Model checking for additive hazards model with multivariate survival data

Guosheng Yin*

Department of Biostatistics and Applied Mathematics, M. D. Anderson Cancer Center, The University of Texas, Houston, TX 77030, USA

Received 7 August 2005
Available online 18 April 2006

Abstract

Multivariate failure time data often arise in biomedical studies due to natural or artificial clustering. With appropriate adjustment for the underlying correlation, the marginal additive hazards model characterizes the hazard difference via a linear link function between the hazard and covariates. We propose a class of graphical and numerical methods to assess the overall fitting adequacy of the marginal additive hazards model. The test statistics are based on the supremum of the stochastic processes derived from the cumulative sum of the martingale-based residuals over time and/or covariates. The distribution of the stochastic process can be approximated through a simulation technique. The proposed tests examine how unusual the observed stochastic process is, compared to a large number of realizations from the approximated process. This class of tests is very general and suitable for various purposes of model fitting evaluation. Simulation studies are conducted to examine the finite sample performance, and the model-checking methods are illustrated with data from an otitis media study.

© 2006 Elsevier Inc. All rights reserved.

AMS 2000 subject classification: Primary 62N03; secondary 62H15

Keywords: Cumulative sum; Marginal model; Martingale residual; Multivariate failure time data; Parallel hazards; Score process

1. Introduction

As opposed to the Cox proportional hazards model [8], the additive hazards model [21] relates the covariate vector $Z_i$ of subject $i$ and the hazard function $\lambda(t|Z_i)$ in a linear form

$$\lambda(t|Z_i) = \lambda_0(t) + \beta'Z_i,$$

(1.1)
where $\lambda_0(t)$ is the unknown and unspecified baseline hazard function and $\beta$ is the regression coefficient vector. As an alternative way to model the hazard function, the additive hazards model yields risk differences instead of risk ratios; this has been studied in various contexts [4,5,19,26,34,16]. The estimate of $\beta$ under model (1.1) has an analytic closed form and its interpretation is intuitively meaningful. Other additive risk models have been explored in the forms of non-parametric or partly parametric structures [1,2,13,22,23,25]. A variety of model-checking and goodness-of-fit tests have been proposed for the additive hazards model [35,14,15,11,10].

A fundamental assumption common to all the aforementioned methods is the independence of the failure times. However, in many biomedical studies, this independence assumption might not hold. For example, during follow-up, each patient may experience multiple distinct events or recurrent events, or the same type of disease may affect clustered organs of the same subject. One interesting example is an otitis media study of young children between 6 months and 8 years of age [17]. The anatomy of the ear includes a small tunnel, connecting the middle ear to the nasopharynx. This tunnel, which is shorter and more horizontal in children than in adults, permits the equalization of air pressure between the middle ear and the outside of the eardrum. However, it can also allow the entrance of bacteria into the middle ear, which may cause an infection of the middle ear (otitis media). Consequently, the tunnel may fill with fluid and pus, which then results in temporary periods of hearing loss. This is particularly worrisome in young children since it may delay behavioral and language development. A popular treatment is to insert ventilating tubes into the infected ears, thereby improving a child’s hearing as long as the tubes are in place and functioning. Tube failure is defined as tube blockage or extrusion from the ear. The objective of the otitis media study was to examine the effectiveness of a combination therapy of prednisone and sulfamethoprim in prolonging the lifetime of the ventilating tubes. The assignment of the treatment was based on an individual level, with both or neither of the ears receiving the treatment for each subject. The paired times to tube failure observed from each child were clearly not independent. Fig. 1 shows the Kaplan–Meier survival curves for the treatment and control arms by combining the paired observations in each arm [33].
Numerous methods have been proposed for multivariate failure time data, among which marginal and frailty models are the two main directions of investigation. Wei et al. [31], Lee et al. [18], Spiekerman and Lin [28] and Clegg et al. [7] among others extended the Cox proportional hazards model to multivariate cases based on the quasi-likelihood or estimating equations. More recently, estimation and asymptotic properties for the marginal additive hazards model have been studied with multivariate failure time data [32]. Various model-checking methods and research have been carried out under the proportional hazards assumption [29,20,27,12,9]. In this paper, we propose a class of model-checking and goodness-of-fit tests for the marginal additive hazards model with correlated survival data. Since the proposed stochastic processes fluctuate randomly around the zero axis under the null hypothesis, the tests are constructed from the maximum deviation of the processes from zero. This class of tests includes the evaluation of the parallel hazards assumption, the covariate functional form and the link function for detecting different aspects of model misspecification.

The rest of the article is organized as follows. In Section 2, we present the basic setup and review the development of the marginal additive hazards model for multivariate survival data. In Section 3, we propose a multiparameter stochastic process and study its weak convergence property. We also introduce a simulation method to approximate the distribution of the process, and propose a series of graphical and numerical model-checking techniques. In Section 4, we conduct simulation studies to investigate the finite sample properties of the test statistics with respect to test sizes and statistical powers. In Section 5, we illustrate the proposed methods with data from the otitis media study, and provide concluding remarks. We briefly outline the technical proofs of the theorems in Appendix.

