models including the Zero Inflated Poisson (ZIP), Zero Altered Poisson (ZAP), Zero Inflated Negative Binomial (ZINB) and Zero Altered Negative Binomial (ZANB). Nekesa [12] concluded that the ZAP was the best model based on AIC values for the different models. Covariates are used to run a regression based on the four models and HEI is shown to be twice likely to detect HIV compared to initial Polymerase chain reaction

(PCR).

Here we provide an extension of Zero-inflated Negative Binomial (ZINB) distribution to Zero Inflated Negative Binomial - Shanker (ZINB-SH) distribution. We have described the properties of ZINB-SH distribution and estimate its parameters. Extensive simulations were conducted and the results in terms of goodness-of-fit, compared to the standard Negative Binomial, Zero-Inflated Negative Binomial and Negative Binomial – Shanker distributions. The ZINB-SH distribution is competitive under different settings of simulation and does well as sample size increases. To validate the distribution we apply real typical HIV-Infant exposed data.

4.1. Conclusions

In this work, the aim was to create a new distribution for zero inflated data and determine whether the new distribution performs better than the standard ZINB. The major difference between the two distributions is that ZINB-SH allows the parameters of the NB to be random and follow a distribution of their own, the parameters for the new distribution are determined using the maximum likelihood method. The new model is used in generating new random variables through simulations. It is also applied to a HEI sero-conversion dataset. In this case the ZINB-SH distribution has been shown to be competitive in performing analysis on zero inflated data. In the context of mixture distributions, a distribution with more parameters being random, with their own distribution, provides greater flexibility. ZINB-SH can be considered as a distribution when fitting models that exhibit excess zeros.

Abbreviation

AIC:	Akaike Information Criteria
AIDS:	Acquired Immune Deficiency Syndrome
EID:	Early Infant Diagnosis
HEI:	HIV Exposed Infant
HIV:	Human Immune Deficiency Virus
MLE:	Maximum Likelihood Estimation
MTCT:	Mother to Child Transmission
MoH:	Ministry of Health
NASCOP:	National AIDS and STI Control Programme
NB:	Negative Binomial
PCR:	Polymerase Chain Reaction
PDF:	Probability Distribution Function
PMTCT:	Prevention of Mother to Child Transmission
PreP:	Pre-Exposure Prophylaxis
SH:	Shanker
STI:	Sexually Transmitted Infections
ZANB:	Zero Altered Negative Binomial
ZAP:	Zero Altered Poisson
ZINB:	Zero Inflated Negative Binomial
ZINB-CR:	Zero-Inflated Negative Binomial- Crack
ZINB-GE:	Zero-Inflated Negative Binomial-

Generalized Exponential
ZINB-S: Zero-Inflated Negative Binomial-Sushila
ZINB-SH: Zero-Inflated Negative Binomial Shanker

ACKNOWLEDGEMENTS

The authors would like to thank the Strathmore Institute of Mathematical Sciences and the faculty who have supported the research by devoting their time and intellectual resources.

REFERENCES

- [1] Williamson, J. M., Lin, H., Lyles, R. H., & Hightower, A. W. (2007). Power calculations for ZIP and ZINB models. *Journal of Data Science*, 5(4), 519-534.
- [2] Lewsey, J. D., & Thomson, W. M. (2004). The utility of the zero-inflated Poisson and zero-inflated negative binomial models: a case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. *Community dentistry and oral epidemiology*, 32(3), 183-189.
- [3] Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of econometrics*, 33(3), 341-365.
- [4] Hu, M. C., Pavlicova, M., & Nunes, E. V. (2011). Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. *The American journal of drug and alcohol abuse*, 37(5), 367-375.
- [5] Yang, S. (2014). A comparison of different methods of zero-inflated data analysis and its application in health.
- [6] Saengthong, P., Bodhisuwan, W., & Thongteeraparp, A. (2015). The zero inflated negative binomial–Crack distribution: some properties and parameter estimation. *Songklanakarin J. Sci. Technol*, 37(6), 701-711.
- [7] Yamrubboon, D., Thongteeraparp, A., Bodhisuwan, W., & Jampachaisri, K. (2017, November). Zero inflated negative binomial-Sushila distribution and its application. In *AIP Conference Proceedings* (Vol. 1905, No. 1, p. 050044). AIP Publishing LLC.
- [8] Aryuyuen, S., Bodhisuwan, W., & Supapakorn, T. (2014). Zero inflated negative binomial-generalized exponential distribution and its applications. *Songklanakarin Journal of Science and Technology*, 36(4), 483-491.
- [9] Warton, D. I. (2005). Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. *Environmetrics: The official journal of the International Environmetrics Society*, 16(3), 275-289.
- [10] Shapiro, R. L., & Lockman, S. (2010). Mortality among HIV-exposed infants: the first and final frontier.
- [11] Mahy, M., Marsh, K., Sabin, K., Wanyeki, I., Daher, J., & Ghys, P. D. (2019). HIV estimates through 2018: data for decision-making.
- [12] Nekesa, F., Odhiambo, C., & Chaba, L. (2019). Comparative Assessment of Zero-Inflated Models with Application to HIV Exposed Infants Data. *Open Journal of Statistics*, 9(6), 664-685. <https://doi.org/10.4236/ojs.2019.96043>.
- [13] Shanker, R. (2015). Shanker distribution and its applications. *International journal of statistics and Applications*, 5(6), 338-348.
- [14] Tlhaloganyang, B. P., Mooketsi, D. R., Leinanyane, L., & Sakia, R. (n.d.). A compound of generalized negative binomial and shanker distribution.
- [15] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1-14.
- [16] Team, R. C. (2013). R: A language and environment for statistical computing.
- [17] Nyamhanga, T., Frumence, G., & Simba, D. (2017). Prevention of mother to child transmission of HIV in Tanzania: assessing gender mainstreaming on paper and in practice. *Health policy and planning*, 32(suppl_5), v22-v30.
- [18] UN-AIDS 2015 Progress Report on the Global Plan, UNAIDS / JC 2774/1/E.
- [19] Blasco-Moreno, A., Pérez-Casany, M., Puig, P., Morante, M., & Castells, E. (2019). What does a zero mean? Understanding false, random and structural zeros in ecology. *Methods in Ecology and Evolution*, 10(7), 949-959.
- [20] Zou, G. (2004). A modified Poisson regression approach to prospective studies with binary data. *American journal of epidemiology*, 159(7), 702-706.