LOG-RANK-TYPE TEST FOR EVOLUTION OF HEALTH RELATED QUALITY OF LIFE

Véronique Boisson and Mounir Mesbah[1]

L.S.T.A. - Université Pierre et Marie Curie - Paris 6
4 place Jussieu, Boîte 158, 75013 Paris, France

Zhiliang Ying

Department of Statistics, MC 4690, Columbia University 1255 Amsterdam Ave, New York, New York 10027, U.S.A.

Abstract. *This paper deals with degradation of health related quality of life (HrQoL) by stochastic process modeling of outcome data. Indeed, an important aspect in many clinical and health studies is the assessment of HrQoL. Researches on HrQoL have drawn much recent interest particularly for chronic diseases of which lack of a definitive cure is a common feature. This paper proposes to convert the HrQoL endpoint into time to degradation of HrQoL. No distributional assumption about this degradation process is assumed. The mathematical relation between the original HrQoL process and the threshold-dependent event times is established. A class of log-rank statistics is derived from the resulting event times. A new test is proposed by combining the log-rank test along with a simulation based approach for p-value approximation. The new method is applied to analyzing a data set from a cancer study.*

Keywords: *Counting process; Cox model; Health related quality of life; Log-rank test; Longitudinal data analysis.*

1. INTRODUCTION

An important aspect in many clinical and health studies is the assessment of health related quality of life (HrQoL). Researches on HrQoL have drawn much recent interest particularly for chronic diseases (such as HIV-infection or cancer) of which lack of a definitive cure is a common feature. The aim of clinical trials carried out in medicine is to estimate efficiency and safety of new therapies. HrQoL is a subjective complex concept reflecting the idea of well-being with respect to the context where the person lives. In clinical studies, HrQoL reflects two notions: well-being physically (not to feel pain, not to be ill) and well-being mentally (to have good relations with the others, not to be stressed). Despite growing importance of HrQoL outcomes and more generally of patient-reported outcomes in medical research satisfactory methods of analysis of these data remain an issue.

[1] Corresponding author Mounir Mesbah email: mounir.mesbah@upmc.fr

Considered here are longitudinal clinical studies, in which patients are asked to fill out HrQoL questionnaires on some dates t. Let $Q(t)$ denote the resulting HrQoL score at time t that can be physical or mental, or a combination of both. They are considered as variables defined on a scale with higher values corresponding to better HrQoL. Assuming that $Q_t < Q0$, define the amount of relative degradation of HrQoL at t to be $D_t = (Q_0 - Q_t)Q_0^{-1}$, which reflects the relative change of HrQoL from the initial date t_0. Some authors in epidemiological or clinical studies choose the definition $D_t = Q_0 - Q_t$.

Awad *et al.* (2002) proposed to analyze degradation of HrQoL by considering the event of first time when degradation of HrQoL is over a prefixed threshold x. Thus, for each value x, there corresponds a time variable, which may be censored if the D_t does not fall below x for the entire study period. This effectively converts the longitudinal observations into survival times. Consequently, classical statistical methods for survival data, including the Kaplan-Meier curve, the log-rank test and the Cox regression, can be made use of. In particular, Boisson and Mesbah (2008) derived a log-rank test for treatment difference by choosing a fixed threshold x of degradation of HrQoL.

This paper is the continuation along the line of Awad *et al.* (2002) and Boisson and Mesbah (2008). By varying the relative degradation threshold x, we obtain a group of survival data sets, which naturally contain more information than any individual one with a single x. We propose non- and semiparametric methods for simultaneously incorporating the entire group of data sets. The theoretical properties of the resulting methods are studied for the continuous time version using empirical process theory.

Under our formulation, analysis of degradation with HrQoL data lies in the interface of survival analysis (Andersen *et al.*, 1993; Cox, 1972) and longitudinal data analysis (Diggle *et al.*, 2002; Verbeke and Molenberghs, 2000). Theoretical properties for the Cox partial likelihood methods (Cox, 1975) can be found in Tsiatis (1981a), Tsiatis (1981b), and Andersen and Gill (1982). Our approach requires understanding of the partial likelihood score process index by an additional argument x. This is similar to the two-parameter score process which arises from group sequential designs for clinical trials as studied in Bilias, Gu and Ying (1997). The formulation of converting longitudinal process to event time via threshold-crossing is similar in spirit to recent development in finance. There, the time to a certain event of interest, such as change of credit rating or bankruptcy, is defined as hitting time of an underlying stochastic process, which indicates the well-being of a company, at a certain value; see Duffie, Saita and Wang (2007);

Hull (2008). Nevertheless, the way that we deal with the problem is slightly different. We focus on the problem of testing the effect of an observed covariate (in the current paper, we assume that it is binary). We then bypass the physical degradation process and directly deal with the log-rank statistics constructed from the event times and indexed by the threshold value x.

The paper is organized as follows. In Section 2 we give some properties of degradation of HrQoL. In Section 3 we introduce some notations and assumptions. In Section 4 we derive a partial likelihood score statistic for a survival time depending on a fixed rate x while the asymptotic properties of the derived parameters estimators are shown in an annexed section. In Section 5 we derive a log-rank statistics $U(x)$ to compare time to degradation of HrQoL between two treatment groups when the rate x is fixed. In Section 6, we show that, for fixed n, $n^{-1/2}U_n(.)$ is a Gaussian process, and in Section 7 we derive a nonparametric statistical test for joint longitudinal evolution of HrQoL and survival of treatment effect called global log-rank test. It is the main result of the paper. The proposed test is applied in Section 8 to a data set from a real cancer study. Concluding remarks are presented in Section 9.

