Curr Pediatr Res 2012; 16 (1): 19-25

Child Survival Model in Reference to First Birth Order: Statistical Validation and Prediction.

Rajvir Singh[1], Abdulbari Bener[2, 3], Kalpana Singh[4], S.N. Dwivedi[5]

[1] Medical Research Centre, Hamad Medical Corporation, Doha, Qatar
[2] Dept. of Medical Statistics & Epidemiology, Hamad Medical Corporation, and Depts. of Public Health, Weill Cornell Medical College, **Qatar**
[3] Dept. Evidence for Population Health Unit, School of Epidemiology and Health Sciences, The University of Manchester, Manchester, UK
[4] Max Super Specialty Hospital, East Wing, Saket, New Delhi-110017, India
5. Department of Biostatistics, All India Institute of Medical Sciences (AIIMS), New Delhi-110029, India

Abstract

Child mortality is very high in India and it varies from state to state. Development of child survival models are lacking for these states. Hence, there is an immense importance and need to develop these models to design better and efficient health care systems and managements to prevent child mortality in India. The main aim of this study was to develop a child survival model to validate and predict survival probabilities for future perspectives. A sample of 627 children for the first birth order born during the four years preceding the survey was included in this study. The complete details of the children have been taken from the National Family Health Survey (NFHS-1992) for ever married women aged 13-49 years of age. Multivariate Cox PH analysis revealed that breastfeeding and immunization were the significant protective factors for child survival. Two hundred random samples with replacement from original sample were taken for the validation of the developed model. Shrinkage coefficient and Somer's D_{xy} rank correlation were 78% and -0.89 having only 22% noise in the model. Validation Indices were found good enough for internal validity and the resulted model is found to be adequate to get accurate predictive survival probabilities. Therefore, this proposed model may be used by health policy planners for better child health care management in the state.

Key words: Bootstrap, Validation, Shrinkage's Coefficient, Somer's D_{xy} rank Correlation.

Accepted November 20 2011

Introduction

An ideal way of model validation is a verification of the developed model to a new data set, which is not a feasible situation most of the time. To overcome this problem, method of internal validation has been recommended [1]. Generally data splitting, cross validation and bootstrapping techniques are used for internal validation. Validation of a model is necessary for its predictive accuracy which quantifies the utility of a developed model to be used for prediction as well as to check over fitting and lack of fit [2,3]. Bootstrapping method i.e. re-sampling method invented by Bradley Efron [4,5,6,7] and further developed by Efron and Tibshirani[8] takes over less bias and more consistent results than the others [2,7,9]. This method is also preferred for internal validation as entire data set is used in re-sampling process and involves a large number of samples with replacement from the original sample.

In developing countries like India the validation and prediction of child survival probabilities of a child survival model will be of immense use for various health care professionals and policy makers to have better and efficient health care management systems. As per the current review search, the numbers of studies on validation and prediction of the prognostic models particularly addressing various aspects related to child's health are rare. Therefore, there is an immense need to perform validation of any prognostic model before its future uses in public health prospect.

Approximately 2.1 million child deaths occur every year in India with mortality rates varying from state to state [10]. To utilize important factors of child survival so as to achieve the millennium development goals on under-five mortality reduction, separate child survival models will be useful for each state.

The objective of the study was to validate child survival model of Tamil Nadu State using validation indices such as shrinkage coefficient, Somer`s Dxy rank correlation and calibration curve through bootstrapping technique for its predictive accuracy so that the developed model can be used for health care professionals and policy planners for better child health care management.

