Revista Colombiana de Estadística
Número especial en Bioestadística
Junio 2012, volumen 35, no. 2, pp. 289 a 307

Determinants of Birth Intervals in Tamil Nadu in India: Developing Cox Hazard Models with Validations and Predictions

Determinantes de los intervalos genésicos en Tamil Nadu (India): desarrollando modelos de riesgos de Cox con validaciones y predicciones

Rajvir Singh[1,a], Vrijesh Tripathi[2,b], Mani Kalaivani[3,c], Kalpana Singh[4,d], S.N. Dwivedi[3,e]

[1]Medical Research Centre- Cardiology, CCS Department, Hamad Medical Corporation, Doha, Qatar

[2]Department of Mathematics & Statistics, Faculty of Science and Agriculture, The University of the West Indies, Trinidad & Tobago, West Indies

[3]Department of Biostatistics, All India Institute of Medical Sciences, New Delhi, India

[4]Clinical Research Department, Max Healthcare Institute Ltd., New Delhi, India

Abstract

The present study uses data from National Family Health Survey (NFHS-1) 1992-93 (International Institute for Population Sciences 1995) conducted in the state of Tamil Nadu, India. Cox models were developed to analyze the effect of breastfeeding as time varying and time dependent factor on birth intervals. Breastfeeding alone improved the log likelihood up to a higher level in each birth interval. Other factors that entered into the models were: at first birth interval, women's education (high school & above) and working status of women; at second birth interval, survival status of index child alive and husband's education (high school & above), and at third birth interval, breastfeeding more than 22 month were found to be protective factors for next births. Validation of the developed models was done through bootstrapping to predict birth intervals.

***Key words*:** Cox model, Multivariate analysis, Validation, Predictions.

[a]Senior consultant. E-mail: rajvir.aiims@gmail.com
[b]Lecturer. E-mail: Vrijesh.Tripathi@sta.uwi.edu
[c]Scientist. E-mail: manikalaivani@gmail.com
[d]Researcher. E-mail: nkalpanasingh@gmail.com
[e]Professor. E-mail: dwivedi7@hotmail.com

Resumen

Este estudio utiliza datos de la Encuesta Nacional de Salud Familiar (International Institute for Population Sciences 1995) realizada en el estado de Tamil Nadu, India. Se desarrollaron modelos de Cox para analizar el efecto de la lactancia materna cuando varía en el tiempo y el factor tiempo depende de los intervalos genésicos. La lactancia materna sólo mejora la probabilidad de acceder a un nivel más alto en cada intervalo de nacimiento. Otros factores que entraron en los modelos fueron en el intervalo del primer parto: nivel educativo de la madre (secundaria y superior) y trabajo de la madre; en el intervalo del segundo parto: nivel de supervivencia en el índice de vida infantil y nivel educativo del padre (secundaria y superior), y en el intervalo del tercer parto: lactancia materna más 22 meses. Cada uno de los anteriores es un factor protector para ampliar el intervalo entre nacimientos en el estudio. Además, este estudio confirma los modelos desarrollados en los servicios públicos de predicción para los intervalos genésicos.

***Palabras clave*:** análisis multivariado, modelo de Cox, predicciones, validación.

1. Introduction

Population change is a global phenomenon. This varies significantly among regions and even among countries within the same region. This also varies significantly among states within the same country. Cognizant of inherent problems in rapid population growth in developing countries like India, epidemiologists including biostatisticians, demographers and social scientists have given high priority to a thorough understanding of the differentials and determinants of this phenomenon. Rates of population change and various aspects of reproductive health need to be understood to understand this phenomenon. Birth interval is defined as interval between termination of one completed pregnancy and the termination of the next. The intrinsic growth rate as well as the mean generational length of any population may get affected by the birth interval pattern (Srinivasan 1980). Thus birth interval can be viewed as a major determinant of population change. The mechanism of reproductive process can be assessed through the analysis of birth interval. This is possible because the disaggregating of the reproductive process is possible into a series of stages, beginning with marriage followed by first birth, second birth, third birth and so on, provides an insight into the fertility behavior of the population which is principally responsible for population change. Emphasis has often been laid on delaying the first birth, interval births, avoiding too many births, and on stopping child bearing in time (UNFPA 1997). An appropriate epidemiological understanding of birth intervals in a region may be helpful to policy planners for an appropriate public health program for the region in the belief that such an attempt is likely to provide more accurate results, which would lead to more appropriate intervention.

2. Material and Methods

The National Family Health Survey (NFHS-I) is a state representative survey of ever-married women aged 13-49 years. Survey period was from 18th April, 1992 to 7th July, 1992 in Tamil Nadu (TN) (International Institute for Population Sciences 1995). Data were collected in the form of systematic, stratified sample of households with two stages in rural areas (selection of villages followed by selection of households) and three stages in urban areas (selection of cities/towns, followed by urban blocks, and finally households) in self weighting fashion. The number of households surveyed was 4,287 having 3,948 ever-married women. Out of them 66.3% were non-sterilized and currently married. The detailed reports covering sampling methods and all other aspects mentioned above were prepared and documented in Population Research Centre, The Gandhigram Institute of Rural Health and Family Welfare Trust, Ambathurai R.S., and International Institute for Population Sciences (1994).

The parity (birth order) specific hazards models for birth intervals in TN have been worked out utilizing available data on 627 women of parity-I, 566 women of parity-II, and 310 women of parity-III. The results provide information on factors associated with experiencing next live birth. The order of the interval was the parity of a woman; e.g., the first birth interval is time interval between effective age at marriage to first parity of women; the second birth interval is time interval between first parity and second parity; and so on. In other words (Trussell, Martin, Fledman, Palmore, Concepcion & Abu Bakar 1985), the order of the interval is the order of birth that would close the interval; e.g., the first birth interval extends from the effective age at marriage to the first birth; the second birth interval extends from the first birth to the second birth; the third birth interval extends from the second birth to third birth; and so on. Birth intervals were considered in months as interval variables in the analysis (Trussell et al. 1985).

Data regarding children from multiple births (including twins) were considered as single birth and included for the analysis. Further, it was decided to exclude birth interval during which there was no possibility of conception because the woman or her husband had been sterilized. As a result of preliminary analysis, the decision to exclude the higher order birth intervals (fourth onwards) from the analysis was taken, mainly because of insufficient number of records/events. Incomplete records, almost negligible in number, were not considered in the analysis.

