International Journal of Community Medicine and Public Health
Kumar D et al. Int J Community Med Public Health. 2026 Jan;13(1):282-287
http://www.ijcmph.com
pISSN 2394-6032 | eISSN 2394-6040

DOI: https://dx.doi.org/10.18203/2394-6040.ijcmph20254438

Original Research Article

Supporting public health efforts in India and Nepal with probabilistic child death modelling

Dileep Kumar[1], Navin Upadhyay[2]*, Himanshu Pandey[2]

[1]Department of Obstetrics and Gynaecology, BRD Medical College Gorakhpur, Uttar Pradesh, India
[2]Department of Mathematics and Statistics, Deen Dayal Upadhyaya Gorakhpur University, Gorakhpur, Uttar Pradesh, India

Received: 06 October 2025
Revised: 28 November 2025
Accepted: 01 December 2025

***Correspondence:**
Navin Upadhyay,
E-mail: navinupadhyayrs@gmail.com

ABSTRACT

Background: Child mortality remains a major public health concern in South Asia, shaping population dynamics and affecting family well-being. Understanding household-level mortality patterns is essential for identifying high-risk groups and developing effective interventions. This study analyzes child mortality data from households in Eastern Uttar Pradesh, India, and Nepal, where deaths are rare but occasionally clustered within families.

Methods: Four probabilistic models were applied to the observed number of child deaths per household: the Geometric distribution, the Inflated Geometric distribution to accommodate excess zeros, and the Himanshu compounded distribution. Model parameters were estimated using Maximum Likelihood Estimation (MLE). Model adequacy was evaluated through Chi-square goodness-of-fit tests comparing expected and observed household mortality counts.

Results: All models showed strong consistency with the empirical data. Chi-square test results produced high p-values (>0.95), indicating that each model successfully captured the zero-heavy structure and the infrequent higher mortality events present in the datasets from both regions.

Conclusions: The findings demonstrate that zero-inflated and compounded probabilistic models provide reliable representations of household-level child mortality in South Asia. These modeling approaches can support better identification of vulnerable households and improve the predictive accuracy of mortality assessments, contributing to more targeted public health strategies aimed at reducing child deaths.

Keywords: Compounded distribution, Geometric distribution, Himanshu distribution, Inflated geometric distribution, Child mortality, Probabilistic modeling, Public health interventions, Statistical modeling

INTRODUCTION

One of the most important measures of the socioeconomic and health status of a population is child mortality. Reducing child fatalities has been a major goal of public health policies and development initiatives worldwide, especially in low- and middle-income nations where the burden is still disproportionately high. Planning health interventions, demographic modeling, population forecasting, and evaluating the efficacy of current policies all depend on an understanding of the trends and causes of child mortality. Cohort survival analysis, mortality rates, and life tables are examples of basic demographic metrics that have historically been used to study child mortality. Even though these methods offer helpful macro-level insights, they frequently fall short of capturing the probability structure of occurrences at the individual level, especially when fatalities are uncommon or occur within families.

A strong foundation for tackling these issues is provided by probabilistic modeling. It is feasible to create statistical models that calculate the likelihood of various mortality outcomes, measure uncertainty, and forecast future trends by considering the number of child deaths in a household as a random variable. The ability of probabilistic models to account for variations in family size, birth order, socioeconomic status, and healthcare access makes them particularly useful in demographic research. Additionally, they enable researchers to account for uncommon or severe occurrences, which are frequently found in child mortality data, as well as compounding effects.

