Demography India Vol. 54, No. 1 (2025) ISSN 0970-454X

Demography India
A Journal of Indian Association of Study of Population Journal
Homepage: https://demographyindia.iasp.ac.in/

A Revised Estimate of NMR with Effect Due to Migration and Discounted Number of Deaths

Ardhendu Banerjee * , Arindam Gupta† and Asis Kumar Chattopadhyay‡

Abstract

The neonatal mortality rate (NMR) is a fundamental gauge of a nation's progress and health. However, the traditional method of measurement provides a summary-picture that might not accurately depict reality. Counting total number of neonates in the denominator continuously till the day 28 also introduces inaccuracy, since the chance of neonatal mortality is highest in the first few hours or days and leave the cohort shortly after death occurs. As we know that the average hourly risk in the United States is approximately 0.07, with the risk being 0.91 per 1000 live births in the first hour and 1.58 per 1000 live births in the next 23 hours. So the actual number of children at daily risk of death is therefore more significant than total number of live births. Furthermore, daily-migration makes it impossible to track cohorts, especially in developing nations, pointing to the incompleteness of the data. As a result, this paper tried to update the traditional approach by fusing migration rate with daily neonatal mortality risk estimates. Consequently, it can be observed that for India, the NMR is 32.74 for year 2015 when it is reported below 30 according to DHS and SRS. It is an accurate representation of the situation rather than an exaggeration. The information provided in this paper may help to better the policies and initiatives that Indian society has access to.

Keywords

Neonatal Mortality Rate, Migration,
Estimation, Measure,
Daily Risk of Death

*Corresponding Author is a Research Scholar, Department of Statistics, Calcutta University, Kolkata, Email: rana.ardh1024@gmail.com

†Associate Professor, Department of Statistics, Burdwan University, Burdwan

‡Professor, Department of Statistics, Calcutta University, Kolkata

Introduction

Does migration have no impact on the NMR (Neonatal Mortality Rate), particularly in the second decade of 21st century? As we know that an essential indicator of a country's development and health is the NMR. It is a serious global public health issue, particularly in developing countries. Neonatal mortality is caused by a multitude of factors, including maternal health, access to high-quality healthcare, socioeconomic status, and environmental conditions. Three reasons, prematurity and low birth weight, newborn infections, birth asphyxia, and delivery trauma, accounted for 78% of all neonatal fatalities in India, according to Million Death Study Collaborators (2010). 50% of all deaths between the ages of 1 and 59 months were related to pneumonia and diarrheal diseases. Additionally, it was discovered that girls in central India had four times the mortality rate from diarrheal disease compared to boys in west India and five times the mortality rate (per 1000 livebirths) from pneumonia compared to boys in south India. Both newborn and child mortality rates have been gradually falling since the turn of the century, when they were 215 fatalities per 1,000 live births (Indian Office of the Registrar General, 1971), but there is still a long way to go. The Millennium Development Goals (MDGs) have spurred efforts to considerably reduce maternal and infant mortality; nevertheless, deaths in the neonatal period, or the first four weeks following delivery, have decreased more slowly Oza et al. (2014).

There is evidence indicating the estimated average annual rate of mortality decrease for newborns between 1990 and 2013 was 2.2%, while the rate for babies aged 1 to 59 months was 4.0% and the rate for maternal deaths was 2.6%. It was discovered that there is an astonishingly high chance of passing away in the first few days after delivery. 2.8 million (or 44%) of the 6.3 million deaths of children under five that happened in 2013 were during the neonatal era, and an additional 1.2 million deaths occurred intrapartum during pregnancy. Additionally, there is an uneven distribution of newborn and infant mortality risk in India.

However, Kanaiaupuni and Donato (1999) assert that there isn't a direct, linear correlation between migration and the likelihood of dying. However, as migration increased and displaced a larger proportion of the community, infant survival significantly decreased, net of migration dollars. While communal life was initially disrupted during a high migration period, this effect changed as the migration process in the communities of origin changed. In communities where at least median migration intensity rates had been in place for 20 years, infant mortality was nearly half as common. Not to mention, after adjusting for every other factor, high annual remittances of increased infant survival. These findings all suggest a strong link between a community's position in the migration process and infant health. The relationship between illness in the nineteenth century and global migration is another area of interest for historical demography. The deaths of the migrants increased rapidly in the late 1800s due to the widespread deployment of European soldiers to India, the Dutch East Indies, and the British West Indies (Curtin, 1980). The troops' vulnerability to tropical illnesses in the locations resulted in these "relocation costs." Due to a lack of data, migration is rarely examined as a factor in the use of healthcare, despite the fact that it has an impact on children as well.

Three groups of children—those left behind in rural areas, those who migrate with their mothers, and those who stay behind after migration has occurred—may be affected by the process of

rural to urban migration to differing degrees, according to Brockerhoff's (1994). According to Brockerhoff (1995), each of these groups is believed to have a different probability of surviving, with the highest mortality risks going to those who remain in rural areas and those who arrive two years prior to or after migration. Children who arrive more than two years after migration have the lowest death rates, and it doesn't appear that a child's mortality rate will decline the longer they live in an urban area Brockerhoff (1990). However, even the children born more than two years after the relocation do not have the low mortality rates of city dwellers Brockerhoff (1995).
Since there is a larger chance of mortality in the first few days, we have considered finding a day-based measure of neonatal deaths in this study. Additionally, we take into account the migration rate as a contributing component to the calculation, which is typically disregarded in favour of the assumption that the cohort is closed under migration. We go through several sections in that order to get to the target state. We will discuss some of the possible measures and their limitations in the second section. We shall next put out our measurement and create the pertinent statistical analysis for India. Finally, we will go over and illustrate the benefits and a few drawbacks of this metric that can be fixed in later research. We hope this trip will be thrilling.

