

Multilevel Survival Modelling of Neonatal Mortality under Some Prognostic Factors in Uttar Pradesh

Shambhavi **Mishra**, Shalini **Jaiswal***

Department of Statistics, University of Lucknow, Uttar Pradesh, India.

ABSTRACT

Introduction: Neonatal mortality rate is a critical indicator of a nation's healthcare system and socio-economic development. While standard survival models assume independence among observations this assumption is inadmissible for hierarchical data structures, limiting their applicability. To address this limitation, this study employed parametric multilevel mixed-effects survival models that explicitly account for clustering and unobserved heterogeneity.

Methods: This study employed unit level data from National Family Health Survey (NFHS-5), 2019-2021 for Uttar Pradesh, which employs a stratified multistage sampling design. Multilevel mixed-effects survival models were fitted (Weibull and exponential distributions) with random intercepts at the Primary Sampling Unit (PSU) and district levels to account for hierarchical clustering. Model fit was assessed using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), while Likelihood Ratio Tests (LRT) compared nested specifications. The Intra-Class Correlation (ICC) was computed to quantify unobserved heterogeneity attributable to clustering at the district and PSU levels.

Results: The Multilevel Weibull Mixed-effects model provided the best fit, revealing that unobserved heterogeneity was predominantly clustered at the PSU level (variance: 1.238) rather than the district level (0.023). Advanced maternal age (45–49 years) was a critical risk factor (HR = 2.943, 95% CI: 1.729–4.713), as was the use of Smokey cooking fuels (HR = 1.304, 95% CI: 1.014–1.677). Neonates not weighed at birth faced a significantly higher mortality risk (HR = 1.612, 95% CI: 1.287–2.019) compared to those with normal birth weight. Maternal education, religion, and place of delivery were also identified as significant determinants of neonatal outcomes.

Conclusion: The study recommends that efforts to reduce neonatal mortality should address not only individual-level risk factors but also community-level disparities in healthcare access and quality.

Key words: Multilevel weibull mixed effects model; Random effect; Akaike information criteria (AIC); Neonatal mortality

***Corresponding Author:**
shalinij343@gmail.com



INTRODUCTION

The first 28 days of life represent the most vulnerable period for child survival, making the reduction of neonatal mortality a paramount global health priority.^{1,2} While under-five mortality rates have declined, neonatal survival remains a persistent public health challenge, accounting for approximately 2.3 million deaths worldwide in 2022.³ India accounts for the largest share of global neonatal deaths.¹⁻³ Improvements have been inconsistent nationwide despite national reports from the Sample Registration System (SRS) and the National Family Health Survey (NFHS) indicating a gradual decline in mortality rates.^{3,4} Uttar Pradesh, India's most populous state, faces a disproportionate share of this burden. With a neonatal mortality rate of 36 per 1,000 live births (NFHS-5), the state significantly exceeds the national average and ranks among one of the highest among high-burden states.^{5,6} This high burden of neonatal mortality in Uttar Pradesh reflects the complex, multifactorial nature of newborn survival. Studies have demonstrated that neonatal mortality is determined by a constellation of bio-demographic factors such as birth interval, birth order, maternal age, biological complications such as low birth weight and prematurity, and socioeconomic determinants such as maternal education and household wealth index at individual level. These determinants operate across multiple levels of causation, from maternal physiological factors to household-level economic constraints that influence healthcare access.⁷⁻⁹

Standard survival analysis techniques, the semiparametric Cox proportional hazards model, parametric models (such as Weibull, Gompertz and log-logistics and Exponential), provide efficient individual-level mortality risk estimates under the assumption of independent observations.⁸⁻¹⁰ However, these approaches fail to account for the hierarchical structure inherent in demographic surveys, where individuals are clustered within communities or districts, leading to correlated survival outcomes. Ignoring such clustering can produce biased estimates and compromised statistical inference¹¹⁻¹⁴

To address these methodological limitations and to identify the best fitted parametric survival model for neonatal mortality, this study applied multilevel mixed effect parametric survival models to the NFHS-5 dataset for Uttar Pradesh. Unlike standard models, this approach explicitly accounts for unobserved heterogeneity by estimating community-level random effects alongside individual-level bio-demographic, biological, and socioeconomic determinant with the parametric baseline hazard functions.

METHODS

Data

This research utilized unit-level data from the fifth round of NFHS, conducted in India between 2019 and 2021. The NFHS is a large-scale, multi-round survey conducted by the International Institute for Population Sciences (IIPS) under the stewardship of the Ministry of Health and Family Welfare (MoHFW), Government of India.⁵ The survey provides comprehensive estimates on key indicators including fertility, child mortality, maternal and child health, and nutrition at the national and district

levels. For this analysis, data specifically for the state of Uttar Pradesh was extracted from the NFHS-5 'child recode' file, obtained from the DHS Program repository.¹⁵

Sampling Design

The NFHS-5, conducted between 2019 and 2021, employed a cross-sectional, stratified multistage sampling design covering all 28 states, 8 union territories, and 707 districts in India. The sampling frame followed a hierarchical structure, with households nested within Primary Sampling Units (PSUs), and PSUs nested within districts. Fieldwork in Uttar Pradesh was carried out in two phases (January–March 2020 and November 2020–April 2021), yielding a final sample of 70,710 households, 93,124 women, and 12,043 men. Data were collected using four standardized instruments Household, Woman's, Man's, and Biomarker questionnaires based on the international Demographic and Health Surveys (DHS) model. To ensure validity and reliability, these instruments were pre-tested and translated into local languages. The data on children was collected through the Women's Questionnaire, with their mothers serving as respondents.⁵

Sample Size

Analysis has been done on 35,766 live births recorded in the five years preceding the survey (2015–2021). Within this cohort, 1,292 neonatal deaths were identified.

Outcome variable

The outcome variable was neonatal death, where deaths occurring within 0–28 days were treated as event (coded as 1), and survival beyond this period was treated as censored observations (coded as 0).

