Statistica Sinica **11**(2001), 723-736

GOODNESS-OF-FIT TESTS FOR SEMIPARAMETRIC MODELS WITH MULTIPLE EVENT-TIME DATA

Suktae Choi*, Zhezhen Jin† and Zhiliang Ying†

**Food and Drug Administration and †Columbia University*

Abstract: A counting process approach to multiple event times modeled by an Andersen-Gill-type extension of the Cox proportional hazards regression model is considered. Tests for checking the validity of such a model against a general frailty model are proposed. These tests are derived from a class of statistics that are connected to Robbins' empirical Bayes estimation of Poisson means. We show that these tests are consistent against any alternative as specified by a nondegenerate frailty. A simple graphical method is introduced to visually check the appropriateness of model assumptions. Simulation studies are reported and a real life example is presented. A similar test for checking the gamma frailty assumption is also introduced.

Key words and phrases: Counting process, Cox regression, empirical Bayes, goodness of fit, martingale, multiple event times, Poisson process, random effect.

1. Introduction

Multiple event-time data have been the focus of many recent investigations in survival analysis. They represent a special kind of multivariate failure time data with each subject experiencing a series of ordered events. Such data arise naturally in biomedical research. Examples include studies of progression of AIDS and other diseases, and of recurrences of chronic diseases. Other examples arise in industrial life tests and software reliability studies.

In principle, one-sample multiple event-time data can be analyzed by nonparametric methods developed for more general multivariate survival time data. Nonparametric estimates of joint survival distribution can be found in Campbell (1981), Tsai, Leurgans and Crowley (1986), Dabrowska (1988), Prentice and Cai (1992) and van der Laan (1996), among others. These estimates are generally complicated, often involving nonparametric density-type estimation. They are also difficult to incorporate into analysis of regression problems.

By modeling inter-event gap times, Prentice, Williams and Peterson (1981) proposed a regression method for multiple event-time data. Parallel and subsequent developments for such data can be found in Gail, Santner and Brown (1980) and Dabrowska, Sun and Horowitz (1992). Wei, Lin and Weissfeld (1989),

on the other hand, proposed a marginal proportional hazards regression model, in which each of the event times is modeled marginally through a Cox model. While their method has the advantage of requiring only mild model assumptions, it also has the disadvantage of not utilizing a dependency structure that can potentially increase efficiency.

A simple model that can handle multiple event-time data is provided by an extension of Cox regression to counting processes through application of a general approach due to Andersen and Gill (1982). This model can be handled in the same way as the Cox model, both numerically and theoretically. Results based on such a model are also easy to interpret. However, as will be elaborated in subsequent sections, the Andersen-Gill model does not take into account the dependence of events experienced by the same subject. To alleviate this shortcoming, one may introduce a random effect (frailty) into the Andersen-Gill model. The resulting formulation is a more general model that allows for suitable inference not only on regression effects but on associations within subject as well. Approaches in this direction can be found in Nielsen, Gill, Andersen and Sorensen (1992) and Oakes (1992).

The main objective of this investigation is to develop statistical methods for checking and testing the presence of a nondegenerate frailty. This is connected to a classical paper on empirical Bayes methodology by Robbins (1955). By utilizing an interesting identity due to Robbins for Poisson random variables with mean parameter following a prior distribution, a class of statistics is derived. The statistics are then used to construct tests for the Andersen-Gill model against alternatives of nondegenerate frailty. It is shown that the tests are consistent against any such alternative. They are also used to construct simple graphical plots for model checking, as well as for visual inspection to check where deviations from the null model may occur. Simulation results indicate that the method performs well in detecting the presence of frailty. A real-life example of tumor occurrence in experimental mice (Gail, Santner and Brown (1980)) is used to illustrate the method. A similar method for checking whether the frailty distribution belongs to the family of gamma distributions is also introduced.

2. Notation and Basic Statistics

Let $N(t), t \geq 0$ denote a counting process for events experienced by a study subject. Thus $T_1 = \inf\{t : N(t) = 1\}$ is the first event time associated with the subject and, in general, $T_k = \inf\{t : N(t) = k\}$ is the kth event time of the subject. We use Z to denote a p-dimensional covariate vector and C the censoring time, which is assumed to be independent of N conditional on Z. Suppose there are n study subjects and (C_i, Z_i, N_i), $i = 1, \ldots, n$, are independent

and identically distributed copies of $\{C, Z, N\}$, representing censoring times, covariates and counting processes of event times. Hence, observations consist of $\{C_i, Z_i, N_i(t), t \le C_i\}$, $i = 1, \ldots, n$. For simplicity, covariates Z will be assumed to be time-independent.

