Tests for Comparing Mark-Specific Hazards and Cumulative Incidence Functions

Peter B. Gilbert, Ian W. McKeague, and Yanqing Sun

December 28, 2001

Abstract

It is of interest in some applications to determine whether there is a relationship between a hazard rate function (or a cumulative incidence function) and a mark variable which is only observed at uncensored failure times. We develop nonparametric tests for this problem when the mark variable is continuous. Tests are developed for the null hypothesis that the mark-specific hazard rate is independent of the mark versus ordered and two-sided alternatives expressed in terms of mark-specific hazard functions and mark-specific cumulative incidence functions. No assumptions are made about the nature of dependence between the risks. The test statistics are based on functionals of a bivariate test process equal to a weighted average of differences between a Nelson–Aalen-type estimator of the mark-specific cumulative hazard function and a nonparametric estimator of this function under the null hypothesis. The weight function in the test process can be chosen so that the test statistics are asymptotically distribution-free. Since the limiting covariance structure of the test process is complicated, asymptotically correct critical values are obtained through a simple simulation procedure. Numerical studies show that the testing procedure has good size and power characteristics at moderate sample sizes. We present an application to viral genetics data collected in an AIDS clinical trial. Specifically, the tests are used to assess if the instantaneous or absolute risk of treatment failure depends on the amount of accumulation of drug resistance mutations in

[1]Peter B. Gilbert is Assistant Member, Fred Hutchinson Cancer Research Center and Research Assistant Professor of Biostatistics, University of Washington, Seattle, WA 98105 (E-mail: pgilbert@scharp.org); Ian W. McKeague is the Ralph A. Bradley Professor of Statistics, Florida State University, Tallahasee, FL 32306 (E-mail: mckeague@stat.fsu.edu); and Yanqing Sun is Associate Professor of Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223 (E-mail: yasun@uncc.edu). This research was partially supported by NIH grant 1 U01 A146703-01 (Gilbert), NSF grant 9971784 (McKeague), and funds provided by the University of North Carolina at Charlotte (Sun).

a subject's HIV virus. This assessment helps guide development of anti-HIV therapies that surmount the problem of drug resistance.

KEY WORDS: Competing Risks; Continuous Mark; Distribution-Free Tests; Genetic Data, Nonparametric Statistics; Right-Censored Data.

1. INTRODUCTION

Many studies of survival data involve mark variables that are only observed at an endpoint event and it is of interest to investigate whether there is any relationship between the time to endpoint and the mark variable. For example, in a clinical trial of drug regimens for treating HIV infection, the time to treatment failure (typically defined by levels of viral load rising above a threshold (Gilbert, DeGruttola, Hammer, and Kuritzkes 2001)) can decrease with increases in a distance measure describing the extent of drug-selected HIV genetic evolution within a patient between baseline and the time of failure. Detecting such an association can help in designing anti-HIV treatments that overcome the problem of drug resistance, which represents one of the greatest barriers to achieving durably efficacious treatment of HIV infection (HHS Report 1998; Hirsch et al. 1998).

In this article we develop tests for detecting whether a mark-specific hazard rate (or cumulative incidence function) depends on the mark, and apply the tests to HIV genetic data collected in an AIDS clinical trial. If we denote the time to endpoint T and the mark variable V, the observable random variables are $(X, \delta, \delta V)$, where $X = \min\{T, C\}$, $\delta = I(T \leq C)$, and C is a censoring random variable that is assumed to be independent of T and V. When the failure time T is observed, $\delta = 1$ and the mark V is also observed, whereas if T is censored, the mark is unknown. Statistical interest focuses on the mark-specific hazard rate function

$$\lambda(t, v) = \lim_{h_1, h_2 \to 0} P\{T \in [t, t+h_1), V \in [v, v+h_2) | T \geq t\}/h_1 h_2, \tag{1}$$

and the cumulative incidence function

$$F(t, v) = \lim_{h_2 \to 0} P\{T \leq t, V \in [v, v+h_2)\}/h_2, \tag{2}$$

with t ranging over a fixed interval $[0, \tau]$. If V is discrete, the limit $h_2 \to 0$ is not needed, and the definitions (1) and (2) simplify respectively to the discrete cause-specific hazard function and the discrete cumulative incidence function, which have received much attention in the competing risks literature. In this article, the mark variable V is assumed to be continuous, in which case the functions (1) and (2) are the natural analogs of their discrete counterparts, with similar interpretations. In particular, $\lambda(t,v)$ is the instantaneous risk of failure by a cause V in a small interval $[v, v+h_2)$ in the presence of all other causes, and $F(t,v)$ is the probability that failure with V in a small interval $[v, v+h_2)$ will occur before the specified time t. Furthermore, just as the cause-specific hazard functions are the basic estimable quantities forming the basis for inference when the mark variable is discrete (as originally pointed out by Prentice et al., 1978), the mark-specific hazard function (1) is estimable from the available data and forms the basis for inference when the mark variable is continuous. These considerations motivate us to develop an inferential procedure based on the function (1). Note that the mark variable V appears *unconditionally* in this framework, in contrast to covariates that appear *conditionally* in hazard function regression modeling; different methods are needed for studying mark-specific hazard rates.

The continuous mark variable V is assumed to have known bounded support; rescaling V if necessary, this support is taken to be $[0,1]$. Our interest centers on testing the null hypothesis

$$H_0\colon\ \lambda(t,v) \text{ does not depend on } v \text{ for } t \in [0,\tau]$$

against the following alternative hypotheses:

$$\begin{aligned} H_1\colon\ F(t,v_1) &\leq F(t,v_2) \text{ for all } v_1 \leq v_2,\ t \in [0,\tau];\\ H_2\colon\ \lambda(t,v_1) &\leq \lambda(t,v_2) \text{ for all } v_1 \leq v_2,\ t \in [0,\tau];\\ H_3\colon\ \lambda(t,v_1) &\neq \lambda(t,v_2) \text{ for some } v_1 \leq v_2,\ t \in [0,\tau] \end{aligned}$$

with strict inequalities for some t, v_1, v_2 in both H_1 and H_2. Equivalently, H_0 can be expressed in terms of the cumulative incidence function $F(t,v)$ not depending on v for $t \in [0,\tau]$. The null hypothesis H_0 can also be written as $H_0 : \lambda(t,v) = \lambda(t)$ for all $t \in [0,\tau]$ and $v \in [0,1]$, where $\lambda(t)$ is the overall hazard function of T. Expressed in this way, H_0 is the continuous version of the null hypothesis considered by Aly, Kochar, and McKeague (1994), who developed a test of equality of two discrete cause-specific hazard rates, $H_0 : \lambda_1(t) = \lambda_2(t)$ for all $t \in [0,\tau]$.

As for the case of discrete competing risks, the interpretation of inferences on the mark-specific hazard function $\lambda(t,v)$ is restricted to actual study conditions (i.e., is "crude" or "gross"), and there is no implication that the same inference would be made under a new set of conditions in which, for example, certain causes of failure v were not present. With T_v denoting the latent (i.e. notional) failure time for mark v (see Prentice et al., 1978, for discussion of latent failure times), the assumption of mutual independence of the T_v for all $v \in [0,1]$ is needed for $\lambda(t,v)$ to possess the stronger interpretation as the hazard function for cause v given that all other causes are inoperative. As in the discrete case, the independence assumption is untestable from the available competing risks data (c.f., Tsiatis, 1975); additional data such as observations of marks beyond the first failure time are needed. Thus, tests of H_0 have an interpretation in terms of association, and cannot be used for causal inference of the predictive effect of a mark variable on the risk of failure. The fact that the mark variable is only observed simultaneously with failure makes clear the impossibility of causal inference.

In the AIDS clinical trial example, V is a measure of the accumulated HIV genetic resistance resulting from exposure to an antiretroviral treatment, which is measured only on subjects who fail treatment, at the time of treatment failure. The test of H_0 versus the monotone alternative $H_1(H_2)$ assesses whether the absolute (instantaneous) risk of treatment failure increases with the level of acquired drug resistance. If V is a reliable measure of the "resistance cost" of the regimen, i.e., if the risk of treatment failure is higher for larger values of V, then we would expect to reject H_0 in favor of H_2. Thus, the test is useful for evaluating if V is a clinically relevant measure of a treatment's resistance cost (see Gilbert et al. 2000 for a discussion of relevant resistance cost metrics). Knowledge of clinically meaningful genetic resistance cost metrics would be helpful for identifying combination drug regimens that do not select for drug resistant virus, and thus provide long-lasting treatment efficacy. Note that an alternative approach would attempt to treat V as a covariate and would test if the level of V predicts the risk of treatment failure. This approach is not possible, however, because V is only observed in subjects who fail treatment.

A second example in which the proposed approach would be of interest is a prospective cohort study of a population at risk for acquiring HIV infection. In this application, T is the time from cohort entry until HIV infection, and V is the value of a metric measuring genotypic or phenotypic dissimilarity of the HIV virus that infects a study participant from a reference HIV strain. For example, V could be Hamming's genetic distance and the reference strain could be the prototype virus contained in an HIV vaccine that is under development for field testing in the cohort population. The test of H_0 versus the two-sided hypothesis H_3 assesses whether the HIV metric V is associated with the instantaneous risk of HIV infection. Finding evidence for H_3 may suggest that the metric V can be used to guide selection of the types of HIV antigens to include in HIV vaccines (Gilbert, Self, Rao, Naficy, and Clemens 2001). For example, if the infection risk appears particularly high for v in the sub-interval [0.7,0.8], then it may behoove vaccine researchers to insert HIV antigens characterized by v contained in [0.7,0.8]. Carrying out the test for multiple metrics in multiple genes could help identify the metric(s) that should be focused on for optimizing the breadth of expected protective coverage of the vaccine. This application is important because the broad genotypic and phenotypic diversity of HIV poses one of the greatest challenges to developing an effective AIDS vaccine (Graham and Karzon 1998; Gilbert, Self, Rao, Naficy, and Clemens 2001).

In the case of a discrete mark variable, tests for comparing mark-specific hazards can be found in the literature on competing risks, see, e.g., Aly, Kochar, and McKeague (1994), Sun and Tiwari (1995), Lam (1998), Hu and Tsai (1999), Luo and Turnbull (1999) and Sun (2001). To the best of our knowledge, however, these tests have not been developed for continuous mark variables. As commented on earlier, our testing procedure can be viewed as a continuous extension of the procedure of Aly, Kochar, and McKeague (1994).

A nonparametric estimator of the joint distribution of a failure time and a failure mark which may be continuous has been introduced by Huang and Louis (1998), with a view to applications such as evaluating the relationship between a quality of life score and survival time (Olschewski and Schumacher 1990), or between lifetime medical cost and survival time. Their estimator could be used to test whether T and V are independent by comparing it with the product of its marginals. A test statistic based on this approach would have a complex asymptotic distribution, however, and it is not clear that a tractable testing procedure could be formulated. Furthermore, given the interpretability of the mark-specific hazard function in terms of the instantaneous risk of failure, we argue that in some biomedical problems testing $\lambda(t, v)$ independent of v is more directly relevant than testing T and V independent. For example, for an HIV infected patient receiving effective antiretroviral treatment at a given time, the risk of treatment failure over the next month is of primary clinical interest, and is measured by the hazard function; accordingly the relationship between the mark variable and the hazard function is of direct clinical interest.

