*Demography India Vol. 28, No. I (1999),* pp. 13-21

**S. Krishnamoorthy* and P. N. Rajna***

# Graduating Childhood Deaths and Survivorship Function with the Model by Mukherjee and Islam

## Introduction

THE force of mortality is relatively high at early ages of life than at the teenages. Intervention to reduce mortality at early ages is found to be easier and effective. The reduction in wastages of life at these ages is important on grounds of economic benefit to the society. Hence, both mathematical and applied study of infant and childhood mortality were extensive. Graduating the survival function at these young ages have been attempted by many. Use of hyperbolic function (Keyfitz, 1977), logarithmic function (Hartman, 1982) and Weibull function (Choe, 1981) are few such attempts. Attracted by the ability of Weibull in effectively describing the age pattern of mortality at early ages, Krishnamoorthy (1982), Pathak *et al*. (1991), Krishnamoorthy and Mathew (1994) and Mathew and Krishnamoorthy (1995) have applied Weibull function in their recent works. Recently, Chauhan (1997) suggested use of a model proposed by Mukherjee and Islam in reliability analysis. Chauhan concludes that this model effectively describes the age distribution of early age deaths. This paper re-examines the adequacy of the model in describing the distribution of deaths at early ages of life and also modifies the model for graduating the survivorship function.

## The Model and Application

The density due to Mukherjee and Islam (1983) with parameter p is given by

*Department of Population Studies. Bharathiar University, Coimbatore 641 046.

$$f(x) = \begin{cases} \frac{p}{x}\left(\frac{x}{\theta}\right)^p, & 0 < x < \theta \\ 0, & \text{otherwise} \end{cases}$$

The corresponding distribution function is

$$F(x) = \begin{cases} 0, & x \le \theta \\ \left(\frac{x}{\theta}\right)^p, & 0 < x \le \theta \\ 1, & x > \theta \end{cases}$$

From this $F(1) = \left(\frac{1}{\theta}\right)^p$, which yields

$$P = \frac{-\ln F(1)}{\ln \theta}$$

Assume a cohort of $N$ births followed for a period of θ months. Let $n$ deaths are observed over the period and $n_1$ of them in the first month. Chauhan suggests to estimate $p$ by

$$\hat{p}_1 = \frac{-\ln(n_1/n)}{\ln \theta}$$

and obtain the graduated density of deaths in the period of observation (0, θ). In this approach there is a possibility of error in $n_1$ due to the preference of many people reporting deaths occurred just below one month as one month. The other alternate approaches are using either median or mean to estimate $p$.

Given the median $M$, the estimate of $p$ is

$$\hat{p}_M = \frac{-\ln(2)}{\ln(M/\theta)}$$

Similarly, as the mean is

$$\mu = \frac{p}{p+1}\theta,$$

the estimate of $p$ from the observed mean $\overline{X}$ is

$$\hat{p}_{\bar{X}} = \frac{\overline{X}}{\theta - \overline{X}}$$

All the three estimates of $p$ are obtained for the data from National Family Health Survey (NFHS) of Uttar Pradesh which provides the information on the births that occurred in the household during the previous 61 months and deaths among them. As the survey data truncate the observations, we took only those births which took place 24 months before the date of interview and obtained the distribution of deaths in the first 24 months to avoid truncate due to survey date.

To avoid heaping error entering the analysis, $\theta$ is set as 22 months. The deaths distributed during the first 22 months is extremely skewed to the extent that even within class intervals the assumption of uniform distribution is not valid. Hence, an iterative procedure is developed to obtain a better estimate of the mean and the parameter, $p$ (see Appendix).

TABLE 1 : OBSERVED AND EXPECTED DISTRIBUTION OF DEATHS FOR UTTAR PRADESH DATA FROM NFHS

| *Month* | *Observed No. of Deaths* | *Estimated Frequencies when p is* | | |
|---|---|---|---|---|
| | | $\hat{p}_1$ *0.194191* | $\hat{p}_m$ *0.185167* | $\hat{p}_{\bar{x}}$ *0.201261* |
| 0 | 310 | 310.00 | 318.77 | 303.30 |
| 1 | 33 | 44.67 | 43.66 | 45.41 |
| 2 | 26 | 29.05 | 28.26 | 29.65 |
| 3 | 16 | 22.05 | 21.38 | 22.55 |
| 4 | 14 | 17.97 | 17.38 | 18.42 |
| 5 | 16 | 15.27 | 14.75 | 15.67 |
| 6 | 24 | 13.34 | 12.86 | 13.71 |
| 7 | 9 | 11.88 | 11.44 | 12.22 |
| 8 | 15 | 10.74 | 10.33 | 11.06 |
| 9 | 14 | 9.82 | 9.43 | 10.12 |
| 10 | 21 | 9.06 | 8.69 | 9.34 |
| 11 | 11 | 8.42 | 8.07 | 8.68 |
| 12 | 16 | 7.87 | 7.54 | 8.12 |
| 13 | 7 | 7.39 | 7.08 | 7.64 |
| 14 | 3 | 6.98 | 6.68 | 7.21 |
| 15 | 3 | 6.61 | 6.33 | 6.84 |
| 16 | 4 | 6.29 | 6.01 | 6.51 |
| 17 | 2 | 5.99 | 5.73 | 6.21 |
| 18 | 17 | 5.74 | 5.48 | 5.94 |
| 19 | 1 | 5.50 | 5.25 | 5.69 |
| 20 | 2 | 5.28 | 5.04 | 5.47 |
| 21 | 1 | 5.08 | 4.85 | 5.27 |
| | 565 | 565.00 | 565.01 | 565.03 |