An application of Andersen and Gill's extension of the Cox model to multiple event times specifies that N has a nonrandom intensity function of form $\lambda_0(t)e^{\beta' Z}$, where $\lambda_0(\cdot)$ is an unspecified baseline intensity function and β is an unknown regression parameter vector. Consequently, $N(t) - \int_0^t \lambda_0(u)e^{\beta' Z}du$ is a martingale with respect to a suitable σ-filtration. By Watanabe's characterization theorem (Bremaud (1981, p.25)), N must be a nonhomogeneous Poisson process. Since C is independent of N, it follows that the stopped process $N(t \wedge C)$, where $t \wedge C = \min\{t, C\}$, has intensity $I(C \ge t)\lambda_0(u)e^{\beta' Z}$. Let $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ be the baseline cumulative intensity. The fact that N is a nonhomogeneous Poisson process entails that its time-transformed version, $N\{\Lambda_0^{-1}(t)\}$, is a homogeneous Poisson process with intensity $e^{\beta' Z}$. Therefore, $X_k = \Lambda_0(T_k) - \Lambda_0(T_{k-1})$ are independent exponential random variables with mean $e^{-\beta' Z}$.

A commonly used frailty model builds on the Andersen-Gill model by including an additional latent variable W into the intensity function in such a way that, given W and Z, the conditional intensity of N is of form

$$W\lambda_0(t)e^{\beta' Z}. \tag{2.1}$$

We refer to Nielsen, Gill, Andersen and Sorensen (1992) and Andersen, Borgan, Gill and Keiding (1993, Chapter 9) for details. To ensure identifiability, we require $EW = 1$ as $\lambda_0(\cdot)$ is completely unspecified. The frailty W represents proneness to event occurrence of a subject. It gives rise to dependency among event times of the subject, with the strength of dependency corresponding to the level of variability of W. In general, a larger variance entails a stronger dependency.

It is not difficult to see that the mean of $N(t)$ given Z has the same form as that without the frailty, i.e., $E[N(t)|Z] = \Lambda_0(t)e^{\beta' Z}$. Therefore, following Pepe and Cai (1993) and Lawless and Nadeau (1995), the Cox-type estimating equation still gives a valid estimator for β and Λ_0, but their variances may be handled by the robust approach of Wei, Lin and Weissfeld (1989). The resulting estimator for β, however, does not take the dependency into consideration and could be inefficient. On the other hand, one can specify the distribution of W in (2.1) and integrate out the frailty to obtain a parametric or semiparametric likelihood function, depending on whether the baseline intensity is specified or not. Likelihood function-based inference procedures can then be developed (Nielson, Gill, Andersen and Sorensen (1992) and Parner (1998)).

In his seminal paper on empirical Bayes methodology, Robbins (1955) derived an interesting formula for Poisson random variables. Suppose that Y given θ follows a Poisson distribution with mean θ and that θ has a prior distribution G. Then the marginal distribution of Y is given by

$$P_G(Y=k)=\int P(Y=k|\theta)dG(\theta)=\int \frac{\theta^k e^{-\theta}}{k!}dG(\theta). \tag{2.2}$$

Let (θ_1,Y_1),(θ_2,Y_2),...,(θ_n,Y_n) be a sequence of i.i.d. copies of (θ,Y). As usual, only Y_1, Y_2, ...,Y_n are observed, not $\theta_1,\theta_2,\ldots,\theta_n$. The problem considered by Robbins is to estimate the posterior mean $E(\theta|Y=k)=\int \theta^{k+1}e^{-\theta}dG(\theta)$ $/\int \theta^k e^{-\theta}dG(\theta)$. In view of (2.2), it is equal to $(k+1)P_G(Y=k+1)/P_G(Y=k)$ for all G. Therefore, a natural estimator of $E(\theta|Y=k)$ is

$$(k+1)\frac{\text{number of } Y_i \text{ such that } Y_i=k+1}{\text{number of } Y_i \text{ such that } Y_i=k}. \tag{2.3}$$

Note that the estimator in (2.3) is formed without any knowledge about the prior distribution G. Furthermore, it follows from the Law of Large Numbers that the estimator converges to $E(\theta|Y=k)$ as $n\to\infty$ and from the Central Limit Theorem that the estimator is asymptotically normal.

When the prior distribution G is degenerate, i.e., $\theta=\theta_0$ for some constant θ_0, (2.3) estimates the same quantity θ_0 for every k. Therefore, for any $m_0<m_1$,

$$\hat{\theta}(m_0,m_1)=\frac{\sum_{i=1}^n I\{m_0+1\le Y_i\le m_1+1\}}{\sum_{i=1}^n I\{m_0\le Y_i\le m_1\}/(Y_i+1)} \tag{2.4}$$

estimates θ_0. It is worth pointing out that (2.3) could lead to estimators that are variations of (2.4). For example, we can multiply each term of both summations in (2.4) by Y_i+1 and the resultant estimator, which is obviously consistent by the Law of Large Numbers, is

$$\frac{\sum_{i=1}^n (Y_i+1)I\{m_0+1\le Y_i\le m_1+1\}}{\sum_{i=1}^n I\{m_0\le Y_i\le m_1\}}.$$

Our experience with simulations indicates, however, that (2.4) generally has good efficiency.

