Time effects of bacterial vaginosis on infant morbidities in Kenya assessed using modified skewed generalized estimating equations

Ngugi Mwenda (samwenda87@gmail.com)
Moi University https://orcid.org/0000-0001-9610-3124
Ruth Nduati
University of Nairobi
Mathew Kosgey
Moi University
Gregory Kerich
Moi University

Research Article

Keywords: HIV, Bacterial Vaginosis, GEE, SGEE

Posted Date: June 15th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-35335/v1

Time effects of bacterial vaginosis on infant morbidities in Kenya assessed using modified skewed generalized estimating equations

Abstract

Infant morbidity and mortality are indicators used globally as measures of a country's health status. Among the 8 millennium development goals (MDGs), this study aimed to address goal four (***MDG 4***) on the reduction of child mortality and six (***MDG 6***) on combating HIV and other diseases. We assessed different health conditions caused by bacterial vaginosis (BV) that could have life-long effects among infants. We aimed to address the time effects of BV on the long-term cause of infants' morbidities when asymmetry is assumed. We analyzed infant data from HIV-positive mothers with known BV status from a randomized controlled trial study conducted in Nairobi, Kenya. We aimed to investigate the effect of BV on infant morbidity with time from birth up to the age of 6 months. We derived a score for morbidity incidences depending on illnesses reported in the register during scheduled visits only. By adjusting for the mother's BV status, child's HIV status, sex, feeding status, and weight for age, we used two approaches for analysis. We considered and fitted the traditional generalized estimating (GEE) equations and our proposed skewed generalized estimating equations (SGEE). Overall, we included information on 327 infants. One thousand nine hundred sixty-two repeated measurements were available for analysis. Among the 327 mothers, 148 (45%) tested positive for BV, while 179 (55%) tested negative. We found that BV, gender, and time were associated with multiple health conditions in infants. Infants of women who tested positive for BV, at month 1, had 4.46 higher odds of various health conditions compared to infants of mothers who tested negative. The effects of BV tended to decrease with time, and at 5 months of age, children in the BV group had 1.10 times the odds of experiencing morbidity incidence. In the SGEE model, BV was statistically significant at the 0.05 level with a positive coefficient, indicating that children in the BV group had a higher probability of experiencing multiple

morbidities. BV is a significant predictor of infant morbidity because its effects on exposed infants could persist over time. In contrast, the traditional GEE results showed an insignificant positive coefficient. The results indicate the need to factor in the skewness during analysis in case of data transformation, especially when converting from continuous to binary data for parsimony and straightforward interpretation of the effects of covariates. Maternal BV status was positively associated with morbidity incidences, which highlights the need for early intervention for infected women. Accelerated programs promoting access to BV treatment with proper infant handling practices that better deal with emerging multiple health conditions in infants may prove useful in reducing the incidence of infant morbidity in Kenya. Emphasis on care to promote better health for infants during growth is necessary to achieve the MDGs.

Introduction

Skewed and non-normal data are commonly observed in health research. Usually, these data are often distorted, censored, or truncated to impose normality rather than modeling the data in its natural state [1]. Many conventional approaches would lead to incorrect estimates of parameters and standard errors due to the assumptions imposed. A typical assumption for the distribution of the error term in logistic analysis is the logistic function that is commonly applied to data with standard binomial distributions. Although the assumption is viewed as a reasonable compromise between mathematical simplicity and parsimonious results, its suitability has been doubted of late. Several studies have investigated various ways of handling non-normal data; however, few have focused on the methodology. For example, a paper published by Bono *et al.* gives several non-normal distributions typical in health, education, and social science [2]; however, their substantiation in the literature remains scarce. Several

distributions do not feature in the work of Bono *et al.*, showing that they were not common in the said period of reference. However, they could be vital in answering some important scientific questions. A critical issue of interest could be how we can model a binomial outcome that longitudinally violates symmetry. In response to the question of interest, some analysts use conventional ways to model this outcome, ignoring the consequences of mis-specifying the distribution by ignoring the asymmetry in the response. This has led to the availability of numerous standard models and researchers routinely treat the dependent variable as regular or symmetrical without scientific backing. Some researchers use the logit or probit models, which assume symmetry and tend to ignore skewness; yet, the supporting literature has shown that this could be inefficient for parameter estimation in some settings. The importance of normality and symmetry in data analysis cannot be underrated and there is a need for compromise between statistical simplicity and plausible estimates of parameters. However, statisticians have become over-reliant on the assumptions and questions regarding the suitability of the said methods have been raised in the literature [3]. Put differently, although numerous probability distribution functions can fit data quite well, the data need to speak for themselves, rather than being forced into a model with assumptions [1].

There is mounting scientific evidence regarding the inconsistency of the logistic distribution for binary response data. Recent studies have proposed alternative methods for handling binomial responses: assuming a gamma generated logistic distribution [4], gamma and log-normal distributions [5], improved analysis for skewed continuous responses [6], skewed Weibull regression model [7], generalized logistic distribution [8], and skewed logit [3]. This application shows that modeling non-normality continues to be appreciated in recent research; however, few methods have been considered and applied in health research.

Notwithstanding, most of the literature and applications have primarily focused on cross-sectional data in social, political, and economics research [9, 10, 11, 12, 13].

The purpose of the present study was to propose a new method of handling asymmetry in response to applications to infant morbidity using longitudinal datasets. Morbidity is defined as a state of body weakness due to illnesses and diseases that can be experienced at any stage in life during growth or maturity. We were motivated to analyze different morbidities among children born to women whose vaginal flora was tested for bacterial vaginosis (BV) during pregnancy and followed up to two years in Nairobi. Morbidity from childhood illnesses due to BV remains a major point of concern globally and particularly in Africa, where most of the cases occur. The scientific literature has established a link between BV and adverse outcomes in mothers and their children. Most studies have established the occurrence of health disparities [14, 15], pregnancy loss, labor complications and preterm delivery [16, 17, 18], and spontaneous and recurrent abortions [19] among mothers, while others have reported adverse outcomes such as neonatal malformations [20] and low birth weight [21] among infants. While some studies have tried to investigate the effects of BV in the context of human immunodeficiency virus (HIV) infection [15, 22, 23], there exists a lacuna in the literature regarding the time effects under such a scenario. Therefore, the present study aimed to assess the effects of BV in the context of HIV over time.

Materials and methods

Ethical approval

The study protocol was approved by the institutional review boards of the University of Washington and University of Nairobi. The original trial was supported through Fogarty grant

registration number D43-TW00007 and T22-TW00001. However, there was no RCT number provided at the time of conducting the study. Full description of the study is available at https://jamanetwork.com/journals/jama/fullarticle/192449 . Verbal consent was obtained from all mothers prior to inclusion in the present study.

Methodology

Monthly morbidity rates from birth up to six months of age were analyzed under the generalized estimating equations (GEE) framework. With wide applications in politics [3], transport [11], accidents [9], and sociology [13], we proposed to model our outcome assuming asymmetry in the response. We proposed a model that assumed a skewed logit distribution and an identity link function. The advantage of the assumed distribution is its ability to relax the strong assumption of symmetry in the logistic model. Second, the proposed model will still give plausible parameter estimates when the data are symmetrical because the logistic distribution is nested within the skewed logit distribution. The GEE framework is based on quasi-likelihood, and therefore, it is not mandatory to specify the full likelihood. However, a correlation in the dependent variable needs to be specified. In our application, we will consider several correlation structures. These include the unstructured, where every measure between two points is assigned its association parameter; autoregressive (AR-1) with $lag = 1$, in which correlation decreases exponentially with the differences in measurements; independence in which we use the identity matrix as the correlation structure; and exchangeable in which correlation is assumed to be equal across different measurements. The reasons for considering these is that using a robust sandwich estimator implies that even if the correlation structure is mis-specified, the parameter estimates remain valid.

