A model checking method for the proportional hazards model with recurrent gap time data

CHIUNG-YU HUANG*
Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA
huangchi@niaid.nih.gov

XIANGHUA LUO
Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA

DEAN A. FOLLMANN
Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

SUMMARY
Recurrent events are the natural outcome in many medical and epidemiology studies. To assess covariate effects on the gaps between consecutive recurrent events, the Cox proportional hazards model is frequently employed in data analysis. The validity of statistical inference, however, depends on the appropriateness of the Cox model. In this paper, we propose a class of graphical techniques and formal tests for checking the Cox model with recurrent gap time data. The building block of our model checking method is an averaged martingale-like process, based on which a class of multiparameter stochastic processes is proposed. This maneuver is very general and can be used to assess different aspects of model fit. Numerical simulations are conducted to examine finite-sample performance, and the proposed model checking techniques are illustrated with data from the Danish Psychiatric Central Register.

Keywords: Correlated failure times; Induced-dependent censoring; Kaplan–Meier estimator; Renewal processes.

1. INTRODUCTION
Recurrent events are frequently encountered in many biomedical and epidemiology studies, where clinical important events occur repeatedly over the course of follow-up. In many applications, the time between consecutive events, that is the gap time, is a natural outcome of interest. For example, the gap between psychiatric admissions can be viewed as one dimension of the course of schizophrenia, as a decreasing trend in the length between admissions suggests a progressive course of the disorder. When events are

*To whom correspondence should be addressed.
of the same type, Wang and Chang (1999) and Peña and others (2001) generalized the Kaplan–Meier estimator to estimate the common marginal survival function of gap times, and Huang and Chen (2003), Chang (2004), Lu (2005), and Sun and others (2006) considered semiparametric estimation problems. When events are of different types, various nonparametric methods, such as Visser (1996), Wang and Wells (1998), Lin and others (1999), and Huang and Wang (2005), as well as semiparametric methods, such as Huang (1999) and Chang (2000), have been developed in the literature.

This work is motivated by the analysis of psychiatric admissions from the Danish Psychiatric Central Register (DPCR), a nationwide psychiatric case registry in Denmark that recorded all psychiatric contacts since 1969. Researchers have shown that the risk of readmission due to relapse of schizophrenia is stable over time (Olesen and Mortensen, 2002; Olesen and Parner, 2006). Moreover, marginal Cox proportional hazards models have been employed to assess the effects of baseline covariates, such as onset age, gender, and marital status, on the risk of rehospitalization. The validity of statistical inference, however, depends critically on the adequacy of the assumed Cox model. While various model checking methods have been suggested for univariate survival time data (Barlow and Prentice, 1988; Therneau and others, 1990; Lin and others, 1993) and for clustered survival time data (Spiekerman and Lin, 1996), these techniques are not appropriate for recurrent gap time data due to its peculiar stochastic structure. To our knowledge, no formal model checking procedures have been studied in the literature for assessing the Cox model in the gap time setting.

In this article, we introduce diagnostic methodologies for the Cox model with recurrent gap times. In Section 2, we briefly review the pseudo-partial likelihood estimator studied in Huang and Chen (2003). The model checking methods based on the averaged martingale-like process are described in Section 3. We present results of simulation studies in Section 4 and illustrate the proposed methods with the DPCR data in Section 5. Some concluding remarks are given in Section 6.

2. RECURRENT EVENT DATA ANALYSIS

2.1 Model and data structure

Suppose \( n \) independent subjects are recruited into a longitudinal study after experiencing an initial event and the recurrences of the same event are recorded during the study period. Let \( i \) be the index for a subject, and let \( j \) index the sequence of recurrent events of an individual, where \( j = 0 \) indexes the initial event. Denote by \( T_{ij} \) the gap time between the \( j - 1 \)th and \( j \)th events of the \( i \)th subject.

In practice, subjects are usually sampled independently, but gap times from the same subject could be correlated. We assume that there exists a subject-specific frailty variable \( \gamma_i \) so that, conditioning on \( \gamma_i \), the individual recurrent event process is a renewal process; that is, on the individual level, the gap times \( T_{i1}, T_{i2}, \ldots \) are independent and identically distributed (iid). Thus, unconditioning on \( \gamma_i \), the recurrent gap times, \( T_{i1}, T_{i2}, \ldots \) are identically (but not independently) distributed. Note that, under our setting, both the distribution of the subject-specific frailty variable \( \gamma_i \) and the relationship between \( \gamma_i \) and \((T_{i1}, T_{i2}, \ldots)\) are left unspecified. Let \( \mathbf{Z}_i = (Z_{i1}, \ldots, Z_{pi})^T \) be a vector of time-independent covariates associated with subject \( i \). To assess the effect of \( \mathbf{Z}_i \) on the distribution of the gap times, we further assume that the common hazard function of the gap times \( T_{ij}, j = 1, 2, \ldots \) satisfies the Cox proportional hazards model

\[
\lambda(t \mid \mathbf{Z}_i) = \lambda_0(t) \exp(\mathbf{Z}_i^T \beta),
\]

