

## AGING PERSPECTIVES IN SOME NONDEGRADATION STOCHASTIC PROCESSES

Sen, Pranab K.  
University of North Carolina

<https://doi.org/10.5109/12586>

---

出版情報 : Bulletin of informatics and cybernetics. 37, pp.1-11, 2005-12. Research Association  
of Statistical Sciences

バージョン :

権利関係 :



**AGING PERSPECTIVES IN SOME NONDEGRADATION  
STOCHASTIC PROCESSES**

by

**Pranab K. SEN**

---

*Reprinted from the Bulletin of Informatics and Cybernetics  
Research Association of Statistical Sciences, Vol.37*

---

**FUKUOKA, JAPAN  
2005**

# AGING PERSPECTIVES IN SOME NONDEGRADATION STOCHASTIC PROCESSES

By

Pranab K. SEN\*

## Abstract

In a stochastic environment, a degradation process, inspite of showing a monotone trend, may contain stochastic variations which may camouflage the statistical picture to a certain extent. There are, however, some other processes which may not exhibit a degradation phenomenon. For some of these nondegradation stochastic processes, associated aging perspectives are appraised, without being confined to a semiparametric fashion, and their application in health related quality of life assessment are considered.

*Key Words and Phrases:* DMRL, degradation and failure, HRQoL, mean residual life, recurrent events, semiparametrics, stopped counting process.

## 1. Introduction

An observed degradation process  $\mathbf{Y} = \{Y_t, t \in \mathbf{R}^+\}$  is a nonnegative stochastic process characterized by a nonincreasing trend of the associated intensity process. Bagdonavičius and Nikulin (2002) have considered some parametric as well as semiparametric degradation models for *aging*. Typically, it is assumed that

$$Y_t = g(t, \mathbf{Z})U_t, \quad t \in \mathbf{R}^+, \quad (1)$$

where  $\mathbf{Z} = (Z_1, \dots, Z_r)'$  is a (possibly) stochastic vector, representing auxiliary (explanatory) or concomitant variables,  $g(t, \cdot)$ , in a parametric setup, is a given monotone decreasing and continuously differentiable function of  $t$ , and

$$U_t = \exp\{\sigma t^{-1/2}W(t)\}, \quad t \in \mathbf{R}^+, \quad (2)$$

where  $\mathbf{W} = \{W(t), t \in \mathbf{R}^+\}$  assumed to be a standard Wiener process (independent of  $\mathbf{Z}$ ), and  $\sigma$  a positive (unknown) scale parameter. It is also desired, in some cases, to allow the covariates to be possibly time-dependent, so that in (1.1),  $\mathbf{Z}$  needs to be replaced by an auxiliary stochastic process  $\mathbf{Z}_t, t \geq 0$ . In a semiparametric formulation, less stringent assumptions have been made on the form of  $g(\cdot)$ , mostly along the lines of the classical Cox (1972) *proportional hazard model* (PHM). If  $g(\cdot, \cdot)$  is not monotone, the process  $\mathbf{Y}$  is not characterized as a degradation process; in real applications, a degradation model may only be adopted under such a characteristic feature. If  $Y_t$  is observable, under the above setup, we have

$$\log Y_t = \log g(t, \mathbf{Z}) + \sigma t^{-1/2}W(t), \quad (3)$$

---

\* University of North Carolina, Chapel Hill, NC 27599-7420 pksen@bios.unc.edu

so that conventional stochastic partial differential equations (SPDE) can be incorporated to prescribe statistical resolutions. However, in a class of problems arising in the context of *health related quality of life* (HRQoL) assessments, the survival time is not the variable  $Y_t$  in (1.1), and a somewhat different approach than the SPDE is needed. The degradation phenomenon also needs to be appraised in those contexts.

There are some notable instances in survival analysis, as well as reliability theory, where we may have a stochastic process of the type (1.1) though the degradation phenomenon may not be apparent or even tenable. In the next section, we motivate such a nondegradation process with a noteworthy hereditary disease, *Thalassemia minor* (or Cooley's anemia; also known as Mediterranean anemia), and appraise the plausibility of a stochastic degradation model, paying due emphasis on HRQoL perspectives. In such a case, the survival time and the QoL state during the survival both need to be taken into account in formulating the stochastic flow of the events of interest. Motivated by this example, in Section 3, we proceed to formulate a class of nondegradation processes, and in Section 4, we appraise their aging perspectives with an eye on the *quality of life adjusted* (QAL) *mean residual life* (MRL) analysis. Some of these statistical analyses are presented in Section 5. Some general statistical results and broad summaritative remarks are made in the concluding section.

