

RESEARCH

Open Access



The best statistical model to estimate predictors of under-five mortality in Ethiopia

Setegn Muche Fenta^{1*} , Haile Mekonnen Fenta² and Girum Meseret Ayenew³

*Correspondence:
setegn14@gmail.com

¹ Department of Statistics,
College of Science, Debre
Tabor University, Debre Tabor,
Ethiopia

Full list of author information
is available at the end of the
article

Abstract

The under-five mortality rate is one of the most important indicators of the socio-economic wellbeing and public health conditions of a country. Under-five death in Ethiopia has reduced, but the rate is still higher than the sustainable development goal target of 20 deaths per 1000 live births. This study aimed to identify the best statistical model to estimate predictors of under-five mortality in Ethiopia. Ethiopian demography and health survey of 2016 data were accessed and used for the analysis. A total of 14,370 women were included. Various count models (Poisson, Negative Binomial, Zero-Inflated Poisson, Zero-Inflated Negative Binomial, Hurdle Poisson, and Hurdle Negative Binomial) were considered to identify risk factors associated with the death of under-five in Ethiopia. The mean number of under-five death was 0.9 and its variance was 1.697. The hurdle negative binomial model had the smallest AIC, Deviance, and BIC, suggesting the best goodness of fit. Besides, the predictive value and probabilities for many counts in the hurdle negative binomial model fitted the observed counts best. The result of hurdle negative binomial model showed that region, mother's age, educational level of the father, education level of the mother, father's occupation, family size, age of mother at first birth, vaccination of child, contraceptive use, birth order, preceding birth interval, twin children, place of delivery, antenatal visit predict under-five death in Ethiopia. The rate of Under-five death remains high. Concerned governmental organizations should work properly to reduce under-five mortality through encouraging child vaccinations and antenatal care visits. Attention should also be provided to multiple births and the spacing among order of birth. The Hurdle negative binomial model provided a better fit for the data. It is argued the Hurdle negative binomial model for count data with excess zeros of unknown sources such as the number of under-five death should be fitted.

Keywords: Under five mortality, Ethiopia, Hurdle negative binomial, Count models

Introduction

Child mortality is defined as the probability of live births per household in a specific year who die before celebrating the age of five. Child mortality has received international attention through Sustainable

Development Goals (SDGs) [1]. Every year, under-five children die in millions [2]. In 2016, for instance, 5.6 million children died, and globally, still about 15,000 children die every single day. The burden of under-five deaths remains unevenly distributed

and in southern Asia and sub Saharan Africa in particular, remains high. For instance, about 80 percent of under-five deaths occur in two regions, sub-Saharan Africa and South Asia. Including Ethiopia, six countries account for half of the global under-five deaths [1, 3, 4].

In Ethiopia, the mortality rate of under-five is 67 per 1000 live births, with large disparities in across regions. Every 15 children die before reaching their fifth year. Every year, more than 257,000 children die under the age of five, of which 120,000 die in the neonatal period. If present trends continue, more than 3,084,000 children will die until the 2030 [5, 6].

To identify the risk factor of under-five mortality, different countries conducted various studies [7–12]. Many small-scale surveys were done on a specific set of variables. These studies investigated the risk factors of under-five mortality through binary logistic and survival analysis [13–15]. Though, binary logistic regression undercounts the total number of mortality since multiple mortalities are collapsed into a single unit to fulfill the requirements of binary logistic regression, provides sufficient information for studying the pattern of multiple child deaths. In this study, the count regression model is the preferred model of analysis. The response variable is the number of under-five death (count) and the main aim is to see how this count changes as the explanatory variable increase. Classical Poisson regression is the most well-known method for modeling count data. Yet its underlying assumption of equal-dispersion (i.e., an equal mean and variance) limits its use in many real-world applications with over or under dispersed data (i.e., the variance is larger than the mean or smaller than the mean). Excess variation may result in incorrect inference about parameter estimates, standard errors, tests, and confidence intervals. Overdispersion frequently arises for various reasons. One is; excessive zero counts or censoring. Over-dispersed count data are common in many areas which in turn, lead to the development of the statistical methodology for modeling over-dispersed data [16]. The negative binomial distribution looks like the Poisson distribution, except that it has a longer, fatter tail to the extent the variance exceeds the mean. Depending on the degree of overdispersion, the negative binomial model could capture more zeros than the Poisson model [17]. Nevertheless, the model may be insufficient in with respect to empirical applications bearing zero values in the data. Zero-inflated models provide a way of modeling the excessive proportion of zero values by allowing overdispersion. When number of zeros is large, provides a good fit than Poisson or negative binomial model [18]. However, these models are not suitable for under-dispersion. A flexible alternative that captures both over and under-dispersion is the hurdle model [19]. On the other hand, most of the scholars on child mortality considered either prevalence alone, Chi square association, or logistic and survival models to analyze such [11, 12, 20–22] outcomes. In logistic regression, the dependent variable is dichotomized to be either dead or alive, which undercounts the total number of under-five mortality. This implies that multiple child deaths are collapsed into a single unit for doing a logistic regression. Therefore, this study tried to identify the best statistical model to estimate predictors of under-five mortality in Ethiopia. The main strength of the model used in this case accounts for survey design features such as weighting, clustering, and stratification since a failure to account for design features leads to invalid statistical inference such as standard errors, and over or under estimation.