Explanatory Variables

The analysis incorporated a wide range of individual- and community-level factors. The selection of variables was based on a comprehensive literature review to ensure the inclusion of key factors influencing neonatal mortality.⁷⁻⁹ To better align with the study's objectives, several variables from the original NFHS-5 dataset including religion, caste, birth size, place of delivery, type of birth, and parity were recategorized. Additionally, continuous quantitative variables such as maternal age, BMI, and the child's birth weight were categories to better capture potential non-linear risks.^{9,12,13} A stepwise regression approach was employed to identify relevant variables for the model, variables with a high proportion of missing data or those exhibiting high multicollinearity were excluded to maintain the robustness and stability of the model. Individual-level factors included socioeconomic and demographic characteristics such as place of residence (rural, urban), mother's education, which was categorised into no education, primary (1-5 years of schooling), secondary (6-12 years of schooling) and higher education (more than 12 years of schooling). This distinct grouping captures the gradient of health literacy impact on mortality.^{16,17} Economic status was assessed using the NFHS Wealth Index, a composite measure calculated via Principal Component Analysis (PCA) based on household assets

and housing characteristics. The population was divided into five quintiles: Poorest (lowest 20%), Poorer, Middle, Richer, and Richest (highest 20%).¹⁸ The other socioeconomic variables included religion (Hindu, Muslim and Other), caste (Scheduled Caste (SC) & Schedule Tribe (ST), Other Backward Class (OBC) and General), place of delivery (Home, Government, Private and Other), distance to health facility (no problem, big problem and not a big problem) and covered by health insurance (yes/no). The type of cooking fuel used was grouped into non-smoke (LPG/electricity) and smoke (solid fuels).¹⁷ Mother-related factors included mother's age (15-19, 20-24, 25-29, 30-34, 35-39, 40-44 and 45-49 years) to capture non-linear risk patterns, particularly among adolescent and older mothers.^{16,19} Mother's nutritional status (BMI) was categorised according to WHO Asian standards (underweight, normal weight, overweight and obesity).^{16,18} Additional maternal variables included current breastfeeding status (yes and no), type of delivery (caesarean and non-caesarean), mother's anaemia level (severe, mild, not anaemic and moderate) and Parity of mother (1 to 2 children, 3 to 4 children, and more than 4 children). Child-related variables included birth weight (low (less than 2.5 kg), average (2.5-4 kg), high (more than 4 kg), not weighted at birth, and don't know), sex of child (male and female), birth size (large, average and small) and type of birth (single and multiple). Community-level factors were represented by Primary Sampling Units (PSUs) and Districts. In this large-scale survey, PSUs correspond to villages, urban blocks, or census units, while districts represent administrative regions with distinct healthcare resources, economic conditions, and policies that shape neonatal survival chances.⁵

Exclusion and Inclusion Criteria

Inclusion Criteria: In NFHS-5 for the Women's Questionnaire, women aged 15 to 49 years and birth of mothers in last five year were included in the study. Eligible participants were those who were usual residents of the selected households or visitors who had stayed in the household the night before the survey.

Exclusion Criteria: Women who were not in the age criteria (15 to 49 years), not usual residents, or had not stayed in the household the night before the survey were excluded from the study.

Statistical Analysis

Descriptive statistics were summarized using frequencies and percentages. Bivariate associations between prognostic factors and neonatal mortality were assessed using the chi-square test. Kaplan-Meier survival curve was used to examine the survival pattern of the outcome variable. Multicollinearity among independent variables was assessed using Variance Inflation Factors (VIF). The Proportional Hazards (PH) assumption was assessed using the Schoenfeld residuals-based test. The Schoenfeld residuals-based test evaluates the PH assumption by examining whether residuals show a systematic trend over time. A non-significant test result indicates that the assumption holds, ensuring the validity of the proportional hazard assumption.^{20,21} Individual characteristics represented the primary unit of analysis, nested within Primary Sampling Units (PSUs), which were further nested within the districts of Uttar Pradesh. Four distinct models were specified to identify significant predictors. Standard

single-level models estimated hazard ratios based solely on individual fixed effects using Weibull and Exponential distributions, assuming that neonatal mortality risks are independent across clusters. multilevel mixed-effects models incorporated PSUs and districts as random effects to account for clustering and unobserved heterogeneity under both distribution types. Hazard ratios (HRs) with 95% confidence intervals (CIs) were computed, with p-values indicating statistical significance. The Log-Likelihood Ratio (LR) test was used to compare the fit of single-level versus multilevel models. Subsequently, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were applied to select the final optimal model among the multilevel specifications. the Intraclass Correlation Coefficient (ICC) was calculated to quantify the proportion of variance in neonatal mortality attributable to differences at the PSU and district level. Data analysis was performed using Stata v.13 (StataCorp, 2013) incorporating sample weights for population representativeness ensuring robust and comprehensive evaluation of factors contributing to neonatal mortality.²²

Weibull and Exponential Survival Models

In survival analysis, individual-levels models assume that all observations are independent. A basic Weibull and exponential parametric survival model for individual i is given by equation 1 and 2 respectively:

$$h_i(t) = \lambda p t^{p-1} e^{X_i \beta} \quad (1)$$

$$h_i(t) = \lambda e^{X_i \beta} \quad (2)$$

Where: $h_i(t)$ is the hazard function for individual i at time t , λ and p are scale and shape parameters of the Weibull distribution and $X_i \beta$ represents the fixed effects, where X_i is a vector of individual-level covariates. The baseline hazard function $h_0(t)$ for the Weibull and Exponential Models are $\lambda p t^{p-1}$ and λ respectively as the exponential distribution is special case of Weibull distribution when $p=1$.^{23,24}

However, these models assumes that all individuals with the same covariates share the same risk, ignoring unobserved heterogeneity across different groups or levels.

