INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

Received: 03-08-2023

Accepted: 28-09-2023

Published: 28-10-2023

Citation: Baruah R, Barman MP, Bora K, Hazarika J (2023) Determinants of Infant Mortality in Assam, India: Modelling with Generalised Linear Regression Model for Count Data. Indian Journal of Science and Technology 16(40): 3538-3548. https://doi.org/10.17485/IJST/v16i40.1948

* **Corresponding author**.

manash_stats@dibru.ac.in

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Determinants of Infant Mortality in Assam, India: Modelling with Generalised Linear Regression Model for Count Data

Rituraj Baruah[1], Manash Pratim Barman[2]*, Krishnarjun Bora[3], Jiten Hazarika[4]

1 Assistant Professor, Department of Statistics, Jagannath Barooah College, Jorhat, Assam, India
2 Head of the Department, Department of Statistics, Dibrugarh University, Dibrugarh, Assam, India
3 Research Scholar, Department of Statistics, Dibrugarh University, Dibrugarh, Assam, India
4 Vice-Chancellor, Dibrugarh University, Dibrugarh, Assam, India

Abstract

Objectives: Compared to other states of the country, infant mortality situation in Assam is still substandard. This study attempts to model infant mortality counts, determine the associated risk factors in Assam. **Methods:** Data have been extracted from the NFHS-5 of India. The "number of infant deaths", is the explained variable used in the present study with several explanatory variables. 34,979 mothers in 33 districts of Assam are included to train the count model. The fitted models are compared by various techniques like Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC), Receiver Operating Characteristic (ROC) curves and Area under the curve (AUC). **Findings:** In the data set, 89.72% of the mothers have not faced any infant deaths in their lifetime. The average number of infant deaths is 0.39, with a variance of 0.689, indicating that overdispersion is present. However, the overdispersion could be due to observational variation or excess zeros. The likelihood ratio chi-square test is used to determine the significance of the inflation parameter. The result of the statistic is 312.32 with an associated P-value of 0.000, so the inflation parameter is significant. Hence, it is better to use a model that considers excess zeros. Among the models, the zero-inflated negative binomial model with the least AIC (34781.56) and BIC (35217.86) is considered to be a better-fitting model than the other candidate models to meet the objective. Moreover, the zero-inflated negative binomial model (ZINB) has the greatest AUC (0.7196). The regression analysis indicates that residential status, religion, wealth index, mother's age at first birth, mother's educational level, and birth order significantly influence the risk of infant mortality. **Novelty:** The findings of this study might help policymakers to identify which socio-economic and demographic groups should be given preference to encourage women and to reduce infant mortality.

Keywords: Count; Overdispersion; Infant Mortality; Assam; Zero Inflated Negative Binomial Regression

1 Introduction

The infant mortality rate is regarded as a leading public health problem and a sensitive indicator of the health status in low and middle-income countries. It is the number of deaths of children under one year old in a given year per 1,000 live births in the same year. In a developing country like India, it reflects the general standard of living of the people and is frequently regarded as a barometer of societal progress. Over the past few decades, India has made substantial progress in reducing infant mortality. Due to the rapid strides that India has taken in social, economic, health, and educational development, the infant mortality rate decreased from 81 to 34 per 1,000 live births between 1990 and 2016 [1]. Recently, according to the National Family Health Survey-5 conducted during 2019–21, the infant mortality rate in India has declined from 41 deaths per 1,000 live births to 35 deaths per 1,000 live births in the five years before the survey [1].

Assam is a state in north-eastern India, and the population of Assam is 3.12 crore, according to the 2011 census. Like the country, Assam has also made remarkable progress over the past few years in reducing the state's infant mortality. According to the National Family Health Survey-5 report, the infant mortality rate in Assam is 31.9 per 1,000 live births, down from 84 per 1,000 in 2005-06 [2]. Figure 1 shows the declining trend of infant mortality in Assam.

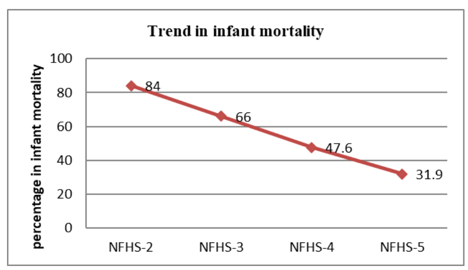

Fig 1. Percentage decline trend in infant mortality rate

Figure 1 depicts that there is a percentage decline in the trend of infant mortality in Assam from National Family Health Survey-2 to National Family Health Survey-5.

1.1 Importance of the Study

Assam is not only geographically isolated from the rest of the country due to inadequate infrastructure; it also has diverse socio-cultural child-rearing customs, which directly or indirectly contribute to child mortality in the region [3]. The Assam government has introduced a number of welfare and health schemes to reduce infant mortality. But women in rural areas of Assam are not sufficiently aware of the various health and welfare programmes connected to maternal and child health. These schemes can be rated partially successful in part because the people of Assam from remote areas still prefer institutional delivery (63.8 percent of women had delivery at home) to local midwives (4.80 percent of home deliveries were assisted by skilled persons) [4]. Assam's infant mortality rate has decreased over time; however, the situation is still unsatisfactory compared to some other regions of the country [4]. So it is urgent to analyse the associated risk factors of infant mortality in Assam.