2. Additive hazards model

Suppose that there are \( n \)-independent clusters and each cluster has \( K \) exchangeable subjects. Let \( T_{ik} \) \((k = 1, \ldots, K; i = 1, \ldots, n)\) be the failure time for subject \( k \) in cluster \( i \), \( C_{ik} \) be the corresponding censoring time, and \( Z_{ik} \) be the associated \( p \times 1 \) bounded covariate vector. We observe \( X_{ik} = \min(T_{ik}, C_{ik}) \) and the censoring indicator \( \Delta_{ik} = I(T_{ik} \leq C_{ik}) \), where \( I(\cdot) \) is the indicator function. We allow the cluster size, \( K \), to vary by setting \( C_{ik} = 0 \) whenever \( T_{ik} \) is missing. Let \( T_i = (T_{i1}, \ldots, T_{iK})' \), with \( C_i \) and \( Z_i \) defined similarly. We assume that \( \{(T_i, C_i, Z_i); i = 1, \ldots, n\} \) are independent and identically distributed (i.i.d.), and that \( T_i \) and \( C_i \) are conditionally independent given \( Z_i \).

Motivated by the work of Spiekerman and Lin [27] for the marginal Cox-type regression model, we focus on the derivation of the model-checking and goodness-of-fit tests for

\[
\lambda(t|Z_{ik}) = \lambda_0(t) + \beta'Z_{ik}. \tag{2.1}
\]

We define the counting process \( N_{ik}(t) = I(X_{ik} \leq t, \Delta_{ik} = 1) \), and the at-risk process \( Y_{ik}(t) = I(X_{ik} > t) \).

If we denote \( \hat{\beta} \) as the consistent estimator of \( \beta \), then the baseline cumulative hazard function \( \Lambda_0(t) = \int_0^t \lambda_0(u) \, du \) can be estimated by

\[
\hat{\Lambda}_0(t) = \sum_{i=1}^n \sum_{k=1}^K \int_0^t \frac{dN_{ik}(u) - Y_{ik}(u) \hat{\beta}'Z_{ik}}{\sum_{i=1}^n \sum_{k=1}^K Y_{ik}(u)} \, du.
\]
Observing that a martingale integral has a mean zero, we define

$$U(\beta, \tau) = \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{\tau} (Z_{ik} - \bar{Z}(t)) dM_{ik}(t),$$  

(2.2)

where $\tau$ is the end time of a study, $M_{ik}(t) = N_{ik}(t) - \int_{0}^{t} Y_{ik}(u) d\Lambda_0(u) + \beta' Z_{ik} du$, and

$$\bar{Z}(t) = \sum_{i=1}^{n} \sum_{k=1}^{K} \frac{Y_{ik}(t) Z_{ik}}{\sum_{i=1}^{n} \sum_{k=1}^{K} Y_{ik}(t)}.$$

Replacing $M_{ik}(t)$ in (2.2) by its empirical counterpart, we obtain an analytic closed form of

$$\hat{U}(\beta, \tau) = \left[ \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{\tau} Y_{ik}(t) (Z_{ik} - \bar{Z}(t)) d\hat{M}_{ik}(t) \right]^{-1} \left[ \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{\tau} (Z_{ik} - \bar{Z}(t)) dN_{ik}(t) \right],$$

(2.3)

where $a^{\otimes 2} = aa'$. Using the empirical process theories, $U(\beta, \tau)$ is shown to be a sum of i.i.d.

random vectors, and thus follows a zero-mean normal distribution by the multivariate central limit theorem (CLT). By Taylor’s series expansion and some probability arguments, $n^{1/2} (\hat{U} - U)$ converges in distribution to a zero-mean $p$-variate normal random vector [32].

3. Model-checking techniques

Let $A(t) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{t} Y_{ik}(u) (Z_{ik} - \bar{Z}(u))^{\otimes 2} du$. We assume that $A(t)$ uniformly converges in probability to a non-singular deterministic matrix, $\tilde{A}(t)$, and $\bar{Z}(t)$ uniformly converges to $\bar{Z}(t)$ for $t \in [0, \tau]$.

Under model (2.1), we define the martingale residual as

$$\hat{M}_{ik}(t) = N_{ik}(t) - \int_{0}^{t} Y_{ik}(u) d\hat{\Lambda}_0(u) + \hat{\beta}' Z_{ik} du$$

which can be viewed as the difference at time $t$ between the observed and expected number of failures for subject $k$ in cluster $i$, as in the ordinary linear regression. Conventionally, by plotting $\hat{M}_{ik}(t)$ with respect to the observed times, we may reveal any misspecification of the model to some extent [29]. However, this cannot be used as an objective diagnostic test. We study the multiparameter stochastic process, involving various forms of cumulative sums of $\hat{M}_{ik}(t)$, defined as

$$W(t, z) = \sum_{i=1}^{n} \sum_{k=1}^{K} f(Z_{ik}) I(Z_{ik} \leq z) \hat{M}_{ik}(t),$$

(3.1)

where $f(\cdot)$ is a known vector-valued bounded function, and $I(Z_{ik} \leq z) = I(Z_{ik1} \leq z_1, \ldots, Z_{ikp} \leq z_p)$.