All the variables in the study satisfied proportional hazard (PH) assumption except breastfeeding and were considered as fixed covariates with fixed effect. Age at index child was taken as continuous variable and did not satisfy linear assumption. Age was considered as time varying with fixed effect and age^2 was added to overcome the non-linearity. Breastfeeding was considered as time varying time dependant factor (Dwivedi & Rajvir 2003). As per the method followed by Trussell et al. (1985), the interval for first birth interval was divided into five categories: $\leq$ 15 months/16-21 months/22-27 month/28-33 month/$\geq$ 34 months. On account of lesser proportion of women experiencing next live birth under extreme categories, the live birth interval related to the second birth interval was divided into three categories: $\leq$ 21 months/22-27 months/$\geq$ 28 months. Similarly, live birth interval under third birth interval was categorized as $\leq$ 21 months and $\geq$ 22 months for meaningful analysis. All the variables having ($p < 0.25$) at univariate analysis were selected for multivariate Cox analysis and the variables ($p < 0.10$) at multivariate analysis were considered in the models.

Being breastfeeding as time varying covariate with time dependent effect, an extended Cox hazards model suggested by Trussell et al. (1985) was used. If birth interval categorized into k categories, general form of the extended Cox Hazards considered as under:

$$\lambda_k(t, X(t)) =$$
$$\lambda_{0k} t \exp\left[\sum_{i=1}^{p_{11}} \beta_{1i} X_{1i} + \sum_{i=1}^{p_{12}} \beta_{2ik}(t) X_{2i} + \sum_{j=1}^{p_{21}} \beta_{1j} X_{1jk}(t) + \sum_{j=1}^{p_{22}} \beta_{2jk}(t) X_{2jk}(t)\right]$$

where, $\lambda_{0k}(t)$ is category-specific baseline hazard; β_{1i} are respective fixed (time-independent) effects of fixed covariates X_{1i}; $\beta_{2ik}(t)$ are respective category-specific (time-dependent) effects of fixed covariates X_{2i}; β_{1j} are respective fixed effects of time-varying covariates $X_{1jk}(t)$; and $\beta_{2jk}(t)$ are respective category-specific (time-dependent) effects of time-varying covariates $X_{2jk}(t)$.

Maximum likelihood functions for extended Cox model were calculated to produce estimates of the coefficients and their standard errors (Trussell & Charles 1983). Using regression coefficients and respective standard errors, Risk Ratio or Hazard Ratio (HR) in the form of $\exp(\beta)$ related to an exposure variable and its 95% confidence interval were calculated and interpreted using standard convention followed in the case of Cox Proportional Hazard model (Kleinbaum 1996).

To satisfy the linearity assumption in the Cox PH models, at each time t, $\log \lambda(t)$ and equivalently $\log[-\log(S(t))]$ were linearly related to covariates, where $\lambda(t)$ was the hazard function or instantaneous event rate at time t and $S(t)$ was the probability of surviving until time t (not having next birth in the study).

Log-log survival curves (Cox 1972, Namboodiri & Suchindran 1987, Kleinbaum 1996) were assessed to check PH assumption of proportionality for each fixed effect with fixed covariate whereas; for continuous covariate i.e. woman's age, birth spacing was categorized as $\leq$ 15 months, 16-21 months, 22-27 months, 28-33 months, and $\geq$ 34 months, based on an exploratory analysis. For a procedure involving time dependent variable, presence of breastfeeding for these categories was specified as > 0, ≥ 16, ≥ 22, ≥ 28 and ≥ 34 months, respectively.

First order interactions between covariates were tested using stratified analysis and no interaction was found. Collinearity among the covariates was checked through correlation analysis (Fox 2008, pp. 307-331).

All covariates considered in the multivariate analysis were followed by stepwise method to select variables for inclusion or exclusion from the model in a sequential fashion. For this, a forward selection with a test for backward elimination was used with probability levels for entry and removal as 0.15 and 0.10, respectively. This was done in view of the fact that early deletion of covariates with little chance of being measured reliably or of being predictive would result in models with less overfitting and more generalization.

In order to test validation of developed models, bootstrapping was applied (Efron & Tibshirani 1993). Calibration curve with 200 re-samples were used to estimate the optimism between predicted survival probability estimates from the developed Cox model and the corresponding Kaplan Meier survival probability (Kaplan & Meier 1958). Shrinkage coefficient was calculated to check for overfitting of the model (Van Houwelingen & Cessie 1990), and discrimination aspect of the model was measured through Somer's D_{xy} rank correlation between predicted log hazard and observed survival time (Harrell, Lee & Mark 1996, Harrell 2001).

Predictive probabilities for a woman not attaining next live birth for a particular variable or combination of variables by holding other variables at their mean levels were estimated. The exponential expression of the Cox model, also known as "Risk score" and generally denoted by R, may be defined as follows: $R = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$ (Dickson, Grambsch, Fleming, Fisher & Langworthy 1989, Singh, Begum, Ahuja, Chandra & Dwivedi 2007), where, $X_1, X_2, \ldots, X_p$ are the considered levels of p predictor variables and $\beta_1, \beta_2, \ldots, \beta_p$ are respective unknown regression coefficients.

Thus, using maximum likelihood estimates of regression coefficients for the model being used and substituting the observed values of the covariates for each individual, risk score is obtained for every person (woman) included in the data analysis. The arithmetic mean of these risk scores provides an average risk score R_1 and hence R_1 is constant for a given data set. Risk score R_2 is obtained again by using the equation substituting again the estimated values of the regression coefficients and changed levels of selected variable/set of variables (same level for every woman) but retaining other variables at their mean level. $S_0(t)$, the baseline survival probabilities at different points of time for a person with average risk score R_1 may be worked out using Kaplan Meier method. Thus, $S_0(t)$ at a given point of time is nothing but the survival probability obtained through Kaplan Meier method at that point of time. Gain in survival probability after adjustment in relation to considered levels of selected covariates has been obtained by $S(t) = S_0(t)^{\exp(R_2 - R_1)}$.

In the present study, for each model, survival probabilities in relation to R_1 are listed under first row of the concerned table, whereas those related to R_2 are listed in successive rows. Thus, differences between these two probabilities provide gain/loss as a result of proposed change in the levels of selected variable/set of variables. BMDP 7.0, University of California, 1992; S-Plus 4.0, 1988-97, Mathsoft Inc. Seatle , WA 98109-3044, USA and Excel 2000 Statistical Software were employed for the analysis.