## 2. Thalassemia Minor

Anemia is a condition in which the number of red blood cells per cu mm, the amount of hemoglobin in 100ml of blood, and the volume of packed red blood cells per 100 ml of blood are less than normal. Clinically, anemia generally pertains to the oxygen-transporting material in a designated volume of blood, in contrast to total quantities as in oligocythemia, oligochromemia and oligemia. Anemia is frequently manifested by paller skin and mucous membrane, shortness of breath, palpitation of the heart, soft systolic murmurs, lethargy and fatigability. Among the varieties of anemia, we may mention (i) hypochromic anemia and (ii) thalassemia, both being marked by deficient hemoglobin and usually microcytic blood cells; microcyte relates to small red blood cell present especially in some anemia. Splenomegaly, i.e., the enlargement of the spleen, is also observed in some case.

Thalassemia or Thalassanemia: a group of inherited disorders of hemoglobin metabolism in which there is a decrease in net synthesis of a particular globin chain without change in the structure of that chain; several genetic types exist, and the corresponding clinical picture may vary from barely detectable hematologic abnormality to severe and fatal anemia. The Lepore thalassemia syndrome is due to production of abnormally structured Lepore (a group of abnormal) hemoglobin which are clinically indistinguishable, but the non  $\alpha$ -globin chains are structurally altered.  $\beta$ -Thalassemia relates to heterozygous state.  $\alpha$ -Thalassemia is due to one of two or more genes that depress (partially to completely) synthesis of  $\beta$ -globin chains by the chromosome bearing the abnormal gene. In a homozygous state, one may have a severe type with erythroblastosis fetalis and fatal death, only Hb Barts and Hb H present; a mild-type is not clinically defined. In a heterozygous state, severe type, Thalassemia minor with 5 - 15 per cent of Hb Barts at birth and only traces of Hb Barts in adult; in mild-type, 1-2 per cent Hb Barts at birth, not detectable in adults. *Thalassemia minor is thus the heterozygous state of a Thalassemia gene or a hemoglobin Lepore gene, usually asymptomatic, and mild hypochromic microcytosis; often slightly reduced hemoglobin level with*

*slightly increased erythrocyte count. Types of hemoglobin are variable and depend on the gene involved. There may be a production of about 10 per cent of the Hb Lepore, Hb F moderately increased, and Hb A<sub>2</sub> normal.*

It is clear from the above description that the type of the disorder and degree of severity can vary considerably, and as a result, the clinical picture may vary considerably; we therefore need to focus on a specific case. In this study, we specifically keep the Talassemia minor disorder (TMD) in mind, and proceed to assess its impact on HRQoL as well as longevity (MRL), following a clinical detection of TMD. A particular measure of the hemoglobin level is the primary variable, denoted by  $Y_t$ , while the other recordable characteristics are to be treated as covariates. Familial factors as well as other clinical observations are also included in the set of covariates and explanatory variables, which is denoted by  $\mathbf{Z}_t$ ,  $t \in \mathbf{R}^+$ . Let us focus primarily on  $Y_t, t \geq 0$  and note that there is generally a normal hemoglobin level,  $N$  which, for people not afflicted with the disorder, is a central value of the distribution, and for a person in the TMD group,  $Y_t$  consistently lies somewhat below this level.  $N$  is also subject to small interpersonal variation even among the TMD-free people, and so also  $Y_t$  among the TMD-identified people. There is also a threshold level, denoted by  $L$ , such that as soon as  $Y_t$  goes below  $L$ , there is some clinical symptom which calls for a medical attention. Generally, following an effective, brief treatment,  $Y_t$  jumps to its preepisode level, and fluctuates around it until the next episode when it dips again below  $L$ . This process continues and apart from these possible episodes, the survival picture is not that much affected, albeit the anemic condition may be reflected in some living characteristics, in the manner described before. There is also a level  $C$ , the comatic state level, and a lower level  $D$ , the death state level. If  $Y_t$  plunges below  $C$ , it needs a serious medical attention and a comatic state may evolve. Further, such a treatment may not be very effective in the longrun, and as  $Y_t$  approaches the level  $D$ , the individuals survival is at stake. In this way, the survival time  $X$  is defined to be the time until a person with TMD enters the absorbing state  $D$ , and the number of episodes occurring prior to entering the absorption state, denoted by  $M$ , though stochastic in nature, may cast valuable information on the severity of the TMD. Further, in this setup, generally one does not bother to record  $Y_t$  as long as  $Y_t > L$ , so that essentially, the observable random element are the epoch times  $\tau_j, j \geq 0$  along with some little observations on the  $Y_t$  in the clinical stage when it dips below  $L$ . Also, the episodes are generally associated with high fever or some other disease factors, and hence, that information being generally available, is an important explanatory variable. From HRQoL perspectives, the events of interest are the inter-episode times  $T_j = \tau_j - \tau_{j-1}, j \geq 1$ , along with the clinical information for the sub-threshold state and  $X$  itself. Simply the survival time itself may not capture the whole picture. Figure 1 pertains to this phenomenon.