2. BASIC PROPERTIES OF RELATIVE DEGRADATION OF HRQOL AND TIME TO DEGRADATION

Recall that Q_t denotes the HrQoL of a patient measured at time t and that $D_t = (Q_0 - Q_t)Q_0^{-1}$ (or $D_t = Q_0 - Q_t$) is the relative degradation of HrQoL. For mathematical convenience, we may consider that time t is measured continuously over interval $[O,C]$. So, the time of degradation of HrQoL is defined by

$$T(x) = \inf\{t \leq C : D_t \geq x\}.$$

Due to HrQoL questionnaires higher scores correspond to better HrQoL, some basic properties of degradation of HrQoL can be obtained.

Proposition 1:

For the threshold induced event time $T(x)$, the following results hold. Here t can be either discrete or continuous.

1. $D_0 = 0$, i.e. the rate of degradation of HrQoL is 0 at origin;
2. $D_t > 0$ if and only if the HrQoL score at t is smaller than that at the baseline 0;

3. $x_1 > x_2$ implies $P(D_t > x_1) \leq P(D_t > x_2)$.

4. $T(x)$ is monotone increasing in x, meaning that a higher threshold produces a larger event time.

Proof. Straightforward from the definition of D_t.

3. NOTATIONS AND ASSUMPTIONS

Consider a clinical study with n independent patients. Let $\mathbf{Z}_i(t)$ denote time-dependent covariate vector and $D_{i,t}$ the rate of degradation of HrQoL at time t for patient i, $i = 1,\ldots,n$. Time t is assumed to be continuous. Furthermore, let C_i denote the (right) censoring time. For a prefixed threshold value x, the associated survival time is denoted by $T_i(x)$ for patient i, as introduced in the preceding section. We shall assume that, for each i, censoring times C_i is independent of $\{D_{i,t}, t \geq 0\}$. Consequently, C_i and $T_i(x)$ are independent for all x. Define $\tilde{T}_i(x) = \min\{T_i(x), C_i\}$ and $\delta_i(x) = \mathbf{1}_{(T_i(x) \leq C_i)}$, which characterizes whether or not the degradation of HrQoL is observed. Thus the observations consist of n triplets containing a duration, a censoring indicator and the observed path of the covariate process till event of interest.

It is well known that information included in the pair $(\tilde{T}_i(x), \delta_i(x))$ is equivalent to information included in the corresponding counting processes $N_i(.,x)$ and $Y_i(.,x)$, which are defined respectively by

$$N_i(t,x) = \mathbf{1}_{(\tilde{T}_i(x) \leq t, \delta_i(x)=1)}$$

and

$$Y_i(t,x) = \mathbf{1}_{(\tilde{T}_i(x) \geq t)}.$$

We consider a generalization of the classical framework of the proportional hazard regression model of Cox (1972) in order to take into account rate x of degradation of HrQoL. The hazard function, for a patient i, λ_i is defined by

$$\lambda_i(t,x|\mathbf{Z}_i(t)) = \lambda_0(t,x)\exp\{\beta'\mathbf{Z}_i(t)\}$$

where β is the unknown vector of regression parameters and λ_0 is the unspecified baseline hazard function. We can now introduce an extension of the multiplicative intensity model in order to take into account rate x of degradation of HrQoL.

DEFINITION 1: The multiplicative intensity model for a random time depending on a critical value x of HrQoL for observations from n independent patients (or items) consists of n triplets $\{N_i(.,x), Y_i(.,x), \mathbf{Z}_i\}_{i=1,\ldots,n}$, a right-continuous filtration $\{\mathscr{F}_{t,x}; t \geq 0\}$ representing the statistical information occurring over time and n intensity processes $l_i(t,x) = \lambda_0(t,x)e^{\beta'\mathbf{Z}_i(t)}Y_i(t,x)$, $i = 1,\ldots,n$, along with the additional assumptions,

1. $\mathbf{N}(.,x) = (N_1(.,x),\ldots,N_n(.,x))'$ is a multivariate counting process, from which it follows that for any $t \geq 0$ and $i \neq j$,

$$\mathbb{P}(\Delta N_i(t,x) = \Delta N_j(t,x) = 1) = 0.$$

2. For each i, $M_i(.,x) = N_i(.,x) - A_i(.,x)$ is a local martingale with respect to $\{\mathscr{F}_{t,x}; t \geq 0\}$, where $A_i(.,x)$ is the continuous compensator

$$A_i(t,x) = \int_0^t \lambda_0(u,x)e^{\beta'\mathbf{Z}_i(t)}Y_i(u,x)du.$$

3. Each of the censoring processes $Y_i(.,x)$ and covariate processes $\mathbf{Z}_i$ is predictable with respect to $\{\mathscr{F}_{t,x}; t \geq 0\}$ and $\mathbf{Z}_i$ are locally bounded processes.

We also introduce a generalization of some classical and important notations and assumptions in order to take into account a rate of degradation of HrQoL.

