SOME PRIOR AND POSTERIOR DISTRIBUTIONS IN SURVIVAL ANALYSIS, AND THEIR APPLICATIONS*

by

Regina C. Elandt-Johnson

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1206

JANUARY 1979

SOME PRIOR AND POSTERIOR DISTRIBUTIONS IN SURVIVAL ANALYSIS, AND THEIR APPLICATIONS*

Regina C. Elandt-Johnson
Department of Biostatistics
University of North Carolina
Chapel Hill, N.C. 27514, U.S.A.

SUMMARY

Concomitant variables $(\underset{\sim}{z})$ introduced into survival models are often regarded as risk or aging factors which contribute to the mortality patterns in various study groups. Concomitant variables observed at the time point $t = 0$ (and having values $\underset{\sim}{z}_0$) can also be regarded as random variables, $\underset{\sim}{Z}_0$, with some *prior* joint distribution. If the Z's do, indeed, contribute to mortality, their *posterior* distribution among the survivors to time t will be different from that at $t = 0$, even when the values of the z's do not depend on t. General formulae for the posterior distributions are given in Section 2 (assuming z's not to depend on t), and in Section 4 (for z's varying with t). Our interest is especially focussed on models with *linear additive* hazard functions of the form (3.2). It is shown that in such cases the posterior distribution does not depend on the "underlying" hazard; its general form is given by (3.6), and by (3.7) where

*This work was supported by the National Heart and Lung Institute contract NIH-NHLI-712243 and by NIH research grant number 1 R01 CA17107 from the National Cancer Institute.

the concomitant variables are independent. Of some interest is the case when a univariate variable Z_0 is a binary variable - an application of posterior probabilities in clinical trials is suggested in Section 3. Of special interest is the case when Z is a continuous function of t. Level of serum cholesterol, or blood pressure are examples. Posterior distributions and posterior regression of Z on t among the survivors to time t are those which are usually observed. Assuming an additive model for the hazard function (6.2), one may infer, under certain conditions, the prior distribution of Z, without using repeated measurements (Section 6).

Key Words & Phrases: Concomitant variables; Prior and posterior distributions, and regression functions; Additive hazard rate function; Survival function.

1. INTRODUCTION

Assessment of the role of concomitant variables as risk factors of mortality in various diseases is the subject of many current longitudinal studies and clinical trials.

Let $T > 0$ denote the survival time (age) and let $\underset{\sim}{z}$ be a $k \times 1$ vector of concomitant variables which can be measured on each individual (unit) under consideration.

Let $\underset{\sim}{z}_0$ be a $k \times 1$ vector of concomitant variables at initial time $t = 0$, and

$$S_T(t;\ \underset{\sim}{z}_0) = \Pr\{T > t;\ \underset{\sim}{z}_0\}\ , \tag{1.1}$$

be the corresponding survival distribution function (SDF).

The usual approach to estimation of SDF defined in (1.1) is to assume a parametric or semiparametric form for the hazard function of (1.1), in which $\underset{\sim}{z}_0$ is treated as a vector of parameters [e.g. Cox (1972), Prentice (1973), Byar and Corle (1977), Kay (1977), and many others).

However, in a given study, one may consider a random vector $\underset{\sim}{Z}_0$ having a specified *prior* distribution (CDF), $F_{\underset{\sim}{Z}_0}(\underset{\sim}{z}_0)$. If the $\underset{\sim}{z}_0$'s play a significant role in mortality (i.e. are risk or aging factors), then the *posterior* distribution among the survivors to age t, $F_{\underset{\sim}{Z}_0|t}(\underset{\sim}{z}_0|t)$, would be different from $F_{\underset{\sim}{Z}_0}(\underset{\sim}{z}_0)$ even though z_0's are not functions of t.

The purpose of this paper is to investigate the properties of certain models of prior and/or posterior distributions of Z_0's, and their use in estimating the SDF's. Though the problem is of Bayesian nature, it should be pointed out that the forms of the prior distributions need not always be assumed 'a priori' — they can be estimated

from actually observed distributions.

The cases when the z_0's are independent of time t (Sections 2, 3, and 4), and when they are functions of t (Sections 5 and 6) are treated separately, for convenience.

2. POSTERIOR DISTRIBUTIONS: CONCOMITANT VARIABLES INDEPENDENT OF TIME

We discuss here some models, in which the concomitant variables do not depend on time t.