Method of Estimation

According to Shryock and Siegel (1973) the standard infant mortality rate typically provides a sufficiently accurate estimate of the likelihood of passing away between the time of birth and the first birthday for the year to which the basic death statistics pertain. Although it may not be particularly suitable for this purpose in developed places, it has been commonly employed as an indicator of a community's health status and, consequently, of its standard of life. So the conventional infant mortality rate may be usefully "broken up" into a rate covering the first month or so and a rate for the remainder of the year due to the extremely high rate of mortality in the first hours, days, and weeks of life as well as the difference in the causes accounting for infant deaths at the earlier and later ages of infancy. Neonatal mortality rate refers to the rate for the first period, and post-neonatal mortality rate refers to the rate for the second period. There are different measures to find out the vivid picture through some figures.

Direct Method

The number of newborn deaths(D) under 4 weeks (28 days) or under 1 month of age per 1000 live births(B) in a year is known as the neonatal mortality rate, i.e.,

$$NMR = \frac{D_{0-3weeks}}{B} \times 1000 \qquad (1)$$

This direct estimating method makes the assumption that all live births and neonatal deaths have been accurately recorded, and that the data are comprehensive and accurate. However, vital registration systems are frequently insufficient in many developing nations, particularly in rural regions, and births and deaths are frequently only partially recorded. Therefore, in such circumstances, direct estimation may overestimate the true NMR. The neonatal mortality rate is close to a probability of neonatal death because nearly all (over 95%) fatalities of newborns under one month of age occur to infants born in the same year. However, this conventional mortality rate will give a misleading estimate of the level and trend of NMR if there are significant variations in the number of births between and within years. So

the conventional rate should be adjusted to account for the actual population at risk.

On the other hand, in general cases instead of NMR, the population of an age cohort will be less in the middle of the year than it was at the beginning of the year, barring net immigration and changes in the number of births, hence central death rates typically tend to be greater than mortality rates. The discrepancies between the mortality rates and central death rates for people aged 60 to 69 years old point to this. For instance, by using the relevant probabilities for this 5-year period, it is possible to determine the likelihood that a 5-year-old will pass away within that time. The odds at increasing ages in increasing years of the calendar would be used. Combining the relevant single-year-of-age probabilities for a single calendar year can yield a synthetic probability of this type. The equation, according to Shryock and Siegel (1973), would look like this:

$$ {}_5q_5^{'} = 1 - \prod_{i=5}^{9}(1 - q_i^{'}) \qquad (2) $$

where ${}_5q_5^{'}$ is the likelihood that a 5-year-old would die during the following 5 years, and $q_i^{'}$ is the 1-year mortality rate between ages i and $i+1$. Typically, single-year-of-age death statistics tabulations are either unavailable or too imprecise to be used as a foundation for calculating death rates and mortality rates for single years of age. Different interpolation techniques can be used with 5-year age data to produce adjusted estimates of fatalities and death rates in single ages a. The following can be used to get a reasonable estimate of the annual mortality rate for a 5-year age group which can ignores the annual impact of net immigration on mortality and population:

$$ {}_5^{y+1}q_x^y = \frac{D_a}{({}_5P_a + \frac{1}{2}{}_5D_a)} \qquad (3) $$

Indirect Method

Sample Registration System (SRS)

The Registrar General of India created the Sample Registration System (SRS), a demographic monitoring system, to produce accurate estimates of infant and child mortality rates. The SRS comprises registering all births and deaths in the chosen locations, covering a representative sample of the population. By dividing the number of newborn fatalities in the sample by the total number of live births within the same time period, the NMR is calculated. SRS encompasses a representative sample of the population and adjusts for underreporting of births and deaths, resulting in estimates of NMR that are more accurate than those obtained from direct estimation. The SRS, however, calls on a robust demographic surveillance system, which may not be practical in many low-income nations.

According to Dandona et. al.(2023) with an average of 22.4 newborn deaths per 1000 live births over this time period the overall neonatal mortality rate from the sample registration system ranged from 24 neonatal deaths per 1000 in 2016 to 20 per 1000 live births in 2020 (2016 to 2020 the NMR is 24, 23, 23, 22, 20 respectively). They concluded by saying that India has to make steps to improve the documentation of stillbirths in its data collection systems in order to reach its 2030 aim of a single-digit stillbirth rate and monitor initiatives to stop preventable stillbirths.

Demographic and Health Surveys (DHS)

For the purpose of gathering and distributing precise, nationally representative data on population and health in developing nations, the Demographic and Health Surveys (DHS) Programme is in charge. ICF International is in charge of carrying out the project, which is

supported by the United States Agency for International Development (USAID) and other donors like UNICEF, UNFPA, WHO, and UNAIDS. More than 300 demographic and health surveys in more than 90 countries have received technical support from The Demographic and Health Surveys (DHS) Programme since 1984. DHS surveys gather data on total fertility rate (TFR), reproductive health, maternal and child health, immunisation and survival, HIV/AIDS, malaria, and nutrition among women and children who are stunted. The DHS Program's strategic goal is to enhance and institutionalise the data gathering and usage by host nations for programme monitoring and evaluation as well as for policy development decisions.