where \( \beta \) is a \( p \times 1 \) vector of parameters and the baseline hazard function \( \lambda_0(t) \) is an arbitrary function with \( \Lambda_0(t) = \int_0^t \lambda_0(u) \, du \). The observation of recurrent events is subject to right censoring. We denote by \( C_i \) the time between the initial event and the end of follow-up, and assume that \( C_i \) is independent.
of \( \gamma_i \) and \((T_{i1}, T_{i2}, \ldots)\) conditional on \( Z_i \). Let \( m_i \geq 1 \) denote the index satisfying \( \sum_{j=1}^{m_i-1} T_{ij} \leq C_i \) and \( \sum_{j=1}^{m_i} T_{ij} > C_i \), with \( \sum_{j=1}^{0} \cdot = 0 \). Thus, \( m_i \) is the number of observed gap times, where the first \( m_i - 1 \) gap times are completely observed and the last gap time is censored by \( C_i - \sum_{j=1}^{m_i-1} T_{ij} \). Denote the observed data for subject \( i \) by \( \{Y_{i1}, \ldots, Y_{im_i}; m_i, Z_i\} \), where \( Y_{ij} = T_{ij} \) for \( j = 1, \ldots, m_i - 1 \) and \( Y_{im_i} = C_i - \sum_{j=1}^{m_i-1} T_{ij} \). The observed data across the \( n \) subjects are assumed to be iid.

Intuitively, the observed recurrent gap time data can be viewed as clustered survival time data. However, as pointed out by Wang and Chang (1999), the sequential structure of recurrent events presents many challenges for statistical analysis. For example, only the last gap time is censored, and the censored gap time \( T_{im_i} \) tends to be longer than the uncensored ones due to intercept sampling. Moreover, the gap time \( T_{ij}, j > 1 \), is subject to dependent censoring \( C_i - T_{i1} - \ldots - T_{ij-1} \), as the gap times are correlated. Finally, subjects who are at a higher risk of recurrent event tend to have more recurrent events; thus the cluster size \( m_i \) is informative about the gap time distribution. Hence, methods for clustered survival time data cannot be directly applied to analyze recurrent gap time data due to its peculiar structure.

### 2.2 Estimating the Cox model

We introduce 2 empirical processes for the estimation procedure of the Cox model with recurrent gap time data. Define \( \Delta_i = I(m_i > 1) \) and \( m_i^* = \max\{m_i - 1, 1\} \). Hence, for people with no observed events \( \Delta_i = 0 \) and \( m_i^* = 1 \), and for people with observed events \( \Delta_i = 1 \) and \( m_i^* \) is the number of uncensored gap times. Let \( N_{ij}(t) = \Delta_i I(Y_{ij} \leq t) \) and \( R_{ij}(t) = I(Y_{ij} \geq t) \). For subject \( i \), we define the “averaged counting process” and the “averaged at-risk process” as

\[
N_i^*(t) = \frac{1}{m_i^*} \sum_{j=1}^{m_i^*} N_{ij}(t) \quad \text{and} \quad R_i^*(t) = \frac{1}{m_i^*} \sum_{j=1}^{m_i^*} R_{ij}(t),
\]

respectively. Note that if subject \( i \) has observed events, a weight \( 1/m_i^* \) is assigned to each of the \( m_i^* \) uncensored gaps, and the last censored gap is not used in the formulation of \( N_i^*(t) \) and \( R_i^*(t) \). Thus, \( N_i^*(t) \) and \( R_i^*(t) \) have jump sizes of \( 1/m_i^* \) and \( -1/m_i^* \), respectively, at each uncensored gap time, which is different from the univariate survival analysis where the jump size of the individual counting and at-risk processes is always 1 and -1.

To estimate the Cox model, Huang and Chen (2003) proposed to solve the pseudo-partial score equation

\[
U^*(b) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \left( Z_i - \frac{S^{(1)}(b, u)}{S^{(0)}(b, u)} \right) dN_i^*(u),
\]

where \( S^{(k)}(b, t) = n^{-1} \sum_{i=1}^{n} Z_i \otimes_k \exp(Z_i^T b) R_i^*(t), \) \( k = 0, 1, 2 \) and \( \tau \) satisfies \( \Pr(Y_{i1} \geq \tau) > 0 \). Note that we define \( a \otimes 0 = a, a \otimes 1 = a \) and \( a \otimes 2 = aa^T \) for a column vector \( a \). The pseudo-partial likelihood estimator \( \hat{\beta}^* \) is defined as the solution to \( U^*(b) = 0 \). As pointed out by Wang and Chang (1999), for an individual with \( m_i > 1 \), the uncensored gap times \( Y_{i1}, \ldots, Y_{im_i} \) of subject \( i \) are identically distributed conditional on \( \gamma_i, m_i \) and \( C_i \). By utilizing this property, it can be shown that \( U^* \) and the partial score function derived by using only the first gap times \( \{(Y_{i1}, Z_i, \Delta_i), i = 1, \ldots, n\} \) converge uniformly to the same limit, ensuring the consistency of \( \hat{\beta}^* \).
3. Model checking techniques

3.1 The averaged residual process and the multiparameter stochastic processes

Following the spirit of conventional counting process martingales, for subject \( i \) we define the averaged martingale-like process

\[
M^*_i(t) = N^*_i(t) - \int_0^t \exp(Z_i^T \beta) R^*_i(u) \, d\Lambda^*_0(u).
\]

With the use of \( M^*_i(t) \), the asymptotic representation of the pseudo-partial likelihood estimator \( \hat{\beta}^* \) can be expressed as

\[
\hat{\beta}^* - \beta = \Gamma(\beta)^{-1} \times \frac{1}{n} \sum_{i=1}^n \int_0^t \{ Z_i - E(\beta, u) \} \, dM^*_i(u) + o_p(n^{-1/2}),
\]

where \( \Gamma(\beta) = -E[dU^*(\beta)/d\beta] \) is assumed to be positive definite at \( \beta = \beta^* \) and \( E(\beta, t) \) is the limit of \( S(1)(b, t)/S(0)(b, t) \) as \( n \to \infty \).