Contribution

To our knowledge, this is the first paper assessing the effects of different covariates on the number of under-five mortality rate using the count regression models such as hurdle negative binomial at a multi-regional level with a huge dataset. Besides, identifying the specific predictor associated with under-five is important in prioritizing interventions and revealing patterns for better results. Further, to offer flawless information to researcher how to use over dispersed, under dispersed and zero-inflated regression models.

Methods

Data source

The data used for this study was taken from the 2016 EDHS a nationally representative survey of women’s aged (15–49 years) groups drawn from the Central Statistics Agency (CSA), Ethiopia. The survey is the fourth comprehensive survey designed to provide estimates in all urban and rural areas of Ethiopia about the targeted health and demographic variables.

Dependent variable

The response variable of this study is number of under-five deaths per mother.

Independent variables

Possible predictors of under-five death use such as region, mother’s age, educational level of the father, education level of the mother, father’s occupation, mother’s occupation, family size, age of mother at first birth, religion, vaccination of child, contraceptive use, birth order, preceding birth interval, child twin, place of delivery, antenatal visit, breastfeeding, and residence were included in the analysis.

Statistical software

The secondary data were recorded in SPSS software version 21 and then exported to R software version 3.5.3 and analyzed by using R software.

Statistical models

In some epidemiological studies, the response of interest consists of a count; such as number of the under-five deaths. In this case, the count regression model better analyzes the response measurements. From count regression models, this study applied Poisson, Negative Binomial, Zero-Inflated Poisson, Zero-Inflated Negative Binomial, Hurdle Poisson, and Hurdle Negative Binomial models.

Poisson regression model

The most popular model for count data is the Poisson model, which assumes that the mean and variance of the dependent variables to be equal [23]. The probability function for Y is given by

$$p(Y = y_i/\mu_i) = \frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!} \quad \begin{matrix} y_i = 0, 1, 2, \dots, \\ \mu_i > 0, i = 0, 1, 2, \dots \end{matrix} \tag{1}$$

where, y_i is the number of under-five deaths the i th mother in a given time with a mean parameter μ_i . The mean and variance of the Poisson distribution is given as

$$E(y_i) = Var(y_i) = \mu_i.$$

Negative binomial regression model (NB)

Sometimes the variance exceeds the mean which is referred to as over dispersion [24]. Various models exist that account for overdispersion and for this study it can be modelled using a negative binomial (NB) regression model. The probability function is given by

$$P(y_i, \mu_i, \alpha) = \frac{\Gamma(y_i + 1/\alpha)}{y_i! \Gamma(1/\alpha)} (1 + \alpha \mu_i)^{-\frac{1}{\alpha}} \left(1 + \frac{1}{\alpha \mu_i}\right)^{-y_i} \quad y_i \geq 0 \text{ and } \alpha > 0 \quad (2)$$

where α is the over dispersion parameter, $\Gamma(\cdot)$ is the gamma function, $\alpha = 0$ and, the Negative Binomial distribution is the same as Poisson distribution. The mean and variance are expressed as:

$$E(y_i) = \mu_i = \exp(x_i^T \beta), \quad Var(y_i) = \mu_i(1 + \alpha \mu_i).$$

Zero-inflated poisson (ZIP) regression model

The zero-inflated Poisson (ZIP) regression model is a modification of this familiar Poisson regression model that allows for an over-abundance of zero counts in the data [18]. Specifically, if Y_i is the number of under-five mortality per mothers is independent random variables having a zero-inflated Poisson distribution, the zeros are assumed to arise in two ways corresponding to distinct underlying states. The first state occurs with probability π_i and produces only zeros (mothers who are never born), while the other state occurs with probability $1 - \pi_i$ and leads to a standard Poisson count with mean μ and hence a chance of further zeros (mothers whose child may not be dead). In general, the zeros from the first state are called structural zeros and those from the Poisson distribution are called sampling zeros [25]. This two-state process gives a simple two-component mixture distribution with probability mass function

$$p(Y_i = y_i) = \begin{cases} \pi_i + (1 - \pi_i) \exp(-\mu_i), & \text{if } y_i = 0 \\ (1 - \pi_i) \frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!}, & \text{if } y_i = 1, 2, \dots \end{cases} \quad 0 \leq \pi_i \leq 1 \quad (3)$$

The parameter μ_i and π_i depends on the covariates x_i and z_i , respectively. The mean and the variance of the ZIP regression model, respectively, are:

$$E(y_i) = (1 - \pi_i) \mu_i \text{ and } Var(y_i) = \mu_i(1 - \pi_i)(1 + \pi_i \mu_i).$$

