A Methodology for Studying Child Mortality Differentials in Populations with Limited Death Registration.

Claire Noël-Miller
Douglas Ewbank

Indirect methods of estimation remain an important tool for gaining an understanding of child mortality conditions in areas with limited, deficient or non-existent vital registration systems. The now commonly used Brass method estimates mortality from reports of children ever born and children deceased organized by duration of their exposure to the risk of death, as approximated for instance by the age of the mother (Brass 1975; Brass and Coale 1968; United Nations 1983). Central to the Brass child mortality estimation is the idea that the proportions of children dead to women aged 15-19, 20-24, 25-29, 30-34, 35-39, 40-44 and 45-49 approximate the life table values of q(1), q(2), q(3), q(5), q(10), q(15) and q(20) respectively. The basic Brass equation:

$$q(x) = k_i \times D_i \tag{1.1}$$

uses age group-specific multipliers (commonly known as "k" coefficients) to translate data on the reported proportion deceased in a given age group i, D_i, into life table estimates of the proportions deceased by various ages, q(x).

An important drawback of the Brass method lies in its inability to allow for the formal test of differentials in mortality across groups since the mortality schedule underlying the calculation of k_i is typically considered to be universal.

In an attempt to address this limitation, Preston and Trussell introduced a multivariate approach to studying mortality differentials from data on children ever born and children surviving (Trussell and Preston 1984; Trussell and Preston 1982). Despite some important statistical problems related to its intrinsic assumption of linearity between covariates of mortality and mortality outcomes (Hill 1989), this method now constitutes the cornerstone of analyses of child mortality differentials for many populations.

The Preston-Trussell method usually relies on published tables of multipliers (Brass 1975; Sullivan 1972; Trussell 1975) themselves based on an index of the shape of the fertility schedule and on a standard life table (Coale and Demeny 1983). In practice, published tables of multipliers can be of limited use when the actual age pattern of mortality differs substantially from the pattern underlying the calculation of the multipliers.

We propose an alternative to the Preston-Trussell method for the multivariate analysis of associations between child mortality as measured by "Brass questions" and various characteristics of the environment into which the children are born. We apply the method to data from the 1993 Gambian census and estimate the effect of various covariates of mortality via maximum likelihood. Particular attention is paid to the effects of education and rural or urban locations. Our model assumes a dichotomous dependent variable recording failures (survivorships in our case) and successes (deaths in our case) following a given number of trials (children ever born). Our approach is based on a skewed logistic regression (also referred to as a scobit model), an elaboration upon the standard logit model. The scobit model relaxes the assumption that individuals with initial probability $P_i = .5$ of experiencing a success are most sensitive to changes in the independent variable (Nagler 1994). The analysis offers a second innovation by relying on multipliers more carefully tailored to local mortality conditions. In lieu of the North model variant of the regional model life tables (Coale and Demeny 1983), typically assumed for African countries, the calculation of the k_i coefficients is based on a life table more closely reflecting the mortality experience of the Gambia. In addition, greater accuracy is obtained by taking into account possible variations in the fertility schedule across specific sub-samples of the population. We examine how well the method performs in small samples relative to the Preston-Trussell method and to the logit model.

Similarly to the logit model, skewed logistic regression belongs to a class of models known as generalized linear models (GLM), which move beyond ordinary regression models to accommodate non-normal response distributions and modeling functions of the mean (Agresti 2002; Dobson 2001; Hardin and Hible 2001; McCulloch and Searle 2001). The outcome variable, i.e. the death or the survival of a child, is assumed to result from a *fixed* number of trials (the number of children ever born) and to follow a binomial distribution[1]. The link function of the scobit model is the logit or the log odds of the probability of experiencing a success.

The logit and the scobit models differ in that the former assumes that individuals with a probability of success equal to 0.5 are most sensitive to changes in the independent variables. Then, should women with probability of experiencing a child death different than 0.5 be in reality most sensitive to changes in the specified covariates of mortality, the logit model would be misspecified. Rather than imposing an unrealistically high value at which individuals' probability of experiencing a child death is most sensitive to stimulus, our model estimates an additional parameter for the maximum point of sensitivity.

[1] A Poisson distribution for the response variable has been proposed as an alternative to the normal distribution assumed in the Preston-Trussel method (see for instance MacLeod 1999). The Poisson distribution is attractive because it is particularly suited for non-negative integers representing counts of a rare occurrence following a number of events. However, the Poisson distribution assumes that the count data result from a small risk in a large population (i.e., number of trials). This is not the case when considering the number of children ever born. Thus, the Poisson distribution is appropriately used when there is no upper limit to n, the number of trials and for counts of events that occur randomly either in time or in space. In some cases, when n is very large and when the probability of success following each trial ,π, is very small, the Poisson distribution can be applied as an approximation for the binomial. This is however not the case in our context.

The skewed logistic model presented in the paper is given by:

$$\gamma_{\mu_i} = \beta_0 + \beta_1 x_{i1} + \beta_i x_{i2} + ... + \beta_k x_{ik} + \gamma_{EPD_i} \tag{1.2}$$

, where γ_{μ_i} represents the logit of μ_i, the probability of experiencing a child death, β_0 is the constant term, the x_i represent covariates of mortality, and γ_{EPD_i} represents the logit of the expected proportion deceased (EPD).