For all $k = 0, 1, 2$,

$$\mathbf{S}^{(k)}(\beta,t,x) = \frac{1}{n}\sum_{i=1}^{n}\{\mathbf{Z}_i(t)\}^{\otimes k}Y_i(t,x)\exp\{\beta'\mathbf{Z}_i(t)\}, \tag{1}$$

$$\mathbf{E}(\beta,t,x) = \frac{\mathbf{S}^{(1)}(\beta,t,x)}{S^{(0)}(\beta,t,x)} \tag{2}$$

and

$$\mathbf{V}(\beta,t,x) = \frac{\mathbf{S}^{(2)}(\beta,t,x)}{S^{(0)}(\beta,t,x)} - \{\mathbf{E}(\beta,t,x)\}^{\otimes 2}. \tag{3}$$

with for all vector $\mathbf{X}$, $\mathbf{X}^{\otimes 0} = 1$, $\mathbf{X}^{\otimes 1} = \mathbf{X}$ and $\mathbf{X}^{\otimes 2} = \mathbf{X}\mathbf{X}'$.

Assumptions The following assumptions will be used.

(A1) The time C is so that $\int_0^C \lambda_0(u,x)du < \infty$.

(A2) For $\mathbf{S}^{(j)}$, $j = 0,1,2$ defined in (1) there exists a neighborhood $\mathscr{B}$ of β_0, a neighborhood $\mathscr{X}$ of x and scalar $s^{(0)}$, vector $\mathbf{s}^{(1)}$ and matrix functions $\mathbf{s}^{(2)}$ defined on $\mathscr{B} \times [0,C] \times \mathscr{X}$ so that, for $j = 0,1,2$,

$$\sup_{(\beta,u,x)\in[0,C]\times\mathscr{B}\times\mathscr{X}} ||\mathbf{S}^{(j)}(\beta,u,x) - \mathbf{s}^{(j)}(\beta,u,x)||_\infty$$

converges in probability to 0,

(A3) There exists a $\delta > 0$ such that

$$\frac{1}{\sqrt{n}} \sup_{(i,u,x)\in[0,n]\times[0,C]\times\mathscr{X}} |\mathbf{Z}_i(u)| Y_i(u,x) \mathbf{1}_{\{\beta_0'\mathbf{Z}_i(u) > -\delta|\mathbf{Z}_i(u)|\}}$$

converges in probability to 0,

(A4) Let $\mathscr{B}$, $\mathscr{X}$ and $\mathbf{s}^{(j)}$, $j = 0,1,2$, be as defined in (A2) and let

$$\mathbf{e} = \frac{s^{(1)}}{s^{(0)}}$$

and

$$\mathbf{v} = \frac{s^{(2)}}{s^{(0)}} - \mathbf{e}^{\otimes 2}.$$

Then, for all $\beta \in \mathscr{B}$, $x \in \mathscr{X}$ and $0 \le u \le C$,

$$\frac{\partial}{\partial\beta} s^{(0)}(\beta,u,x) = \mathbf{s}^{(1)}(\beta,u,x)$$

and

$$\frac{\partial^2}{\partial\beta^2} s^{(0)}(\beta,u,x) = \mathbf{s}^{(2)}(\beta,u,x).$$

(A5) The functions $\mathbf{s}^{(j)}$, $j = 0,1,2$, are bounded on $\mathscr{B} \times [0,C] \times \mathscr{X}$ and, moreover, $s^{(0)}$ is bounded away from 0 on $\mathscr{B} \times [0,C] \times \mathscr{X}$; for $j = 0,1,2$, the family of functions $\mathbf{s}^{(j)}(.,u,x), 0 \le u \le C$ and $x \in \mathscr{X}$ is an equicontinuous family (Rudin, 1974) at β_0.

(A6) The matrix

$$\Sigma(\beta_0,C,x) = \int_0^C \mathbf{v}(\beta_0,u,x) s^{(0)}(\beta_0,u,x) \lambda_0(u,x) du$$

is strictly positive definite.

4. PARTIAL LIKELIHOOD SCORE STATISTICS FOR FIXED x

Let $D[0,C]$ be the space of functions on $[0,C]$ which are right continuous with left hand limits. We can now state the partial likelihood score statistic for fixed x.

Proposition 2:

Suppose that Assumptions (A1)-(A6) hold for the multiplicative intensity model. We have the following properties.

1. The normalized vector score processes $\{n^{-1/2}\mathbf{U}(\beta_0,t,x);0\leq t\leq C\}$ whose value at time t is

$$\frac{1}{\sqrt{n}}\mathbf{U}(\beta_0,t,x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t\{\mathbf{Z}_i(u)-\mathbf{E}(\beta_0,u,x)\}N_i(du,x),$$

converges weakly in $(D[0,C])^p$ to a mean zero p-variate Gaussian process so that each component process has independent increment and the covariance function at instant t for components l and l' is

$$\Sigma(\beta_0,t,x)_{ll'}=\int_0^t\{\mathbf{v}(\beta_0,u,x)\}_{ll'}s^{(0)}(\beta_0,s)\lambda_0(u,x)du;$$

2. If $\hat{\beta}$ is a consistent estimator of β_0, then

$$\sup_{0\leq t\leq C}||\frac{1}{n}\int_0^t\sum_{i=1}^{n}\mathbf{V}(\hat{\beta},u,x)N_i(du,x)-\Sigma(\beta_0,t,x)||_\infty,$$

where $\mathbf{V}(\hat{\beta},.,x)$ is defined like in (3), converges in probability to 0.