While many authors have assumed a restrictive symmetric relationship when modeling the association between morbidities and the presence of BV using the logit or probit as the link function, we took a different approach of assuming asymmetry, which relaxes these restrictions. A standard approach to dealing with such a response, as has been the norm in previous studies, would be to use the generalized linear mixed model (GLMM). However, the weakness of this approach is that it assumes that the response is independent and homogeneous. A different approach would be to consider the standard GEE using the logit link, which relaxes the assumption of independence; however, if there is an asymmetry in the response, we end up with incorrect parameter estimates. Therefore, we proposed a skewed GEE (SGEE) to address this type of response.

First, we used the skewed logit transformation model under the generalized linear models (GLMs), on the dependent variable, assuming a variance cluster estimate (VCE) with the covariates of interest. The VCE was used to relax the strong assumption of independence in the response to allow dependency within the subject during the calculation of the skewness parameter. Independence is unrealistic in longitudinal data when dealing with the same subject. Second, from the skewness parameter estimate obtained, we modified the binomial response in the traditional GEE to allow for skewness and relax the assumptions of symmetry.

This was then implemented in R version 3.6.3 (The R Development Core Team, Vienna, Austria) using the *GEEM* package [24]. Thereafter, motivated by the fact that this methodology has not been implemented in analyzing longitudinal morbidity data, we analyzed a real dataset to provide a straightforward interpretation of the effect of the covariates on the response. Application to the Nairobi study infant morbidity dataset demonstrated that the SGEE outperforms the standard GEE in detecting a significant interaction between time and BV.

Notations and statistical methods

Consider n independent subjects observed at a specified time t. Given the number of subjects in our data, let, $i = i, .., n$ where $n = 327$ represents the total number of subjects and we let the i^{th} infant be observed n_j times, where $j = 1, .., 6$. A subject could be observed at a common set of time $t = 1, \dots, m$ up to a maximum of 6 $times$. The methods can also be applied to unequally spaced time $t_1 < \cdots < t_m$ points (see Hardie and Hilbe [25]. pg 75). Let $x_i = \left(x_{i1}, \dots, x_{in_j}\right)^T$ represent an $n_i \times p$ matrix-vector of covariates and let y_{it} be an $n_i \times 1$ vector of responses. This study assumes that our response variable, Y_{ij} has a Bernoulli distribution, *i.e.,* $y_{ij}|\rho_{ij} \sim Bern(\rho_{ij})$ with unknown $E(y_{ij}) = \rho_{ij}$. The outcome at time t which is morbidity can be represented as $Y_i = \begin{cases} 1, & if\ Yes\ for\ morbidity \\ 0, & otherwise \end{cases}$. This dependent variable is related to the covariates through a link function given by $g(\rho_{ij}) = x'_{ij}\beta$, where $g(.)$ is a logit link function, β is a p dimensional vector of regression coefficients, and x_{ij} is an n dimensional vector of covariates.

To model the marginal and subject specific probability of this type of response, authors have suggested we parametrize using a probit link $\Phi^{-1}(\mu)$ or a logit link $ln\ (\frac{\mu}{1-\mu})$. Subject specific analyses are important since their interpretation is at a lower level, but are complex to handle when there is a large number of respondents. A probit link function given by $\Phi\left(\frac{x_{it}\beta^{ss}}{\sqrt{1+\sigma_v^2}}\right)$ is a good candidate for this type, but has computation complexity for modeling higher order associations, while the logit link function given by $\Phi\left(\frac{x_{it}\beta^{ss}}{\sqrt{1+c\sigma_v^2}}\right)$ is easy to implement but has no closed form. The constant is expressed by $c = 16\sqrt{3/15\pi}$ as suggested

by Hilbe and Hadie [25]. This poses a challenge in the interpretability and practical applicability of the model.

Furthermore, the link functions are widely applicable in GLMs, which require that the full likelihood be specified. A major drawback to this approach for repeated measures is that an increase in the number of the measures results in an exponential increase in the number of parameters to be specified in the model and estimated. Both the logit and probit have conditional probability distributions, which are maximum at 0 such that P_i for $i \in (0,1)$ is 0.5 and thus has a fixed symmetry at 0.5.

However, symmetry may not be realistic to all Bernoulli or continuous responses as demonstrated by the works of different researchers in different fields such as political science by John Nagler [3], social and behavioral sciences by Golet *et al.* [26], and fisheries by Coelho *et al.* [10]. In these entire analyses, the researchers obtained better results by ignoring normality in the response. The methods used by Nagler followed an asymmetric logistic distribution and his results hold for models without repeated measures. Therefore, there was an assumption of independence among the responses. The restriction to models with independence is unappealing to models in a longitudinal set up where there is correlation within the subject measurements with time and interaction of covariates with time is of essence. The research conducted by Coelho *et al.* [10] was based on the GEE framework but their response was continuous.

As far as we are concerned, we have not come across any work on asymmetric binary under the GEE framework. We constructed a flexible link function that can accommodate both symmetric and asymmetric binary responses. Consider the model in which the response Y_i relates to the latent Y_i^* as follows:

$$Y_i = \begin{cases} 1, & if\ Y_i^* > 0 \\ 0, & otherwise \end{cases} \quad (1)$$

The probability density function (PDF) of subject i falling in the category of morbidity incidence as given by Nagler is $P_i = pr[X_i\beta + \mu_i > 0]$ and the cumulative distribution function (CDF) is given by $P_i = 1 - F(-X_i\beta)$ such that Y_i can be expressed as $P_i(Y_0 = 1) = F(-X_i\beta)$.

The marginal effect on P_i for a change in X_m is expressed as

$$\frac{\partial P_i}{\partial (X_m)} = \frac{\partial [1 - F(-X_i\beta)]}{\partial (X_m)} = f(-X_i\beta)\beta_m \quad (2)$$

We estimate the probability P_i for which the sensitivity $\partial P_i/\partial (X_m)$ is the maximum.

To relax the strong conditional probability on a binary response, accommodate the heterogeneity of repeated measures on the subjects, and put up for interaction time effects in the selected covariates, we employ the Burr type 10 distribution in the logit link under the GEE. This caters for the disturbance introduced in the logit during this process. Therefore, as proposed by Nagler, we have a more flexible predictor that can handle both symmetric and asymmetric responses in the binary variable.

However, to modify the link function, the logit link is usually preferred as it is easier to modify and can easily be generalized to imitate or mimic the Burr type 10 distribution proposed by Irving Burr in 1942 [27] which is a desired characteristic in this work. We introduce another parameter φ that will be referred to as the skewness parameter, and will be used to modify our response curve. This variation implies that the maximum is no longer restricted to $P = 0.5$.

The disturbance term estimated as $\hat{\boldsymbol{\alpha}}$ is independent of time and therefore its GLM estimate is assumed to be unbiased for a true α. This is achieved through an iterative weighted least square method put forth by McCullagh and Nelder [28]. An advantage of this model extension is that the disturbance term is assumed to be a constant. Therefore, our model can

still be easily generalized to conform to the exponential dispersion model (EDM) which can then be easily adopted in the GEE framework.

Irving Burr [27] developed the Burr type 10 distribution given a random variable *X* with a cumulative distribution $F(x)$ with its CDF distribution defined as $F(x) = \begin{cases} 1 - \frac{1}{(1+x^c)^k}, & x \geq 0 \\ 0 & otherwise \end{cases}$ with PDF given by $F'(x) = f(x) = 1 - \frac{kcx^{c-1}}{(1+x^c)^{k+1}}$. John Nagler recently proposed a new family of distributions called the skewed logit in which he modified the Burr type 10 CDF to mimic the logit such that the characteristic *"S"*-shaped curve of the sigmoid function $S(x) = \frac{1}{1+e^{-x}}$ is retained to accommodate a binary response. This was achieved by adding a constant parameter $F(k;\varphi) = \frac{1}{(1+e^{-k})^{\varphi}}$, for $\varphi > 0$, which is non-zero, non-negative and continuous, to the logit distribution function. This makes the estimator more flexible in modeling real binary data, since the logit is now nested in the scobit such that when the parameter is 1, then the proposed estimator conforms to the logistic distribution (see Fig 1 for various values of the parameter). As can be seen, the sigmoid slope takes on its maximum values at different probability levels depending on the parameter of choice.