Multilevel Mixed- Effect Parametric Models

The inclusion of random effects in survival models has a significant impact on the estimation of baseline risks by accounting for unobserved heterogeneity and reducing bias in the parameter estimates. The Multilevel Mixed-Effects Parametric Survival Models integrate fixed effects and random intercept effects both, enabling the correlation of observations within the same community level due to shared community-level random effects. The study here used the individual level models and Mixed-effects model with PSUs and districts as community factor. The proportional hazard (PH) models assume that the dependent variables have proportional effect on the hazard function for some baseline hazard $h_0(t)$, which follows a parametric survival function.²⁵

The Proportional Hazards Mixed-effect survival model can be written as

$$h_{ijk}(t) = h_0 \exp(y_{ijk} \beta + U_i + V_{j(i)}) \quad (3)$$

Where $h_0(t)$ is the baseline hazard function of a standard parametric model and $k=1,2,3,\dots,n$ sample of neonates clustered within $j=1,2,\dots,m_j$ PSUs included within $i=1,2,3,\dots$ districts of Uttar Pradesh. The vector y_{ijk} contains the covariates for the fixed effect with regression coefficient of fixed effect β . To account the clustering, the random effects U_i and $V_{j(i)}$, representing the districts and PSUs nested within Districts, reflects random intercept and random coefficient effect respectively. These effects are assumed to follow a multivariate normal distribution with mean zero i.e. $N(0, \Sigma)$ where Σ denotes the variance covariance matrix of random effects. The commonly used parametric survival regression models are Exponential, Weibull, Gamma, Log-logistic and Lognormal. The multilevel Weibull and Exponential models were employed in this study due to their distinct advantages in handling varying hazards rates.^{24,26} The Exponential model is characterized by its assumption of a constant hazard rate over time and in contrast, the Weibull model extends the framework of Exponential model by introducing a shape parameter of varying hazard rates either increasing or decreasing over time. By incorporating random effects, these models account for unobserved heterogeneity across clusters, allowing the capture of variability that may influence survival outcomes. This hierarchical approach not only enhances the robustness of the findings but also maintains flexibility in modelling covariate effects across different levels of analysis.^{27,28}

Intraclass Correlation Coefficient

The intraclass correlation Coefficient (ICC), is a measure of within cluster variation, the variation between individuals within the same cluster, and it was calculated using the formula

$$ICC \text{ (for two level) } \rho^{(2)} = \frac{\sigma_A^2}{(\delta^2 + \sigma_A^2)} \quad (4)$$

$$ICC \text{ (for three level) } \rho^{(3)} = \frac{\sigma_A^2 + \sigma_B^2}{(\delta^2 + \sigma_A^2 + \sigma_B^2)} \quad (5)$$

Where $\delta^2 = \frac{\pi^2}{6p^2}$, where p is the shape parameter for the Weibull distribution. For Exponential distribution, this shape parameter $p=1$. Thus, $\delta^2 = \frac{\pi^2}{6}$.²⁹

Model Selection Criteria

Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC)

AIC and BIC were used to find the best fitted model. AIC, introduced by Akaike, (1973), is a widely used statistical tool for selecting the best-fitting model, while taking into account the number of parameters. AIC aims to choose the model that strikes a balance between maximizing the likelihood of the data fit and minimizing the penalty imposed by the number of parameters, and BIC prioritizes model parsimony, imposing a stronger penalty on complexity, which helps prevent overfitting, especially in large samples. as expressed in the equations 6 and 7 respectively.^{28,30}

$$AIC = 2p - 2 * \ln(\hat{L}) \quad (6)$$

$$BIC = -2 * \ln(\hat{L}) + p * \ln(n) \quad (7)$$

Where, p is the number of parameters in the model and $\hat{L}$ is the likelihood under the fitted model and n is the sample size. We select the final model based on the minimum AIC and BIC values.

The Likelihood Ratio Test

The likelihood ratio (LR) test was used for comparing the fit of two nested models, to assesses whether the inclusion of random effects significantly improves the model's fit to the data in comparison of fixed effect. When applied to individual and Mixed-effects models, the LR test evaluates the null hypothesis that the simpler model (e.g., an individual-level model) is sufficient compared to a more complex model that includes community-level factors.³¹

RESULTS

The study analysed 35,766 neonatal cases from the NFHS-5 dataset of Uttar Pradesh. Of these 1,292 (3.61%) experienced mortalities within 28 days of birth while 34,474 (96.39%) survived the neonatal period. Statistically significant associations were observed for socio-demographic factors including residence, wealth, and education, as well as critical biological determinants such as maternal age, parity, birth weight, and multiple birth status. (Table 1)

Figure 1 presents the Kaplan-Meier survival curve, which shows survival probability declines sharply in the initial days, indicating that neonatal mortality is highest immediately after birth. Following this steep initial decline, the survival curve flattens, suggesting that the risk of neonatal death decreases as the newborn survives the critical early period. The median survival death was 2 days showing that 50 % of neonatal deaths occur within the first two days of life.

The distribution of survival time was assessed for Weibull and Exponential models. In the Exponential model, the baseline hazard was estimated at 0.235, assuming a constant hazard rate over time. In contrast, the Weibull model estimated a baseline hazard of 0.254 with a shape parameter of 0.953 (less than 1), indicating a decreasing hazard rate over time. The log-pseudo-likelihood (-1907.343) for the Weibull model was higher than that of the Exponential model (-1909.228), indicating a better survival distribution for the data as compared to the exponential model.

Figure 2 and Figure 3 represent the survival curves from the Weibull and Exponential regression models, respectively. Both curves show an initial steep decline, indicating that neonatal mortality risk is highest immediately after birth, followed by a levelling off over time.

The mean VIF was found to be 1.24 shows there is no significant multicollinearity between the independent variables. The proportionality-hazards assumption results shows that the hazard ratio

remains constant over time ($p\text{-value} > 0.05$) (Table 2).

