The Annals of Statistics
1984, Vol. 12, No. 1, 142–160

NONPARAMETRIC INFERENCE FOR A CLASS OF SEMI-MARKOV PROCESSES WITH CENSORED OBSERVATIONS[1]

BY JOSEPH G. VOELKEL AND JOHN CROWLEY

University of Wisconsin-Madison and Fred Hutchinson Cancer Research Center, Seattle

A class of semi-Markov models, those which have proportional hazards and which are forward-going (if state j can be reached from i, then i cannot be reached from j), are shown to fit into the multiplicative intensity model of counting processes after suitable random time changes. Standard large-sample results for counting processes following this multiplicative model can therefore be used to make inferences on the above class of semi-Markov models, including the case where observations may be censored. Large-sample results for a four-state model used in clinical trials are presented.

1. Introduction. Counting process techniques (Aalen, 1975, 1978) are valuable tools in the nonparametric analysis of right-censored data. The large-sample theory for the Kaplan-Meier estimator and its associated hazard estimator (Breslow and Crowley, 1974), for the logrank test (Mantel, 1966), and for a censored-data generalization of the Wilcoxon test (Gehan, 1965) were all derived without counting process techniques. Nevertheless, such techniques provide an elegant and unifying approach for, and lead to a deeper understanding of, such results.

In addition, these techniques have also been used recently in their own right to analyze censored data. These include a modification of the Kolmogorov-Smirnov test (1980), and the introduction of a family of two-sample tests (1979) by Fleming, et al.; Aalen's (1980) development of a regression model that complements the proportional-hazard model proposed by Cox (1972); and Aalen and Johansen's (1978) study of the (nonhomogeneous) finite-state Markov model. For an excellent review of counting process techniques, see Andersen, Borgan, Gill, and Keiding (1982).

All the above uses of counting-process theory rely on the multiplicative-intensity model for making inferences: roughly speaking, if $N(\cdot)$ is a counting process with intensity $L(\cdot)$, the multiplicative-intensity model is said to hold if $L(\cdot) = Y(\cdot)a(\cdot)$, where $Y(\cdot)$ is a stochastic process and $a(\cdot)$ is a hazard function for which inferences are sought. For example, in the one-sample case $N(t)$ is the number of events up to time t, $Y(t)$ is the number at risk at t, and $a(t)$ is the hazard associated with the distribution in question. In the Markov model men-

Received June 1982; revised July 1983.

[1]Research supported in part by NIH grants P30-CA-14520, R01-CA-18332, and R01-GM-28314; and a University of Wisconsin WARF Fellowship.

AMS 1970 *subject classifications.* Primary 62G05; secondary 60K15.

Key words and phrases. Nonparametric inference, semi-Markov models, censored observations, clinical trials.

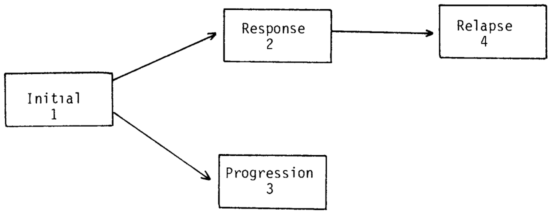

FIG. 1. *State space of a model used in clinical trials.*

tioned above, $N = N_{ij}$ counts the $i \to j$ (state i to state j) transitions, $Y = Y_i$ counts the number in state i, and $a = a_{ij}$ is the $i \to j$ transition intensity.

The semi-Markov model does not readily fit into the multiplicative-intensity framework, precisely because of its renewal nature. In order to circumvent this, we first establish some notation and preliminary results in Section 2. In Section 3 we define the class of semi-Markov models considered here: the class consists of finite-state semi-Markov models that (a) are extended to have (possibly stochastic) proportional hazards and (b) are restricted to be forward-going, in the sense that if state j can be reached from state i, then state i cannot be reached from state j. In Section 4 we introduce random time changes and show how these can be used to transform our original counting processes to fit into the multiplicative-intensity model. Finally, in Section 5 we consider a four-state semi-Markov model that has proven useful in certain clinical trials. This model, whose Markov analog was studied by Temkin (1978), assumes that from an initial state either a progression state or a response state may be entered, and from the response state a relapse state may be entered (see Figure 1). Using this model, we establish the large-sample properties for an estimator of a useful measure of a treatment's efficacy, the probability-of-being-in-response function. All proofs are relegated to the appendix.