Second, Cox proportional regression models or binary logistic regression models are the most popular and useful techniques to investigate the factors that influence the survival of infants. These models might be more recommended to determine the incidence of infant mortality a woman has experienced in her lifetime and contributing factors [5]. But, binary logistic regression undercounts the overall number of fatalities as multiple deaths are combined into a single unit to meet the requirements of the model. Moreover, most recommended binary logistic and Cox proportional regression models are not capable of handling excess zeros, overdispersion and heteroskedasticity present in the count data set [6]. Further, there is no evidence to model the zero-inflated, over-dispersed count data of infant mortality using recently released NFHS data in Assam. The novelty of the current study is that it fills this gap. Moreover, the National Family Health Survey (NFHS)-5 survey data (2019–2021), which were just released, makes this paper relevant.

Further, the general decline trend in the count of infant mortality has also created a renewed interest in analysing the risk factors behind the causes of infant mortality in Assam using count regression models such as the poisson regression model, the negative binomial regression model, zero-inflated models and hurdle models.

Keeping this points in mind, this work aims to model the counts and identify the risk factors of infant mortalities using the rounds of the National Family Health Survey-5 in Assam with the help of a generalized linear regression model for count data and draw inferences accordingly.

1.2 Related works

A number of countries conducted studies to evaluate the risk factors for infant mortality. The majority of this research employed Cox proportional regression models or binary logistic regression to investigate the factors that influence the mortality of infants [5]. Although binary logistic regression gives sufficient information on the factors causing numerous infant deaths, the model underestimates the risk of a count of infant deaths. Logistic regression approach splits the dependent variable into two groups: dead or alive. Multiple fatalities are combined into a single unit to meet the requirements for binary logistic regression. As a result, the number of deaths is underestimated. So, to overcome the problem, researchers now perhaps prefer the count regression technique to analyse the risk of infant mortality data [6]. For instance, Darnah et al. conducted a poisson regression analysis using a nonparametric regression approach based on a local linear estimator. This study aimed to model the count of infant mortality and maternal mortality cases in East Kalimantan, Indonesia. The modelling of both the maternal mortality case and the infant mortality scenario using the shaman number resulted in two estimated models of exponential functions meeting the goodness of fit criterion. This study observed that the model explained the association between the infant mortality case and the shaman number as well as the relationship between the maternal mortality case and the shaman number in East Kalimantan quite well [7]. Kumar and Shivgotra also carried out a research in some selected district of Jammu division, where the seasonal effect on infant death as a whole was measured using poisson and negative binomial regression analysis. This study concluded that seasonal variations significantly affect the infant mortality. The findings demonstrated that the incidence of infant deaths for the month of January was high when compared to the reference month of June. However, compared to the reference month, the mean incidence of deaths was a little higher in the months of March and October [8]. Prahutama et al. developed a bivariate poisson regression model for maternal and infant mortality in Central Java. The expectation maximisation algorithm was employed to estimate the model. The percentage of pregnant women using vitamin K1, the percentage of pregnant women using vitamin K4, the percentage of pregnant women who received Fe3 tablets, the percentage of births assisted by medical personnel, the percentage of obstetric complications handled, the percentage of childbirth women who had puerperal health services, and the percentage of households with clean and healthy behaviour were the variables in this model that had a significant impact on infant mortality [9]. Saputro and Qudratullah carried out a case study in Wonogiri, Indonesia, to determine the factors influencing infant mortality counts. A zero-inflated negative binomial model was employed for the study. The findings revealed an association between a high risk of pregnancy and infant mortality counts [10]. To determine the contributing variables to perinatal death in Ethiopia, Aragaw et al. used the poisson-logit hurdle model. A cross-sectional study was carried out in Ethiopia. The main protective factors associated with this study were a long preceding birth interval and husband education. Higher parity, rural residence, Caesarean section delivery, multiple pregnancies, institutional delivery, and having a history of abortion increased the perinatal mortality factor per mother [11]. To calculate the risk factors for infant mortality in Ethiopia, Mulugeta et al. employed a multilevel count regression model. Among the multilevel log linear models, the zero-inflated negative binomial model was chosen as the best model to estimate the risk factors associated with infant mortality. The outcomes of this study provided insight into Ethiopia's infant mortality rates. Infant mortality was found to be significantly influenced by the residential status, the mother's age at delivery, the size of the household, the mother's age at her first childbirth, breastfeeding, the weight of the child at birth, the use of contraceptives, the birth order, the wealth index, the father's education level, multiple births, and the birth interval [6]. To reduce and analyse the causative factors of infant mortality in Indonesia, a geographically weighted Poisson regression (GWPR) with a weighted bi-square kernel function was employed. Several independent variables like low birth weight, receiving vitamin A, exclusive breast milk, immunization and medical assistant. The number of infant deaths in South Sulawesi Province was divided into seven categories based on relevant variables, according to the partial estimation with the geographically weighted Poisson regression (GWPR) model. The Selayar Islands were the first of the seven groups to be created, and all of the factors there had a significant effect [12].

2 Methodology

2.1 Sources of data and variables in the study

For analysing the count of infant deaths in Assam, data have been extracted from the fifth round of the National Family Health Survey-5 (NFHS-5) of India during 2019–21 [2]. National Family Health Survey-5 fieldwork was conducted in 33 districts of Assam from 17th June, 2019 to 21st December, 2019 by Nielsen India Pvt. Ltd. The National Family Health Survey has been conducted under the stewardship of the Ministry of Health and Family Welfare (MoHFW), Government of India. The Ministry of Health and Family Welfare designated the International Institute for Population Sciences (IIPS), Mumbai, as the nodal agency for the surveys.