In the case of finitely many causes of failure, test statistics can be based on differences between Nelson–Aalen estimators of the cumulative cause-specific hazard functions, see Sun (2001). Generalizing this approach, our test procedure is based on estimates of the doubly cumulative mark-specific hazard function $\Lambda(t, v) = \int_0^v \int_0^t \lambda(s, u)\, ds\, du$; estimation of this function was also used in a fundamental way by Huang and Louis (1998). The idea of our testing procedure is to compare a nonparametric estimate of $\Lambda(t, v)$ with an estimate under H_0. We show that the comparison can be weighted to make the test statistics asymptotically distribution-free.

This article is organized as follows. In Section 2 we introduce the test statistics and describe a Monte Carlo procedure for approximating critical values. In Section 3 we derive the asymptotic null distributions of the test statistics, and show that the Monte Carlo-derived critical values are asymptotically accurate. The results of a simulation study and the AIDS clinical trial example are presented in Sections 4 and 5, and proofs of results are given in the Appendix.

2. TEST PROCEDURE

Given observation of i.i.d. replicates $(X_i, \delta_i, \delta_i V_i),\ i = 1, \ldots, n$ of the (possibly right-censored) marked failure times, a suitable nonparametric estimator of $\Lambda(t, v)$ is provided by the Nelson–Aalen-type estimator

$$\hat{\Lambda}(t, v) = \int_0^t \frac{N(ds, v)}{Y(s)},\ t \geq 0,\ v \in [0, 1], \tag{3}$$

where $Y(t) = \sum_{i=1}^n I(X_i \geq t)$ is the size of the risk set at time t, and

$$N(t, v) = \sum_{i=1}^n I(X_i \leq t, \delta_i = 1, V_i \leq v)$$

is the marked counting process with jumps at the uncensored failure times X_i and associated marks V_i, cf. Huang and Louis (1998, eq. 3.2). A closely related doubly cumulative hazard function estimator was introduced by McKeague and Utikal (1990) and McKeague, Nikabadze, and Sun (1995) for testing independence of a covariate from a failure time. The two estimators are not interchangeable, however, because the covariate and the mark variable play different roles in each setting.

From the representation $H_0 : \lambda(t, v) = \lambda(t)$ for all $t \in [0, \tau]$ and $v \in [0, 1]$, it can be shown that H_0 holds if and only if T and V are independent and V is uniformly distributed over $[0, 1]$. Because of this fact, we can write $\Lambda(t, v) = v\Lambda(t, 1)$, where $\Lambda(\cdot, 1)$ is the cumulative hazard function of T under H_0. Thus, under H_0 we can estimate the doubly cumulative hazard function by $\bar{\Lambda}(t, v) = v\hat{\Lambda}(t, 1)$.

2.1 Test Processes and Test Statistics

We consider test processes of the form

$$L_n(t, v) = \sqrt{n} \int_0^t H_n(s)\, (\hat{\Lambda} - \bar{\Lambda})(ds, v) \tag{4}$$

for $t \geq 0, 0 \leq v \leq 1$, where $H_n(\cdot)$ is a suitable weight process. The weight process $H_n(\cdot)$ provides a flexible way to specify the relative importance attached to differences in the mark-specific hazards at different times, and is useful for controlling instability in the tails. The bivariate test process

$L_n(t,v)$ is similar to the univariate test process $L_n(t)$ used by Aly, Kochar, and McKeague (1994, p. 996) for comparing two competing risks 1 and 2, given by $L_n(t) = \int_0^t w(s) d(\hat{\Lambda}_1 - \hat{\Lambda}_2)(s)$, with $\hat{\Lambda}_j(\cdot)$ the cause-j-specific Nelson-Aalen estimator.

Let $y(t) = P(X \geq t)$ and $\tilde{\tau} = \sup\{t: y(t) > 0\}$ and assume $\tau < \tilde{\tau}$. We propose the following test statistics to measure departures from H_0 in the direction of H_1, H_2 and H_3:

$$U_1 = \sup_{v_1 < v_2} \sup_{0 \leq t < \tau} \Delta(t, v_1, v_2) \tag{5}$$

$$U_2 = \sup_{v_1 < v_2} \sup_{0 \leq s \leq t < \tau} (\Delta(t, v_1, v_2) - \Delta(s, v_1, v_2)) \tag{6}$$

$$U_3 = \sup_{v_1 < v_2} \sup_{0 \leq t < \tau} |L_n(t, v_1) - L_n(t, v_2)| \tag{7}$$

where $\Delta(t, v_1, v_2) = L_n(t, v_1) + L_n(t, v_2) - 2L_n(t, (v_1 + v_2)/2)$.

In the next section we show that $L_n(t,v)$ converges weakly to a Gaussian process under H_0. We also show that the proposed tests based on the U_j are consistent against their respective alternatives. Since each U_j is a continuous functional of $L_n(t,v)$, its limiting null distribution is the distribution of the corresponding functional of the limiting Gaussian process. These distributions are intractable, however, so the critical values of the U_j need to be determined using a Monte Carlo procedure.

2.2 Monte Carlo Procedure

The procedure is based on a randomized version U_j^* of U_j defined by replacing each V_i by V_i^*, where $V_1^*, \ldots, V_n^*$ are i.i.d. uniform [0,1] random variables. This yields a randomized version of the test process given by

$$L_n^*(t,v) = \sqrt{n} \int_0^t H_n(s) \, (\hat{\Lambda}^*(ds, v) - v\hat{\Lambda}^*(ds, 1)), \tag{8}$$

where $\hat{\Lambda}^*(t,v) = \int_0^t N^*(ds,v)/Y(s)$ and $N^*(t,v) = \sum_{i=1}^n I(X_i \leq t, \delta_i = 1, V_i^* \leq v)$. Exploiting the property that T and V are independent and V is uniformly distributed over $[0,1]$ under H_0, in Section 3 we show that the null distribution of U_j coincides in the limit with the conditional distribution of U_j^* given the observed data. Therefore a critical value of U_j can be approximated via a Monte Carlo estimate of the quantile of U_j^* corresponding to a given level of the test.

2.3 Choice of Weight Process and a Graphical Procedure

The simplest weight process, $H_n(t) = 1$, yields a test process equal to a normalized difference of estimated doubly cumulative mark-specific hazard functions evaluated at v and at 1:

$$L_n(t,v) = \sqrt{n}\left(\hat{\Lambda}(t,v) - v\hat{\Lambda}(t,1)\right). \tag{9}$$

This process is useful for a graphical procedure, in which the surface $L_n(t,v)$ is plotted together with 10 or 20 realizations of the simulated null surface $L_n^*(t,v)$. Relative to the reference processes $L_n^*(t,v)$, large positive values of $L_n(t,v)$ increasing with v suggest H_1, an increasing trend of $L_n(t,v)$ with time suggests H_2, and large absolute differences in $L_n(t,v)$ over different mark values suggest H_3. The graphical procedure is illustrated in the example given in Section 5.

To give the tests maximal power, the weight process should be chosen to downweight the comparison of mark-specific hazards at larger times, where the test process is most variable. A weight process that accomplishes this is given by $H_n(t) = \hat{S}_C(t-)\hat{S}_T^{1/2}(t)$, where $\hat{S}_C$ and $\hat{S}_T$ are the Kaplan–Meier estimators of S_C and S_T, respectively, S_C being the survivor function of C. As shown in the next section, this weight process has the added advantage of making the test statistics asymptotically distribution-free.

3. LARGE-SAMPLE RESULTS

We begin by defining notation that is used in the sequel. Let $\gamma(t,v) = P(X \le t, \delta = 1, V \le v)$. By the Glivenko–Cantelli Theorem, $N(t,v)/n$ and $Y(t)/n$ converge almost surely to $\gamma(t,v)$ and $y(t)$, uniformly in $(t,v) \in [0,\infty) \times [0,1]$ and $t \in [0,\infty)$, respectively. Note that we may write $\lambda(t,v) = f(t,v)/S_T(t)$, where $S_T(t) = P(T > t)$ and $f(t,v)$ is the joint density of (T,V). Let $\lambda(t) = f_T(t)/S_T(t)$ be the hazard function of the failure time T, where $f_T(t)$ is the density of T. Let $D(I)$ be the Skorohod space for a k-dimensional rectangle I (Bickel and Wichura 1971), and $C(I)$ be the subspace of continuous functions on I. Also, let $x \wedge y$ and $x \vee y$ denote the minimum and maximum of x and y, respectively.

Our first result describes the limiting null distribution of the test process.

Theorem 1. Let the weight process $H_n(t)$ be a continuous functional of the processes $N(t,1)$ and $Y(t)$, $t \in [0,\tau]$, $\tau < \tilde{\tau}$. Assume there exists a measurable function $H(t)$ such that $\sup_{0 \le t \le \tau} |H_n(t) - H(t)| \xrightarrow{\text{a.s.}} 0$ and both H_n and H have bounded variation independent of n almost surely. Then, under H_0

$$L_n(t,v) \xrightarrow{\mathcal{D}} \int_0^t H(s)y(s)^{-1}(G_1(ds,v) - vG_1(ds,1)) \equiv L(t,v) \tag{10}$$

in $D([0,\tau] \times [0,1])$ as $n \to \infty$, where $G_1(t,v)$ and $G_2(t)$ are continuous mean zero Gaussian processes with covariances

$$\begin{aligned}
\mathrm{Cov}(G_1(s,u), G_1(t,v)) &= \gamma(s \wedge t, u \wedge v) - \gamma(s,u)\gamma(t,v), \\
\mathrm{Cov}(G_2(s), G_2(t)) &= y(s \vee t) - y(s)y(t), \\
\mathrm{Cov}(G_1(t,v), G_2(s)) &= (\gamma(t,v) - \gamma(s-,v))I(s \le t) - \gamma(t,v)y(s).
\end{aligned}$$

The limiting process $L(t,v)$ is a mean zero Gaussian process with covariance

$$\mathrm{Cov}(L(s,u), L(t,v)) = (u \wedge v - uv) \int_0^{s \wedge t} \frac{H(r)^2}{y(r)^2}\,\gamma(dr,1). \tag{11}$$

The process $L(t,v)$ resembles the Kiefer–Müller process (van der Vaart and Wellner 1996, p. 226).

We next establish that the randomized version of the test process $L_n^*(t,v)$, introduced in Section 2.2, has the same limiting null distribution as $L_n(t,v)$.

Theorem 2. Under the conditions of Theorem 1, conditional on the observed data sequence,

$$L_n^*(t,v) \xrightarrow{\mathcal{D}} L(t,v) \tag{12}$$

in $D([0,\tau] \times [0,1])$ under H_0, where $L(t,v)$ is given in Theorem 1.

Theorem 2 justifies the Monte Carlo procedure described in Section 2.2, showing that it yields asymptotically correct critical values of the tests. Furthermore, under mild conditions the tests are consistent against their respective alternatives, as stated in Theorem 3.