Aalen (1975) presented an example of a simple semi-Markov model. Although his example contains some minor errors—his L_1 process is in fact not a counting process—the idea of random time changes is inherent in his paper. Our paper formalizes this idea, extends it to a larger class of models, and uses an example to show how asymptotic theory may be employed in this larger class. Nonparametric inference for semi-Markov models has been studied in a more general setting by Gill (1980). This generality is achieved primarily by relying on relatively sophisticated machinery and fairly technical assumptions. In contrast, our more restrictive class of processes relies on less machinery for its development. It also allows us to use martingale theory directly, which results in milder conditions and a simpler proof of convergence

2. Notation and preliminary results.

We first introduce some notation and results for general multivariate counting processes and then establish the framework for the semi-Markov model.

Let $(\Omega, \mathscr{F}, P)$ be a complete probability space and let $\{\mathscr{F}_t\}$ for $t \in [0, 1]$ be a history, i.e., an increasing, right-continuous family of sub-sigma-fields of $\mathscr{F} =$

$\mathscr{F}_1$. All the processes below are defined on this space and have either real-valued or vector-valued outcomes. To this end, let $D[0, 1]$ and $C[0, 1]$ be defined in the usual way (see Billingsley, 1968, for example), and let $S[0, 1]$ be the left-handed partner of $D[0, 1]$; i.e., $S[0, 1]$ is the set of real-valued functions on [0, 1] that are left-continuous and have right-hand limits. A stochastic process $X(\cdot)$ with outcomes in one of these spaces is said to be a random element of that space, and, if $X(t)$ is $\mathscr{F}_t$-measurable for each t in [0, 1], it is said to be adapted (to $\{\mathscr{F}_t\}$). Finally, let ℓ be Lebesgue measure on [0, 1].

The following definition and theorems may be found in Aalen (1975, 1978); see Brémaud and Jacod (1977) for a more comprehensive treatment.

DEFINITION 1. A stochastic process $\mathbf{N} = (N_1, N_2, \cdots, N_p)$ is a multivariate counting process if

I) The sample paths of each N_i are right-continuous step functions with a finite number of jumps, each positive and of size 1, and $N_i(0) = 0$.
II) Two component processes, N_i and N_j, $i \neq j$, cannot jump at the same time.
III) $\mathbf{N}$ is adapted to $\{\mathscr{F}_t\}$.

Let $S_1 < S_2 < \cdots$ be the jump times of $\Sigma_i N_i$—note that the S_m are stopping times—and let $V_m = i$ if N_i jumps at S_m. In the sequel it will always hold that $E(N_i(\cdot)) < \infty$ for each i (this assumption can be dropped if recourse is made to local martingales, but this is not necessary in our applications), and that $P(S_{m+1} - S_m \leq t, V_{m+1} = i \mid \mathscr{F}_{S_{(m)}})$ is absolutely continuous in t for all m, and has derivatives in $S[0, 1]$. Under these conditions, one can prove the following theorem.

THEOREM 1. *There exists a unique (up to equivalence) nonnegative adapted process* $\mathbf{L} = (L_1, L_2, \cdots, L_p)$ *such that each* L_i *is a random element of* $S[0, 1]$ *and* $M_i(t) = N_i(t) - \int_0^t L_i \, d\ell$ *are square-integrable martingales. In addition,* M_i *and* M_j *for* $i \neq j$ *are orthogonal (i.e., their product is a martingale) and* $M_i^2(t) - \int_0^t L_i d\ell$ *are martingales.*

The process $\mathbf{L}$, called the intensity process of $\mathbf{N}$ with respect to $\{\mathscr{F}_t\}$, can be viewed as a vector of conditional hazard functions, since one can show that $L_i(t^+) = \lim_{h\downarrow 0} h^{-1} P(N_i(t+h) - N_i(t) = 1 \mid \mathscr{F}_t)$, whenever L_i is bounded by an integrable random variable.