The model specified by (2.1) can now be dealt with using the above analysis. Suppose there is no frailty, i.e., $W\equiv 1$, and that the censoring variable $C_i\ge t$ for all i. Following (2.4), it is easy to see that if there is no covariate in (2.1), i.e., $\beta=0$, the baseline cumulative hazard function $\Lambda_0(t)$ can be estimated, with $m_0=0$ and $m_1=m-1$, by

$$\frac{\sum_{i=1}^n I\{1\le N_i(t)\le m\}}{\sum_{i=1}^n I\{0\le N_i(t)\le m-1\}/[N_i(t)+1]}.$$

With covariates in (2.1), the preceding estimator needs to be extended. Specifically, let $\hat{\beta}$ be the estimator of β using the partial likelihood estimating equation (Andersen and Gill (1982)). Then $\Lambda_0(t)$ can be estimated by $\hat{\Lambda}_R(\hat{\beta}, t)$, where

$$\hat{\Lambda}_R(\beta, t) = \frac{\sum_{i=1}^n I\{1 \le N_i(t) \le m\}}{\sum_{i=1}^n e^{\beta' Z_i} I\{0 \le N_i(t) \le m-1\}/[N_i(t)+1]}. \tag{2.5}$$

One could add a small number, such as 1/4 or 1/2, to the denominator of the above expression to make the estimator more stable when the sample size is small. Such a modification clearly does not affect the large sample properties. Here, note that we estimate β based on all available information, but estimate $\Lambda_0(t)$ based only on the information up to time t.

3. Goodness-of-fit Tests

In the preceding section, we introduced $\hat{\Lambda}_R(\hat{\beta}, t)$ as an estimator of $\Lambda_0(t)$ when there is no frailty. It will be shown in this section that the estimator converges to a quantity that is always smaller than $\Lambda_0(t)$ when there is frailty. On the other hand, $\Lambda_0(t)$ can be estimated consistently by the Nelson-Aalen estimator $\hat{\Lambda}(\hat{\beta}, t)$, where

$$\hat{\Lambda}(\beta, t) = \frac{\sum_{i=1}^n N_i(t)}{\sum_{i=1}^n \exp(\beta' Z_i)};$$

cf. Andersen and Gill (1982) and Pepe and Cai (1993). Throughout this section, we assume that $C_i \ge t$ for all i.

The next theorem shows that $\hat{\Lambda}_R(\hat{\beta}, t)$ is asymptotically biased downward when the frailty W is nondegenerate. Because $\hat{\Lambda}(\hat{\beta}, t)$ is always asymptotically unbiased, a natural way to check for the presence or absence of the frailty is to contrast the two estimators of $\Lambda_0(t)$. Since $\hat{\beta}$ is consistent, note that the empirical Bayes estimator $\hat{\Lambda}_R(\hat{\beta}, t) \to \bar{\Lambda}_R(t)$ as $n \to \infty$ by the Strong Law of Large Numbers, where

$$\bar{\Lambda}_R(t) = \frac{EI\{1 \le N(t) \le m\}}{E(I\{N(t) \le m-1\} e^{\beta' Z}/[N(t)+1])}.$$

Theorem 3.1. *Under the degenerate frailty assumption, $W \equiv 1$, we have $\bar{\Lambda}_R(t) = \Lambda_0(t)$. Furthermore, if $Var(W) > 0$, then $\bar{\Lambda}_R(t) < \Lambda_0(t)$.*

The theorem will be proved in the Appendix. Because the Nelson-Aalen estimator $\hat{\Lambda}(\hat{\beta}, t)$ is asymptotically unbiased, $\hat{\Lambda}_R(\hat{\beta}, t) - \hat{\Lambda}(\hat{\beta}, t)$ (or more precisely, its standardized version) may be used to test for frailty. It is clear from Theorem 3.1 that such a test will be consistent against any frailty alternative. The critical

region can be specified by the following theorem on asymptotic normality of the test statistic.

Theorem 3.2. *Under the degenerate frailty assumption,* $W \equiv 1$, $\sqrt{n}[\hat{\Lambda}_R(\hat{\beta},t) - \hat{\Lambda}(\hat{\beta},t)]$ *converges to a normal distribution with mean 0 and variance* $V(\beta,t)$.

A consistent estimator of $V(\beta,t)$ is $\hat{V}(\hat{\beta},t)$, where

$$\hat{V}(\beta,t) = \frac{1}{n}\sum_{i=1}^{n}\Big\{\Big(\frac{\partial\hat{\Lambda}_R(\beta,t)}{\partial\beta} - \frac{\partial\hat{\Lambda}(\beta,t)}{\partial\beta}\Big)' I^{-1}(\beta)\Big(Z_i - \frac{\sum_{j=1}^n Z_j e^{\beta' Z_j}}{\sum_{j=1}^n e^{\beta' Z_j}}\Big)N_i(t) - \frac{N_i(t) - \hat{E}N_i(t)}{\sum_{j=1}^n e^{\beta' Z_j}} + \frac{I\{1 \le N_i(t) \le m\} - \hat{\Lambda}(\beta,t) I\{1 \le N_i(t) \le m\} e^{\beta' Z_i}/[N_i(t)+1] e^{\beta' Z_i}}{\sum_{j=1}^n e^{\beta' Z_j} I\{N_j(t) \le m-1\}/[N_j(t)+1]}\Big\}^2,$$

with $\hat{E}N_i(t) = \hat{\Lambda}(\beta,t)e^{\beta' Z_i}$, and $I(\beta)$ is the limit of