Let $dF_{\underset{\sim}{Z}_0}(\underset{\sim}{z}_0)$ be the probability element of the random vector $\underset{\sim}{Z}_0' = (Z_{01}, \ldots, Z_{0k})$, where $\underset{\sim}{z}_0 \in \Omega_k$. We call $dF_{\underset{\sim}{Z}_0}(\underset{\sim}{z}_0)$ the *prior* probability element of $\underset{\sim}{Z}_0$ in Ω_k.

The *posterior* probability element *among the survivors to age* t with the SDF defined in (1.1) is

$$dF_{\underset{\sim}{Z}_0|t}(\underset{\sim}{z}_0|t) = \frac{S_T(t;\ \underset{\sim}{z}_0)dF_{\underset{\sim}{Z}_0}(\underset{\sim}{z}_0)}{\int_{\Omega_k} S_T(t;\ \underset{\sim}{z}_0)dF_{\underset{\sim}{Z}_0}(\underset{\sim}{z}_0)} \ . \tag{2.1}$$

Note that

$$\int_{\Omega_k} S_T(t;\ \underset{\sim}{z}_0)dF_{\underset{\sim}{Z}_0}(\underset{\sim}{z}_0) = E_{\underset{\sim}{Z}_0}[S_T(t;\ \underset{\sim}{Z}_0)] \tag{2.2}$$

is the *average* SDF over the whole set of z_0's. It can be estimated from survival data.

Sometimes we might be able to estimate the prior as well as the posterior distributions of $\underset{\sim}{Z}_0$ and then the SDF from the formula

$$S_T(t;\ \underset{\sim}{z}_0) = \frac{dF_{\underset{\sim}{Z}_0|t}(\underset{\sim}{z}_0|t)}{dF_{\underset{\sim}{Z}_0}(\underset{\sim}{z}_0)} E_{\underset{\sim}{Z}_0}[S_T(t;\ \underset{\sim}{Z}_0)] \ , \tag{2.3}$$

3. MODELS WITH ADDITIVE HAZARD RATE FUNCTIONS

We define the general *additive* model of hazard rate function as

$$\lambda(t;\ z_0) = \lambda(t) + \sum_{i=1}^{k} h_i(t) \cdot g_i(z_{0i}) \ , \tag{3.1}$$

where $\lambda(t)$ is the so called *underlying* hazard rate. Note that the $h_i(t)$'s are entirely functions of t, while the $g_i(z_{0i})$'s are not dependent on t.

For convenience, however, and without loss of generality (by appropriate definition of z_0's) we confine our discussion to *linear additive* models of the form

$$\lambda(t;\ \underset{\sim}{z}_0) = \lambda(t) + \sum_{i=1}^{k} h_i(t) z_{0i} \ , \tag{3.2}$$

and in special cases

$$\lambda(t;\ \underset{\sim}{z}_0) = \lambda(t) + \sum_{i=1}^{k} \alpha_i z_{0i} \ , \tag{3.3}$$

where the α_i's are constants.

Further, define

$$\Lambda(t) = \int_0^t \lambda(u)du, \quad H_i(t) = \int_0^t h_i(u)du \ ,$$

and denote by $M_{\underset{\sim}{Z}_0}(\underset{\sim}{s})$ the joint moment generating function (MGF) of the distribution of concomitant variables $F_{\underset{\sim}{Z}_0}(\underset{\sim}{z}_0)$.

Then the cumulative hazard function (CHF) of (3.2) is

$$\Lambda(t;\ \underset{\sim}{z}_0) = \Lambda(t) + \sum_{i=1}^{k} H_i(t) z_{i0} \ ; \tag{3.4}$$

the survival function is

$$S_T(t;\ \underset{\sim}{z}_0) = \exp[-\Lambda(t)]\exp[-\sum_{i=1}^{k} H_i(t) z_{0i}] \ ; \tag{3.4a}$$

and (from (2.2))

$$E_{\underset{\sim}{Z}_0}[S_T(t;\ \underset{\sim}{Z}_0)] = \exp[-\Lambda(t)]\int_{\Omega_k} \exp[-\sum_{i=1}^{k} H_i(t)z_{0i}]dF_{\underset{\sim}{Z}_0}(\underset{\sim}{z}_0)$$

$$= \exp[-\Lambda(t)]M_{\underset{\sim}{Z}_0}[-\underset{\sim}{H}(t)], \tag{3.5}$$

where $\underset{\sim}{H}(t) = (H_1(t), \ldots, H_k(t))$. (See Elandt-Johnson (1976).)