Theorem 3. Suppose the conditions of Theorem 1 hold.

(a) If there exist (t_0, v_1, v_2), $0 < t_0 < \tau$, $(v_1, v_2) \in [0,1]$, such that H_1 holds with strict inequality, $H(t)/S_T(t)$ is decreasing and $H(t_0) > 0$, then the test based on U_1 is consistent against H_1.

(b) If there exist (t_0, v_1, v_2), $0 < t_0 < \tau$, $(v_1, v_2) \in [0,1]$, such that H_2 holds with strict inequality, $H(t)$ and $\int_0^v \lambda(t,u)\,du$ are continuous in t in a neighborhood of t_0, and $H(t_0) > 0$, then the test based on U_2 is consistent against H_2.

(c) If $\lambda(t,v)$ and $\lambda(t)$ are continuous on $[0,\tau] \times [0,1]$ and $[0,\tau]$, respectively, and $H(t) \ge c > 0$ on $[0,\tau]$, then the test based on U_3 is consistent against H_3.

We now show that the test statistics are asymptotically distribution-free when given the weight process $H_n(t) = \hat{S}_C(t-)\hat{S}_T^{1/2}(t)$ introduced in Section 2.3. Let $\tilde{U}_1, \tilde{U}_2$, and $\tilde{U}_3$ denote the test statistics with this weight process. With $H(t) = S_C(t-)S_T^{1/2}(t)$, simple calculation shows that $\int_0^t H(s)^2/y(s)^2\,\gamma(ds,1) = \int_0^t f_T(s)\,ds = F_T(t)$. If $\tilde{\tau} = \sup\{t\colon S_T(t) > 0\}$, then, from the previous discussion and by the continuous mapping theorem, under H_0,

$$\begin{aligned}
\tilde{U}_1 &\equiv \sup_{v_1<v_2}\sup_{0\le t<\tilde{\tau}} \Delta(t,v_1,v_2) \xrightarrow{\mathcal{D}} \sup_{v_1<v_2}\sup_{0\le t<1} \Delta K(t,v_1,v_2)\\
\tilde{U}_2 &\equiv \sup_{v_1<v_2}\sup_{0\le s\le t<\tilde{\tau}} (\Delta(t,v_1,v_2) - \Delta(s,v_1,v_2))\\
&\xrightarrow{\mathcal{D}} \sup_{v_1<v_2}\sup_{0\le s\le t<1} (\Delta K(t,v_1,v_2) - \Delta K(s,v_1,v_2))\\
\tilde{U}_3 &\equiv \sup_{v_1<v_2}\sup_{0\le t<\tilde{\tau}} |L_n(t,v_2) - L_n(t,v_1)| \xrightarrow{\mathcal{D}} \sup_{v_1<v_2}\sup_{0\le t<1} |K(t,v_2) - K(t,v_1)|,
\end{aligned}$$

where $\Delta(t,v_1,v_2)$ is defined following (7), $\Delta K(t,v_1,v_2) = K(t,v_1) + K(t,v_2) - 2K(t,(v_1+v_2)/2)$, and $K(t,v)$ is a Kiefer process with $\mathrm{Cov}(K(s,u),K(t,v)) = (s\wedge t)(u\wedge v - uv)$. Therefore, the $\tilde{U}_j$ are asymptotically distribution-free test statistics.

Remark. In some applications, it is of interest to evaluate whether the instantaneous or absolute risk of failure depends on a continuous mark variable in a given time interval, say $[t_1,t_2)$, rather than over the entire time range $[0,\tau)$. The null and alternative hypotheses, and the test statistics U_1, U_2, and U_3, can be modified straightforwardly to address this problem. All of the results given in this section carry over to this case, by replacing $[0,\tau)$ everywhere with $[t_1,t_2)$. In addition, the results continue to hold if the time range $[0,\tau)$ is replaced with the possibly larger range $[0,\tilde{\tau})$; see the remark following the proof of Theorem 2 in the Appendix.

4. SIMULATION RESULTS

We describe results of a simulation study of the test statistics $\tilde{U}_1, \tilde{U}_2$, and $\tilde{U}_3$.

First we consider a case with T and V independent. The cumulative incidence function is then $F(t,v) = P\{T \le t\} f_V(v)$, where f_V is the density of V. We specify T to be exponential with mean 1, and $f_V(v) = (1/\beta)v^{(1/\beta)-1}$ for $0 \le v \le 1$. Here $\beta = 1.0$ corresponds to the null hypothesis H_0 and $\beta = 0.75, 0.5, 0.25$ correspond to three different alternative hypotheses under the monotone alternatives H_1 and H_2. The extent of departure from the null hypothesis increases as β decreases. We also consider a two-sided alternative with $f_V(v) = 12(v-0.5)^2$, $0 \le v \le 1$ (results in this case are given under the heading "two-sided" in Tables 1 and 2). Next, we consider a case with T and V dependent. For the monotone alternatives H_1 and H_2, we use $F(t,v) = P\{T \le t | V = v\} f_V(v) = (1 - \exp(-t/(v+1))) f_V(v)$, with $f_V(v) = (1/\beta)v^{(1/\beta)-1}$ for

$0 \leq v \leq 1$ and $\beta = 0.5$ and 0.25. For a two-sided alternative, we select V from uniform $(0, 1)$ and $F(t, v) = 1 - \exp(-v^4 t)$.

We choose $n = 50, 100$ and use a 30% censoring rate for the failure times. The sizes and powers of the tests are calculated based on 1000 samples. The nominal level is set at 0.05 in each case. The critical level for each test is calculated using 1000 independent replicates of $\{V_1^*, \ldots, V_n^*\}$. The results in Table 1 indicate that the proposed tests perform well at moderate sample sizes. The estimated sizes are all within 2.0% of the nominal 5.0% (range: 3.9% to 7.0%), and the estimated powers are high for detecting $\beta = 0.25$ when $n = 50$ (range: 81.3% to 100.0%) and for detecting $\beta = 0.50$ when $n = 100$ (range: 72.0% to 86.6%).

To this point, we have assumed that T and V are jointly continuous. In many applications, however, some ties may be present in the data. To study the sensitivity of the tests to the presence of ties, we use the same simulated data that yielded Table 1, and group the failure times into 25 tied values $x_m = 0.05 + 0.1(m - 1)$ for $0.1(m - 1) < x \leq 0.1m$, $m = 1, \ldots, 24$ and $x_{25} = 2.45$ for $x > 2.4$ (or $x_m = 3 + 6(m - 1), m = 1, \ldots, 24, x_{25} = 147$ for the two-sided alternative model under dependent T and V), and group the failure marks into 20 tied values $v_m = 0.025 + 0.05(m - 1)$ for $0.05(m - 1) < v \leq 0.05m$, $m = 1, \ldots, 20$. The sizes and powers of the tests for the grouped data at a 30% censoring rate and the 5.0% nominal level are given in Table 2. We note that the presence of ties causes only a slight decrease in the power of the tests, and the levels become slightly more conservative than in the untied case. For the tests based on $\tilde{U}_1$ and $\tilde{U}_2$, when $n = 100$, the levels are actually closer to nominal in the tied case. The test based on $\tilde{U}_3$ is quite conservative for tied data, though, with estimated size 1.8% when $n = 50$. The test becomes less conservative when the sample size is increased to $n = 100$, with estimated size 2.7%.

Table 1. Observed Levels and Powers (%) of the Test Statistics $\tilde{U}_1, \tilde{U}_2, \tilde{U}_3$ for Testing H_0 versus H_1, H_2, H_3, respectively, at the 5.0% Nominal Level.

		Independent T and V					Dependent T and V		
				β			β		
Size n	Test	1	0.75	0.5	0.25	two-sided	0.5	0.25	two-sided
	$\tilde{U}_1$	5.9	21.1	55.2	88.2	19.2	49.5	85.0	17.9
50	$\tilde{U}_2$	5.5	17.5	48.6	83.1	14.9	46.0	81.3	12.1
	$\tilde{U}_3$	3.9	12.5	60.0	100.0	99.0	55.8	99.9	56.3
	$\tilde{U}_1$	7.0	27.7	78.2	99.2	62.1	80.2	98.9	29.4
100	$\tilde{U}_2$	6.3	23.9	72.0	98.4	54.3	76.0	98.0	19.7
	$\tilde{U}_3$	6.0	20.5	84.5	100.0	100.0	86.6	100.0	85.0

Table 2. Observed Levels and Powers (%) of the Test Statistics $\tilde{U}_1, \tilde{U}_2, \tilde{U}_3$ for Testing H_0 versus H_1, H_2, H_3, respectively, with Tied Data at the 5.0% Nominal Level.

		Independent T and V					Dependent T and V		
				β			β		
Size n	Test	1	0.75	0.5	0.25	two-sided	0.5	0.25	two-sided
	$\tilde{U}_1$	4.5	16.8	50.3	85.7	17.4	40.5	82.1	9.8
50	$\tilde{U}_2$	4.2	13.9	41.7	77.3	11.8	36.2	74.5	3.0
	$\tilde{U}_3$	1.8	8.6	52.4	99.8	96.6	39.4	99.7	42.0
	$\tilde{U}_1$	5.0	24.9	74.0	99.0	55.3	71.1	97.3	25.0
100	$\tilde{U}_2$	4.5	19.6	65.2	97.7	45.7	64.6	95.8	11.0
	$\tilde{U}_3$	2.7	13.5	79.1	100.0	100.0	78.3	100.0	86.3

5. APPLICATION

In 1995 and 1996, the Adult AIDS Clinical Trials Group (AACTG) conducted a randomized trial (Study 241) of 400 HIV infected adults to evaluate two combination antiretroviral treatments by their ability to suppress HIV viral load (D'Aquila et al. 1996). The drug regimens contained zidovudine and didanosine plus either nevirapine or nevirapine placebo. Gilbert et al. (2000) analyzed the data from this trial with the failure time T defined as the time from randomization until plasma HIV levels rose above 1000 copies/ml. The available genotypic data from the study are the amino acids at 19 codons in the reverse transcriptase of HIV isolated from peripheral blood mononuclear cells at baseline and at or after the time of failure from 12 patients on the dual-drug arm and 33 patients on the triple-drug arm who failed. The 19 codons were chosen on the basis of information from published studies that mutations in these positions confer resistance to at least one of the studied drugs (Gunthard et al. 1999; Leigh-Brown et al. 1999; Hanna et al. 2000). For the present analysis, codons with a resistance mutation are coded as ones while codons with nonresistant (whether wildtype or variant) or ambiguous amino acids are coded as zeros.

Let V_b be the mutational distance of a subject's virus sequence measured at baseline relative to the "wildtype" virus with no mutations, defined by

$$V_b = \sum_{i=1}^{19} w_i I(\text{mutation at codon } i) \Big/ \sum_{i=1}^{19} w_i,$$

where the weight w_i measures the amount of resistance conferred by a mutation at the ith position, as measured by a drug resistance assay. Define V_f similarly for a subject's virus sequence measured at or after the time of failure (we refer to this time as the "late week"). Then, we take $V = V_f - \frac{2}{3}V_b$ as the measure of acquired mutational distance during the trial, which emphasizes new mutations more than baseline mutations. The weights $\{w_i\}$ are taken to be those used by Gilbert et al. (2000). Note that V is only defined and measured on subjects who fail treatment, and therefore is appropriately viewed as a mark accompanying failure events rather than as a covariate. In the analysis we consider both treatment arms in a single group. Pooling the arms is meaningful because the accumulated resistance metric V is relevant for both arms, as they share the nucleoside inhibitors zidovudine and didanosine.