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^\infty \Big(\frac{\sum_{j=1}^n Z_j^{\otimes 2} I(C_j \ge u) e^{\beta' Z_j}}{\sum_{j=1}^n I(C_j \ge u) e^{\beta' Z_j}} - \Big(\frac{\sum_{j=1}^n Z_j I(C_j \ge u) e^{\beta' Z_j}}{\sum_{j=1}^n I(C_j \ge u) e^{\beta' Z_j}}\Big)^{\otimes 2}\Big) dN_i(u),$$

as $n \to \infty$ (here $a^{\otimes 2} = aa'$).

Theorem 3.2 shows that we can use $S(t) = n^{\frac{1}{2}}[\hat{\Lambda}_R(\hat{\beta},t) - \hat{\Lambda}(\hat{\beta},t)]/\Big(\hat{V}(\hat{\beta},t)\Big)^{\frac{1}{2}}$ to test the hypothesis of no frailty. Such a test is consistent against any alternative with a nondegenerate frailty. Because both $\hat{\Lambda}_R$ and $\hat{\Lambda}$ are positive, we can take log-transformations to reduce skewness. Thus, alternatively we can use the test statistic

$$S_1(t) = \frac{n^{\frac{1}{2}}\hat{\Lambda}(\hat{\beta},t)\Big[\log(\hat{\Lambda}_R(\hat{\beta},t)) - \log(\hat{\Lambda}(\hat{\beta},t))\Big]}{\Big(\hat{V}(\hat{\beta},t)\Big)^{\frac{1}{2}}}.$$

Note that the test based on S or S_1 is one-sided since $\hat{\Lambda}_R(\hat{\beta},t)$ has a negative bias when W is nondegenerate.

In principle, for any t, the constructed tests are valid and consistent. In practice, we may need to choose a suitable t. One approach is to choose t to be large, but with limited censoring up to that time point. In the case of type I simple censorship, i.e., $C_i \equiv c_0$, as in the example given in the next section, our empirical findings suggest that the choice of t at which approximately 1/2 to 2/3 of events have occurred usually gives satisfactory results in terms of accurate approximation of the type I error and reasonably good power.

Another natural way is to summarize them by integration. For example, we can use $\int_0^{t_0}\{\hat{\Lambda}_R(\hat{\beta},t) - \hat{\Lambda}(\hat{\beta},t)\}d\mu(t)$ or $\int_0^{t_0}\{\log(\hat{\Lambda}_R(\hat{\beta},t)) - \log(\hat{\Lambda}(\hat{\beta},t))\}d\mu(t)$ for

some suitably chosen measure μ. These statistics will have to be standardized, which will involve the covariance function for the limiting process $\sqrt{n}\{\hat{\Lambda}_R(\hat{\beta}, t) - \hat{\Lambda}(\hat{\beta}, t)\}$. The covariance function may be quite complicated to estimate. To avoid this, resampling methods may be used.

4. Simulation Results and a Real Example

In the preceding section we derived theoretical properties for an estimator of the baseline cumulative intensity function under the assumption of no frailty, and then proposed asymptotically valid tests for the presence of frailty. To investigate the finite-sample behavior of the proposed procedure, we conducted simulation studies.

Tables 1 and 2 report some of the simulation results using 1,000 runs per case. In Table 1, data were generated from (2.1) with frailty W following a gamma distribution with mean 1 and variance σ^2, $\lambda_0(t) \equiv 1$, and no covariates. The data generation in Table 2 is similar to that in Table 1 except the model includes a covariate Z that was generated from the uniform distribution $U[0, 1]$. The coefficient β was set to 0.25. In both cases, the censoring time is set at 3 so that, on average, there will be about 3 events for each subject at time $t = 3$.

Table 1. Simulation results for testing frailty in model without covariate using S and S_1 at the time 2/3 of events occurred.

Sample	Type I $\sigma^2 = 0$		Power $\sigma^2 = 0.25$		Power $\sigma^2 = 0.5$		Power $\sigma^2 = 1$	
n	S	S_1	S	S_1	S	S_1	S	S_1
50	0.029	0.036	0.376	0.446	0.747	0.800	0.976	0.987
75	0.032	0.045	0.558	0.602	0.917	0.935	0.999	1.000
100	0.038	0.048	0.665	0.697	0.966	0.970	1.000	1.000
125	0.043	0.048	0.780	0.801	0.989	0.992	1.000	1.000
150	0.043	0.056	0.813	0.844	0.997	0.997	1.000	1.000

Table 2. Simulation results for testing frailty in model with a covariate using S and S_1 at the time 2/3 of events occurred.