The information matrix $\mathscr{I}$ can be defined by

$$\mathscr{I}(\beta,t,x)=\int_0^t\sum_{i=1}^{n}\mathbf{V}(\beta,u,x)N_i(du,x).$$

So part (2) of *Theorem* 5.3.5 in Fleming and Harrington (1991) shows that $n^{-1}\mathscr{I}(\beta_0,t,x)$ and $n^{-1}\mathscr{I}(\hat{\beta},t,x)$ are uniformly on $[0,C]$ consistent estimators of the variance function $\Sigma(\beta_0,C,x)$ of the score process at $\beta=\beta_0$.

The consistency and asymptotic normality of the maximum partial likelihood estimator $\hat{\beta}$ of β and the asymptotic distribution of $\sqrt{n}(\hat{\Lambda}_0-\Lambda_0)$, where $\hat{\Lambda}_0$ is a generalization of the estimator proposed by Breslow (1972), are obtained (Boisson and Mesbah, 2008).

5. LOG-RANK STATISTICS FOR FIXED x

Let A and B be two groups of patients (undergoing different treatments for example). In this part covariates are reduced to $Z_i = \mathbf{1}_{(i\in B)}$ the indicator variable of group B. Let $n_A(t,x) = \sum_{i\in A} \mathbf{1}_{(T_i(x)\geq t)}$ (respectively $n_B(t,x) = \sum_{i\in B} \mathbf{1}_{(T_i(x)\geq t)}$) the number of patients at risk (i.e. not degraded) in group A (respectively B) at time t for a fixed rate x of degradation of HrQoL and let $n(t,x) = n_A(t,x) + n_B(t,x)$ be the total number of not degraded patients. The null hypothesis of test is $H_0 : S_A(t,x) = S_B(t,x)$ where the survival functions are identical for the two groups versus the alternative hypothesis $H_1 : S_A(t,x) \neq S_B(t,x)$ where the survival functions are not the same.

Define the log-rank statistic $U_n(x)$ and its variance estimator $\hat{Var}(U_n(x))$

$$U_n(x) = \sum_{i=1}^{n} \left\{ Z_i - \frac{n_B(T_i(x),x)}{n(T_i(x),x)} \right\}$$

and

$$\hat{Var}(U_n(x)) = \sum_{i=1}^{n} \frac{n_A(T_i(x),x) n_B(T_i(x),x)}{n^2(T_i(x),x)}.$$

The asymptotic distribution of the test statistic and consistency of its variance estimator under the null hypothesis can be obtained similarly to those given in Fleming and Harrington (1991).

Proposition 3:

For any fixed x, $n^{-1/2}U_n(x)$ converges in distribution to $\mathcal{N}(0, \sigma^2_{U(x)})$. Furthermore, $n^{-1}\hat{Var}(U_n(x))$ converge to $\sigma^2_{U(x)}$.

6. LOG-RANK STATISTIC AS A STOCHASTIC PROCESS WHEN x IS VARYING

We now consider the case that x, the rate of degradation of HrQoL, is varying. We establish below that the test statistic $U_n(x)$, as a function of x and normalized by $\sqrt{n}$, converges weakly to a zero-mean Gaussian process. Instead of applying the martingale central limit theorem, the proof involves using empirical process theory.

Proposition 4:

The score process $n^{-1/2}U_n(.)$ converges weakly to a Gaussian process with mean 0 and covariance function Γ, which is defined below.

Proof. Since patients are independent, the random variables Z_i are independent. We have

$$U_n(x) = \sum_{i=1}^{n} \left\{ Z_i - \frac{\sum_{j=1}^{n} Z_j \mathbf{1}_{(T_j(x) \geq T_i(x))}}{\sum_{j=1}^{n} \mathbf{1}_{(T_j(x) \geq T_i(x))}} \right\} = \sum_{i=1}^{n} \int \left\{ Z_i - \frac{\sum_{j=1}^{n} Z_j \mathbf{1}_{(T_j(x) \geq T_i(x))}}{\sum_{j=1}^{n} \mathbf{1}_{(T_j(x) \geq T_i(x))}} \right\} d\mathbf{1}_{(T_i(x) \leq t)}$$

$$= \sum_{i=1}^{n} \int \{Z_i - \overline{Z}(t,x)\} \{ d\mathbf{1}_{(T_i(x) \leq t)} - \mathbf{1}_{(T_i(x) \geq t)} d\Lambda_x(t) \}$$

where $\overline{Z}(t,x) = \frac{\sum_{j=1}^{n} Z_j \mathbf{1}_{(T_j(x) \geq t)}}{\sum_{j=1}^{n} \mathbf{1}_{(T_j(x) \geq t)}}$ and $\Lambda_x(t)$ is the cumulative hazard of the random time $T_i(x)$. It is easy to see that $M_i(t,x) = \mathbf{1}_{(T_i(x) \leq t)} - \int_0^t \mathbf{1}_{(T_i(x) \geq s)} d\Lambda_x(s)$ is a martingale. Consequently

$$\frac{1}{\sqrt{n}} U_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \{Z_i - \overline{Z}(t,x)\} M_i(dt,x).$$

For all fixed rate of degradation of HrQoL x, $n^{-1/2} U_n(x)$ is asymptotically a Gaussian variable and we will prove below that $n^{-1/2} U_n(.)$ is asymptotically a Gaussian process.

