THE EGYPTIAN POPULATION AND
FAMILY PLANNING REVIEW.
ISSR,CAIRO UNIV. Vol.33,No.1,2000

HAZARD MODEL ANALYSIS:
SOME APPLICATIONS IN DEMOGRAPHY

By

Farouk Shoieb

1999

I Introduction

In this section, we shall present definitions, notation, and basic facts used throughout.

1.1 Preliminaries

Let T be the waiting time to the occurrence of an event of interest: death, birth, pregnancy, etc.). **T is usually called the survival time.** The distribution of T can be characterized by the following three equivalent functions.

1- **Survivorship Function** (or Simply, Survival Function). This function, denoted by S(t), is defined as the probability that an individual survives longer than t:

$$\begin{aligned} S(t) &= P\text{ (an individual survives longer than t)} \\ &= P(T > t) = 1\text{-}F(t) \qquad (1.1) \end{aligned}$$

In practice, the survivorship function is estimated as the proportion of patients surviving longer than t:

$$\hat{S}(t) = \frac{\text{Number of patients surviving longer than t}}{\text{Total number of patients}} \qquad (1.2)$$

Where the symbol "^" denotes an estimate of the function.

The function S(t) is also known as the cumulative survival rate. To depict the course of survival, **Berkson (1942)** recommended a graphic presentation of S(t). The graph of S(t) is called the survival curve.

2- **Probability Density function** (or simply, density function). Like any other continuous random variable, the survival time T has a probability density function defined as the limit of the probability that an individual fails in the short interval t to t + Δt per unit width Δt, or simply the probability of an event in a small interval per unit time. It can be expressed as:

$$f(t) = \lim_{\Delta t \to 0} \frac{P\{\text{a patient dying in the interval } (t, t+\Delta t)\}}{\Delta t} \qquad (1.3)$$

In practice, the probability density function f(t) is estimated as the proportion of events in an interval per unit width:

$$\hat{f}(t) = \frac{\text{Number of patients dying in the interval begining at time t}}{(\text{Total number of patients})\ (\text{Interval width})} \qquad (1.4)$$

3. **The hazard function** h(t) of survival time T gives the conditional failure rate. This is defined as the probability of failure during a very small time interval, assuming that the individual has survived to the beginning of the interval,i. e.,

$$h(t) = \lim_{x \to 0} \frac{1}{x} P(t < T \le t + x \mid T > t)\ .$$

The hazard function can also be defined in terms of the cumulative distribution function F(t) and the probability density function f(t) as:

$$h(t) = f(t) / \{1 - F(t)\} \qquad (1.5)$$

In practice, the hazard function is estimated as the proportion of patients dying in an interval per unit time, given that they have survived to the beginning of the interval:

$$\hat{h}(t) = \frac{\text{Number of patients dying in the interval begining at time t}}{(\text{Number of patients surviving at t})(\text{Interval width})}$$

$$= \frac{\text{Number of patients dying per unit time in the interval}}{\text{Number of patients surviving at t}} \qquad (1.6)$$

The hazard function is also known as the instantaneous failure rate, force of mortality and conditional mortality rate. The hazard function thus gives the risk of failure per unit time during the aging process. It plays an important role in survival data analysis. The hazard function may increase, decrease, remain constant or indicate a more complicated process.

The statistical analysis of lifetime or response time data has become a topic of considerable, interest to statisticians and workers in different areas such as engineering, medicine, biological and demographic sciences.

In recent years, the use of statistical procedures has thrown substantial research which uses event history models, survival analysis, and hazard rate models to examine sociological phenomena such as retirement, unemployment, and teenage pregnancy, birth interval, marriage duration, mortality, migration and reproductive health studies.

Thus, **failure time analysis and hazard models** have become popular in demographic studies. It can be viewed as a part of regression analysis with limited dependent variables as well as a special case of event history analysis and multistage demography. The idea of hazard function and failure time analysis, however, have not been properly introduced nor commonly discussed by demographers until the celebrated **Cox (1972)** was published.

1.2. **The Proportional Hazard Regression Model (Cox, 1972)**

Studies of association between a random variables X and, the survival time T, may only be partially observable due to censoring has been the focus of many investigations starting with the historical breakthrough by **Cox (1972).** The so-called **Cox regression model or proportional hazards model** (PHM) expresses a log-linear relation between X and the hazard function of T:

$$h(t|\ X=x)=\lim_{\delta\downarrow 0}\frac{P(t\le T\le t\ +\ \delta|\ T\ge t, X=x)}{\delta}$$

$$=h_0(t)\ e^{\beta x}$$

In this model, $h_0(t)$ is an unspecified baseline hazard, i.e., hazard at x=0, and β is an unknown regression coefficient. **Cox (1972, 1975)** mentioned the following important point regarding PHM. $h_o(t)$

might be identically zero in some time intervals in which no events (failures) occur. Thus, in some cases, one has to argue **conditionally** on the set of instances at which failures occur.

The Proportional Hazard model, introduced by **Cox (1972)**, is a general non-parametric model appropriate for the analysis of survival data with and without censoring. In the following we first introduce **Cox's** original model and then discuss some of its extensions to bring the reader up to date.