The Geometric distribution is one of the most popular classical probabilistic models for simulating the number of attempts until the first success or failure. A kid surviving is considered a "success" in the context of child mortality, whereas a child dying is considered a "failure." When it is assumed that the likelihood of infant mortality is constant for every birth and irrespective of the number of children, the geometric distribution is especially appropriate. It is a useful tool for public health researchers and demographers because to its interpretability and simplicity. Furthermore, the Geometric distribution offers a foundation for expanding to more intricate models that may account for observed anomalies in actual data, including clustered mortalities or excess zero deaths. However, the straightforward geometric distribution frequently fails to adequately describe the features of real-world child mortality data. The overabundance of zero deaths is one notable characteristic; households without child fatalities are generally overrepresented in comparison to what a conventional geometric model would predict. In order to solve this, the Inflated Geometric distribution adds an inflation parameter, usually represented by α, that indicates the likelihood of a structural zero, or the likelihood that a family will not have any child fatalities as a result of favorable socioeconomic circumstances, access to healthcare, or good maternal health. While offering flexibility to simulate the overdispersion of zeros frequently seen in demographic data, the Inflated Geometric model maintains the simplicity of the Geometric distribution.

More extended models have been created to take into consideration compounding and probabilistic interactions in addition to the Geometric and Inflated Geometric distributions. The Himanshu distribution is one example of such a model.[1] By include the influence of N compounding trials, this distribution extends the Geometric distribution and makes it possible to simulate families with numerous child fatalities while still capturing the likelihood of zero deaths. When environmental circumstances, maternal health, and genetic predisposition all have an impact on child mortality and might lead to a clustering of fatalities within certain families, the Himanshu distribution is very helpful. The model may be modified to account for different levels of mortality risk and reliance patterns across households by changing the value of N.

There are several benefits to using these distributions to data on child mortality. First of all, they offer a probabilistic framework for calculating important parameters that might guide health policy, such the chance of zero deaths and the probability of child death. Second, researchers may determine expected counts, compare them with actual frequencies, and use goodness-of-fit tests (such Chi-square tests) to evaluate each model's adequacy by fitting these models to observed data. Thirdly, the probabilistic method makes it possible to predict future patterns in child mortality, which is essential for effectively allocating resources and organizing treatments.

A strong, adaptable, and understandable framework for examining child mortality statistics is offered by the use of probabilistic distributions, particularly the Geometric, Inflated Geometric, and Himanshu distributions.[2] Essential characteristics that are not captured by straightforward statistical measurements, such as excess zero deaths, mortality clustering, and compounding effects, are captured by these models. Researchers can guarantee precise fitting, produce trustworthy predictions, and promote evidence-based public health actions by estimating parameters using techniques like the Maximum Likelihood Estimation (MLE) and the Method of Moments (MoM). In general, applying probabilistic modeling to demographic studies of child mortality improves our knowledge of population health dynamics, directs the distribution of resources, and informs policies meant to lower avoidable child fatalities. These factors ultimately help to raise child survival rates in South Asia and other low- and middle-income areas.

In South Asia, child mortality is still a major public health concern that calls for sophisticated statistical models in order to comprehend and forecast death trends. The intricacies present in child mortality data, such as excess zeros and overdispersion, are frequently ignored by traditional demographic approaches. To overcome these problems, recent research has used probabilistic models, which provide more accurate depictions of mortality trends.

The number of tries until the first success or failure has been modeled using the geometric distribution. It can be used to indicate how many children a family has before someone dies in the context of child mortality. In order to examine newborn fatalities by age, an inflated geometric model, emphasizing how well it captures mortality trends in South Asian communities.[3]

The inflated geometric distribution adds an inflation parameter to compensate for the high frequency of zero deaths in some populations. This model, which offers a more accurate fit for data with extra zeros, has been used to investigate the number of child deaths given stable parity. The usefulness of this approach in reflecting the distribution of child fatalities in households was shown.

Child mortality trends have been modeled using the Himanshu distribution, which is an extension of the Geometric distribution. In order to investigate child death trends, a probability model based on zero truncation of the Himanshu distribution, providing a more sophisticated comprehension of mortality distributions.[4]

Overdispersed count data, such child mortality, has been modeled using the Negative Binomial distribution. examined infant mortality data in Padang, Indonesia, using Negative Binomial regression, proving that it was a useful tool for managing mortality data overdispersion.[5]

To model count data with extra zeros, zero-inflated models have been used, such as the Zero-Inflated Negative Binomial Regression (ZINBR). By using ZINBR to forecast under-five mortality in Nepal, Bhusal et al. To pinpoint important risk factors for child mortality and offer suggestions for practical solutions.[6]