PLACE FIGURES 1-3 HERE

As depicted in Figure 1, the mutational distance at baseline V_b ranges between 0.0 and 0.187 in the 45 subjects who failed treatment, and increases to 0.0 to 0.435 by the late week, indicating a trend of increase in mutational distance during the trial. The observed mark variable V takes 28 unique values for the 45 failures, ranging from 0.0 to 0.358, and appears to be approximately uniformly distributed (Figure 1c). A scatterplot of the mark versus failure time does not reveal a systematic pattern (Figure 2). To implement the tests, we first normalize V by its maximum observed value (0.358).

The tests confirm what is suspected from the descriptive plots, yielding nonsignificant results for the three alternative hypotheses, with test statistics $\tilde{U}_1 = 0.628 (p = 0.59), \tilde{U}_2 = 0.487 (p = 0.64)$, and $\tilde{U}_3 = 0.353 (p = 0.76)$. We next implement the graphical procedure, which uses a unit weight process $H_n(t) = 1$ in the test process $L_n(t, v)$ (as in (9)). When comparing the surface $L_n(t, v)$ to eight simulated surfaces $L_n^*(t, v)$ (Figure 3), they appear similar except that $L_n(t, 0)$ rises above zero for increasing t while the processes $L_n^*(t, 0)$ tend to remain closer to zero. Other than this caveat, which can be explained by the fact that four trial participants had a tied mark value $V = 0$, the graphical comparison suggests that the observed test process does not behave unusually compared to the behavior expected under the null hypothesis. We conclude that there is no evidence that the instantaneous or absolute risk of virological failure depends on the level of the resistance mutational distance variable V as defined above. Thus, V may not be useful as a marker of drug resistance. It would be of interest to apply the testing procedure for several other metrics V, as an exploratory search for marks that indicate drug resistance.

6. CONCLUDING REMARKS

The problem addressed here, evaluating whether there is a significant association between the instantaneous or absolute risk of failure and a continuous mark variable observed only at uncensored failure times, has broad application. The two cited applications in AIDS research, in which a time to disease or infection is measured and the mark describes a feature of the agent that causes or is associated with the failure event, arises in many biomedical applications. For one non-AIDS example, in studies evaluating survival of cancer patients, tumor mass might be measured in patients at baseline and at the time of death, and the tests can be used to evaluate a possible link between the growth rate of the tumor and the risk of death. In addition to many other biomedical applications, including the aforementioned problems of assessing the relationship between the risk of death and a quality of life score or a lifetime medical cost, there are clearly a broad variety of applications in other scientific fields. Advantages of the tests developed here for addressing these problems include that they are based on a continuous generalization of the widely-applied and well-understood discrete cause-specific Nelson–Aalen estimator, and they are nonparametric and asymptotically distribution-free.

The authors thank the AACTG Study 241 Virology Team for providing the data for the example, namely George Hanna, Victor DeGruttola, Andrew Leigh Brown, Daniel Kuritzkes, Victoria Johnson, Douglas Richman, and Richard D'Aquila.

APPENDIX: PROOFS OF THEOREMS

The following two lemmas are needed in the proofs of the main results.

Lemma 1. Assume that C is independent of (T, V). Then

$$\int_0^t \frac{\gamma(ds, v)}{y(s)} = \int_0^t \int_0^v \lambda(s, u)\, du\, ds \tag{A.1}$$

$$\int_0^t \frac{\gamma(ds, 1)}{y(s)} = \int_0^t \lambda(s)\, ds. \tag{A.2}$$

Proof of Lemma 1.

Let $F_C(t)$ be the distribution function of the censoring variable C. Recalling the notation $f(t, v)$ for the joint density of (T, V), we have

$$\begin{aligned}
\gamma(t, v) &= P(X \le t, \delta = 1, V \le v) \\
&= P(T \le t, T \le C, V \le v) \\
&= \int_0^\infty P(T \le t, T \le s, V \le v)\, dF_C(s) \\
&= \int_0^\infty \left(\int_0^{s \wedge t} \int_0^v f(r, u)\, du\, dr \right) dF_C(s) \\
&= \int_0^t \int_r^\infty \int_0^v f(r, u)\, du\, dF_C(s)\, dr \\
&= \int_0^t P(C \ge r) \int_0^v f(r, u)\, du\, dr.
\end{aligned}$$

It follows that

$$\int_0^t \frac{\gamma(ds, v)}{y(s)} = \int_0^t \left(\int_0^v f(s, u)\, du \Big/ P(T \ge s) \right) ds = \int_0^t \int_0^v \lambda(s, u)\, du\, ds.$$

This proves (A.1). The result (A.2) follows by letting $v = 1$ and using $\int_0^1 f(s, u)\, du = f_T(s)$.

The following lemma of Bilias, Gu, and Ying (1997) is stated here for convenience.

Lemma 2. Let $D = [a, b] \times [c, d] \subset \mathbb{R}^2$. Suppose that

$$\lim_{n \to \infty} \sup_{(t,s) \in D} \{ |h_n(t, s) - h(t, s)| + |J_n(t, s) - \tilde{J}_n(t, s)| \} = 0,$$

where h_n and h are continuous functions on D, and, for each fixed t, $J_n(t, \cdot)$ and $\tilde{J}_n(t, \cdot)$ are left continuous functions, with their total variations bounded by a constant, independent of n and t. Then

$$\lim_{n \to \infty} \sup_{(t,s) \in D} \left| \int_c^s h_n(t, u) J_n(t, du) - \int_c^s h(t, u) \tilde{J}_n(t, du) \right| = 0,$$

$$\lim_{n \to \infty} \sup_{(t,s) \in D} \left| \int_c^s h_n(t, u) J_n(t, du) - \int_c^s h_n(t, u) \tilde{J}_n(t, du) \right| = 0.$$

Proposition 1 describes the limiting distribution of the test process $L_n(t,v)$ when the weight function is unity.

Proposition 1. For $\tau < \tilde{\tau}$,

$$\sqrt{n}(\hat{\Lambda}(t,v) - \bar{\Lambda}(t,v)) - \sqrt{n}\left(\Lambda(t,v) - v\int_0^t \lambda(s)\,ds\right) \tag{A.3}$$
$$\xrightarrow{\mathcal{D}} \int_0^t y(s)^{-1}(G_1(ds,v) - vG_1(ds,1)) - \int_0^t G_2(s)y(s)^2(\gamma(ds,v) - v\gamma(ds,1))$$

in $D([0,\tau]\times[0,1])$.

Proof of Proposition 1.

By the empirical central limit theorem

$$\sqrt{n}(N(t,v)/n - \gamma(t,v), Y(t)/n - y(t)) \xrightarrow{\mathcal{D}} (G_1(t,v), G_2(t)) \tag{A.4}$$

in $D([0,\tau]\times[0,1])\times D[0,\tau]$, where $G_1(t,v)$ and $G_2(t)$ are continuous mean zero Gaussian processes with covariances

$$\begin{aligned}
&\mathrm{Cov}(G_1(s,u), G_1(t,v))\\
&= E\{(I(X\le s,\delta=1,V\le u) - \gamma(s,u))(I(X\le t,\delta=1,V\le v) - \gamma(t,v))\}\\
&= E\{I(X\le s,\delta=1,V\le u)I(X\le t,\delta=1,V\le v)\} - \gamma(s,u)\gamma(t,v)\\
&= \gamma(s\wedge t, u\wedge v) - \gamma(s,u)\gamma(t,v),\\
&\mathrm{Cov}(G_2(s), G_2(t)) = E\{(I(X\ge s) - y(s))(I(X\ge t) - y(t))\}\\
&= y(s\vee t) - y(s)y(t),\\
&\mathrm{Cov}(G_1(s,u), G_2(t)) = E\{(I(X\le s,\delta=1,V\le u) - \gamma(s,u))(I(X\ge t) - y(t))\}\\
&= E\{I(X\le s,\delta=1,V\le u)I(X\ge t)\} - \gamma(s,u)y(t)\\
&= E\{I(X\le s,\delta=1,V\le u) - I(X<t,\delta=1,V\le u)\}I(t\le s) - \gamma(s,u)y(t)\\
&= (\gamma(s,u) - \gamma(t-,u))I(t\le s) - \gamma(s,u)y(t).
\end{aligned}$$

Let $D = D([0,\tau]\times[0,1])\times D[0,\tau]$ be the product space. The Nelson–Aalen-type estimator $\hat{\Lambda}(t,v) = \int_0^t N(ds,v)/Y(s)$, $0\le t\le\tau$, depends on the pair $(n^{-1}N(t,v), n^{-1}Y(t))$ through the following map from the domain of the type $D_\phi = \{(A(t,v),B(t))\colon \int |A(dt,v)| \le M, B(t)\ge\epsilon\} \subset D$ for given M and $\epsilon > 0$:

$$\phi\colon\ (A(t,v),B(t)) \longrightarrow \int_0^t \frac{1}{B(s)}A(ds,v).$$

Let $D_0 = C([0,\tau] \times [0,1]) \times C[0,\tau]$. First, we show that the map ϕ is Hadamard-differentiable tangentially to the set D_0 at every point (A, B) such that $1/B(t)$ is of bounded variation. Let $t_n \to 0$ be any converging sequences and let $(\alpha_n, \beta_n) \to (\alpha, \beta) \in D_0$ such that $(A+t_n\alpha_n, B+t_n\beta_n) \in D_\phi$. Then

$$
\begin{aligned}
& t_n^{-1}(\phi(A + t_n\alpha_n, B + t_n\beta_n)(t, v) - \phi(A, B)(t, v)) \\
&= t_n^{-1}\left(\int_0^t \frac{1}{B(s) + t_n\beta_n(s)} (A(ds, v) + t_n\alpha_n(ds, v)) - \int_0^t \frac{1}{B(s)} A(ds, v)\right) \\
&= \int_0^t \frac{1}{B(s) + t_n\beta_n(s)} \alpha_n(ds, v) + t_n^{-1} \int_0^t \left(\frac{1}{B(s) + t_n\beta_n(s)} - \frac{1}{B(s)}\right) A(ds, v) \\
&= \int_0^t \frac{1}{B(s) + t_n\beta_n(s)} \alpha_n(ds, v) - \int_0^t \frac{\beta_n(s)}{B(s)(B(s) + t_n\beta_n(s))} A(ds, v) \\
&= \int_0^t \left(\frac{1}{B(s) + t_n\beta_n(s)} - \frac{1}{B(s)}\right) \alpha_n(ds, v) + \int_0^t \frac{1}{B(s)} \alpha_n(ds, v) \qquad \text{(A.5)} \\
&\quad - \int_0^t \frac{\beta_n(s)}{B(s)(B(s) + t_n\beta_n(s))} A(ds, v) \\
&\to \int_0^t \frac{1}{B(s)} \alpha(ds, v) - \int_0^t \frac{\beta(s)}{B(s)^2} A(ds, v), \quad \text{uniformly in } (t, v) \\
&= \phi'_{A,B}(\alpha, \beta)(t, v),
\end{aligned}
$$

where the limit is obtained by applying Lemma 2 to each of the three terms involved since the conditions on the total variation are satisfied for $\alpha_n(\cdot, v)$, $1/B(\cdot)$ and $A(\cdot, v)$, respectively. Let

$$Z_n(t, v) = \sqrt{n}(\hat{\Lambda}(t, v) - \Lambda(t, v)). \qquad \text{(A.6)}$$