DHS surveys come in two varieties - Standard DHS Surveys and Interim DHS surveys. Standard DHS Surveys are normally conducted every 5 years and feature large sample sizes (often between 5,000 and 30,000 households) to allow for comparisons across time. While data for all impact evaluation metrics, including mortality rates, may not always be included in interim DHS surveys, they do concentrate on gathering information on essential performance monitoring indicators. These surveys, which are carried out in between rounds of DHS surveys, have shorter questionnaires. These surveys typically have smaller samples than DHS surveys, despite being nationally representative.

Classical Methods

As the deaths are mostly occurred on early days of birth it is necessary to model the NMR on the basis of days.In the United States, the risk is 0.91 per 1000 live births in the first hour and 1.58 per 1000 live births in the following 23 hours, translating to an average hourly risk of roughly 0.07. Accordingly so Oza et. al.(2014) proposed in order to estimate the percentage of deaths occurring on each day of the newborn period for nations lacking adequate VR data, a three-parameter model for the daily risk of neonatal mortality. They then applied this model to the DHS data. The model used the assumption that, provided one survives until day t, the likelihood of dying that day decreases exponentially. The model also allowed the likelihood of passing away on day 0 to deviate from this pattern. This can be written mathematically as:

$$h_t = \begin{cases} \alpha\ , & t = 0 \\ \beta\gamma^{t-1}\ , & 1 \leq t \leq 27 \end{cases} \tag{4}$$

where, h_t is the probability of dying on day t conditional on survival until that day. The unconditional probability of dying on day t of the neonatal period, p_t, can be derived from the multinomial distribution. The likelihood of observing $n_0, \ldots, n_{27}$ deaths in the neonatal period conditional on N livebirths, and the proportion surviving the neonatal period, p_s, can be expressed as:

$$\begin{aligned} p_0^{n_0} \times p_1^{n_1} \times p_2^{n_2} \times \ldots \times p_{27}^{n_{27}} \times p_S^{N-\sum_0^{27} n_t} \\ = p_S^{N-\sum_0^{27} n_t} \times \Pi_0^{27} p_t^{n_t} \end{aligned} \tag{5}$$

To deal with potential misclassifi cation between days 0 and 1 in the DHS data, they combined observed deaths on days 0 and 1 and rewrote the likelihood calculation as:

$$\begin{aligned} &(p_0+p_1)^{n_0+n_1} \times p_2^{n_2} \times \ldots \times p_{27}^{n_{27}} \times p_S^{N-\sum_0^{27} n_t} \\ &= p_S^{N-\sum_0^{27} n_t} \times (p_0+p_1)^{n_0+n_1} \times \Pi_2^{27} p_t^{n_t} \end{aligned} \tag{6}$$

The model utilizes Maximum Likelihood Estimation (MLE) to estimate the parameters α, β, and γ. However, it has certain limitations, primarily the assumption that deaths are solely dependent on time. This assumption leads to the drawback of using an exponential distribution, which

may not accurately reflect the true nature of death rates. Additionally, when incorporating deaths on day 0 and day 1, distributional errors arose, further highlighting the model's limitations. Although the model provides a better fit to the data, this does not necessarily imply that it is a valid or reliable model. The model's assumptions and the resulting errors suggest that it may not fully capture the complexities of the real-world data, and relying on it without addressing these issues could lead to misleading conclusions. It is important to recognize that fitting a model well does not always mean it is the correct or most appropriate model for the data.

Bayesian Methods

It is crucial to have precise numbers, be able to anticipate death levels, and have some idea of the uncertainty in the estimates and projections when assessing a nation's success in lowering child mortality. In reality, developing nations with relatively high death rates, a dearth of effective vital registration systems, and frequent substantial sample mistakes and/or poor quality data are most hit in getting credible mortality estimates. To determine underlying mortality trends in this context, statistical models are required.

That's why Alexander and Alkema (2018) provided a novel model that addresses certain issues with the prior IGME NMR model for estimating the NMR for all nations worldwide. To estimate and predict the NMR and to determine the level of uncertainty surrounding these estimates and projections, they utilise a penalised splines regression model within a Bayesian hierarchical framework. Estimates in the model are based on the link between NMR and U5MR, and country-specific trends are captured using a spline regression model. A Bayesian technique provides an understandable means to share information across multiple countries and time periods for modelling mortality levels across nations, and a data model can incorporate various types of error into the results. The model can be summarize as follows:

$$
\begin{aligned}
r_{c,i} &\sim N(R_{c,t[c,i]}, \delta_i^2) \\
\delta_i^2 &= \begin{cases} \tau_{c,i}^2 , & \text{(VR and SVR Data)} \\ \nu_{c,i}^2 + \omega_{s[c,i]}^2 , & \text{(non-VR Data)} \end{cases} \\
R_{c,t} &= f(U_{c,t}) P_{c,t} \\
log(f(U_{c,t})) &= \beta_0 + \beta_1 (log(U_{c,t}) - log(\theta))_{[U_{c,t} > \theta]} \\
log(P_{c,t}) &= \sum_{k=1}^{K_c} B_k(t) \alpha_{c,k} \\
\alpha_{c,k} &= \lambda_c + [\boldsymbol{D}_{\boldsymbol{K_c}}^{`} (\boldsymbol{D}_{\boldsymbol{k_c}} \boldsymbol{D}_{\boldsymbol{K_c}}^{`})^{-1} \epsilon_c]_k \\
\lambda_c &\sim N(0, \sigma_\lambda^2) \\
\epsilon_{c,q} &\sim N(0, \sigma_\epsilon^2) \\
log(\sigma_{\epsilon_c}^2) &\sim N(\chi, \psi^2)
\end{aligned}
$$

where,

- $R_{c,t}$ is the true ratio in country c at time t, $R_{c,t} = \dfrac{N_{c,t}}{U_{c,t} - N_{c,t}}$, where, $N_{c,t}$ and $U_{c,t}$ are the NMR and U5MR for country c at time t, respectively.
- $r_{c,i}$ is observation i of the ratio in country c.
- $\tau_{c,i}$ is the stochastic standard error, $\nu_{c,i}$ is the sampling error, and $\omega_{s[c,i]}^2$ is nonsampling error for series type s.
- β_0 is the global intercept, β_1 is the global slope with respect to U5MR, θ is the level of U5MR at which β_1 begins to act.
- $P_{c,t}$ is a country-specific multiplier for country c at time t.
- $B_k(t)$ is the k^{th} basis spline evaluated at time t and $\alpha_{c,k}$ is splines coefficient k.
- λ_c is the splines intercept for country c.

- $\boldsymbol{D_{k_c}}$ is a $K_c \times (K_c - 1)$ first-order difference matrix.
- $\epsilon_{c,q}$ are fluctuations around the country-specific intercept.
- $\sigma^2_{\epsilon_c}$ is the country-specific smoothing parameter, modeled hierarchically on the log-scale with mean χ and variance ψ^2.

These model based estimation with multiple assumption may lead to difficulty in understanding along with some underestimation of NMR.

Proposed Measure and Statistical Analysis

As we discussed for estimating infant and child death rates, there are primarily two categories: direct and indirect. Data on a child's birthdate, survival status, and date of death or age at death are all used in direct methods of calculation. Information on children's survival status is provided to moms in particular age cohorts via indirect approaches. Contrary to the direct methods, the indirect methods heavily rely on a number of hypotheses that may or may not come to pass. These assumptions include: little or no change in fertility levels and age patterns; no change or a linear decline in mortality; and a pattern of mortality by age that conforms to known "families," largely derived from European experience.
Data mistakes can affect both kinds of approaches. Both approaches are likely equally flawed by leaving out children who have passed away. The accurate reporting of the age at death as being under or over one year is necessary for the estimation of infant mortality using direct methods. On the other hand indirect method assumptions are frequently violated as well. The position of the estimate in time is another issue with indirect approaches.

In reality, the indirect approaches calculate the likelihood of passing away based on experience that may span many years, yielding an average throughout that time. The methods used to situate the mortality estimate in time can be somewhat inaccurate depending on changes in fertility and death rates. To overcome all these issues we are here to propose a new measure which can reflect the scenario more clearly.

Measure

As per McCutheon (1973) let us assume the IMR as q_0. So that,

$$q_0 = \frac{D}{E} \tag{7}$$

Where,

- D = Number of deaths in the first year of life and
- E = The correct exposed-to-risk, to be used in conjunction with the observed number of deaths D.

Here, we supposed to have a cohort of $R_{i=0}(= r_0)$ in a particular area in a particular period of new-born child i.e., the total number of child exposed to risk initially when $t = 0$ and we will follow them up to reaching their first four weeks of birth (i.e., $T = 28$ days) or first birth anniversary (i.e., $T = 365$ days) or at any desired time point ($T = t$). Now let the number of deaths(D_i) corresponding to i^{th} day from day 1(sometimes data recorded as the day 0) to day 28 or 365 or $T = t$ be realized like $d_0, d_1, d_2, ..., d_{28}, ..., d_{364}, ..., d_{T_t}, ...,$ (observed values of D). The number of neonates left at the end of first day after birth in the study-cohort is $(r_0 - d_0)$ which is exposed to the risk for the next day. Following so, we can say that the number of risk on i^{th} day is $(r_{i-1} - d_{i-1})$. Also we assume that after 28 days all the

survivors will be obsolete for the cohort-study. Now if we consider that a migration is happened during this interval, that is few members of the cohort left the study, then this can be a factor affecting the number of child exposed to risk. Obviously one can argue that why is the number of child who joined the cohort coming from outside not considered? Firstly for many developing and under-developed countries where the registration process is not well-structured, to keep the track to every member may not be possible. But the incoming member can easily be identified as the name and other relevant details are completely new to the register, and while putting the death record beside the name the mismatch can not be overlooked. So let the outgoing migration be M_i realised similarly as $m_1, m_2, ..., m_{T_t}...$ i.e., the number of child left the study for that interval. Now the number of child actually exposed to risk on i^{th} day become $(r_{i-1} - d_{i-1} - m_{i-1})$. Here from the notion given by Cox and Oakes (1963) we can find the probability that any individual will survive the neonatal period if there be no migration i.e., none left the study during the period $T = 28$ is,

$$P(T > t \mid t \leq 28) = L_1(T) = \Pi_{i:t_i \leq 28}(1 - \frac{D_i}{R_{i-1} - D_{i-1}}) \quad (8)$$