3. Results

There is similarity in distribution of women for each birth interval in terms of religion/caste, place of residence, (ever) contraceptive use, (ever) fetal loss, sex of index child, survival status of index child, husband's occupation, type of house, media exposure, and distance of primary health center (Table 1). As the parity increased, there was an increase in the proportion of women in categories characterized by illiteracy, working status of women, illiteracy of husband and breastfeeding for 22 and more months. Women with high school (and above) education were significantly less likely to experience the next live birth in comparison to illiterate women in case of first and second birth interval, whereas education was not a significant factor in case of the third birth Interval (HR: 0.79; C.I.: 0.24-2.63). In addition, women with middle education were also significantly less likely to the experience the next live birth in case of second birth interval (HR: 0.50; C.I.: 0.27-0.92). Ever contraceptive use was a significant protective factor in case of first and second birth interval in contrast to third birth interval (HR: 0.83; C.I.: 0.38-1.81). Ever fetal loss was a significant protective factor only in case of first birth interval (HR: 0.70; C.I.: 0.50-0.97). Previous birth interval was not used in case of first birth interval. But, women with more than 36 months of previous birth interval were significantly less likely to experience the next live birth in case of second birth interval (HR: 0.54; C.I.: 0.33-0.88) but previous birth interval was not a significant factor for third birth interval (HR: 0.50; C.I.: 0.22-1.10). Further, women with surviving index child were significantly less likely to experience next live birth in case of first and second birth interval, but this was not the case for third birth interval (HR: 0.45; C.I.: 0.16-1.26). Husband's education (high school and above) was a significant protective factor against the next live birth only in the case of second birth interval (HR: 0.41; C.I.: 0.24-0.72). Shorter distance from primary health center was also a significant protective factor only in the case of second birth interval (HR: 0.68; C.I.: 0.47-0.99). Surprisingly, breastfeeding did not emerge as a significant predictor of birth interval in most cases. But, the period 1-15 months of breastfeeding did predict the first birth interval where it was noticed to be a significant risk factor (HR: 1.92; C.I.: 1.04-3.58).

Table 1: Covariates associated with parity specific birth intervals: definitions and means.

Variables	Category	Parity		
		1st Interval	2nd Interval	3rd Interval
Religion/caste	SC/ST Hindu	0.15	0.18	0.23
	Other Hindu	0.70	0.67	0.65
	Non-Hindu	0.15	0.15	0.12
Place of residence	Rural	0.61	0.61	0.67
	Urban	0.39	0.39	0.33
Women's education	Illiterate	0.35	0.43	0.58
	Primary	0.29	0.25	0.23
	Middle	0.17	0.15	0.08
	≥High school	0.19	0.17	0.11
Ever contraceptive use	No	0.69	0.66	0.72
	Yes	0.31	0.34	0.28
Ever fetal loss	No	0.79	0.76	0.72
	Yes	0.21	0.24	0.28
Previous birth interval	< 24 Months	- -	0.28	0.34
	24-36 Month	- -	0.39	0.26
	> 36 Months	- -	0.33	0.40

Table 1: Continue.

Variables	Category	Parity		
		1st Interval	2nd Interval	3rd Interval
Sex of index child	Male	0.49	0.49	0.45
	Female	0.51	0.51	0.55
Survival status of index child	Alive	0.94	0.94	0.96
	Dead	0.06	0.06	0.04
Women's occupation	Not Working	0.72	0.68	0.54
	Working	0.28	0.32	0.46
Husband's occupation	Not working	0.03	0.03	0.02
	Working	0.97	0.97	0.98
Husband's education	Illiterate	0.19	0.20	0.27
	Primary	0.31	0.34	0.36
	Middle	0.18	0.16	0.16
	≥High School	0.32	0.30	0.21
Type of house	Kuchha	0.32	0.37	0.39
	SemiPucca+Pucca	0.68	0.63	0.61
Media exposure	No	0.16	0.19	0.22
	Yes	0.84	0.81	0.78
Distance of primary health Center	$\geq$ 2 km	0.54	0.54	0.60
	$<$ 2 km	0.46	0.46	0.40
Breastfeeding (months)	$\geq$ 1	0.92		
	$\geq$ 16	0.29	0.38*	
	$\geq$ 22	0.11	0.17	0.22***
	$\geq$ 28	0.04	0.05**	
	$\geq$ 34	0.02		
Age of women at index child (years) (X SD)	Continuous	20.41 ± 3.51	22.36 ± 3.72	24.0 ± 4.01

*: for period of birth interval $\leq$ 21 months; **: for period of birth interval $\geq$ 28 months
***: for period of birth interval $\geq$ 22 months

Table 2: Univariate analysis and multivariate with extended cox model of first birth interval according to different variables in TN.

Variables	Categories	Univariate		Multivariate	
		exp(β)	C.I. 95%	exp(β)	C.I. 95%
Women's age at index child	Continuous			1.33	0.88 - 2.04
Women's age^2 at index child	Continuous			0.99	0.98 - 1.00
Religion/caste[a]	Non-Hindu	0.86	0.59 - 1.25		
	Other Hindu	1.21	0.76 - 1.91		
Place of residence[b]	Urban	0.96	0.73 - 1.27		
Women's education[c]	Primary	1.24	0.89 - 1.72	0.96	0.65 - 1.42
	Middle	1.05	0.71 - 1.54	0.92	0.58 - 1.45
	≥High school	0.58	0.38 - 0.90	0.40	0.22 - 0.75
Ever contraceptive use[d]	Yes	0.70	0.52 - 0.93		
Ever fetal loss[e]	Yes	0.70	0.50 - 0.97	0.63	0.44 - 0.88
Sex of index child[f]	Male	0.95	0.73 - 1.25	0.52	0.27 - 1.02
Survival status of index child[g]	Alive	1.86	1.18 - 2.96		
Women's occupation[h]	Working	0.76	0.56 - 1.03	0.72	0.52 - 1.00

Table 2: Continue.