Material and Methods

Data Collection Methods and Description of Variables

The data sets used under this study are from the National Family Health Survey (NFHS- 1992-93) of the Tamil Nadu (TN) State, India [11]. The sample design adopted by the National Family Health Survey was systematic, two-stage stratified sample. The detailed methodology of data collection is reported in the Tamil Nadu report of NFHS-1992-93[11]. A total of 627 children of first birth order born during four years preceding the survey were studied for the analysis. The sample was self weighting for the state. In case of multiple births i.e. triplets or twins, only the first one was included in the analysis. Duration of child survival (in months) with child status (alive/dead) was considered as dependent or outcome variable. Co-variates such as religion/caste (SC/ST Hindu/ other Hindu / non-Hindu); place of residence (rural/ urban); mother's education (illiterate/ primary/ middle/ ≥ high school); breast feeding (no/ yes); sex of index child (male/ female); mother's occupation (not working/working); father's occupation (not working/ working); type of house (kuchha/ Semi pucca + Pucca); media exposure (no/ yes); distance from primary health center (≥ 2 kms/ < 2 kms); antenatal care (no/ yes); immunization of child (no/ yes); place of delivery (at home/ at hospital); complication at delivery (no/ yes); premature birth (no/ yes) and age of mother at index child in complete years were considered as independent variables. In view of its non-linear relationship with child survival, mother's age at index child was squared and added to the model to fulfill the linearity assumption for the Cox PH model. All the variables mentioned above were in the form of fixed covariates with fixed effects, except the age of mother at index child, which was the time varying covariate with fixed effect.

Multivariate Cox Regression Model

Cox regression assumptions of linearity, proportionality, interaction effect and multi co-linearity were checked using exploratory analysis. The Cox PH model equation of the hazard at time t, λ (t),

$$\lambda(t) = \lambda_0(t)\exp(\beta_1X_1+\beta_2X_{2+}............._+\beta_pX_p)$$

where, $\lambda_0(t)$ is called baseline hazard function and $\beta_{1,}$ $\beta_{2,}$ ....β_p are unknown regression coefficients has been used [12].

Regression coefficients of the model have been derived through maximizing likelihood function. Taking all the covariates considered in the multivariate analysis, a step-wise method is used to select variables for inclusion or exclusion from the model in a sequential fashion. For this, a forward with a test for backward elimination is used with probability levels for entry and removal as 0.15 and 0.10 respectively.

Model Validation Method

Validation of model was undertaken by use of bootstrap re-sampling method. For each group of 200 bootstrap samples, the model was refitted and tested against the observed sample in order to derive an estimate of the predictive accuracy and bias. Two important components of predictive accuracy i.e. calibration and discrimination were used for the validation [2]. The detailed steps and explanation is given in few studies [2, 13, 14].

The shrinkage coefficient was used to quantify the over fitting of the model. The heuristic shrinkage estimator [15] equation

$$\Upsilon = (\text{model } \chi^2 - p)/ \text{ model } \chi^2$$

Where, p is the number of regression parameters including all non-linear and interaction effects and the model χ^2 is the total likelihood ratio of χ^2 statistics was used to quantify the over fitting of the model.

Discrimination aspect of the validation of model, was measured through Somer's D_{xy} rank correlation between predicted the log hazard and the observed survival time using 2(C-0.5) formula, where C was concordance index and was performed using various steps described in [16,17] .

Predicted Survival Probabilities of the Developed Child Survival Model

The prediction of survival probabilities have been calculated by the exponential expression of the Cox model, also known as 'Risk score' and generally denoted by R, is defined as follows:

$$R = \beta_1X_1+ \beta_2X_2+.......+ \beta pXp$$

Where X_1 , X_2 , ..., Xp are the considered levels of p predictor variables and β_1 ,β_2 , ….βp are respective unknown regression coefficients. The details of steps and procedure are explained by [18] and gain in survival probability after adjustment in relation to considered levels of selected covariates was obtained by

$$S(t) = S_0(t)^{\exp(R_2 - R_1)}$$

A complete analysis under the present study was accomplished with the help of various packages namely BMDP version 7.0, University of California, 1992 [19] ,S-plus 6.0, 1988-97, Mathsoft Inc., Seatle, WA 98109-3044 USA [20]. These packages were either available in the Department of Biostatistics, All India Institute of Medical Sciences (AIIMS), New Delhi or used after due permission of the concerned authority. Predicted probabilities of survival were performed through Macros on Excel 2000.