The averaged martingale-like process \( M^*_i(t) \) is expected to be informative about model misspecifications because it is the difference between the observed and the expected proportion of uncensored gaps that are shorter than \( t \). Note, however, that \( M^*_i(t) \) is not a martingale because of the sequential structure of recurrent events. The stochastic process \( M^*_i(t) \) can be estimated by the averaged residual process

\[
\hat{M}^*_i(t) = N^*_i(t) - \int_0^t \exp(Z_i^T \hat{\beta}^*) R^*_i(u) d\hat{\Lambda}^*_0(u),
\]

where the estimated baseline cumulative hazard function \( \hat{\Lambda}^*_0(t) \) is given by the Breslow-type estimator \( \hat{\Lambda}^*_0(b, t) = \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{dN^*_i(u)}{S^*_0(u)} \).

The averaged residual process is different from the martingale residual processes considered by Lin and others (1993) and Spiekerman and Lin (1996), as it is constructed based on correlated gap times and is inversely weighted by the corresponding cluster size \( m^*_i \). Interestingly, the averaged residual process \( \hat{M}^*_i(t) \) behaves similarly to the ordinary martingale residual process in that, for any \( t \), \( \sum_{i=1}^n \hat{M}^*_i(t) = 0 \) and \( E[\hat{M}^*_i(t)] \approx 0 \) for large \( n \).

We now describe objective diagnostic techniques for the Cox model by grouping the averaged residuals cumulatively with respect to follow-up time and/or covariate values. Specifically, we consider a general class of multiparameter stochastic processes

\[
W^*(t, z) = \frac{1}{n} \sum_{i=1}^n f(Z_i) I(Z_i \leq z) \hat{M}^*_i(t),
\]

where \( f(\cdot) \) is a prespecified bounded function and \( I(Z_i \leq z) = I(Z_{1i} \leq z_1, \ldots, Z_{pi} \leq z_p) \). Under the assumed Cox proportional hazards model, the processes are expected to fluctuate randomly around zero. We show in Appendix A of the supplementary material (available at Biostatistics online) that the process \( \sqrt{n}W^*(t, z) \) is asymptotically equivalent to \( \sqrt{n} \Psi^*_i(t, z) = n^{-1/2} \sum_{i=1}^n \Psi^*_i(t, z) \), where

\[
\Psi^*_i(t, z) = \int_0^t \{ f(Z_i) I(Z_i \leq z) - \hat{w}(\beta, u, z) \} \, dM^*_i(u)
\]

\[
-\hat{H}(\beta, t, z) \Gamma(\beta)^{-1} \int_0^t \{ Z_i - E(\beta, u) \} \, dM^*_i(u).
\]
Note that \( \tilde{w}(b, t, z) \) and \( \tilde{H}(b, t, z) \) are the \( n \)-limits of
\[
\tilde{w}^*(b, t, z) = \frac{1}{n} \sum_{i=1}^{n} f(Z_i) I(Z_i \leq z) \exp(Z_i^T b) R_i^*(t) / S^{(0)*}(b, t)
\]
and
\[
\tilde{H}^*(b, t, z) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{t} f(Z_i) I(Z_i \leq z) \exp(Z_i^T b) R_i^*(u) \left\{ Z_i - \frac{S^{(1)*}(b, u)}{S^{(0)*}(b, u)} \right\} d\hat{\Lambda}_i^*(b, u).
\]

It can be shown that, as \( n \to \infty \), the multiparameter process \( \sqrt{n} W^*(\cdot, \cdot) \) converges to a zero-mean Gaussian process with covariance function \( E\{\Psi^*_i(t_1, z_1) \Psi^*_i(t_2, z_2)^T\} \) for any pairs \( (t_1, z_1) \) and \( (t_2, z_2) \). Replacing the unknown quantities in \( \Psi^*_i \) by the corresponding sample estimates, we define the function
\[
\hat{\Psi}^*_i(t, z) = \int_{0}^{t} \{ f(Z_i) I(Z_i \leq z) - \tilde{w}^*(\hat{\beta}^*, u, z)\} d\hat{M}_i^*(u) - \hat{H}^*(\hat{\beta}^*, t, z)^T \hat{\Gamma}^* \hat{\beta}^* dN_i^*(u),
\]
where
\[
\hat{\Gamma}^* = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{t} \left\{ \frac{S^{(2)*}(\hat{\beta}^*, u)}{S^{(0)*}(\hat{\beta}^*, u)} - \left\{ \frac{S^{(1)*}(\hat{\beta}^*, u)}{S^{(0)*}(\hat{\beta}^*, u)} \right\}^2 \right\} dN_i^*(u).
\]

In Appendix B of the supplementary material (available at Biostatistics online), we show that \( E\{\Psi^*_i(t_1, z_1) \Psi^*_i(t_2, z_2)^T\} \) can be consistently estimated by \( \frac{1}{n} \sum_{i=1}^{n} \hat{\Psi}^*_i(t_1, z_1) \hat{\Psi}^*_i(t_2, z_2)^T \).