The number of infant deaths, Y, is the explained variable used in the present study. Several explanatory variables in this study are considered as predictors, which are commonly reported in infant mortality studies. The study's explanatory variables include the mother's educational status (literate / illiterate), residential status (rural / urban), wealth index (poorest/ poorer/ middle/richer /richest), mother's age (in years) at first birth of her child (19-35/<19 and >35), caste (SC/ST / others), religion (Hindu/ Muslim/ /others), and birth order (1st/ 2nd/ 3rd/ 4th or more).

2.2. Methods of count data analysis

The count data like "number of infant deaths" are often almost skewed, non-negative, overdispersed, and include too many zeros. The use of diverse techniques and models for count data regressions has been prompted by these characteristics. As a result, this study offer a useful methods for modelling count data that focuses on data with extra zeroes and overdispersion [5,13]. Poisson regression analysis is base model to deal with count data. Due to the over-dispersion and/or excess number of zeros that are frequently present in empirical count data sets, the standard poisson regression model for count data is frequently of little utility in these fields. As a result, when modelling count data, it is customary to go on to the analysis of correcting for overdispersion if it occurs after developing a poisson regression model. The Hurdleand zero-inflated poisson and negative binomial regression models can each be used to handle excess zeros in their own unique ways, while the negative binomial mode can cope with overdispersion. In the beginning, descriptive statistics were used to gather data on socioeconomic factors as well as demographic characteristics. To select the appropriate count model and evaluate the models performance, tests like (i) Vuong test (ii) Likelihood Ratio test (iii) Akaike Information Criteria (AIC)(iv) Bayesian Information Criteria (BIC) (v) Receiver Operating Characteristic curve (ROC) and (vi) the Area Under the Curve(AUC) for different models are performed [13,14]. The Area Under the Curve(AUC) [14] is widely used to measure the accuracy of diagnostic tests. The closer the Receiver Operating Characteristic (ROC) curve is to the upper left corner of the graph, the higher the accuracy of the test. So, the Area Under the Curve = 1.0 is the value of the ideal Receiver Operating Characteristic curve. Additionally, the curve with the largest the area under the curve (AUC) is thought to have a better diagnostic performance. Following that, bi-variate analysis based on Chi square [15] is performed to illustrate the variations infant mortality caused by the chosen predictor variables. Finally, appropriate count model analysis is conducted to show the association between socio-economic and demographic variables and incidence of under five mortality.

3 Results and Discussion

3.1 Background Characteristics of the respondents

So far, socio economic and demographic characteristics are considered, in this study, rural areas have a higher percentage of women overall (85.01%) than urban areas (14.9%). Hinduism makes up the majority of the women's faiths (62.01%). Compared to ST and SC, the bulk of the women (73.89%) come from other castes (general and OBC). According to the wealth index, the greatest percentage of women (38.41%) is in the poorer category, while the lowest percentage (4.15%) is in the richest category. When a mother's age during her first birth is taken into account, the age range of (19 – 35) years accounts for the highest percentage (60.73%). The majority of women (72.09%) are literate. When the birth order of their child is taken into account, the majority of respondents (40.09%) fall into the first-order birth group

3.2 Development of count data regression models for determinants of infant mortality: Model identification

In this study, count of infant deaths is the dependent variable with zero responses. If the respondents/mothers didn't experience the death of their child before their first birthday, then the only possible outcome is zero. If the respondents experienced a death, then it is a counting process. In the present study, it is seen that 89.72% of the mothers have not faced any infant deaths in their lifetime. Figure 2 depicts the percentage distribution of count of infant deaths a mother has experienced in her lifetime.

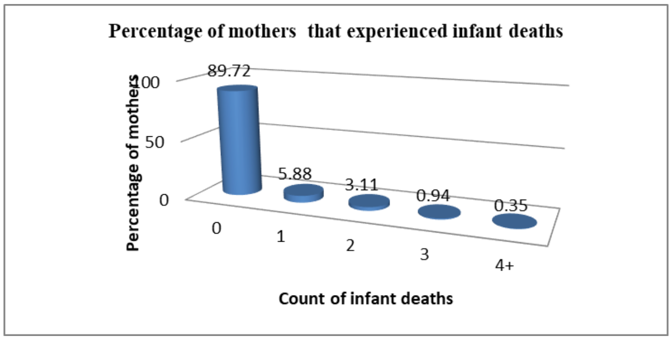

Fig 2. Percentage distribution of number of infant deaths a woman has experienced in her life time

The preponderance of zeros in the observed data is an indication that the data can be fitted better by the count regression model, which takes into account excess zeros. The mean and variance of the dependent variable (number of infant deaths) are.39 and.689, respectively, so, the dependent variable is over-dispersed.