Variables	Categories	Univariate		Multivariate	
		exp(β)	C.I. 95%	exp(β)	C.I. 95%
Husband's occupation[i]	Working	0.56	0.30 - 1.06		
Husband's education[j]	Primary	1.05	0.73 - 1.51	1.08	0.72 - 1.61
	Middle	0.86	0.56 - 1.32	0.66	0.41 - 1.08
	≥High school	0.70	0.48 - 1.03	0.75	0.45 - 1.26
Type of house[k]	Pucca+Semi Pucca	0.79	0.60 - 1.04		
Media exposure[l]	Yes	1.00	0.70 - 1.44		
Distance primary health centre[m]	< 2 km	0.90	0.69 - 1.19		
Birth interval	1-15 months	0.07	0.02 - 0.18	0.07	0.03 - 0.19
Breastfeeding[n]	≥ 1 months	1.92	1.04 - 3.58	3.08	1.44 - 6.60
Birth interval	16-21 months	0.29	0.13 - 0.62	0.30	0.14 - 0.66
Breastfeeding[n]	≥ 16 months	0.82	0.56 - 1.19	0.74	0.50 - 1.09
Birth interval	22-27 months	0.39	0.13 - 1.16	0.40	0.13 - 1.19
Breastfeeding[n]	≥ 22 months	0.98	0.52 - 1.86	0.90	0.46 - 1.73
Birth interval	28-33 months	0.46	0.04 - 4.86	0.51	0.05 - 5.51
Breastfeeding[n]	≥ 28 months	0.54	0.14 - 2.07	0.45	0.11 - 1.74
Birth interval	≥ 34 months	0.85	0.04 - 16.4	0.74	0.04 - 14.70
Breastfeeding[n]	≥ 34 months	0.47	0.05 - 4.70	0.45	0.04 - 4.69
Reference Categories:	a) SC/ST Hindu, b) Rural, c) Illiterate, d) No, e) No, f) Female, g) Dead, h) Not working, i) Not working, j) Illiterate, k) Kuccha, l) No, m) >= 2 km , n) Less than the given.				

Table 3: Univariate analysis and multivariate with extended cox model of second birth interval according to different variables in TN.

Variables	Categories	Univariate		Multivariate	
		exp(β)	C.I. 95%	exp(β)	C.I. 95%
Women's age at index child	Continuous			0.80	0.49 - 1.29
Women's age^2 at index child	Continuous			1.00	0.99 - 1.01
Religion/caste[a]	Non-Hindu	0.68	0.44 - 1.05		
	Other Hindu	0.67	0.36 - 1.26		
Place of residence[b]	Urban	0.70	0.48 - 1.03		
Women's education[c]	Primary	0.72	0.47 - 1.11		
	Middle	0.50	0.27 - 0.92		
	≥High school	0.46	0.25 - 0.83		
Ever contraceptive use[d]	Yes	0.65	0.44 - 0.97		
Ever fetal loss[e]	Yes	1.00	0.67 - 1.50		
Previous birth interval[f]	24-36 Months	0.87	0.58 - 1.31		
	≥ 36 Months	0.54	0.34 - 0.88		
Sex of index child[g]	Male	1.11	0.78 - 1.58		
Survival status of index child[h]	Alive	4.30	2.45 - 7.54	0.40	0.22 - 0.72
Women's occupation[i]	Working	1.01	0.70 - 1.47		
Husband's occupation[j]	Working	1.22	0.38 - 3.85		
Husband's education[k]	Primary	0.87	0.56 - 1.37	0.72	0.46 - 1.13
	Middle	0.90	0.52 - 1.56	0.62	0.35 - 1.11
	≥High school	0.41	0.24 - 0.72	0.34	0.19 - 0.61

Table 3: Continue.

Variables	Categories	Univariate		Multivariate	
		exp(β)	C.I. 95%	exp(β)	C.I. 95%
Type of house[l]	Pucca+Semi Pucca	0.94	0.65 - 1.36		
Media exposure[m]	Yes	0.82	0.53 - 1.25		
Distance primary health centre[n]	$<$ 2 km	0.68	0.47 - 0.99		
Birth interval	1-21 month	0.10	0.02 - 0.47	0.06	0.02 - 0.23
Breastfeeding[o]	$\geq$ 1 month	0.86	0.53 - 1.39	0.99	0.59 - 1.68
Birth interval	22-27 month	0.07	0.01 - 0.57	0.04	0.01 - 0.57
Breastfeeding[o]	$\geq$ 22 month	0.99	0.47 - 2.08	1.17	0.54 - 2.52
Birth interval	$\geq$ 28 month	1.61	0.17 - 15.4	0.48	0.04 - 5.94
Breastfeeding[o]	$\geq$ 28 month	0.55	0.07 - 4.47	1.07	0.10 - 11.0
Reference Categories:	a) SC/ST Hindu, b) Rural, c) Illiterate, d) No, e) No, f) $<$ 24 Month, g) Female, h) Dead, i) Not working, j) Not working, k) Illiterate, l) Kuccha, m) No, n) $\geq$ 2 km, o) Less than the given.				

Table 4: Univariate analysis and multivariate with extended cox model of third birth interval according to different variables in TN.

Variables	Categories	Univariate		Multivariate	
		exp(β)	C.I. 95%	exp(β)	C.I. 95%
Women's age at index child	Continuous			0.74	0.43 - 1.26
Women's age^2 at index child	Continuous			1.00	0.99 - 1.02
Religion/caste[a]	Non-Hindu	1.72	0.66 - 4.45		
	Other Hindu	1.52	0.41 - 5.66		
Place of residence[b]	Urban	1.20	0.60 - 2.38		
Women's education[c]	Primary	0.76	0.33 - 1.75		
	Middle	0.63	0.15 - 2.65		
	$\geq$High school	0.79	0.24 - 2.63		
Ever contraceptive use[d]	Yes	0.83	0.38 - 1.81		
Ever fetal loss[e]	Yes	0.50	0.22 - 1.14		
Previous birth interval[f]	24-36 Months	0.90	0.42 - 1.95		
	$\geq$ 36 Months	0.50	0.22 - 1.10		
Sex of index child[g]	Male	1.47	0.75 - 2.86		
Survival status of index child[h]	Alive	2.24	0.79 - 6.33		
Women's occupation[i]	Working	0.81	0.42 - 1.54		
Husband's occupation[j]	Working	0.49	0.07 - 3.61		
Husband's education[k]	Primary	1.45	0.62 - 3.40		
	Middle	0.91	0.30 - 2.79		
	$\geq$High school	1.13	0.42 - 3.02		
Type of house[l]	Pucca+Semi Pucca	1.10	0.57 - 2.13		
Media exposure[m]	Yes	0.53	0.26 - 1.07	0.50	0.24 - 1.02
Distance primary health centre[n]	$<$ 2 km	0.99	0.51 - 1.92		

Table 4: Continue.