Now, the Neonatal Period Mortality Rate without Migration(NPMRWOM) can be measured by,

$$Q_1 = 1 - L_1(T) = 1 - \Pi_{i:t_i \leq 28}(1 - \frac{D_i}{R_{i-1} - D_{i-1}}) \quad (9)$$

Similarly, if we consider m_i as the number of migration during the period the above expression (9) becomes,

$$P(T > t \mid t \leq 28) = L_2(T) = \Pi_{i:t_i \leq 28}(1 - \frac{D_i}{R_{i-1} - D_{i-1} - M_{i-1}}) \quad (10)$$

And the NPMRWM(i.e., with Migration) will be,

$$Q_2 = 1 - L_2(T) = 1 - \Pi_{i:t_i \leq 28}(1 - \frac{D_i}{R_{i-1} - D_{i-1} - M_{i-1}}) \quad (11)$$

Estimation

When the sample size is large and the data is grouped into some specific fixed interval of size M, $[t_i, t_{i+1}), i = 0, 1, ..., M-1$. As d_i is the observed number of deaths in the interval and r_i is the number of survivors at the begining of the interval we can write that,

$$r_i = \sum_{j=i}^{M-1} d_j \quad (12)$$

In particular,

$$r_0 = \sum_{j=0}^{M-1} d_j \quad (13)$$

is the total sample size, that is, the size of the group that starts at time t_0 and is followed over the period during which all members of the group die as per our assumption mentioned already in the above section. Now the probability of death can be estimated in the interval by,

$$\hat{q}_i = \frac{d_i}{r_i} \quad (14)$$

While finding the joint distribution of the numbers of deaths we can consider a multinomial variables as,

$$Pr[d_0, d_1, ..., d_{M-1}] = r_0! \prod_{i=0}^{M-1} = \frac{(p_i q_i)_i^d}{d_i!} \quad (15)$$

where, p_i = the probability of survivors. It follows that, $E(d_i) = r_0 p_i q_i$ and $Var(d_i) = r_0 p_i q_i (1 - p_i q_i)$. The is the very special case of complete mortality data. In particular, the relation $r_{i+1} = (r_i - d_i)$ holds for every i. More generally if we

consider the case of distribution of total observed number at the begining of the interval it can be claimed more accuracy according to Elandt-Johnson and Johnson, (1980). Conditional on r_i, the proportion of deaths in the interval, $\hat{q}_i = \frac{d_i}{r_i}$ is distributed as binomial proportion with parameters r_i, q_i. We the have, $E(\hat{q}_i \mid r_i) = q_i$ and $E(\hat{p}_i \mid r_i) = p_i$ along with $Var(\hat{q}_i \mid r_i) = Var(\hat{p}_i \mid r_i) = \frac{p_i q_i}{r_i}$. We must take a note that conditional on r_i the random variables, $\hat{q}_i$ are mutually independent. Similar note can be considered for the joint distribution of deaths. So that we can write,

$$\hat{L}_1(T) = \Pi_{i:t_i \leq 28}(1 - \hat{q}_i)$$
$$= \Pi_{i:t_i \leq 28}(1 - \frac{d_i}{r_{i-1} - d_{i-1}}) \quad (16)$$

and

$$\hat{L}_2(T) = \Pi_{i:t_i \leq 28}(1 - \hat{q}_i)$$
$$= \Pi_{i:t_i \leq 28}(1 - \frac{d_i}{r_{i-1} - d_{i-1} - m_{i-1}}) \quad (17)$$

So that, Q_1 and Q_2 can be obtained from the similar notion.

Few properties of the Estimator

As we can see from the above equation that,

1. When there is no migration, the results from (10) reduces to (8).

2. A summary estimate of the mortality experience of a given population is provided by the estimator. The matching standard error gives a little bit of insight into the accuracy of the estimate. For the construction of confidence interval (see, Klein and Moeschberger (1997)).

3. The estimator can also be used to provide estimates of quantiles of the distribution of the time-to-event distribution (see, Hall and Wellner (1980)).

4. With jumps at the observed event times t_i, the estimator is a step function. The magnitude of these leaps relies on the pattern of the censored observations made before each event time t_i as well as the number of events observed at each event time t_i (see, Kaplan and Meier (1958)).

5. The variance can be estimated from Greenwood's formula as,

$$\hat{V}(\hat{P}(T)) = \hat{P}(T)^2 \sum_{t_i \leq T} \frac{d_i}{r_i(r_i - d_i)}$$

On average (see, Greenwood (1926) and Klein (1991)), this estimator tends to come closest to the true variance and has a smaller variance except when r_i is very small.

6. The estimator is predicated on the premise that noninformative migration occurred, meaning that knowing a person's migrating period tells us nothing more about that person's chances of surviving had they continued the study.

7. For all time points smaller than the longest observed study time T_{max}, the survival function estimator is well defined. The projected survival curve is 0 beyond this point if the longest study time also corresponds to a death time (see, Efron (1967)). Because we don't know when the final survivor would have passed away if the survivor hadn't been censored, the value of $P(T)$ beyond the biggest time point cannot be calculated. There are several nonparametric theories that have been proposed to explain this ambiguity (see, Gill (1980)).