Suppose that on each of the n individuals involved in the study, in addition to the survival time t_i, one or more measurements are available, on variables $x_1, x_2, \ldots x_p$ (Cox calls these explanatory variables). For the ith individual let values of the p variables be $x_{1i}, x_{2i}, \ldots, x_{pi}$. The x's may be specific patient characteristics, such as age and white cell count, or functions of time. The problem is to assess the relation between the distribution of survival time t and the x's. **Cox** suggests that the **hazard function** be used. Let $h_i(t)$ be the <u>hazard</u> function of the ith patient. When survival times are continuously distributed ties can be ignored, and the hazard function is

$$h_i(t) = h_0(t) \exp\left(\sum_{j=1}^{p} \beta_j X_{ji}\right) \qquad (1.7)$$

Where $h_0(t)$ is the hazard function of the underlying survival distribution (arbitrary) when all the x variables are ignored, i.e., all x's equal zero, and the β's are regression coefficients. In fact, $\sum_{j=1}^{p} \beta_j X_{ji}$ can be replaced by any known function of x's and β's. It is clear that Cox's model assumes that the hazard of the study group is proportional to that of the underlying survival distribution $h_0(t)$.

Equation (1.7) has many uses. Some of which are summarized below.

1- <u>**Two-Sample Problems**</u>. Suppose that p = 1, i.e., there is only one X variable, X_1 which is an indicator variable,

$$X_{1i} = 0 \text{ if the ith individual is from sample 1}$$
$$= 1 \text{ if the ith individual is from sample 2}$$

Then according to Equation (1.7) the hazard function of samples 1 and 2 are, respectively, $h_0(t)$ and $h_0(t) \exp(\beta_1)$. The hazard function of sample 2 is equal to the hazard function of sample 1 multiplied by a constant $\exp(\beta_1)$ or the two hazard functions are proportional. In terms of the survivorship function:

$$S_2(t) = c.[\, S_1(t)\,],$$

Where the constant $c = \exp(\beta_1)$ **(Nadas, 1970).**

2- **Two-Sample Problems with covariates.** The x variables in Equation (1.7) can either be indicator variables such as x_1 in the two - sample problem above or concomitant variables (patient characteristics). Having one or more x variables representing concomitant variables in Equation (1.7) enables us to examine the relation between two samples adjusting for the presence of concomitant variables.

3- **Two-Sample Problems with Time – Dependent Covariate.** In Equation (1.7) one – or more x variables can be functions of time. For example, suppose that, in addition to x_1 above, time dependent variable, $x_2 = tx_1$, is introduced. According to Equation (1.7), the hazard in sample 2 is

$$h_i(t) = h_0(t) \exp(\beta_1 + \beta_2 t)$$
$$= ch_0(t) \exp(\beta_2 t)$$

and that in sample 1 remains $ch_0(t)$.

4- **Regression Problems**. Dividing both sides of Equation (1.7) by $h_0(t)$ and taking logarithm, we obtain

$$\log_e \frac{h_i(t)}{h_0(t)} = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots \beta_p x_{pi} = \sum_{j=1}^{p} \beta_j x_{ji} \qquad (1.8)$$

The left- hand side of Equation (1.8) is a function of the hazard for the ith patient, and the right-hand side is a linear combination of the concomitant variables $x_{1i}, \ldots, x_{pi}$ with coefficients $\beta_1, \ldots, \beta_p$ respectively. The X's can be indicator variables, covariates, and time- dependent covariates. If we let $Y_i = \log_e [h_i(t) / h_0(t)]$ Equation (1.8) is simply

$$Y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} \qquad (1.9)$$

which is a standard multiple regression equation with the concomitant variables as independent variables and a function of the hazard as the dependent variable.

In addition to identifying important prognostic factors, **Cox's** regression model can also define a prognostic index or hazard ratio, namely $\log_e [h_i(t) / h_0(t)]$, for each patient. This index or ratio can be used to compare two treatment groups as well as prognoses between patients. As mentioned earlier, $h_0(t)$ is the hazard function when all of the independent variables are ignored. If the independent variables are standardized about the mean, the following model is used.

$$\log_e \frac{h_i(t)}{h_0(t)} = \beta_1(x_{1i} - \bar{x}_1) + \ldots + \beta_p(x_{pi} - \bar{x}_p) \qquad (1.10)$$

where $\overline{X}$ the average of the ith independent variable for all patients, then $h_0(t)$ is the hazard function when all variables are at their average values.

5-**Estimation of the parameters**. Now the question is how to estimate the coefficients, $\beta_1, \ldots, \beta_p$. **Cox** suggests a maximum likelihood procedure where the likelihood function is based on a conditional probability of failure. Suppose that $t_{(1)} < t_{(2)} < \ldots t_{(k)}$ are the k exact failure times . Let $R(t_{(i)})$ be the risk set at time $t_{(i)}$. ($R(t_{(i)})$ consists of all individuals whose survival times are at least $t_{(i)}$), then the log of the likelihood function is:

$$LL(\beta) = \sum_{i=1}^{k} \sum_{j=1}^{p} \beta_j x_{ji} - \sum_{i=1}^{k} \log\left[\sum_{l \in R(t_{(i)})} \exp\left(\sum_{j=1}^{p} \beta_j x_{jl}\right)\right] \qquad (1.11)$$

Suppose that among the survival times $t_1, \ldots, t_n$ there are k distinct times. Let $t_{(1)} < \ldots < t_{(k)}$ be the k distinct failure times (uncensored observations). Let $m_{(i)}$ be the multiplicity of $t_{(i)}$, $m_{(i)}>1$ if there is more than one observation with $t_{(i)}$, $m_{(i)} = 1$ if there is only one observation with value $t_{(i)}$. Let $R(t_{(i)})$ denote the set of individual at risk at time $t_{(i)}$. Let $r_{(i)}$ be the number of such individuals.