The values of the $\{EPD_i\}$ used in equation (1.2) were obtained by turning equation (1.1) around so that:

$$EPD_i = \frac{q^S(A_i)}{k_i} \tag{1.3}$$

where, $q^S(X)$ is a standard proportion of children who died by exact age X and k_i is derived to more closely reflect age patterns of mortality and fertility in the Gambia. Figures for the standard mortality schedule (i.e. the $q^S(X)$ values in equation (1.3)) were obtained from a single year of age life table of the Mlomp area in Sénégal, covering the period 1985-1999 for both males and females (Pison, Gabadinho, and Enel 2001). The Mlomp Demographic Surveillance Survey (DSS) site is a rural, mostly rice cultivating area located in Southwest Sénégal, near the border between Sénégal and Guinea-Bissau (for more details see INDEPTH 2002). Because of its relative proximity to the Gambia, use of the Mlomp data insures a similar pattern of mortality to that found in the Gambia. However, the level of mortality in the Mlomp data was adjusted for that in the Gambia.

The scobit model is also estimated with values of the EPD_i derived a) by single year of age and b) for different sub-groups of the population. The latter is done by assuming not one single fertility schedule for the entire population, but rather different fertility schedules for different sub-groups of the population defined according to the various combinations of literacy levels and rural or urban locations.

The paper compares the performances of the skewed logistic regression, the logit model and the Preston-Trussell method in estimating covariates of mortality based on small sample sizes. We run Monte Carlo simulations of the three methodologies applied to 1,000 samples of size N=1,000 and 1,000 with N=5,000 randomly drawn from the entire Gambia census. Results of the simulations are compared to the estimates obtained with the scobit model applied to the entire census. In particular, the Monte Carlo simulations examine the sensitivity of the estimated effects of education and rural or urban locations. Preliminary results suggest that in samples of 1,000 both the scobit and the logit coefficients are biased upwards, with the logit coefficients displaying a comparatively smaller root mean squared error. In samples of N=5,000 the bias is reduced in the scobit model estimates and their variance around the true estimate is very small. These results seem to indicate that the logit model remains preferable in small samples while scobit performs better in relatively larger samples.

Reference List

1. Agresti, A. 2002. *Categorical Data Analysis*. Hoboken, N.J: John Wiley & Sons, Inc.

2. Brass, W. 1975. *Methods for Estimating Fertility and Mortality From Limited and Defective Databased on Seminars Held 16-24 September 1971 at the Centro Latinoamericano De Demografía (CELADE) San José, Costa Rica* . Chapel Hill, North Carolina: Carolina Population Center, Laboratories for Population Statistics.

3. Brass, W. and A. J. Coale. 1968. "Methods of Analysis and Estimation." Pp. 88-139 in *The Demography of Tropical Africa*, editor W. Brass. Princeton, N.J.: Princeton University Press.

4. Coale, A. J. and P. Demeny. 1983. *Regional Model Life Tables and Stable Populations*. New York: Academic Press.

5. Dobson, A. J. 2001. *An Introduction to Generalized Linear Models*. Second ed. Chapman and Hall/CRC.

6. Hardin, J. and J. Hible. 2001. *Generalized Linear Models and Extensions*. College Station, Texas: Stata Press.

7. Hill, K. 1989. "Socio-Economic Differentials in Child Mortality: the Case of Jordan." Pp. 91-124 in *Infant and Childhood Mortality in Western Asia.* United Nations. Economic and Social Commission for Western Asia.

8. INDEPTH. 2002. *Population and Health in Developing Countries*, vol. 1. Population, health and survival at INDEPTH sites. Ottawa: Canada: International Development Research Center.

9. MacLeod, W. B. 1999. "The Decline of Child Mortality in the Gambia." Ph.D. Dissertation, Harvard University , Cambridge, Massachusetts.

10. McCulloch, C. and S. Searle. 2001. *Generalized, Linear and Mixed Models*. New York: Wiley.

11. Nagler, J. 1994. "Scobit: An Alternative Estimator to Logit and Probit. " *American Journal of Political Science* 38:230-255.

12. Pison, G., A. Gabadinho, and C. Enel. 2001. *Mlomp (Sénégal): Niveaux Et Tendances Démographiques 1985- 2000*. Dossiers et recherches No. 103. Institut National d'Etudes Démographiques.

13. Sullivan, J. M. 1972. "Models for the Estimation of the Probability of Dying Between Birth and Exact Ages of Early Childhood." *Population Studies* 26(1):79-97.

14. Trussell, J. 1975. "A Re-Estimation of the Multiplying Factors for the Brass Technique for Determining Childhood Survivorship Rates." *Population Studies* 29(1):97-107.

15. Trussell, J. and S. H. Preston. 1984. "Estimating the Covariates of Childhood Mortality From Retrospective Reports of Mothers." *Methodologies for the Collection and Analysis of Mortality Data*, editors Vallin J., J. H. Pollard, and L. Heligman. Liege, Belgium: Ordina Editions.

16. Trussell, T. and S. H. Preston. 1982. "Estimating the Covariates of Childhood Mortality From Retrospective Reports of Mothers." *Health Policy and Education* 3(1):1-36.

17. United Nations. 1983. "Indirect Techniques for Demographic Estimation." Pp. 73-81 in New York: United Nations.