Under the hypothesis H_0 the random variables $T_i(x)$ are independent and identical distributed. Let $g_Z(t,x) = \lim_{n \to \infty} \overline{Z}(t,x)$, we obtain

$$\frac{1}{\sqrt{n}} U_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \{Z_i - g_Z(t,x)\} M_i(dt,x) + o_p(1),$$

a sum of independent random variables of mean zero, ignoring the $o_p(1)$. Furthermore $T_i(x)$ monotone in x and for $Z_i = \mathbf{1}_{(i \in B)}$, there exists a constant C such that $\left\| Z_i \right\| 1 + \int_0^C \left\| dZ_i \right\| 1 \leq C$ and for each k=0,1,2, for all t,x and $\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[Z_i^{\otimes k} Y_i(t,x) \exp(\beta_0' Z_i) \right]$ exists by the law of large numbers. So by applying Theorem 3.2 of Bilias, Gu and Ying (1997), n-$^{1/2} U_n(.)$ converges weakly to a Gaussian process with covariance function Γ defined as

$$\Gamma(x,x') = \lim_{n \to \infty} \mathbb{E}\left(\frac{1}{n} U_n(x) U_n(x') \right)$$

$$= \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\int \{(Z_i - \overline{Z}(t,x)) M_i(t,x) dt\} \times \int \{(Z_i - \overline{Z}(t,x')) M_i(t,x') dt\} \right].$$

Let $\hat{\Gamma}$ be an estimator of Γ defined as

$$\hat{\Gamma}(x,x') = \frac{1}{n} \sum_{i=1}^{n} \left[\int \{(Z_i - \overline{Z}(t,x)) \hat{M}_i(t,x) dt\} \times \int \{Z_i - \overline{Z}(t,x) \hat{M}_i(t,x') dt\} \right], \quad (4)$$

where $\hat{M}_i$ are the same as M_i except with Λ replaced by the usual Nelson-Aalen estimator.

Proposition 5:

The estimator $\hat{\Gamma}$ is consistent, i.e. $\hat{\Gamma}(x,x')$ converges in probability to $\Gamma(x,x')$ for all x and x'.

Proof. The result is a consequence of the consistency of the Nelson-Aalen estimator and repeated applications of the law of large numbers.

7. GLOBAL LOG-RANK TEST

Although we can estimate Γ by $\hat{\Gamma}$, most statistics based upon the score process will have complicated limiting distributions. Specifically, we are interested in Kolmogorov-Smirov-type supremum test statistic $\sup_{0<x<1}|n^{-1/2}U_n(x)|$. Indeed, its limiting distribution does not have any readily usable analytical form so that its p-value can be calculated. We propose to a simple simulation based approach for p-value approximation.

Consider $(\xi_1,\ldots,\xi_n)$ a vector of n independent and identically distributed random variables of $\mathscr{N}(0,1)$ and let

$$\frac{1}{\sqrt{n}}U_n^*(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\int\{Z_i-\overline{Z}(t,x)\}d\hat{M}_i(t,x).$$

The covariance matrix

$$Cov\left(\frac{1}{\sqrt{n}}U_n^*(x)\ ,\frac{1}{\sqrt{n}}U_n^*(x')\right)=$$

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\int\{Z_i-\overline{Z}(t,x)\}d\hat{M}_i(t,x)\ \times\int\{Z_i-\overline{Z}(t,x')\}d\hat{M}_i(t,x')\right]$$

$$=\hat{\Gamma}(x,x')\rightarrow\Gamma(x,x'),\tag{5}$$

as $n\rightarrow\infty$, where the first equality follows since ξ_i are i.i.d. with mean zero and variance 1 and are independent of the data. Thus, it follows that $n^{-1/2}U_n^*(x)$ has the same asymptotic covariance and consequently the same asymptotic distribution as $n^{-1/2}U_n(x)$. Thus we can approximate the distribution of $n^{-1/2}U_n(x)$ by simulating a large number (say 1000 or 10000) of realizations of $n^{-1/2}U_n^*(x)$ by repeatedly generating random vectors $(\xi_1,\ldots,\xi_n)$ while holding the observed data $(X_i(x),\delta_i(x),\mathbf{Z}_i)$ fixed. We then obtain a nonparametric statistical test for treatment effect of longitudinal evolution of HrQoL and survival.

With the above results, we now define our global log-rank test as the supremum test statistic $S \equiv \sup_{0<x<1} \left| n^{-1/2} U_n(x) \right|$. To find its p-value, suppose that s is the observed value of S. Let $\hat{S} = \sup_{0<x<1} \left| n^{-1/2} U_n^*(x) \right|$. The p-value of the global log-rank test $P(S > s)$ can be approximated by $P(\hat{S} > s)$, which can be estimated by a great number of realizations (say 1000 or 10000) from $\hat{S}$.
The idea of using simulation based processes to find the p-value of a test statistic whose distribution may not be analytically tractable dates back to Lin, Wei and Ying (1993). See also Lin, Wei and Ying (2002), Pan and Lin (2005) and Martinussen, Aalen and Scheike (2008).