Hence, the posterior distribution of Z_0's among the survivors to time t is (from (2.1))

$$dF_{\underset{\sim}{Z}_0|t}(\underset{\sim}{z}_0|t) = \frac{\exp[-\sum_{i=1}^{k} H_i(t)z_{0i}]dF_{\underset{\sim}{Z}_0}(\underset{\sim}{z}_0)}{M_{\underset{\sim}{Z}_0}[-\underset{\sim}{H}(t)]}\ . \tag{3.6}$$

In particular, when the Z_0's are *mutually independent*,

$$dF_{\underset{\sim}{Z}_0}(\underset{\sim}{z}_0) = \prod_{i=1}^{k} dF_{Z_{0i}}(z_{0i})\ ,$$

and

$$M_{\underset{\sim}{Z}_0}[-\underset{\sim}{H}(t)] = \prod_{i=1}^{k} M_{Z_{0i}}[-H_i(t)]\ ,$$

so that (3.6) takes the form

$$dF_{\underset{\sim}{Z}_0|t}(\underset{\sim}{z}_0|t) = \prod_{i=1}^{k} \frac{\exp[-H_i(t)]dF_{Z_{0i}}(z_{0i})}{M_{Z_{0i}}[-H_i(t)]}$$

$$= \prod_{i=1}^{k} dF_{Z_{0i}|t}(z_{0i}|t)\ , \tag{3.7}$$

where

$$dF_{Z_{0i}|t}(z_{0i}|t) = \frac{\exp[-H_i(t)]dF_{Z_{0i}}(z_{0i})}{M_{Z_{0i}}[-H_i(t)]}\ , \tag{3.8}$$

is the posterior probability element with respect to variable Z_{0i}.

Summarizing these results:

If the hazard rate function, $\lambda(t; \underset{\sim}{Z}_0)$ *is of the linear additive form (3.2), and the* Z_0*'s are random variables with joint distribution* $dF_{\underset{\sim}{Z}_0}(\underset{\sim}{z}_0)$*, then the posterior distribution of the* Z_0*'s among the survivors to time* t *does not depend on the underlying hazard* $\lambda(t)$*, and its explicit form is given by (3.6). If additionally, the* Z_0*'s are mutually independent, then the* Z_0*'s, given* t*, are also mutually independent.*

The general *multiplicative* model of hazard rate function can be defined as

$$\lambda(t; \underset{\sim}{z}_0) = \lambda(t) g(\underset{\sim}{z}_0) , \tag{3.9}$$

or more specifically as

$$\lambda(t; \underset{\sim}{z}_0) = \lambda(t) \prod_{i=1}^{k} g(z_{0i}) . \tag{3.10}$$

A special case, the so called *multiplicative exponential* model

$$\lambda(t; \underset{\sim}{z}_0) = \lambda(t) \exp\left(\sum_{i=1}^{k} \beta_i z_{0i} \right) , \tag{3.10a}$$

is in common use (e.g. Cox (1972)).

General multiplicative models will not be discussed in detail, though special cases will occasionally be used for comparisons.

4. SOME APPLICATIONS IN CLINICAL TRIALS

Consider a simple clinical trial in which only two groups are distinguished: control and experimental. Let

$$z_0 = \begin{cases} 0 \text{ if control} \\ 1 \text{ if experimental} . \end{cases}$$

Suppose that the patients are assigned to these groups in the initial ratio

Control : Experimental $= p : (1-p) = c \quad (p,\ 1-p > 0)$,

Then the *prior* distribution of Z_0 among the survivors to time t is

$$\Pr\{Z_0 = 0\} = p = c(1+c)^{-1} \quad \text{and} \quad \Pr\{Z_0 = 1\} = 1 - p = (1+c)^{-1} ,$$