Sample	Type I $\sigma^2 = 0$		Power $\sigma^2 = 0.25$		Power $\sigma^2 = 0.5$		Power $\sigma^2 = 1$	
n	S	S_1	S	S_1	S	S_1	S	S_1
50	0.017	0.027	0.389	0.459	0.798	0.835	0.989	0.994
75	0.035	0.044	0.588	0.629	0.937	0.954	1.000	1.000
100	0.025	0.033	0.714	0.739	0.982	0.985	1.000	1.000
125	0.031	0.036	0.786	0.804	0.998	0.998	1.000	1.000
150	0.033	0.041	0.853	0.872	0.997	0.999	1.000	1.000

Tables 1 and 2 summarize empirical type I error and power at time t when 2/3 of events have occurred. The nominal level of significance was set at the usual 0.05. The results show that the test without log-transformation is rather conservative which, we believe, results from the skewness of $\hat{\Lambda}_R$. But the test gives reasonably good power in detecting frailty. As expected, the empirical type I errors in the test with log-transformation are much closer to the nominal level 0.05 than those in the test without log-transformation.

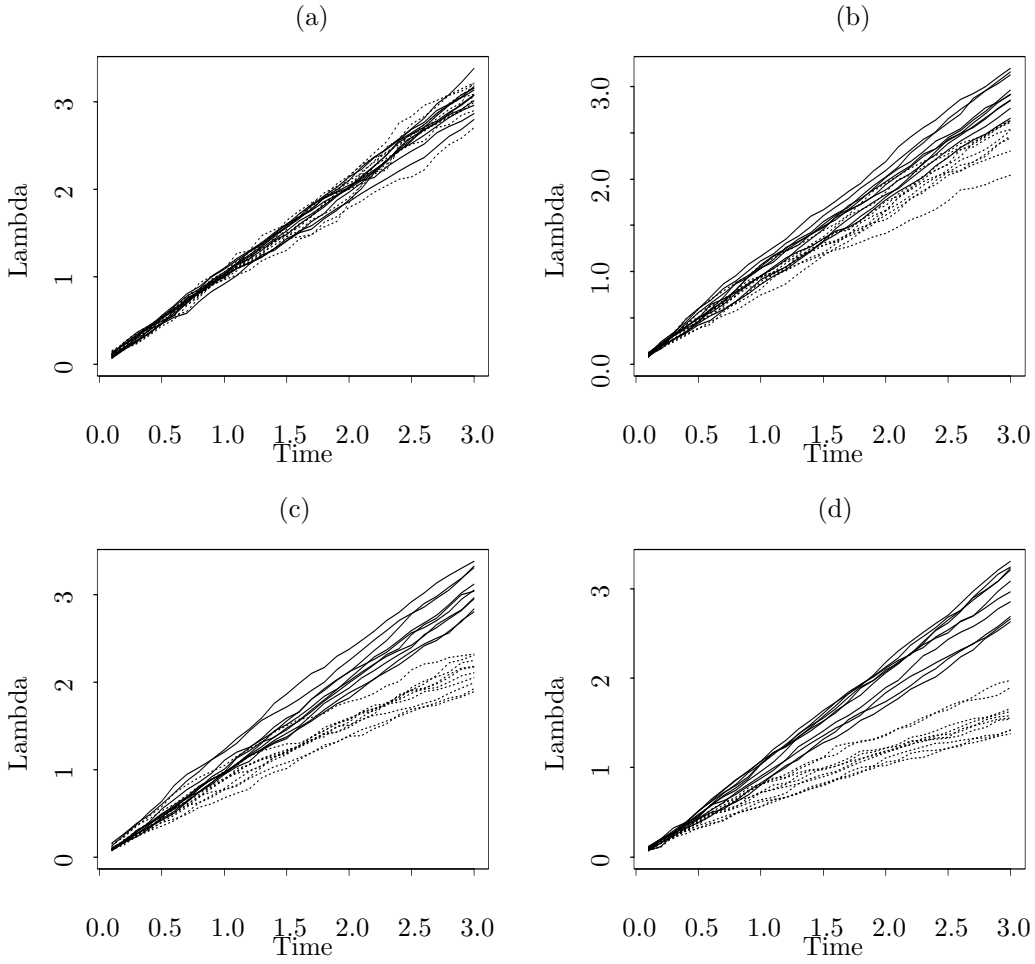

Figure 1. The Nelson-Aalen estimate $\hat{\Lambda}$ (solid lines) and the alternative estimate $\hat{\Lambda}_R$ (dotted lines). (a) $\sigma^2 = 0$; (b) $\sigma^2 = 0.25$; (c) $\sigma^2 = 0.5$; (d) $\sigma^2 = 1$.

From Theorem 1, we see that $\hat{\Lambda}_R$ should lag behind $\hat{\Lambda}$ when there is frailty.