Frequently, L_i can be written in the form $L_i(t) = a_i(t) Y_i(t)$, all i, where each Y_i is an adapted random element of $S[0, 1]$ (and is often, from a statistician's point of view, observable), while the a_i's, deterministic functions of $S[0, 1]$ (usually unknown to the statistician), are underlying hazard functions associated with the counting processes. When $\mathbf{L}$ has such a form, the multiplicative intensity model is said to hold. As we shall see below, such a model allows us to make inferences about $A_i(t) = \int_0^t a_i \, d\ell$ and related functions.

We examine certain integrals below, and to ensure that they are all well-defined we will assume in the sequel that positive $Y_i(t)$ values are bounded away from 0, for all i and t. Note that when Y_i has the interpretation of "number at risk" this assumption is automatically satisfied. Define $K_i(t) = I\{Y_i(t) > 0\}$, where I is the standard indicator function.

THEOREM 2. *Define* **N**, **L**, *and* **M** *as in Theorem* 1 *and assume that the multiplicative intensity model holds with* $L_i = a_i Y_i$ *for all i. Let* H_i, *all i, be adapted random elements of* $S[0, 1]$ *that satisfy* $\sup_t |H_i(t)K_i(t)(Y_i(t))^{-1}| < d$ a.s. *for some finite d. Then*

$$\hat{B}_i(t) - B_i^+(t) = \int_0^t K_i H_i Y_i^{-1}\, dN_i - \int_0^t a_i K_i H_i\, d\ell = \int_0^t K_i H_i Y_i^{-1}\, dM_i$$

are orthogonal square-integrable martingales, and

$$(\hat{B}_i(t) - B_i^+(t))^2 - \int_0^t a_i H_i^2 K_i Y_i^{-1}\, d\ell$$

are martingales.

The conditions place on the H_i allow the stochastic integrals $\int_0^t K_i H_i\, dM_i$ to be interpreted as Lebesgue-Stieltjes integrals (Doléans-Dadé and Meyer, 1970).

An example of the multiplicative-intensity model. Let $U_1, U_2, \cdots, U_p$ be nonnegative i.i.d. random variables on $(\Omega, \mathscr{F}, P)$ and have survival function F, where $F(t) > 0$ for $t \le 1$ and the hazard function $a(t) = -(dF(t)/d\ell)F(t)^{-1}$ exists and is in $S[0, 1]$. Let $N_i^*(t) = I\{U_i \le t\}$ and $Y_i^*(t) = I\{U_i \ge t\}$ be adapted. It is easy to show that each N_i^* is a counting process and that each $L_i^* = aY_i^*$ satisfies the conditions of Theorem 1; hence, $L_i^*(t)$ is the intensity process of $N_i^*(t)$ with respect to $\{\mathscr{F}_t\}$. This choice of L_i^* is intuitively appealing: if U_1 is the death time of the ith patient in a clinical trial, the above says that the individual is subjected to the hazard $a(t)$ of dying until death itself occurs, at which point the hazard ceases. To include a simple form of censoring in the above scheme (a more general formulation is possible), let ($\mathbf{C} = C_1, \cdots, C_p$) be $\mathscr{F}_0$-measurable random variables such that $\mathbf{C}$ is independent of $U_1, \cdots, U_p$. The C_i are to be considered latent censoring times: to this end, define the censoring processes J_i by $J_i(t) = I\{C_i \ge t\}$, and also define the censored counting processes corresponding to the above, $N_i(t) = \int_0^t J_i\, dN_i^*$. Since $M_i^*(t) = N_i^*(t) - \int_0^t L_i^*\, d\ell$ are orthogonal square-integrable martingales with respect to $\{\mathscr{F}_t\}$, so are $M_i(t) = \int_0^t J_i\, dM_i^* = N_i(t) - \int_0^t J_i L_i^*\, d\ell$, by Theorem 2; i.e., N_i has intensity $L_i = J_i L_i^*$. Note that in this setup $\mathbf{C}$ is $\mathscr{F}_0$-measurable, and this can be interpreted to mean that the latent censoring times are known at time zero. Usually this is not reasonable; however, if we let