4.1. Models with additive hazard function

Suppose that the hazard rate has the additive form

$$\lambda(t;\ z_0) = \lambda(t) + h(t)z_0 . \tag{4.1}$$

Thus, (from (3.4))

$$S_T(t;\ z_0) = \exp[-\Lambda(t)]\exp[-H(t)z_0] , \tag{4.2}$$

and

$$E_{Z_0}[S_T(t;\ Z_0)] = \sum_{z_0=0}^{1} S_T(t;\ z_0)\Pr\{Z_0 = z_0\}$$

$$= (1+c)^{-1}\exp[-\Lambda(t)]\{c + \exp[-H(t)]\} . \tag{4.3}$$

The posterior probabilities among the survivors to time t are

$$\Pr\{Z_0 = 0|t\} = c\{c + \exp[-H(t)]\}^{-1} ,$$

and

$$\Pr\{Z_0 = 1|t\} = \exp[-H(t)]\{c + \exp[-H(t)]\}^{-1} ,$$

Their ratio is

$$\frac{\Pr\{Z_0 = 0|t\}}{\Pr\{Z = 1|t\}} = R(t) = c\cdot\exp[H(t)], \tag{4.4}$$

or

$$\log[R(t)/c] = H(t). \tag{4.5}$$

In particular, when $h(t) = \alpha$,

$$\log[R(t)/c] = \alpha t , \tag{4.6}$$

Estimation and fitting.

Suppose that at time $t = 0$, there are N_{00} and N_{10} individuals in the control and experimental groups, respectively. Clearly $c = N_{00}/N_{10}$. Suppose that there are no new entries or withdrawals during

the observation period, so that we may consider the N_{00} and N_{10} individuals as two 'cohorts' observed over a certain period.

Let N_{0t} and N_{1t} be the numbers, and $\hat{P}_{0t} = N_{0t}/N_{00}$ and $\hat{P}_{1t} = N_{1t}/N_{10}$ — the corresponding proportions of survivors to time t in the control and experimental groups, respectively. Then the estimated $R(t)$ is clearly, $\hat{R}(t) = N_{0t}/N_{1t}$, and

$$\frac{\hat{R}(t)}{c} = \frac{N_{0t}}{N_{1t}} \Big/ \frac{N_{00}}{N_{10}} = \frac{\hat{P}_{0t}}{\hat{P}_{1t}} , \tag{4.7}$$

so that (from (4.5) and (4.7))

$$\log(\hat{P}_{0t}/\hat{P}_{1t}) = H(t) , \tag{4.8}$$

and in particular (for (4,6))

$$\log(\hat{P}_{0t}/\hat{P}_{1t}) = \alpha t . \tag{4.9}$$

Note that the estimated relative risk is $(1 - \hat{P}_{0t})/(1 - \hat{P}_{1t})$; the ratio $\hat{P}_{0t}/\hat{P}_{1t}$ might be thought of as *"relative survival"*.

4.2. Models with proportional hazard rates

In a simple form of multiplicative model, we assume

$$\lambda(t;\ z_0) = \lambda(t)e^{\beta z_0} , \tag{4.10}$$

so that

$$\lambda(t;\ 1)/\lambda(t;\ 0) = e^{\beta} = \theta \qquad \theta > 0 , \tag{4.11}$$

— the hazard rate in the experimental group is proportional to that in the control group.

Then

$$S_T(t;\ z_0) = \exp[-\Lambda(t)e^{\beta z_0}] , \tag{4.12}$$

and

$$E_{Z_0}[S_T(t;\ Z_0)] = (1 + c)^{-1}\exp[-\Lambda(t)]\{c + \exp[(1 - \theta)\Lambda(t)]\} , \tag{4.13}$$

The *posterior* probabilities among the survivors to time t are

$$\Pr\{Z_0 = 0 | t\} = c\{c + \exp[(1-\theta)\Lambda(t)]\}^{-1},$$

and

$$\Pr\{Z_0 = 1 | t\} = \exp[(1-\theta)\Lambda(t)]\{c + \exp[(1-\theta)\Lambda(t)]\}^{-1},$$

so that

$$\log[R(t)/c] = (1-\theta)\Lambda(t) \ . \tag{4.14}$$

Note that where $\lambda(t) = \lambda$, (4.14) takes the form

$$\log[R(t)/c] = (1-\theta)\lambda t = \alpha t, \tag{4.15}$$

which is essentially the same as (4.9).

Therefore, (4.9) cannot be used to infer about the appropriateness of the additive model $\lambda(t; z_0) = \lambda(t) + \alpha z_0$, unless it can be assumed that $\lambda(t)$ is not constant.