Therefore, by plotting the two estimates simultaneously we will be able to see the impact of frailty. This is demonstrated in Figure 1, where each panel plots 10 pairs of the estimates from 10 runs. The data are generated from (2.1) with $\lambda_0(t) = 1$ and W having gamma distribution with $EW = 1$ and variance σ^2, and with no covariate included. The four panels correspond to four different values of σ^2: (a) $\sigma^2 = 0$ (no frailty); (b) $\sigma^2 = 0.25$; (c) $\sigma^2 = 0.5$; (d) $\sigma^2 = 1$. The sample size is 200 for each run. It is clear that as σ^2 increases, the two types of curves, representing $\hat{\Lambda}_R$ and $\hat{\Lambda}$, drift apart.

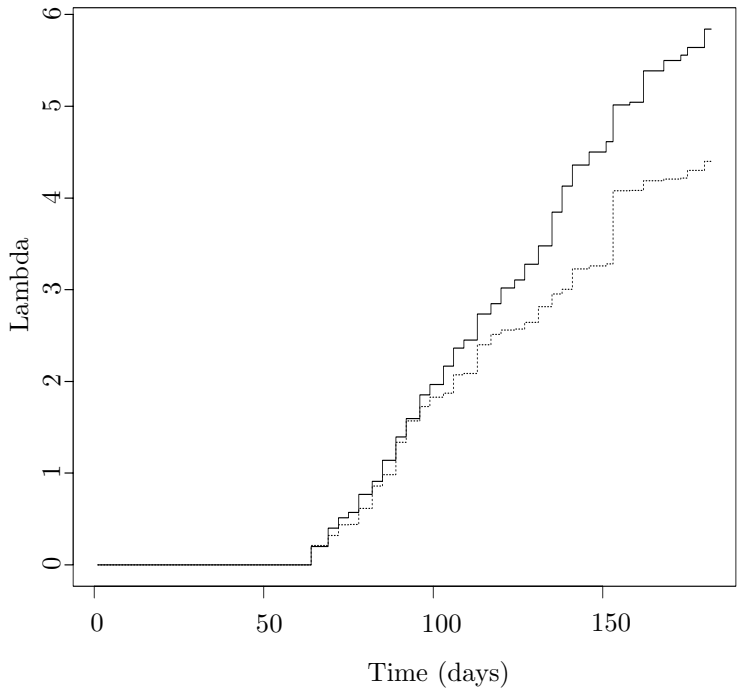

Figure 2. The Nelson-Aalen estimate $\hat{\Lambda}$ (solid line) and the alternative estimate $\hat{\Lambda}_R$ (dotted line) of baseline cumulative intensity function for tumor occurrence with the carcinogenesis experiment as reported in Gail, Santner and Brown (1980).

We next illustrate the method with a carcinogenesis experiment as reported and analyzed in Gail, Santner and Brown (1980). There were 48 female rats injected with a carcinogen for mammary cancer at day 0. They were treated with retinyl acetate to prevent cancer for 60 days, after which the treatment was continued for 23 rats and discontinued for the remaining 25. The entire

experiment lasted 182 days, thus the censoring times are a constant $C_i \equiv 182$. Each experimental animal could have multiple tumors. The response variables are the event times for the appearance of tumors. It was shown in Oakes (1992) that the Andersen-Gill-type model is not suitable and that an obvious frailty exists. We apply test statistics S and S_1 as defined in Section 3, with a covariate Z and with time $t = 137$ at which about 2/3 of events occurred. The covariate Z takes the value 1 if the rat was treated and the value 0 if it was not. The estimate of the coefficient β is -0.823 based on the Andersen-Gill model. The standardized test statistics S and S_1 are -2.16 and -2.46, respectively. The resulting p-values are 0.015 and 0.007. Therefore, there is a strong evidence for the presence of frailty, in agreement with the finding of Oakes (p-value = 0.005).

We plot the two estimates as given in Figure 2. Because of the presence of frailty, we expect the two curves to be separated and they clearly deviate from each other.

5. Testing Gamma Frailty

If there is a nondegenerate frailty that follows a gamma distribution, inference can be carried out easily with well-established asymptotic theory (Murphy (1994) and Parner (1998)). Although Parner (1998) provided asymptotic theory for the general gamma frailty model, his result cannot be used to test the gamma frailty assumption. In practice, however, it is very important to check this assumption.

In this section, we show that a similar idea can be used to construct a test or tests for the null hypothesis that the frailty distribution belongs to the family of gamma distributions. Here, we will only outline the approach. For this method to become effective, we find that the number of events experienced by each subject should not be small. For simplicity, we consider the one-sample problem, it is straightforward to extend the approach to the case when Z takes a finite number of values.