### 3. Aging Perspectives

In characterizing aging aspects of a life distribution, generally, the hazard rate is more commonly used instead of the survival function itself. The celebrated Cox (1972) model allows incorporation of explanatory and concomitant variables in a semiparametric way, though there is a need to check the validity of such a model in a specific case, as may be the present one. We denote the survival time by  $X$ , and all other explanatory

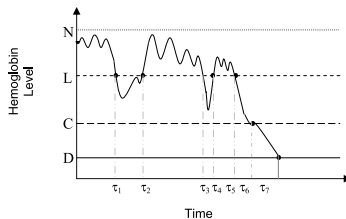


Figure 1: A TMD Process with its level crossings

or auxiliary variables (including the QoL states) by  $\mathbf{Z}$ . Generally,  $\mathbf{Z}$  is, at least partly, time-dependent, and that needs to be taken into account too. The covariate vector by  $\mathbf{Z}_t$ , and the conditional survival function, given  $\mathbf{Z}_t$  by  $S(t|\mathbf{Z}_t)$ ; we assume that it has a continuous density function  $f(t|\mathbf{Z}_t)$ . Then the conditional hazard function, given  $\mathbf{Z}_t$ , is defined as

$$\lambda(t|\mathbf{Z}_t) = \frac{f(t|\mathbf{Z}_t)}{S(t|\mathbf{Z}_t)}, \quad (4)$$

which is nonnegative for all  $t \geq 0$  and all  $t$ . In a semiparametric modeling, we conceive of a baseline hazard function  $\lambda_o(t)$ , independent of  $\mathbf{Z}_t$ , possibly of nonparametric nature, and a nonnegative regression function, possibly of a parametric form,  $g(\mathbf{Z}_t, \beta)$  involving a parameter (possibly a vector)  $\beta$ , and write

$$\lambda(t|\mathbf{Z}_t) = \lambda_o(t) \cdot g(\mathbf{Z}_t, \beta), \quad (5)$$

where the aging perspectives are then studied in terms of the intrinsic aging aspects depending on the baseline hazard function alone, and the impact of extraneous factors, through the regression function  $g(\cdot, \cdot)$ . In a degradation model,  $\lambda_o(t)$ , though unspecified, is taken to be nondecreasing in  $t$ . This is referred to as the *increasing failure rate* (IFR) case. As has been illustrated in the preceding section, for a nondegradation process, this IFR nature of the baseline hazard function may not be reasonable to assume.

The MRL  $\mu(x)$  of a life time (nonnegative) r.v.  $X$ , at age  $x$ , is defined as  $E(X - x|X > x)$ . A less restrictive (than the IFR) characterization of aging is the (decreasing) DMRL property which relates to  $\mu(x)$  being nonincreasing in  $x$ . In the presence of concomitant variables, the picture is more complex (Sen (1999)). If the concomitant variable  $\mathbf{Z}_t$  is not time-varying (and we denote it by  $\mathbf{Z}$ ), the MRL with a concomitant  $\mathbf{Z}$  is defined as

$$\mu(x|\mathbf{Z} = \mathbf{z}) = E(X - x|X > x; \mathbf{Z} = \mathbf{z}), \quad x \geq 0, \quad \mathbf{z} \in \mathbf{R}^q, \quad (6)$$