In practice, we can approximate the limiting distribution of \( W^*(t, z) \) through Monte-Carlo simulations. Let \( G_1, \ldots, G_n \) be independent standard normal random variables which are independent of the data. Define \( \tilde{W}^*(t, z) = n^{-1} \sum_{i=1}^{n} \hat{\Psi}^*_i(t, z) G_i \). Conditional on the data, and thus conditioned on \( \hat{\Psi}^*_i(t, z) \), \( \tilde{W}^*(t, z) \) is a sum of \( n \) independent, zero-mean random vectors for every \( (t, z) \). Arguing as in section 2.4 of Spiekerman and Lin (1996), we can show that \( \sqrt{n} W^*(t, z) \) and \( \sqrt{n} \tilde{W}^*(t, z) \) converge weakly to the same limiting process. Hence, to approximate the null distribution of \( W^*(t, z) \), we can simulate a number of realizations from \( \tilde{W}^*(\cdot, \cdot) \) by repeatedly generating standard normal random samples \( (G_1, \ldots, G_n) \).

To assess how unusual the observed residual pattern is under the assumed model, we plot the observed multiparameter process along with a few, say 20, processes simulated from \( \tilde{W}^*(t, z) \). An approximate \( P \)-value for the supremum test \( \sup_{t, z} |W^*(t, z)| \) can be estimated by the empirical probability that the observed value of \( \sup_{t, z} |\tilde{W}^*(t, z)| \) exceeds the observed value of \( \sup_{t, z} |W^*(t, z)| \) through simulation.

### 3.2 Checking the Cox model

Two important aspects of the Cox model can contribute to the lack of fit to the data: the functional form of covariates might be misspecified, and the proportional hazards assumption may fail. In this section, we demonstrate the use of the results described in Section 3.1 in examining these 2 aspects of the Cox model by considering 2 special cases of \( W^*(\cdot, \cdot) \).
Functional form of covariates. When fitting a Cox model, one would like to determine the functional form of a covariate that best explains its effect on the gap time distribution. For example, we may want to assess whether the effect of schizophrenia onset age is linear on the logged hazard. The appropriateness of the functional form of the $k$th component of $Z$ can be examined by plotting the partial-mean process for $Z_{ki}$,

$$W^*_{z}(z) = \frac{1}{n} \sum_{i=1}^{n} I(Z_{ki} \leq z) \hat{M}_i^*(\tau),$$

against $z$. If the functional form of $Z_{ki}$ is appropriate, then the observed process is expected to fluctuate around zero for all values of $z$; otherwise one would expect a systematic trend. Note that the partial-mean process $W^*_{z}(z)$ is a special case of $W^*(t, z)$, with $f(\cdot) = 1$ and $t = \tau$ in (3.1). Hence, the null distribution of $W^*_z$ can be approximated through simulating the corresponding zero-mean Gaussian process $\hat{W}^*_z(z)$, and the $P$ value can be obtained by using the supremum test $\sup_z |W^*_z(z)|$.

Figure 1 shows plots of the partial-mean process, along with 20 realizations of the null distribution, when fitting a proportional hazards model $\lambda(t|Z) = \lambda_0(t) \exp(\beta Z)$ to data generated from various alternative models. Figure 1(a) shows that the observed partial-mean process falls well inside the range determined by the simulated processes when fitting the Cox model with the correct functional form. Figure 1(b) shows that the observed partial-mean process has a curvilinear pattern when a quadratic term is missing from the fitted model. The fitted model overestimates the hazards for lower end values of $Z$.

Fig. 1. Plots of the partial-mean process $W^*_z(z)$ based on fitting the model $\lambda(t|Z) = \lambda_0(t) \exp(Z\beta)$ for a single data set of 100 subjects with correlated gap times generated from various models. In all scenarios, the proportion of subjects without any complete gap times is about 25%. The covariate $Z \sim$ uniform $(0, 10)$ in models (a), (b), and (f); $\sqrt{Z} \sim$ uniform$(0, 10)$ in model (c); $\log(Z) \sim$ uniform$(0, 5)$ in model (d); and $\exp(Z) \sim$ uniform$(0, 10)$ in model (e). The bold curve corresponds to the observed process and the gray curves are 20 (out of 1000) approximating processes.
and underestimates the hazards for large values of Z, suggesting that a concave function of Z could fit better than the linear function of Z. Figure 1(c) and (d) consider 2 other examples of concave functions, \( \sqrt{Z} \) and \( \log(Z) \), and it is easy to see that the pattern is similar to Figure 1(b). Figure 1(e) gives an example of convex function \( \exp(Z) \), and its pattern is opposite to Figure 1(b). Finally, Figure 1(f) considers a dichotomized Z. Interestingly, the observed partial-mean process has a sharp drop near the cutoff value of Z, and falls inside the range determined by the simulated processes when moving away from the cutoff value.