The conditional log-likelihood function is then

$$LL(\beta) = \sum_{i=1}^{k} (\beta_1 z_{1i} + \ldots + \beta_p z_{pi}) - \sum_{i=1}^{k} \log\left[\sum_{l \in R(t_{(i)})} \exp(\beta_1 z_{1l} + \ldots + \beta_p z_{pl}) \right] \quad (1.12)$$

where z_{1i} is the sum of x_{1i}'s over the $m_{(i)}$ individuals failing at $t_{(i)}$, z_{2i} is the sum of x_{2i}'s over the $m_{(i)}$ individuals failing at $t_{(i)}$.

1.3 General Historical Developments

Cox's arguments, particularly the conditional likelihood function (Equation (1.11)) and the logistic model for discrete case, are more intuitive than formal. They have been formalized when no covariates depend on time in terms of a rank- like procedure by **Kalbfleisch and Prentice (1973)** and in terms of an approximate likelihood by **Breslow (1974)** . The rank-like procedure is not applicable in the case of time-dependent covariate, but an adaptation of **Breslow's** procedure is possible **(Crowley and Hu, 1977).** Three particular cases can be identified:

(a) <u>Uncensored Data without Ties.</u> Suppose that $t_1, t_2, \ldots, t_n$ are the survival times of n individuals and $x_1, \ldots, x_n$ the corresponding covariate values. Let $t_{(1)} < t_{(2)} < t_{(n)}$ be the n survival times rearranged in ascending order of magnitude. The $t_{(1)}, t_{(2)} \ldots t_{(n)}$ are called ordered survival times . The ranks corresponding to these order statistics i.e., (1), (2) , … , (n) are rank statistics . On the basis of rank statistics, **Kalbfleisch and Prentice** derive the following marginal likelihood of β from the original distribution of the marginal distribution of the rank

$$\left(\beta\sum_{i=1}^{n} x_i\right) / \prod_{i=1}^{n}\left[\sum_{\iota\in R(t_{(i)})} \exp(\beta x_\iota)\right] \quad (1.13)$$

(b) Censored Data without Ties. When there are censored observations in the sample, the rank statistics are not completely available. Suppose, for example, that there are four survival times 1, 2+, 3, 4. We may order the uncensored times as $1 < 3 < 4$, however 2+ may be equal to or greater than 3, or may be equal to or greater than 4. Thus, we have only partial information on the ranks. The structure of the model becomes more complicated.

Suppose that k individuals are observed to failure and the ordered survival times are $t_{(1)} < t_{(2)} < \ldots < t_{(k)}$ with corresponding covariates $x_1, x_2, \ldots, x_k$. Suppose further that q_i individuals with covariates $x_{i1}, \ldots, x_{iqi}$ are censored between $t_{(1)}$ and $t_{(i+1)}$, where $t_{(0)} = 0$ and $t_{(k+1)}, = \infty$. The marginal likelihood function in this case is.

$$L_2(\beta) = \exp\left(\beta\sum_{i=1}^{n} x_j\right) / \prod_{i=1}^{n}\left[\sum_{\iota\in R(t_{(i)})} \exp(\beta x_j)\right] \quad (1.14)$$

Where $R(t_{(i)})$ consists of individuals whose survival times, censored and uncensored, are at least $t_{(i)}$. That is, $R(t_{(i)})$ includes the t's whose subscripts are $(i), i_1, \ldots i_{qi}, (i+1), \ldots, (k), k_1, \ldots, k_{qk}$. The logarithm of the marginal likelihood function in equation (1.13) is exactly the same as **Cox's** is Equation (1.11)

(c) Data With Ties. When data include tied observations, **Cox's** likelihood is derived by consideration of logistic model for discrete survival time. The likelihood appears to be a statement of inference about regression coefficients in the logistic model rather than in model (1.7). **Kalbfleisch and Prentice (1980b)** derive a different likelihood to cover this case.

Breslow (1974) suggests another approach to the estimation of β and $h_0(t)$. It differs from both **Kalbfleisch and Prentices** arguments and those of **Cox** in that simultaneous estimation of β and $h_0(t)$ is made through a joint likelihood function. The underlying survival distribution is assumed continuous having constant hazard $h_i = \exp(a_i)$ between each pair $(t_{(i)})$, $(t_{(i+1)})$ of distinct failure times. All censored observations that occur in the interval $(t_{(i)})$, $(t_{(i+1)})$ are assumed to have occurred at $(t_{(i)})$.

Log – likelihood function is

$$L_3(\beta) = \sum_{i=1}^{k} \left[\beta z_i - m_i \log \sum_{j \in R(t_{(i)})} \exp(\beta x_j) \right] \qquad (1.15)$$

Where k is the number of distinct failure times. When there are no ties ($m_i = 1$ for $i = 1, \ldots, k$), Equation (1.15) is equivalent to the likelihood set forth by **Cox and kalbfleisch and Prentice.** Otherwise, Equation (1.15) is an approximation of Cox's.
Kalbfeisch and Prentice (1979) justified the use of partial-likelihood functions under the assumption of no ties and **Breslow (1974 and 1975)** tried with the piecewise exponential baseline. **Tsiates (1981)** gives a proof of the asymptotic normality of $\hat{\beta}$ (see also **Efron, 1977).**

On practical level, the **piecewise proportional hazards model** is familiar to demographers and suitable for demographic analysis given the nature of demographic data **(Allison 1982).** Demographic analysis typically involves very large data sets, and the information on time is available only in unit interval, such as months or years . It is therefore, easy to group time into a number of segments and to handle efficiently the computation required by very large data sets using the piecewise proportional hazards model.