Inference about proportional hazard rate model.

If the mortality data are complete, and no parametric function for $\Lambda(t)$ is assumed, $\Lambda(t)$ can be estimated from the mortality data in the control group, using the formula

$$\hat{\Lambda}(t) = \sum_{j=1}^{i} \frac{1}{N_{00} - j + 1} \quad \text{for} \quad t_i' \le t < t_{i+1}' \ , \tag{4.16}$$

where t_i' is the ith ordered time at death (Nelson (1972)).

In view of (4.7) and (4.14), one may study (graphically) the relation

$$\log(\hat{P}_{0t}/\hat{P}_{1t}) \doteq (1-\theta)\hat{\Lambda}(t) \ , \tag{4.17}$$

to investigate whether a multiplicative model (4.10) is appropriate in a given clinical trial.

5. POSTERIOR DISTRIBUTIONS: TIME DEPENDENT CONCOMITANT VARIABLES

Continuous variables which might be thought of as risk or aging factors for mortality are often *time dependent*.

We will restrict ourselves to the case of a single concomitant variable z. Let z_0 denote the value of z at time $t=0$, and z_t the value of z_0 which would be reached at time t, *assuming survival to that time*. For example,

(i) If z_0 denote initial age (at $t=0$), then at time t, the age of an individual is clearly

$$z_t = z_0 + t,$$

if he is alive at time t.

(ii) Suppose that t represents age and z the level of cholesterol (or blood pressure). It is often assumed that z is linearly related to age, that is

$$z_t = z_0 + \beta t \ ,$$

where z_0 is the initial level of z in an individual, and z_t would be the value of z at time t in the same individual if he had not died.

(iii) It is often assumed that concentrations of certain cell constituents increase exponentially with time (e.g. Arley (1961)). If z_0 is the initial concentration of z, then at time t, we have (for the same individual),

$$z_t = z_0 e^{\alpha t} \ .$$

In general case, z_t can be a function of t, z_0 and m additional parameters, $\underline{\eta}' = (\eta_1, \ldots, \eta_m)$, say,

$$z_t = \psi(t; z_0, \underline{\eta}) \ . \tag{5.1}$$

We may consider Z_0 and the η's to be continuous random

variables with the joint density $f_{Z_0,\underline{\eta}}(z_0, \underline{\eta})$. Further, the hazard rate function at time t is a function of z_t

$$\lambda(t; z_t) = \lambda[t, \psi(t; z_0, \underline{\eta})] = \lambda^*(t; z_0, \underline{\eta}), \tag{5.2}$$

In particular, the general additive model is of the form

$$\lambda(t; z_t) = \lambda(t) + \alpha z_t . \tag{5.4}$$

Thus the survival function $S_T(t; z_t)$ is equal to

$$S_T^*(t; z_0, \underline{\eta}) = \exp[-\Lambda^*(t; z_0, \underline{\eta})], \tag{5.5}$$

where

$$\Lambda^*(t; z_0, \underline{\eta}) = \int_0^t \lambda^*(u; z_0, \underline{\eta})du .$$

The posterior joint density of Z_0 and $\underline{\eta}$ among the survivors to time t is

$$f_{Z_0,\underline{\eta}|t}(z_0, \underline{\eta}|t) = \frac{S_T^*(t; z_0, \underline{\eta}) f_{Z_0,\underline{\eta}}(z_0, \underline{\eta})}{\int\cdots\int_{\Omega_{m+1}} S_T^*(t; z_0, \underline{\eta}) f_{Z_0,\underline{\eta}}(z_0, \underline{\eta}) dz_0 d\underline{\eta}}, \tag{5.6}$$

where the $(m+1)$-tuple integral in the denominator of (5.6) over the parameter space Ω_{m+1}, represents the *average* survival function, $E_{Z_0,\underline{\eta}}[S_T^*(t; Z_0, \underline{\eta})]$.