where  $q$  stands for the dimension of the concomitants. Since the coordinates of  $\mathbf{Z}$  may not all be continuous or count variables (some may even be binary or purely qualitative ones), expressing  $\mu(x|\mathbf{z})$  in a simple parametric form may not be very realistic Sen (1999), and for that reason, nonparametric as well as semiparametric ones are usually advocated. Corresponding to the hazard  $\lambda_o(t)$ , we define the cumulative hazard function as

$$\Lambda_o(t) = \int_0^t \lambda_o(y) dy, \quad t \geq 0, \quad (7)$$

which is a monotone nondecreasing function of  $t$  with  $\Lambda_o(0) = 0$  and  $\Lambda_o(\infty) = \infty$ . Incorporating the Cox (1972) PHM and using the equivalence of the negative logarithm of the (conditional) survival function and the (conditional) cumulative hazard function, we may write

$$P\{X > t|\mathbf{Z}\} = \exp\{-e^{\beta'\mathbf{Z}}\Lambda_o(t)\}. \quad (8)$$

As a result, we obtain that

$$\mu(x|\mathbf{Z}) = \int_0^\infty \exp\{-e^{\beta'\mathbf{Z}}[\Lambda_o(x+u) - \Lambda_o(x)]\}du. \quad (9)$$

It follows from Sen (2003) that if  $\Lambda_o(\cdot)$  is convex, i.e., the baseline distribution function belongs to the IFR class, then adjusted by the concomitants,  $\mu(x|\mathbf{Z})$  is nonincreasing in  $x$ , so that the DMRL property remains in tact for the conditional survival function as well. However, such a DMRL property may not hold when the basic IFR property is replaced by DMRL property (on  $\mu(x)$ ). Sen (2003) also discussed the roadblocks for the general case where the concomitants are possibly time-varying. In the most general case where  $\lambda(x|\mathbf{Z}_x) = \lambda_o(x)e^{\beta'\mathbf{Z}_x}$ , the DMRL property of  $\mu(x|\mathbf{Z}_x)$  may not follow. Sen has actually considered an intermediate case wherein he lets

$$\Lambda(x|\mathbf{Z}_s, s \leq x) = \Lambda_o(x).e^{\beta'\mathbf{Z}_x}, \quad \forall x \geq 0. \quad (10)$$

Even so, as the conditional survival function  $S(x|\mathbf{Z}_s, s \leq x)$  must be nonincreasing in  $x$ , we need a compatibility condition that  $\Lambda_o(x)\exp\{\beta'\mathbf{Z}_x\}$  is nondecreasing in  $x$ . Although such a condition may hold for a degradation model, for nondegradation models this is unlikely to be true. Hence, we need to explore alternative approaches.

#### 4. HRQoL Perspectives

In a HRQoL study, typically, based on the vector  $\mathbf{Z}_t$  as well as  $Y_t$ , a composite score, termed a QAL score, is formulated. We denote this score by  $Q_t$ ,  $t \geq 0$ . Keeping in mind the TMD discussed in Section 2, we may gather that there may be a basic problem in precisely formulating the QAL score. For example, when  $Y_t > L$ , usually no observation is available on it, albeit it is presumed that it lies somewhat below the normal level  $N$ . Therefore, there has to be a QOLD (*quality of life deficiency*) factor associated with this stage. Secondly, only after an episode time measurement of  $Y_t$ , albeit below the level  $L$  is available. The severity of TMD is perceived more with lower values of  $Y_t$  in this phase. Hence, here the actual measurement on  $Y_t$  (not a dichotomization or polychotomization) should be incorporated in  $Q_t$ . More important is the factor that an episode is generally triggered by some virus (e.g., flu) or sudden rise in the body temperature. Such covariates are only sporadic over time, and their incorporation in  $Q_t$  may not be very easy in a statistical formulation. Further adjustment for  $Y_t < C$  may also be statistically complicated. A multi-state model for  $Q_t$ , though ideal, may not be easy to implement.