Variables	Categories	Univariate		Multivariate	
		exp(β)	C.I. 95%	exp(β)	C.I. 95%
Birth interval	1-21 months	0.73	0.04 - 12.8	0.79	0.04 - 13.98
Breastfeeding^{o}	$\geq$ 1 months	1.73	0.23 -13.2	2.15	0.25 - 18.14
Birth interval	$\geq$ 22 months	1.63	0.15 - 17.4	1.50	0.14 - 16.16
Breastfeeding^{o}	$\geq$ 22 months	0.13	0.02 - 1.00	0.13	0.05 - 1.00

Reference Categories: a) SC/ST Hindu, b) Rural, c) Illiterate, d) No, e) No, f) < 24 Month, g) Female, h) Dead, i) Not working, j) Not working, k) Illiterate, l) Kuccha, m) No, n) $\geq$ 2 km, o) Less than the given.

4. Multivariate Analysis

The final models consisted of varying subsets of variables for first, second, and third birth intervals. Variables that entered partially are considered fully in the presentation of final models for a meaningful presentation. In order to account for age which is a well-known confounder, woman's age at index child was forced into the model. Square of woman's age at index child was also considered in order to overcome the problem of non-linear relationship.

The first variable to enter in the model for each birth interval was breastfeeding. Also, for each birth interval, breastfeeding alone improved the log likelihood up to a higher level, clearly showed the inclusion of breastfeeding even partially at first step itself significantly improved the model. High improvement in chi-square with one degree of freedom was seen for each birth interval, the improvement being 28.3 for first birth interval, 39.9 for second birth interval and 17.7 for third birth interval. Surprisingly, under the first birth interval, the effect of breastfeeding persisted only during the period 0-15 months. Also, effect of breastfeeding under the second birth interval disappeared during each of the periods considered in the analysis. However, its effect again persisted under the third birth interval during the period 22 and more months. It may be worth reporting that the role of breastfeeding fell in line with that reported based on univariate analysis.

Before comparison of variables entered in the final extended Cox models related to various birth intervals, it may be noted that subsets of variables considered in the data analysis vary from first birth interval to third birth interval because varying periods of classification of breastfeeding were considered. Strictly speaking, this may prohibit a comparison among the models. However, a qualitative comparison of results presented in Tables 2 to 4 reveals that high school (and above) education of women was a significant protective factor under the first birth interval analysis (HR: 0.40; C.I.: 0.22-0.75). On the other hand, high school (and above) education of father (HR: 0.34; C.I.: 0.19-0.61) and survival status of index child (HR: 0.40; C.I.: 0.22-0.72) were significant protective factors under the second birth interval analysis. Media exposure entered into the model for the third birth interval. Ever-fetal loss, survival status of index child, occupation of woman and husband's education also entered into the model for the first birth interval. Hence, variables that entered into the models varied from the first birth interval to the third birth interval.

5. Validation of the Models

Calibration curves for extended Cox models for the birth intervals are shown in Figures 1 to 3. Except for one group with extremely bad prognosis in each figure, bias corrected calibrations are very good. Shrinkage coefficients related to first to third birth interval are 0.90, 0.92 and 0.78, respectively (Table 5). This clearly reveals that 10%, 8% and 22% of the model fitting will be noisy in relation to first to third birth intervals, respectively. Thus, especially in case of third birth interval, the shrinkage coefficient could easily be used to shrink predictions to yield better calibration. Table 5 also shows that the discrimination accuracy in terms of the calculated Somer's D_{xy} rank correlation related to first to third birth interval are -0.56, -0.62 and -0.68, respectively. This index provides good predictive accuracy especially in case of third birth interval. In summary, these models are good enough to describe the parity specific birth intervals in Tamil Nadu.

TABLE 5: Validity indices of extended cox hazard models developed for parity specific birth intervals.

Shrinkage Coefficient and D_{xy}	Index Original	Training	Test	Optimism	Index Corrected	Resample
Parity-I	1.00	1.00	0.90	0.10	0.90	200
	−0.58	−0.59	−0.57	−0.02	−0.56	200
Parity-II	1.00	1.00	0.92	0.08	0.92	200
	−0.63	−0.64	−0.62	−0.01	−0.62	200
Parity-III	1.00	1.00	0.78	0.22	0.78	200
	−0.72	−0.74	−0.70	−0.04	−0.68	200

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

6. Prediction from the Final Models

Prediction from the final model may be used to provide important clues to policy planners through predicted survival probabilities at considered level of a variable by holding all other variables at their average level in the model. In the present prediction analysis, the possible selected variables(s) and some combination of variables are: women's primary education; women's middle education; women's high school (and above) education; survival of index child; working women; husband's education of high school (and above); media exposure; women's high school (and above) education and survival index of child; and women's high school (and above) education and husband's high school (and above) education. On account of varying subsets of variables in the models, only results possible under each model are presented in the Tables 6 to 8 that deal with first to third birth intervals.

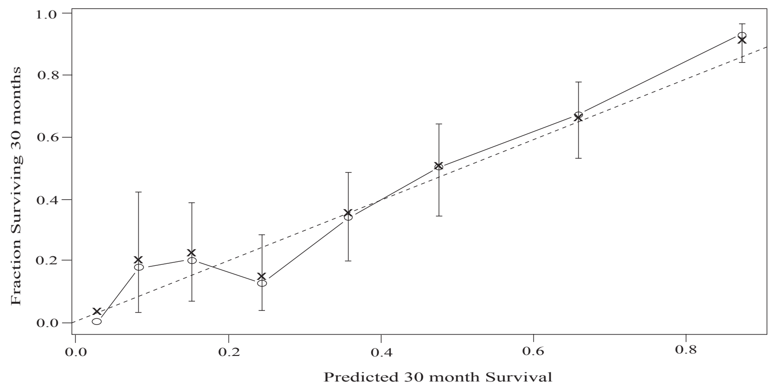

Figure 1: Bootstrap estimates of calibration accuracy for 30 months estimates from the final extended Cox model for 1st birth interval. Dots correspond to apparent predictive accuracy. X marks the bootstrap-corrected estimates.