There is a myriad of approaches to estimating skewness particularly under the GLM framework in which independence is assumed and the conventional way is to model binary "conditional independence models". We propose a new way of modeling binary data referred to as the "unconditional dependence model" a few methods of which exist.

This paper proposes a modification of the link function in the GEE proposed by Liang and Zeger to handle asymmetry in data with dependence. The skewed logit proposed under the GEE framework has a multiplier with a constant, meaning it can still be expressed as an EDM. This property makes it very easy to integrate in the GEE as the mean-variance relationship can easily be estimated. This property also implies that we do not have a difficult task in specifying

a full likelihood (in which we arrive at wrong conclusions when we specify a wrong one) but we could flexibly select any correlation structure and still obtain plausible results and reduction in the margin of error and bias. Since we are dealing with a binary response calculated as a score from a continuous response, the common approach would be to assume the logit model given by

$$log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta \mathbf{X_i^T} \in \mathfrak{R} \quad (3)$$

where $X_i's$ are the model covariates that include the weight, mother's BV status, HIV status of the infants, and feeding status in our data and the β values are the coefficients to be estimated. The logit assumes that the probability of success or failure is the same, maintaining symmetry assumptions and the maximum of the logit is achieved at $p = 0.5$. In this work, we aim to consider a response that violates the symmetry assumption, but still within the same framework.

Let $k(.)$ be the link function and $E(Y) = \mu$ such that $k(\mu) = X\beta$. The k^{-1} is the logit link for a binary response.

$$k^{-1}(\mu_{ij}) = \eta_{ij} = X_{ij}^T \quad (4)$$

Fitting a GLM to the data to obtain the disturbance parameter

After specifying the parameters, the initial estimate of β and the disturbance term were obtained using the GLM approach using scobit as the link function. However, the β values are just a proxy of association or what researchers refer to as "starting values" and not correct since they assume the presence of independence. The assumption of independence implies that the standard errors could be underestimated or overestimated, because we ignore within-subject dependency. Assuming that the parameter is a constant and that the scobit and logit are related,

then by a direct relationship, a scobit belongs to the EDM defined by the marginal density of the response belonging to the family of exponential distributions.

For the repeated Bernoulli response where measurements for subject i are taken repeatedly at time t, the PDF is given by:

$$\exp\left\{\frac{y_{it}\theta_{it} - b(\theta_{it})}{a(\phi)} + c(y_{it}, \phi)\right\} \tag{5}$$

where θ is the natural or canonical parameter, $a_i(\phi) > 0$ is the scale parameter, and $c(y, \phi)$ is the normalizing constant to ensure the PDF integrates to one. From basic principles, it can easily be shown that the variance is a function of the mean using $V(y_i) = v(\mu_i) = \mu_i(1 - \mu_i)$ where $\mu_i \in (0,1)$ depends on the expected value of the response, and to ensure that the CDF integrates to 1, the normalizing constant is independent of the natural parameter. Being an EDM means it is easy to form estimating equations to estimate the β values.

Estimating the $\boldsymbol{\beta}$ values

From a series of several equations and the chain rule, we differentiate the log-likelihood $\mathcal{L}(\theta, \phi \mid y_{1,\dots,}y_n) = \sum_{i=1}^{n}\left\{\frac{y_i - b(\theta)}{a(\phi)} + c(y_i, \phi)\right\}$ for $1, \dots, n$ with respect to β values through a chain of equations $\frac{\partial \mathcal{L}}{\partial \beta} = \left(\frac{\partial \theta}{\partial \mu}\right)\left(\frac{\partial \mathcal{L}}{\partial \theta}\right)\left(\frac{\partial \eta}{\partial \beta}\right)\left(\frac{\partial \mu}{\partial \eta}\right)$ for the EDM. For the estimating equations as defined by Liang and Zeger [cite] for a population average for GLM, the quasi-likelihood is given by

$$\Psi(\beta) = \sum_{i=1}^{n}\left\{\frac{1}{v(\mu_i)}\frac{y_i - \mu_i}{a(\phi)}\mathbf{x}'_{\mathbf{ij}}\boldsymbol{D}\right\} = 0 \tag{6}$$

whereby the first part of the equation is a generalization of the estimating equations of a GLM. The variance $V(\boldsymbol{\mu_i})$ can be decomposed into

$$V(\mu_i) = \boldsymbol{D}\big(V(\mu_{it})\big)^{1/2}\boldsymbol{I}_{(\boldsymbol{n_i}*\boldsymbol{n_i})}\mathbf{D}\big(V(\mu_{it})\big)^{1/2}. \quad (7)$$

We replace the identity matrix with a more general correlation matrix, say $R_i(\alpha)$, since the variance matrix for correlated data does not have a closed form.

Wedderbun showed that for any choice of V_i, the estimate of β (say $\widehat{\beta}$) is asymptomatically normal and consistent such that mis-specification of the variance is not an issue in parameter estimation [29].

By modifying the Newton–Raphson algorithm using the Fisher scoring criteria, we can estimate the $\beta's$. This procedure replaces the observed Hessian matrix with the expected Hessian matrix. This is achieved by setting $R_i(\alpha)$ as an identity matrix and the scale parameter ϕ as estimated from the GLM.

To solve the estimating equations, we employ the iterative reweighted least squares algorithm, which is a modification of the Newton–Raphson algorithm such that the observed Hessian matrix replaces the expected Hessian matrix. The following approach is used to estimate β.

$$\hat{\boldsymbol{\beta}}^{(r)} = \hat{\boldsymbol{\beta}}^{(r-1)} - \{\sum \mathbf{D}_i^T v(\mu)_i^{-1}\mathbf{D}_i\}\{\sum \mathbf{D}_i^T v(\mu)_i^{-1}\mathbf{S}_i\} \quad (8)$$

$$D_i = D\big(V(\mu_{it})\big)D\left(\frac{\partial \mu_i}{\partial \eta}\right)X_i \quad (9)$$

$$S_i = y_i - g^{-1}\big(\hat{\eta}_i\big) \quad (10)$$

The iteration continues until the convergence set by the researcher is achieved. Since this paper compares the two models with a modified link function under the GEE, our choice of model is the one that is better at predicting the association of BV with what is supported by the literature.

To make the GEE more efficient, we include several hypothesized structures in the subject correlations. Several correlation structures are considered for $B_i(\alpha)$, such as the independent, AR-1, m-dependent, exchangeable, and unstructured forms.

The algorithm for our proposed GEE is summarized as follows.

Step 1

Rather than assuming an extremely restrictive distribution for the binary data (which restricts the maximum change to probability 0.5), we propose that a distribution such as the Burr type 10 distribution be chosen, as it allows for the inclusion of a disturbance term without strong assumptions of symmetry. This is advantageous in that, as shown in the equation, when $\varphi = 1$, the distribution transforms itself to the logit. This means that our model is flexible in modeling both symmetrical and asymmetrical binomial data.

Step 2

Choose an initial $\varphi^{(0)}$ for the skewness parameter to estimate the true φ. We use the estimate from the skewed logit regression from the binomial GLM, which includes the same fixed effect (model covariate) and different time intercepts. We also use the VCE to relax the assumption of independence inherent in the GLMs. The VCE is given by $(X'\,X)^{-1}\sum\mu_i'\mu_j(X'\,X)^{-1}$. Our data were collected over time and are therefore, not independent.

Step 3

Confirm that $kJ^{-1}k' < \epsilon$ where ϵ is a small constant, such as 0.001 . Repeat step 2 using these educated guesses. Once the conditions are satisfied, it means that convergence has been achieved and the estimated φ is the most robust estimate to be used in the GEE.