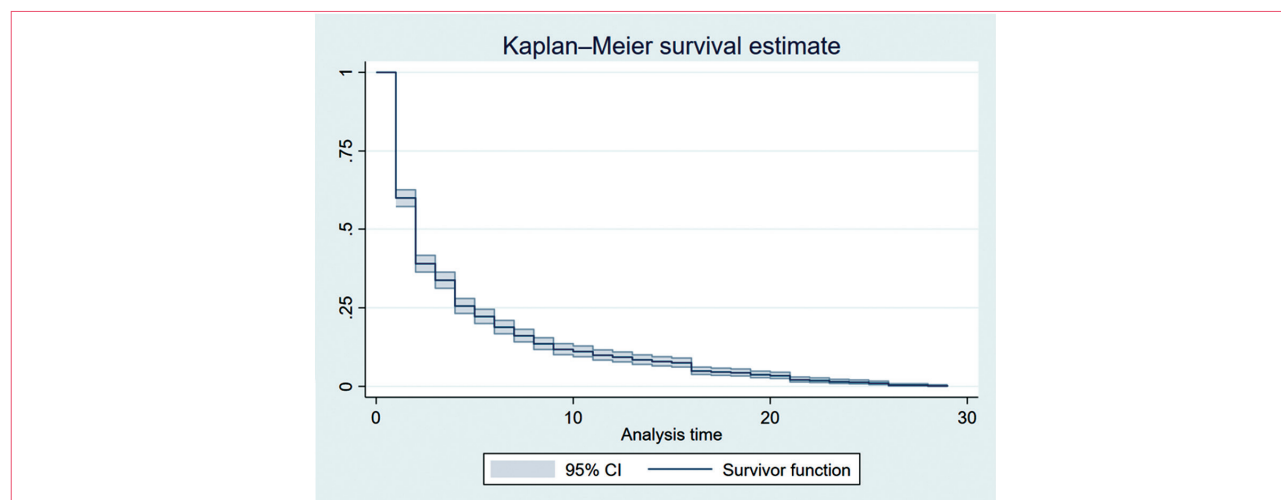


Figure 1. Kaplan Meier Survival Curve

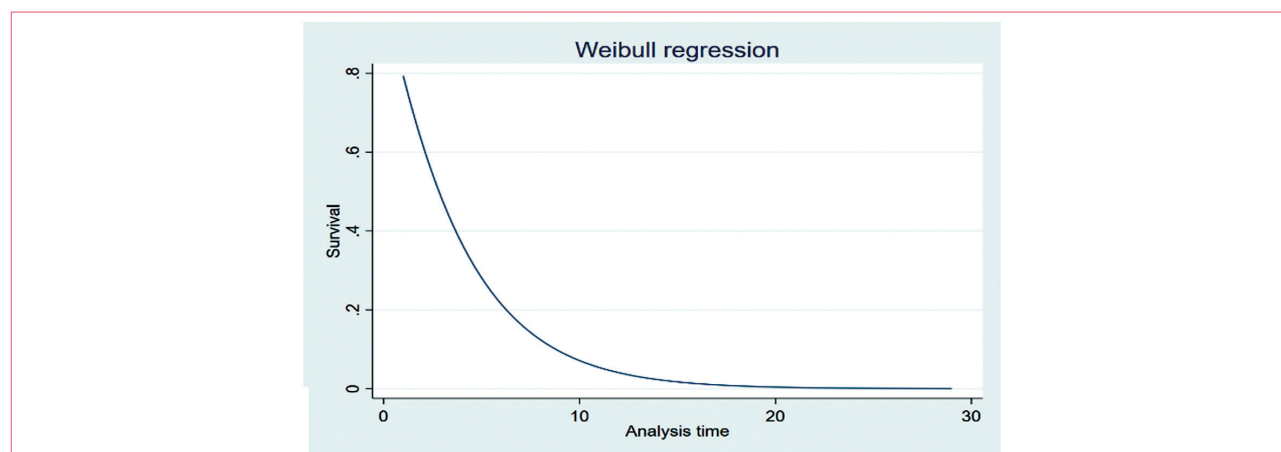


Figure 2. Weibull Survival Curve

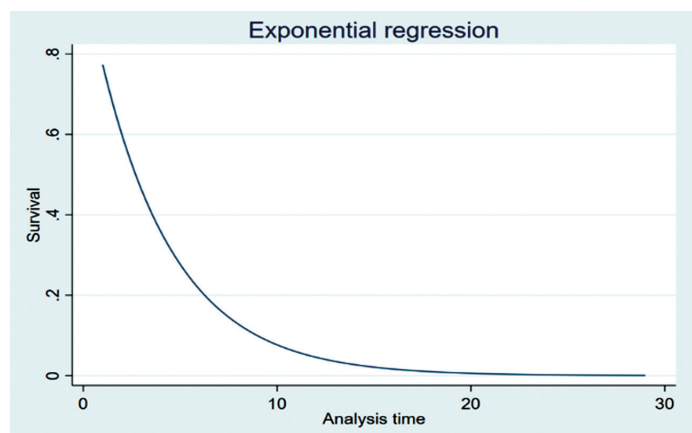


Figure 3. Exponential Survival Curve

Table 1. Distribution of Prognostic Factors and Neonatal Mortality in Uttar Pradesh.