Since the pair $(n^{-1}N(t, v), n^{-1}Y(t))$, $(t, v) \in [0, \tau] \times [0, 1]$, is contained in the domain D_ϕ with probability tending to 1 for $M \geq 1$ and sufficiently small $\epsilon > 0$, applying the functional delta method theorem (van der Vaart and Wellner, 1996, p. 374), (A.1) and (A.5), we have

$$Z_n(t, v) \xrightarrow{\mathcal{D}} Z(t, v) \ , \qquad \text{(A.7)}$$

where

$$Z(t, v) = \int_0^t \frac{1}{y(s)} G_1(ds, v) - \int_0^t \frac{G_2(s)}{y(s)^2} \gamma(ds, v). \qquad \text{(A.8)}$$

Now, consider the following continuous map from $D([0, \tau] \times [0, 1])$ into itself,

$$\psi_1\text{:} \ \ \psi_1(g)(t, v) = g(t, v) - vg(t, 1), \ \ (t, v) \in [0, \tau] \times [0, 1]. \qquad \text{(A.9)}$$

Applying the continuous mapping theorem, we get

$$Z_n(t, v) - vZ_n(t, 1) \xrightarrow{\mathcal{D}} Z(t, v) - vZ(t, 1), \ \ . \qquad \text{(A.10)}$$

Proposition 1 follows by plugging the specific forms (A.6) and (A.8) of the processes Z_n and Z into (A.10) .

Proposition 2 extends the result of Proposition 1 to general weight processes.

Proposition 2. Given the conditions expressed in Theorem 1,

$$L_n(t,v) - \sqrt{n}\left(\int_0^t \int_0^v H_n(s)\lambda(s,u)\,du\,ds - v\int_0^t H_n(s)\lambda(s)\,ds\right) \tag{A.11}$$
$$\xrightarrow{\mathcal{D}} \int_0^t H(s)y(s)^{-1}(G_1(ds,v) - vG_1(ds,1)) - \int_0^t H(s)G_2(s)y(s)^2(\gamma(ds,v) - v\gamma(ds,1))$$

in $D([0,\tau]\times[0,1])$.

Proof of Proposition 2.

By the almost sure representations theorem (Shorack and Wellner 1986, p. 47), there exist $\tilde{N}(t,v)$, $\tilde{Y}(t)$, $\tilde{G}_1(t,v)$ and $\tilde{G}_2(t)$ on some probability space such that $\tilde{N}(t,v)$, $\tilde{Y}(t)$, $\tilde{G}_1(t,v)$ and $\tilde{G}_2(t)$ are equal in law to $N(t,v)$, $Y(t)$, $G_1(t,v)$ and $G_2(t)$, respectively, and (A.4) holds almost surely uniformly in (t,v). Furthermore $\tilde{N}(t,v)$, $\tilde{Y}(t)$, $\tilde{G}_1(t,v)$ and $\tilde{G}_2(t)$ can be chosen to have the same sample paths as the original processes. Let $\tilde{Z}_n$ and $\tilde{Z}$ be the corresponding representations of Z_n and Z defined in (A.6) and (A.8), respectively. Repeating the steps of (A.5) with $t_n = n^{-1/2}$, $\alpha_n(t,v) = \sqrt{n}(N(t,v)/n - \gamma(t,v))$ and $\beta_n(t) = \sqrt{n}(Y(t)/n - y(t))$ and applying Lemma 2, we have

$$\tilde{Z}_n(t,v)\xrightarrow{\text{a.s.}}\tilde{Z}(t,v), \quad \text{uniformly in } (t,v)\ .$$

Consequently,

$$\tilde{Z}_n(t,v) - v\tilde{Z}_n(t,1)\xrightarrow{\text{a.s.}}\tilde{Z}(t,v) - v\tilde{Z}(t,1).$$

Let $\tilde{H}_n(t)$ be the process $H_n(t)$ redefined in terms of $\tilde{N}(t,v)$ and $\tilde{Y}(t)$. Then $\tilde{H}_n(t)$ has the properties of $H_n(t)$ assumed in the theorem. Applying Lemma 2, we have

$$\int_0^t \tilde{H}_n(s)(\tilde{Z}_n(ds,v) - v\tilde{Z}_n(ds,1))\xrightarrow{\text{a.s.}}\int_0^t H(s)(\tilde{Z}(ds,v) - v\tilde{Z}(ds,1)).$$

Hence

$$\int_0^t H_n(s)(Z_n(ds,v) - vZ_n(ds,1))\xrightarrow{\mathcal{D}}\int_0^t H(s)(Z(ds,v) - vZ(ds,1)), \tag{A.12}$$

in $D([0,\tau]\times[0,1])$. Proposition 2 follows by plugging the specific forms (A.6) and (A.8) of the processes Z_n and Z into (A.12) and by (4).

Proof of Theorem 1.

Under H_0, the failure time T and the failure mark V are independent and V is uniformly distributed on $[0,1]$. Consequently, $\lambda(t,v) = \lambda(t)$ for all (t,v). Further, since the censoring variable C is independent of (T,V), V is independent of (T,C) under H_0. Hence, $\gamma(t,v) = v\gamma(t,1)$ under H_0. The result (10) follows by applying Proposition 2.

Let $N_i(t) = I(X_i \le t, \delta_i = 1)$. Since

$$\sqrt{n}(N(t,v)/n - \gamma(t,v)) = n^{-1/2}\sum_{i=1}^{n}(N_i(t)I(V_i \le v) - \gamma(t,v)) \xrightarrow{\mathcal{D}} G_1(t,v)$$

in $D([0,\tau]\times[0,1])$, under H_0 we have

$$n^{-1/2}\sum_{i=1}^{n} N_i(t)(I(V_i \le v) - v)) \xrightarrow{\mathcal{D}} G_1(t,v) - vG_1(t,1)$$

in $D([0,\tau]\times[0,1])$ through the continuous map ψ_1 given in (A.9). Since $H(t)$ and $1/y(t)$ are of bounded variation over $t \in [0,\tau]$, the map ψ_2: $g(t,v) \to \int_0^t y(s)^{-1}\, g(ds,v), (t,v) \in [0,\tau]\times[0,1]$ from $D([0,\tau]\times[0,1])$ into itself is continuous on $C([0,\tau]\times[0,1])$. Hence, by the continuous mapping theorem,

$$n^{-1/2}\sum_{i=1}^{n}(I(V_i \le v) - v)) \int_0^t \frac{H(s)}{y(s)}\, N_i(ds) \xrightarrow{\mathcal{D}} \int_0^t \frac{H(s)}{y(s)}(G_1(ds,v) - vG_1(ds,1)) = L(t,v).$$

The covariance of $L(t,v)$ is given by

$$\begin{aligned}
&\mathrm{Cov}(L(s,u), L(t,v)) \\
&= E\left\{(I(V_i \le u) - u))(I(V_i \le v) - v)) \int_0^s \frac{H(r)}{y(r)}\, N_i(dr) \int_0^t \frac{H(r)}{y(r)}\, N_i(dr)\right\} \\
&= E\{(I(V_i \le u) - u))(I(V_i \le v) - v))\}E\left\{ \int_0^s \frac{H(r)}{y(r)}\, N_i(dr) \int_0^t \frac{H(r)}{y(r)}\, N_i(dr)\right\} \\
&= (u \wedge v - uv) \int_0^{s\wedge t} \frac{H^2(r)}{y(r)^2}\, \gamma(dr,1),
\end{aligned}$$

where the second equality is obtained by the independence between V and (X,δ).

Proof of Theorem 2.

Let $L_1^*(t,v) = \sqrt{n}(\hat{\Lambda}^*(t,v) - v\hat{\Lambda}^*(t,1))$. Then $L_n^*(t,v) = \int_0^t H_n(s)\, L_1^*(ds,v)$. Let $L_1(t,v) = \int_0^t y(s)^{-1}(G_1(ds,v) - vG_1(ds,1))$. It is sufficient to show $L_1^* \xrightarrow{\mathcal{D}} L_1$ conditionally in $D([0,\tau]\times[0,1])$.

Let $N_i(t) = I(X_i \le t, \delta_i = 1)$ and $N(t) = \sum_{i=1}^n N_i(t)$. Then

$$L_1^*(t,v) = \sqrt{n}\sum_{i=1}^{n}(I(V_i^* \le v) - v) \int_0^t \frac{dN_i(s)}{Y(s)}. \tag{A.13}$$

To establish the conditional weak convergence of $L_1^*(t,v)$, we shall show that the finite distributions of $L_1^*(t,v)$ converge weakly to those of $L_1(t,v)$ given the data sequence, and that $L_1^*(t,v)$ is asymptotically tight given the data sequence; see van der Vaart and Wellner (1996, p. 183).

By the central limit theorem, the finite dimensional distributions of $L_1^*(t,v)$ given the data converge weakly to Gaussian distributions with covariance

$$\begin{aligned}
&\lim_{n\to\infty} E\{L_1^*(s,u)L_1^*(t,v)|\text{data}\}\\
&= \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\Big\{(I(V_i^* \le u)-u))(I(V_i^*\le v)-v))\\
&\quad \int_0^s \frac{1}{Y(r)/n} N_i(dr)\int_0^t \frac{1}{Y(r)/n} N_i(dr)\Big|\text{data}\Big\}\\
&= \lim_{n\to\infty}(u\wedge v-uv)n^{-1}\sum_{i=1}^{n}\int_0^s \frac{1}{Y(r)/n} N_i(dr)\int_0^t \frac{1}{Y(r)/n} N_i(dr)\\
&= \lim_{n\to\infty}(u\wedge v-uv)n^{-1}\int_0^{s\wedge t}\frac{1}{(Y(r)/n)^2} N(dr)\\
&= (u\wedge v-uv)\int_0^{s\wedge t}\frac{\gamma(dr,1)}{y(r)^2} = \text{Cov}(L_1(s,u),L_1(t,v)).
\end{aligned}$$

Now, it is left to prove that $L_1^*(t,v)$ is asymptotically tight given the data sequence. We shall apply the tightness criteria of Bickel and Wichura (1971, eq. 3, p. 1658) based on neighboring blocks.