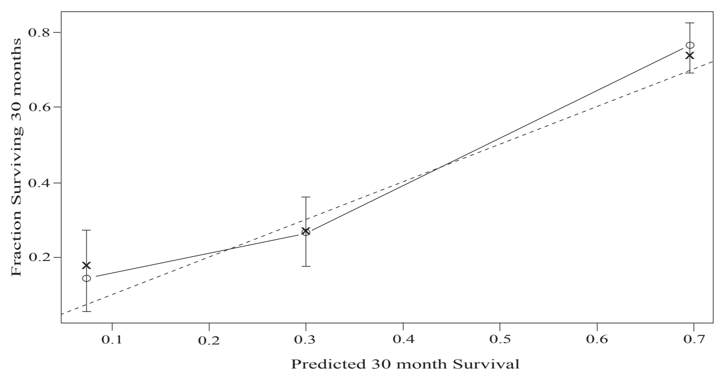

Figure 2: Bootstrap estimates of calibration accuracy for 30 months estimates from the final extended Cox model for 2nd birth interval. Dots correspond to apparent predictive accuracy. X marks the bootstrap-corrected estimates.

There is a decreasing trend in probability of not having next child over a period of time related to each birth interval. This is more evident in relation to first birth interval. There is no specific trend with increasing period of breastfeeding. However, within each category, there is an increasing trend in not having next child probability in relation to increasing education of women. Women's high school (and above) education was noticed to provide maximum benefit. This is in further evidence if women have a surviving index child.

Very few predictions were possible in relation to second and third birth intervals (Tables 7-8). High school (and above) education of husband provided the maximum benefits up to the category 22-27 months under the second birth interval. Similar results were obtained in relation to survival of index child. Surprisingly, these probabilities were lower during the period of 28 and more months. Under third birth interval, prediction was possible only in relation to media exposure.

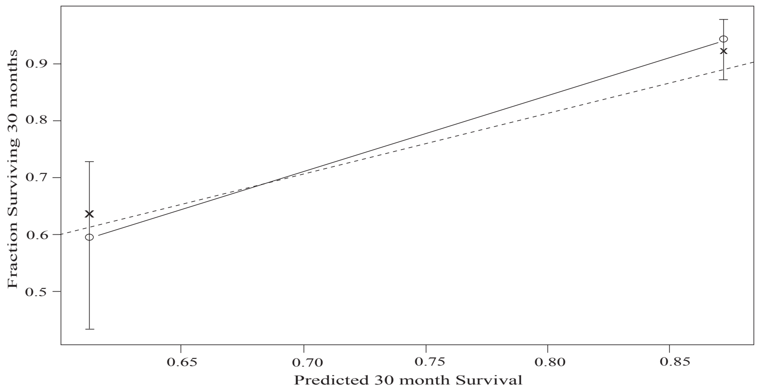

FIGURE 3: Bootstrap estimates of calibration accuracy for 30 months from the final extended Cox model for 3rd birth interval. Dots correspond to apparent predictive accuracy. X marks the bootstrap-corrected estimates.

This was not possible under the first and second birth intervals. Interestingly, media exposure showed maximum benefit during 22 and more months of breastfeeding (Table 8).

Table 6: Estimated probabilities of not having second live birth at specific months after first live birth, by selected characteristics, according to model (Ist Birth Interval).

Characteristics	Probability of not having births at months						
	12	18	24	30	36	42	48
Breastfeeding (0-15 months)							
Average	0.99	0.97	0.92	0.87	0.82	0.73	0.68
Primary educated women	0.99	0.96	0.91	0.85	0.79	0.70	0.64
Middle educated women	0.99	0.96	0.91	0.85	0.80	0.71	0.65
High school and above educated women	1.00	0.98	0.96	0.93	0.90	0.86	0.83
Index child alive	0.99	0.97	0.92	0.87	0.82	0.74	0.69
Working women	0.99	0.97	0.94	0.89	0.85	0.78	0.74
High school and above educated husband	0.99	0.97	0.93	0.88	0.84	0.77	0.72
High school and above educated women + index child alive	0.99	0.99	0.97	0.94	0.92	0.88	0.85
High school and above educated women & husband	1.00	0.98	0.96	0.94	0.91	0.86	0.83
Breastfeeding (16-21 months)							
Average	0.99	0.96	0.92	0.86	0.81	0.73	0.67
Primary educated women	0.99	0.96	0.90	0.84	0.78	0.69	0.63
Middle educated women	0.99	0.96	0.91	0.85	0.79	0.70	0.64
High school and above educated women	1.00	0.98	0.96	0.93	0.90	0.86	0.82
Index child alive	0.99	0.97	0.92	0.87	0.82	0.74	0.69
Working women	0.99	0.97	0.93	0.89	0.85	0.79	0.73
High school and above educated husband	0.99	0.97	0.93	0.88	0.83	0.76	0.71
High school and above educated women + index child alive	1.00	0.98	0.96	0.93	0.91	0.86	0.83
High school and above educated women & husband	1.00	0.98	0.96	0.94	0.92	0.87	0.85

Table 6: Continue.

Characteristics	Probability of not having births at months						
	12	18	24	30	36	42	48
Breastfeeding (22-27 months)							
Average	0.99	0.96	0.87	0.79	0.72	0.61	0.67
Primary educated women	0.98	0.94	0.85	0.76	0.68	0.56	0.48
Middle educated women	0.98	0.94	0.86	0.77	0.69	0.57	0.50
High school and above educated women	1.00	0.97	0.94	0.90	0.85	0.79	0.83
index child alive	0.99	0.95	0.88	0.80	0.73	0.62	0.55
Working women	0.99	0.96	0.90	0.83	0.77	0.67	0.61
High school and above educated husband	0.99	0.95	0.89	0.82	0.75	0.65	0.58
High school and above educated women + index child alive	0.99	0.97	0.94	0.90	0.86	0.79	0.75
High school and above educated women & husband	0.99	0.98	0.94	0.91	0.87	0.81	0.77
Breastfeeding (28-33 months)							
Average	0.99	0.96	0.92	0.86	0.81	0.73	0.67
Primary educated women	0.99	0.96	0.90	0.84	0.78	0.69	0.63
Middle educated women	0.99	0.96	0.91	0.85	0.79	0.70	0.64
High school and above educated women	1.00	0.98	0.96	0.93	0.90	0.86	0.83
Index child alive	0.99	0.97	0.92	0.87	0.82	0.73	0.68
Working women	0.99	0.97	0.93	0.89	0.85	0.78	0.73
High school and above educated husband	0.99	0.97	0.93	0.88	0.83	0.76	0.71
High school and above educated women + index child alive	1.00	0.98	0.96	0.94	0.91	0.87	0.84
High school and above educated women & husband	1.00	0.98	0.96	0.93	0.90	0.86	0.83
Breastfeeding ($\geq$ 34 months)							
Average	0.99	0.95	0.88	0.81	0.73	0.63	0.56
Primary educated women	0.98	0.94	0.86	0.78	0.70	0.58	0.51
Middle educated women	0.99	0.94	0.87	0.79	0.71	0.59	0.52
High school educated women	0.99	0.97	0.94	0.90	0.86	0.80	0.76
Index child alive	0.99	0.95	0.89	0.81	0.74	0.64	0.57
Working women	0.99	0.96	0.90	0.84	0.78	0.69	0.63
High school and above educated husband	0.99	0.96	0.90	0.73	0.77	0.67	0.61
High school and above educated women + index child alive	0.99	0.98	0.94	0.90	0.86	0.80	0.76
High school and above educated women & husband	0.99	0.98	0.94	0.90	0.86	0.80	0.76