Moreover, with the **piecewise proportional hazard model,** it is easy to separate the effects of covariates on the probability and the timing of an event as suggested by **Yamaguchi (1995).**

We point out that, in a review paper, **Kumar and Klefsjo (1994)** surveyed the existing literature on the proportional hazard model (PHM). At first, the characteristics of the method are explained and its importance in **reliability analysis** is presented. Subsequently, methods for estimating parameters, along with the small and large sample properties of the estimators, are briefly discussed. They also described some possible extensions of the model considered so far and available computer programs and packages for estimating the parameters of these models. However, the article did not cover the developments and applications associated with this model to **DEMOGRAPHY**.

Teachman and Hayward (1993) in this regard published an excellent article. In this article, basic hazard rate models are surveyed, and survival functions and their relationship to hazard rates are described. Ways in which survival functions can be used to expand the information provided by hazard rates (from relatively simple to more complex hazard rate models) are illustrated. Away from **Teachman (1993)** article, the reader might be interested in going back in time looking for specific results (both on the theoretical and applied levels). Therefore, the objective of the present paper is to survey the most recent developments and applications contributed thus far to demographic sciences and data analysis.

The historical developments presented above have paved the way towards the following sections.

II. Theoretical Developments

In this section, we present most recent theoretical results, which followed **Cox's (1972)** model, starting 1982.

1. **Andersen and Gill (1982)** discussed how the **Cox's** model can be extended to a model where covariate processes have a proportional effect on the intensity process of a multivariate counting process. They considered the large sample properties of a counting process model with intensity given by:

$$\lambda(t;z) = \lambda_0(t) \exp (\beta_0 z(t)), \; t \geq 0,$$

where β_0 is a p - vector of unknown regression coefficients.

2. **Ciampi and Etezadi (1985)** presented a method for testing the proportional hazard (PH) and accelerated failure times (AFT) hypotheses against a general model for the hazard function. They presented the Cox's model as

$$\lambda(t;z) = g(\beta z)\lambda_0(t),$$

where g is a positive function with g(0) =1, most often the exponential function.

3. **Doksum (1987)** considered a transformation model when the response variable follows a linear model. The PHM with time-independent covariates of Cox (1975) is a special case of such transformation model, and since partial likelihood methods have proved to be so useful. Therefore the author investigates proportion of partial likelihood method in the transformation model and introduces a likelihood sampling method to compute partial likelihood and then maximum partial likelihood estimates. Linear hazard models have been

considered legally as alternates to the well – known **Cox PHM** for the repression analysis of censored survival data. He examines how individual observation influences cumulative hazard estimates.

4. Etezadi and Ciampi (1987) developed extended hazard regression (EHR) for censored data. They defined the extended hazard regression model as:

$$\lambda(t;z) = g_1(\alpha.z)\, \lambda_0(g_2(\beta.z)\, t),$$

where $\lambda_0(t)$ is the baseline hazard function.

5. Dorota and Kjell (1987) presented some estimates and confidence intervals for median and mean life in the proportional hazard model.

6. Ghorai (1987) studied the nonparametric estimation of the mean residual life with censored data under the PHM and obtained some convergence properties and an asymptotic confidence interval for the mean residual life.

7. Maller (1987) derived necessary and sufficient conditions for the existence of the maximum-partial likelihood estimation of the regression model associated with **Cox** PHM.

8. Efron (1988) used partial logistic regression techniques to fit parametric survival curves to censored data. These constructions were used to compare two treatments for the cancer study in terms of their estimated hazard rates in an effort to show how some familiar theoretical ideas (logistic regression, hazard rate analysis, and partial likelihood) can be combined to give a simple, insightful analysis of censored data.

9. Bagai (1989) proposed a distribution-free test for testing the equality of two failure rates in the competing risk setup. He uses the only information about the course of failure.

10. **Pena (1989)** gave an estimate of the survival for a generalized PHM with random censorship.

11. **Han (1989)** developed an approximate formula for the asymptotic relative efficiency of the partial likelihood estimation in the proportional hazard model of **Cox (1972).** His work extends the results of **kalbfeisch (1974) and Kay (1977)** to multiple covariates and censoring.

12. **Yashin (1991)** explained how to choose the parametric form of a hazard rate in the existence of partially observed covariates. To this end, he discusses the problem of parameterization of the conditional survival function and the respective hazard. The case of partially observed randomly changing covariates is considered.

13. **Lin (1992)** designed a computer program, namely, **COFCOX** to examine the adequacy of the Cox proportional hazard model. The underlying methodology is based on the comparison of the maximum partial likelihood estimator and a weighted parameter estimator.

14. **Barnhart (1994)** related a number of multinomial models currently in use for ordinal response data in a unified manner, through the use of generalized logit models, PHM under different parameterization and computed mle for these models.

15. **Cheng and Ying (1995)** considered a class of semiparametric transformation models, under which an unknown transformation of the survival time is linearly, related to the covariates with various completely specified error distributions. This class of regression models includes the **proportional hazard models.** These transformation models, coupled with the new simple inference procedures, provide many useful alternatives to **the Cox regression model** in survival analysis. Also they showed that **Cox model (1972)** can be written as:

$$\text{Log}\,[-\log\{S_z\,(t)\}] = \lambda(t) + Z^T\,\beta$$

Where $\lambda(t)$ is a completely unspecified strictly increasing function, and β is a p x 1 vector of unknown parameters.'

16. Li, Klein and Moeschberger. (1996) Presented the results of a Monte Carlo study of the size and power of parametric and semi-parametric approaches to influence the covariate effects in survival (time-to-events) models in the presence of model misspecification and an independent censoring mechanism are reported. Basic models employed are a parametric model, where both a baseline distribution and the dependence structure of covariates on the failure times are fully specified (exponential, Weibull, logistic, log normal, and normal regression models are studied), and a semi-parametric approach(due to **Cox)** in which the baseline distribution is unspecified. Appropriate parametric models have the potential of improving the size and power of the tests, although overall they are not appreciably better than the **Cox** model.