In a conventional setup, the (quality of life adjusted) QLAMRL at age  $x$  is defined as

$$\mu_Q(x) = \{-\int_0^\infty Q(x+u)u dS_o(x+u)\}/S_o(x)$$

$$\begin{aligned}
&= \int_0^\infty Q(x+u)e^{-[\Lambda_o(x+u)-\Lambda_o(x)]}du \\
&\quad + \int_0^\infty ue^{-[\Lambda_o(x+u)-\Lambda_o(x)]}dQ(x+u)
\end{aligned} \tag{11}$$

when  $Q(\cdot)$  is nonstochastic [Sen (2003)]. However, generally  $Q(\cdot)$  is stochastic, and hence, as in Sen (2002), we need to replace it by its expectation counterpart. There is still another complication we should keep in mind. The survival function  $S_o(x+u)$  needs to be adjusted by the covariates at that time. If the covariates  $\mathbf{Z}$  are not time-varying and if a PHM formulation is deemed appropriate, this adjustment can be made in a reasonably manageable way. For example, as in (3.5), we need to adjust  $S_o(x)$  by the nonnegative factor  $g(\mathbf{Z})$ , so that in (4.1), we need to bring this additional factor in the exponent. For time-varying covariates, this simple prescription may not work out well (Sen (1999)), and in general, a simplification would deem the knowledge of the process  $\{\mathbf{Z}_s, s < x+u\}$ , and this is only known up to the present time  $x$ . Hence, we need to make certain regularity assumptions regarding the covariates' temporal variation. While this can be done with reference to a PHM formulation for certain degradation processes when  $\mathbf{Z}$  is not time-varying, there may be genuine roadblocks for other processes, particularly for nondegradation ones. The difficulty is that  $Q(x+u), u \geq 0$  involve the future course, so that it may depend as well on the time-varying covariates, and hence, in order to compute its expected quantity, certain restrictive assumptions may need to be made on the sample space of  $Q(\cdot)$  along with certain temporal-stationarity conditions (to facilitate predictability over a long range of time).

To illustrate this point, suppose that  $Q(x)$  is nonnegative and time-varying while the rest not. In this case, in a PHM setup, for the regression of the log-hazard on the covariates, we have  $\beta_0 Q(x) + \beta' \mathbf{Z}$ . Further, in such a context, we expect that with increasing QoL the hazard should be decreasing, so that it may not be unreasonable to assume that  $\beta_0 \leq 0$ , while the other regression parameter (vector)  $\beta$  may be unrestricted. As such, we have for given  $Q(x)$  and  $\mathbf{Z}$ ,

$$\begin{aligned}
\mu_Q(x|Q(x), \mathbf{Z}) &= \{-\int_0^\infty uQ^o(x+u)dS(x+u|Q^o(x+u), \mathbf{Z})\}/S(x|Q(x), \mathbf{Z}) \\
&= \frac{\int_0^\infty S(x+u|Q^o(x+u), \mathbf{Z})[Q^o(x+u)du + udQ^o(x+u)]}{S(x|Q(x), \mathbf{Z})}, \tag{12}
\end{aligned}$$

where we define  $Q^o(x+u)$  as the solution of

$$Q^o(x+u)e^{e^{\beta_0 Q^o(x+u)}\Lambda_o(x+u)} = E[Q(x+u)e^{e^{\beta_0 Q(x+u)}\Lambda_o(x+u)}|Q(x)], \text{ for } u \geq 0. \tag{13}$$

If we had a degradation process, we would have  $Q^o(x+u)$  nonincreasing in  $u \geq 0$ , leading to  $dQ^o(x+u) \leq 0, \forall u \geq 0$ , so that from (4.2) we obtain that

$$\begin{aligned}
\mu_Q(x|Q(x), \mathbf{Z}) &\leq \frac{\int_0^\infty Q^o(x+u)S(x+u|Q^o(x+u), \mathbf{Z})du}{S(x|Q(x), \mathbf{Z})} \\
&\leq Q(x) \int_0^\infty \exp\{-e^{\beta' \mathbf{Z}}[e^{\beta_0 Q^o(x+u)}\Lambda_o(x+u) - e^{\beta_0 Q(x)}\Lambda_o(x)]\}du. \tag{14}
\end{aligned}$$