Suppose the counting process N is modeled by (2.1), with W a gamma distribution with mean 1 and variance σ^2. Define $H_j(t) = E[W\Lambda_0(t)|N(t) = j-1]$, $j \geq 1$. By Bayes formula, it is not difficult to show that

$$H_j(t) = \Lambda_0(t)\frac{E\{W^j \exp[-W\Lambda_0(t)]\}}{E\{W^{j-1}\exp[-W\Lambda_0(t)]\}} = \frac{\Lambda_0(t)}{1+\sigma^2\Lambda_0(t)}\left[1+(j-1)\sigma^2\right]. \quad (5.1)$$

Since the middle expression in (5.1) is equal to $jP(N_i(t) = j)/P(N_i(t) = j-1)$, $H_j(t)$ can be consistently estimated by

$$\hat{H}_j(t) = \frac{j\sum_{i=1}^n I\{N_i(t) = j\}}{\sum_{i=1}^n I\{N_i(t) = j-1\}}.$$

It is obvious that $H_j(t)/H_1(t) = 1 + (j-1)\sigma^2$, which does not involve t. Therefore, in large samples, $\hat{A}(j,t) = \hat{H}_j(t)/\hat{H}_1(t)$ should be approximately constant in t and linear in j. Significant deviation from such a pattern could indicate departure from the gamma frailty assumption.

Further modifications to $\hat{A}(j,t)$ are possible. For example, $\hat{H}_1(t)$ may be replaced by the Nelson estimator of the cumulative hazard function of the first event time. One could also take the average of $\hat{H}_j(t)$ over time to eliminate t, for example, $\tilde{H}_j = \int \hat{H}_j(t)d\mu(t)$ for some known measure $\mu(t)$, and then compute the ratios $\tilde{H}_j/\tilde{H}_0$. In doing so, the pattern to be checked against will be the linearity of the ratios over j.

Preliminary simulation results indicate that for this approach to perform reasonably well the number of subjects and the number of events per subject need to be large. Further research on this matter is currently underway and results will be reported elsewhere.

6. Remarks

The test statistics $S(t)$ and $S_1(t)$ depend on t. Note that the assumption $C_i \geq t$ was made to ensure its validity. When the C_i can take small values, modifications to S or S_1 are needed. One simple measure is to first partition subjects into several groups according to their C_i values, and then construct S or S_1 separately for each group. If C is independent of $\{Z, N(t)\}$, a possible modification is to introduce $I\{C_i \geq t\}$ to each term in the estimators $\hat{\Lambda}_R(\beta,t)$ and $\hat{\Lambda}(\beta,t)$.

Modifications of test statistic are needed when there is random censoring. One possible approach is to use the idea outlined in previous paragraph. However, the investigation of this approach is very much incomplete.

Acknowledgements

We are very grateful to the editor, an associate editor and referees for helpful comments on the paper.

Appendix

We prove Theorems 3.1 and 3.2. For Theorem 3.1, we need an inequality which can be found in Casella and Berger (1990, p.184).

Lemma A.1. *Suppose that f is an increasing function, g a decreasing function. Then for any random variable X, $E\{f(X)g(X)\} \leq Ef(X)Eg(X)$. The inequality is strict if f is strictly increasing, g strictly decreasing and X nondegenerate.*

Proof of Theorem 3.1. It is easy to check that

$$\bar{\Lambda}_R(t) = \frac{EI\{1 \le N(t) \le m\}}{E(I\{N(t) \le m-1\}e^{\beta' Z}/[N(t)+1])} = \Lambda_0(t)$$

when W is degenerate. We only need to show that

$$\bar{\Lambda}_R(t) < \Lambda_0(t) \tag{A.1}$$

when W is nondegenerate. Now

$$EI\{1 \le N(t) \le m\} = E\Big[\sum_{j=1}^{m} \frac{1}{j!}\Big(W\Lambda_0(t)e^{\beta' Z}\Big)^j \exp\{-W\Lambda_0(t)e^{\beta' Z}\}\Big].$$

Let $f(W) = W$ and

$$g(W) = \sum_{j=1}^{m} \frac{1}{j!} E\Big[\Big(W\Lambda_0(t)e^{\beta' Z}\Big)^j \exp\{-W\Lambda_0(t)e^{\beta' Z}\}\frac{1}{W}\,|W\Big].$$

Then f is strictly increasing. In addition, it is not difficult to show that $g'(W) < 0$. By Lemma A.1, $E[f(W)g(W)] < Ef(W)Eg(W)$, which clearly entails (A.1).

Proof of Theorem 3.2. Write

$$\sqrt{n}\Big[\hat{\Lambda}_R(\hat{\beta}, t) - \hat{\Lambda}(\hat{\beta}, t)\Big] = I_1 + I_2 + I_3 + I_4, \tag{A.2}$$

$$\begin{aligned}
I_1 &= \sqrt{n}\Big(\hat{\Lambda}_R(\hat{\beta}, t) - \hat{\Lambda}_R(\beta, t) - \hat{\Lambda}(\hat{\beta}, t) + \hat{\Lambda}(\beta, t)\Big),\\
I_2 &= \sqrt{n}\Big(\hat{\Lambda}_R(\beta, t) - \frac{E\sum_{i=1}^n I\{1 \le N_i(t) \le m\}}{\sum_{i=1}^n I\{N_i(t) \le m-1\}e^{\beta Z_i}/(N_i(t)+1)}\Big),\\
I_3 &= \sqrt{n}\Big(\frac{E\sum_{i=1}^n I\{1 \le N_i(t) \le m\}}{\sum_{i=1}^n I\{N_i(t) \le m-1\}e^{\beta' Z_i}/(N_i(t)+1)} - \Lambda_0(t)\Big),\\
I_4 &= \sqrt{n}\Big(\Lambda_0(t) - \hat{\Lambda}(\beta, t)\Big).
\end{aligned}$$