$$\mathscr{G}_t = \sigma(\mathbf{N}(s),\ Y_i^*(s)J_i(s),\ s \le t,\ i = 1, \cdots, p),$$

then $N_i(t)$ has intensity $L_i(t)$ with respect to $\{\mathscr{G}_t\}$, since Theorem 1 holds using $\{\mathscr{G}_t\}$ by the Innovation Theorem (Aalen, 1978). This is the history one would want to use because it is the observable one. This argument is merely a precise way of saying that the martingale property, and hence inferences on the $a_i(t)$, are not affected by whether or not we know the latent censoring times at time zero.

If we let $\bar{N} = \Sigma_i N_i$, $\bar{Y} = \Sigma_i Y_i^* J_i$, and $\bar{L} = \Sigma_i L_i$, then, since a sum of martingales is again a martingale, $\bar{N}$ has intensity $\bar{L} = a\bar{Y}$. By Theorem 2,

$$\hat{A}(t) - A^+(t) = \int_0^t \bar{Y}^{-1} I\{\bar{Y} > 0\}\, d\bar{N} - \int_0^t I\{\bar{Y} > 0\} a d\ell$$

is a martingale, so a reasonable estimator of $A(t) = \int_0^t a d\ell$ is $\hat{A}(t)$. (This estimator has been suggested on intuitive grounds by Nelson, 1969.) In the case where the latent censoring variables C_i are i.i.d. with survival function $\mathscr{G}$, it follows that $E(\hat{A}(t)) = A(t) - \int_0^t a(s)(1 - F(s)G(s))^p\, ds$. That the expected value of the estimator converges to $A(t)$ at this exponential rate (when $G(t) > 0$) was pointed out by Aalen (1976) in the context of no censoring.

3. A class of semi-Markov models. We restrict the class of semi-Markov models that we will examine in three ways. First, we assume that the number of states in the model is finite. Second, we assume that the underlying distributions of the model are absolutely continuous, with left-continuous hazard functions. This allows us to view the model purely in terms of these hazard functions a_{ij}: if T and T' are the (first, say) entry and exit times for state i in the customary semi-Markov process, V is the next state visited, and $\{\mathscr{G}_t\}$ is the history of the process, then, in obvious notation,

$$a_{ij}(t) = \lim_{h \downarrow 0} \frac{1}{h} P(t < T' - T \le t + h,\ V = j \mid \{T' - T > t\},\ \mathscr{G}_t).$$

Third, we assume that if state j can be reached from state i, then state i cannot be reached from state j. (This assumption, apparently needed for the martingale theory to function properly, was also made by Aalen (1975) in his attempt to model a specific semi-Markov process.) We use this assumption to simplify notation by numbering the states in the model such that an $i \rightarrow j$ transition is possible only if $i < j$. We also generalize the model, in the customary proportional-hazards manner: in a sense to be made more precise below, the $i \rightarrow j$ hazard for a particular observation in state i at time t is $a_{ij}(t - T_i)Z_i(t)$, where T_i is the time state i was entered and Z_i is a stochastic process.

To put the above ideas on a more concrete footing, we now develop notation for the kth of n observations.

Define the following:

N_{ijk}^0: A counting process which counts the actual number (i.e., in the presence of censoring) of transitions from state i to state j. We assume that the histories generated by observations, denoted by $\{\mathscr{F}_{kt}: t \in [0, 1]\}$, are independent so, e.g., information on one observation cannot be used to censor another observation. We show below that this restriction is needed to preserve the martingale property in the randomly changed time.

T_{ik}: The time at which state i is entered. This is a stopping time with respect to $\{\mathscr{F}_{kt}\}_t$ for each i.

T'_{ik}: The time at which state i is left, also a stopping time for each i.