**Proportional hazards assumption.** In evaluating the effect of electroshock therapy on schizophrenia patients, it is often assumed that the hazard ratio of the intervention group versus the non-intervention group is constant over time. However, the results of Cox regression can be misleading if proportional hazards assumption is not valid. To check the proportional hazards assumption, we consider the pseudo-score process

\[
\hat{U}^* (\hat{\beta}^*, t) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{t} \left\{ Z_i - \frac{S^{(1)}(\hat{\beta}^*, u)}{S^{(0)}(\hat{\beta}^*, u)} \right\} dN_i^* (u),
\]

which is also a special case of \( W^*(t, z) \), with \( f(Z_i) = Z_i \) and \( z = \infty \) in (3.1). The results stated before for \( W^*(t, z) \) can be used to simulate the distributions of the standardized pseudo-score process \( \left\{ \hat{\Omega}^* (\hat{\beta}^*) \right\}^{-1/2} \sqrt{n} \hat{U}^* (\hat{\beta}^*, t) \), where \( \hat{\Omega}^* (\hat{\beta}^*) \) is the estimated variance–covariance matrix for \( \hat{U}^* \) given by

\[
\hat{\Omega}^* (\hat{\beta}^*) = \frac{1}{n} \sum_{i=1}^{n} \left[ \int_{0}^{t} \left\{ Z_i - \frac{S^{(1)}(\hat{\beta}^*, u)}{S^{(0)}(\hat{\beta}^*, u)} \right\} dM_i^* (u) \right] \bigotimes^2.
\]

Graphical inspections of the proportional hazards assumption with respect to the \( k \)th component of \( \hat{U}^* \) can be performed based on \( \left\{ \left[ \hat{\Omega}^* (\hat{\beta}^*) \right]_{kk}^{-1} \right\}^{1/2} \sqrt{n} \hat{U}^* (\hat{\beta}^*, t) \), where \( \hat{U}^*_k \) is the \( k \)th component of \( \hat{U}^* \) and \( \left\{ \hat{\Omega}^* (\hat{\beta}^*) \right\}^{-1} \) is the \( k \)th diagonal element of \( \left\{ \hat{\Omega}^* (\hat{\beta}^*) \right\}^{-1} \). Moreover, the \( P \) value for the overall assessment of the proportional hazards assumption can be obtained by using the supremum test \( \sup_t |W_{ph}^*(t)| \), where

\[
W_{ph}^*(t) = \sum_{k=1}^{p} \left[ \left[ \hat{\Omega}^* (\hat{\beta}^*) \right]_{kk}^{-1} \right]^{1/2} \sqrt{n} \hat{U}^*_k (\hat{\beta}^*, t).
\]

Figure 2 shows plots of standardized pseudo-score processes when fitting the proportional hazards model, \( \lambda(t|Z) = \lambda_0(t) \exp(\beta Z) \), to data generated from alternative models where the proportionality assumption may fail to hold. Figure 2(a) shows that the observed process falls well inside the range determined by the simulated processes when the proportional hazards assumption holds. Figure 2(b) considers the covariate effect to be increasing with time, while Figure 2(c) considers the scenario where there are no covariate effects before \( t = 0.5 \) and the covariate effect follows proportional hazards assumption after \( t = 0.5 \). Both plots show evidence of nonproportional hazards for the covariate. It is interesting to see that in Figure 2(c) the observed pseudo-score process for large \( t \) falls well inside the range determined by the approximating processes, suggesting that the proportional hazards assumption may hold for later time points.

4. **Simulations**

To simulate correlated gap times for subject \( i \), we generate a subject-specific variable \( A_i \) and episode-specific variables \( B_{ij} \) independently from zero-mean normal distributions with variances \( \rho \) and \( 1 - \rho \),
Fig. 2. Plot of the standardized pseudo-score process based on fitting the $\lambda(t|Z) = \lambda_0(t) \exp(Z\beta)$ to a single data set of 100 subjects with correlated gap times generated from various models. In all scenarios, the proportion of subjects without any complete gap times is about 25%. $Z \sim $ uniform $(0, 10)$ in models (a) and (b), and $Z \sim $ Bernoulli$(0.5)$ in model (c). The bold curve corresponds to the observed process and the gray curves are 20 (out of 1000) approximating processes.

respectively, where $\rho \in [0, 1]$. Let $T_{ij}^{(0)} = -\log[1 - \Phi (A_i + B_{ij})]$, where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution. It can be verified that $T_{i1}^{(0)}, T_{i2}^{(0)}, \ldots$ are realizations of correlated exponential random variables with unit means. Thus by applying a transformation to $T_{ij}^{(0)}$, we can easily impose covariate effects and/or change the shape of the hazard function. As an example, $T_{ij} = (T_{ij}^{(0)}/a)^{1/\kappa}$ has a marginal cumulative baseline hazard function $\Lambda(t) = a \times t^\kappa$.