Prognostic Factors	Status of Neonates	
	Survival (Censored)	Death (Outcome)
	N (%)	N (%)
Place of Residence***		
Urban	5,464(97.29)	152(2.71)
Rural	29,010(96.22)	1,140(3.78)
Highest Education Level of Mother***		
No Education	10,433(96.15)	418(3.85)
Primary	4,665(95.61)	214(4.39)
Secondary	13,951(96.43)	516(3.57)
Higher	5,425(97.41)	144(2.59)
Wealth Index***		
Poorest	9,664(96.17)	385(3.83)
Poorer	9,040(95.72)	404(4.28)
Middle	6,314(96.24)	247(3.76)
Richer	5,062(97.20)	146 (2.80)
Richest	4,394(97.56)	110(2.44)
Religion		
Hindu	28,271(96.34)	1,075(3.66)
Muslim	6,103(96.58)	216(3.42)
Other	100(99.01)	1(0.99)
Caste**		
SC & ST	9,870(96.00)	411(4.00)
OBC	18,782(96.53)	676(3.47)
General	5,627(96.67)	194(3.33)
Type of Cooking Fuel use in Household***		
Smoke	13,238(97.05)	403(2.95)
Non-Smoke	21,236(95.98)	889(4.02)
Age of Mother***		
15-19 years	374(92.12)	32(7.88)
20-24 years	8,879(95.57)	412(4.43)
25-29 years	14,485(96.78)	486(3.22)
30-34 years	6,957(96.83)	228(3.17)
35-39 years	2,792(96.74)	94(3.26)
40-44 years	707(95.80)	31(4.20)
45-49 years	180(95.24)	9 (4.76)
Mothers Nutritional Status (BMI)		
Underweight	5,570(96.33)	212(3.67)
Normal weight	21,728(96.47)	794(3.53)
Overweight	4,776(96.21)	188(3.79)
Obesity	1,119(95.31)	59(4.69)
Distance to Health Facility		
No Problem	13,162(96.61)	462(3.39)
Big Problem	7,604(96.23)	298(3.77)
Not a big Problem	13,708(96.26)	532(3.74)
Covered by Health Insurance		
No	32,385(96.43)	1,200(3.57)
Yes	2,089 (95.78)	92(4.22)
Place of Delivery***		
Home	5,697(96.28)	220(3.72)
Government	20,679(96.83)	676(3.17)
Private	8,057(95.33)	395(4.67)
Other	41(97.62)	1(2.38)
Type of Delivery		
Non- Caesarean	30,165(96.53)	1,083(3.47)
Caesarean	4,309(95.37)	209(4.63)
Anaemia Level of Mother***		
Severe	830 (94.10)	52(5.90)
Moderate	9,198(96.19)	364(3.81)
Mild	8,159(96.61)	286(3.39)
Not Anaemic	14,486(96.51)	524(3.49)
Parity ***		
1 to 2 children	18,336(96.90)	587(3.10)
3 to 4 children	12,138(96.11)	491(3.89)
More than 4 children	4,000(94.92)	214(5.08)
Birth Weight***		
Low birth weight	13,253(96.68)	455(3.32)
Normal birth weight	15,765(98.10)	306(1.90)
High birth weight	558(91.58)	33(5.58)
Not weighted at birth	3,192(90.56)	408(9.44)
Don't know	986(91.64)	90(8.36)
Sex of Child**		
Male	17,870(96.19)	708(3.81)
Female	16,604(96.60)	584(3.40)
Birth Size***		
Large	5,736(96.18)	228(3.82)
Average	24,717(96.91)	787(3.09)
Small	3,577(10.38)	199(5.27)
Don't Know	444(85.06)	78(14.94)
Currently Breastfeeding***		
No	11,100(92.97)	839(7.03)
Yes	23,374(98.10)	453(1.90)
Status of Birth***		
Single birth	33,995(96.65)	1,178(3.35)
Multiple birth	479(80.78)	114(19.22)

***p-value<0.001, **p-value<0.01, *p-value<0.01

Table 2. Test of Proportional-Hazard Assumption

	χ^2	Degree of freedom	p-value
Global test	33.89	43	0.838*

*p-value>0.05

Model Comparison and Best Fitted Model

The comparison between individual level models and the multilevel mixed effects models. From the Likelihood Ratio Test, the Multilevel Mixed-effects Models provided a significantly better fit than individual-level models, justifying the inclusion of hierarchical structures in the analysis for both Weibull and exponential distributions. Among the Mixed-effects models, the Multilevel Weibull Mixed-Effect Model was the best fit, as indicated by the lowest Akaike Information Criterion and Bayesian Information Criteria (AIC: 5358.462, BIC 5593.139) respectively, compared to the Multilevel Exponential Mixed-effect Model (AIC: 5536.926, BIC 5771.980). The lower AIC and BIC value showed that the Weibull model better captures the underlying variability in neonatal mortality risk. (Table 3).

Table 3. Comparison of Multilevel Parametric Survival Models

Models	LR test (Individual vs Mixed-effects Model)	Loglikelihood	AIC	BIC
Weibull	159.62 **	-2633.231	5358.462	5593.139
Exponential	12.12 **	-2722.463	5536.926	5771.980

***p-value>0.05

Table 4 presents the hazard rates for the Multilevel Weibull Mixed Effect model (Model 2), contrasting them with the individual-level Weibull model (Model 1) to highlighting the impact of adjusting for clustering.

After adjusting for community-level clustering in Model 2, religion emerged as a significant predictor. Muslim mothers exhibited a hazard ratio of 1.35 (95% CI: 1.019–1.788) compared to Hindu mothers, an association that was masked in the Model 1. Conversely, the elevated risk initially observed for mothers with secondary education in Model 1 (HR: 1.30) lost statistical significance after accounting for clustering, suggesting that educational disparities may be confounded by unobserved community-level characteristics.

Exposure to smoke-producing cooking fuels remained a robust risk factor. In the final multilevel model, households using these fuels showed a 1.30 times higher hazard of neonatal death (HR: 1.30; 95% CI: 1.014–1.677) than those using clean fuels. Delivery in non-government institutions was associated with the most severe risk, with the HR increasing from 2.73 in Model 1 to 3.24 in Model 2 (95% CI: 1.615–6.502). Advanced maternal age (45–49 years) was a critical determinant, associated with a nearly threefold increase in mortality risk (HR: 2.93) compared to the reference group (20–24 years) in the final model. Similarly, neonates not weighed at birth faced significantly higher mortality,

with the hazard ratio rising from 1.79 in the individual model to 2.12 in the multilevel model (95% CI: 1.471–3.05). The counter-intuitive risk associated with single births observed in Model 1 (HR: 1.45) became non-significant in the multilevel specification, implying that the initial finding was driven by regional or facility-level confounders. Place of residence, wealth index, and maternal BMI did not show statistically significant associations in the final model.