We define

$$g(t, z) = \frac{\sum_{i=1}^{n} \sum_{k=1}^{K} f(Z_{ik}) I(Z_{ik} \leq z) Y_{ik}(t)}{\sum_{i=1}^{n} \sum_{k=1}^{K} Y_{ik}(t)}$$
and
\[ h(t, z) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{t} f(Z_{ik}) I(Z_{ik} \leq z) Y_{ik}(u)(Z_{ik} - \bar{Z}(u))' \, du, \]

and denote \( \tilde{g}(t, z) \) and \( \tilde{h}(t, z) \) as the limits of \( g(t, z) \) and \( h(t, z) \). By Taylor’s series expansions of \( W(t, z) \) and \( U(\hat{\beta}, \tau) \) around \( \hat{\beta} \), and through some probability arguments, we can show that \( n^{-1/2} W(t, z) \) is asymptotically equivalent to \( n^{-1/2} \hat{W}(t, z) \), where
\[ \hat{W}(t, z) = \sum_{i=1}^{n} Q_i(t, z) \]

and
\[ Q_i(t, z) = \sum_{k=1}^{K} \int_{0}^{t} \{ f(Z_{ik}) I(Z_{ik} \leq z) - \tilde{g}(u, z) \} \, dM_{ik}(u) \]
\[ -\tilde{h}(t, z) A^{-1}(\tau) \sum_{k=1}^{K} \int_{0}^{\tau} \{ Z_{ik} - \tilde{Z}(t) \} \, d\hat{M}_{ik}(t). \]

The stochastic process \( W(t, z) \) accommodates several specific tests for different aspects of model misspecification and its asymptotic property is presented in the following theorem.

**Theorem 1.** Under the regularity conditions given in Appendix, \( n^{-1/2} W(t, z) \) converges weakly to a zero-mean Gaussian random field with the covariance function between \( (t, z) \) and \( (t^*, z^*) \) given by \( E\{Q_1(t, z)Q_1'(t^*, z^*)\} \).

The key steps in the proof are to verify the finite-dimensional distribution convergence and the tightness condition, as outlined in Appendix. The covariance function can be consistently estimated by \( n^{-1} \sum_{i=1}^{n} \hat{Q}_i(t, z)\hat{Q}_i'(t^*, z^*) \), where
\[ \hat{Q}_i(t, z) = \sum_{k=1}^{K} \int_{0}^{t} \{ f(Z_{ik}) I(Z_{ik} \leq z) - g(u, z) \} \, d\hat{M}_{ik}(u) \]
\[ -h(t, z) A^{-1}(\tau) \sum_{k=1}^{K} \int_{0}^{\tau} \{ Z_{ik} - \hat{Z}(t) \} \, d\hat{M}_{ik}(t). \]

We can approximate the limiting distribution of \( n^{-1/2} W(t, z) \) through a Monte Carlo simulation technique. By independently generating a simple random sample \( (\xi_1, \ldots, \xi_n) \) from the standard normal distribution, \( N(0, 1) \), we obtain the perturbed version of the stochastic process
\[ \hat{W}(t, z) = \sum_{i=1}^{n} \hat{Q}_i(t, z) \xi_i. \quad (3.2) \]

The next theorem provides the theoretical justification for this perturbing procedure.

**Theorem 2.** Given the observed data \( \{(N_{ik}(t), Y_{ik}(t), Z_{ik}); \ t \in [0, \tau]; \ i = 1, \ldots, n; \ k = 1, \ldots, K\} \), \( n^{-1/2} \hat{W}(t, z) \) converges weakly to the same zero-mean Gaussian random field as that of \( n^{-1/2} W(t, z) \).
Conditional on the observed data, the only random variables in (3.2) are the \( \xi_i \)'s, and thus \( \hat{W}(t, z) \) can be viewed as a sum of the independent random variables for each fixed time \( t \) and covariate \( z \). This is one of the critical arguments of the proof in Appendix.

We illustrate how \( \hat{W}(t, z) \) can be utilized for different purposes of model-fitting evaluation in the following derivations. To check the additive or parallel hazards assumption under model (2.1), we consider the score-type process

\[
\hat{U}(\hat{\beta}, t) = \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{t} \{ \mathbf{Z}_{ik} - \mathbf{\bar{Z}}(u) \} \{ dN_{ik}(u) - Y_{ik}(u) \hat{\beta}' \mathbf{Z}_{ik} \, du \}.
\]

Clearly, \( \hat{U}(\hat{\beta}, t) \) is a special case of \( W(t, z) \) with \( f(\mathbf{Z}_{ik}) = \mathbf{Z}_{ik} \) and \( z = \infty \).