*Source:* National Family Health Survey, Uttar Pradesh, 1992.

Table 1 provides the observed frequencies of deaths in the first 22 months and the estimated frequencies based on the parameter $p$ estimated by the three procedures. All the three methods yield reasonably good graduated distribution. It is very difficult to judge which one of these is the best as the observed distribution contains lot of errors due to time preference in reporting of events.

Our suggestion is to use the estimate of $p$ from the mean as the mean depends on all observations. Further, though there would be large gross error due to mistreatment of age at death in less developed countries, the net error in the estimate of mean is likely to be small if over reporting and under reporting of age cancel out substantially.

## Modification and Application to Survivorship Function

Let us assume a cohort of $l_0$ births are followed up to a period of $\theta$ and the number of survivors to age $\theta$ is $l_\theta$. Now $l_0 - l_\theta$, the number of deaths within age $\theta$, is distributed according to the density proposed by Mukherjee and Islam. The number surviving at least up to age $x$ is

$$l_x = l_0 - (l_0 - l_\theta)\, F(x)$$

$$= l_0 - (l_0 - l_\theta) \left(\frac{x}{\theta}\right)^p$$

The force of mortality is,

$$\mu_x = \frac{-dl_x}{l_x d_x}$$

$$= \frac{(l_0 - l_\theta)\, p x^{p-1}}{l_0 \theta^p - (l_0 - l_\theta) x^p}$$

The person years lived in the age range $(x, x + n)$ is

$$nL_x = \int_0^n l(x+t)dt$$

$$= nl_0 - \frac{(l_0 - l_0)}{(p+1)\theta^p}\left[(x+n)^{p+1} - x^{p+1}\right]$$

It is now clear from the above that the model due to Mukherjee and Islam can be used not only to graduate the deaths in the early ages but also to the survivorship function once $l_0$, $l_0$ and $p$ are estimated. If $l_0$ is set as 1, the only parameters required to graduate the survivorship function are $l_0$ and $p$.

When $l_0 = 1$.

$$l_x = 1 - (1 - l_\theta)\left(\frac{x}{\theta}\right)^p$$

After slight modification and taking logarithm

$$\ln(1 - l_x) = \ln(1 - l_\theta) + p\ln\left(\frac{x}{\theta}\right)$$

This may be written as

$$y = A + pX \text{ where } y = \ln(1 - l_x),\ A = \ln(1 - l_\theta) \text{ and } X = \ln\left(\frac{x}{\theta}\right)$$

Given observed $lx$ for different values of $x$ the above equation may be solved for $A$ and $p$ by the method of least squares. From this $l_\theta$ is also estimated by

$$l_\theta = 1 - e^{\hat{A}}$$

For the data on child survival on births in 1915–24 in Sweden which was used by Chauhan to show how Mukherjee and Islam model fits, the above model is fit. The results are given in Table 2. The fit is extremely good. The difference between the predicted and the observed proportion of survivors at various months of age is quite small. Hence, we conclude that Mukherjee and Islam model is good not only to graduate the deaths at early ages but also the survivorship function.

TABLE 2 : THE FIT OF THE MODEL DUE TO MUKHERJEE AND ISLAM FOR THE SWEDISH DATA

| *Age in exact month* | *Observed proportion of survivors*[a] | *Predicted proportion of survivors* | *Predicted - Observed* | |
|---|---|---|---|---|
| 0 | 1.00000 | 1.00000 | | |
| 1 | .97370 | .97383 | 13 | |
| 2 | .96620 | .96611 | –9 | |
| 3 | .96040 | .96057 | 17 | |
| 4 | .95610 | .95610 | 0 | $\hat{p} = .37323$ |
| 5 | .95250 | .95229 | –21 | |
| 6 | .94930 | .94893 | –37 | |
| 7 | .94630 | .94590 | –40 | |
| 8 | .94340 | .94314 | –26 | $\hat{l}_{12} = .93385$ |
| 9 | .94060 | .94058 | –2 | |
| 10 | .93800 | .93820 | 20 | |
| 11 | .93560 | .93596 | 36 | $R$-square = .99964 |
| 12 | .93330 | .93385 | 55 | |

[a]*Source*: Swedish data for 1915-24 is drawn from Chauhan (1997).