L^0_{ijk}: The intensity process of N^0_{ijk}. Assume the L^0_{ijk} have the form $L^0_{ijk}(t) = a_{ij}(t - T_{ik})Y^0_{ik}(t)$, where $a_{ij}(t)$ is the underlying hazard function of the $i \to j$ transition (and hence is in $S[0, 1]$), and $Y^0_{ik}(t) = I_{ik}(t)X^0_{ik}(t)$ is defined via

I_{ik}: The indicator function for state i, defined so that its sample paths are in $S[0, 1]$. Note that I_{ik} is adapted.

X^0_{ik}: Some non-negative process such that $I_{ik}X^0_{ik}$ is adapted and is a random element of $S[0, 1]$. In the simplest case, the one-sample semi-Markov model without censoring, $X^0_{ik}(t) = 1$ for all i, k and t. More generally, it can be constructed to include a censoring scheme (as in Section 5) and it also can attach weights to the observations' hazard functions as done in the Cox model. (See Andersen and Gill, 1982, for an elaboration of the counting process approach to the Cox model.) Note that X^0_{ik} may also be a function of transition times up to T_{ik}, e.g., $X^0_{ik} = e^{-\beta\tau_{ik}}$. (The corresponding $I_{ik}X^0_{ik}$ is adapted.)

S_{mk}: The time of the mth transition. This is also a stopping time for each m. If the mth transition has not been reached by time 1, set $S_{mk} = 1^+$.

V_{mk}: The mth state visited, where V_{0k} is the initial state. Note that if $J(t)$ were to be defined as the state the observation is in at time t, then J is an adapted random element of $D[0, 1]$ and $V_{mk} = J(S_{mk})$; hence (Neveu, 1965), V_{mk} is $\mathscr{F}_{kS_{mk}}$-measurable. Also note that $V_{mk} = i$ if and only if $S_{mk} = T_{ik}$, and that, because of the convention used in numbering the states, V_{mk} increases in m.

Note that L^0_{ijk} is in fact an intensity process, since it is adapted and has sample paths in $S[0, 1]$.

Now consider the state space in Figure 1 and assume in the one-sample setting that three observations make $1 \to 2$ transitions at times .2, .3, and .6. At time .7, if no other entries and no exits from the state have occurred, then the intensity of a $2 \to 4$ transition is $a_{24}(.5) + a_{24}(.4) + a_{24}(.1)$ (as opposed to the corresponding Markov intensity of $3a_{24}(.7)$). Clearly, the multiplicative-intensity model does not hold here.

A natural approach to putting the above into the multiplicative-intensity framework would be to gather all the observations that enter state 2, keep them there without risk until a suitable starting time to align their hazard functions with each other, and then expose them to the appropriate hazards. (It is here that our third restriction is required: if an observation could enter state 2 more than once, the two hazards could not be aligned.) This can be done quite easily heuristically, but we prefer to examine these so-called random time changes in a more rigorous manner. This rigor enables us to preserve the martingale structure, which we use to find the joint asymptotic distribution of certain estimators. One can also use this structure to examine extensions of the semi-Markov model, for example by letting hazards for a particular state depend on the past in certain ways.

The case in which i.i.d. observations, possibly right-censored, follow a semi-Markov process was examined by Lagakos, Sommer and Zelen (1978). They wrote down a likelihood for the observations and proceeded to maximize it in the sense of Kaplan and Meier (1958). Then, formally assuming that transitions occur only at a finite number of times, they indicated the asymptotic mean and covariance structure of certain estimators. Our results put this earlier work into a more general and rigorous framework.

4. The random time change and the transformed counting processes. Define the random time change function $\Psi_k(t)$ by

$$\Psi_k(t) = t - S_{mk} + V_{mk} - 1, \; t \in [S_{mk}, S_{m+1,k}). \tag{3.1}$$

Thus, for a fixed outcome ω, $\Psi_k(\omega, \cdot)$: $[0, 1] \to [0, m']$, where m' is the number of states in the model; in addition, note that $\Psi_k(S_{mk}) = V_{mk} - 1$. Also define $\Psi_k^{-1}(u) = \inf\{t{:}\Psi_k(t) \geq u\}$. This is almost a bona fide inverse, since $\Psi_k^{-1}\Psi_k(t) = t$, and $\Psi_k\Psi_k^{-1}(u) \geq u$, with equality if and only if the right-hand derivative of $\Psi_k^{-1}(u)$ is 1. It also follows from the definition of Ψ_k^{-1} that