Table 5. Multilevel Weibull Model with Fixed and Random Effects

Prognostic Factors	Model 1 HR (95 % CI)	Model 2 HR (95 % CI)	Prognostic Factors	Model 1 HR (95 % CI)	Model 2 HR (95 % CI)
Fixed Effects			Fixed Effects		
Place of Residence ® Rural			Anaemia Level of Mother ® Not Anaemic		
Rural	1.065(0.845,1.342)	1.004(0.676,1.492)	Severe	0.966(0.701,1.333)	0.805(0.474,1.368)
Highest Education Level of Mother ® Higher Education			Moderate	1.065(0.898,1.264)	1.192(0.775,1.834)
No Education	0.893(0.674,1.184)	0.832(0.534,1.297)	Mild	0.965(0.806,1.155)	1.059(0.801,1.399)
Primary	0.801(0.592,1.084)	0.749(0.496,1.132)	Currently Breastfeeding Mother ® No		
Secondary	1.306*(1.009,1.689)	1.292(0.899,1.858)	Yes	0.935(0.803,1.088)	0.878(0.686,1.124)
Caste ® General			Parity ® 1 to 2 children		
SC & ST	1.271(0.991,1.629)	1.575(0.991,2.503)	3 to 4 Children	0.872(0.734,1.035)	0.813(0.646,1.024)
OBC	1.238(0.993,1.543)	1.425(0.907,2.24)	More than 4 Children	0.881(0.672,1.155)	0.847(0.569,1.261)
Wealth Index ® Richest			Birth Weight of Child ® Average Birth Weight		
Poorest	1.198(0.835,1.718)	0.982(0.589,1.638)	Low Birth Weight	0.821(0.67,1.005)	0.668**(0.514,0.868)
Poorer	1.235(0.894,1.705)	1.13(0.686,1.861)	High Birth Weight	1.497(0.95,2.36)	1.837(0.972,3.472)
Middle	0.911(0.658,1.26)	0.803(0.507,1.27)	Not Weighted at Birth	1.793**(1.435,2.24)	2.118**(1.471,3.05)
Richer	1.033(0.735,1.45)	0.878(0.531,1.45)	Don't know	1.285(0.94,1.755)	1.535(0.952,2.476)
Religion® Hindu			Gender of Child ® Male		
Muslim	1.188(0.952,1.482)	1.35*(1.019,1.788)	Female	0.913(0.791,1.055)	0.87(0.699,1.082)
Type of Cooking Fuel use in Household ® Non-Smoke			Brith Size ® Average		
Smoke	1.22*(1.016,1.465)	1.304*(1.014,1.677)	Large	1.096(0.905,1.328)	1.158(0.829,1.619)
Distance to Health Facility ® No Problem			Small	0.923(0.751,1.135)	1.047(0.759,1.446)
Big Problem	0.898(0.732,1.102)	1.039(0.758,1.423)	Don't Know	1.038(0.758,1.421)	1.051(0.687,1.607)
Not a Big Problem	1.025(0.874,1.202)	1.029(0.771,1.373)	Status of Birth ® Multiple birth		
Covered By Health Insurance ® Yes			Single Birth	1.45*(1.081,1.945)	0.99(0.55,1.784)
No	0.975(0.722,1.317)	0.848(0.558,1.29)	Constant	0.094**(0.048,0.186)	0.068**(0.025,0.186)
Place of Delivery ® Government			Shape parameter (ln(p))	0.056(0.021,0.092)	0.484(0.31,0.658)
Home	0.821(0.657,1.028)	0.619*(0.42,0.912)	Random Effects		
Private	1.22*(1.019,1.461)	1.289(0.989,1.681)	Variance (Districts)	-	0.023(0.004,1.321)
Other	2.734**(1.938,3.857)	3.241**(1.615,6.502)	Variance (Districts>PSUs)	-	1.238(0.636,2.410)
Age of Mother ® 20-24 years			Mothers Nutritional Status (BMI) ® Normal Weight		
15-19 Years	0.786(0.459,1.347)	0.685(0.371,1.266)	Underweight	0.984(0.799,1.213)	1.111(0.778,1.588)
25-29 Years	0.954(0.804,1.132)	1.071(0.845,1.357)	Overweight	0.936(0.765,1.145)	0.829(0.648,1.062)
30-34 Years	1.176(0.929,1.489)	1.356(0.934,1.969)	Obesity	0.886(0.637,1.233)	0.704(0.423,1.173)
35-40 Years	0.974(0.673,1.41)	1.271(0.753,2.145)	Type of Delivery ® Caesarean		
41-44 Years	1.465(0.887,2.42)	1.622(0.702,3.747)	Non-Caesarean	1.084(0.878,1.337)	1.355(0.998,1.839)
45-49 Years	3.103**(1.807,5.327)	2.927**(1.241,6.901)			

*p-value<0.05, **p-value<0.01, HR: Hazard Ratio, CI: Confidence Interval.

Model 1: Individual Weibull model, Model 2: Multilevel mixed effect Weibull model

Variance Components and Intraclass Correlation Coefficient (ICC)

For the best-fitting Multilevel Weibull Mixed-Effect Model, the district-level variance was 0.023, whereas the PSU-level variance was 1.28, indicating substantial heterogeneity at the PSU level within districts (Table 4). The Intraclass Correlation Coefficient (ICC) analysis further confirmed this pattern, with a district-level ICC of 3.55 % and a significantly higher PSU-level ICC of 67.58% (Table 5). This showed that while district-level variations contribute minimally to neonatal mortality, a large proportion of the variance is explained by differences between PSUs nested within the district, highlighting the strong influence of localized community-level factors.