Naturally, the question is whether the proposed model is better than the earlier ones already available. Krishnamoorthy (1982) tried to compare the models proposed by

Keyfitz. Hartman and Choe. In the Table 3 we present earlier findings of Krishnamoorthy and our results on the application of the present model to the same set of data. Based on the differences between observed and expected frequencies, we may conclude that the present model fits much better than the Hyperbolic function proposed by Keyfitz (1966), probably even better than the Logarithmic function proposed by Hartman (1982) and probably almost as good as but not so good as the Weibull function proposed by Choe (1981). While Weibull survivorship function is not easily integrable, the present function is easily integrable which is definitely an advantage in computing person years survived in small intervals of age with greater accuracy and in incorporating this in larger models.

TABLE 3 : OBSERVED AND GRADUATED SURVIVORSHIP OF AUSTRALIAN MALE CHILDREN" USING HYPERBOLIC, LOGARITHMIC. WEIBULL, MUKHERJEE AND ISLAM MODELS

| *Age* | *Observed survivorship*[1'] | *Differences between Observed and Graduated survivorship* | | | |
|---|---|---|---|---|---|
| | | *Hyperbolic function* | *Logarithmic function* | *Weibull function* | *Mukherjee & Islam function*[^rd] |
| Days | | | | | |
| 0 | 100,000 | 0 | 0 | 0 | 0 |
| 28 | 99.175 | 494 | -41 | -41 | 97 |
| Months | | | | | |
| 2 | 99,083 | 333 | -45 | -45 | - 107 |
| 3 | 99,005 | 249 | -22 | -22 | -76 |
| 4 | 98.942 | 193 | 1 | 1 | -48 |
| 5 | 98.883 | 160 | 27 | 27 | - 16 |
| 6 | 98,848 | 122 | 35 | 35 | -5 |
| 7 | 98.822 | 88 | 37 | 37 | 1 |
| 8 | 98.807 | 54 | 32 | 32 | - 1 |
| 9 | 98,792 | 27 | 28 | 28 | -2 |
| 10 | 98.778 | 6 | 25 | 25 | -2 |
| 11 | 98,761 | -8 | 26 | 26 | 2 |
| Years | | | | | |
| 1 | 98.734 | - 17 | 30 | 30 | 8 |
| 2 | 98,649 | -94 | 3 | 3 | 2 |
| 3 | 98,594 | - 106 | - 18 | -18 | -6 |
| 4 | 98,539 | -87 | -20 | -20 | 4 |
| 5 | 98,488 | -59 | - 14 | - 14 | 18 |
| 6 | 98.459 | -45 | -23 | -23 | 16 |
| 7 | 98,425 | -22 | -22 | -22 | 24 |
| 8 | 98.389 | 5 | - 15 | - 15 | 37 |
| 9 | 98,362 | 25 | - 14 | - 14 | 43 |
| 10 | 98.337 | 45 | -13 | -13 | 49 |

a. *Source:* Australian Bureau of Statistics. 1981 a. Deaths Australia 1979.
b. Survivorships are given per 100,000 births for the sake of convenience.
c. Except the last column, others are from Krishnamoorthy (1982).
d. Mukherjee and Islam model fit estimates are $p$ = 0.11118, $i_{12}$ = 0.987507 and R-square = 0.95047.

## Conclusion

The model proposed by Mukherjee and Islam in the study of reliability and introduced by Chauhan to graduate infant deaths in tested for its efficacy in describing, not on\v the age pattern of childhood deaths but also the survivorship probabilities. It is found that the model is quite satisfactory in both the situations.

## References

Choc. M. K.. 1981. Fitting the ;igc patterns of infant and child mortality with the Weibull Survival Function, *Asian and I'acijic Census f'orum,* 7(4): 10-13.