Step 4

From the Burr type 10 distribution, $F(k;\varphi) = \frac{1}{(1+\exp -k)^{\varphi}}$, $k^{-1}(\mu_{ij}) = \eta_{ij} = X_{ij}^{T}$ will be the link function that relates the first moments to the covariates of interest.

Step 5

After obtaining the estimated value of φ, we use it to modify the skewness parameter for the binomial response, (usually assumed to be 1 in the *GEEM* package in the R software). Use the estimated φ to update the estimated values of β.

Step 6

Run the GEEM model using morbidity as the response and BV, feeding group, HIV status, weight, and gender as the covariates. To obtain an adequate model, we systematically added the predictors in order of importance while updating the β values as shown by the equation. Once convergence was attained, the estimated β values and their standard errors were obtained considering significance at $p = 0.05$.

Results

Application to real dataset and interpretation

The model was applied to the Nairobi infant morbidity dataset, comprising 327 infants, with a total of 1,962 measures. These were recorded from birth up to six months of age. These data were obtained from a longitudinal study designed to monitor mortality and morbidity among children born to women infected with HIV-1. The key outcomes were whether a child experienced any disease during the course of growth. These were captured during a scheduled or non-scheduled clinic visit and were previously analyzed by Mbori-Ngacha *et al.* [30] and Nduati *et al.* [31] who reported that morbidities varied at different times during the maturity of

331 the children. These diverse diseases, including early HIV infection, were more pronounced,
332 persistent, and fatal during early ages, particularly before the age of six months.

333 The preliminary analysis showed that 148 (45%) infants were born to women who tested
334 positive for BV while the remaining 179 (55%) were born to women who tested negative for
335 BV, 185 (57%) were in the breastfeeding arm while 142 (43%) were in the formula feeding
336 arm, 168 (51%) were males while 159 (49%) were female, and 61 (19%) were HIV-positive
337 while 266 (81%) were HIV-negative.

338 It was of scientific interest to model the effects of BV on the marginal probability of an
339 infant suffering from different morbidities in the first six months of life. Assuming morbidity
340 incidence as the response, our data had between zero and seven morbidity incidences recorded
341 in a given month for each infant. Morbidity incidences were recorded as continuous data;
342 however, in our analysis, we converted them into binary data for ease of model formulation
343 and interpretation. Other details regarding randomization, follow-up, and ethical review have
344 been described elsewhere [31]. We sought to assess whether children born to women who
345 tested positive for BV were likelier to have a higher morbidity incidence compared with their
346 counterparts and if the said effects would change with time. The literature has shown that BV
347 has more effects during the first months after birth as the child continues to build immunity as
348 they grow up. Also, we expect children who gain weight standardized for age to have fewer
349 morbidity incidences than those children who take time to gain weight. Abnormal weight for
350 age gain could be a sign of morbidity.

351

352 **Table 1: Distribution of bacterial vaginosis and disease incidence assessed between birth**
353 **and age six months**

Time	Morbidity incidences?	BV-Present	BV-Absent
Month 1	yes	115(35%)	85(26%)
	no	33(10%)	94(29%)
Month 2	yes	97(30%)	84(26%)
	no	51(16%)	95(29%)
Month 3	yes	97(30%)	85(26%)
	no	51(16%)	94(29%)
Month 4	yes	92(28%)	101(31%)
	no	56(17%)	78(24%)
Month 5	yes	86(26%)	96(29%)
	no	62(19%)	83(25%)
Month 6	yes	79(24%)	103(31%)
	no	69(21%)	76(23%)

The frequency of morbidity incidence seemed to decrease in both groups, but rose in the BV-absent group in months 4 and 6 (**Table 1**). It was, therefore, important to examine the effect of BV on child morbidity; thus, we considered the following marginal model:

$$logit\ (p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 BV_i + \beta_2 HIV_i + \beta_3 feeding_i + \beta_4 weight_i + \beta_5 weight_i + \beta_6 male_i + \beta_7 time_i + \beta_{17} time_i \times BV_i + \epsilon_i \quad (11)$$

where $time_i = 1, \ldots, 6$, $gender_i = 1$ if the i^{th} child is male and 0 if female, $HIV = 1$ if the child tests positive and 0 otherwise, $feeding_i = 1$ if the child was randomized to the formula feeding group and 0 if randomized to the breastfeeding group, $BV_i = 1$ if the mother tested positive for BV and 0 if she tested negative, and $weight_i$ is recorded continuously for 6 months.

In this paper, we were interested in comparing our proposed model to the standard GEE when the response was likely to be asymmetric. We wanted to show that our model fits better for binary data that violate symmetry. We tested different correlation structures for comparison purposes; however, because our model was concerned with the time effect, we used the AR (1) output for interpretation.

Table 2: Inference from the GEE and the proposed SGEE model for the Nairobi infant morbidity data with several correlation structures. We wanted to show that our model performs better than the standard model even when different correlation structures are used.

		GEE			SGEE		
Effect	Corr	Estimates	SE	p-value	Estimates	SE	p-value
Intercept	IND	0.253	0.228	0.359	0.176	0.209	0.461
	EXCH	0.088	0.263	0.738	0.024	0.242	0.920
	AR1	0.119	0.267	0.667	0.043	0.245	0.858
	M-DEP	0.129	0.261	0.640	0.050	0.239	0.836
	UNSTR	0.029	0.272	0.914	-0.038	0.249	0.873
BRESTFED	IND	-0.057	0.108	0.718	-0.062	0.099	0.649
	EXCH	-0.022	0.146	0.883	-0.027	0.134	0.838
	AR1	-0.052	0.137	0.740	-0.058	0.126	0.671
	M-DEP	-0.051	0.131	0.747	-0.057	0.121	0.678
	UNSTR	-0.027	0.149	0.863	-0.031	0.137	0.823
WEIGHT	IND	-0.112	0.057	0.114	-0.125	0.052	0.041
	EXCH	-0.068	0.067	0.326	-0.087	0.062	0.148
	AR1	-0.062	0.067	0.384	-0.076	0.062	0.214
	M-DEP	-0.068	0.065	0.341	-0.080	0.060	0.190
	UNSTR	-0.034	0.068	0.631	-0.050	0.063	0.421
HIV	IND	0.189	0.179	0.541	0.222	0.159	0.363
	EXCH	0.248	0.250	0.406	0.273	0.225	0.273
	AR1	0.216	0.234	0.508	0.253	0.208	0.317
	M-DEP	0.212	0.224	0.518	0.250	0.198	0.323
	UNSTR	0.193	0.253	0.551	0.232	0.225	0.371
MALE	IND	-0.370	0.117	0.028	-0.382	0.107	0.008
	EXCH	-0.358	0.156	0.024	-0.358	0.144	0.013
	AR1	-0.359	0.148	0.031	-0.369	0.135	0.010
	M-DEP	-0.363	0.142	0.030	-0.373	0.130	0.010
	UNSTR	-0.319	0.159	0.053	-0.329	0.145	0.024
TIME	IND	0.178	0.062	0.019	0.192	0.057	0.004
	EXCH	0.146	0.067	0.052	0.165	0.061	0.011
	AR1	0.135	0.071	0.075	0.150	0.065	0.022
	M-DEP	0.143	0.070	0.061	0.156	0.064	0.018
	UNSTR	0.110	0.069	0.145	0.126	0.063	0.055
BV	IND	1.086	0.431	0.170	1.495	0.348	0.000
	EXCH	1.049	0.470	0.191	1.475	0.371	0.000

	AR1	1.000	0.533	0.272	1.494	0.421	0.001
	M-DEP	1.017	0.526	0.272	1.513	0.417	0.001
	UNSTR	0.901	0.530	0.347	1.287	0.440	0.020
BV:TIME	IND	-0.199	0.091	0.169	-0.275	0.077	0.001
	EXCH	-0.198	0.088	0.171	-0.277	0.072	0.001
	AR1	-0.191	0.107	0.240	-0.280	0.088	0.001
	M-DEP	-0.196	0.106	0.238	-0.285	0.088	0.001
	UNSTR	-0.176	0.099	0.300	-0.246	0.084	0.017

The results were generally similar and well within the sampling random error in terms of parameter estimates and standard errors. When we chose a level of significance of $\alpha = 0.05$, the effects of the standard GEE were not significant and only time and gender were significant. Using our proposed SGEE model, gender, time, BV, and the interaction between time and BV were significant.