Table 4. Intraclass correlation coefficients for Weibull Mixed-effect Model

Model	Variance Component	Variance	ICC (%)
Weibull Mixed-effect Model	Individual (δ^2)	0.625	
	District (σ_A^2)	0.023	3.549
	PSUs within the District (σ_B^2)	1.28	67.583

***p-value>0.05

DISCUSSION

This study provides critical insights into the determinants of neonatal mortality using Multilevel Weibull and Exponential, Individual and Mixed-effects models. The survival estimates revealed that mortality risk is most acute immediately following birth, emphasizing the neonate's vulnerability during the early perinatal period due to complications, a finding consistent with established literature.³²

While the study identified several influential factors including socio-religious background, household environment, and maternal characteristics adjusting for community-level clustering provided clarifying insights into how these factors operate. The Influence of mother education became insignificant after account for unobserved heterogeneity in the multilevel model.¹⁶ This highlights the importance of adjusting for variability across Primary Sampling Units (PSUs) and districts, mirroring earlier research suggesting community factors may mediate this relationship. Conversely, the elevated risk observed among Muslim mothers in the final model points to potential socio-cultural barriers in healthcare utilization that require targeted exploration.

Environmental and healthcare factors also emerged as critical drivers. Reliance on smoke-producing fuels consistently increased mortality risk, corroborating evidence on the adverse effects of indoor air pollution.¹⁷ Regarding healthcare access, the significant hazard associated with non-institutional deliveries reinforces the global consensus advocating for skilled birth attendance.³³ Biological factors remained robust predictors as well; maternal age was a significant determinant, consistent with the known obstetric complications in higher age group.³² Finally, the heightened risk for neonates not weighed at birth highlights a critical gap in postnatal care. This aligns with studies linking unmonitored birth weight to higher infant mortality in India, emphasizing the urgent need for rigorous neonatal surveillance.⁷

The study determined the superiority of the mixed-effects model over standard individual-level approaches. A pivotal finding was the dominance of PSU-level variance over district-level variance. This disparity indicates that local healthcare accessibility, socioeconomic conditions, and specific cultural practices within communities play a far more substantial role in influencing survival than broader administrative district factors. Furthermore, the distinct shift in the Weibull shape parameter between the individual and mixed-effects models underscores the necessity of accounting for this unobserved heterogeneity; failing to do so obscures the true temporal nature of the mortality risk.³⁴

Thus, by incorporating random effects as both district and primary sampling unit (PSU) levels, the study captures significant heterogeneity in neonatal mortality risk, providing a more nuanced understanding of the interplay between individual, household, and community-level factors.

Strengths and Limitations

This study's strength lies in its multilevel modelling approach with Mixed-effects survival analysis, which accounts for heterogeneity at individual, household, and community levels while addressing clustering effects within PSUs and districts. However, limitations include missing data and non-response in certain variables, as the study relies on secondary data from NFHS-5. Additionally, the dataset lacks detailed health system and geospatial factors, limiting a comprehensive assessment of healthcare accessibility. Future research might integrate health system variables and spatial analysis to enhance policy-relevant insights.

CONCLUSION

This study highlights the critical role of maternal education, household environment, delivery place, and neonatal health at birth in determining neonatal mortality. Addressing these significant determinants requires multi-tiered interventions targeting individual, household, and community-level factors. By accounting for hierarchical structures, the Multilevel Weibull Mixed-effects model provides robust insights into the complex determinants of neonatal mortality, outperforming individual-level and Exponential Mixed-effects models. The inclusion of random effects in analytical models offers a more nuanced understanding of these risks, emphasizing the need for localized, data-driven, and equity-focused strategies to effectively reduce neonatal mortality.

Potential Conflicts of Interest

There is no conflict of interest.

Ethical consideration

This study is based on secondary data from the National Family Health Survey fifth round (NFHS-

5, 2019–21), conducted by the International Institute for Population Sciences (IIPS), Mumbai, under the Ministry of Health and Family Welfare, Government of India. The dataset is publicly available through the DHS Program (www.dhsprogram.com), and access was obtained via data-use agreement. Since the analysis relies solely on anonymized, publicly archived data with no direct human subject involvement, institutional ethical clearance was not required.

ACKNOWLEDGMENT

We express our sincere gratitude to the Demographic and Health Surveys (DHS) Program for granting access to the National Family Health Survey (NFHS-5) 2019-2021 dataset, which formed the basis of this research. We also have our appreciation to the journal editorial team and reviewers for their constructive feedback and the opportunity to share our work.

REFERENCES

1. WHO. Newborn mortality: World Health Organization; 2024 [WHO]. Available from: <https://www.who.int/news-room/fact-sheets/detail/newborn-mortality>.
2. UNICEF Data: Neonatal Mortality New York: UNICEF; 2024 [Available from: Available from: <https://data.unicef.org/topic/child-survival/neonatal-mortality/>].
3. Duke T. Levels and trends in child mortality estimation. Archives of Disease in Childhood. 2024;109(8):620–1.
4. Office of the Registrar General & Census Commissioner, India. Sample Registration System (SRS) Statistical Report 2020. New Delhi: Government of India; 2022.
5. Government of India, Ministry of Health and Family Welfare. Compendium of Fact Sheets India and 14 States/UTs (Phase-II): National Family Health Survey (NFHS-5) 2019-21. New Delhi: MoHFW; 2021.
6. National Health Systems Resource Centre (NHSRC). Health Dossier 2021: Reflections on Key Health Indicators – Uttar Pradesh New Delhi: NHSRC; 2021.
7. Jana A, Saha UR, Reshmi RS, Muhammad T. Relationship between low birth weight and infant mortality: evidence from National Family Health Survey 2019-21, India. Archives of Public Health. 2023;81(1).
8. Y G, S B. Survival Analysis of Under-Five Mortality of Children and its Associated Risk Factors in Ethiopia. Journal of Biosensors & Bioelectronics. 2016;7(3).
9. Saroj RK, Murthy KHHVSSN, Kumar M, Singh R, Kumar A. Survival Parametric Models to