At the outset, a poisson regression model is fitted, yielding model deviance and Pearson chi-square values of.873 and 1.636, respectively. Since these values are not so near 1(one), the fit of the poisson regression is not entirely satisfactory, and overdispersion is present in the data set. However, the overdispersion could be due to observational variation or excess zeros. As the reason for overdispersion is not known, it is needed to fit the zero-inflated model to identify whether the overdispersion is due to the presence of excess zeros. The likelihood ratio chi-square test statistic value for testing the inflation parameter is 312.32 with an associated P-value of.000, so it is significant. Hence, it is better to use a model that considers excess zeros. Different zero-inflated models are considered in this study, namely: the zero-inflated poisson model, the zero-inflated negative binomial model, the poisson hurdle model, and the negative binomial hurdle model. By examining the null hypothesis that both models are equally comparable to the observed distribution, the Vuong test compares the zero inflated poisson model to the normal poisson model. The obtained statistic is statistically significant (V=8.24, p-value 0.001) and showed that the zero inflated poisson model, which includes excess zeroes, better represents the observed data than normal poisson model. Similar results are obtained using the Vuong test to compare zero inflated negative binomial model to negative binomial model, which produces a test statistic of V=2.73 and a p-value of 0.001. As a result, when compared to the conventional negative binomial model, the Zero-inflated negative binomial regression model more accurately reflects the data on the number of infant deaths. In addition, Table 1 summarises the statistics to compare the goodness of fit of the different zero-inflated models.

Table 1. Model Comparison Criteria

Models	Criteria	
	AIC	BIC
Zero inflated poisson regression model (ZIP)	35227.51	35852.31
Zero-inflated negative binomial regression model(ZINB)	34781.56	35217.86
Poisson hurdle model (HP)	35974.98	35877.69
Negative binomial hurdle model (HNB)	35208.70	35229.80

Source: Estimated by the authors from the National Family Health Survey, 2019-21.
Notes: AIC-Akaike Information Criteria & BIC-Bayesian Information Criteria

The likelihood test can be used to compare the zero-inflated negative binomial and zero-inflated poisson models because they are nested. The zero-inflated poisson regression model yielded a log-likelihood of -17293.02, while the zero-inflated negative binomial model yielded -17039.73. The likelihood ratio test statistic is 506.58 with an associated P-value .000 and provides evidence for preferring the zero-inflated negative binomial over the zero-inflated poisson regression model. Further Table 1 provides the fact that zero inflated negative binomial model better predicts the data than zero-inflated poisson regression models, poisson hurdle models, and negative binomial hurdle models, as the values for the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) are smaller than the other models. Moreover, the outcomes in Figure 3 also demonstrate that the zero inflated negative binomial model better predicts the number of infant deaths compared to other models.

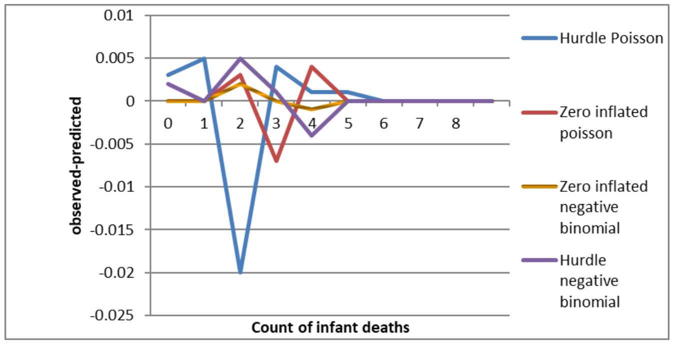

Fig 3. Plots of difference between the observed probability of each count and the prediction from each of the four models

Additionally, Receiver Operating Characteristic (ROC) curves and Area under the curve (AUC) are utilised in this study to evaluate the model's fitness or predictive capacity. Figure 4 displays the ROC curve and AUC of all the models under study.

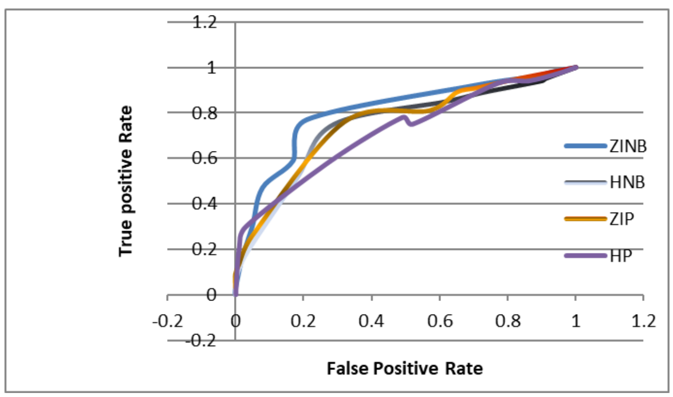

Fig 4. ROC Curve **Notes:** ZINB: Zero inflated negative binomial, HNB: Hurdle negative binomial, ZIP: Zero inflated poisson, HP: Hurdl poisson regression

Here also, it is found that, zero inflated negative binomial model(ZINB) has the greatest AUC (0.7196), followed by negative binomial hurdle(HNB) model (0.7018), zero inflated poisson(ZIP) model (0.6974), and poisson hurdle(HP) model (0.6189). Therefore, the zero-inflated negative binomial regression model is found to be the most appropriate model that fits the data for the present study. It can be regarded as the best and most superior model among the other potential candidate models.

3.3 Bi-variate analysis

To investigate the relationship between the dependent variable and each of the chosen predictors, bivariate analyses (based on the Pearson Chi square test) are carried out. All the predictors-location (χ^2=55.75), religion (χ^2=129.85), caste (χ^2=34.86), wealth index (χ^2= 194.76), mother's age during her first birth (χ^2= 21.85) educational status of mother (χ^2= 49.32) and birth order (χ^2= 12.97) have a meaningful relationship to the outcome variable (p<.001) and so, every relevant predictor is incorporated into the zero inflated negative binomial model.