By Taylor’s series expansion

\[
n^{-1/2} \hat{U}(\hat{\beta}, t) = n^{-1/2} U(\beta, t) - n^{-1/2} A(t)(\hat{\beta} - \beta) + o_{p}(1)
\]

\[
= n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \left[ \int_{0}^{t} \{ \mathbf{Z}_{ik} - \mathbf{\bar{Z}}(u) \} \, dM_{ik}(u) \right.
\]

\[
- \mathbf{A}(t)\mathbf{A}^{-1}(\tau) \int_{0}^{\tau} \{ \mathbf{Z}_{ik} - \mathbf{\bar{Z}}(u) \} \, dM_{ik}(u) \left] + o_{p}(1). \right.
\]

A consistent estimator of the covariance matrix of \( n^{-1/2} \hat{U}(\beta, \tau) \) is \( \hat{\Sigma} = n^{-1} \sum_{i=1}^{n} \hat{\Phi}_{i}(\hat{\beta}, \tau) \hat{\Phi}'_{i}(\hat{\beta}, \tau) \), where

\[
\hat{\Phi}_{i}(\hat{\beta}, t) = \sum_{k=1}^{K} \int_{0}^{t} \{ \mathbf{Z}_{ik} - \mathbf{\bar{Z}}(u) \} \, d\hat{M}_{ik}(u). \tag{3.3}
\]

Then, we propose the test statistic for checking the additive hazards structure of the \( j \)th covariate \((j = 1, \ldots, p)\) as

\[
S_{j} = \sup_{t \in [0, \tau]} \{ \hat{\Sigma}^{-1} \}_{jj}^{1/2} n^{-1/2} U_{j}(\hat{\beta}, t), \tag{3.4}
\]

where \( U_{j}(\hat{\beta}, t) \) denotes the \( j \)th component of \( \hat{U}(\hat{\beta}, t) \) and \( \{ \hat{\Sigma}^{-1} \}_{jj} \) denotes the \( jj \)th diagonal element of \( \hat{\Sigma}^{-1} \). The \( p \)-value = \( \Pr(\hat{S}_{j} > s_{j}) \) for the goodness-of-fit test can be approximated by \( \Pr(\hat{S}_{j} > s_{j}) \), where \( s_{j} \) is the observed value of \( S_{j} \),

\[
\hat{S}_{j} = \sup_{t \in [0, \tau]} \{ \hat{\Sigma}^{-1} \}_{jj}^{1/2} n^{-1/2} \hat{U}_{j}(\hat{\beta}, t),
\]

and \( \hat{U}_{j}(\hat{\beta}, t) \) is the \( j \)th component of the perturbed score process \( \hat{U}(\hat{\beta}, t) \). The \( p \)-value can be empirically estimated by the percentage of \( \hat{S}_{j} > s_{j} \) through generating many realizations of \( \hat{S}_{j} \).

The overall test statistic for the joint additivity of all the \( p \) covariates is given by

\[
S_{a} = \sup_{t \in [0, \tau]} \sum_{j=1}^{p} \{ \hat{\Sigma}^{-1} \}_{jj}^{1/2} n^{-1/2} U_{j}(\hat{\beta}, t), \tag{3.5}
\]
In order to check the functional form of a covariate, e.g., the jth component \( Z_j \), we take \( f(Z_{ik}) = 1, t = \tau \) and \( z_l = \infty \) for all \( l \neq j (l = 1, \ldots, p) \), then

\[
W_j(\tau, z) = \sum_{i=1}^{n} \sum_{k=1}^{K} I(Z_{ikj} \leq z) \hat{M}_{ik}(\tau).
\]

The null distribution of \( W_j(\tau, z) \) can be approximated by simulating the corresponding zero-mean Gaussian process. One can thus obtain a \( p \)-value for the supremum test \( \sup_z |W_j(\tau, z)| \) by generating a large number of realizations of \( \hat{W}_j(\tau, z) \), where \( \hat{W}_j(\tau, z) \) is the jth component of (3.2). The test based on \( U_j(\hat{\beta}, t) \) is consistent against a non-additive hazards structure for \( Z_j \), and that based on \( W_j(\tau, z) \) is consistent against an incorrectly specified functional form of \( Z_j \), under the conditions that \( Z_j \) is independent of other covariates and that no other type of model misspecification exists. For the case with more than one covariate (\( p > 1 \)), the examination of the link function can be accomplished by \( \sup_z |W(\tau, z)| \). It is consistent against the alternative model that \( \lambda(t|Z_{ik}) = \lambda_0(t) + \phi(Z_{ik}) \), under which there does not exist a \( \beta \) such that \( \phi(Z_{ik}) - \beta'Z_{ik} \) is constant over all the possible values of \( Z_{ik} \).

The consistency of each model-checking test statistic can be established using arguments similar to those of Lin et al. [20]. In Appendix, we briefly outline the proof for the consistency of the test for the additive hazards structure based on the score process.

4. Simulation studies

We conducted simulation studies to investigate the properties of our proposed method with finite sample sizes. We simulated the failure times under the Clayton and Cuzick [6] model with the baseline hazard function from a Weibull distribution, e.g., \( \lambda_0(t) = t \). The cluster size was two, i.e., \( K = 2 \). Under the null model of, \( H_0 : \lambda(t|Z_{ik}) = t + \beta Z_{ik} \), the joint survival function for the bivariate failure times \( (T_{i1}, T_{i2}) \) in cluster \( i \) was

\[
Pr(T_{i1} > t_{i1}, T_{i2} > t_{i2}|Z_{i1}, Z_{i2}) = \left\{ \sum_{k=1}^{2} \exp \left( \frac{t_{ik}^2}{2\theta} + \frac{\beta Z_{ik}t_{ik}}{\theta} \right) - 1 \right\}^{-\theta},
\]

where \( \theta > 0 \), and a smaller \( \theta \) would induce a larger correlation. The correlation parameter \( \theta \) was preset for achieving the within-cluster failure time correlation \( \rho = .2 \) and .5, respectively. The censoring times were generated independently from uniform distributions to achieve approximately 30% and 60% censoring percentages. Cases without any censoring were also simulated. We examined the situation with one binary covariate which took a value of 0 or 1 with probability .5. Two different experimental designs were carried out: one corresponded to a pair-matched study \( (Z_{i1} = 1 - Z_{i2}) \), with \( \beta = .5 \); and the other was a common clusterwise-covariate study \( (Z_{i1} = Z_{i2}) \), with \( \beta = .2 \). The number of clusters was \( n = 100 \) and 150, and we performed 1000 simulations for each configuration.