In infant mortality research, multilevel models have been used to take hierarchical data structures into consideration. Highlighted the value of maternal health services in lowering infant mortality by using multilevel Negative Binomial analysis to find characteristics related with prenatal care contacts among expectant women in low- and middle-income countries.[7]

Geographic and temporal differences in child mortality have been studied using spatial and temporal modeling tools. In order to shed light on regional differences and guide focused actions. This research performed space-time modeling of child mortality at the administrative level in Nigeria.[8]

Data on child mortality has been subjected to uncertainty modeling using Bayesian techniques. In order to account for overdispersion and excess zeros in the data, the article used Bayesian zero-inflated regression models to evaluate under-five child mortality in Ethiopia.[9]

The prediction of death among children under five and the identification of important parameters linked to child mortality have both been investigated using machine learning algorithms. The accuracy of machine learning models in forecasting mortality among children under five was evaluated in article, indicating their potential to improve prediction skills.[10]

In Eastern Uttar Pradesh, India, and Nepal, where the majority of families have no child deaths while a small number have multiple fatalities, the study seeks to understand the distribution of child deaths across households. It aims to assess how well a number of probabilistic models geometric, inflated geometric, and Himanshu compounded distributions capture these patterns. The study seeks to identify the model that best captures the observed mortality distribution using maximum likelihood estimation and Chi-square goodness-of-fit testing. The ultimate goal is to improve child mortality prediction and create more trustworthy instruments for identifying high-risk households.

METHODS

Geometric distribution

$$P(X = x) = p(1 - p)^x; 0<p<1 \quad (1)$$

Inflated Geometric Distribution

$$P(X = 0) = 1 - \alpha \quad (2)$$

$$P(X = x) = \alpha p(1 - p)^{x-1}; 0<p<1, \quad (3)$$

Himanshu Distribution

$$P(X = x) = p^N(1 - p^N)^x ; 0<p<1, N=\{1,2,3\ldots\} \quad (4)$$

Inflated Himanshu Distribution

$$P(X = 0) = 1 - \alpha \quad (5)$$

$$P(X = x) = \alpha p^N(1 - p^N)^{x-1} ; 0<p <1 , N =\{1,2,3\ldots\} \quad 6)$$

Maximum Likelihood Estimation (MLE)

A statistical technique called Maximum Likelihood Estimation (MLE) is used to estimate model parameters by selecting values that maximize the probability of observing the provided sample data. To put it simply, MLE determines the parameter values that make the observed data most likely. The present research is a secondary data analysis (retrospective analytical study) using published demographic datasets from India (1995) and Nepal (2000).

Two previously published demographic datasets from rural Nepal (Palpa and Rupandehi districts, 2000) and North Rural India (1995) were used in this study's secondary data analysis design. The current analysis was carried out between January and March 2024, whereas the original data, which showed the number of child deaths per family, was gathered between January and December 1995 in India and January and December 2000 in Nepal. While the current analysis was carried out between September and October 2024, all families included in the original surveys were taken into account for analysis; no additional inclusion or exclusion criteria were applied beyond the original studies; families with missing or incomplete child-death information were excluded in accordance with the primary dataset protocols. The datasets came from studies "Effect of Breastfeeding on Fertility in North Rural India (1995)" and "A Demographic Survey on Fertility and Mortality in Rural Nepal (2000)." The data were cleaned, coded, and arranged into grouped frequency distributions according to the number of child deaths per family. Observed frequencies were compared with expected frequencies generated using probability models such as the Poisson and Negative Binomial distributions. Model fit was evaluated using chi-square goodness-of-fit statistics, degrees of freedom, and associated p values. As the study utilized publicly available and fully de-identified secondary datasets, no ethical approval was required; however, institutional exemption for secondary data use was obtained where applicable. All statistical analyses performed using standard Python software, and a significance level of $p<0.05$ adopted for inferential tests.