8. SIMULATIONS

We conducted simulations to evaluate performance of the proposed global log-rank-type test. They are implemented using SAS. To that end different data sets which represent longitudinal HrQoL evolution are simulated for two groups. We consider that patients' HrQoL evolves linearly, i.e. $Q_t = a_0 + b_0 t$ for group 0 and $Q_t = a_1 + b_1 t$ for group 1. We also consider two scenarios to convey variability of patients' answers compared with the theoretical straight line. In the first one, slope and initial score of HrQoL are disrupted by a white noise consisting of both random effect and fixed effect, whereas in the other one, slope and initial score of HrQoL are disrupted by an autoregressive process (AR(1)) (Doob, 1953). We also simulate missing data and dropout by Bernoulli law of parameters 0.90 and 0.95 respectively. Furthermore, we consider two cases, in the first one evolutions of the two groups of patients are different with $a_0 = 28$, $a_1 = 40$, $b_0 = 1$ and $b_1 = 15$ whereas in the second we make the null hypothesis H_0 with $a_0 = a_1 = 28$, $b_0 = b_1 = 1$. We examine different situations in varying one of the following parameters: size of sample, number of dates, percentage of each group, percentage of dropout and missing data and number of simulated curves. We also consider each definition of degradation of HrQoL, $D_t = (Q_0 - Q_t)Q_0^{-1}$ and $D_t = Q_0 - Q_t$. We consider the following values of reference: 300 patients, 7 dates of visits, 100 simulated curves, same number of patients in each group and with presence of dropout and missing data. Indeed these values correspond to classical situations in practice (Protopopescu, 2007).

Results of simulations (see Table 1) show that the test is operational. Despite the fact that there is often a misunderstanding in its interpretation, we choose to present results of our simulations in terms of estimated p-values, which remain, in real applied statistics, the gold standard. The right part of current Table 1, is built, to estimate the p-value, when data come from a true underlying distribution

Table 1: Results of simulations for the following values of reference: 300 patients, 7 dates, 100 simulated curves, same number of patients in each group and with presence of dropout and missing data

	First case: $a_0 = 28$, $a_1 = 40$, $b_0 = 1$ and $b_1 = 15$				Second case: $a_0 = a_1 = 28$, $b_0 = b_1 = 1$			
	$\hat{p}$-value for white noise		$\hat{p}$-value for AR(1)		$\hat{p}$-value for white noise		$\hat{p}$-value for AR(1)	
Degradation	$(Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$	$(Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$	$(Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$	$(Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$
Number of patients								
30	0,05	0,02	0,02	0,02	0,83	0,95	0,41	0,56
300	0,07	0,01	0,03	$< 0,01$	0,72	0,74	0,48	0,49
500	0,06	0,01	0,01	0,03	0,31	0,47	0,47	0,46
1000	0,04	0,02	0,03	0,03	0,23	0,16	0,55	0,43
Number of dates								
5	0,08	0,02	0,06	0,02	0,71	0,60	0,78	0,70
7	0,07	0,01	0,03	$< 0,01$	0,72	0,74	0,48	0,49
10	0,02	0,01	$< 0,01$	0,01	0,61	0,43	0,26	0,29
Percentage of patients in group 0								
20%	0,06	$< 0,01$	0,01	0,10	0,77	0,47	0,07	0,04
33%	0,02	0,02	0,03	0,03	0,34	0,32	0,40	0,47
50%	0,07	0,01	0,03	$< 0,01$	0,72	0,74	0,48	0,49
Dropout								
without dropout	0,05	0,07	0,04	0,06	0,48	0,32	0,32	0,24
with dropout	0,07	0,01	0,03	$< 0,01$	0,72	0,74	0,48	0,49
Number of simulated curves								
100	0,07	0,01	0,03	$< 0,01$	0,72	0,74	0,48	0,49
1000	0,05	0,05	0,03	0,05	0,21	0,34	0,45	0,50

H_0 two confounded straight lines: $a_0 = a_1 = 28$, $b_0 = b_1 = 1$), i.e., to estimate: p = Prob (to reject H_0, using our method, while data are truly H_0) while its left part is built to estimate the p-value when data comes from a true underlying distribution H_1 (two different straight lines: $a_0 = 28, a_1 = 40$, $b_0 = 1, b_1 = 15$), i.e., to estimate: p = Prob (to reject H_0, using our method, while data are truly under a specific alternative H_1). We see "significant" p-value (small) when the two lines are theoretically confounded, and non significant p-values, when they are different. Most of the time, our method detect such departure from the null hypothesis. When simulations are done under the null hypothesis, we see a kind of decreasing of the p-value, when the number of patients or its percentage in group 0 increase.

9. A REAL EXAMPLE

In this section we applied the proposed test to real data from a clinical trial of treatment for metastatic colorectal cancer. In this trial, HrQoL was measured using the QLQ-C30 instrument (Aaronson *et al.*, 1993). It is a questionnaire with thirty items self completed by patients. The QLQ-C30 permits by linear transformations to obtain fifteen scales for analysis. These scales include five functional scales, nine symptom scales and one global health status/quality-of-life scale (QL). We only analyze the QL scale and we characterize degradation of HrQoL by an absorbing state like defined in Section 2. A more detailed description of the study was reported by Awad *et al.* (2002) and Mesbah *et al.* (2004).

In our implementation, the $\hat{p}$-value for supremum test is based on 1000 realizations from $\hat{S}$ when the degradation of HrQoL is defined by $D_t = (Q_0 - Q_t)Q_0^{-1}$. In each graphical display, the observed process is indicated by a solid curve while thirty simulated processes are plotted in dotted curves.