8. First, a reduced-sample method was used to build the estimator. In this method, it should be noted that since no information about events occurring at other times is available, $P(T)$ should be a step function with leaps only at the times t_i. At the different time points t_is, we can use a discrete distribution with mass to estimate (see, Klein and Moeschberger (1997)). The percentage of people who are at risk at time t_i but do not pass away at this time is how we can calculate the probability,

$$Pr[T > t_i | T \geq t_i] = \frac{r_i - d_i}{r_i}; \quad i = 1, ..., n \quad (18)$$

And hence $P(0) = 1$ and $P(\infty) = 0$. Here ∞ is impossible for human life, so that for our case we can put the value of t as 29 or 366 or as desired.

9. The estimator of the survival function at a time t is the fraction of observations which are larger than t if we had no migration. We intend to build our estimator similarly for migrated data by redefining the scoring function. So that the estimator can be claimed as self-consistent (see, Cox and Oakes (1963)).

10. Under certain regularity conditions, one can show that the estimator is nonparametric maximum likelihood estimators (see, Cheng Wang, M. (1987)and Prentice, R. L. (2014)).

Statistical Analysis

Data Sources

The fourth National Family Health Survey (NFHS-4) was conducted in India in 2014–2015. With 29 states, NFHS-4 for the first time, included all six of the union territories. It also offered estimates of the majority of indicators at the district level for all 640 districts in the nation as of the 2011 Census. The number of homes in the NFHS-4 sample was predicted to be around 568,200, an increase over the NFHS-3 sample size of nearly 109,000. This resulted in a final sample of 625,014 women and 93,065 males who were qualified for the interview. The study gathered data on 265,653 kids under the age of five living in these households. Data gathered utilising mini-notebook PCs and computer-assisted personal interviews (CAPI). For our purpose we have created a cohort of births on January 2015 and extracted required data along with some associated variable i.e., d for description.

Methods

In the first stage, a cohort of live births is created. As the last census of India happened in 2011 which implied that the next census is to be conducted in 2021, we took a mid-point of this decade and choose January of 2015 as our time-point. Any interested researcher can choose any time-period as desired. In the second stage, various descriptive statistics are used to show the various condition of the cohort with different socio-economic characteristics. These two stages of analysis is done using STATA 17.0 statistical package. In the third stage, the required result as per equation (12) is calculated through R software using the first and second stage data. Here we project the inter-state migration rate for the same time-period with the help of Migration Tables available in Census 2011 data. In table D1 of Census 2011 we get the number of people counted who born in states in India beyond the state of enumeration is 56,297,563 and the total population of India was 1,210,854,977. So that the nation rate of migration for any age becomes 4.65. As there is no avail-

ability of Migration data associated with Neonatal Mortality the National Migration Rate in general is applied uniformly all over the ages and all states. This may be a kind of limitation for this study but we can not help it. Also we assumed that the births are given in institutes whether it is private or public. Accordingly we put the migration rate after five days of giving birth because during this period mother and the neonate are under observation of the institute (see Table 2). Hoping that in near future data on migration of neonates be collected.

Simulation Study

Computer experiments called simulation studies use pseudo-random sampling to produce data. Because some "truth" (often some interesting parameter(s)) is known from the process of creating the data, understanding the behaviour of statistical approaches is one of the primary strengths of simulation studies. Here we regenerate the situation of real life through a computer program by using R-software. First, we have generated a sample for the total number of live births. Secondly we have generated the number of deaths per day basis from an Binomial Distribution. Thirdly, the total number of migration is generated through a Generalized Poisson Distribution. The study performed 100000 times for checking the convergence. The result is almost converge to the real-life value (see Table 3). The corresponding Histogram and Exposed to risk are shown on Figure 1, and 2 respectively.

Results and Discussion

Table 1 provided presents neonatal deaths categorized by the age at death, ranging from 0 to 27 days. With a total of 39 newborn fatalities, Uttar Pradesh had the highest number. The majority of deaths—14 at 0 days and 8 at 1 days—occurred within the first few days of life. With 15 neonatal deaths recorded from Bihar, a similar pattern can be seen, with the most deaths occurring at 0 days (9 deaths). 16 newborn fatalities, spread across a range of age groups, occurred in Madhya Pradesh. The distribution of newborn fatalities among various age groups was reported to vary in other regions as well. According to the research, a sizable percentage of neonatal deaths are thought to happen within the first few days of life. To improve newborn health and lower mortality rates, early neonatal care and interventions are crucial, as this illustrates.

As a baby's days get longer, the probability of newborn death gradually goes down (see Table 2). By day 0, there is a 4,669 percent danger, and by day 28, there is a 1,513 percent risk. On various days of life, there are varying numbers of neonatal deaths. Day 0 reports 56 newborn deaths, which is the largest number of fatalities. With no deaths reported on days 13 and 14, the number of fatalities gradually declines over time. Information on how many newborns migrated is also included in the statistics. The accuracy of newborn fatalities reported in a given area can be impacted by migration. On days 5 and 6, there were 211 and 201 migrations, respectively, which is interestingly a large number. This information can be utilised to pinpoint variables linked to newborn mortality at various time intervals. Researchers can look into possible causes for the high mortality rate at day 0 and look into how migration affects the correct reporting of newborn fatalities. The effectiveness of healthcare systems, socioeconomic conditions, and access to healthcare in various regions can be understood by analysing regional variations in neonatal mortality risk and deaths.