As such, if  $\beta_0 \leq 0$  and as for a degradation process,  $Q^o(x+u) \leq Q(x), \forall u \geq 0$ , we



would have  $\beta_0 Q(x) \leq \beta_0 Q^o(x+u) \leq 0$ ,  $\forall u \geq 0$ . Therefore, we would have

$$\begin{aligned} \mu_Q(x|Q(x), \mathbf{Z}) &\leq Q(x) \int_0^\infty \exp\{-[\Lambda_o(x+u) - \Lambda_o(x)]e^{\beta' \mathbf{Z} + \beta_0 Q(x)}\} du \\ &= Q(x) E[X^o - x | X^o > x, Q(x), \mathbf{Z}], \end{aligned} \quad (15)$$

where  $X^o$  stands for a r.v. whose survival function, given  $\mathbf{Z}$  and  $Q^o(x+u) = Q(x)$ ,  $\forall u \geq 0$ . This shows that by simply adjusting by the current QoL factor  $Q(x)$  for the remaining life, the QLAMRL  $\mu_Q(x|Q(x), \mathbf{Z})$  may be well below the quantity by assuming that for the rest of life, the QoL factor does not change (a contradiction to the degradation assumption). Moreover, such a formulation is unlikely to be very realistic in the present context of nondegradation processes arising in HRQoL studies. Hence, we proceed to formulate statistical modeling and analysis protocols in an alternative way, putting proper emphasis on the episode times and their interdependence (as explained in earlier sections).

## 5. A Stopped Counting Process Approach

Consider a study plan involving  $n$  individual having the TMD and as in Sections 2 and 3, we define the episode times, and other events. Then, corresponding to a time interval  $(0, t]$ ,  $t > 0$ , for the  $k$ th individual, the vector of episode times is denoted by  $\mathbf{T}_k = (\tau_{k0}, \dots, \tau_{kj_k(t)})'$ , where  $\tau_{k0} = 0$  and

$$j_k(t) = \max\{j(\geq 0) : \tau_{kj} \leq t\}, \quad (16)$$

for  $k = 1, \dots, K$ . In that way,  $j_k(t)$  is the number of episode times (if any) for the  $k$ th individual preceding the time point  $t$ , and as such,

$$[j_k(t) \geq m] \Leftrightarrow [\tau_{km} \leq t], \quad \forall t \geq 0, \quad m \geq 0, \quad k = 1, \dots, K. \quad (17)$$

Due to time-sequential nature of the outcome, *censoring*, due to competing causes, possible withdrawal from the study, death due to the TMD, or due to other causes, is expected, and we denote the possible censoring time of the  $k$ th individual by  $C_k$ . We associate a random vector  $\mathbf{A} = (A_1, \dots, A_q)'$  of causes of censoring with the nonnegative r.v.  $C_k$ , the particular (stochastic) outcome being denoted by  $A_k$ . Let then

$$X_{kj_k(t)} = \min\{\tau_{kj_k(t)}, C_k\}, \quad \delta_{kj_k(t)} = I(X_{kj_k(t)} = \tau_{kj_k(t)}), \quad k = 1, \dots, K, \quad t \geq 0. \quad (18)$$

Note that although  $(X_{kj_k(t)}, \delta_{kj_k(t)}), k = 1, \dots, K$  capture most of the statistical information, there is some contained in following vector valued counting processes

$$\mathbf{N}_k(t) = (N_{k0}(t), N_{k1}(t), \dots, N_{kj_k(t)}(t))', \quad k = 1, \dots, K, \quad (19)$$

where  $N_{k0}(t) = 1$  for all  $t > 0$ ,  $k \geq 1$ , and

$$N_{ki}(t) = I(\tau_{ki} \leq t), \quad t > 0, \quad i \geq 0, \quad k \geq 1. \quad (20)$$

These observable random elements lead us to the following array of counting processes

$$\mathbf{N}(t) = (\mathbf{N}_1(t), \dots, \mathbf{N}_K(t))', \quad t \geq 0. \quad (21)$$

Note that the  $j_k(t)$  are not necessarily all equal, and each being stochastic, we have an array with stochastic and heterogeneous number of elements in each column of  $\mathbf{N}(t)$ , and these elements are all nondecreasing in  $t$ . In this way, there is a need for a statistical formulation of a manageable model for the entire picture, and hence, we shall mainly confine ourselves to a partial set that seems to capture a good deal of statistical information.