For each data realization, we obtained the pointwise estimate of the regression coefficient and the sandwich-type variance estimator. For the coefficient estimators, we calculated the sample standard deviations (SD), the average of the estimated standard errors (SE) and the coverage rates (CR) of the 95% confidence intervals. Table 1 summarizes the results under the null hypothesis for our evaluation of the test size. Apparently, under each scenario, the test size is close to the nominal significance level of .05, indicating that the proposed test preserves the type I error rate.
Table 1
Test sizes for the marginal additive hazards model under two designs based on the score process

<table>
<thead>
<tr>
<th>n</th>
<th>ρ</th>
<th>c%</th>
<th>Matched pairs (Z1 = 1 – Z2)</th>
<th></th>
<th>Common covariates (Z1 = Z2)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bias</td>
<td>SD</td>
<td>SE</td>
<td>CR</td>
</tr>
<tr>
<td>100</td>
<td>.2</td>
<td>0</td>
<td>.003</td>
<td>.111</td>
<td>.107</td>
<td>93.1</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.004</td>
<td>.124</td>
<td>.125</td>
<td>94.4</td>
<td>.048</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>.003</td>
<td>.142</td>
<td>.140</td>
<td>95.1</td>
<td>.060</td>
</tr>
<tr>
<td>.5</td>
<td>0</td>
<td>.013</td>
<td>.096</td>
<td>.089</td>
<td>93.9</td>
<td>.067</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.001</td>
<td>.114</td>
<td>.113</td>
<td>94.2</td>
<td>.064</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>.002</td>
<td>.135</td>
<td>.132</td>
<td>95.2</td>
<td>.061</td>
</tr>
<tr>
<td>150</td>
<td>.2</td>
<td>0</td>
<td>.007</td>
<td>.087</td>
<td>.088</td>
<td>94.5</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.006</td>
<td>.103</td>
<td>.102</td>
<td>95.3</td>
<td>.056</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>.005</td>
<td>.118</td>
<td>.115</td>
<td>94.1</td>
<td>.064</td>
</tr>
<tr>
<td>.5</td>
<td>0</td>
<td>.009</td>
<td>.075</td>
<td>.072</td>
<td>95.5</td>
<td>.050</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.005</td>
<td>.093</td>
<td>.092</td>
<td>94.6</td>
<td>.051</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>.006</td>
<td>.098</td>
<td>.100</td>
<td>93.8</td>
<td>.056</td>
</tr>
</tbody>
</table>

SD is the standard deviation, SE is the average of estimated standard errors, and CR is the 95% confidence interval coverage rate.

Table 2
Statistical powers under two alternative models based on the score process

<table>
<thead>
<tr>
<th>n</th>
<th>c%</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>H₀₁</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ρ = .2</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>.614</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.480</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>.296</td>
</tr>
<tr>
<td>150</td>
<td>0</td>
<td>.762</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.646</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>.416</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n</th>
<th>c%</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>H₀₁</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ρ = .5</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>.630</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.508</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>.314</td>
</tr>
<tr>
<td>150</td>
<td>0</td>
<td>.766</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.632</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>.408</td>
</tr>
</tbody>
</table>

Alternative models, H₀₁ : \( \hat{\lambda}(t|Z_{ik}) = t \exp(Z_{ik}) \) and H₀₂ : \( \hat{\lambda}(t|Z_{ik}) = 1 + 2Z_{ik}t \).