Application

An essential part of comprehending demographic and health trends within communities is the examination of child mortality. In this study, we looked at the distribution of child fatalities in households in two different areas: Nepal and Eastern Uttar Pradesh, India. These kinds of statistics are very important for determining the burden of child mortality, creating public health initiatives, and pinpointing areas that need specific health regulations. In example, the high frequency of zero fatalities and the uncommon occurrence of greater child mortality within families are two trends found in such data that are frequently missed by traditional statistical methodologies. This study used four statistical distributions to address these issues: the Inflated Geometric distribution, which adds an extra inflation parameter (α) to account for an excess of zero deaths; the Himanshu distribution, which generalizes the geometric distribution to account for compounded probabilities; and the Geometric distribution, which models the probability of the first "failure" (or child death) in a series of independent trials. Maximum Likelihood Estimation (MLE) was used to estimate the parameters for each distribution, guaranteeing that the model appropriately captured the variance and mean of the observed data. The purpose of the research was to provide light on each distribution's fit as well as how these models may influence public health policy, predict future demographic changes, and pinpoint high-risk groups. Using Chi-square goodness-of-fit tests and comparing observed frequencies with projected counts from each distribution, the study assesses each statistical model's adequacy and its use in health policy planning and demographic forecasting.

RESULTS

Four different probability distributions geometric, inflated geometric, Himanshu, and inflated Himanshu were fitted to observed data on child fatalities per household in this study using the Maximum Likelihood Estimation (MLE) approach. Finding the best model to depict these intricate patterns in two different geographical locations Nepal (Table 2) and Eastern Uttar Pradesh, India (Table 1) was the goal. In order to provide a thorough assessment from a public health standpoint, the model's adequacy was evaluated using a combination of Chi-square (χ^2) goodness-of-fit tests, related p values, and information criteria such as the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC).

Table 1: Observed and expected frequencies of child deaths per family of eastern Uttar Pradesh, India with model fit statistics.

No. of child death	Observed no. of families	Expected no. families			
		Geometric distribution	Inflated geometric distribution	Himanshu distribution	Inflated Himanshu distribution
0	506	506.07	508.41	505.63	507.92
1	178	188.70	185.32	190.14	184.90
2	76	70.41	73.18	71.55	74
3	32	26.25	27.01	26.88	27.40
4	8	9.78	8.90	9.90	8.85
5	6	3.64	3.33	3.73	3.29
6+	4	1.15	0.85	1.17	0.84
Total	810	810	810	810	810
Mean	0.63	0.59	0.58	0.59	0.58
SD	1.04	0.94	0.93	0.95	0.93
$\hat{p}$		0.706	0.719	0.745	0.758
$\hat{\alpha}$			0.081		0.073
χ^2		2.16	1.67	1.92	1.38
df		4	3	4	3
P value		0.83	0.76	0.78	0.85
AIC		1057.4	1055.2	1056.9	1054.1
BIC		1061.3	1059.1	1060.8	1058

There is a noteworthy pattern of high frequency at zero child fatalities (506 homes) followed by a declining trend for the Eastern Uttar Pradesh, India dataset (Table 1), which includes 810 families and records child deaths up to 6+. According to the fitting findings, all four models fit the observed data satisfactorily, as shown by the p values, which are much higher than the traditional significance level of 0.05. The p values for the Geometric distribution, Inflated Geometric distribution, Himanshu distribution, and Inflated Himanshu distribution were as follows: 0.83 (χ^2=2.16, df=4), 0.76 (χ^2=1.67, df=3), 0.78 (χ^2=1.92, df=4), and 0.85 (χ^2=1.38, df=3), respectively. The best model for this area among them is the inflated Himanshu distribution. The greatest statistical agreement with the measured frequencies is shown by its p value of 0.85, which suggests little disagreement. Additionally, this model had the lowest AIC (1054.1) and BIC (1058) values, indicating that it provides the most effective description of the underlying data creation process for child fatalities in Eastern Uttar Pradesh by achieving the greatest fit with optimum parsimony. The significance of taking into consideration the inflated number of families with zero infant fatalities was demonstrated by the Inflated Geometric distribution, which came in second.