To apply the zero-inflated Poisson model in practical modeling situations, [18, 26, 27] suggested the following joint models for μ and π

$$\ln(\mu) = X^T \beta \text{ and } \ln\left(\frac{\pi}{1 - \pi}\right) = Z^T \gamma. \quad (4)$$

where X and Z are covariate matrices and β, γ are $(p + 1) \times 1$ and $(q + 1) \times 1$ vectors of unknown parameters respectively. The two sets of covariates may or may not coincide. For a random sample of observations y_1, y_2, \dots, y_n the, log-likelihood function is given by

$$l(\mu; \pi; y) = \sum_{i=1}^n \{ I_{(y_i=0)} \log [\pi_i + (1 - \pi_i) \exp(-\mu_i)] + I_{y_i>0} [\log(1 - \pi_i) + y_i \log(\mu_i) - \mu_i - \log(y_i!)] \} \tag{5}$$

where $I(\cdot)$ is the indicator function for the specified event, i.e. equal to 1 if the event is true and 0 otherwise.

Zero-inflated negative binomial (ZINB) regression model

The ZIP model may sometimes, fail to fit such data either because of over-dispersion in relation to the Poisson distribution. We extend the ZIP mixed regression model to ZINB mixed regression model. The ZINB regression is used for count data that exhibit over-dispersion and excess zeros [28, 29].

Suppose Y_i is the number of under-five mortality per mother then, the probability mass function of ZINB is given by:

$$p(Y_i = y_i) = \begin{cases} \pi_i + \frac{(1 - \pi_i)}{(1 + \alpha\mu_i)^{-\frac{1}{\alpha}}}, & \text{if } y_i = 0 \\ 1 - \pi_i \frac{\Gamma(y_i + 1/\alpha)}{y_i! \Gamma(1/\alpha)} (1 + \alpha\mu_i)^{-\frac{1}{\alpha}} \left(1 + \frac{1}{\alpha\mu_i}\right)^{-y_i}, & \text{if } y_i > 0 \end{cases}, \quad 0 \leq \pi_i \leq 1 \tag{6}$$

where μ_i the mean of the underlying negative binomial distribution and α is the over-dispersion parameter. The ZINB distribution reduces to the ZIP distribution as $\alpha \rightarrow 0$. The mean and variance, $E(Y_i) = (1 - \pi_i)\mu_i$ and $Var(Y_i) = (1 - \pi_i)(1 + \pi_i\mu_i + \alpha\mu_i)$ respectively [29].

In the terminology of generalized linear models (GLMs) $\ln(\mu_i)$ and $\log it(\pi_i)$ are the natural links for the negative binomial mean and Bernoulli probability of success [18, 30]

$$\ln(\mu_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \text{ and } \log it(\pi_i) = \gamma_0 + \gamma_1 z_{i1} + \dots + \gamma_q z_{iq} \tag{7}$$

where x_i and z_i are respectively vectors of covariates for the negative binomial and the logistic components, β and γ are the corresponding vectors of regression coefficients.

The hurdle poisson model

The hurdle regression handles the excess zeros by relaxing the assumption that zeros and positives come from a single data generating process [18]. A hurdle model is introduced by [30] for the analysis of over-dispersed or under-dispersed count data. Hurdle Poisson (HP) model is a two-component model consisting of a hurdles component models zero versus non-zero counts. A truncated Poisson count component is employed for the non-zero counts [19, 31]. Its probability density function is given as:

$$p(Y_i = y_i) = \begin{cases} \pi_0 & \text{if } y_i = 0 \\ (1 - \pi_0) \frac{\exp(-\mu_i) \mu_i^{y_i}}{(1 - \exp(-\mu_i)) y_i!} & \text{if } y_i = 1, 2, \dots \end{cases} \quad 0 \leq \pi_0 \leq 1 \tag{8}$$

where $\pi_0 = p(y_i = 0)$ and $\mu_i = \text{Exp}(x_i \beta)$.

In the PLH model, the most natural choice to model the probability of zeros is a logistic regression model.

$$\log \text{it}(\pi_0) = \log \left(\frac{\pi_0}{1 - \pi_0} \right) = z_i^T \gamma \tag{9}$$

where $z_i = (1, z_{i1}, z_{i2}, \dots, z_{iq})$ is the i th row of covariate matrix Z and $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_q)$ are unknown q -dimensional column vector of parameters. While the effect of covariates z_i on strictly positive (that is censored). Count data are modeled through Poisson regression:

$$\log(\mu_i) = x_i^T \beta \tag{10}$$

where $x_i = (1, x_{i1}, x_{i2}, \dots, x_{ip})$ is the i th row of covariate matrix X and $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ are unknown p -dimensional column vector of parameters. This model was proposed originally by [31].