One hundred and twenty-two patients were randomized in group A and one hundred and twenty-one were randomized in group B during one year. We consider time to event of degradation of HrQoL for a fixed rate x as duration in days between baseline and occurrence of the event of degradation of HrQoL for this rate x. Figure 1 plots process $n^{-1/2}U_n$ (solid curve) along with thirty simulated processes $n^{-1/2}U_n^*$ (dotted curves) when the rate x goes over from 0 to 1. It appears that the solid line lies well inside the “band” formed by the dotted lines, except for x close to 0 where the solid line is close to but still somewhat below the upper boundary. This indicates that the proposed global log-rank test is most likely to result in a non-significant p-value at 5% level. Indeed, the $\hat{p}$-value of the global log-rank test is 0.094.

Due to the duration of protocol cycle and therefore the frequency of HrQoL

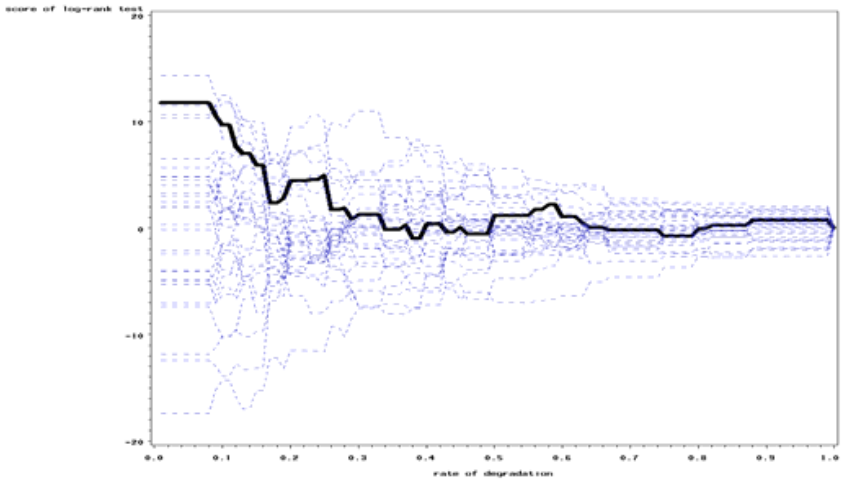

Figure 1: Process $n^{-1/2}U_n$ score according to rate of degradation of HrQoL. Time to event of degradation of HrQoL expressed in days.

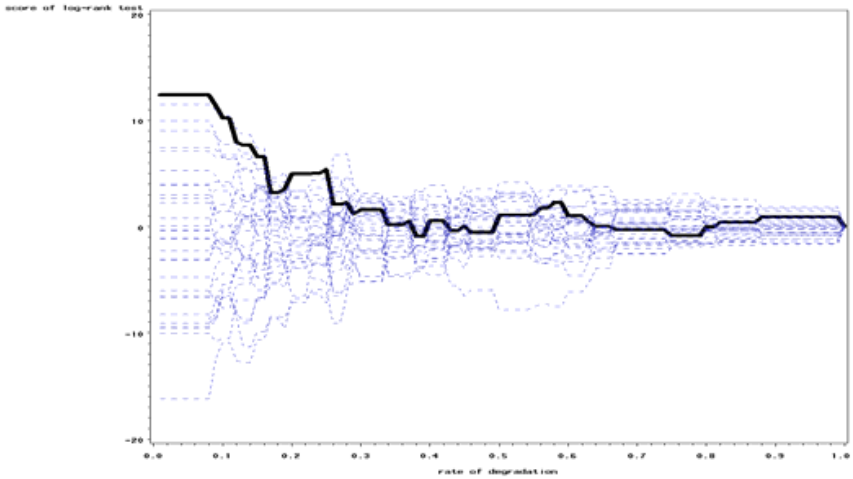

Figure 2: Process $n^{-1/2}U_n$ score according to rate of degradation of HrQoL. Time to event of degradation of HrQoL expressed in cycle.

evolution different for the two groups, analysis of degradation of HrQoL was performed according to time interval which corresponds to seven-week period of chemotherapeutic treatment. This allows the response to questionnaires process to be more balanced between the two treatment groups and more close to the designed clinical trial protocol. Nevertheless, the recorded date in the data is the real date of medical visit, which in practice cannot be scheduled at a pre-specified date for all patients, but is adapted to each patient and doctor. Moreover, the real date of HrQoL's questionnaire is the exact date when the patient fills out the form. This date is not necessary the date of medical visit. Our method allows us to use those exact dates, instead of using specifically chosen cycles as in Awad *et al.* (2002). In Figure 2, the same graphic as in Figure 1 is plotted, with a different time scale, which is now expressed in seven-week periods. In this case, the solid line lies almost at the upper extreme for x values up to about 0.3. It indicates that the resulting test probably attains the statistical significance at 5% level. Indeed, the $\hat{p}$-value of the global log-rank test is 0.032 and therefore the test gives more significant result. This is consistent with Awad *et al.* (2002) results, where no global test was used.