To examine the statistical power of the test, we generated data from two different alternative models. One was based on the Cox proportional hazards model with the marginal hazard given by, H₀₁ : \( \hat{\lambda}(t|Z_{ik}) = \lambda_0(t) \exp(\beta Z_{ik}) \), with \( \lambda_0(t) = t \) and \( \beta = 1 \); and the other was H₀₂ : \( \hat{\lambda}(t|Z_{ik}) = \lambda_0(t) + (\beta Z_{ik})t \), with \( \lambda_0(t) = 1 \) and \( \beta = 2 \). We generated a continuous covariate from Uniform(0, 1). The power was evaluated under different scenarios by varying the underlying correlation, censoring rates, and sample sizes. As shown in Table 2, when the modeling structure is deviated from the truth (parallel hazards), the proposed test has adequate power to reject the null hypothesis. The power decreases as the censoring rate increases, and increases as the sample size increases, while the correlation does not have much effect on the power.
To examine the functional form of the covariate, we generated the paired failure times from the Clayton–Cuzick model with the marginal hazards given by \[
\lambda(t|Z_{ik}) = \lambda_0(t) + \beta Z_{ik} + \gamma Z_{ik}^2 \quad (k = 1, 2),
\] where \(Z_{ik}\) was simulated from Uniform(0, 2). Under the null hypothesis with \(\gamma = 0\), we took \(\lambda_0(t) = 1\), \(\beta = .5\) and set the correlation between the paired failure times to be .4. The censoring times were generated from a uniform distribution to yield a censoring rate of 30%. Then, we obtained the test size of .061 for \(n = 100\) and .048 for \(n = 150\) at the .05 significance level. Clearly, the type I error rate is well preserved using the proposed test statistic. Under the alternative model with \(\gamma = 2\), we took \(\lambda_0(t) = .5\) and \(\beta = 1\). The underlying correlation was .3 and the censoring rate was set at 10%. The power of the proposed test was .624 for \(n = 100\) and .808 for \(n = 150\). The results demonstrate that the test based on the cumulative sum of martingale-based residuals over a covariate has adequate power against the misspecification of the covariate functional form. Fig. 2 presents graphical views of the observed partial sum of the...
martingale-based residuals versus the first 50 simulated ones under the null and alternative models for \( n = 100 \). When the model is incorrect, the curve of the observed cumulative sum of martingale residuals cannot be completely covered by the simulated processes.

5. Example

As an illustration, we applied the proposed model-checking methods to the data from the otitis media study. The objective of the study was to investigate the effectiveness of the medical treatment in prolonging the lifetime of the ventilating tubes in children’s ears. In this analysis, there were 78 children of 6 months to 8 years of age who had chronic otitis media with effusion (after deleting two ineligible observations from the original data set). By randomization, 40 children were assigned to the medical treatment and 38 served as controls. Approximately, 8% of the data were censored by loss to follow-up.

The treatment estimate under model (2.1) is \( \hat{\beta} = -0.0384 \), with the estimated standard error \( \text{SE}(\hat{\beta}) = 0.0204 \). This yielded a \( p \)-value of 0.060, indicating that the treatment did not significantly prolong the lifetimes of the tubes. Since there was only one dichotomous covariate in the data set, we focused on the examination of the parallel hazards assumption using the score process. Based on 1000 simulated score processes, the proposed model adequacy test had a \( p \)-value of 0.503, which supported the appropriateness of the additive hazards model to fit the data. Fig. 3 shows that the observed score process can be completely covered by the first 50 simulated score processes, which graphically demonstrates that the observed score process is not that unusual compared to the simulated ones.

We assumed that the covariates were time independent throughout the development of the model-checking methods under the additive hazards model. Incorporating time-dependent covariates would allow the additive hazards model to be more flexible. In this case, the score
process would take the form of

$$U(\hat{\beta}, t) = \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{t} [Z_{ik}(u) - \bar{Z}(u)] \{dN_{ik}(u) - Y_{ik}(u) \hat{\beta}' Z_{ik}(u) \, du\},$$

which could be used to check the parallel hazards assumption. Although the theoretical derivation would still hold, it might be difficult to plot the partial sum process against a time-varying covariate.

We have investigated a class of residual-based model-checking procedures for the marginal additive hazards model with multivariate failure time data. On the basis of simulating the proposed stochastic process, we can construct the maximal deviation test in the same spirit as the well-known Kolmogorov–Smirnov statistic. Within the additive hazards modeling framework, we developed several martingale residual-based test statistics, which are sensitive to any departures from the additive structure in the hazards, the linear covariate forms or the identity link function. We examined the test size under the null model, and the power of the test under several different alternative models in the simulation study. For sample sizes of practical use, the proposed tests preserve the type I error rate and possess adequate power to detect model misspecification.

Acknowledgments

We would like to thank the associate editor and two anonymous referees for their constructive suggestions which led to substantial improvement of the article.

Appendix A.

Proof of Theorem 1. We assume the following regularity conditions: there exists a constant, \( \tau > 0 \), such that, \( \Pr\{Y_{ik}(t) = 1, t \in [0, \tau]\} > 0 \); \( \int_{0}^{\tau} \lambda_{0}(t) \, dt < \infty \); the covariate vector \( Z_{ik} \) is bounded; and \( \tilde{A}(t) \) is positive definite.

By the functional CLT [24, Theorem 10.6], we can show that

$$n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{t} [\tilde{Z}(u) - \bar{Z}(u)] \, dM_{ik}(u) \rightarrow 0,$$

and

$$n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{t} [\bar{g}(u, z) - \bar{\tilde{g}}(u, z)] \, dM_{ik}(u) \rightarrow 0,$$

in probability. Note that \( A(t) \) uniformly converges in probability to \( \tilde{A}(t) \) and \( h(t, z) \) to \( \tilde{h}(t, z) \). If we take Taylor’s series expansions of \( W(t, z) \) and \( U(\hat{\beta}, \tau) \) around \( \beta \), and apply (A.1), \( n^{-1/2} W(t, z) \) is asymptotically equivalent to \( n^{-1/2} \tilde{W}(t, z) = n^{-1/2} \sum_{i=1}^{n} Q_{i}(t, z) \). For any fixed \( t \) and \( z \), \( \{Q_{i}(t, z), i = 1, \ldots, n\} \) are i.i.d. random vectors. It follows from the multivariate CLT that the finite-dimensional distribution of \( n^{-1/2} W(t, z) \) converges to a zero-mean normal distribution.