In both the GEE and proposed SGEE, the signs of the effects were similar. Using our proposed model, we could say that gender is a good predictor of morbidity and the odds for males versus females with the SGEE model was $\exp(\beta_6) = \exp(-0.372) = 0.7$. Those with BV had an $\exp(\beta_1) = \exp(1.494) = 4.48$ odds of morbidity compared with their counterparts.

Effects of time on BV

We calculated the effects of BV with time among infants given by $\exp(\boldsymbol{\beta_1 + \beta_{17} \times time})$ (Table 3).

Table 3: Coefficients of bacterial vaginosis with time calculated from $\boldsymbol{exp^{(\beta_1+\beta_{17}time)}}$ by replacing the respective values from the SGEE model with the AR-1 correlation structure

Time	Coefficient of effects of Bacterial Vaginosis
Birth	4.48

Month 1	3.37
Month 2	2.54
Month 3	1.92
Month 4	1.45
Month 5	1.11
Month 6	0.83

The effects of BV on morbidity tended to decrease from birth with time (**Table 3**). We observed higher morbidity effects at birth, and the effects decreased with time. For example, comparing months 1 and 5, we can conclude that at month 1, the odds of morbidity incidence was 2.72 for children whose mothers had BV compared to those whose mothers did not have BV. At month 5, the odds decreased to 1.04 for mothers who tested positive for BV versus those who tested negative.

Discussion

The present study utilized the skewed logit technique under the GEE framework to analyze the risk factors associated with BV. We built on the existing contributions put forth by Nagler [3] and Liang and Zeger [32]. The model proposed in the present study is based on logistic regression, modified assuming a parameter for skewness, to allow it to accommodate both symmetric and asymmetric responses. There are several situations in which the relationship between the function of the response and covariates is not strictly symmetric. The asymmetric model is a class of models that borrows strength from both symmetric and asymmetric forms and can be applied in both scenarios while maintaining parsimony. Furthermore, the most encountered assumption of symmetry is very restrictive, unrealistic, and can lead to incorrect conclusions regarding the parameter estimates. The present study focused on investigating commonly neglected “minor diseases” and proved they should not be ignored

411 at the expense of the so-called “major causes” of infant morbidity and mortality such as mother
412 to child transmission of HIV [33, 34].

413 We found that gender is a good predictor of infant morbidity. Precisely, girls were more
414 likely to be healthy than boys. This finding is supported by previous studies and adds to the
415 large body of knowledge indicating that boys require more attention healthwise than girls, and
416 girls had a higher survival probability, consistent with the reports of Stevenson and colleagues
417 [35]. Regarding this finding, there will be hopes of a decline in mortality among boys as a result
418 of better interventions targeting their health.

419 BV was proven to have a significant relationship with infant morbidities controlled for
420 other covariates. Infants whose mothers tested positive for BV were found to have higher
421 morbidity incidences compared to those whose mothers tested negative. The effect of BV on
422 infant health has been reported in several studies, but with different conclusions [36, 37]. The
423 most important finding in this work was the degree of significance observed in the SGEE model
424 on the interaction between BV and time. This finding would be of interest to doctors, as it
425 indicates the need to plan for proper treatment and monitoring of the infant’s health after
426 confirming the maternal BV status, particularly during the first six months. This finding can
427 also inform targeted infant morbidity campaigns depending on the mother’s BV status and the
428 age of the infant. The effects of BV on time tended to decrease with infant age; however, at six
429 months, infants of mothers without BV tended to show a reversal in the odds of morbidity
430 incidence. This was an unexpected finding and may be attributed to reverse causality, in which
431 at first, infants who were exposed could have a higher number of morbidities than unexposed
432 infants. This would trigger more hospital visits by this group in search of treatment compared
433 with their counterparts. Therefore, with time, there would be a reduction in infant morbidities

in the exposed group compared with the unexposed group. This would likely tilt the odds of morbidity incidence in favor of the exposed group.

Some of the six covariates that were analyzed with controls were not statistically significant but could assist in detecting a trend of association with infant morbidities. Some of these covariates include the mode of breastfeeding, which had a negative relation with infant morbidities. This finding could reflect behaviors that have been reported in other studies whereby breastfed infants were more healthy than their counterparts who were formula-fed [31, 38].

Another covariate of interest was HIV status, which showed a positive coefficient with infant morbidity. Infants who tested positive for HIV showed signs of morbidity, consistent with the results obtained by Kartik *et al.* [38]. Morbidities associated with HIV were found to increase infant mortality risk according to studies conducted in Kenya [30], Botswana [39], Cameroon [36], and South Africa [38].

Finally, the negative coefficient of weight and infant morbidities could indicate that weight gain could reduce morbidity. Weight gain standardized for age is a good measure of infant health, and the results were consistent with those reported by Margarita *et al*. [40] and Sarah *et al*. [41].

In addition, our study found a decline in the number of morbidities from birth, although this was in sharp contrast to the report published by Verma *et al*. [42], who found that morbidity would increase during infant growth. Our results could be attributable to the fact that there were more hospitalizations in the first months, which decreased with time.

To our knowledge, this is the first health research that considers a skewed logit under the GEE framework. Our study is one of the few studies that specifically explores the effect of BV on infants with time. We considered and compared two models: a standard GEE and a standard

458 skewed GEE. Since the beginning of the twenty-first century, there has been a lot of
459 development in the field of statistics. The nature of statistical complexities, advances in
460 statistical software, and sophisticated methodologies indicate that previous unrealistic
461 assumptions about data can now be easily approached and solved. Asymmetry in the binary
462 outcome is a phenomenon that should be appropriately accounted for in analytical models to
463 avoid biases in final parameter estimates, as has been established in **Table 2**. It should be noted
464 that the proposed skewed logit density converges to a logit density when the skewness
465 parameter $\varphi = 1$, because the logit is nested within the skewed logit. Based on this, we
466 recommend and advise researchers to evaluate the data before analysis to ensure that they use
467 not only a correct model, but a parsimonious one. Although our findings cannot be extrapolated
468 to all morbidity studies, they have shed light on situations in which the skewed logit can be
469 applied. Moreover, we have introduced a new mathematical approach to appropriately
470 addressing asymmetric issues with binary outcomes.

471 We recommend policies targeting BV among HIV-infected pregnant women. Finally, the
472 choice of model selection has evolved over time and different researchers will have their own
473 preferences. For example, for full likelihood models, the Akaike information criterion,
474 corrected Akaike information criterion, Bayesian information criterion [43], and deviance
475 information criterion [44] are used; and for the quasi-likelihood model, the quasi-information
476 criterion was proposed by Pan [45]. However, other researchers have adopted models that yield
477 superior predictions of model effects by covariates, supported by the literature. For example,
478 Wu *et al*. proposed a beta-binomial model under the GEE and GLMM frameworks [46],
479 whereby the former was preferred due to its ability to predict the index of cumulative toxicity
480 of the mutant protein characteristic of Huntington's disease which was consistent with other
481 studies regarding the Functional Assessment Scale of the Unified Huntington's Disease Rating

Scale. Similarly, the literature supports an association between BV and morbidity among infants [47]; thus, we conclude that our model is better than the conventional model in predicting a BV-time interaction.