Results

Univariate Analysis

Table 1 describes the distribution of children and percentage of deaths among them according to different categories of the variables. Mortality was higher among children who were not breast fed (69.2%) and who had no antenatal care (37.5%), no immunization (12.6%) and delivery at home (25.8%). The percentages of deaths were almost similar, approximately 6% in terms of residence in rural, mother’s literacy, Father’s education (middle school), mother’s media exposure, no complications

Table 1. *Distribution of Children and their Percentage of Deaths across the Different Variables (N=627)*

Variables	**Category**	**No. of Children**	**% of deaths**
Religion/ caste	SC/ST Hindu	92	5.2
	Other Hindu	442	5.9
	Non-Hindu	93	4.0
Place of residence	Rural	381	6.0
	Urban	246	4.9
Mother’s education	Illiterate	223	7.6
	Primary	179	5.6
	Middle	105	2.9
	≥High School	120	4.2
Breastfeeding	No	39	69.2
	Yes	588	1.4
Sex of the index child	Male	310	3.8
	Female	317	7.4
Mother’s occupation	Not Working	449	4.2
	Working	178	7.4
Father’s occupation	Not Working	18	11.1
	Working	609	5.4
Father’s education	Illiterate	124	6.5
	Primary	193	8.3
	Middle	110	6.4
	≥High School	200	2.0
Type of house	Kuchha	228	8.3
	Semipucca+Pucca	399	4.0
Mother’s media exposure	No	100	6.0
	Yes	527	5.5
Distance of primary health Center	≥2 KM	337	6.5
	<2 KM	290	4.5
Antenatal care	No	16	37.5
	Yes	611	4.8
Immunization of the child	No	255	12.6
	Yes	372	1.0
Place of delivery	At home	31	25.8
	At hospital	596	4.5
Complications at delivery	No	462	6.1
	Yes	165	4.2
Premature birth	No	592	3.9
	Yes	35	34.3
Total		627	5.6

Table 2. *Bi-variate Analysis- First Birth Order Child Survival Model*

Variables	**Categories**	**Cox Regression**			
		Coefficient(β)	**S.E.**	**RR**	**95% CI**
Religion/caste[a]	Non-Hindu	0.1339	0.4258	1.14	0.50 – 2.63
	Other Hindu	-0.2660	0.8018	0.77	0.16 – 3.69
Place of residence[b]	Urban	-0.2094	0.3561	0.81	0.40 – 1.62
Mother's education[c]	Primary	-0.3062	0.3985	0.73	0.33 – 1.61
	Middle	-0.9907	0.6262	0.37	0.11 – 1.27
	≥High school	-0.6093	0.5087	0.54	0.20 – 1.47
Breastfeeding[d]	Yes	-4.4505	0.4093	0.01	0.001 – 0.03
Sex of index child[f]	Male	0.6908	0.3561	1.99	0.99 – 4.01
Mother's occupation[g]	Working	0.7631	0.3393	2.14	1.10 – 4.17
Father's occupation[h]	Working	-0.7255	0.7282	0.48	0.12 – 2.02
Father's education[i]	Primary	0.2576	0.4330	1.29	0.55 – 3.02
	Middle	-0.0073	0.5176	0.99	0.36 – 2.73
	≥High school	-1.1833	0.6124	0.31	0.09 – 1.02
Type of house[j]	Pucca+Semi Pucca	-0.7530	0.3393	0.47	0.24 – 0.92
Media exposure[k]	Yes	-0.836	0.4485	0.92	0.38 – 2.21
Distance primary[l] health center	<2 KM	-0.3799	0.3498	0.68	0.34 – 1.36
Antenatal care[m]	Yes	-2.2632	0.4496	0.10	0.04 – 0.25
Immunization[n]	Yes	-2.8219	0.6040	0.06	0.02 – 0.19
Place of delivery[o]	At home	-1.8450	0.4030	6.25	2.85 – 14.3
Complications at delivery[p]	Yes	-0.3511	0.4226	0.70	0.31 – 1.61
Premature birth[q]	Yes	2.3524	0.3573	10.51	5.22 – 21.17

Reference Categories:
a) SC/ST Hindu, b) Rural, c) Illiterate, d) No, e) < 24 Month, f) Female, g) Not working,, h) Not working,, i) Illiterate, j) Kuchha, k) No l) ≥2 KM, m) No, n) No, o) At Hospital, p) No, q) No.
SE:Standard Error

at delivery and 5% for residence in urban areas, father's working, more than 2km from health center, antenatal care, delivery at hospital.