- Estimate the Factors of Under-Five Child Mortality. *Journal of Health Research and Reviews*. 2019;6(2):82–8.
10. Ayiko R, Antai D, Kulane A. Trends and determinants of under-five mortality in Uganda. *East African journal of public health*. 2009;6 2:136–40.
 11. Pandey A, Sharma R. Neonatal Deaths in Empowered Action Group States: Multilevel Modeling to Study the within District Education Effect on Neonatal Survival. *International Journal of Scientific Research*. 2022:54–7.
 12. Kumar M, Patel K. Differential and determinants of neonatal mortality: A comparative study in Northern and Southern Regions of India. *Indian Journal of Community Medicine*. 2021;46(3).
 13. Moise IK, Bora JK, Saikia N. Neonatal and under-five mortality rate in Indian districts with reference to Sustainable Development Goal 3: An analysis of the National Family Health Survey of India (NFHS), 2015–2016. *Plos One*. 2018;13(7).
 14. Jaiswal AK, Alagarajan M, Meitei WB. Survival among children under-five in India: a parametric multilevel survival approach. *BMC Public Health*. 2024;24(1).
 15. The DHS Program [Internet]. 2024. Available from: Available from: <https://dhsprogram.com/data/available-datasets.cfm>.
 16. Patel N, Olickal JJ. Maternal and child factors of under-five mortality in India. Findings from NFHS-4. *Clinical Epidemiology and Global Health*. 2021;12.
 17. Patel AB, Meleth S, Pasha O, Goudar SS, Esamai F, Garces AL, et al. Impact of exposure to cooking fuels on stillbirths, perinatal, very early and late neonatal mortality - a multicenter prospective cohort study in rural communities in India, Pakistan, Kenya, Zambia and Guatemala. *Maternal Health, Neonatology and Perinatology*. 2015;1(1).
 18. Rutstein SO JK. The DHS Wealth Index. DHS Comparative Reports No. 6. Calverton: ORC Macro. 2004.
 19. Cleary-Goldman J, Malone FD, Vidaver J, Ball RH, Nyberg DA, Comstock CH, et al. Impact of Maternal Age on Obstetric Outcome. *Obstetrics & Gynecology*. 2005;105(5, Part 1):983–90.
 20. Kuitunen I, Ponkilainen VT, Uimonen MM, Eskelinen A, Reito A. Testing the proportional hazards assumption in cox regression and dealing with possible non-proportionality in total joint arthroplasty research: methodological perspectives and review. *BMC Musculoskeletal Disorders*. 2021;22(1).

21. Andrade C. Survival Analysis, Kaplan-Meier Curves, and Cox Regression: Basic Concepts. *Indian Journal of Psychological Medicine*. 2023;45(4):434–5.
22. StataCorp. Stata Statistical Software: Release 13.1 [Computer program]. College Station (TX): StataCorp LLC; 2013.
23. Zhang Z. Parametric regression model for survival data: Weibull regression model as an example. *Annals of Translational Medicine*. 2016;4(24):484–.
24. Oyato B, Zakir H, Hussein D, Lemma T, Awol M. Time to Death and Its Predictors Among Infants in Ethiopia: Multilevel Mixed-Effects Parametric Survival Analysis Using the 2019 Ethiopian Mini Demographic Health Survey. *Pediatric Health, Medicine and Therapeutics*. 2023;Volume 14:169–83.
25. Crowther MJ. Multilevel mixed-effects parametric survival analysis: Estimation, simulation, and application. *The Stata Journal: Promoting communications on statistics and Stata*. 2019;19(4):931–49.
26. Tesema GA, Teshale AB, Tessema ZT. Incidence and predictors of under-five mortality in East Africa using multilevel Weibull regression modeling. *Archives of Public Health*. 2021;79(1).
27. Taketomi N, Yamamoto K, Chesneau C, Emura T. Parametric Distributions for Survival and Reliability Analyses, a Review and Historical Sketch. *Mathematics*. 2022;10(20).
28. Zhan Y, Fagbamigbe AF, Norrman E, Bergh C, Wennerholm U-B, Petzold M. Comparison of the performances of survival analysis regression models for analysis of conception modes and risk of type-1 diabetes among 1985–2015 Swedish birth cohort. *Plos One*. 2021;16(6).
29. Masoomnikbakht KFVa. APPLICATION REMLMODEL AND DETERMINING CUT OFF OF ICC BY MULTI-LEVEL MODEL BASED ON MARKOV CHAINS SIMULATION IN HEALTH. *Indian Journal of Fundamental and Applied Life Sciences*. 2015;5(S2):1432–48/ Vajargah and Nikbakht.
30. Cavanaugh JE, Neath AA. The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Computational Statistics*. 2019;11(3).
31. Lewis F, Butler A, Gilbert L. A unified approach to model selection using the likelihood ratio test. *Methods in Ecology and Evolution*. 2010;2(2):155–62.
32. Subramanian SV, Kumar A, Pullum TW, Ambade M, Rajpal S, Kim R. Early-Neonatal, Late-Neonatal, Postneonatal, and Child Mortality Rates Across India, 1993-2021. *JAMA Network Open*. 2024;7(5).

33. Goudar SS, Goco N, Somannavar MS, Kavi A, Vernekar SS, Tshefu A, et al. Institutional deliveries and stillbirth and neonatal mortality in the Global Network's Maternal and Newborn Health Registry. *Reproductive Health*. 2020;17(S3).
34. A.U C, S.O O, V.E J-D, Li B. A Multilevel Hazards Model for Child Mortality In Nigeria. *Journal of Child and Adolescent Psychiatry*. 2019;1(1):9–23.