Risk factors (based on zero inflated negative binomial model)

Now let us look into the findings of the zero inflated negative binomial regression model, which have been presented in Tables 2 and 3 and results are interpreted accordingly.

Zero inflated negative binomial regression analysis (A: Count Group (Table 2) & B: Inflation Zero Group (Table 3)) for assessing the socioeconomic and demographic correlates of infant mortality.

Table 2. A: Count Group. Zero inflated negative binomial regression analysis for assessing the socioeconomic and demographic correlates of infant mortality

Count Group		**IRR**	**95% Confidence interval**	
			Lower	**Upper**
Variables				
Constant		0.568***	0.248	0.752
Demographic variables of Mothers				
Residence	Urban (Ref.)			
	Rural	1.282***	1.169	2.235
Religion	Hindu (Ref.)			
	Muslim	0.832***	0.586	0.905
	Others	1.130	0.932	1.760
Cast	Other (Ref.)			
	Scheduled Tribe	0.762	0.744	0.881
	Scheduled Caste	0.653	0.527	0.719
Age of mother at her first birth	19-35 (Ref.)			
	<19 years & >35 years	1.322***	1.084	1.742
Socioeconomic Variables of Mothers				
Wealth Index	Richest (Ref.)			
	Richer	1.033***	0.809	1.109
	Middle	1.090***	0.882	1.105
	Poorer	1.196**	0.933	1.295
	Poorest	1.365*	1.084	1.514
Education	Literate (Ref.)			
	Illiterate	1.190**	0.944	1.623
Child related Variable				
Birth Order	1^{st} Order (Ref.)			
	2^{nd} Order	1.109	0.976	1.254
	3^{rd} Order	1.169**	0.078	1.298
	$\geq 4^{th}$ order	1.243***	1.049	1.697

Place of residence: The expected count of infant mortality among rural mothers is 28.2% higher compared to urban mothers, while holding all other variables constant. When the inflated zero group is considered, the risk of infant mortality for mothers who belong to a rural locality is ($e^{.473}$ (OR) = 1.60, P =.000) 1.60 times higher than for mothers in an urban locality.

Religion: While holding all other variables constant, the expected count of infant deaths is 17% lower for mothers who belong to Muslim religions compared to Hindu religions. For the inflated zero group, when religion is taken into account, the risk of infant mortality is .858 times lower for the Muslim community ($e^{-.152}$ (OR) =.858, P =.000) compared to Hindu mothers.

Wealth Index: The expected count of infant deaths for women belonging to the richer, middle, poorer, and poorest categories is respectively 3%, 9%, 19%, and 36% more than those from the reference category (the richest category), while holding all other variables constant. Again, for the zero inflation group, the risk of infant mortality is respectively 1.120 times higher, 1.190 times higher, 1.302 times higher, and 1.436 times higher for mothers in richer ($e^{.114}$(OR) = 1.120, P =.000), middle ($e^{.174}$ (OR) = 1.190, P =.000), poorer ($e^{.264}$ (OR)= 1.302, P =.029), and poorest ($e^{.362}$ (OR)=1.436 , P =.048) families compared to richest category.

Age: The expected count of infant deaths for mothers aged less than 19 years and more than 35 years at their first birth is 32% higher than for the reference group (19–35). Again, the risk of infant mortality is 1.27 times higher ($e^{.242}$ = 1.27, P =.000)

Table 3. B:Inflation Zero Group: Zero inflated negative binomial regression analysis for assessing the socioeconomic and demographic correlates of infant mortality

Inflation Zero Group			95% Confidence interval	
		Coefficient	**Lower**	**Upper**
Variables				
Constant		-.992**	-1.72	-0.261
Demographic variables of Mothers				
Residence	Urban (Ref.)			
	Rural	0.473***	0.387	0.852
Religion	Hindu (Ref.)			
	Muslim	-0.152***	-0.212	-0.072
	Others	-1.670	-3.874	0.084
Caste	Other (Ref.)			
	Scheduled Tribe	-0.164	-0.375	0.125
	Scheduled Caste	-0.339	-0.365	0.559
Age of mother at her first birth	19-35 (Ref.)			
	<19 years & >35 years	0.242***	0.165	0.528
Socioeconomic Variables of Mothers				
Wealth Index	Richest(Ref.)			
	Richer	0.114***	0.108	0.229
	Middle	0.174***	0.122	0.334
	Poorer	0.264*	0.209	0.428
	Poorest	0.362*	0.232	0.542
Education	Literate (Ref.)			
	Illiterate	0.124***	0.096	0.473
Child related Variable				
Birth Order	1^{st} Order (Ref.)			
	2^{nd} Order	0.117	-0.039	0.137
	3^{rd} Order	0.184***	0.074	0.575
	$\geq 4^{th}$ order	0.261**	0.125	0.339

Source: Estimated by the authors from the National Family Health Survey, 2019-21. **Notes:** IRR-Incidence Rate Ratio; CI-Confidence interval; Ref-reference category; ***p<0.001, **p<0.01, * p<0.05 significant.

for mothers aged less than 19 years and more than 35 years at the time of their first birth in the zero inflation group compared to mothers aged (19–35) years.