10. SUMMARY AND CONCLUSION

In clinical trials the length of specific disease or treatment stages and evolution of HrQoL are of interest to practitioners. Degradation of HrQoL enables us to reflect evolution of HrQoL as compared to the baseline. This article is motivated by the need to develop statistical methods in order to analyze degradation of HrQoL independently of any previously fixed rate x of degradation of HrQoL. The primary goal of this article is to perform a nonparametric statistical test for evolution of HrQoL to be used in clinical research by practitioners. The resulting log-rank-type test does not make any assumption on process of degradation of HrQoL and on threshold x.

We started by extending the partial likelihood score statistic for survival time depending on a fixed rate x of degradation of HrQoL, under a generalization of classical notations and assumptions, in order to take into account a rate x of degradation of HrQoL, of counting process. We then derived a log-rank statistic which permits to compare time to degradation of HrQoL between two kinds of treatments when rate x of degradation of HrQoL is fixed. We proved that when rate x of degradation of HrQoL is varying this log-rank statistic converges to a Gaussian process. This permits to combine the set of global log-rank tests to an overall global test for treatment effect on the longitudinal evolution of HrQoL. A simula-

tion based resampling method is proposed to approximate the underlying p-value. Simulation results show that this nonparametric test performs reasonably well under practical situations in terms of type I error and power. The method is also applied to a data set from a cancer study, with results consistent with previous analysis.

Because for each individual we have a set of event times $T(x)$, it is obvious that these times are dependent. In fact, $T(x)$ as a function of x is monotone increasing. Therefore the corresponding set of log-rank tests are much related. It would be of interest to investigate different ways to combine these test statistics, as the proposed global test is of Kolmogorov-Smirnoff type. For example, instead of supreme, we may consider an integral type analogous to the Cramer-von Mises test.

References

Aaronson, N.K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez N.J., Filiberti, A., Flechtner, H., Fleishman, S.B., de Haes, J.C.J.M., Kaasa, S., Klee, M.C., Osoba, D., Razavi, D., Rofe, P.B., Schraub, S., Sneeuw, K.C.A., Sullivan, M. and Takeda, F. (1993) The European Organisation for Reserach and Treatment of Cancer Study Group on Quality of Life. The European Organisation for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. In *Journal of the National Cancer Institute*, 85: 365–376.

Andersen, P. and Gill, R. (1982). Cox regression for counting processes: a large sample study. In *Annals of Statistics*, 10: 110–1120.

Andersen, P., Borgan, O., Gill, R. and Keiding, N. (1993).'*Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

Awad, L., Zuber, E. and Mesbah, M. (2002). Applying survival data methodology to analyse longitudinal quality of life. In M. Mesbah, B. Cole and M. Lee, editors, *Statistical Methods for Quality of Life Studies: Design, Measurements and Analysis*, 231-243. Kluwer Academic, Boston.

Bilias, Y., Gu, M. and Ying, Z. (1997). Toward a general asymptotic theory for Cox model with staggered entry. In *Annals of Statistics*, 25(2): 662–682.

Boisson, V. and Mesbah, M. (2008). Test du log-rank et statistique du score de vraisemblance partielle pour une durée de survie dépendant d'un taux de dégradation. In *Comptes Rendus de l'Académie des Science Paris, Series I*, 346: 567–570.

Breslow, N. (1972). Covariance analysis of censored survival data. In *Biometrics*, 30: 89–99.

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

Cox, D.R. (1975). Partial likelihood. In *Biometrika*, 62: 269–276.

Diggle, P., Heagerty, P., Liang, K.-Y. and Zeger, S. (2002). *Analysis of Longitudinal Data* (second edition). Oxford University Press, Oxford.

Doob, J. (1953). *Stochastic Processes*. Wiley, New York.

Duffie, D., Saita, L. and Wang, K. (2007). Multi-period corporate default prediction with stochastic

covariates. In *Journal of Financial Economics*, 83: 635– 665.

Fleming, T. and Harrington, D. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

Hull, J. (2008). *Options, Futures and Other Derivatives*. Prentice Hall, New York.

Lin, D.Y., Wei, L. and Ying, Z. (1993). Checking the Cox model with cumulative sum of martingale-based residuals. In *Biometrika*, 80: 557–572.

Lin, D.Y., Wei, L. and Ying, Z. (2002). Model checking techniques based on cumulative residuals. In *Biometrics*, 58: 1–12.

Martinussen, T., Aalen, O.O and Scheike, T. (2008). The Mizon-Richard encompassing test for the Cox and Aalen additive hazards models. In *Biometrics*, 64: 161–171.

Mesbah, M., Dupuy, J.-F., Heutte, N. and Awad, L. (2004). Joint analysis of longitudinal quality of life and survival processes. In *Advances in survival analysis*, volume 23 of *Handbook of Statistics*, pages 689–728. Elsevier, Amsterdam.

Pan, Z. and Lin, D. (2005). Goodness-of-fit methods for generalized linear mixed models. In *Biometrics*, 61: 1000–1009.

Protopopescu, C. (2007). Health-related quality of life in hiv-1-infected patients on HAART: A five-years longitudinal analysis accounting for dropout in the aproco-copilote cohort anrs co-8. In *Quality of Life Research*, 16: 557–591.

Rudin, W. (1974). *Real and Complex Analysys*. McGraw-Hill, New York.

Tsiatis, A. (1981a). The asymptotic distribution of the efficient scores test for the proportional hazards model calculated over time. In *Biometrika*, 68: 311–315.

Tsiatis, A. (1981b). A large sample study of cox's regression model. In *Annals of Statistics*, 9: 93–108.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.