References

[1] B. Manandhar and B. Nandram, "Hierarchical bayesian models for continuous and positively skewed data from small areas," *Communications in Statistics - Theory and Methods*, pp. 1–19, Aug. 2019. [Online]. Available: https://doi.org/10.1080/-03610926.2019.1645853

[2] R. Bono, M. J. Blanca, J. Arnau, and J. Gomez-Benito, "engNon-normal distributions commonly used in health, education, and social sciences: A systematic review," *engFrontiers in psychology*, vol. 8, no. 28959227, pp. 1602–1602, Sep. 2017. [Online]. Available: https://-www.ncbi.nlm.nih.gov/pmc/articles/PMC5603665/

[3] J. Nagler, "Scobit: An alternative estimator to logit and probit," *American Journal of Political Science*, vol. 38, no. 1, pp. 230–255, 1994. [Online]. Available: http://-www.jstor.org/stable/2111343

[4] F. Castellares, M. A. C. Santos, L. C. Montenegro, and G. M. Cordeiro, "A gamma-generated logistic distribution: Properties and inference," *American Journal of Mathematical and Management Sciences*, vol. 34, no. 1, pp. 14–39, Jan. 2015. [Online]. Available: https://-doi.org/10.1080/01966324.2014.954296

[5] M. Faddy, N. Graves, and A. Pettitt, "Modeling length of stay in hospital and other right skewed data: Comparison of phase-type, gamma and log-normal distributions," *Value in Health*, vol. 12, no. 2, pp. 309 – 314, 2009. [Online]. Available: http://-www.sciencedirect.com/science/article/pii/S1098301510607097

[6] A. A. Afifi, J. B. Kotlerman, S. L. Ettner, and M. Cowan, "engMethods for improving regression analysis for skewed continuous or counted responses." *engAnnual review of public health*, vol. 28, pp. 95–111, 2007.

[7] R. Caron, D. Sinha, D. K. Dey, and A. Polpo, "Categorical data analysis using a skewed weibull regression model," *Entropy*, vol. 20, no. 3, 2018. [Online]. Available: https://www.mdpi.com/1099-4300/20/3/176

[8] P. N. Rathie, P. Silva, and G. Olinto, "Applications of skew models using generalized logistic distribution," *Axioms*, vol. 5, no. 2, 2016. [Online]. Available: https://-www.mdpi.com/2075-1680/5/2/10

[9] R. Tay, "Comparison of the binary logistic and skewed logistic (scobit) models of injury severity in motor vehicle collisions," *Accident Analysis & Prevention*, vol. 88, pp. 52–55, Mar. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/-S0001457515301615

[10] R. Coelho, P. Infante, and M. N. Santos, "Application of generalized linear models and generalized estimation equations to model at-haulback mortality of blue sharks captured in a pelagic longline fishery in the atlantic ocean," *Fisheries Research*, vol. 145, pp. 66 – 75, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/-S0165783613000532

[11] J. Zhang and H. Timmermans, “Scobit-based panel analysis of multitasking behavior of public transport users,” *Transportation Research Record*, vol. 2157, no. 1, pp. 46–53, Aug. 2019. [Online]. Available: https://doi.org/10.3141/2157-06
[12] N. D. Wright, M. Symmonds, L. S. Morris, and R. J. Dolan, “Dissociable influences of skewness and valence on economic choice and neural activity,” *PLOS ONE*, vol. 8, no. 12, p. e83454, Dec. 2013. [Online]. Available: https://doi.org/10.1371/journal.pone.0083454
[13] J. Hay, A. Walker, K. Sanchez, and K. Thompson, “Abstract social categories facilitate access to socially skewed words,” *PLOS ONE*, vol. 14, no. 2, p. e0210793, Feb. 2019. [Online]. Available: https://doi.org/10.1371/journal.pone.0210793
[14] M. L. Alcaide, M. Chisembele, E. Malupande, K. Arheart, M. Fischl, and D. L. Jones, “engA cross-sectional study of bacterial vaginosis, intravaginal practices and hiv genital shedding; implications for hiv transmission and women’s health.” *engBMJ open*, vol. 5, p. e009036, Nov 2015.
[15] D. J. Alcendor, “engEvaluation of health disparity in bacterial vaginosis and the implications for hiv-1 acquisition in african american women.” *engAmerican journal of reproductive immunology (New York, N.Y. : 1989)*, vol. 76, pp. 99–107, Aug 2016.
[16] P. Brocklehurst, A. Gordon, E. Heatley, and S. J. Milan, “engAntibiotics for treating bacterial vaginosis in pregnancy.” *engThe Cochrane database of systematic reviews*, p. CD000262, Jan 2013.
[17] J. C. Carey, M. A. Klebanoff, J. C. Hauth, S. L. Hillier, E. A. Thom, J. Ernest, R. P. Heine, R. P. Nugent, M. L. Fischer, K. J. Leveno, R. Wapner, M. Varner, W. Trout, A. Moawad, B. M. Sibai, M. Miodovnik, M. Dombrowski, M. J. O’Sullivan, J. P. VanDorsten, O. Langer, and J. Roberts, “Metronidazole to prevent preterm delivery in pregnant women with asymptomatic bacterial vaginosis,” *New England Journal of Medicine*, vol. 342, no. 8, pp. 534–540, 2000, pMID: 10684911. [Online]. Available: https://doi.org/-10.1056/NEJM200002243420802
[18] S. Guaschino, F. De Seta, M. Piccoli, G. Maso, and S. Alberico, “engAetiology of preterm labour: bacterial vaginosis.” *engBJOG : an international journal of obstetrics and gynaecology*, vol. 113 Suppl 3, pp. 46–51, Dec 2006.
[19] G. Isik, S. Demirezen, H. G. Donmez, and M. S. Beksac, “engBacterial vaginosis in association with spontaneous abortion and recurrent pregnancy losses,” *engJournal of cytology*, vol. 33, no. 27756985, pp. 135–140, 2016. [Online]. Available: https://-www.ncbi.nlm.nih.gov/pmc/articles/PMC4995870/
[20] A. S. Dingens, T. S. Fairfortune, S. Reed, and C. Mitchell, “engBacterial vaginosis and adverse outcomes among full-term infants: a cohort study,” *engBMC pregnancy and childbirth*, vol. 16, no. 27658456, pp. 278–278, Sep. 2016. [Online]. Available: https://-www.ncbi.nlm.nih.gov/pmc/articles/PMC5034665/
[21] S. L. Hillier, R. P. Nugent, D. A. Eschenbach, M. A. Krohn, R. S. Gibbs, D. H. Martin, M. F. Cotch, R. Edelman, J. G. Pastorek, A. V. Rao, D. McNellis, J. A. Regan, J. C. Carey, and M. A. Klebanoff, “Association between bacterial vaginosis and preterm delivery of a low-birth-weight infant,” *N Engl J Med*, vol. 333, no. 26, pp. 1737–1742, Nov. 2018. [Online]. Available: https://doi.org/10.1056/NEJM199512283332604
[22] D. N. Burns, R. Tuomala, B. H. Chang, R. Hershow, H. Minkoff, E. Rodriguez, C. Zorrilla, H. Hammill, and J. Regan, “engVaginal colonization or infection with candida albicans in human immunodeficiency virus-infected women during pregnancy and during the postpartum period. women and infants transmission study group.” *engClinical infectious*

diseases : an official publication of the Infectious Diseases Society of America, vol. 24, pp. 201–10, Feb 1997.

[23] D. J. Jamieson, A. Duerr, R. S. Klein, P. Paramsothy, W. Brown, S. Cu-Uvin, A. Rompalo, and J. Sobel, "engLongitudinal analysis of bacterial vaginosis: findings from the hiv epidemiology research study." *engObstetrics and gynecology*, vol. 98, pp. 656–63, Oct 2001.

[24] L. S. McDaniel, N. C. Henderson, and P. J. Rathouz, "Fast pure R implementation of GEE: application of the Matrix package," *The R Journal*, vol. 5, pp. 181–187, 2013. [Online]. Available: https://journal.r-project.org/archive/2013-1/mcdaniel-henderson-rathouz.pdf

[25] J. W. Hardin and J. M. Hilbe, *Generalized Estimating Equations*, 2, Ed. Chapman and Hall/CRC; 2nd edition (December 10, 2012), 2013.