Without loss of generality, the covariates are assumed to be bounded in \([-1, 1]\). Following the argument by Spiekerman and Lin [27], we next show the tightness condition for \( W(t, z) \).
in \( \mathcal{D}([0, \tau] \times [-1, 1]^p) \). We rewrite
\[
n^{-1/2} W(t, z) = n^{-1/2} W^{(1)}(t, z) - n^{-1/2} W^{(2)}(t, z) - n^{-1/2} W^{(3)}(t, z) + o_p(1),
\]
where \( W^{(1)}(t, z) = \sum_{i=1}^n \sum_{k=1}^K \int_0^t g(u, z) dM_{ik}(u) \), \( W^{(2)}(t, z) = \sum_{i=1}^n \sum_{k=1}^K \int_0^t g(u, z) dM_{ik}(u) \), and \( W^{(3)}(t, z) = h(t, z) A(t) n (\hat{\beta} - \beta) \). By Lemma 1 of Lin et al. [20], \( n^{-1/2} W^{(1)}(t, z) \) is tight. It follows from the weak convergence of \( n^{-1/2} \sum_{i=1}^n \sum_{k=1}^K M_{ik}(t) \) and (A.1), that \( n^{-1/2} W^{(2)}(t, z) \) converges weakly to a zero-mean Gaussian random field. Thus, \( n^{-1/2} W^{(2)}(t, z) \) is tight by Theorem 10.2 of Pollard [24]. The tightness of \( n^{-1/2} W^{(3)}(t, z) \) follows from the uniform convergence of \( h(t, z) \) and \( A(t) \), and the asymptotic normality of \( n^{-1/2} (\hat{\beta} - \beta) \). This completes the proof of the tightness of \( n^{-1/2} W(t, z) \), and hence its weak convergence property. \( \square \)

**Proof of Theorem 2.** The perturbed version of the stochastic process is \( \hat{W}(t, z) = \sum_{i=1}^n \hat{Q}_i(t, z) \hat{z}_i \), where the \( \hat{z}_i \)'s are generated independently from \( \mathcal{N}(0, 1) \). Define \( W^*(t, z) = \sum_{i=1}^n \hat{Q}_i(t, z) \hat{z}_i \), and we have that \( n^{-1/2} \sum_{i=1}^n Q_i(t, z) \) converges weakly to a zero-mean Gaussian random field unconditionally from Theorem 1. Based on the conditional multiplier CLT in van der Vaart and Wellner [30, Theorem 2.9.6], \( n^{-1/2} W^*(t, z) \) converges weakly to the same Gaussian random field, given the data. Therefore, it suffices to prove that \( n^{-1/2} ||W^*(t, z) - \hat{W}(t, z)|| \to 0 \) in probability, where \( ||f(t, z)|| = \max_j \sup_{t \in [0, \tau], z \in [-1, 1]} |f_j(t, z)|, j = 1, \ldots, p \) for a function \( f(t, z) = (f_1(t, z), \ldots, f_p(t, z))' \) and \( f_j : ([0, \tau] \times [-1, 1]^p) \to \mathcal{R} \).

Observing \( \Phi_1(\beta, t) \) in (3.3), let
\[
\Phi_1(\beta, t) = \sum_{k=1}^K \int_0^t \{Z_{ik}(u) - \bar{z}(u)\} dM_{ik}(u).
\]
Some algebraic manipulation yields that
\[
\left\| n^{-1/2} \sum_{i=1}^n \hat{z}_i \{\Phi_1(\beta, t) - \hat{\Phi}_1(\hat{\beta}, t)\} \right\|
\leq \left\| n^{-1/2} \sum_{i=1}^n \sum_{k=1}^K \hat{z}_i [\hat{Z}(X_{ik}) - \bar{Z}(X_{ik})] \Delta_{ik} I(X_{ik} \leq t) \right\|
\]
\[
+ \left\| n^{-1/2} \sum_{i=1}^n \sum_{k=1}^K \int_0^t \hat{z}_i Y_{ik}(u) \left[\{Z_{ik} - \bar{z}(u)\} d\Delta_0(u) - \{Z_{ik} - \bar{z}(u)\} d\Delta_0(u)\right] \right\|
\]
\[
+ \left\| n^{-1/2} \sum_{i=1}^n \sum_{k=1}^K \int_0^t \hat{z}_i Y_{ik}(u) \left[\{Z_{ik} - \bar{z}(u)\} \hat{\beta}' - \{Z_{ik} - \bar{z}(u)\} \beta' \right] \right\|\left\|Z_{ik}(u) du\right\|.
\]

(A.2)