India reported a newborn mortality rate of 20 per 1,000 live births in 2020. This number was lower than it was in 2015

as 25 as per Sample Registration System (SRS). The numbers were still high, though, and they showed that the nation's healthcare system for pregnant women and new mothers has to be improved. Contrarily, the infant mortality rate was 41 deaths per 1,000 live births over the five years prior to the survey (NFHS 4), and the under-five mortality rate was 50. There were 30 fatalities per 1,000 live births, or neonatal mortality rate. In the five years prior to the 1992–1993 survey, there were 109 fatalities per 1,000 live births; in the five years prior to the 2015–16 study, there were 50 deaths per 1,000 live births. During the same time period, the infant mortality rate decreased from 79 deaths per 1,000 live births to 41 deaths per 1,000 live births. Uttar Pradesh has the highest baby and under-five death rates, while Kerala has the lowest. In comparison to urban regions, rural areas have significantly higher baby and under-five mortality rates. The Neonatal Mortality rate for 2015-16 is 29.5 as per the fact sheet for India provided by NFHS 5 (2019-21). And through the computation of proposed measure it is 32.74 per 1000 live births in January 2015. From Table 3 the comparison can be observed. The proposed measure of 32.74 deaths per 1,000 live births suggests a higher neonatal mortality rate than both the SRS and NFHS 4 data. This is not the case of overestimation. The higher number is anticipated to be more accurate than other values since the suggested metric takes death risk into account rather than using the total live births as the denominator. Due to the fact that children who have already passed away cannot in any way be linked to live-birth numbers up to 28 days. However, on the other hand, it is nearly impossible to locate those who moved to different locations (read: states) after birth. But contrary to reality, our traditional methodologies assume that the cohort is closed under migration. So that the neonatal mortality rate is lightly higher through the proposed measure but closer to greater accuracy along with the number of risk at death and the predicted rate of migration.

Acknowledgment

The authors gratefully acknowledge the comments of the referee and the editor, which led to an improved version of the article. I would like to sincerely thank my beloved friend, Shoummo Sen Gupta, for his invaluable help in retrieving data using various statistical software.

References

Alexander, M., & Alkema, L. (2018). Global estimation of neonatal mortality using a Bayesian hierarchical splines regression model. Demographic Research, 38, 335-372.

Ali, I., Jaleel, C. A., Maheshwari, N., & Rahman, H. (2019). Migration and maternal health care services utilisation in Uttar Pradesh, India. Social Science Spectrum, 4(3), 136-146.

Brockerhoff, M. (1990). Rural-to-urban migration and child survival in Senegal. Demography, 27(4), 601-616.

Brockerhoff, M. (1994). The impact of rural-urban migration on child survival. Health transition review, 127-149.

Brockerhoff, M. (1995). Child survival in big cities: the disadvantages of migrants. Social Science & Medicine, 40(10), 1371-1383.

Cox, D. R., & Oakes, D. (1984). Analysis of survival data Chapman and Hall. New York.

Curtin, P. D., & Curtin, P. D. A. (1989). Death by migration: Europe's encounter with the tropical world in the nineteenth century. Cambridge University Press.

Cheng Wang, M. (1987). Product limit estimates: a generalized maximum likelihood study. Communications in Statistics-Theory and Methods, 16(11), 3117-3132.

Dandona R, George S, Majumder M, Akbar M, Kumar GA. Stillbirth undercount in the sample registration system and national family health survey, India. Bull World Health Organ. 2023 Mar 1;101(3):191-201. doi: 10.2471/BLT.22.288906. Epub 2023 Jan 26. PMID: 36865608; PMCID: PMC9948498.

Duryea, S., Cordova, E. L., & Olmedo, A. (2005). Migrant remittances and infant mortality: Evidence from impact of migration. New Delhi: Orient Blackswan.

Efron, B. (1967, January). The two sample problem with censored data. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability (Vol. 4, No. University of California Press, Berkeley, CA, pp. 831-853).

Elandt-Johnson, R. C., & Johnson, N. L. (1980). Survival models and data analysis (Vol. 110). John Wiley & Sons.

Gill, R. D. (1980). Censoring and Stochastic Integrals, Mathematical Centre Tracts 124. Amsterdam: Mathematisch Centrum. Mathematical Reviews (MathSciNet): MR596815 Zentralblatt MATH, 456.

Greenwood, M. (1926). Reports on public health and medical subjects. The errors of sampling of the survivorship tables. London: HM Stationery Office.

Hall, W. J., & Wellner, J. A. (1980). Confidence bands for a survival curve from censored data. Biometrika, 67(1), 133-143.

Indian Institute for Population Sciences (IIPS) and MoHFW. National Family Health Survey - 4. 2017. Available from: http://rchiips.org/nfhs/pdf/NFHS4/India.pdf.AccessedApril25,2023

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American statistical association, 53(282), 457-481.

Klein, J. P., & Moeschberger, M. L. (1997). Statistics for biology and health. Stat. Biol. Health, New York, 27238.