Let $B=[t_1,t_2)\times[v_1,v_2)$, $C_1=[t_1,t_2)\times[v_0,v_1)$ and $C_2=[t_2,t_3)\times[v_1,v_2)$ be the blocks in $[0,\tau]\times[0,1]$. The pairs B,C_1 and B,C_2 are both neighboring blocks. They exhaust all the possible positions of two neighboring blocks. We have

$$L_1^*(B)=\sqrt{n}\sum_{i=1}^{n}(I(v_1<V_i^*\le v_2)-(v_2-v_1))\frac{\delta_i I(t_1<X_i\le t_2)}{Y(X_i)}$$

and

$$\begin{aligned}
&E\{(L_1^*(B))^2(L_1^*(C_1))^2|\text{data}\}\\
&= n^2\sum_{i,j=1}^{n}\sum_{k,l=1}^{n}\frac{\delta_i\delta_j\delta_k\delta_l I(t_1<X_i\le t_2)I(t_1<X_j\le t_2)I(t_1<X_k\le t_2)I(t_1<X_l\le t_2)}{Y(X_i)Y(X_j)Y(X_k)Y(X_l)}\\
&\quad *E\{[I(v_1<V_i^*\le v_2)-(v_2-v_1)][I(v_1<V_j^*\le v_2)-(v_2-v_1)]\\
&\qquad *[I(v_0<V_k^*\le v_1)-(v_1-v_0)][I(v_0<V_l^*\le v_1)-(v_1-v_0)]\}.
\end{aligned}$$

Simple calculations show that the expression for the expectation is equal to

$$(v_2-v_1)(1-(v_2-v_1))(v_1-v_0)(1-(v_1-v_0)),\quad \text{for } i=j,k=l,i\ne k;$$

$$
\begin{aligned}
&(v_2-v_1)^2(v_1-v_0)^2, \qquad\qquad \text{for } i=k, j=l, i\neq j, \text{ or } i=l, j=k, i\neq j;\\
&(v_2-v_1)(v_1-v_0)^2(1-2(v_2-v_1))+(v_1-v_0)(v_2-v_1)^2(1-2(v_1-v_0))\\
&+(v_2-v_1)^2(v_1-v_0)^2 \qquad\qquad\qquad\qquad\qquad \text{for } i=j=k=l.
\end{aligned}
$$

These expectations are bounded by $(v_2-v_1)(v_1-v_0)$. The expectations are zero for all other combinations of i, j, k, l. Hence

$$
\begin{aligned}
&E\{(L_1^*(B))^2(L_1^*(C_1))^2|\text{data}\}\\
&\leq (v_2-v_1)(v_1-v_0)n^2\left[3\left(\sum_{i=1}^n \frac{\delta_i I(t_1<X_i\leq t_2)}{Y(X_i)^2}\right)^2+\sum_{i=1}^n \frac{\delta_i I(t_1<X_i\leq t_2)}{Y(X_i)^4}\right]\\
&=(v_2-v_1)(v_1-v_0)n^2\left[3\left(\int_{t_1}^{t_2}\frac{dN(s)}{Y(s)^2}\right)^2+\int_{t_1}^{t_2}\frac{dN(s)}{Y(s)^4}\right]\\
&\xrightarrow{\text{a.s.}} 3(v_2-v_1)(v_1-v_0)\left(\int_{t_1}^{t_2}\frac{\gamma(ds,1)}{y(s)^2}\right)^2.
\end{aligned} \tag{A.14}
$$

$$
\begin{aligned}
&E\{(L_1^*(B))^2(L_1^*(C_2))^2|\text{data}\}\\
&=n^2\sum_{i,j=1}^n\sum_{k,l=1}^n \frac{\delta_i\delta_j\delta_k\delta_l I(t_1<X_i\leq t_2)I(t_1<X_j\leq t_2)I(t_2<X_k\leq t_3)I(t_2<X_l\leq t_3)}{Y(X_i)Y(X_j)Y(X_k)Y(X_l)}\\
&\quad *E\{[I(v_1<V_i^*\leq v_2)-(v_2-v_1)][I(v_1<V_j^*\leq v_2)-(v_2-v_1)]\\
&\qquad *[I(v_1<V_k^*\leq v_2)-(v_2-v_1)][I(v_1<V_l^*\leq v_2)-(v_2-v_1)]\}\\
&=(v_2-v_1)^2(1-(v_2-v_1))^2n^2\sum_{i,k=1,i\neq k}^n \frac{\delta_i\delta_k I(t_1<X_i\leq t_2)I(t_2<X_k\leq t_3)}{Y(X_i)^2Y(X_k)^2}\\
&\leq (v_2-v_1)^2n^2\int_{t_1}^{t_2}\frac{dN(s)}{Y(s)^2}\int_{t_2}^{t_3}\frac{dN(s)}{Y(s)^2}\\
&\xrightarrow{\text{a.s.}}(v_2-v_1)^2\int_{t_1}^{t_2}\frac{\gamma(ds,1)}{y(s)^2}\int_{t_2}^{t_3}\frac{\gamma(ds,1)}{y(s)^2}.
\end{aligned} \tag{A.15}
$$

Let μ be the finite Borel measure on $[0,\tau]\times[0,1]$ generated by $\mu(B)=2(v_2-v_1)\int_{t_1}^{t_2}\gamma(ds,1)/y(s)^2$. The asymptotic tightness of $L_1^*(t,v)$ follows from (A.14) and (A.15); see Bickel and Wichura (1971).

Remark. It is clear from the proof of Theorem 2 that the process L_n^* is tight over $[0,\tilde\tau)\times[0,1]$ if $\int_0^{\tilde\tau} H^2(s)/y^2(s)\,\gamma(ds,1)<\infty$. By Theorem 1 of Bickel and Wichura (1971), with $T=[\tau,\tilde\tau)\times[0,1]$, this implies that, for every $\epsilon>0$,

$$
\lim_{\tau\to\tilde\tau}\limsup_{n\to\infty} P\big(\sup_{0\leq v\leq 1}\sup_{\tau\leq t<\tilde\tau}|L_n^*(t,v)-L_n^*(\tau,v)|>\epsilon|\text{data}\big)=0. \tag{A.16}
$$

By Theorem 2 and applying Theorem 4.2 of Billingsley (1968), we have $L_n^*(t,v) \xrightarrow{\mathcal{D}} L(t,v)$ conditionally in $D([0,\tilde{\tau}] \times [0,1])$.

By going through the proof of Theorem 2, replacing the conditional expectation by the unconditional expectation and L_1^* by L_n, we have that under H_0, the process L_n is tight over $[0,\tilde{\tau}) \times [0,1]$ if $\int_0^{\tilde{\tau}} H(s)^2/y(s)^2\, \gamma(ds,1) < \infty$. The weak convergence of L_n to L in $D([0,\tilde{\tau}] \times [0,1])$ under H_0 then follows from Theorem 1.

Proof of Theorem 3.

(a) Let

$$m(t,v) = \int_0^t \int_0^v H_n(s)\lambda(s,u)\,du\,ds - v\int_0^t H_n(s)\lambda(s)\,ds.$$

Denote the left and the right side of (A.11) by $L_n^a(t,v)$ and $L^a(t,v)$, respectively. Then

$$\begin{aligned} U_1 &= \sup_{v_1<v_2} \sup_{0\le t<\tau} \Delta(t,v_1,v_2) \\ &= \sup_{v_1<v_2} \sup_{0\le t<\tau} [(L_n^a(t,v_1) + L_n^a(t,v_2) - 2L_n^a(t,(v_1+v_2)/2)) \\ &\quad +\sqrt{n}(m(t,v_1) + m(t,v_2) - 2m(t,(v_1+v_2)/2))] \\ &\ge \sqrt{n} \sup_{v_1<v_2} \sup_{0\le t<\tau} ((m(t,v_1) + m(t,v_2) - 2m(t,(v_1+v_2)/2))) \\ &\quad - \sup_{v_1<v_2} \sup_{0\le t<\tau} [-(L_n^a(t,v_1) + L_n^a(t,v_2) - 2L_n^a(t,(v_1+v_2)/2))] \end{aligned} \tag{A.17}$$

Note that

$$\begin{aligned} & m(t,v_1) + m(t,v_2) - 2m(t,(v_1+v_2)/2) \\ &= \int_0^t H_n(s)\Big(\int_0^{v_1} \lambda(s,u)\,du + \int_0^{v_2} \lambda(s,u)\,du - 2\int_0^{(v_1+v_2)/2} \lambda(s,u)\,du\Big)\,ds \\ &\xrightarrow{\text{a.s.}} \int_0^t H(s)\Big(\int_0^{v_1} \lambda(s,u)\,du + \int_0^{v_2} \lambda(s,u)\,du - 2\int_0^{(v_1+v_2)/2} \lambda(s,u)\,du\Big)\,ds \\ &= \int_0^t [H(s)/S_T(s)]\,d\Big(\int_0^{v_1} F(s,u)\,du + \int_0^{v_2} F(s,u)\,du - 2\int_0^{(v_1+v_2)/2} F(s,u)\,du\Big) \\ &= [H(t)/S_T(t)]\Big(\int_0^{v_1} F(t,u)\,du + \int_0^{v_2} F(t,u)\,du - 2\int_0^{(v_1+v_2)/2} F(t,u)\,du\Big) \\ &\quad - \int_0^t \Big(\int_0^{v_1} F(s,u)\,du + \int_0^{v_2} F(s,u)\,du - 2\int_0^{(v_1+v_2)/2} F(s,u)\,du\Big)\,d(H(s)/S_T(s)) \\ &\ge [H(t)/S_T(t)]\Big(\int_0^{v_1} F(t,u)\,du + \int_0^{v_2} F(t,u)\,du - 2\int_0^{(v_1+v_2)/2} F(t,u)\,du\Big), \end{aligned}$$

where the last inequality is obtained since $\int_0^{v_1} F(s,u)\,du + \int_0^{v_2} F(s,u)\,du - 2\int_0^{(v_1+v_2)/2} F(s,u)\,du \ge 0$ under H_1 and $H(s)/S_T(s)$ is decreasing. Under H_1 with the given t_0 such that the inequality of the alternative H_1 holds strictly for some $(v_1, v_2) \in [0,1]$, $\int_0^v F(t_0,u)\,du$ is a strictly concave function. Hence

$$\int_0^{v_1} F(t_0,u)\,du + \int_0^{v_2} F(t_0,u)\,du - 2\int_0^{(v_1+v_2)/2} F(t_0,u)\,du > 0,$$

for some $v_1, v_2 \in [0,1]$. By Proposition 2, the second term of (A.17) converges in distribution to a finite random variable. This yields $U_1 \xrightarrow{P} \infty$. Since U_1^* converges in distribution to a finite random variable, U_1 is consistent against H_1.