Table 2: Observed and expected frequencies of child deaths of Nepal with model fit statistics.

No. of child death	Observed no. of families	Expected no. families			
		Geometric distribution	Inflated geometric distribution	Himanshu distribution	Inflated Himanshu distribution
0	669	662.8	666.9	660.4	665.3
1	137	138.2	138.7	139.4	139
2	32	30.9	31.1	32.7	32.1
3	6	6.5	6.4	7.1	6.7
4+	7	12.6	7.9	11.4	7.9
Total	851	851	851	851	851
Mean	0.29	0.32	0.30	0.32	0.30
SD	0.65	0.72	0.66	0.71	0.67
	$\hat{p}$	0.829	0.841	0.872	0.876
	$\hat{\alpha}$		0.091		0.084
	χ^2	3.96	1.88	3.21	1.44
	df	3	2	3	2
	P value	0.26	0.61	0.36	0.70
	AIC	1023.4	1019.2	1024.8	1018.5
	BIC	1026.8	1022.7	1028.3	1021.9

A similar distribution profile is seen when focusing on the data from Nepal (Table 2), which was gathered from 851 households with child fatalities ranging from 0 to 4+. This distribution is marked by a significant number of families (669 families) that had no child deaths and a subsequent fall. All four models had P-values over 0.05, demonstrating their statistical validity in explaining the observed frequencies and being in line with the results from India. A p value of 0.26 (χ^2= 3.96, df=3) was found for the Geometric distribution, 0.61 (χ^2=1.88, df=2) for the Inflated Geometric distribution, 0.36 (χ^2=3.21, df=3) for the Himanshu distribution, and 0.70 (χ^2 = 1.44, df=2) for the Inflated Himanshu distribution. With the highest p value of 0.70, the Inflated Himanshu Distribution once more showed the greatest match in this case.

In addition, this model has the best AIC (1018.5) and BIC (1021.9) values, making it the most suitable and economical model for this dataset's description of Nepal's child mortality trends. Following this, the Inflated Geometric distribution displayed competitive AIC/BIC values and a very strong match (p value = 0.61). The Inflated Himanshu Distribution's consistently excellent performance in Eastern Uttar Pradesh, India, and Nepal is extremely important from the perspective of public health. Because it specifically takes into consideration the phenomena of "excess zeros" families reporting no child deaths which frequently occurs in health-related count data, the "inflated" feature of this model is essential. This zero count is frequently underestimated by standard distributions such as the straightforward Geometric or Himanshu, which results in estimates of population health that are less precise. High p values and low AIC/BIC suggest that the Inflated Himanshu Distribution fits the data well, capturing both the prevalence of households unaffected by child mortality and the decreasing likelihood of rising child fatalities.

Because it helps policymakers and health professionals better understand the true burden of child mortality, target interventions more precisely, and allocate resources for maternal and child health programs in these regions, this increased modeling accuracy is crucial for public health efforts. The model's potential generalizability and usefulness in comparable demographic circumstances are shown by the consistent results obtained across two different populations.

DISCUSSION

All four probabilistic models the Geometric, Inflated Geometric, Himanshu, and Inflated Himanshu distributions offer a strong fit to the observed data, according to an analysis of child mortality counts per family in Eastern Uttar Pradesh and Nepal. The household-level mortality pattern is significantly zero-inflated in both regions, with a small percentage of families reporting multiple losses and the vast majority reporting no child deaths. The parameter estimates, where the fitted values are comparatively high and show a strong tendency toward zero counts, reflect this distributional feature. The presence of excess zeros beyond what is predicted by standard geometric structures is further confirmed by the zero-inflation parameters $\hat{\alpha}$ for the inflated models.