Chauhan. Rajesh Kumar. 1997. Graduation of Infant Deaths by Age. *Demography India,* 26(2): 261-274. Hartirian. M.. 1980. *Infant and Child Mortality.* Dept. of Statistics. University of Lund. Sweden. Keyfitz. N.. 1977. *Introduction to llie Mathematics of Population with Revisions.* Addison-Wesley Publishing Company. Reading. Massachusetts. Krishnamoorthy. S.. 1982. On Graduating Infant and Child Mortality with Mathematical Curves. *Asian and Pacific Census Forum.* 8(3): 14-15. KrishnamoortlTi. S. and Mathew. Anil C.. 1994. Estimating the Person Years Lived in Infant and Childhood Ages. *Janasamkhya.* 12(1&2): 151-157. Mathew. A. C.. and Krishnamoorthy. S.. 1995, Refined Estimates of Person Years at Infant and Childhood Ages of Model Life Tables. *Contributions to Applied and Mathematical Statistics,* **1:** 67-76. Mukherjee. S. P. and Islam. 1983. A finite range distribution of failure time. *Naval Research Logistic Quarterly.* 30: 487-491. Pathak. K. R.. Pandcy. A. and Mishra. U. S.. 1991, Estimating Current Levels of Fertility and Child Mortality from Data on Open Birth Interval and Survival Status of the Last Child. *Janasamkhya,* **15:** 15-23.

## Conclusion

The model proposed by Mukherjee and Islam in the study of reliability and introduced by Chauhan to graduate infant deaths in tested for its efficacy in describing not only the age pattern of childhood deaths but also the survivorship probabilities. It is found that the model is quite satisfactory in both the situations.

## References

Choc. M. K.. 1981. Fitting the age patterns of infant and child mortality with the Weibull Survival Function. *Asian and Pacific Census l-'orum,* 7(4): 10-13.

Chauhan. Rajesh Kumar. 1997. Graduation of Infant Deaths by Age. *Demography India.* 26(2): 261-274. Hartman. M.. 1980. *Infant and Cliiid Mortality.* Dept. of Statistics. University of Lund. Sweden. Kcyfitz. N.. 1977. *Introduction to llie .Mathematics of Population with Revisions.* Addison-Wesley Publishing Company. Reading. Massachusetts. Krishnamoorthy. S.. 1982. On Graduating Infant and Child Mortality with Mathematical Curvts. *Asian and Pacific Census Forum.* 8(3): 14-15. . ' Krishnamoorthy. S and Mathew. Anil C.. 1994. Estimating the Person Years Lived in Infant and Childhood Ages. *Janiisamkhya.* 12(1&2): 151-157. Mathew. A. C.. and Krishnamoorthy. S.. 1995. Refined Estimates of Person Years at Infant and Childhood Ages of Model Life Tables. *Contributions to Applied and Mathematical Statistics,* **1:** 67-76. Mukherjee. S. P. and Islam. 1983. A finite range distribution of failure time. *Naval Research Logistic Quarterly.* 30: 487-491. Pathak, K. *V,*.. Pandey. A. and Mishra. U. S.. 1991, Estimating Current Levels of Fertility and Child Mortality from Data on Open Birth Interval and Survival Status of the Last Child. *Janasamkhya,* **15:** 15-23.

## Appendix

Number of deaths at early ages of childhood is extremely skewed. Mean computed from the observed data is likely to contain certain error if the mid of the class interval is taken for the computation of mean as the assumption is uniform distribution of deaths within the class interval is not appropriate. In many other situations positive and negative biases may cancel out to a large extent. But in the case of distribution of deaths at early ages it does not, as the number of deaths are monotonically declining. To reduce such a bias in computing the mean age at death we resort to the following procedure.

First we compute the mean using the mid point of the class interval and treat this as our first approximation. That is

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i},$$

where $x_i$ is the mid point of the class interval and $f_i$ is the corresponding frequency, we assume that the model proposed by Mukherjee and Islam is applicable to data. We then estimate the parameter $p$ from

$$\hat{p} = \frac{\bar{x}}{\theta - \bar{x}}$$

The theoretical density is now known as

$$f(x) = \frac{\hat{p}}{\theta^{\hat{p}}} x^{\hat{p}-1}$$

Then the mean of the density within the class interval $(t, t+n)$ is

$$x_i^* = \frac{\int_t^{t+n} x f(x) dx}{\int_t^{t+n} f(x) dx}$$

Substituting the theoretical density function and evaluating, we get

$$x_i^* = \frac{\hat{p}\left[(t+n)^{\hat{p}+1} - t^{\hat{p}+1}\right]}{(\hat{p}+1)\left[(t+n)^{\hat{p}} - t^{\hat{p}}\right]}$$

Now the improved mean of the density is obtained from

$$\bar{x}^* = \frac{\sum x_i^* f_i}{\sum f_i},$$

The parameter estimate is revised using

$$\hat{p} = \frac{\bar{x}^*}{\theta - \bar{x}^*}$$

This process is repeated until the estimate of $p$ converges. For the data on Uttar Pradesh described in the body of the paper it took hardly three iterations to converge to a reasonable level of accuracy. Indeed the results of the fifth and sixth iterations did not differ even at the sixth decimal place.

The first approximation of the mean was 0.203387 and the parameter $p$ was 0.533296. After using the above iterative procedure we obtained improved estimate of mean as 0.201261 and of the parameter $p$ as 0.536813.