$$\Psi_k^{-1}(u) \leq t \text{ iff } \Psi_k(t) \geq u, \tag{3.2}$$

and that

$$\begin{aligned}\Psi_k^{-1}(u) &= \begin{cases} 0, & u \in [0, V_{0k} - 1] \\ u + S_{mk} - V_{mk} + 1, & u \in (V_{mk} - 1, V_{mk} - 1 + S_{m+1,k} - S_{mk}] \\ S_{m+1,k}, & u \in (V_{mk-1} + S_{m+1,k} - S_{mk}, V_{m+1,k} - 1], \end{cases} \\ &= \begin{cases} 0, & u \in [0, V_{0k} - 1] \\ u + T_{ik} - i + 1, & u \in (i - 1, i - 1 + T'_{ik} - T_{ik}] \\ T'_{ik}, & u \in (i - 1 + T'_{ik} - T_{ik}, i]. \end{cases}\end{aligned} \tag{3.3}$$

The relationship (3.2) implies

$$\begin{aligned}\{\Psi_k^{-1}(u) \leq t\} &= \cup_m \{t - S_{mk} + V_{mk} - 1 \geq u\} \cap \{S_{m+1,k} > t\} \cap \{S_{mk} \leq t\}, \\ &= \cup_m \cup_i \{S_{mk} \leq t - u + i - 1\} \cap \{V_{mk} = i\} \cap \{S_{m+1,k} > t\} \cap \{S_{mk} \leq t\}.\end{aligned}$$

Now S_{mk} is a stopping time for each m, and V_{mk} is $\mathscr{F}_{kS_{mk}}$-measurable, so $\Psi_k^{-1}(u)$ is a stopping time for each u. Also, using (3.2) again, $\Psi_k(t)$ is $\mathscr{F}_{kt}$-measurable, and it is also a random element of $D[0, 1]$ because of its sample-path properties.

Before proceeding further we give a simple example. Using the state space in Figure 1, assume for a particular outcome that $1 \to 2$ and $2 \to 4$ transitions occurred at times .1 and .5, respectively. Then the transition number index m assumes the values 0, 1, and 2, the respective values of S_{mk} are $0(T_{1k})$, $.1(T_{2k})$, and $.5(T_{4k})$, and the respective values of V_{mk} are 1, 2, and 4. The corresponding values of Ψ_k and Ψ_k^{-1} are graphed in Figure 2.

We now apply the random time change to the counting processes and then derive the corresponding intensity processes. These new intensity processes are in the form of a multiplicative-intensity model and also conform to one's intuition.

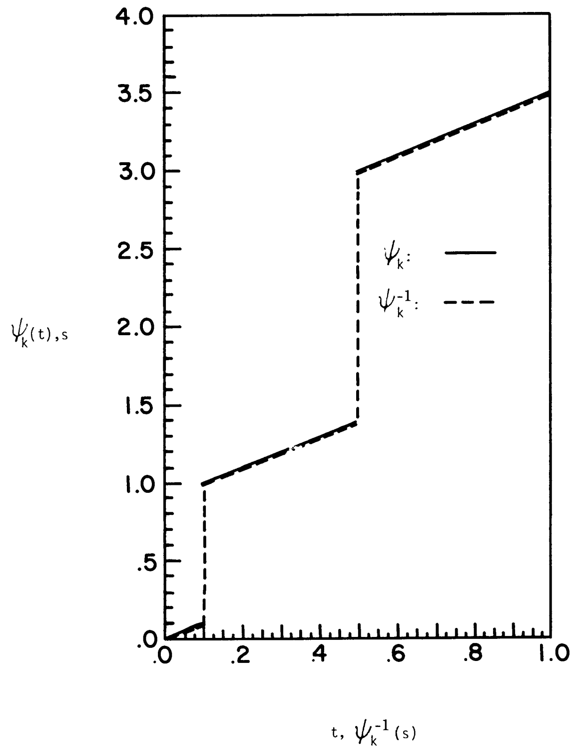

FIG. 2. *Values of Ψ_t and Ψ_t^{-1} for a 1 → 2 transition at $t = 0.1$ and a 2 → 4 transition at $t = 0.5$.*

Define $M^0_{ijk}(t) = N^0_{ijk}(t) - \int_0^t L^0_{ijk}\, d\ell$, all (i, j); these are square-integrable orthogonal martingales with respect to $\{\mathscr{F}_{kt}\}_t$.