Two sets of simulation studies are conducted to examine the performance of the proposed supremum tests in checking the functional form of a covariate and the proportional hazards assumption. For each set of simulations, we consider combinations of different sample sizes $n = 50, 100$ and $\rho = 0, 0.25, 0.5, 0.75$. The censoring times are generated from a uniform $(0, \tau_c)$ distribution, where $\tau_c$ is chosen such that the proportion of subjects with zero events is approximately 25%. The size and power of the test statistics are estimated using 1000 simulated data sets. The significance level of the supremum tests is set at 0.05.

The first set of simulations assesses the performance on checking the functional form of a covariate $Z$ in the Cox model. We fit a proportional hazards model $\lambda(t|Z) = \lambda_0(t) \exp(Z\beta)$ to recurrent gap time data generated under various models described in Table 1. We compare the performance of the proposed method with the diagnostic test proposed by Lin and others (1993), henceforth LWY, that only uses the first gap times $\{(Y_{i1}, \Delta_i, Z_i), i = 1, \ldots, n\}$. Table 1 summarizes the estimated power and size of the proposed supremum test $sup_z |W_z^*(z)|$ and the LWY testing procedure. When the functional form of the covariate is correctly specified, the estimated size of the proposed test is close to the predetermined significance level (0.05) across various degrees of within-subject association in recurrent gap times. As expected, the proposed method always yields greater power than the LWY method. Moreover, the power of the proposed supremum test decreases when the within-subject correlation increases, despite the increase in the average number of gaps, whereas the power of LWY test remains constant. Such a phenomenon is frequently observed in correlated data where the power increases with a decrease in within-subject correlation.

The second set of simulations assesses the performance on checking the proportional hazards assumption. We generate gap times with marginal hazard functions $\dot{\lambda}(t|Z) = 2t \exp(0.3Z)$, $\lambda(t|Z) = \exp[Z(0.8 + 0.3 \log(t))]$ and $\dot{\lambda}(t|Z) = \exp[ZI(t > 0.5)]$, where the assumption of proportionality in $Z$ holds for the former and fails for the other 2 scenarios. The supremum test based on the standardized
Table 1. Size and power of the supremum test \( \sup_{z} |W^*_z(z)| \) for checking the functional form of a covariate

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \rho )</th>
<th>( \hat{m} )</th>
<th>Proposed</th>
<th>LWY</th>
<th>( \hat{m} )</th>
<th>Proposed</th>
<th>LWY</th>
<th>( \hat{m} )</th>
<th>Proposed</th>
<th>LWY</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0</td>
<td>3.01</td>
<td>0.031</td>
<td>0.049</td>
<td>2.90</td>
<td>0.430</td>
<td>0.293</td>
<td>2.97</td>
<td>0.424</td>
<td>0.294</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.25</td>
<td>3.18</td>
<td>0.045</td>
<td>0.055</td>
<td>3.04</td>
<td>0.399</td>
<td>0.314</td>
<td>3.12</td>
<td>0.383</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>3.38</td>
<td>0.063</td>
<td>0.053</td>
<td>3.23</td>
<td>0.366</td>
<td>0.292</td>
<td>3.31</td>
<td>0.343</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.75</td>
<td>3.67</td>
<td>0.051</td>
<td>0.048</td>
<td>3.53</td>
<td>0.343</td>
<td>0.302</td>
<td>3.61</td>
<td>0.327</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>3.02</td>
<td>0.041</td>
<td>0.037</td>
<td>2.90</td>
<td>0.797</td>
<td>0.644</td>
<td>2.97</td>
<td>0.848</td>
<td>0.659</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.25</td>
<td>3.18</td>
<td>0.044</td>
<td>0.043</td>
<td>3.04</td>
<td>0.780</td>
<td>0.641</td>
<td>3.12</td>
<td>0.810</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>3.38</td>
<td>0.054</td>
<td>0.040</td>
<td>3.24</td>
<td>0.752</td>
<td>0.644</td>
<td>3.32</td>
<td>0.786</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.75</td>
<td>3.69</td>
<td>0.052</td>
<td>0.046</td>
<td>3.53</td>
<td>0.705</td>
<td>0.661</td>
<td>3.65</td>
<td>0.711</td>
</tr>
</tbody>
</table>

\( \hat{m} \) is the average number of censored and uncensored gap times per subject; “Proposed” is the proposed supremum test and “LWY” is the test studied by Lin and others (1993). The proportion of subjects without any complete gap times is about 25% in all scenarios. The covariate \( Z \sim \text{uniform}(0, 10) \) in models (a-c), \( \exp(Z) \sim \text{uniform}(0, 10) \) in model (d), \( \sqrt{Z} \sim \text{uniform}(0, 10) \) in model (e), and \( \log(Z) \sim \text{uniform}(0, 5) \) in model (f). The empirical calculations were based on 1,000 replicates each.

The pseudo-score process \( \sup_{t} |W^*_b(t)| \) is applied to the simulated data and is compared with the LWY test using the first gap times. Table 2 summarizes the estimated sizes and powers of the 2 model checking methods. When the assumption of proportional hazards holds, the estimated size of both tests are close to the nominal level. As expected, the proposed test consistently yields higher power than the LWY test that only uses data on the first gap times.