Table 7: Estimated probabilities of not having third live birth at specific months after second live birth in TN, by selected characteristics, according to model (II^{nd} Birth Spacing).

Characteristics	Probability of not having births at months						
	12	18	24	30	36	42	48
Breastfeeding (16-21 months)							
Average	1.00	1.00	0.99	0.98	0.98	0.97	0.95
High school and above educated women	1.00	1.00	0.99	0.99	0.99	0.98	0.97
Index child alive	1.00	1.00	0.99	0.98	0.98	0.97	0.95
Breastfeeding(22-27 months)							
Average	1.00	1.00	1.00	0.99	0.99	0.99	0.98
High school and above husband	1.00	1.00	1.00	1.00	1.00	0.99	0.99
Index child alive	1.00	1.00	1.00	0.99	0.99	0.99	0.98
Breastfeeding ($\geq$ 28 months)							
Average	0.99	0.91	0.79	0.68	0.62	0.50	0.33
High school and above husband	0.99	0.95	0.87	0.80	0.76	0.68	0.53
Index child alive	0.99	0.91	0.80	0.69	0.63	0.52	0.35

Table 8: Estimated probabilities of not having fourth live birth at specific months after third live birth in TN, by selected characteristics, according to model (III^{rd} Birth Spacing).

Characteristics	Probability of not having births at months						
	12	18	24	30	36	42	48
Breastfeeding (0-21 months)							
Average	0.99	0.95	0.90	0.83	0.78	0.71	0.71
Media exposure	0.99	0.96	0.91	0.85	0.81	0.75	0.75
Breastfeeding ($\geq$ 22 months)							
Average	1.00	0.99	0.97	0.95	0.94	0.91	0.91
Media exposure	1.00	0.99	0.98	0.96	0.95	0.93	0.93

7. Discussion

To our knowledge, there is no study on birth interval, in which an exercise related to the validity of the developed Cox hazards models has been carried out. Therefore, there is no scope to compare the developed models in the present study with those reported under other studies, especially with regard to validity of the models. However, if necessary, one could examine the reported likelihood values for models under other studies to provide for a comparison with models developed in the present study.

Breastfeeding is the only covariate, which is noticed to be a significant protective factor associated with each birth interval. Education of women was significantly associated with first birth interval only while husband's education was significantly associated with first and second birth interval. Survival status of index child emerged as an important associated factor at second birth interval only while fetal loss was associated at the first birth interval. However, contraceptive use did not emerge as a significant associated factor at any birth interval.

Breastfeeding is the most important and significant factor for extending the birth interval at all the parities in TN. Trussell et al. (1985) in a study done in the Philippines, Malaysia, and Indonesia also found breastfeeding beyond 11 months to be a significant protective factor on birth interval. Anderson & Bean (1985) also support the relation between ever breastfeeding and exclusive breastfeeding and birth interval. Thus, though the relation between breastfeeding and birth interval is already documented, this study is able to predict the precise nature of this effect.

Education of woman high school (and above) was a protective factor for the first birth interval. This finding is supported by Rajaram, Rao & Pandey (1994) and Gandotra, Retherford, Pandey, Luther & Mishra (1998), who found that education of woman led to reduction in fertility, probably due to increase in awareness and choice. In contrast, Rodriguez, Hobcraft, McDonald, Menken & Trussell (1984) found little association between education and birth interval except at higher parities. Ojha (1998) and Richter, Podhisita, Chamratrithirong & Soonthorndhada (1994) support Rodriguez's findings. A similar reasoning can be attributed to the fact that media exposure had a protective impact on higher order birth interval. This finding is supported by Gandotra et al. (1998).

This study has clearly indicated that working status of women was a significant protective factor specifically for the first birth interval in TN. While this is not supported by Trussell et al. (1985), Richter et al. (1994) noticed that women employed as salesgirls and manufacturing laborer and in self-employment were significantly less likely to go for the next birth.

Ojha (1998) and Blanchard & Bogaert (1997) reported that birth intervals are comparatively longer following the birth of a male in comparison to female child. However, sex of index child did not emerge as a significant associated factor. The present study did not indicate the likelihood of it being a protective factor at any birth interval.

Survival of index child emerged as a significant protective factor for the first and second birth intervals. This was in line with many other studies like Oheneba-Sakyi & Heaton (1993); Rehman & DaVanzo (1993) Rajaram et al. (1994); Ojha (1998); and Palloni & Hantamala (1999). This shows that this factor is not a country or region specific determinant.

Other documented factors such as contraceptive use (Rajaram et al. 1994, Mahmud & Islam 1995), place of residence (Swenson & Thang 1993) and importance of previous birth interval in extending succeeding birth intervals (Rodriguez et al. 1984, DaVanzo & Starbird 1991, Miller, Trussell, Pabley & Vaughan 1992, Swenson & Thang 1993, Trussell et al. 1985) were not supported in this study.

8. Limitations

First National Health and Family Survey (NFHS) was conducted in 1992-93 in India and data was available to use in 1995. However, no study is available till now on breastfeeding as a time varying covariate with time dependent effect using bootstrap technique for validations and predictions. These techniques have been used for the first time on birth interval data. Our internet search has not revealed any similar study. Therefore, we felt that the study has valuable information for strategic and policy planners and gives more occasion for readership.