(b) Note that, for $s < t$,

$$\begin{aligned}
& m(t,v_1) + m(t,v_2) - 2m(t,(v_1+v_2)/2) - m(s,v_1) + m(s,v_2) - 2m(s,(v_1+v_2)/2) \\
&= \int_s^t H_n(r)\Big(\int_0^{v_1} \lambda(r,u)\,du + \int_0^{v_2} \lambda(r,u)\,du - 2\int_0^{(v_1+v_2)/2} \lambda(r,u)\,du\Big)\,dr \\
&\xrightarrow{\text{a.s.}} \int_s^t H(r)\Big(\int_0^{v_1} \lambda(r,u)\,du + \int_0^{v_2} \lambda(r,u)\,du - 2\int_0^{(v_1+v_2)/2} \lambda(r,u)\,du\Big)\,dr.
\end{aligned}$$

Under H_2, with the given t_0 such that the inequality of the alternative H_2 holds strictly for some $(v_1, v_2) \in [0,1]$, $\int_0^v \lambda(t_0,u)\,du$ is a strictly concave function. Hence

$$\int_0^{v_1} \lambda(t_0,u)\,du + \int_0^{v_2} \lambda(t_0,u)\,du - 2\int_0^{(v_1+v_2)/2} \lambda(t_0,u)\,du > 0,$$

for some $v_1, v_2 \in [0,1]$. Following a similar argument as in part (a) of the proof for the consistency of U_1, applying Proposition 2, and by the continuity assumptions of the theorem, we have $U_2 \xrightarrow{P} \infty$. Since U_2^* converges in distribution to a finite random variable, U_2 is consistent against H_2.

(c) Note that the alternative H_3 that $\lambda(t,v)$ does not depend on v for all $t \in [0,\tau]$ is equivalent to $\int_0^t H(s)\Big[\int_0^v (\lambda(s,u) - \lambda(s))\,du\Big]\,ds$ not depending on v for all $t \in [0,\tau]$. Thus, under H_3 there exists (t_0, v_1, v_2), $t_0 \in [0,\tau]$, $(v_1, v_2) \in [0,1]$ such that

$$m(t_0,v_2) - m(t_0,v_1) \xrightarrow{\text{a.s.}} \int_0^{t_0} H(s)\Big[\int_{v_1}^{v_2} (\lambda(s,u) - \lambda(s))\,du\Big]\,ds \neq 0. \tag{A.18}$$

Since

$$\begin{aligned}
U_3 \;\geq\; & \sqrt{n} \sup_{v_1<v_2} \sup_{0\le t<\tau} |m(t,v_2) - m(t,v_1)| \\
& - \sup_{v_1<v_2} \sup_{0\le t<\tau} |(L_n(t,v_2) - \sqrt{n}m(t,v_2)) - (L_n(t,v_1) - \sqrt{n}m(t,v_1))|,
\end{aligned}$$

it follows by Proposition 2, the continuous mapping theorem and (A.18) that $U_3 \xrightarrow{P} \infty$ under H_3. Since U_3^* converges in distribution to a finite random variable, U_3 is consistent against H_3.

REFERENCES

Aly, E.-E., Kochar, S. C. and McKeague, I. W. (1994), "Some tests for comparing cumulative incidence functions and cause-specific hazard rates," *Journal of the American Statistical Association*, 89, 994–999.

Bickel, P. J. and Wichura, M. J. (1971), "Convergence criteria for multiparameter stochastic processes and some applications," *Annals of Mathematical Statistics*, 42, 1656–1670.

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

Billingsley, P. (1968), *Convergence of Probability Measures*, New York, Wiley.

D'Aquila, R. T., Hughes, M. D., Johnson, V. A., Fischl, M. A., Sommadossi, J.-P., Liou, S., Timpone, J., Myers, M., Basgoz, N., Niu, M., Hirsch, M. S. and the National Institute of Allergy and Infectious Diseases AIDS Clinical Trials Group Protocol 241 Investigators (1996), "Nevirapine, zidovudine, and didanosine compared with zidovudine and didanosine in patients with HIV-1 infection," *Annals of Internal Medicine*, 124, 1019–1031.

Gilbert, P. B., Hanna, G. J., De Gruttola, V., Martinez-Picado, J., Kuritzkes, D. R., Johnson, V. A., Richman, D. D. and D'Aquila, R. T. (2000), "Comparative analysis of HIV type 1 genotypic resistance across antiretroviral trial treatment regimens," *AIDS Research and Human Retroviruses*, 16, 1325–1336.

Gilbert, P. B., DeGruttola, V., Hammer, S. M. and Kuritzkes, D. R. (2001), "Virological and regimen termination surrogate endpoints in AIDS clinical trials," *Journal of the American Medical Association*, in press.

Gilbert, P. B., Self, S. G., Rao, M., Naficy, A. and Clemens, J. D. (2001), "Sieve analysis: Methods for assessing how vaccine efficacy depends on genotypic and phenotypic pathogen variation from vaccine trial data," *Journal of Clinical Epidemiology*, in press.

Graham, B. S. and Karzon, D. T. (1998), "AIDS vaccine development," in *Textbook of AIDS Medicine, Second Edition*, eds. T.C. Merigan, Jr., J.G. Bartlett and D. Bolognesi, Baltimore, Williams and Wilkins, pp. 689–724.

Gunthard, H. F., Leigh-Brown, A. J., D'Aquila, R. T., Johnson, V. A., Kuritzkes, D. R., Richman,

D. D. and Wong, J. K. (1999), "Higher selection pressure from antiretroviral drugs in vivo results in increased evolutionary resistance in HIV-1 pol," *Virology*, 259, 154–165.

Hanna, G. J., Johnson, V. A., Kuritzkes, D. R., Richman, D. D., Leigh-Brown, A. J., Savara, A. V., Hazelwood, J. D. and D'Aquila, R. T. (2000), "Patterns of resistance mutations selected by treatment of human immunodeficiency virus type 1 infection with zidovudine, didanosine, and nevirapine," *The Journal of Infectious Diseases*, 181, 904–911.

HHS Report (1998), "Guidelines for the use of antiretroviral agents in HIV-infected adults and adolescents," Department of Health and Human Services and Henry J. Kaiser Family Foundation, *Morbidity and Mortality Weekly Report*, 47 (RR-5), 43–82.

Hirsch, M. S., Conway, B., D'Aquila, R. T., Johnson, V. A., Brun-Vezinet, F., Clotet, B., Demeter, L. M., Hammer, S. M., Jacobsen, D. M., Kuritzkes, D. R., Loveday, C., Mellors, J. W., Vella, S. and Richman, D. D. for the International AIDS Society USA Panel (1998), "Antiretroviral drug resistance testing in adults with HIV infection," *Journal of the American Medical Association*, 279, 1984–1991.

Hu, X. S. and Tsai, W. Y. (1999), "Linear rank tests for competing risks model," *Statistica Sinica*, 9, 971–983.

Huang, Y. and Louis, T. A. (1998), "Nonparametric estimation of the joint distribution of survival time and mark variables," *Biometrika*, 85, 785–798.

Lam, K. F. (1998), "A class of tests for the equality of k cause-specific hazard rates in a competing risks model," *Biometrika*, 85, 179–188.

Leigh-Brown, A. J., Gunthard, H. G., Wong, J. K., D'Aquila, R. T., Johnson, V. A., Kuritzkes, D. R. and Richman, D. D. (1999), "Sequence clusters in human immunodeficiency virus type 1 reverse transcriptase are associated with subsequent virological response to antiretroviral therapy," *The Journal of Infectious Diseases*, 180, 1043–1049.

Luo, X. and Turnbull, B. (1999), "Comparing two treatments with multiple competing risks endpoints," *Statistica Sinica*, 9, 985–997.

McKeague, I. W., Nikabadze, A. M. and Sun, Y. (1995), "An omnibus test for independence of a survival time from a covariate," *Annals of Statistics*, 23, 450–475.

McKeague, I. W. and Utikal, K. J. (1990), "Identifying nonlinear covariate effects in semimartingale regression models," *Probability Theory of Related Fields*, 87, 1–25.

Olschewski, M. and Schumacher, M. (1990), "Statistical analysis of quality of life in cancer clinical trials," *Statistics in Medicine*, 9, 749–763.

Prentice, R.L., Kalbfleisch, J.D., Peterson, A.V., Flournoy, N., Farewell, V.T. and Breslow, N.E. (1978), "The analysis of failure times in the presence of competing risks," *Biometrics*, 34, 541–554.

Shorack, G. R. and Wellner, J. A. (1986), *Empirical Processes with Applications to Statistics*, New York, Wiley.

Sun, Y. (2001), "Generalized nonparametric test procedures for comparing multiple cause-specific hazard rates," *Journal of Nonparametric Statistics*, 13, 171-207.

Sun, Y. and Tiwari, R. C. (1995), "Comparing cause-specific hazard rates of a competing risks model with censored data," *IMS Lecture Notes-Monograph Series*, 27, 255–270. *Analysis of Censored Data*, edited by H. L. Koul and J. V. Deshpandé.

Tsiatis, A.A. (1975), "A nonidentifiability aspect of the problem of competing risks," *Proceedings of the National Academy of Sciences USA*, 72, 20-22.

Van der Vaart, A. W. and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes with Applications to Statistics*, New York, Springer.

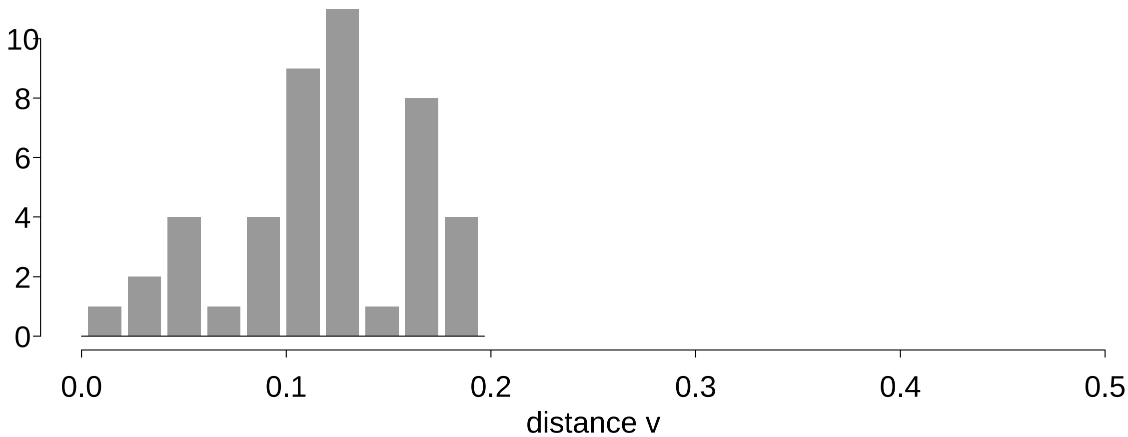
FIGURE CAPTIONS

Figure 1. Genetic data from the 45 trial participants in AACTG Study 241 who failed antiretroviral therapy. The frequency distribution of the mutational distance is shown (i) at baseline (V_b), (ii) at the late week (V_f), and (iii) accumulated between baseline and the late week ($V = V_f - \frac{2}{3}V_b$). The mark V is the scaled weighted sum of indicators of zidovudine or didanosine resistance mutations at positions 41, 65, 67, 69, 70, 74, 210, 215, and 219 in the reverse transcriptase gene, and of the nevirapine resistance mutations at positions 98, 100, 101, 103, 106, 108, 179, 181, 188, and 190 in the reverse transcriptase gene. Details about the types of mutations, including the selected weights based on the level of *in vitro* drug susceptibility, can be found in Gilbert et al. (2000).