Level of Education: It is observed that the expected count of infant deaths for illiterate women is 19% higher than that of literate women. For the zero inflation group, when mothers' education level has been considered, infant mortality risk is 1.13 times higher for illiterate women ($e^{.124}$ (OR) = 1.13; P = .000) than their literate counterparts.

Birth Order: The expected count of infant deaths with birth order 3^{rd} and 4^{th} or more is respectively 16% and 24% higher compared to the reference category, i.e., for 1^{st} order birth. When the zero inflation group is considered, the risk of infant mortality is respectively 1.20 and 1.29 times higher for third ($e^{.184}$ = 1.20 (OR), P =.000) and fourth or more ($e^{.261}$ = 1.29, P =.002) birth orders compared to first-order births.

This research work employed a National Family Health Survey data set of 2019 to assess the risk factors associated with infant mortality in Assam, India using a zero inflated negative binomial model. The findings of this study shed light on causes of infant mortality in Assam.

From Tables 2 and 3, it is observed that there is a significant association between infant mortality and the place of residence of respondents, which is as commonly anticipated. This means that there is a difference in the risk of infant mortality depending on whether a woman belongs to a rural or urban area. Although rural scenarios in the rate of infant mortality have gotten better over the past few years (4), this study reveals that there is still a variation in the risk of infant mortality between rural and urban areas of Assam. This gap in the risk of infant mortality has been documented in numerous research studies (3,4). The Indian government has introduced a welfare and health scheme the National Health Mission (NHM), to reduce infant mortality in rural area [16]. Under the National Health Mission, schemes like JSY (Janani Suraksha Yojana) and JSSK (Janani Sishu Suraksha Karyakaram) are being implemented to lower the infant mortality rate, under-five mortality rate, and maternal mortality rate by providing women from caste groups and low-income households with conditional cash transfers for institutional deliveries [16]. But women in rural areas of Assam are not sufficiently aware of the various health and welfare programmes connected to maternal and child health. Some of the major barriers to access health services in rural areas of Assam include a lack of communication, ignorance, poor infrastructure, and sufficient education [4].

Second, the association of infant mortality with Muslim religion appears to be statistically significant in this study. According to this study, Muslims were more likely than Hindus to be alive at birth. This result is in line with the earlier findings of a life expectancy study at birth in India [17]. However, in contrast, an earlier study in northeast India highlighted an insignificant association in the risk of infant mortality between Hindu and Muslim religions [3]. The difference in the risk of infant mortality between the two religious groups may be due to birth weight differences in newborns between Hindu and Muslim married women. Birth weight is important since it has an impact on the child's growth and development, mortality, and morbidity. Muslims are more likely than Hindus to have infants who are normal or very heavy [18]. Moreover, the low use of maternal healthcare services by Muslims is not a recent problem. The National Health Mission (NHM) 2005 was a significant effort by the Indian government to enhance healthcare for pregnant women and infants and to close the gap between the country's many religious communities. This mission plays a significant role in increasing the rate of utilisation of maternal care services among Muslims [19].

However, so far as caste is concerned, the association with infant mortality seems to be statistically insignificant. This implies that women from scheduled castes and scheduled tribes are at the same risk of infant mortality as those from other groups (OBC or general). This may be because tea garden labourers in Assam belong to the OBC category. The infant mortality rate in tea gardens is very high, possibly because tea tribes place much emphasis on child marriage. Most teenage girls from tea communities experience economic and cultural pressure to get married young, between the ages of 15 and 18, after reaching puberty [20]. Adolescent mothers delivered babies who, on average, had lower birth weights, baseline anthropometric measurements, similar growth outcomes from enrolment to 6 months, and increased odds of all-cause mortality at 6 months [21]. Another study about the consequences of child marriage in Assam found that the children of teen mothers experience serious health problems [22]. These reports can justify the possibility of a higher risk of infant deaths among the tea garden labourers in Assam, for which in the present study an insignificant difference may be observed in the incidence of infant mortalities between other castes and scheduled castes or scheduled tribes.

A significant negative association has been observed between the wealth index of the respondents and infant mortality. This study's outcome is comparable to that of the earlier investigation, confirming that women who belong to the richest wealth quintile had a lower likelihood of experiencing infant deaths in the group of underperforming states in India [1]. Perhaps it is because these women have the resources to provide their children with nutritious food and the time to spend with them [6].

There is a significant association between the ages of mothers at their first birth and infant mortality. This finding is consistent with the study undertaken in low- and middle-income nations, including Ethiopia, Kenya, and Nepal [6].On the contrary, mothers' age at birth had no influence in well-performing states, but the chance of infant deaths was low in the (20–29) years age group in states that perform poorly in India [1]. Perhaps younger mothers have an increased likelihood of child death due to undeveloped reproductive systems and less stability in handling all the difficulties of giving birth [6]. Similarly, preterm birth rates and low birth rates are observed to be significantly higher in pregnancies with advanced maternal age [23].

From Tables 2 and 3, it is also revealed that infant mortality among illiterate women is significantly higher than that among literate women, which is also not against our expectations. It confirms the findings from the previous study, which found that compared to mothers with different levels of education, mothers with no formal education reported the greatest rate of infant mortality [24]. A mother's education level has a significant impact on infant mortality because it raises her awareness of how to care for her children both before and after birth and gives her the ability to change feeding and childcare practices by reshaping and changing traditional household relationships. Mother's education helps to improve more effective preventative and health care practices, which increases her productivity and influences infant mortality [13].