### 5. Schizophrenia Data Example

We illustrate the proposed methods using data collected by the nationwide DPCR (Munk-Jørgensen and Mortensen, 1997). The DPCR has monitored contacts with psychiatric institutions throughout Denmark since 1969 and has included all contacts in emergency rooms and outpatient clinics since 1995. Each admission record includes date of admission, date of discharge, one main discharge diagnosis and up to 3 auxiliary discharge diagnoses. The data used in our analysis is from a cohort of 286 individuals with first contact with Danish psychiatric services during the period from April 1, 1970 to December 31, 1970 with a diagnosis of schizophrenia. For our cohort, 63% were male, 20% were married, and the median onset age of schizophrenia was 26 with a range from 14 to 88. We are interested in assessing the effects of onset age, gender, and marital status on the risk of rehospitalization due to a schizophrenic episode. Although previous literature (Olesen and Mortensen, 2002; Olesen and Parner, 2006) shows that the readmission
Table 2. Size and power of the supremum test based on the standard pseudo-score processes for checking the proportional hazards assumption

| $\lambda(t|Z)$ | (a) $2t \exp(0.3Z)$ | (b) $\exp[Z(0.8 + 0.3 \log(t))]$ | (c) $\exp[ZI(t > 0.5)]$ |
|---------------|-------------------|-------------------------------|-------------------|
| $n$ | $\rho$ | $\bar{m}$ | Proposed | LWY | $\bar{m}$ | Proposed | LWY | $\bar{m}$ | Proposed | LWY |
| 50 | 0 | 3.01 | 0.041 | 0.047 | 3.01 | 0.585 | 0.401 | 2.96 | 0.435 | 0.271 |
| | 0.25 | 3.18 | 0.056 | 0.056 | 3.13 | 0.505 | 0.373 | 3.31 | 0.476 | 0.286 |
| | 0.5 | 3.38 | 0.056 | 0.057 | 3.27 | 0.594 | 0.379 | 3.95 | 0.432 | 0.283 |
| | 0.75 | 3.67 | 0.050 | 0.048 | 3.51 | 0.549 | 0.409 | 5.66 | 0.416 | 0.282 |
| 100 | 0 | 3.02 | 0.044 | 0.043 | 3.03 | 0.904 | 0.727 | 2.98 | 0.709 | 0.495 |
| | 0.25 | 3.18 | 0.050 | 0.052 | 3.12 | 0.927 | 0.722 | 3.31 | 0.718 | 0.512 |
| | 0.5 | 3.38 | 0.047 | 0.053 | 3.27 | 0.898 | 0.711 | 3.98 | 0.677 | 0.516 |
| | 0.75 | 3.69 | 0.049 | 0.051 | 3.49 | 0.890 | 0.739 | 5.82 | 0.665 | 0.525 |

$m$ is the average number of censored and uncensored gap times per subject. “Proposed” is the proposed supremum test and “LWY” is the test studied by Lin and others (1993). The proportion of subjects without any complete gap times is about 25% in all scenarios. The covariate $Z \sim \text{uniform}(0, 10)$ in models (a) and (b), $Z \sim \text{Bernoulli}(0.5)$ in model (c). The empirical calculations were based on 1,000 replicates each.

Risk seems to be stable over time, to avoid potential long-term pattern changes in recurrent gap times, we set the maximum follow-up time for each person to be 3 years. Forty percent of these individuals (171 out of 286) did not experience another hospitalization during the 3 year follow-up. On average, each patient experienced 1.7 hospitalizations after the initial contact.

We apply the nonparametric trend test proposed by Wang and Chen (2000) to the recurrent gap times of the study cohort and find that there is no significant change in the distributional pattern of gap times. Univariate Cox regression using the method by Huang and Chen (2003) shows that 3 risk factors, age of schizophrenia onset ($P < 0.001$), male ($P = 0.09$), and unmarried ($P = 0.02$) are associated with a shorter gap time between repeated hospitalizations. When all 3 variables are included in the Cox model, the effects of onset age remains highly significant ($P < 0.001$), while the effects of gender ($P = 0.85$) and marital status ($P = 0.47$) are insignificant. The multivariate Cox model estimates that a person who has onset at 20 has almost a 25% higher instantaneous risk of rehospitalization than a person who has an onset at 30, adjusting for gender and marital status. Further analysis reveals that both gender and marital status were highly correlated with onset age. The median onset age was 24 and 34 for males and females, and was 23 and 38 for unmarried and married patients of the studied cohort. Hence, in the remaining analysis, we will only focus on the effect of onset age on the distribution of recurrent gap times.

To check the assumption of proportional hazards for the candidate model with age at onset, denoted by $Z$, as the only covariate, we plot the standardized pseudo-score process $W^*_ph(t)$ from the fitted model. Figure 3(a) shows the observed standardized pseudo-score process versus follow-up time since last hospitalization along with approximating processes simulated from the null hypothesis of a correct model fit. The observed process seems to be typical of the null hypothesis simulations, and the proportional hazards assumption fails to be rejected with an estimated $P$ value 0.29 that is based on 1000 simulated processes.