It may be helpful now to take demographers a little bit away from theoretical modeling to applications.

III. Applications of Hazard Models in Demography

A wealth of applications of hazard models appear in the literature. The applications cover many vital demographic areas such as birth intervals, age at marriage, marriage duration, mortality, risk analysis, migration issues, reproductive health, among many other areas.
Most of these applications are sequentially (**between 1983 and 1998**) listed in the following.

1. **Foster et al. (1983)** in an unpublished paper, applied the hazard model, to the study of female reproductive development. It is demonstrated that use of this method permits more detailed and penetrating findings to be extracted from the kinds of retrospective have only recently become available.

2. **Rodriguez (1984)** illustrated the application of proportional hazard models or life tables with regression to the analysis of birth intervals, using data from the Colombian National Fertility Survey conducted in 1976 as part of the world Fertility survey. The model describes the family building process as a series of stages where women move progressively for first birth to second birth, and so on, until they reach their completed family size. The author then presents the proportional hazards model, which can be implemented using a software package for log-linear models.

3. **Newman and McCulloch (1984)** discussed two approaches that economists have taken in analyzing the timing of births. This paper formulates an empirical model appropriate for one of these approaches and demonstrates its usefulness using household survey data from Costa Rica. The hazard rate technique employed in this paper is a natural way of modeling a broad class of problems where the occurrence of an event is uncertain.

4. **Trussel and Menken (1985)** in a comparative study of the determinants of birth-interval length using hazard - model analysis and worlds Fertility Survey data, showed that use of contraception and

breastfeeding are important but that length of the previous interval also has an effect. This variable may capture information on length and efficacy of contraception use that is not available from direct measures.

5. **Shoieb** (1985) used the Hazard Model to study the influence of Socio-demographic variables on initial and subsequent childbearing.
6. **Ofosu** (1986) studied the application of hazard models to birth interval data, with the measurement of recent fertility changes as the main objective. The proportional hazards (PH) model (Cox, 1972) is considered in the framework of nonparametric models, as well as in the form of a Weibull failure model. A generalization to non-PH situations is also attempted using the Weibull distribution as baseline. The results are encouraging for both PH models, but inconclusive for the generalized model.

7.**Dewit, et al.** (1987) applied the hazard model to study the Covariates of birth spacing patterns in Panama. In this study data from the 1975 Panama World Fertility Survey were used to identify sources of variation in birth spacing in the 2nd birth interval among 3004 women 20-49 years of age. Except for the age group 20-29 years, there was a clear trend toward longer birth intervals with increasing age at 1st union for women who married at age 18 years or above.

8. **Anderson et al. (1987)** estimated a proportional hazard model for the timing of age at marriage of women in Malaysia. They hypothesize that age at marriage responds significantly to differences in male and female occupations, race, and age. They found considerable empirical support for the relevance of economic variables in determining age at marriage well as evidence of strong differences in marriage patterns across races.

9.**Matsushita and Inaba (1987)** briefly described the concept of hazard function in comparisons with life tables, where the force of mortality is interchangeable with the hazard rate. The basic idea of failure time analysis is summarized for the cases of exponential multiple decrement life table is also introduced as an example of life time data analysis with cause-specific hazard rates.

10. **Gage (1987)** presented an interesting set of mathematical hazard models of mortality. In his study, a five parameter competing hazard model of the pattern of mortality is described, and methods of fitting it to survivorship, death rate, and age structure data are developed and resented. The methods are then applied to published life table and census data to construct life tables.

11. **El Rouby (1990)** presented an interesting review paper on lifetime survival function from a demographic perspective and demonstrates the use of hazard models in mortality and pregnancy wastage analysis of the Egyptian data.

12. **Kitts (1991)** in a hazard model analysis examined inferential of evidence of migration from the port-city of Viana do Castelo, Minho, is considered in). This study is an attempt to analysis the determinants of out - migration of the elite from the Portuguese city of Viana do Castelo [from 1834-1931]. The data used are derived from reconstruction of this electorate using record linkage methods.

13. **Teachman and Schollaert (1991)** studied direct and indirect effects of religion on birth timing: a decomposition exercise using discrete-time hazard-rate models. The authors found that being Catholic has a variable impact upon 1st-birth timing depending upon the manner in which the dependent variable is measured. Where birth timing is measured as age at 1st birth, Catholices' is slower than of non-Catholics.

14. **Shoieb (1991)** compared level of fertility in urban and rural Egypt utilizing a Hazard Model.

15. **Brecht and Michels (1991)** outlined nonparametric techniques for estimating the hazard function and then applied his formulation to the analysis of return migration among guest - workers in West Germany.

16. **Ahn and Shariff (1992)** considered a hazard model analysis together with data from the Demographic and Health Surveys are used to develop a comparative study of fertility in Togo and Uganda.

17. Guo (1993) used the sibling data to estimate family mortality effects in Guatemala. He pointed out that the parameter estimates yielded by the multivariate hazard model are very similar to those yielded by the standard hazard model. He defined the concept of the familial genetic factors in the light of behavioral genetic theory.

18. Swenson and Thang (1993) used hazard model analysis in a study of the determinants of birth intervals in Vietnam. This study assesses the impact of selected determinants (birth order, birth intervals in Vietnam). Retrospective data on 4172 eligible women aged 15-49 years and 13137 children were obtained from the 1988 Vietnam Demographic and Health Survey. Hazard models were constructed for each birth order for birth orders of two through five and over six.