A significant positive association is seen between infant mortality and the birth order of the children. This is in line with the research, which revealed that compared to infants born to mothers with birth orders one, there was a higher mortality risk for infants born to mothers with birth orders two and above [6]. A possible explanation might be that as birth order gets higher, intra-familiar competition for foods and other precious resources that children need will increase. Additionally, children are more likely to experience most of its impacts. Also, as birth order increases, the quality of child care reduces since the mother will have to care for more children [11]. This may also be attributable to higher confidence from prior pregnancies and delivery experiences with restrictions of time and resources among women who had multiple birth orders [25].

4 Conclusion

Although the infant mortality rate in Assam has declined over time, the situation is still poor in comparison to some other parts of the country. Women in rural areas of Assam are unaware about numerous health and welfare initiatives related to maternity and child health. This study aimed to identify the risk factors for the count of infant deaths in families in Assam using a generalised linear regression model for the count data. This study involved 34,979 mothers, and the findings indicate that 89.72% of the mothers did not experience infant deaths. The average number of infant deaths is 0.39, with a variance of .689, indicating overdispersion in the data under study. To test whether overdispersion is due to the presence of excess zeros, a likelihood ratio chi-square test is performed. The result is significant (312.32 with a P-value of 0.000), and there is evidence of the presence of excess zeros. So, zero-inflated models were selected for the study. Among the models, the zero-inflated negative binomial model had the least AIC (34781.56) and BIC (35217.86). Moreover, the zero-inflated negative binomial model (ZINB) has the greatest AUC (0.7196). So, finally, zero-inflated negative binomial model is found to be the most parsimonious way to assess the risk factors associated with infant mortality in Assam. According to the findings, place of residence, religion, wealth index, age of mother at her first birth, level of education of mothers, and birth order were found to be significant risk factors for infant mortality.

The findings presented in this study have the following research and policy implications:

- Improving the education level of mothers is vital. Infant and child mortality will be reduced if education and the empowerment of women are enhanced. The government of India must devise additional programmes to improve education in society, with a particular emphasis on women's education.
- Family welfare services also need to be peaked up as the mother's age at childbirth and birth order influences infant mortality rate. Women are therefore encouraged to use need-based contraception, to maintain sufficient spacing, to prevent higher deliveries, and to avoid getting pregnant before the age of twenty.
- It is important to develop a variety of income-generating activities, particularly in rural and slum areas, so that residents can enhance their quality of life.
- Lastly, although India has a vast PHC (Primary Health Care) infrastructure, its performance requires close monitoring for accountability and to ensure that all employees do their job. The government needs to promote the necessity of antenatal care among pregnant mothers to improve maternal and infant health.

The National Family Health Survey (NFHS)-5 survey data (2019–2021), which were just released, make this work relevant. There is no evidence to model the zero-inflated, over-dispersed count data of infant mortality using recently released NFHS data in Assam. This research will definitely provide a direction for further study to analyse over-dispersed, under-dispersed and zero-inflated count data for multidisciplinary research.

Although the National Family Health Survey provides a comprehensive dataset that is nationally representative, this research may have some limitations. This study is vulnerable to various types of error because it is based on self-reported data from a nationally representative survey. Participants were asked to recall events from the 5 years prior to the survey, and they may have forgotten certain particulars. But, due to the use of identical instruments across countries, there will be a chance to compare the results nationally and globally.

References

1) Bhatia M, Dwivedi LK, Ranjan M, Dixit P, Putcha V. Trends, patterns and predictive factors of infant and child mortality in well-performing and underperforming states of India: a secondary analysis using National Family Health Surveys. *BMJ Open*. 2019;9(3):e023875. Available from: https://doi.org/10.1136/bmjopen-2018-023875.
2) International Institute for Population Sciences (IIPS) and ICF. 2021. National Family Health Survey (NFHS-5), 2019-21: India. Mumbai: IIPS. ;p. 2019–2040. Available from: https://ln.run/iy5Iv.
3) Chiero V, Das KC. Estimation of infant mortality and under-five mortality by place of residence and religion and its predictive factors in Northeast India. *International Journal of Advanced Research*;8(10):830–843. Available from: http://dx.doi.org/10.21474/IJAR01/11907.
4) Gogoi K, Barman MP, Hazarika J. Indirect estimation of infant mortality rate in Assam:a district level analysis. *Advances in Mathematics: Scientific Journal*. 2021;10(3):1675–1689. Available from: https://doi.org/10.37418/amsj.10.3.50.
5) Woldeamanuel BT, Aga MA. Count Models Analysis of Factors Associated with Under-Five Mortality in Ethiopia. *Global Pediatric Health*. 2021;8. Available from: https://doi.org/10.1177/2333794X21989538.
6) Mulugeta SS, Muluneh MW, Belay AT, Moyehodie YA, Agegn SB, Masresha BM, et al. Multilevel log linear model to estimate the risk factors associated with infant mortality in Ethiopia: further analysis of 2016 EDHS. *BMC Pregnancy and Childbirth*. 2022;22(1):1–1. Available from: https://doi.org/10.1186/s12884-022-04868-9.
7) Darnah, Utoyo MI, Chamidah N. Modeling of Maternal Mortality and Infant Mortality Cases in East Kalimantan using Poisson Regression Approach Based on Local Linear Estimator. *IOP Conference Series: Earth and Environmental Science*. 2019;243(1):012023. Available from: https;//doi.org/10.1088/1755-1315/243/1/012023.