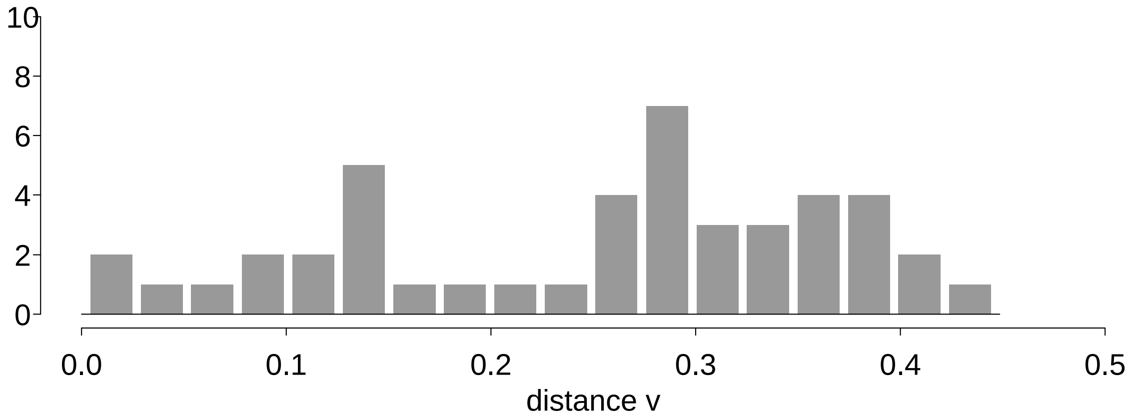
Figure 2. Accumulated mutational distance V versus failure time, for the 45 subjects in clinical trial AACTG Study 241 who failed antiretroviral therapy. The line in the plot is a lowess curve that smooths the data in windows that contain two-thirds of the nearest data-points.

Figure 3. For clinical trial AACTG Study 241, for the testing procedures that account for ties in the mark variable, (a) plots the test process $L_n(t, v)$; (b)-(i) plot individual realizations of the simulated test processes $L_n^*(t, v)$.

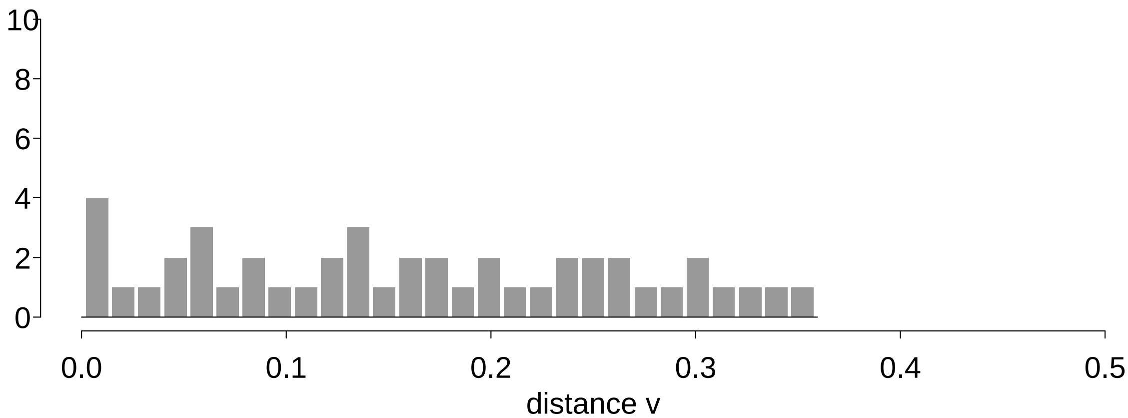
(i) mutational distance at baseline

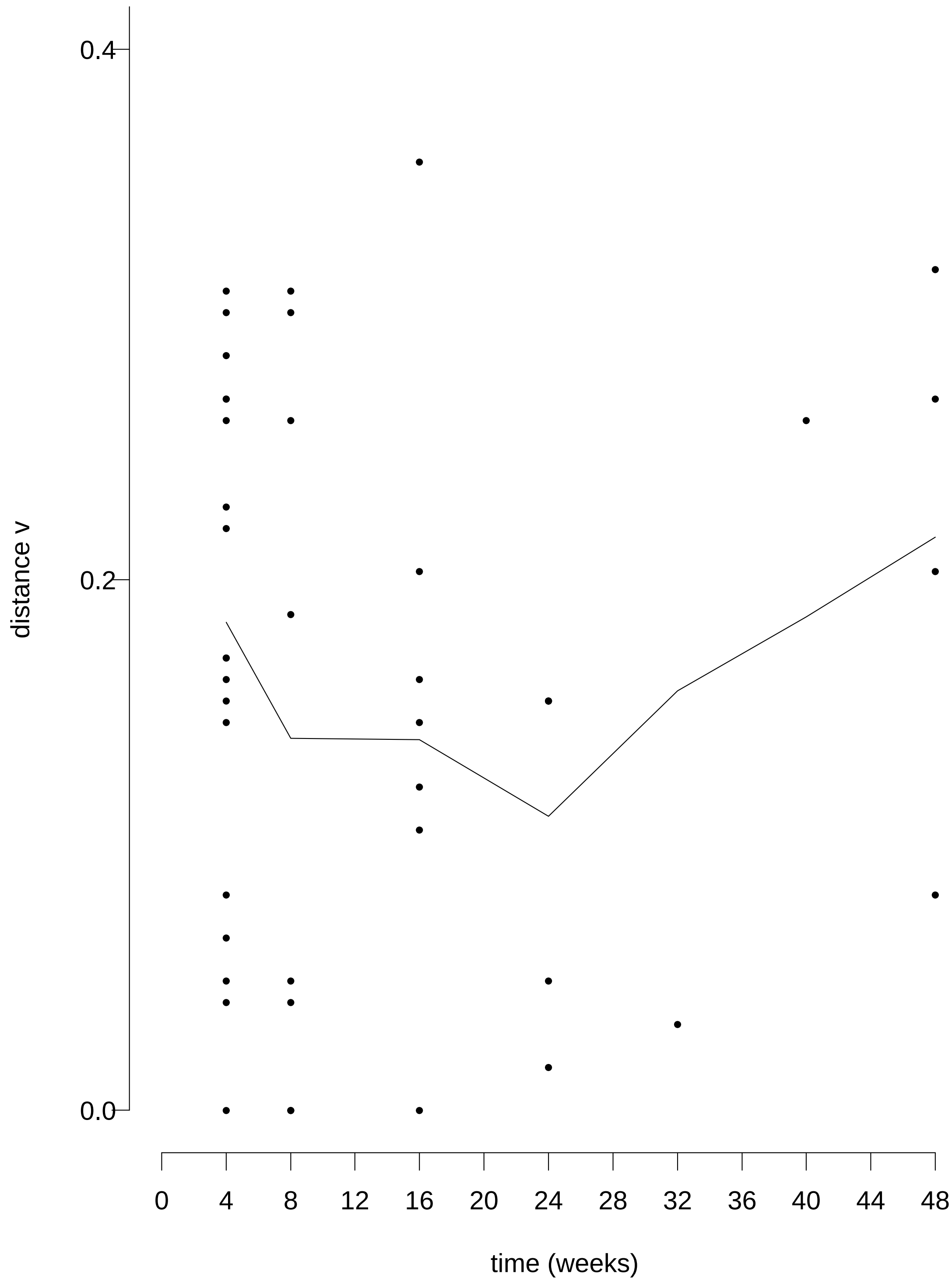


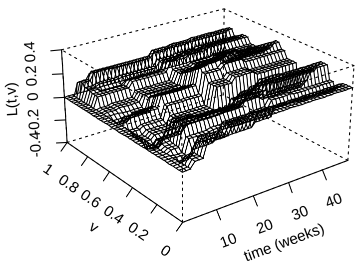
(ii) mutational distance at the late week



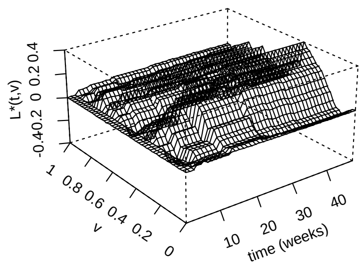
(iii) mutational distance V accumulated during the trial



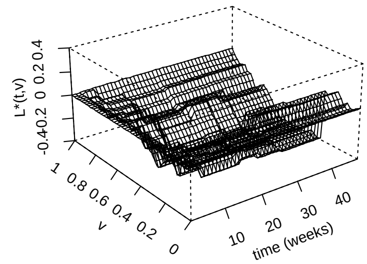

0.4
0.2
0.0
distance v
0
4
8
12
16
20
24
28
32
36
40
44
48
time (weeks)


L(t,v)
-0.4 0.2 0 0.2 0.4
v
1 0.8 0.6 0.4 0.2 0
time (weeks)
10 20 30 40

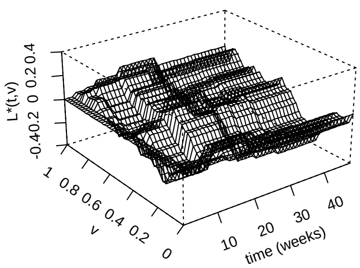
(a)


L*(t,v)
v
time (weeks)

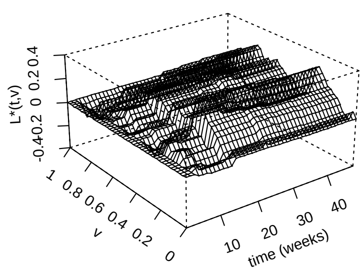
(b)


L*(t,v)
v
time (weeks)

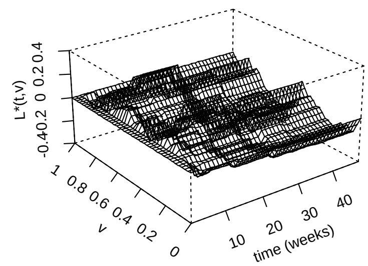
(c)


L*(t,v)
v
time (weeks)

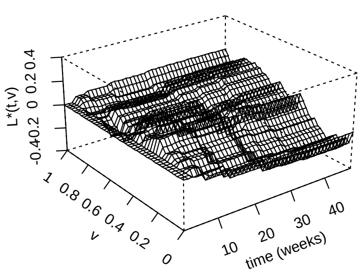
(d)


L*(t,v)
v
time (weeks)

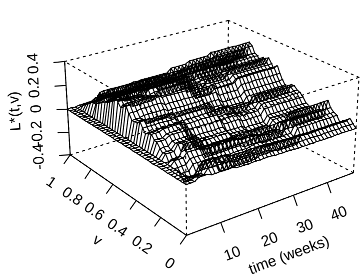
(e)


L*(t,v)
v
time (weeks)

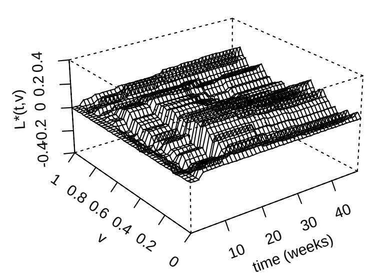
(f)


L*(t,v)
v
time (weeks)

(g)


L*(t,v)
v
time (weeks)

(h)


L*(t,v)
v
time (weeks)

(i)