We are mainly interested in the following stopped counting processes:

$$(j_k(t), \mathbf{Z}_k(t), A_k), \quad t \geq 0, \quad k = 1, \dots, K. \quad (22)$$

At this time, we denote the cumulative distribution function of  $\tau_{kj}$ , given the explanatory variables and the competing censoring factor, by  $F_{(kj)}(x|B_{kr})$  where  $B_{kr}$  stands for the sigma-field generated by  $\mathbf{Z}_k(s), A_k, s \leq \tau_{kr}$ . Then, using (5.2), we obtain that

$$\begin{aligned} \nu_k(t) &= E\{j_k(t)\} = E[E\{j_k(t)|B_{k(t)}\}] \\ &= \sum_{r \geq 0} r E\{P\{j_k(t) = r|B_{kr}\}\} \\ &= \sum_{r \geq 1} E\{P\{j_k(t) \geq r|B_{kr}\}\} \\ &= \sum_{r \geq 1} E\{P\{T_{kr} \leq t|B_{kr}\}\} \\ &= \sum_{r \geq 1} E\{F_{(kr)}(t|B_{kr})\}, \\ &= \sum_{r \geq 1} F_{(kr)}^o(t), \quad \forall t \geq 0, \quad k \geq 1, \end{aligned} \quad (23)$$

where  $F_{(kr)}^o(t)$  denotes the marginal distribution obtained by taking the expectation over the conditioning factors. It is clear that  $\nu_k(t)$  is nondecreasing in  $t \geq 0$ . The analogy of this result with the classical *renewal theorem* is quite clear from above, albeit the present situation is quite complex due to the presence of the concomitant and competing risk variables.

Let us refer back to the TMD problem discussed in Section 2. The severity of the TMD syndrome may be judged with its threshold level crossing  $L$ , so that more frequent dropping below this level is a significant QOLD indicator. This, in turn, makes the inter-episode times stochastically smaller, so that for the episode times  $\tau_{kr}$ ,  $r \geq 1$ ,  $k \geq 1$ , the corresponding distributions  $F_{(kr)}^o(t)$  becomes more left-tilted. As a result, the  $\nu_k(t)$  becomes larger. On the other hand, it is not at all unreasonable to assume that the expected number of episodes (or short excursions below the level  $L$ ) in a finite life time is finite (i.e., exits) and therefore, the  $\nu_k(t)$  have finite asymptotes  $\nu_k(\infty)$  that may of course depend on other extraneous factors depending on the explanatory variables. This feature makes it appealing to incorporate suitable semiparametric models for the distributions  $F_{(kr)}^o(t)$  (or their survival functions) and express the HRQoL perspectives in terms of the auxiliary variables as much as possible.

Let us denote by

$$\bar{J}_K(t) = K^{-1} \sum_{k=1}^K j_k(t), \quad t \geq 0. \quad (24)$$

Side by side, we also, let

$$\bar{\nu}_K(t) = K^{-1} \sum_{k=1}^K \sum_{r \geq 1} F_{(kr)}^o(t), \quad t \geq 0. \quad (25)$$

As in Sen (1999), we consider then a stochastic process:

$$W_K(t) = K^{1/2} \{ \bar{J}_K(t) - \bar{\nu}_K(t) \}, \quad t \geq 0. \quad (26)$$

Realising that the counting processes  $\{j_k(t) - \nu_k(t), \quad t \geq 0\}$  are independent, weak convergence of  $W_K(t), t \geq 0$  to a Gaussian process has been studied in detail in Sen (2003). This enables us to draw statistical inference on the  $\bar{\nu}_K(\cdot)$  through the observed  $\bar{J}_K(\cdot)$ . In this respect, a semiparametric modeling of  $\bar{\nu}_K(\cdot)$  provides great simplification of statistical formulation.

There are, however, some genuine roadblocks in adopting semiparametrics in this context. First and foremost, the stopping numbers  $j_k(t)$  are themselves highly stochastic, and their incorporation in the QAL / QOLD measures requires pertinent information on the concomitant as well as auxiliary variables which are more likely to be time-varying. This alone can create impasses for routine use of PHM. Use of time-varying concomitant variables and more generally, time-varying regression coefficients (Murphy and Sen 1991), although theoretically advantageous, may encounter serious operational setbacks due to sample size and other constraints. Moreover, the very appeal of a PHM is lost if such a more general time-varying Cox-type model is adopted. Finally, genetic as well as environmental factors are generally quite relevant and their incorporation in a statistical modeling requires more delicate and nonstandard tools and techniques.