The predicted frequencies produced by all models for Eastern Uttar Pradesh (Table 1) closely correspond to the observed counts for all death categories, including the higher-order deaths (4-6+). None of the models exhibit signs of lack of fit, as evidenced by the low Chi-square statistics and high p-values (0.76-0.85). The Inflated Himanshu distribution achieves the lowest AIC (1054.1) and BIC (1058) among them, indicating that it offers the most effective depiction of the mortality pattern while taking into consideration excess zeros and sporadic higher death counts.

In a similar vein, all models accurately replicate the observed frequency structure for Nepal (Table 2). No model substantially deviates from the empirical distribution, and the p-values range from 0.26 to 0.70, indicating acceptable to strong fits. Once more, the Inflated Himanshu distribution produces the lowest AIC and BIC values, demonstrating its ability to capture both zero inflation and the existence of households with multiple child deaths. The inflated models' consistent performance on both datasets emphasizes how crucial it is to account for additional zeros when simulating uncommon family-level mortality events. the comparison indicates that the inflated and compounded models especially the Inflated Himanshu distribution offer superior fit and better capture the heterogeneity in child mortality across households, while the basic Geometric distribution offers a reasonable approximation. These results support the value of adaptable probabilistic models in the analysis of demographic data with skewness, zero inflation, and infrequent but significant high-mortality outcomes.

CONCLUSION

The study shows that family-level child mortality trends in Eastern Uttar Pradesh and Nepal are well captured by probabilistic models, such as the Geometric, Himanshu, and Inflated Geometric distributions. These models show that while unusual multiple fatalities are adequately represented by compounded distributions, most families experience zero or one child death. The dependability of the Maximum Likelihood Estimation techniques is confirmed by consistent parameter estimations. The results have important ramifications for public health. Policymakers may identify high-risk families, distribute resources effectively, and create focused interventions like immunization campaigns, nutritional programs, and maternal health efforts by precisely predicting mortality trends. These models may be used to forecast future trends in child mortality, helping to foresee both possible multi-death situations in susceptible groups and recurring zero-death patterns. All things considered, this modeling approach improves evidence-based decision-making, aids in strategic planning for health initiatives, and provides a flexible tool that may be used in other areas with comparable demographics.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Agarwal AB, Pandey HI. Himanshu distribution and its applications. Bullet Mathemat Stat Res. 2022;10(4):24-35.
2. Agarwal A, Pandey H. An inflated model based on sized biased Himanshu distribution and its applications. J Probab Stat Sci. 2023;21(1).
3. Singh BP, Dixit S, Maurya DK. Some probability models for age at infant death and estimation procedure. Demography India. 2022;51(2):25-44.
4. Pandey H, Kishun J. A Probability Model for the Child Mortality in a Family. Revista Colombiana de Estadística. 2010;33(1):1-1.
5. Fitri F, Sari FM, Gamayanti NF, Utami IT. Infant mortality case: an application of negative binomial regression in order to overcome overdispersion in poisson regression. EKSAKTA: Berkala Ilmiah Bidang MIPA. 2021;22(3):200-10.
6. Bhusal MK, Khanal SP. Statistical models for predicting the number of under-five mortality in Nepal. PLoS One. 2025;20(5):e0324321.
7. Alem AZ, Tegegne BA, Aragaw FM, Teklu RE, Baykeda TA. Multilevel negative binomial analysis of factors associated with numbers of antenatal care contacts in low and middle income countries: Findings from 59 nationally representative datasets. Plos one. 2024;19(4):e0301542.
8. Godwin J, Wakefield J. Space-time modeling of child mortality at the Admin-2 level in a low and middle income countries context. Stat Medi. 2021;40(7):1593-638.
9. Workie MS, Azene AG. Bayesian zero-inflated regression model with application to under-five child mortality. J Big Data. 2021;8(1):4.
10. Saroj RK, Yadav PK, Singh R, Chilyabanyama ON. Machine learning algorithms for understanding the determinants of under-five mortality. BioData mining. 2022;15(1):20.

Cite this article as: Kumar D, Upadhyay N, Pandey H. Supporting public health efforts in India and Nepal with probabilistic child death modelling. Int J Community Med Public Health 2026;13:282-7.