8) Kumar M, Shivgotra VK. Effect of seasonal variations on infant mortality rate of some selected districts of Jammu Division. *Indian Journal of Community Medicine*. 2020;45(4):415. Available from: https://doi.org/10.4103/ijcm.IJCM_441_19.
9) Prahutama A, Suparti, Munawaroh DA, Utami TW. Modeling bivariate Poisson regression for maternal and infant mortality in Central Java. *AIP Conference Proceedings* . 2021;2329. Available from: https://doi.org/10.1063/5.0042142.
10) Saputro MI, Qudratullah MF. Estimation of Zero-Inflated Negative Binomial Regression Parameters Using the Maximum Likelihood Method (Case Study: Factors Affecting Infant Mortality in Wonogiri in 2015). *Proceeding International Conference on Science and Engineering*. 2021;4:240–254. Available from: https://ejournal.uin-suka.ac.id/saintek/icse/article/download/2915/2273/7924.
11) Aragaw AM, Azene AG, Workie MS. Poisson logit hurdle model with associated factors of perinatal mortality in Ethiopia. *Journal of Big Data*. 2022;9(1):1–1. Available from: https://doi.org/10.1186/s40537-022-00567-6.
12) Siswanto S, R ES, Sunusi N, Ilyas N. Comparison of Negative Binomial Regression Model and Geographically Weighted Poisson Regression on Infant Mortality Rate in South Sulawesi Province. *Indonesian Journal of Statistics and Its Applications*. 2022;6(2):170–179. Available from: https://stat.ipb.ac.id/journals/index.php/ijsa/article/download/800/376.
13) Alito WA, Girmma AG. Count regression modelling of under-five child mortality in SNNPRS, Ethiopia. 2021. Available from: https://www.allsubjectjournal.com/assets/archives/2021/vol8issue4/7-2-55-101.pdf.
14) Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean Journal of Anesthesiology*. 2022;75(1):25–36. Available from: https://doi.org/10.4097/kja.21209.
15) Odetunmibi OA, Adejumo AO, Anake TA. A study of Hepatitis B virus infection using chi-square statistic. *Journal of Physics: Conference Series*. 2021;1734(1):012010. Available from: https://doi.org/10.1088/1742-6596/1734/1/012010.
16) Bango M, Ghosh S. Reducing infant and child mortality: assessing the social inclusiveness of child health care policies and programmes in three states of India. *BMC Public Health*. 2023;23(1):1–8. Available from: https://doi.org/10.1186/s12889-023-15812-7.
17) Kumari M, Mohanty SK. Caste, religion and regional differentials in life expectancy at birth in India: cross-sectional estimates from recent National Family Health Survey. *BMJ open*. 2020. Available from: https://doi.org/10.1136/bmjopen-2019-035392.
18) Dandapat B, Biswas S, Patra B. Religion, nutrition and birth weight among currently married women (15–49) in India: A study based on NFHS-5. *Clinical Epidemiology and Global Health*. 2023;20:101218. Available from: https://doi.org/10.1016/j.cegh.2023.101218.
19) Mondal NA, Ali B, Sk MI. Has Muslim got benefited from the national health mission? A situational analysis of maternal health services in India. *Ethiopian Journal of Health Sciences*. 2020;30(5). Available from: http://dx.doi.org/10.4314/ejhs.v30i5.19.
20) Subudhi SC. Child worker and child marriage among tea working class communities in Assam: Scheme for engagement. International journal of research culture society. *International journal of research culture society*. 2022;30(4):6–6. Available from: https://doi.org/10.2017/IJRCS/202204014.
21) Gebreegziabher E, Bountogo M, Sié A, Zakane A, Compaoré G, Ouedraogo T, et al. Influence of maternal age on birth and infant outcomes at 6 months: a cohort study with quantitative bias analysis. *International Journal of Epidemiology*. 2023;52(2):414–425. Available from: https://doi.org/10.1093/ije/dyac236.
22) Marphatia AA, Saville NM, Amable GS, Manandhar DS, Cortina-Borja M, Wells JC, et al. How Much Education Is Needed to Delay Women's Age at Marriage and First Pregnancy? *Frontiers in Public Health*. 2009;7:396. Available from: https://doi.org/10.3389/fpubh.2019.00396.
23) Ratiu D, Sauter F, Gilman E, Ludwig S, Ratiu J, Mallmann-Gottschalk N, et al. Impact of Advanced Maternal Age on Maternal and Neonatal Outcomes. *In Vivo*. 2023;37(4):1694–1702. Available from: https://doi.org/10.21873/invivo.13256.
24) Mfateneza E, Rutayisire PC, Biracyaza E, Musafiri S, Mpabuka WG. Application of machine learning methods for predicting infant mortality in Rwanda: analysis of Rwanda demographic health survey 2014–15 dataset. *BMC Pregnancy and Childbirth*. 2022;22(1):388. Available from: https://doi.org/10.1186/s12884-022-04699-8.
25) Chilot D, Belay DG, Ferede TA, Shitu K, Asratie MH, Ambachew S, et al. Pooled prevalence and determinants of antenatal care visits in countries with high maternal mortality: A multi-country analysis. *Frontiers in Public Health*. 2023;11. Available from: https://doi.org/10.3389/fpubh.2023.1035759.