## 6. Concluding Remarks

Two distinct characteristics of the present approach are (i) its adaptability in HRQoL studies without having necessarily a degradation feature, and (ii) deemphasizing semiparametrics to a certain extent (resulting in greater model flexibility and robustness perspectives). In HRQoL studies, the QoL feature may contain some qualitative factors or traits which are difficult to quantify in a continuous scale, and even sometimes, in an ordinal scale. For that reason, a formulation of a QoL score based on item analysis and high-dimensional questionnaire is a primary statistical problem. Though there has been significant developments in psychometry and theory of mental tests in the past forty years, many of these tools and measures may not be suitable for high-dimensional data models. Moreover, these measures may be highly sensitive to specific causes, and hence, needs careful appraisal.

There is an additional feature that merits critical examination. For the TMD problem, as we can see from Figure 1, a failure due to this cause is typically accompanied by a sudden drop of the hemoglobin level all the way to the absorbing stage, and therefore, is likely to be preceded by the comatic stage where the QoL is very poor. Thus, if we regress the (conditional) hazard function on the QoL score at the failure point it will fail to capture the true dependence picture. It may be better to keep track of the passage from the threshold level to the absorbing level (time and intensity both) and use that as a measure of QoL at failure point. On the other hand, if the level is above the threshold

value then generally no measurement of the hemoglobin level is made regularly, and hence one may need to use some sort of truncation modeling for this feature, requiring more specific formulation of QoL factors.

The competing risk factor comprising of drop-outs (withdrawal) and death due to other causes also require serious consideration. A semiparametric modeling may be unsuitable in this respect. For example, for the TMD problem, a withdrawal may generally indicate a physically comfortable state so that the hemoglobin count should be above the threshold level. A cause of failure other than TMD, may not necessarily be associated with a low level (in the comatic state), so that the causes of censoring may not simply relate to the common hemoglobin count feature, and hence, a simple regression model may not suffice.

Finally, the fact that the  $j_k(t)$  are themselves stochastic suggests that if more information is available on the sojourn times (below the threshold level), one could formulate a model treating these sojourn times as concomitant variables, and thereby incorporate semiparametrics more effectively in the modeling part. However, this requires delicate considerations of the underlying biological as well as statistical undercurrents, and we intend to pursue these aspects in greater detail in a subsequent communication.

### Acknowledgement

The author is grateful to the reviewers for their helpful comments on the manuscripts, and also to Professor Antonio Carlos Pedrosa de Lima and Dr. Gisela Tunes da Silva for their help in the electronic preparation. This research was supported by the Boshamer Grant at the University of North Carolina, Chapel Hill.

### References

- Bagdonavičius, V. and Nikulin, M. (2002). *Accelerated Life Models*, Chapman-Hall/CRC Press, Boca Raton, Florida.
- Chen, P. L. and Sen, P. K. (2001). Quality adjusted survival estimation with periodic observations. *Biometrics*, **57**, 814–821.
- Cox, D. R. (1972). Life tables and regression models (with discussion). *J. Roy. Statist. Soc. B.*, **34**, 187–225.
- Murphy, S. A. and Sen, P. K. (1991). Cox regression model with time-dependent covariates. *Stochastic Processes and their Applications*, **39**, 153–180.
- Sen, P. K. (1999). Mean residual life with a concomitant. In *Golden Jubilee Festschrift*, Statistics Department, University of Bombay, (ed. U. L. Dixhit), pp.79–91.
- Sen, P. K. (2002). Measures of quality adjusted life and quality of life deficiency: statistical perspectives. In *Statistica; Methods for Quality of Life Studies: Design, Measurement and Analysis* (ed. M. Mesbah et al.) Kluwer Publ. Boston, Mass. pp.275–286.
- Sen, P. K. (2003). HRQoL and concomitant adjusted mean residual life analysis. In *Parametric and Semiparametric Models with Applications to Reliability, Survival Analysis and Quality of Life*. (eds. M. Nikulin et al.), Birkhauser, Boston, pp.349–362.

*Received November 25, 2003*

*Revised September 8, 2004*