19. Toulemon (1993) presented a method to calculate adjusted rates from hazard regression parameters and adjusted probabilities from logistic regression parameters. Login hazard regression software produced estimates of instant rates. The odes explain cohabitation and marriage rates for childless women who have never lived before within a couple using data from the French Fertility Survey conducted from January to March 1988.

20. Rodriguez (1994) introduced a highly selective review of statistical issues that arise in the application of hazard models to the analysis of reproductive histories, focusing largely on the need to make explicit provision in the model for unobserved sources of heterogeneity.

21. Pinto (1994), in a Doctoral dissertation, submitted to the University of Wisconsin-Madison, used a hazard model to estimate the effects of the independent variables on the resumption of postpartum menses. Simple life-table analyses confirmed longer anovulatory cycles among women with a pattern of intense suckling/prolonged nursing (more than eight times/day), poor nutrition, and a heavy work load. Significant for postpartum contraception.

22. **Kidd (1995)** applied a hazard model to access the impact of legislation on the divorce rate with the role of legislation captured via a time -varying covariant.

23. **Nair (1996)** used Cox's proportional hazard model to estimate the effects of socio-economic, demographic and proximate variables on birth intervals in Kerala (India).

24. **Kravdal (1997)** in studying the sociodemographic differentials in the elevated mortality from cancer, used a mixed additive-multiplicative continuous-time hazard model with categorical covariates is suggested. This model is a simple and plausible extension of the multiplicative hazard model demographers are well acquainted with.

24. **Shoieb (1998)** in a study to determine what Socio- demographic variables affect the onset of childbearing and the pace of subsequent births used a proportional hazard model.

25. **Smith and McClean (1998)** described and apply selected techniques used to analyze paired hazard rates when event times are right censored and the techniques by looking at mortality patterns husbands and wives.

VI. Conclusion

This paper has tried to introduce a self-contained article on hazard models in Demography. To this end, we have introduced the relevant material which we felt would be just enough for the reader to follow the presentation. An almost complete coverage of published developments, results, and applications of value to demographers followed this. We believe that the contents of this paper could serve as a good start for any researcher interested in linking hazard models to a demographic issue.

Bibliography

- **Ahn, N. and Shariff, A. (1992)** A Comparative Study of Fertility Determinants In Togo And Uganda: A Hazard Model Analysis Economic Growth Center, Yale University, Papers No. 665.
- **-Allison, P. (1982).** Discrete – Time Methods For The Analysis Of Event Histories, PP. 61-98 In Sociological Methodology, Edited By S. Leinhardt, San Francisco :
- **Andersen, P. and Gill, R. (1982).** Cox's Regression Model For Counting Processes: A Large Sample Study. Ann. Statist., 10 , 1100-1120
- **Anderson, K., Hill, M. and Butler, J. (1987).** Age At Marriage In Malaysia: A Hazard Model Of Marriage Timing. Journal Of Development Economics, 26 (2), 223 – 234.
- **Bagai, I. (1989).** A Distribution-Free Test For The Equality Of Failure Rates Due To Two Competing Risks. Comm. Statist. Theory And Methods, Vol. 18, No. 1, 107 – 120.
- **Barnhart, H. (1994).** Overview Of Multinomial Models For Ordinal Data. Comm. Statist. Theory And Methods, Vol. 23, No. 12, 3395 – 3416.
- **Berkson, J. (1942).** The Calculation Of Survival Rates, In Carcinoma And Other Malignant Lesions Of The Stomach, Edited By Walkman Walters, Howard K. Cray And James T. Priestly, W. B. Saunders Company, Philadelphia.
- **Brecht B. and Michels, P. (1991).** The Application Of Nonparametric Techniques For Estimating The Hazard Function With Censored Data To The Duration Of Stay Of Guest – Workers In Federal Republic Of Germany. Germany, Konstanz

University, Series II No. 137.
- **Breslow, N. (1972).** Contribution To The Discussion On Professor Cox's Paper. Journal Of The Royal Statistical Society, Series B. 34, 187 – 220.
- __________ **(1974).** Covariance Analysis Of Censored Survived Data. Biometrics, 30, 89- 99.
- __________ **(1975).** Analysis Of Survival Data Under The Proportional Hazards Model. International Statistics Review, 30, 43- 54.
- **Buckley , J. And James, I. (1979)** . Linear Regression With Censored Data. Biometrika, 66, 429 – 436.
- **Cai. J. And Prentice R.L. (1995).** Estimating Equations For Hazard Ratio Parameters Based On Correlated Failure Time Data. Biometrika, 82, 1, 151- 164.
- **Cheng, S. And Ying, Z. (1995).** Analysis Of Transformation Models With Censored Data. Biometrika, 82, 4, 835 – 847.
- **Ciampi, A. And Etezadi, J. (1985).** A General Model For Testing The Proportional Hazard And The Accelerated Failure Time Hypotheses In The Analysis Of Censored Survival Data With Coveriates. Commun. Statist. Theory And Methods, 14 (3), 651 – 667.
- **Cox, D. (1972).** Regression Models In Life Table. Journal Of The Royal Statistical Society, Series B, 34, 187– 220.
- __________ **(1975).** Partial Likelihood. Boimetrika, 62, 269 – 276.
- __________ **(1984).** Analysis Of Survival Data. Monographs On Statistics And Applied Probability. Chapman And Hall.
- **Crowley, John And Marie. (1977).** Covariance Analysis Of Heart Transplant Survival Data, Journal Of The American Statistical Association, Vol. 72, No. 357, 27-36.
- **Desai, J. (1989).** Birth Interval Analysis : Results From A Parametric Proportional Hazard Model, University Of North Carolina At Chapel Hill, Carolina