The log-likelihood function of a Logit-Poisson regression can, therefore, be expressed as the sum of log-likelihood functions of two components as below:

$$l(\mu; \pi; y) = \sum_{i=1}^n \{ I_{(y_i=0)} \log(\pi_0) + I_{y_i>0} [\log(1 - \pi_0) - \mu_i + y_i \log(\mu_i) - \log(1 - \text{Exp}(\mu_i)) - \log(y_i!)] \}. \tag{11}$$

The hurdle negative binomial model

Similarly, for the hurdle models, the Hurdle Negative Binomial can be used instead of Poisson distribution for over-dispersion [19]. We consider a hurdle negative binomial (HNB) regression model in which the response variable $Y_i = (i = 1, 2, 3, \dots, n)$ has the distribution

$$p(Y_i = y_i) = \begin{cases} \pi_0, & \text{if } y_i = 0 \\ 1 - \pi_0 \frac{\Gamma(y_i + 1/\alpha) (1 + \alpha \mu_i)^{-\frac{1}{\alpha}} \left(1 + \frac{1}{\alpha \mu_i}\right)^{-y_i}}{y_i! \Gamma(1/\alpha) \left(1 - (1 + \alpha \mu_i)^{-\frac{1}{\alpha}}\right)}, & \text{if } y_i > 0 \end{cases} \quad 0 \leq \pi_0 \leq 1 \tag{12}$$

where $\alpha \geq 0$ is a dispersion parameter that is assumed not to depend on covariates. In addition, we suppose $0 < \pi_0 < 1$ and $\pi_0 = \pi_0(z_i)$.

The most natural choice to model probability of excess zeros is to use a logistic regression model:

$$\log it(\pi_0) = \log \left(\frac{\pi_0}{1 - \pi_0} \right) = \sum_{j=1}^q z_{ij} \gamma_j \tag{13}$$

where $z_i = (1, z_{i1}, z_{i2}, \dots, z_{iq})$ is the i th row of covariate matrix Z and $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_q)$ are unknown q -dimensional column vector of parameters. Impact of covariates on count data is modeled through NB regression

$$\log(\mu_i) = \sum_{j=1}^p x_{ij} \beta_j. \tag{14}$$

x_{ij} is the covariates, β is the coefficient of the independent variables in the regression model and p is the number of these independent variables.

We can obtain the log-likelihood function for the hurdle negative binomial regression model as

$$LL = \sum_{i=1}^n \left\{ I_{y_i=0} \log(\pi_0) + I_{y_i>0} \left[\log(1 - \pi_0) - \frac{1}{\alpha} \log(1 - (1 + \alpha \mu_i)) - \log y_i! - y_i \log \left(1 + \frac{1}{\alpha \mu_i} \right) - \frac{1}{\alpha} \log(\alpha \mu_i + 1) + \log \left(\frac{\Gamma(y_i + 1/\alpha)}{\Gamma(1/\alpha)} \right) \right] \right\} \tag{15}$$

Assessing model adequacy

We use the likelihood ratio test (LRT) to assess the goodness-of-fit between two nested models. The LRT is only valid test to compare hierarchically nested models. None nested model was compared by using Akaike information [32].

Result

A total of 14,370 women were included of which 7720 (53.3%) of the mothers have not faced any under-five death, only 78 (0.5%) of them lost 7 children. Further screening of the number of death of under-five showed that the variance (1.697) is greater than the mean (0.9) indicating over-dispersion (Table 1).

Table 1 Number of child deaths per mother

Number of deaths	Frequency	Percent
0	7720	53.7
1	3314	23.1
2	1773	12.3
3	798	5.6
4	413	2.9
5	198	1.4
6	76	0.5
7	78	0.5
Total	14,370	100
Mean	0.90	
Variance	1.67	

Table 2 Model selection criteria for the count regression models

Model	Def.	AIC	BIC	Deviance
Poisson	32	30218.3	30460.7	30154.3
NB	33	30198.2	30448.1	30132.2
ZIP	50	29746.6	30125.3	29646.6
ZINB	51	29737.1	30123.3	29635.1
HP	50	29457	29828	29359
HNB	51	29427	29813.2	29325

AIC, Akaike's information criterion; BIC, Bayesian information criterion; HNB, hurdle negative binomial; HP; Hurdle Poisson; NB, negative binomial; ZINB, zero-inflated negative binomial; ZIP, zero-inflated Poisson

Table 3 Zero counts in the count model

	Observed	Poisson	NB	ZIP	ZINB	HP	HNB
No of zeros	7720	7535.94	7629.34	7878.27	7891.22	7720	7720

HNB, hurdle negative binomial; HP; Hurdle Poisson; NB, negative binomial; ZINB, zero-inflated negative binomial; ZIP, zero-inflated Poisson

Socio-demographic and economic characteristics of under-five mortality in Ethiopia

Some of the socioeconomic, demographic, health and environmental-related factors on the child death per mother are summarized in Table 1. The majority, 80.9% of child death occurs with uneducated mothers and 3.7% of deaths occur with mothers who have a secondary level education or above. A part, (66.2%) of the child death occur with uneducated fathers and 7.7% of the deaths occur with parents who attained secondary and above. The highest percent of child death (25.8%) is observed with children whose order of birth is four and above and the lowest percent (38.5%) of the death is observed with children whose order of birth first. It is also noted that (58.6%) of the child death is attributed to poor women. Besides, the lowest percent of child death occurred with mothers whose visit to a health care institution during pregnancy is at least 4 times (85.0%) while the highest percentage of child death occurred with mothers who did not receive any antenatal check during pregnancy (7.3%). Working fathers have a lower percent of child deaths (46.6%) as compared to non-working fathers (53.4%) (Additional file 1: Table S1).