McCutcheon, J. J. (1973). On the measurement of infant mortality. Transactions of the Faculty of Actuaries, 34, 433-442.

Klein, J. P. (1991). Small sample moments of some estimators of the variance of the Kaplan-Meier and Nelson-Aalen estimators. Scandinavian Journal of Statistics, 333-340.

Kanaiaupuni, S. M., & Donato, K. M. (1999). Migradollars and mortality: The effects of migration on infant survival in Mexico. Demography, 36(3), 339-353.

Mexico. Washington: Inter-American Development Bank. Mimeo.

Million Death Study Collaborators. (2010). Causes of neonatal and child mortality in India: a nationally representative mortality survey. The Lancet, 376(9755), 1853-1860.

Oza, S., Cousens, S. N., & Lawn, J. E. (2014). Estimation of daily risk of neona-

tal death, including the day of birth, in 186 countries in 2013: a vital-registration and modelling-based study. The Lancet Global health, 2(11), e635-e644.

Prentice, R. L. (2014). Self-consistent nonparametric maximum likelihood estimator of the bivariate survivor function. Biometrika, 101(3), 505-518.

Shryock, H. S., Siegel, J. S., & Larmon, E. A. (1973). The methods and materials of demography (Vol. 2). US Bureau of the Census.

Zachariah, K. C., & Irudya Rajan, S. (2012). Kerala's Gulf connection 1998-2011: Economic and social

Appendix

Table 1. Age at death in days among cohort of January 2015

States	Age at death of Neonates (Days)																		
	0	1	2	3	4	5	6	7	8	9	10	11	12	15	20	21	22	27	Total
Arunachal Pradesh	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
Assam	2	3	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	7
Bihar	9	1	3	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	15
Chhattisgarh	0	3	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	6
Gujarat	5	1	2	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	10
Haryana	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
Himachal Pradesh	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3
Jammu and Kashmir	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Jharkhand	3	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	8
Karnataka	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Madhya Pradesh	7	4	1	0	0	0	0	0	2	1	0	0	0	1	0	0	0	0	16
Maharashtra	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Manipur	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Meghalaya	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Mizoram	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Nagaland	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Odisha	1	3	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
Punjab	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Rajasthan	4	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	7
Sikkim	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Tamil Nadu	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Tripura	1	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	3
Uttar Pradesh	14	8	2	3	2	1	2	1	0	0	1	2	2	0	0	0	0	0	39
Uttarakhand	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	3
West Bengal	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Total	56	30	14	14	4	3	3	1	3	2	4	2	4	2	1	1	0	1	145

Table 2. Number of Live Births, Risk of deaths, Number of Deaths and Number Migrated among cohort of January 2015

Days	Risk of death(r)	Number of Deaths(d)	Migrated(m)
0	4669	56	0
1	4613	30	0
2	4583	14	0
3	4569	14	0
4	4555	4	0
5	4551	3	211
6	4337	3	201
7	4133	1	192
8	3940	3	183
9	3754	2	174
10	3578	4	166
11	3408	2	158
12	3248	4	151
13	3093	0	144
14	2950	0	137
15	2813	2	131
16	2680	0	124
17	2556	0	119
18	2437	0	113
19	2324	0	108
20	2216	1	103
21	2112	1	98
22	2013	0	93
23	1920	0	89
24	1831	0	85
25	1746	0	81
26	1665	0	77
27	1588	1	74
28	1513	0	70

Table 3. Neonatal Mortality Rate of India in 2015 by various Reports

Report	Neonatal Mortality Rate per 1000 live births
SRS	25
NFHS 4	29.5
Proposed Measure	32.74
Simulation Study	33.33

Figure 1. Histogram(Simulated)

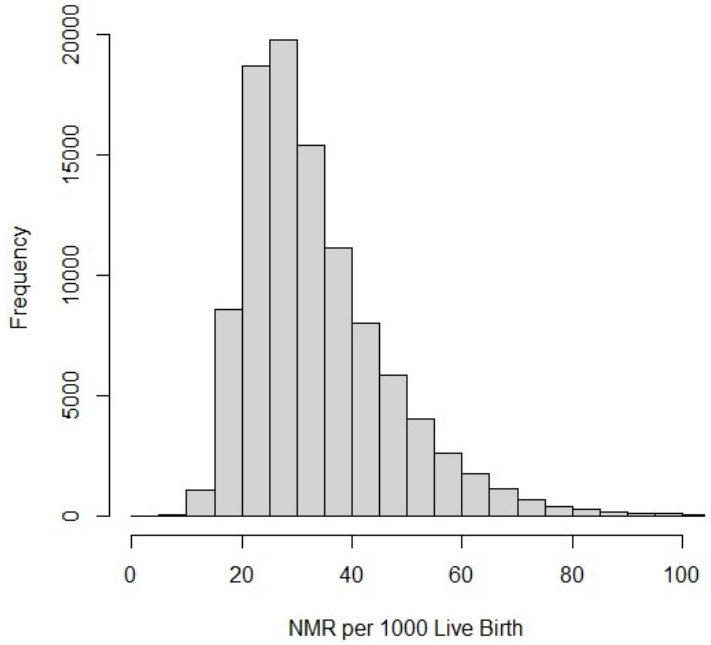

Figure 2. Exposed of Risks(Simulated)