Population Center, Papers No.89–90

- **Dewit, D., Rao, K., And Murty, K. (1987).** Covaviates Of Birth Spacing Patterns In Panama : A Hazard Model Analysis, In : American Statistical Association, 116-7.
- **Doksum, K. (1987).** An Extension of Partial Likelihood Methods for Proportional Hazard Models to General Transformation Models. Ann. Statist, 15, No. 1, 325 – 345.
- **Dorota, M. And Kjell, A. (1987).** Estimates And Confidence Interval For Median And Mean Life In The Proportional Hazard Model. Biometrka. 14 (4), 799 – 807.
- **Efron, B. (1977) .** The Efficiency Of Cox's Likelihood Function For Censored Data. Journal Of The American Statistics Association 72 : 557 – 565.

______________ **(1988)** . Logistic Regression, Survival Analysis, And The Kaplan – Meier Curve. Journal Of The American Statistical Association, 83, 414 – 425.

- **EL- Rouby, M. (1990).** Lifetime Survival Function. In A Demographic Perspective. The 25th Annual Conference I.S.S.R. – Cairo University, Vol. 25.
- **Etezadi, J. And Ciampi, A. (1987).** Extended Hazard Regression For Censored Survival Data With Covariates: A Spline Approximation For The Baseline Hazard Function. Biometrics 43, 181-192.
- **Foster, A. (1983) .** Female Reproductive Development And The Hazard Model. Unpublished Report (Abstract).
- **Gage, T. (1987).** Mathematical Hazard Models Of Mortality: An Alternative To Model Life Tables. Unpublished Report (Abstract).
- **Ghorai, J. (1987).** Estimation Of Mean Residual Life With Censored Data Under The PHM. Comm. Statist. Theory And Methods, Vol. 16, No. 7, 2097 – 2114.

Guo, G. (1993) Use Of Sibling Data To Estimate Family Mortality Effects In Guatemala. Demography, Vol. 30 No. 1, 15 – 32.

- Han, A. (1989). A Symptomatic Efficiency Calculations Of The Partial Likelihood Estimator. J. Econometrics, 41, No. 2, 237 – 250.

– Holford. T. (1976). Life Table With Concomitant Information. Biometrics 32, 587 – 597.

- Jensen, E. (1984). A Hazard Rate Model Of Fertility And Mortality In Guatemala. Paper Presented At The Population Association Of America Annual Meeting, Minneapolis, Minn., May 3 – 5, 1984.

-Jensen, A. (1986). Do Employed Women Have A Different Fertility Pattern From Not Employed Women? An Illustration Of Hazard Rate Methodology. Tidsskrift For Samfunnsforskning, 27 (4),334–356.

-Kalbfleisch, J. D. And Prentice, R. L. (1973). Marginal Likelihood Based On Cox's Regression And Life Table Model, Biometrika, Vol. Go, 267-278.

Kalbfleisch, J. (1974). Some Extensions And Applications Of Cox's Regression And Life Model, Presented At The Joint Meeting Of The Biometric Society And The American Statical Association, Tallahassee, Florida, March 20-22.

- Kalbfleisch, J., And Prentice, R. (1979) . Hazard Rates Models With Covariates . Biometrics, 35, 25 – 39.

________________ (1980a). Marginal Likihoods Based On Cox's Regression And Life Model. Biometrika, 60, 267 – 278.

________________(1980b). The Statistical Analysis Of Failure Time Data. New York: John Wiley And Sons.

- Kay, R. (1977) . Proportional Hazard Regression Models And The Analysis Censored Survival Data. Applied Statistics, 26, 227 – 237.

- **Kidd, M. (1995).** The Impact Of Legislation On Divorce : A Hazard Function Approach . Applied Economics, 27 (1) : 125 – 130.
- **Kitts, A. (1991) .** A Hazard Model Analysis Examining Inferential Evidence Of Migration From The Port – City Of Viana Do Costelo, Minho, 1834 – 1931. European Journal Of Population 7 (2) : 113 – 28.
-**Krvdal, O. (1997).** The Attractiveness Of An Additive Hazard Model An Example From Medical Demography. European Journal Of Population, 13 (1) : 33 – 47.

- **Kumar, D. (1996).** Proportional Hazard Modeling Of Time – Dependent Covariates Using Linear Regression; A Case Study. IEEE Trans. Rel., 45, 368 – 391.
- **Le, C. (1997).** Applied Survival Analysis. John Wiley And Sons.
- **Lee, E. (1980).** Statistical Methods For Survival Data Analysis. Lifetime Learning Publications.
- **Li, Y., Klein, J. And Moeschberger, M. (1996).** Effects Of Model Misspecification In Estimating Covariant Effects In Survival Analysis For Small Sample Sizes. Computational Statistics And Data Analysis . 22, 177 – 192.
- **Li, L. And Shoe, M. (1997).** A Mixture Model For Duration Data: Analysis Of Second Births In China. Demography, Vol. 34, No. 2, 189 – 197.
- **Lin, D. (1992)** COFCOX. A Computer Program For The Goodness – Of Fit Analysis Of The Cox Proportional Hazards Model. Computer Methods And Programs In Biomedicine. 33, 101 – 105.

- **London,D. (1988)** . Survival Models And Their Estimation (Second Edition). ACTEX Publications, Winston And New Britain, Connecticut.
-**Maller, A., (1987).** Relating Cox's Proportional Hazard Model To The Exponential Model For Survival. Biometrical J., 29, No. 2, 231 – 238 .
- **Marini, M. (1981)** . Measuring The Effects Of The Timing Of

Marriage And First Birth. Journal Of Marriage And Family, Pp. 19 – 26.