The posterior density of Z_t among the survivors to age t can be obtained by applying the transformation

$\underline{\eta} = \underline{\eta}$, $z_t = \psi(t; z_0, \underline{\eta})$, with inverse

$\underline{\eta} = \underline{\eta}$, $z_0 = \psi^{-1}(t; z_t, \underline{\eta})$, and integrating out the $\underline{\eta}$'s, giving

$$f_{Z_t|t}(z_t|t) = \int\cdots\int_{\Omega_m} f_{Z_0,\underline{\eta}|t}[\psi^{-1}(t; z_t, \underline{\eta})|t] \left|\frac{d\psi^{-1}}{dz_t}\right| d\underline{\eta} . \tag{5.7}$$

In further generalization, one may consider a random vector of concomitant variables, $\underline{Z}$. This would be a natural multivariate

extension of the model just discussed. In practice, however, the technical problems become rather difficult.

6. APPLICATIONS IN EPIDEMIOLOGY

It is often assumed (though not fully established) that there is a tendency for serum cholesterol level to increase with age for a normal (healthy) individual. Studies of such relations require repeated measurements on the same individuals, under specified conditions, and over long period of time — they are difficult (and costly) to obtain on a mass scale.

The available data are usually cross-sectional population data. It has been shown (e.g. Lewis *et.al.* (1957), Carlson and Lindstedt (1968)) that the distribution of serum cholesterol in each age group is approximately normal, and that a third order polynomial (or linear) regression function, of serum cholesterol on age, for females, and quadratic — for males, is not unreasonable to fit. We now show that these 'posterior' results are in agreement with certain simple 'prior' assumptions.

Let Z_t denote the level of cholesterol at age t, and suppose that

$$Z_t = Z_0 + B\psi(t) \ , \tag{6.1}$$

where Z_0 (the initial level of cholesterol) and B are *independent* normally distributed random variables: $Z_0 \sim N(\zeta_0, \sigma_0^2)$, $B \sim N(\beta, \sigma_1^2)$, and $\psi(t)$ is a certain (specified) function of t.

Suppose that the hazard rate function is of the *additive* form

$$\begin{aligned} \lambda(t; z_t) &= \lambda(t) + \alpha z_t \\ &= \lambda(t) + \alpha z_0 + \alpha\psi(t)b = \lambda^*(t; z_0, b), \end{aligned} \tag{6.2}$$

where α is a constant.

Note that $\lambda^*(t; z_0, b)$ is of the same form as (3.2) for $k = 2$ with $z_{01} = z_0$, $z_{02} = b$, $h_1(t) = \alpha$, $h_2(t) = \alpha\psi(t)$, though it arises from a different biological situation. To evaluate the posterior distribution of Z_0 and B, we can apply the results obtained in Section 3.

We have $H_1(t) = \alpha t$, $H_2(t) = \alpha\int_0^t \psi(u)du$. Recall that the moment generating function of a normal variable X with mean μ and variance σ^2 is $M_X(s) = \exp(\mu s)\exp(\frac{1}{2}\sigma^2 s^2)$.

Thus

$$M_{Z_0}[-H_1(t)] = M_{Z_0}(-\alpha t) = \exp(-\zeta_0\alpha t)\exp(\tfrac{1}{2}\sigma_0^2\alpha^2 t^2),$$

and

$$M_B[-H_2(t)] = \exp[-\beta H_2(t)]\exp[\tfrac{1}{2}\sigma_1^2(H_2(t))^2] \quad . \tag{6.3}$$

From (3.8), the corresponding posterior densities of $Z_0|t$ and $B|t$ are

$$f_{Z_0|t}(z_0|t) = \frac{\exp(-z_0\alpha t)(\sqrt{2\pi}\,\sigma_0)^{-1}\exp[-\frac{1}{2\sigma_0^2}(z_0 - \zeta_0)^2]}{\exp(-\zeta_0\alpha t)\exp(\frac{1}{2}\sigma_0^2\alpha^2 t^2)}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_0}\exp\left\{-\frac{1}{2\sigma_0^2}[z_0 - (\zeta_0 - \alpha\sigma_0^2 t)]^2\right\}; \tag{6.4}$$

— this is the PDF of a normal variate with mean $\zeta_0 - \alpha\sigma_0^2 t$ and variance σ_0^2;

$$f_{B|t}(b|t) = \frac{\exp[-\beta H_2(t)](\sqrt{2\pi}\,\sigma_1)^{-1}\exp[-\frac{1}{2\sigma_1^2}(b - \beta)]^2}{\exp[-\beta H_2(t)]\exp[\frac{1}{2}\sigma_1^2(H_2(t))^2]}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_1}\exp\left\{-\frac{1}{2\sigma_1^2}[b - (\beta - \sigma_1^2 H_2(t))]^2\right\} \tag{6.5}$$

— this is also the PDF of a normal variate with mean $[\beta - \sigma_1^2 H_2(t)]$ and

variance σ_1^2.