[26] I. Golet, "Symmetric and asymmetric binary choice models for corporate bankruptcy," *Procedia - Social and Behavioral Sciences*, vol. 124, pp. 282 – 291, 2014, challenges and Innovations in Management and Leadership. [Online]. Available: http://-www.sciencedirect.com/science/article/pii/S1877042814020357

[27] I. W. Burr, "Cumulative frequency functions," *Ann. Math. Statist.*, vol. 13, no. 2, pp. 215–232, Jun. 1942. [Online]. Available: https://projecteuclid.org:443/euclid.aoms/-1177731607

[28] P. McCullagh and J. Nelder, *Generalized Linear Models*. Chapman & Hall, 1989.

[29] R. W. M. Wedderburn, "Quasi-likelihood functions, generalized linear models, and the gauss-newton method," *Biometrika*, vol. 61, no. 3, pp. 439–447, 1974. [Online]. Available: http://www.jstor.org/stable/2334725

[30] D. Mbori-Ngacha, R. Nduati, G. John, and et al, "Morbidity and mortality in breastfed and formula-fed infants of hiv-1-infected women: A randomized clinical trial," *JAMA*, vol. 286, no. 19, pp. 2413–2420, Nov. 2001. [Online]. Available: http://dx.doi.org/10.1001/-jama.286.19.2413

[31] R. Nduati, G. John, D. Mbori-Ngacha, B. Richardson, J. Overbaugh, A. Mwatha, J. Ndinya-Achola, J. Bwayo, F. E. Onyango, J. Hughes, and J. Kreiss, "engEffect of breastfeeding and formula feeding on transmission of hiv-1: a randomized clinical trial." *engJAMA*, vol. 283, pp. 1167–74, Mar 2000.

[32] K.-Y. LIANG and S. L. ZEGER, "Longitudinal data analysis using generalized linear models," *Biometrika*, vol. 73, no. 1, pp. 13–22, 04 1986. [Online]. Available: https://doi.org/-10.1093/biomet/73.1.13

[33] S. E. Kellerman, S. Ahmed, T. Feeley-Summerl, J. Jay, M. Kim, B. R. Phelps, N. Sugandhi, E. Schouten, M. Tolle, F. Tsiouris, C. S. W. G. of the Interagency Task Team on the Prevention, M. Treatment of HIV infection in Pregnant Women, and Children, "engBeyond prevention of mother-to-child transmission: keeping hiv-exposed and hiv-positive children healthy and alive," *engAIDS (London, England)*, vol. 27 Suppl 2, no. 24361632, pp. S225–S233, Nov. 2013. [Online]. Available: https://www.ncbi.nlm.nih.gov/-pmc/articles/PMC4087192/

[34] J. Ladner, M.-H. Besson, M. Rodrigues, K. Sams, E. Audureau, and J. Saba, "engPrevention of mother-to-child hiv transmission in resource-limited settings: assessment of 99 viramune donation programmes in 34 countries, 2000-2011," *engBMC public health*, vol. 13, no. 23672811, pp. 470–470, May 2013. [Online]. Available: https://-www.ncbi.nlm.nih.gov/pmc/articles/PMC3660172/

[35] D. K. Stevenson, J. Verter, A. A. Fanaroff, W. Oh, R. A. Ehrenkranz, S. Shankaran, E. F. Donovan, L. L. Wright, J. A. Lemons, J. E. Tyson, S. B. Korones, C. R. Bauer, B. J.

618 Stoll, and L.-A. Papile, “Sex differences in outcomes of very low birthweight infants: the
619 newborn male disadvantage,” *Archives of Disease in Childhood - Fetal and Neonatal Edition*,
620 vol. 83, no. 3, pp. F182–F185, 2000. [Online]. Available: https://fn.bmj.com/content/83/3/-
621 F182
622 [36] F. Monebenimp, D. E. Nga-Essono, A.-C. Zoung-Kany Bissek, D. Chelo, and
623 E. Tetanye, “Hiv exposure and related newborn morbidity and mortality in the university
624 teaching hospital of yaounde, cameroon,” *The Pan African medical journal*, vol. 8, no.
625 22121451, pp. 43–43, Apr. 2011. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/-
626 PMC3201607/
627 [37] Z. S. Lassi, A. Majeed, S. Rashid, M. Y. Yakoob, and Z. A. Bhutta, “The
628 interconnections between maternal and newborn health evidence and implications for
629 policy,” *The Journal of Maternal-Fetal & Neonatal Medicine*, vol. 26, no. sup1, pp. 3–53,
630 2013. [Online]. Available: https://doi.org/10.3109/14767058.2013.784737
631 [38] K. K. Venkatesh, G. de Bruyn, E. Marinda, K. Otwombe, R. van Niekerk, M. Urban,
632 E. W. Triche, S. T. McGarvey, M. N. Lurie, and G. E. Gray, “Morbidity and mortality among
633 infants born to hiv-infected women in south africa: implications for child health in resource-
634 limited settings,” *Journal of tropical pediatrics*, vol. 57, no. 20601692, pp. 109–119, Apr.
635 2011. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/PMC3107462/
636 [39] R. L. Shapiro, S. Lockman, S. Kim, L. Smeaton, J. T. Rahkola, I. Thior, C. Wester,
637 C. Moffat, P. Arimi, P. Ndase, A. Asmelash, L. Stevens, M. Montano, J. Makhema,
638 M. Essex, and E. N. Janoff, “engInfant morbidity, mortality, and breast milk immunologic
639 profiles among breast-feeding hiv-infected and hiv-uninfected women in botswana.” *engThe*
640 *Journal of infectious diseases*, vol. 196, pp. 562–9, Aug 2007.
641 [40] M. E. Ahumada-Barrios and G. F. Alvarado, “Risk factors for premature birth in a
642 hospital,” *Revista Latino-Americana de Enfermagem*, vol. 24, 00 2016.
643 [41] S. G. Berger, S. de Pee, M. W. Bloem, S. Halati, and R. D. Semba, “Malnutrition and
644 morbidity are higher in children who are missed by periodic vitamin a capsule distribution for
645 child survival in rural indonesia,” *The Journal of Nutrition*, vol. 137, no. 5, pp. 1328–1333,
646 2007. [Online]. Available: https://doi.org/10.1093/jn/137.5.1328
647 [42] I. C. Verma and S. Kumar, “Causes of morbidity in children attending a primary health
648 centre,” *The Indian Journal of Pediatrics*, vol. 35, no. 12, pp. 543–549, Dec. 1968. [Online].
649 Available: https://doi.org/10.1007/BF02759677
650 [43] M. J. Brewer, A. Butler, and S. L. Cooksley, “The relative performance of aic, aicc and
651 bic in the presence of unobserved heterogeneity,” *Methods in Ecology and Evolution*, vol. 7,
652 no. 6, pp. 679–692, 2016. [Online]. Available: https://besjournals.onlinelibrary.wiley.com/-
653 doi/abs/10.1111/2041-210X.12541
654 [44] N. J. Evans, “Assessing the practical differences between model selection methods in
655 inferences about choice response time tasks,” *Psychonomic Bulletin & Review*, vol. 26, no. 4,
656 pp. 1070–1098, 2019. [Online]. Available: https://doi.org/10.3758/s13423-018-01563-9
657 [45] W. Pan, “Akaike’s information criterion in generalized estimating equations,”
658 *Biometrics*, vol. 57, no. 1, pp. 120–125, 2001. [Online]. Available: http://www.jstor.org/-
659 stable/2676849
660 [46] H. Wu, Y. Zhang, and J. D. Long, “engLongitudinal beta-binomial modeling using gee
661 for overdispersed binomial data.” *engStatistics in medicine*, vol. 36, pp. 1029–1040, Mar
662 2017.
663 [47] N. Mwenda, R. Nduati, M. Kosgey, and G. Kerich, “Morbidities and mortality among
664 infants of HIV-1-infected mothers with bacterial vaginosis in kenya,” jun 2020.