The first term on the right-hand side of (A.2) converges to zero in probability by Lemma A.3 in Spiekerman and Lin [28] and (A.1). The second term can be shown to converge to zero based
on Theorem 2 in Spiekerman and Lin [28], the consistency of \( \hat{\Lambda}_0(t) \) and (A.1). The third term converges to zero by the consistency of \( \hat{\beta} \) and (A.1). Thus, as \( n \to \infty \)

\[
\left\| n^{-1/2} \sum_{i=1}^{n} \xi_i \{ \Phi_i(\beta, t) - \hat{\Phi}_i(\hat{\beta}, t) \} \right\| \to 0 \tag{A.3}
\]

in probability. Similarly, if we define

\[
\Psi_i(\beta, t) = \sum_{k=1}^{K} \int_{0}^{t} \{ f(Z_{ik}) I(Z_{ik} \leq z) - \tilde{g}(u, z) \} \, dM_{ik}(u),
\]

\[
\hat{\Psi}_i(\hat{\beta}, t) = \sum_{k=1}^{K} \int_{0}^{t} \{ f(Z_{ik}) I(Z_{ik} \leq z) - g(u, z) \} \, d\hat{M}_{ik}(u),
\]

then we have

\[
\left\| n^{-1/2} \sum_{i=1}^{n} \xi_i \{ \Psi_i(\beta, t) - \hat{\Psi}_i(\hat{\beta}, t) \} \right\| 
\leq \left\| n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \xi_i \{ g(X_{ik}, Z_{ik}) - \tilde{g}(X_{ik}, Z_{ik}) \} \Delta_{ik} \, I(X_{ik} \leq t) \right\|
\]

\[
+ \left\| n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{t} \xi_i Y_{ik}(u) \left[ f(Z_{ik}) I(Z_{ik} \leq z) \{ d\hat{\Lambda}_0(u) - d\Lambda_0(u) \} \right.
\]

\[
+ \tilde{g}(u, z) \, d\hat{\Lambda}_0(u) - g(u, z) \, d\Lambda_0(u) \right] \right\|
\]

\[
+ \left\| n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{t} \xi_i Y_{ik}(u) \left\{ f(Z_{ik}) I(Z_{ik} \leq z) \left( \hat{\beta}' - \beta' \right) \right.
\]

\[
+ \tilde{g}(u, z) \beta' - g(u, z) \hat{\beta}' \right\} Z_{ik}(u) \, du \right\|. 
\]

Each term on the right-hand side of the above equation can be shown to converge to zero as in (A.2), and thus we have in probability

\[
\left\| n^{-1/2} \sum_{i=1}^{n} \xi_i \{ \Psi_i(\beta, t) - \hat{\Psi}_i(\hat{\beta}, t) \} \right\| \to 0. \tag{A.4}
\]

Finally, we consider

\[
\| n^{-1/2} W^*(t, z) - n^{-1/2} \hat{W}(t, z) \|
\leq \left\| n^{-1/2} \sum_{i=1}^{n} \xi_i \{ \Psi_i(\beta, t) - \hat{\Psi}_i(\hat{\beta}, t) \} \right\|
\]
\[
\begin{align*}
&+ \left| \sum_{i=1}^{n} \frac{1}{\sqrt{n}} \left[ h(t, z) \tilde{A}^{-1}(\tau) \{ \hat{\Phi}_i(\hat{\beta}, t) - \Phi_i(\beta, t) \} \right] \right| \\
&+ \left| \sum_{i=1}^{n} \frac{1}{\sqrt{n}} \left[ h(t, z) \tilde{A}^{-1}(\tau) - h(t, z) \tilde{A}^{-1}(\tau) \{ \hat{\Phi}_i(\hat{\beta}, t) \} \right] \right|.
\end{align*}
\]

By (A.4), the first term on the right-hand side of (A.5) converges to zero, and by (A.3) and the boundness of \( \tilde{h}(t, z) \) and \( \tilde{A}(t) \), the second term converges to zero. The uniform convergence of \( A(t) \) to \( \tilde{A}(t) \) and \( h(t, z) \) to \( \tilde{h}(t, z) \) for \( t \in [0, \tau] \) and \( z \in [-1, 1]^p \), and together with (A.3), entail that the third term in (A.5) converges to zero. That completes the proof. □

Consistency of the proposed tests

Under the alternative model (with non-parallel hazards), we assume that \( \lambda^*(t | Z_{ik}) \) is the hazard corresponding to subject \( k \) in cluster \( i \), and \( \hat{\beta} \) in (2.3) converges in probability to \( \beta^* \). Then,

\[
n^{-1} U(\beta, t) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \left[ \int_{0}^{t} [Z_{ik} - \bar{Z}(u)] dM_{ik}(u) \\
+ \int_{0}^{t} [Z_{ik} - \bar{Z}(u)] Y_{ik}(u) [\lambda^*(u | Z_{ik}) - \beta^* Z_{ik}] du \right].
\]

Note that \( K \) is the cluster size, then \( n^{-1} U(\hat{\beta}, t) \) converges in probability to

\[
K \int_{0}^{t} \left[ E\{Y(u) \lambda^*(u | Z)\} - \frac{E\{Y(u)Z\} E\{Y(u) \lambda^*(u | Z)\}}{E\{Y(u)\}} \right] du \\
- K \int_{0}^{t} \left[ E\{Y(u)Z^{\otimes 2} \beta^*\} - \frac{E\{Y(u)Z^{\otimes 2}\} \otimes 2 \beta^*}{E\{Y(u)\}} \right] du,
\]

which is non-zero under the alternative model. This establishes our claim.

References