The joint posterior PDF $f_{Z_0,B|t}(z_0, b|t)$ is the product of (6.4) and (6.5). But since $Z_t = Z_0 + Bh_2(t)$, where Z_0 and B are independent normal variates, then the posterior distribution of Z_t among the survivors to time t, is normal with mean (posterior regression function)

$$E(Z_t|t) = E(Z_0|t) + [\psi(t)]^2 Var(B|t), \tag{6.6}$$

and variance

$$Var(Z_t|t) = Var(Z_0|t) + [\psi(t)]^2 Var(B|t). \tag{6.7}$$

Special cases

(a) Suppose that $Z_t = Z_0 + \beta\psi(t)$, where β is a constant. Then (from (6.6)) the posterior regression function is

$$E(Z_t|t) = (\zeta_0 - \alpha\sigma_0^2 t) + \beta\psi(t) . \tag{6.8}$$

In particular, when $\psi(t) = t$

$$E(Z_t|t) = \zeta_0 - (\alpha\sigma_0^2 - \beta)t \quad \cdots \textit{ linear;} \tag{6.9}$$

when $\psi(t) = t^2$

$$E(Z_t|t) = \zeta_0 - \alpha\sigma_0^2 t + \beta t^2 \quad \cdots \textit{ quadratic;} \tag{6.10}$$

when $\psi(t) = t^3$

$$E(Z_t|t) = \zeta_0 - \alpha\sigma_0^2 t + \beta t^3 \quad \cdots \textit{ cubic;} \tag{6.11}$$

etc.

(b) Suppose that $Z_t = Z_0 + B\psi(t)$, but both Z_0 and B are independent random variables. Now, however, when $\psi(t) = t$ (or $h_2(t) = \alpha t$), we have $H_2(t) = \frac{1}{2}\alpha t^2$, and so

$$\begin{aligned} E(Z_t|t) &= \zeta_0 - \alpha\sigma_0^2 t + (\beta - \tfrac{1}{2}\alpha\sigma_1^2 t^2)t \\ &= \zeta_0 - (\alpha\sigma_0^2 - \beta)t - \tfrac{1}{2}\alpha\sigma_1^2 t^3 \end{aligned} \tag{6.12}$$

— this is also *cubic* (as for $\psi(t) = t^3$ with B constant), but with

different coefficients.

Similarly, for $\psi(t) = t^2$, the posterior regression function is of the form $E(Z_t|t) = A_0 + A_1 t + A_2 t^2 + A_4 t^4$, etc.

The more general form of the relationship, $Z_t = Z_0\psi_1(t) + B\psi_2(t)$, allows for a variety of posterior regression functions — the mathematics is straightforward.

As has already been mentioned, we can usually observe the posterior, but almost never the prior, distribution and regression function. However, assuming that the hazard rate function, $\lambda(t; z_t)$, is of additive form (6.2), the following information about 'prior' distribution of Z, and regression of Z on t, can be deduced from the available information.

(i) If the posterior distribution of $Z_t|t$ is normal, and $Z_0|t$ and $B|t$ are independent, then the distributions of $Z_0|t$ and $B|t$ for those values of t, where observations are available, are also normal (by Cramér's Theorem — see, for example, Mathai and Pederzoli (1977), p. 6). By inversion of (3.7), Z_0 and B are also independent normal variables, and so Z_t is normal.

(ii) If the posterior regression function of $Z_t|t$ on t is of polynomial form $(z(t) = A_0 + A_1 t + A_2 t^2 + \cdots)$, the prior regression function is also of the polynomial form, but not necessarily of the same order; it depends on further assumptions about the stochastic nature of the prior regression coefficients.

To illustrate that our results can agree with observations, we present in Fig. 1 (taken from Lewis *et.al.* (1957), *Circulation*, 16, p. 236), based on cross-sectional data on cholesterol level as function of age. The authors fit a third order polynomial posterior

regression to the female data, while male data are better fitted by a quadratic regression (solid lines).