Declarations:

We as authors declare no competing interests.

We also declare that this research article is part of the first authors PhD output and the work received no funding from any source.

Appendix 1

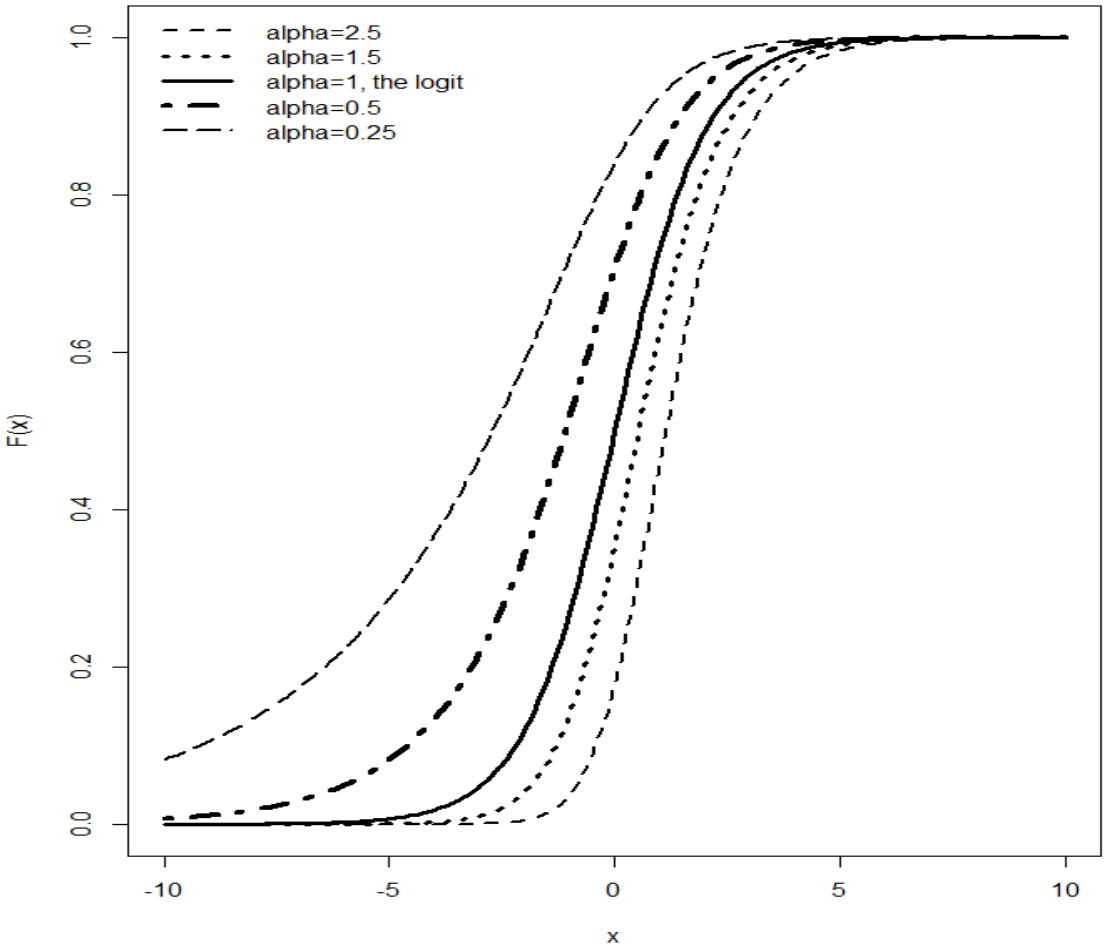

Fig 1: Cumulative density function of the skewed logit with different values of skewness. The bold continuous line represents the logit which assumes symmetry.

Figures

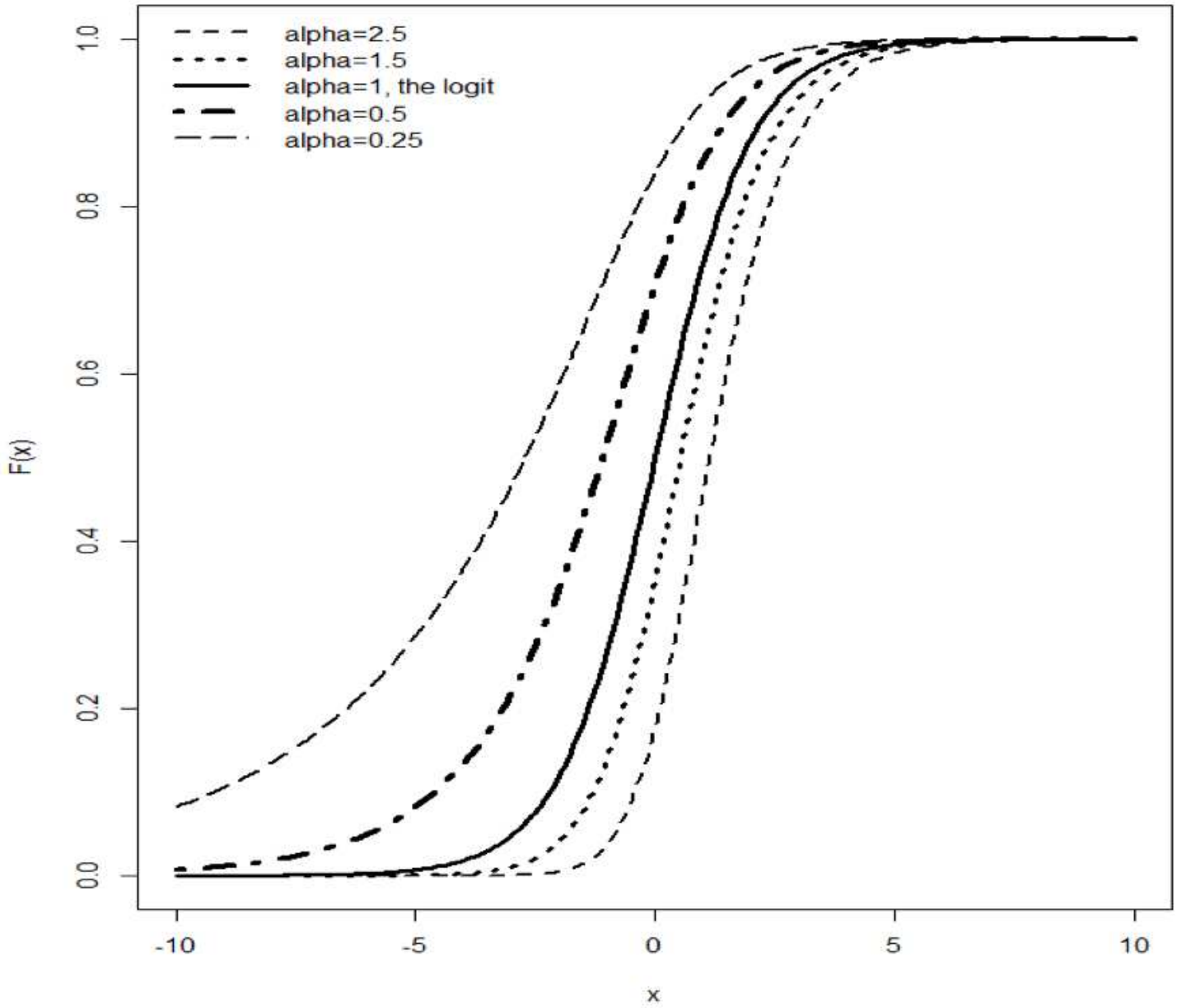

Figure 1

Cumulative density function of the skewed logit with different values of skewness. The bold continuous line represents the logit which assumes symmetry.