- **Matsushita, K. And Inaba, H. (1987)** . Hazard Function And Life Table : An Introduction To The Failure Time Analysis . Journal Of Population Problems, 182, 36 – 50.
- **Moore, K. And Sandra, L. (1978).** The Consequences Of Age At First Childbirth : Final Research Summary. Washington, D. C.: The Urban Institute.
- **Nadas, A. (1970).** On Proportional Hazard Functions, Tachometric, Vol. 12, No. 2, 413-416.
- **Nair, S. (1996)** . Determinants Of Birth Intervals In Kerala : An Application Of Cox's Hazard Model GENUS. Jul-Dec; 52 (3-4) : 47 – 65.
- **Negm, A., Tawila, S. And Sheneity, S. (1995).** Determinants Of Infant Mortality In Egypt; A Parametric Hazards Model Approach. The 30th Annual–Conference ISSR–Cairo University, Vol. 30.
- **Newman, J. And Mcculloch, C. (1984).** A Hazard Rate Approach To The Timing Of Births. Econometrica, 52 (4) : 939 – 961.
- **Ofosu, D. (1986).** Hazard Models As A Methodology For Measuring Recent Fertility Changes Via Birth Intervals: An Exploratory Study. Belgium, University Catholique De Louvain, Department De Demographie, /Working Paper No. 130.
- **Pena, E. (1989)** . Survival Function Estimation For A Generalized Proportional Hazard Model Of Random Censorship. J. Statist. Plan. Inf., Vol. 22, No. 22, 371 – 389.
- **Peto, R. And Peto, J. (1972).** Asymptotically Efficient Rank Invariant Procedures. Journal Of The Royal Statistical Society, A 135, 185 – 207.
- **Pinto, G. (1994).** Breast Feeding Patterns, Nutrition And Postpartum Amenorrhea In Guatemalan Women: A Multi–State Hazard Approach. Doctoral Dissertation, University Of Wisconsin, Madison.

- **Rodriguez, G. (1984)** . The Analysis Of Birth Intervals Using Proportional Hazard Models. Santiago, Chile, Facultad De Matematica.

_______________ **(1994)** Statistical Issues In The Analysis Of Reproductive Histories Using Hazard Models. In : Human Ecology : Interactions Of Environment, Fertility, And Behavior, New York Academy Of Sciences, 709 : 266 – 79.

- **Sahai . H. And Anwer K. (1996)** . Statistics In Epidemiology : Methods, Techniques, And Applications . CRC Press.

- **Shoieb, F. (1985).** Socio – Demographic Influence On Initial And Subsequent Childbearing: A Hazard Model Analysis. Unpublished Dissertation, North Carolina State University.

_______________ **(1991).** Initiation And Subsequent Birth Intervals, And Level Of Fertility In Urban And Rural Egypt. The Egyptian Population And Family Planning Review, Vol. 25, No. 2, 64- 75.

_______________ **(1998)** . Initial And Subsequent Childbearing And Some Fertility Determinants In Egypt, 1995. The Annual Conference On Statistics, Computer Sciences And Operation Research. Institute Of Statistical Studies And Research, Cairo University.

Smith, K. And Zick, C.(1994). Linked Lives, Dependent Demise? Survival Analysis Of Husbands And Wives. Demography, Vol. 31, No.1.

Smith, K. And Mcclean, S. (1998). An Introduction To The Analysis Of Paired Hazard Rates In Studies Of The Family. Journal Of Marriage And The Family, 60 (1), 243 – 257.

- **Steele, F. Diamond, I. And Wang, D. (1996),** The Determinants Of The Duration Of Contraceptive Use In China: A Multilevel Multinomial Discrete – Hazards Modeling Approach. Demography, Vol. 33, No. 1, 12- 23.

- **Swenson, I., Thang, N. (1993).** Determinants Of Birth Intervals In Vietnam: A Hazard Model Analysis, Journal Of Tropical Pediatrics, 39 (3), 163 –167.
- **Teachman, J. And Schollaert, P. (1991).** Direct And Indirect Effects Of Religion On Birth Timing: A Decomposition Exercise Using Discrete–Time Hazard–Rate Models. Sociological Quarterly, 32 (1), 151 –159.
- **Teachman, J. And Hayward, M. (1993).** Interpreting Hazard Rate Models. Sociological Methods And Research, 21 (3), 340 –371.
- **Toulemon, L. (1993).** Standardization From Hazard And Logistic Regression. Presented At The International Population Conference, Montreal, Canada, Aug. 24- Sept. 1, 1993.
- **Trussell, T. And Menken, J. (1978).** Early Childbearing And Subsequent Fertility. Family Planning Perspectives, 10 (4).
- ________________ **(1985).** Determinants Of Birth – Interval Length In The Philippines, Malaysia, And Indonesia: A Hazard – Model Analysis. Demography, 22 (2), 145 – 168.
- **Tsiatis, A. (1981).** A Large Sample Study Of Cox's Regression Model. Annals Of Statistics, 9, 93 – 108.
- **Yamaguchi, K. (1995).** Mover-Styer Models For Analyzing Event Nonoccurrence And Event Timing With Time-Dependent Covariates: Black-White Differences In The 1995 Annual Meeting Of The American Sociological Association, August, Washington, DC.
- **Yashin, A. (1991).** How To Choose The Parametric Form Of A Hazard Rate: Partially Observed Covariates. University Of Minnesota, Center For Population Analysis And Policy, Research Report No. 91– 04–2.