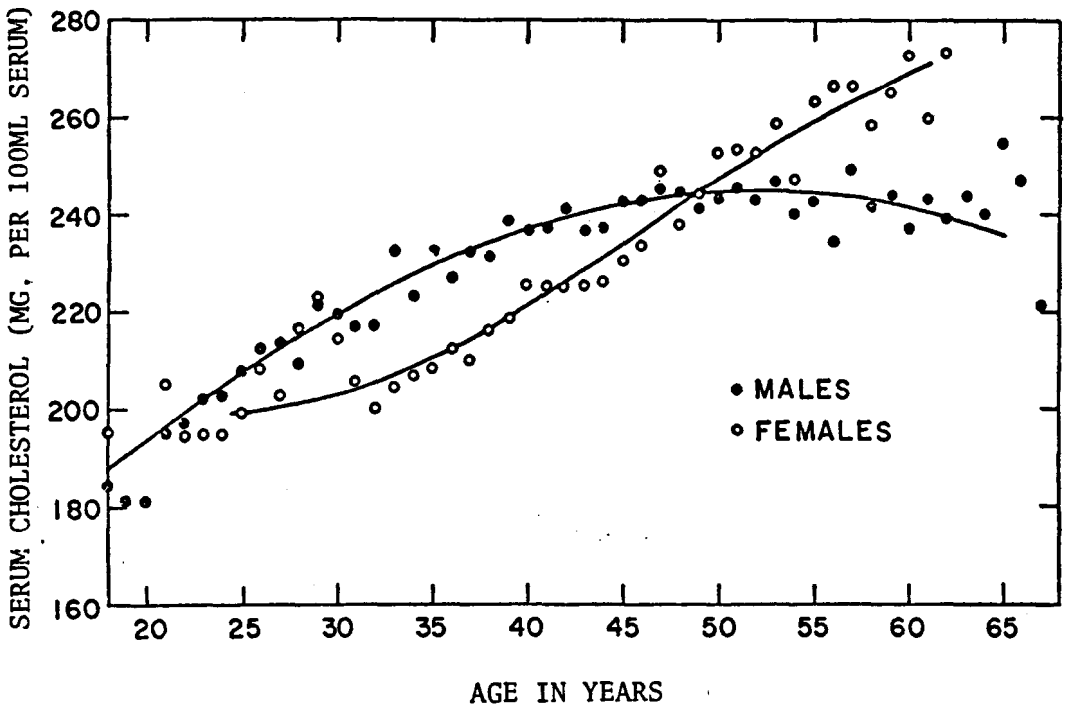

Fig. 1. Mean serum cholesterol level by age and sex (*From:* Lewis *et.al.* (1957), *Circulation* 16, p. 236).

ACKNOWLEDGMENTS

I would like to thank my husband, Dr. N.L. Johnson for a helpful discussion.

REFERENCES

ARLEY, N. (1961). Theoretical analysis of carcinogenesis. *Fourth Berkeley Symposium* Vol. 4, 7-17.

BYAR, D.P. and CORLE, D.K. (1977). Selecting optimal treatment in clinical trials using covariate information. *J. Chron. Dis.* 30, 445-449.

CARLSON, L.A. and LINDSTEDT, S. (1958). The Stockholm prospective study. 1. The initial values for plasma lipids. *Acta. Med. Scand., Suppl.* 493, 1-135.

COX, D.R. (1972). Regression models of life tables (with discussion). *J. Roy. Statist. Soc. Ser. B.* 33, 187-220.

ELANDT-JOHNSON, R.C. (1976). A class of distributions generated from distributions of exponential type. *Nav. Res. Log. Quart.* 23, 131-138.

KAY, R. (1977). Proportional hazard regression models and the analysis of censored survival data. *Appl. Statist.* 26, 227-237.

LEWIS, L.A., OLMSTED, F. et.al. (1957). Serum lipids level in normal persons. Findings of a cooperative study of lipoproteins and arteriosclerosis. *Circulation* 16, 227-245.

MATHAI, A.M. and PEDERZOLI, G. (1977). *Charaterization of the Normal Probability Law*, J. Wiley & Sons, New York.

NELSON, W.A. (1972). Theory and application of hazard plotting for censored failure data. *Technometrics*, 14, 945-966.

PRENTICE, R.L. (1973). Exponential survivals with censoring and explanatory variables. *Biometrika*, 60, 279-288.