Bivariate Analysis

Table 2 describes the results under bi-variate analysis that are in the form of risk ratio (RR), and 95% confidence interval (CI) The table reveals that breast feeding (RR: 0.01; C.I. 0.001-0.03), Type house (Pucca) (RR: 0.47; C.I. 0.24-0.92), antenatal care (RR: 0.10; C.I. 0.04-0.25), and immunization of the child (RR: 0.06; C.I. 0.02-0.19) are found protective risk factor whereas working mothers (RR: 2.14; C.I. 1.10-4.17), place of delivery (at home) (RR: 6.25; C.I.: 2.85-14.3) and premature birth (RR: 10.51; C.I.: 5.22–21.17) were found risk factor for child survival.

Multivariate Analysis

Multivariate Cox regression was performed after all the covariates satisfied the proportionality assumption i.e. log [-log(s (t)] for different subjects at equidistance over time. Consideration of each covariate in data analysis was done in the form of their fixed effects. Some of the variables entered into the model partially. For meaningful presentation, partially entered variables were considered with all the categories of the variables in the presentation in the final model. All the important variables were considered in the model, only variables breast feeding (RR: 0.01; C.I. 0.004-0.02), and immunization of the child (RR: 0.05; C.I. 0.01-0.17) were found significant protective factors. Details are given in table 3.

Table 3. *Multivariate Cox PH Analysis- First Birth Order Child Survival Model*

Variable	**Coefficient**	**S.E.**	**R.R.**	**95% C.I.**

Mother's age at index child	0.1144	0.4472	1.12	0.46 – 2.69
Mother's age^2 at index child	-0.0028	0.0108	1.00	0.98 – 1.02
Breastfeeding[a]	-4.7310*	0.4633	0.01	0.004 - 0.02
Immunization[b]	-3.0061*	0.6355	0.05	0.01 - 0.17

Reference Category: a) No, b) No..
**Significant at < 0.05 level*
RR = Relative Risk
SE: Standard Error

Table 4. *Validation Indices of Cox PH Models Developed for First Birth Order Child Survival*

	Index Original	Training	Test	Optimism	Index Corrected	Re-sample
Birth order I Shrinkage coefficient	1.00	1.00	0.78	0.22	0.78	200
Dxy	-0.93	-0.94	-0.89	-0.05	-0.89	200

D_{xy}: Somer`s D-rank correlation.

Model Validation

Calibration and Discrimination indices are described by Harrell et al [2] for predictive accuracy of a developed Cox hazard model. Calibration is obtained in the form of Shrinkage coefficient to quantify lack of fit of the model and calibration curve to see extent of bias in the model. Discrimination aspect of the model is measured by Somer's D_{xy} rank correlation between the log hazard and the observed survival time through bootstrapping. Shrinkage coefficient was 78% indicating 22% lack of fit in the model whereas Somer's D_{xy} rank correlation was –0.89 indicating good correlation between the log hazard and the observed survival time. Detail results of model validation indices is provided in Table 4. Similar pattern was revealed by calibration curve i.e, to evaluate the accuracy of the model for prediction where dots correspond to apparent predictive accuracy and X marks the bootstrap corrected estimates (Figure 1).

Prediction of survival probabilities through developed child survival model:

Table 5 reveals that the prevailing probability of child survival is constant (95%) at 1 month to 12 months. Should all the children are breastfed, the gain in child survival probability is only 1.25%. On the other hand, 3.6% gain is achieved from the immunization of all children. Should all the children are breastfed as well as fully immunized, the gain in child survival is up to 4% at 1 month and up to 12 months period. Thus, immunization of children is found important irrespective of breastfeeding.