Model selection criteria

AIC, BIC, and Deviance statistics were used to identify the appropriate count model among the six commonly considered count models from the results in Table 2, we can observe that the log-normal survival model has the smallest AIC, BIC, and Deviance statistic. This indicates that the Hurdle negative binomial survival model is a better fit to data as compared to the other count regression models (Table 2).

Table 4 indicated the zero counts captured by six-count regression models. Poisson and the NB model under-estimated zero counts, the zero-inflated models over-estimated zero counts and the hurdle models captured all zero values. Hurdle based model was better estimated than the other count model (Table 3).

Figure 1 and Table 4 shows the predictive probabilities distribution curve of six-count regression models and the observed proportions. The hurdle negative binomial model is

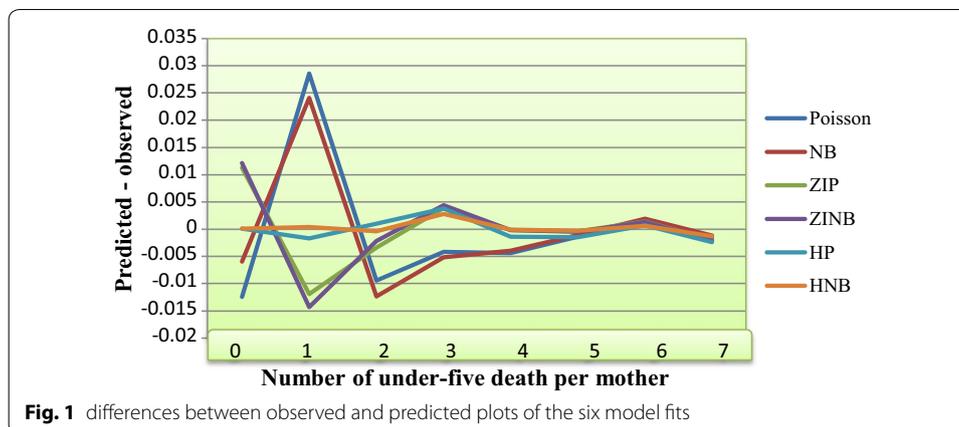


Table 4 Observed and predicted probabilities from the count model for the under-five death

No of under-five death	Observed probability	Predicted probability					
		Poisson	NB	ZIP	ZINB	HP	HNB
0	0.5370	0.5244	0.5309	0.5482	0.5491	0.5370	0.5370
1	0.2310	0.2596	0.2551	0.2189	0.2165	0.2292	0.2313
2	0.1230	0.1134	0.1105	0.1195	0.1207	0.1239	0.1225
3	0.0560	0.0517	0.0507	0.0598	0.0603	0.0597	0.0587
4	0.0290	0.0245	0.0249	0.0287	0.0287	0.0275	0.0288
5	0.0140	0.0125	0.0128	0.0135	0.0134	0.0124	0.0136
6	0.0050	0.0064	0.0068	0.0063	0.0062	0.0056	0.0055
7	0.0050	0.0033	0.0037	0.0029	0.0028	0.0025	0.0064

HNB, hurdle negative binomial; HP; Hurdle Poisson; NB, negative binomial; ZINB, zero-inflated negative binomial; ZIP, zero-inflated Poisson

a better choice than the other count models as the predicted probability for the hurdle negative binomial model is closed to the observed probability. Therefore, it is possible to conclude that the HNB model is more appropriate than other count models in terms of fitting the number of under-five deaths per mother.

Factors associated with under-five mortality in Ethiopia

Table 2 presents summaries of the Negative binomial hurdle model. The negative binomial component shows the magnitude of severity of under-five mortality. Compared with non-vaccinated children, the rates of non-zero under-five death for vaccinated children decreased by 28.5% (IRR = 0.715, 95% CI 0.630, 0.811). compared to children born in Tigray region, the risk of under-five child mortality is 1.475 times (IRR = 1.475, 95% CI 1.286, 1.678) higher among under-five children born to mothers from Benishangul-Gumuz regions. Relative to children whose mothers did not receive any antenatal follow up during pregnancy, the rate of non-zero under-five death whose mothers ANC visit at least 4 times during pregnancy was decreased by 22.3% (IRR = 0.777, 95% CI 0.671, 0.91). As birth order increases the under-five mortality also increases. The death rate of non-zero under-five whose order of birth is four and above increased by 47.9% (IRR = 1.621, 95% CI 1.496, 1.756) as compared to children have a first order of birth.

The result, also indicated that the educational level of mother's and father's was a significant factor for affecting the number of under-five deaths. Compared with non-educated mothers, among mothers who have a primary level of education, the rate of non-zero under-five death decreased by 23.4% (IRR = 0.766, 95% CI 0.700, 0.837). Similarly, compared to non-educated fathers, among fathers with secondary and above level of education, the incidence rate of non-zero under-five death for father's attended secondary and above education was decreased by 32% (IRR = 0.680, 95% CI 0.586, 0.789). The result also showed that the incidence rate of non-zero under-five death in multiple births was 1.471 times (IRR = 1.471, 95% CI 1.359, 1.593) that of the single births. Interpretation of other significant variables was done in a similar way (Additional file 2: Table S2).