Next, we plot the partial-mean process $W^*_z(z)$ and apply the supremum test to check the linear functional form of $Z$. Figure 3(d) displays the observed partial-mean process along with 20 approximating processes generated from the null distribution. In addition to the linear Cox model, we also consider 2 different functional forms, $\sqrt{Z}$ and $\log(Z)$, of $Z$. The corresponding plots are given by Figures 3(e) and 3(f). The estimated $P$ values for the 3 models are 0.42, 0.71, and 0.77, respectively. These results suggest that the linear form of onset age, although reasonable, may not fit as well as the proportional hazards models with either $\sqrt{Z}$ or $\log(Z)$ as the covariate.
Model checking for recurrent gap time data

Fig. 3. Diagnostic plots for the DPCR data. Upper panel: plots of the standardized pseudo-score process $W_{ph}^*(t)$ for checking the proportional hazards assumption. The covariates in the fitted models are (a) $Z$, (b) $\sqrt{Z}$, and (c) log($Z$). Lower panel: plots of the partial-mean process $W^*_z(z)$ for checking the functional forms (d) $Z$, (e) $\sqrt{Z}$, and (f) log($Z$). The bold curve corresponds to the observed process and the gray curves are 20 (out of 1000) approximating processes.

Plots of standardized pseudo-score process for the other 2 candidate models are also obtained to examine the assumption of proportional hazards. As shown in Figures 3(b) and (c), the estimated $p$ values for the models with $\sqrt{Z}$ and log($Z$) as the only covariate are 0.50 and 0.75, respectively. Based on these results, the logarithm transformation of the schizophrenia onset age seems to be more appropriate than the other functional forms considered in this section. The pseudo-partial likelihood method estimates the coefficient of log($Z$) to be $-1.04$ (SE 0.21) in the Cox model.

6. DISCUSSION

The proposed methodology can be employed to develop other model assessment techniques for the Cox model. For example, the link function of the proportional hazards model can be checked by plotting $W_i^*(z) = n^{-1} \sum_{i=1}^n I(Z_i^T \hat{\beta}^* < z) \hat{M}_i^*(t)$ against the set of unique values of $Z_i^T \hat{\beta}^*$, $i = 1, \ldots, n$. A global test can be constructed by considering $W^*_g(z, t) = n^{-1} \sum_{i=1}^n I(Z_{1i} \leq z_1, \ldots, Z_{pi} \leq z_p) \hat{M}_i^*(t)$ for all possible values of $z = (z_1, \ldots, z_p)$ and $t$. Arguing as in Section 3.1, the distributions of $W^*_i(z)$ and $W^*_g(z, t)$ also can be approximated through simulating the corresponding zero-mean Gaussian process. Finally, a smoothed plot of $\hat{M}_i^*(\tau)$ versus a covariate omitted from the fitted model provides approximately the correct functional form of that covariate.
In this article, we introduced 2 stochastic processes, the averaged counting process $N^*_i(t)$ and the averaged at risk process $R^*_i(t)$, to formulate the averaged martingale-like process $M^*_i(t)$. Different graphical techniques and formal tests were proposed by grouping the estimated process $\hat{M}^*_i(t)$, that is the averaged residual process, cumulatively with respect to follow-up time and/or covariate values. The proposed methodology is very flexible. In fact, by properly modifying the definition of $N^*_i(t)$ and $R^*_i(t)$, the proposed methods can be generalized to check non-Cox models. As an example, consider the accelerated failure time (AFT) model $\log(T_{ij}) = Z_i^T \beta + \epsilon_{ij}$, where $\epsilon_{ij}$'s have a common distribution. We can define $N^*_i(t) = \{m^*_i\}^{-1} \sum_{j=1}^{m^*_i} \Delta_i I\{Y_{ij} \exp(-Z_i^T \beta) \leq t\}$ and $R^*_i(t) = \{m^*_i\}^{-1} \sum_{j=1}^{m^*_i} I\{Y_{ij} \exp(-Z_i^T \beta) \geq t\}$. Model checking methods for the AFT model can be developed based on the corresponding averaged martingale-like process. Plots of the multiparameter stochastic processes with simulated trajectories under the AFT model will help assess model fit. If the AFT model does not fit well, we can repeat this exercise for the Cox model and a linear transformation model. At the end, we can select the model that has the best-fitting residual process.

A limitation of the proposed methodology is its requirement that the recurrent gap times are conditional iid. When the conditionally iid assumption fails to hold, the exchangeability among uncensored gap times may not hold and thus the proposed method may not be appropriate. In practice, before fitting the proposed model, one should check whether there is a trend in the distributional pattern of gap times by applying trend tests such as the ones proposed by Wang and Chen (2000). Moreover, one may compare the results using the proposed model checking methods with that using the methods considered by Lin and others (1993) for the first gap times. If the 2 approaches give different conclusions, it is likely that the conditional iid assumption is violated and one should consider other statistical models that better capture the difference in episode-specific distributional patterns.

**SUPPLEMENTARY MATERIAL**

Supplementary material is available at http://biostatistics.oxfordjournals.org.

**ACKNOWLEDGMENTS**

The authors thank Drs Preben Bo Mortensen and William Eaton for generously providing anonymous Denmark schizophrenia data for illustrating the proposed methods. *Conflict of Interest:* None declared.

**REFERENCES**


*[Received July 8, 2010; revised October 22, 2010; accepted for publication November 1, 2010]*