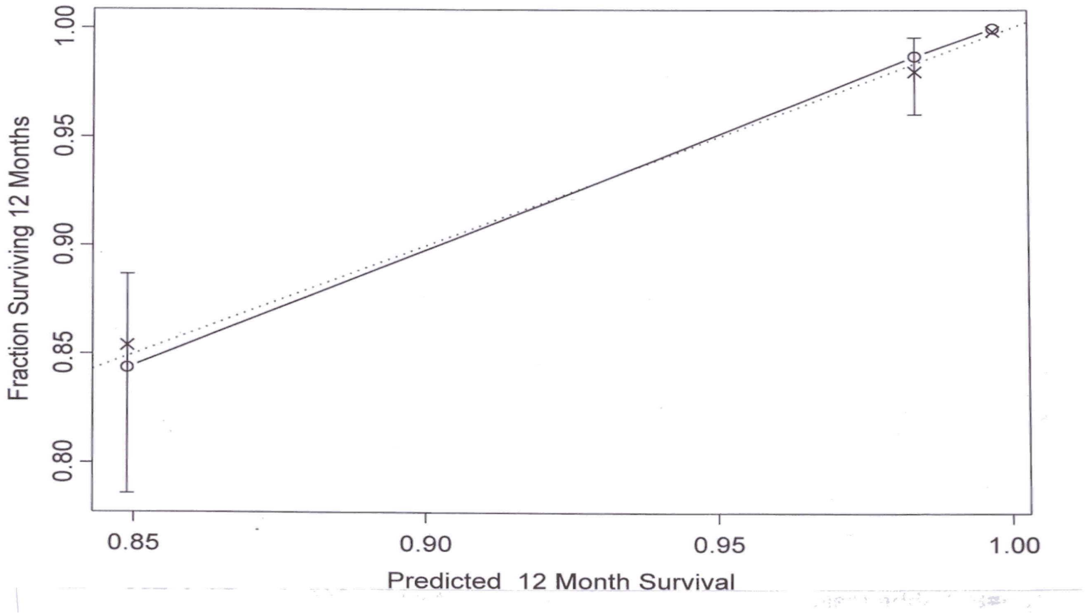

***Figure 1**. Bootstrap estimates of calibration accuracy for 12 months from the final Cox model for first birth order. Dots correspond to apparent predictive accuracy. X marks the bootstrap-corrected estimates.*

***Table 5**. Estimated Probabilities of Survival of Children of First Birth Order at Specific Months after Birth*

Characteristics	**Probability of survival at months**				
	1	3	6	9	12
Average	0.9486	0.9486	0.9486	0.9486	0.9486
Breast feeding	0.9611	0.9611	0.9611	0.9611	0.9611
Immunisation	0.9847	0.9847	0.9847	0.9847	0.9847
Breastfeeding+immunisation	0.9885	0.9885	0.9885	0.9885	0.9985

Discussion and Conclusion

This study of survival rates of children in Tamil Nadu revealed that negligence of immunization and breastfeeding affect survival rates of children. In consistent with our study, another survival cohort study done in western rural India reported that the role of child survival strategies like immunization and early initiation of breast feeding in improving survival cannot be challenged [21]. Interestingly, mortality was higher in children who were not immunized (12.6% vs 1%) and no breast fed children (69.2% vs 1.4%) compared to their counterparts.

Since two decades, the international target of reducing the infant mortality below 70 per 1000 live births has not been achieved [22]. India has highest child deaths world wide approximately 2.1 million child deaths every year [10]. The national under-five mortality rate is around 87 per 1000 live births with wide variation among states. The main causes of deaths in low-income countries are diarrhea, pneumonia, measles, malaria, and HIV/AIDS. The major underlying cause of death is mal nutrition. Deaths among children under-five years can be reduced through achievement of high coverage of basic public health and nutrition interventions. Due to state wise high variations in the death rates, separate child survival models will be needed to get an overall improvement in child survival in India. To apply a developed model for health management, its predictive accuracy is to be checked so that policy planners may get fruitful results to achieve the Millennium Development Goal [23].

Results of validation indices suggest that this present developed model is good enough to describe the predictive accuracy for the target outcome. This can be used for prediction and the predicted survival probabilities of individual as well as a combination of variables. The finding of this present study will be very useful for various health professionals and health policy planners for betterment of child health care management.