The zero-inflated hurdle negative binomial part also indicated that the estimated odds that the number of under-five death becomes zero with vaccinated children increased by 78.1% (AOR = 1.781; 95% CI 1.587, 1.999) as compared to non-vaccinated children. An increase in family size by 1, the estimated odds that the number of under-five deaths become zero increased by 22.6% (AOR = 1.226; 95% CI 1.200, 1.253). The result also revealed that the estimated odds that the number of under-five deaths become zero with mothers who made ANC visit at least 4 times during the pregnancy was 2.316 times (AOR = 1.316; 95% CI 2.019, 2.657) that of children whose mothers who have not have any antenatal follow up. The estimated odds that the number of under-five deaths with mothers using contraceptives was 1.291 times (AOR = 1.291; 95% CI 1.161, 1.437) that of mothers who didn't use contraceptive. The analysis further indicated that the estimated odds the number of under-five death becomes zero with children who are born in the private sector was 2.083 times (AOR = 2.083; 95% CI: 1.546, 2.807) that of children born at home. The probability of under-five deaths decreased with increase in the educational level of the mother. The estimated odds that the number of under-five deaths become zero with mothers who have secondary and above education was 1.500 times (AOR = 1.500; 95% CI 1.212, 1.855) that of the non-educated mothers (Table 2).

Discussion

The objective of this study was to investigate the important risk factors of under-five child mortality in Ethiopia using the latest (EDHS-2016) dataset. Identified these factors help policymakers, expedite the efforts made towards making life better and evaluate progresses made towards achieving the MDG4. The outcomes of this study shed more light on the risk factors of under-five death in Ethiopia. The study for instance determined that parent level of education is an important socio-economic predictor of under-five mortality, that is, mortality rate decreased with increase in the level of education of the parents. Thus, it is found that an educated father takes a better care of the under-five child and his wife during antenatal and postnatal periods. This is line with a finding from a previous study which found that higher level of maternal and father education, lowers child mortality [21, 33–36]. Relative to single births, the risk of under-five deaths due to multiple births is very high, this study is similar to previous studies in finding out the link between birth type and higher risk of child mortality [12, 21, 35–37]. The study found that under-five mortality decreased as the length of the preceding birth interval increased which is consistent with different findings [11, 12, 21].

Moreover, the study revealed that death of under-five children from mothers who use contraceptive was significantly less than that of children from mothers who do not use contraceptive which agree with prior study findings by [11, 15]. Similar to findings of study by [37], vaccinated children had a risk of mortality that is lower than that of the non-vaccinated.

The estimated result showed that as mothers age at first birth increases, the risk of under-five mortality reduced, and similar to findings from previous studies by [11, 12, 15, 34, 35] Mothers who had their first child at a younger age have a higher under-five mortality risk. In addition to this, similar to reports of findings by [34, 35], this reported that for every unit increase in the ages of a mother, the risk of under-five mortality increased. The result also revealed that children born from working fathers had a higher risk of mortality than those of from non-working fathers. In a developing country like Ethiopia, a father has more responsibility for household income and didn't have enough time to care for his children. More importantly, in Ethiopia more than 80% the population is rural and their farming activities denies them time for caring their children, a finding that is consistent with [11]. Further, it is learnt that increase the number of antenatal visits during pregnancy reduces the risk of under-five mortality, a finding that is also confirmed by previous researches [11]. Children born in a healthcare facility that is in the public or private sectors were at lower risk than those born at home. This might be due to the proper health care and attention these facilities provide them during and after delivery, a finding which is confirmed by other studies as well [8, 11, 12].

The study also revealed that household size is an important variable affecting the number of under-five mortality. As household size increased the risk of under-five mortality decreased significantly, and this finding is consistent with [15, 35, 37, 38]. Similar to studies by [12, 37, 38] which found under-five mortality increase with increase in household size. The under-five child mortality in this study significantly and positively associated with the birth order.

In this study, the Deviance, AIC, and BIC statistic and predictive probability curve indicated that the Hurdle negative binomial model was the best model for the number of under-five death with about 53.7% zero counts. Several studies reported similar results that the Hurdle negative binomial model was the best model for count outcomes [39–42].