9. Conclusion

This study showed that subsets of important covariates, which entered into the final models, varied among the birth intervals within the state. However, the assessment of predictive accuracy clearly established the suitability of the parity specific developed models in describing respective birth interval. Thus, the present study emphasizes the need for regional studies in planning public health programs as per needs of the region. Further, this study also demonstrates the importance of parity specific analysis of birth interval and may assist in working out parity specific strategies in the considered region. Breastfeeding emerged as an important protective covariate that extended the birth interval irrespective of parity. Further, education of women, sex of index child, husband's education, and media exposure also demonstrated an important protective role for extending birth interval in the study.

[Recibido: agosto de 2011 — Aceptado: marzo de 2012]

References

Anderson, D. L. & Bean, L. L. (1985), 'Birth spacing and fertility limitation: A behavioral analysis of nineteenth century populationl', *Demography* **22**, 169–183.

Blanchard, R. & Bogaert, A. F. (1997), 'Additive effects of older brothers and homosexual brothers in the prediction of marriage and cohabitation', *Behavior Genetics* **27**, 45–54.

Cox, D. R. (1972), 'Regression models and life tables (with Discussion)', *Journal of the Royal Statistical Society, Series B* **34**, 187–220.

DaVanzo, J. & Starbird, E. (1991), 'Correlates of short interbirth intervals in peninsular Malaysia: Their pathways of influence through breastfeeding and contraceptive use', *Studies in Family Planning* **22**(4).

Dickson, E., Grambsch, P., Fleming, T., Fisher, L. & Langworthy, A. (1989), 'Prognosis in primary biliary cirrohsis: Model for decision making', *Hepatology* **10**(1), 1–7.

Dwivedi, S. & Rajvir, S. (2003), 'On assessing the child spacing effect of breastfeeding using cox proportional hazards model with nfhs data', *Demography India* **32**(2), 215–224.

Efron, B. & Tibshirani, R. (1993), *An Introduction to the Bootstrap*, Chapman and Hall, New York.

Fox, J. (2008), *Applied Regression Analysis, Linear Models, and Related Methods*, Sage Publications, Inc., London.

Gandotra, M., Retherford, R., Pandey, A., Luther, N. & Mishra, V. (1998), Fertility in India, National Family Health Survey Subject Reports 9, Mumbai: International Institute for Population Sciences; and Honolulu.

Harrell, F. E. (2001), *Regression Modeling Strategies with Application to Linear Models, Logistic Regression, and Survival Analysis*, Springer-Verlag, Berlin.

Harrell, F. E., Lee, K. L. & Mark, D. B. (1996), 'Tutorial in Biostatistics Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and meassuring and reducing errors', *Statistics in Medicine* **15**, 361–387.

International Institute for Population Sciences (1994), National Family Health Survey (MCH and Family Planning), Tamil Nadu 1992, Summary report, Population Research Centre, The Gandhigram Institute of Rural Health and Family Welfare Trust, Ambathurai R.S. (PRC, Gandhigram), and International Institute for Population Sciences (IIPS), Bombay, India.

International Institute for Population Sciences (1995), National Family Health Survey (MCH and Family Planning): India 1992-93, Summary report, International Institute for Population Sciences (IIPS), Bombay, India.

Kaplan, E. & Meier, P. (1958), ' Nonparametric estimation from incomplete observations', *Journal of the American Statistical Association* **53**, 457–481.

Kleinbaum, D. G. (1996), *Survival Analysis, A Self Learning Text*, 1 edn, Springer-Verlag, New York.

Mahmud, M. & Islam, M. (1995), 'Adolescent contraceptive use and its determinants in Bangladesh: Evidence from Bangladesh Fertility Survey 1989.', *Contraception* **52**(3), 181–186.

Miller, J. E., Trussell, J., Pabley, A. R. & Vaughan, B. (1992), 'Birth spacing and child mortality in Bangladesh and the Philippines', *Demography* **29**(2), 305–316.

Namboodiri, K. & Suchindran, C. M. (1987), *Life Table Techniques and Their Applications Studies in Population*, Academic Press, Orlando, Florida.

Oheneba-Sakyi, Y. & Heaton, T. B. (1993), 'Effects of socio-demographic variables on birth intervals in Ghana', *Journal of Comparative Family Studies* **24**(1), 113–135.

Ojha, A. (1998), 'The effect of sex preference on fertility in selected states of India', *The Journal of Family Welfare* **44**(1), 42–48.

Palloni, A. & Hantamala, R. (1999), 'The effects of infant mortality on fertility revisited: New evidence from Latin America', *Demography* **36**(1), 41–75.

Rajaram, S., Rao, S. & Pandey, A. (1994), 'Birth interval dynamics in Goa: A parity specific analysis', *Demography India* **23**(1), 67–81.

Rehman, M. & DaVanzo, J. (1993), ' Gender preference and birth spacing in Matlab, Bangladesh', *Demography* **30**(3), 315–332.

Richter, K., Podhisita, C., Chamratrithirong, A. & Soonthorndhada, K. (1994), 'The impact of child care on fertility in urban Thailand', *Demography* **31**(4), 651–662.

Rodriguez, G., Hobcraft, J., McDonald, J., Menken, J. & Trussell, J. A. (1984), 'Comparative analysis of the determinants of birth intervals', *WFS Comparative Studies* (30).

Singh, R., Begum, S., Ahuja, R. K., Chandra, P. & Dwivedi, S. N. (2007), 'Prediction of child survival in India using developed Cox PH model: a utility for health policy programmers', *Statistics in Transition* **8**(1), 97–110.

Srinivasan, K. (1980), *Birth Interval Analysis in Fertility Surveys*, Scientific reports (World Fertility Survey), Voorburg and International Statistical Institute, Voorburg and London.

Swenson, I. & Thang, N. M. (1993), 'Determinants of birth intervals in Vietnam: A hazard model analysis', *Journal of Tropical pediatrics* **39**, 163–167.

Trussell, J. & Charles, H. (1983), 'A hazards model analysis of the covariates of infant and child mortality in Sri Lanka', *Demography* **20**(1), 1–24.

Trussell, J., Martin, L., Fledman, R., Palmore, J., Concepcion, M. & Abu Bakar, D. (1985), 'Determinants of birth interval length in the Phillipines, Malaysia and Indonesia: A hazard model analysis', *Demography* **22**(2).

UNFPA (1997), Reproductive Rights, Reproductive Health and Family Planning, Population issues, United Nations Fund for Population Activities (UNFPA).

Van Houwelingen, J. C. & Cessie, S. (1990), 'Predictive value of statistical models', *Statistics in Medicine* **8**, 1303–1325.