Acknowledgement

We would like to thank Prof. Arvind Pandey, Director, NIMS, ICMR, New Delhi, India for providing Statistical Package S-Plus 6.0 and Frank E. Harrell Jr, Division of Biostatistics and Epidemiology, Department of Health Evaluation Sciences, School of Medicine, University of Virginia, USA for the permission to use Library in S-Plus.

References

1. Harrell FE. Regression Modeling Strategies with Application to Linear Models, Logistic Regression, and Survival Analysis. 2001; Springer-Verlag, New York Berlin Heidelberg.
2. Harrell FE, Lee KL, Mark DB. Tutorial in Biostatistics Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine 1996; 15: 361-387.
3. Singh R, Begum S, Ahuja RK, Chandra P, Dwivedi SN. Prediction of Child Survival in India using developed Cox PH model: a utility for health policy programmers. Statistics in Transition 2007; 8 (1): 97-110.
4. Efron B. Bootstrap methods: Another look at the Jackknife. The Annals of Statistics 1979; 7(1): 1-26.
5. Efron B. Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. Biometrika 1981; 63: 589-599.
6. Efron B. The jackknife, the bootstrap, and other resampling plans, Society of Industrial and Applied Mathematics CBMS-NSF Monographs 1982; 38.
7. Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. Journal of the American Statistical Association 1983; 78: 316-331.
8. Efron B, Tibshirani RJ. An Introduction to Bootstrap. Chapman and Hall 1993, New York 1983.
9. Fan X; Wang L. Comparability of jackknife and bootstrap results: An investigation for a case of canonical correlation analysis. Journal of Experimental Education 1996; 64:173-189.
10. UNICEF. State of the World's Children 2005, New York: UNICEF 2005.
11. International Institute for population Sciences (IIPS), 1995. National Family Health Survey, Tamil Nadu (TN), India, 1992-1993.
12. Kleinbaum DG. Survival Analysis. Statistics in Health Sciences. Springer-Verlag, New York, Inc 1996.
13. Harrell FE, Lee K, Califf R, Proyor D, Rosati R. Regression modeling strategies for improved prognostic prediction, Statistics in Medicine 1984; 3(2): 143-152.
14. Harrell FE. Design: S-Plus functions for biostatistical/-epidemiologic modeling, testing, estimation, validation, graphics, prediction, and typesetting by sorting enhanced model design attributes in the fit. Programs available from statlib@lib.stat.cmu.edu 1994.
15. Van Houwelingen JC, Le Cessie S. Predictive value of statistical models. Statistics in Medicine 1990; 9 (11): 1303-1325.
16. Harrell, F.E.; Califf, R.M.; Pryor, D.B.; Lee, K.L.; Rosati, R.A. Evaluating the yield of medical tests. JAMA 1982; 247: 2543-2546.
17. Kaplan ME. Non parametric estimation from incomplete observations. J Amer Statist Assoc 1958; 53: 457-481.
18. Dickson ER, Grambsch PM, Fleming TR,; Fisher LD, Langworthy A. Prognosis in Primary Biliary Cirrhosis: Model doe Decision-Making. Hepotolog 1980; 10: 1-7.
19. BMDP 7.0 Statistical Software, Inc, 1440 Sepulveda Boulevard Suite 316, Los Angeles, CA 90025 1992.
20. S-plus 6.0, Mathsoft Inc., Seatle, WA 98109-3044 USA 1988-97.
21. Hirve S, Ganatra B. A Prospective Cohort Study on the Survival Experience of Under Fve children in Rural Western India, Indian Pediatrics 1997; 34: 995-1001.
22. UNICEF. Progress since the World Summit for children: a statistical review, New York: UNICEF 2001.
23. UN. General assembly, 56th session, Road Map towards the Implementation of the United Nations Millennium Declaration: report of the Secretary General (UN document no. A/56/326), New York: United Nations 2001.

***Correspondence to:**
Abdulbari Bener
Department of Medical Statistics & Epidemiology
Hamad Medical Corporation and
Department of Public Health, Weill Cornell Medical College
PO Box 3050, Doha- State of Qatar