Conclusions

This study aimed to visualize and identify the responsible factors associated with the number of under-five child mortality in Ethiopia using the latest DHS data, which is essential for formulating appropriate health programs and policy implications. In general, the report showed that despite a concerted effort by the government of Ethiopia and several stakeholders in the health sector to improve under-five child survival, the child death rate did not decrease. The mean number of under-five deaths was 0.9 with a variance of 1.697, indicating that the data is over dispersed. The hurdle negative binomial model had the smallest AIC, Deviance, and BIC indicating that it is the best goodness of fit. Besides, the predictive value and probabilities for many counts in the hurdle negative binomial model fitted the observed counts best. This study revealed that region, mother's age, education level of a father, education level of the mother, fathers occupation, family

size, age of mother at first birth, vaccination of the child, contraceptive use, birth order, preceding birth interval, child twins, place of delivery, and antenatal visit were important determinants of number of under-five death in Ethiopia. The Hurdle negative binomial model provided a better fit for the data. We applied the Hurdle negative binomial model for count data with excess zeros of unknown sources such as the number of under-five deaths. The government, policymakers, and health care department are recommended that they should focus on the parental education and awareness, contraceptive use and breastfeeding practices is important to reduce under-five mortality and be in line with MDG4. Moreover, the study found that early marriage is associated with child mortality, and the government and the ministry of health should aware and educate about the impacts of early marriage on child mortality.

Limitations

Some variables are not included because of the large number of missing values like the weight of child at birth, anemia, and the size of child at birth. The interaction term is not considered in this study due to the convergence issue.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s40537-020-00339-0>.

Additional file 1: Table S1. Socio-demographic and economic characteristics of under-five mortality in Ethiopia.

Additional file 2: Table S2: Hurdle negative binomial regression coefficients for the number of under-five mortality in Ethiopia.

Abbreviations

ANC: Antenatal care; CSA: Central statistical agency; EDHS: Ethiopian Demographic and Health Survey; HNB: Hurdle negative binomial; HP: Hurdle poisson; LRT: Likelihood ratio test; NB: Negative binomial; SDG: Sustainable development goal; ZINB: Zero-inflated negative binomial; ZIP: Zero-inflated poisson; MDG4: Millennium development goal 4.

Acknowledgements

We would like to thank the Ministry of Health and Central Statistical Agency, Government of Ethiopia, for making the data freely available for research purposes. In addition, authors would like to acknowledge Berhanu Engidaw (Assist. Professor) from the English department of Bahir Dar University for providing language reviewing and editing service for the manuscript.

Future research implications

While several studies that deployed different statistical models and techniques were made on determinants of mortality among under-five children through different statistical models and techniques, little attention is provide to a comprehensive study that examines the patterns and associated risk factors of the under-five mortality, as well as its change over time and space. Therefore, we plan to conduct further study to determine the factors and trends of the number of under-five children using spatiotemporal multivariate decomposition analysis. This is useful to locate the hot and cold area of the under-five mortality of children in Ethiopia in order to be able to formulate appropriate health programs and policies.

Authors' contribution

SMF drafted the proposal, did the analysis, wrote the results, and prepared the manuscript. HMF participated in editing, analysis, and write up of the result and GMA critically revised the manuscript for its scientific content. All authors read and approved the final manuscript.

Funding

None.

Availability of data and materials

The data set was accessed from the Measure DHS website (<http://www.measuredhs.com>).

Ethics approval and consent to participate

The study used available secondary data accessed under the National Data Sharing and Accessibility Policy (NDSAP) of the Government of Ethiopia. The data set had no identifiable information on the survey participants; therefore, no ethical approval is required for this work.

Consent for publication

Not applicable.

Competing interests

The authors declare that no competing interests exist.

Author details

¹ Department of Statistics, College of Science, Debre Tabor University, Debre Tabor, Ethiopia. ² Department of Statistics, College of Science, Bahir Dar University, Bahir Dar, Ethiopia. ³ Research and technology transfer directorate, Amhara Public Health Institute, P.O. Box 477, Bahir Dar, Ethiopia.