Note that

$$\begin{aligned}
I_1 &= \Big[\frac{\partial\hat{\Lambda}_R(\beta, t)}{\partial\beta} - \frac{\partial\hat{\Lambda}(\beta, t)}{\partial\beta}\Big]\sqrt{n}(\hat{\beta} - \beta) + o(\|\hat{\beta} - \beta\|),\\
I_2 &= \frac{\eta(t)}{\sqrt{n}}\sum_{i=1}^{n}\Big[I\{1 \le N_i(t) \le m\} - EI\{1 \le N_i(t) \le m\}\Big] + o_p(1),\\
I_3 &= \frac{\eta(t)\Lambda_0(t)}{\sqrt{n}}\sum_{i=1}^{n}\Big[E\Big(\frac{I\{N_i(t) \le m-1\}e^{\beta' Z_i}}{N_i(t)+1}\Big) - \frac{I\{N_i(t) \le m-1\}e^{\beta' Z_i}}{N_i(t)+1}\Big] + o_p(1),
\end{aligned}$$

and $\quad I_4 = \dfrac{1}{\sqrt{n}Ee^{\beta' Z}}\displaystyle\sum_{i=1}^{n}[EN_i(t) - N_i(t)] + o_p(1),$

where $\eta(t) = \Big\{EI\{N(t) \le m-1\}e^{\beta' Z}/(N(t)+1)\Big\}^{-1}$.

By the standard result on $\hat{\beta}$ (Andersen and Gill (1982)), $\sqrt{n}(\hat{\beta}-\beta) = I^{-1}(\beta)U(\beta) + o_p(1)$, where $U(\beta)$ is the partial likelihood score function for β and is equal to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\infty}\Big(Z_i - \frac{\sum_{j=1}^n Z_j I(C_j \ge u)e^{\beta' Z_j}}{\sum_{j=1}^n I(C_j \ge u)e^{\beta' Z_j}}\Big)dN_i(u).$$

Therefore, (A.2) is asymptotically a sum of independent zero-mean random variables. The standard Multivariate Central Limit Theorem can be used to get the convergence in distribution. Finally, $\hat{V}(\hat{\beta}, t)$ converges to the limiting variance by the Law of Large Numbers and the consistency of $\hat{\beta}$.

References

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

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-1120.

Bremaud, P. (1981). *Point Processes and Queues: Martingale Dynamics*. Springer-Verlag, New York.

Casella, G. and Berger, R. L. (1990). *Statistical Inference*. Wadsworth and Brooks/Cole, Pacific Grove, California.

Campbell, G. (1981). Nonparametric bivariate estimation with randomly censored data. *Biometrika* **68**, 417-522.

Dabrowska, D. M. (1988). Kaplan-Meier estimate on the plane. *Ann. Statist.* **16**, 1475-1489.

Dabrowska, D. M., Sun, G. and Horowitz, M. M. (1992). Cox regression in a Markov renewal model with an application to the analysis of bone marrow transplant data. *J. Amer. Statist. Assoc.* **89**, 867-877.

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

Gail, M. H., Santner, T. J. and Brown, C. C. (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics* **36**, 255-266.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37**, 158-168.

Nielsen, G. G., Gill, R. D., Andersen, P. K. and Sorensen, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.* **19**, 25-43.

Oakes, D. (1992). Frailty models for multiple event-time data. In *Survival Analysis: State of the Art* (Edited by J. P. Klein and P. K. Goel), 471-480. Kluwer, Boston.

Parner, E. (1998). Asymptotic theory for the correlated Gamma-frailty model. *Ann. Statist.* **26**, 183-214.

Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *J. Amer. Statist. Assoc.* **88**, 811-820.

Prentice, R. L. and Cai, J. (1992). Covariance and survival function estimation using censored multivariate failure time data. *Biometrika* **79**, 495-512.

Prentice, R. L., Williams, B. J. and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68**, 373-379.

Robbins, H. (1955). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symposium on Math. Statist. and Prob.* **1**, 157-164.

Tsai, W-Y., Leurgans, S. and Crowley, J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring. *Ann. Statist.* **14**, 1351-1365.

van der Laan, M. J. (1996). Efficient estimation in the bivariate censoring model and repairing NPMLE. *Ann. Statist.* **24**, 596-627.

Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.* **84**, 1065-1073.

Food and Drug Administration, 9201 Corporate Blvd, Rockville, MD 20850, U.S.A.

E-mail: chois@cder.fda.gov

Division of Biostatistics, The Joseph L. Mailman School of Public Health, Columbia University, 600 West 168th Street, New York, NY 10032, U.S.A.

E-mail: zjin@biostat.columbia.edu

Department of Statistics, Columbia University, 2990 Broadway, New York, NY 10027, U.S.A.
E-mail: zying@stat.columbia.edu

(Received December 1999; accepted March 2001)