Received: 20 March 2020 Accepted: 30 July 2020

Published online: 17 August 2020

References

1. You D, et al. Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Inter-agency Group for Child Mortality Estimation. *Lancet*. 2015;386(10010):2275–86.
2. WHO. Under five mortality rates. Global Health Observatory (GHO) Data. 2016.
3. Hug L, Sharrow D, You D. Levels & trends in child mortality: report 2017. Estimates developed by the UN Inter-agency Group for Child Mortality Estimation. 2017.
4. WHO, Children: reducing mortality. 2017.
5. USAID, Maternal, Neonatal and Child Health In Ethiopia 2018.
6. EDHS, Ethiopian Demographic and Health Survey 2016.
7. Nasejje JB, Mwambi HG, Achia TN. Understanding the determinants of under-five child mortality in Uganda including the estimation of unobserved household and community effects using both frequentist and Bayesian survival analysis approaches. *BMC public health*. 2015;15(1):1003.
8. Muriithi DM, Muriithi DK. Determination of infant and child mortality in Kenya using cox-proportional hazard model. *Am J Theoret Appl Stat*. 2015;4(5):404–13.
9. Mekonnen D. Infant and child mortality in Ethiopia. The role of socioeconomic, demographic and biological factors in the previous 5 years period of 2000 and 2005, 2011.
10. Madise NJ, Banda EM, Benaya KW. Infant mortality in Zambia: socioeconomic and demographic correlates. *Soc Biol*. 2003;50(1–2):148–66.
11. Getachew Y, Bekele S. Survival analysis of under-five mortality of children and its associated risk factors in Ethiopia. *J Biosens Bioelectron*. 2016;7(213):2.
12. Bereka SG, Habtewold FG. Under-five mortality of children and its determinants in Ethiopian Somali Regional State, Eastern Ethiopia. *Health Sci J*. 2017;11(3):1.
13. Adeolu M, et al. Environmental and socioeconomic determinants of child mortality: evidence from the 2013 Nigerian demographic health survey. *Am J Public Health Res*. 2016;4(4):134–41.
14. Kanmiki EW, et al. Socio-economic and demographic determinants of under-five mortality in rural northern Ghana. *BMC Int Health Human Rights*. 2014;14(1):24.
15. Bedada D. Determinant of under-five child mortality in Ethiopia. *Am J Stati Probabil*. 2017;2(2):12–8.
16. Sellers KF, Shmueli G. Data dispersion: now you see it... now you don't. *Commun Stat-Theory Methods*. 2013;42(17):3134–47.
17. Hilbe JM. Negative binomial regression. Cambridge: Cambridge University Press; 2011.
18. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*. 1992;34(1):1–14.
19. Gurmu S. Generalized hurdle count data regression models. *Econ Lett*. 1998;58(3):263–8.
20. Ayele DG, Zewotir TT, Mwambi H. Survival analysis of under-five mortality using Cox and frailty models in Ethiopia. *J Health Popul Nutr*. 2017;36(1):25.
21. Gebretsadik S, Gabreyohannes E. Determinants of under-five mortality in high mortality regions of Ethiopia: an analysis of the 2011 Ethiopia Demographic and Health Survey data. *Int J Popul Res*. 2016. 2016.
22. Aheto JMK. Predictive model and determinants of under-five child mortality: evidence from the 2014 Ghana demographic and health survey. *BMC Public Health*. 2019;19(1):64.
23. Hoffman JP. Generalized linear models: an applied approach. Harlow: Pearson Education; 2004.
24. Molla, D.T. and B. Muniswamy. Power of tests for negative binomial regression coefficients in count data. *Int J Mathe Archive*. ISSN 2229–5046 [A UGC Approved Journal], 2012. 3(8).
25. Jansakul N, Hinde J. Score tests for zero-inflated Poisson models. *Comput Stat Data Anal*. 2002;40(1):75–96.
26. Afifi AA, et al. Methods for improving regression analysis for skewed continuous or counted responses. *Annu Rev Public Health*. 2007;28:95–111.
27. Agarwal DK, Gelfand AE, Citron-Pousty S. Zero-inflated models with application to spatial count data. *Environ Ecol Stat*. 2002;9(4):341–55.
28. Long JS, Freese J. Regression models for categorical dependent variables using Stata. 2006: Stata press.
29. Zuur AF, et al. Zero-truncated and zero-inflated models for count data, in Mixed effects models and extensions in ecology with R. Cham: Springer; 2009. p. 261–93.
30. Lawless JF. Negative binomial and mixed poisson regression. *Canadian J Stat*. 1987;15(3):209–25.
31. Mullahy J. Specification and testing of some modified count data models. *J Econ*. 1986;33(3):341–65.

32. Ismail N, Zamani H. Estimation of claim count data using negative binomial, generalized poisson, zero-inflated negative binomial and zero-inflated generalized poisson regression models. In *Casualty Actuarial Society E-Forum*. 2013.
33. Khan JR, Awan N. A comprehensive analysis on child mortality and its determinants in Bangladesh using frailty models. *Arch Public Health*. 2017;75(1):58.
34. Yaya S, et al. Prevalence and determinants of childhood mortality in Nigeria. *BMC Public Health*. 2017;17(1):485.
35. Alam M, et al. Statistical modeling of the number of deaths of children in Bangladesh. *Biom Biostat Int J*. 2014;1(3):00014.
36. Rahman MS, Rahman MS, Rahman MA. Determinants of death among under-5 children in Bangladesh. *J Res Opinion*. 2019;6(3):2294–302.
37. Berhie KA, Yirtaw TG. Statistical analysis on the determinants of under five mortality in Ethiopia. *Am J Theoret Appl Stat*. 2017;6(1):10–21.
38. Ahmed Z, Kamal A, Kamal A. Statistical analysis of factors affecting child mortality in Pakistan. *JCPSP*. 2016;26(6):543.
39. Yusuf Olushola K. Statistical modeling of fertility experience among women of reproductive age in Nigeria. *Int J Stat Appl*. 2018;8(1):23–33.
40. Hidayat B, Pokhrel S. The selection of an appropriate count data model for modelling health insurance and health care demand: case of Indonesia. *Int J Environ Res Public Health*. 2010;7(1):9–27.
41. Kazembe LN. A Bayesian two part model applied to analyze risk factors of adult mortality with application to data from Namibia. *PLoS ONE*. 2013;8(9):e73500.
42. Gebremedhin TA, Mohanty I. Child schooling in Ethiopia: the role of maternal autonomy. *PLoS ONE*. 2016;11(12):e0167639.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