THEOREM 3. *For $u \in [0, m']$,*

$$M^*_{ijk}(u) = M^0_{ijk}(\Psi_k^{-1}(u)), \quad \textit{all } (i, j),$$

are square-integrable orthogonal martingales with respect to $\{\mathscr{F}^*_{ku}\}_u$, where $\mathscr{F}^*_{ku} = \mathscr{F}_{k\Psi_k^{-1}(u)}$. Also,

$$M^*_{ijk}(u) = N^*_{ijk}(u) - \int_0^u a^*_{ij} Y^*_{ik}\, d\ell$$

where

$$N^*_{ijk}(u) = I\{u \geq i - 1 + T'_{ik} - T_{ik}\} I\{N^0_{ijk}(1) = 1\},$$

$$a^*_{ij}(s) = a_{ij}(s - (i - 1)),$$

and

$$Y^*_{ik}(s) = X^0_{ik}(s + T_{ik} - (i - 1)) I\{i - 1 < s \leq i - 1 + T'_{ik} - T_{ik}\}.$$

In particular, N^*_{ijk} is a counting process with intensity process $L^*_{ijk} = a^*_{ij} Y^*_{ik}$ with respect to $\{\mathscr{F}^*_{ku}: u \in [0, m']\}$.

We now introduce a bit more notation, most of which we use repeatedly in the sequel. Let $N_{ij}(t) = \Sigma_k N^*_{ijk}(t + i - 1)$, the number of $i \to j$ transitions whose transition times from i to j (i.e., amount of time spent in i before going to j) is less than or equal to t—here we begin to let "t" represent time other than real time. Also, let $Y_i(t) = \Sigma_k Y^*_{ik}(t + i - 1)$. In the one-sample case with right-censoring, this is the number of observations at risk in state i just before time t, where the time is "local" to state i. Finally, define $L_{ij} = a_{ij} Y_i$, $M^*_{ij} = \Sigma_k M^*_{ijk}$, $\mathscr{F}^*_t = V_k \mathscr{F}^*_{kt}$, and $M_{ij}(t) = M^*_{ij}(t + i - 1) = N_{ij}(t) - \int_0^t L_{ij}\, d\ell$. (To fix ideas, we continue the example where $1 \to 2$ transitions were made at .2, .3, and .6, and suppose the corresponding $2 \to 4$ transitions were made at .9, .8, and 1.0. Then the amounts of time spent in state 2 are .7, .5, and .4, so $N_{24}(.3) = 0$, $N_{24}(.6) = 2$, $Y_2(.3) = 3$, $Y_2(.6) = 1$, $L_{24}(.3) = a_{24}(.3) Y_2(.3)$, and $L_{24}(.6) = a_{24}(.6) Y_2(.6)$.) The M^*_{ij} are square-integrable orthogonal martingales with respect to $\{\mathscr{F}^*_t\}$; however, we have abused notation in defining the M_{ij}, for there may well be no history with respect to which these are orthogonal square-integrable martingales. Nevertheless, for the large-sample theory of the next section, we may endow the M_{ij}'s with this property: from a mathematical point of view, this is equivalent to proving the results for the M^*_{ij} processes and then employing shift operators (Billingsley, 1968, Section 17); from a notational point of view the M_{ij} processes are the natural ones to use because they are easily interpretable. Without loss of generality, we assume these processes are defined on [0, 1], rather than on [0, m'].

To indicate that the independence of the observations' histories is an important assumption, suppose otherwise. Specifically, for the state space in Figure 1, let us follow two observations for $t \in [0, 1]$ and assume that if one reaches state 4 while the other is in state 1 then the latter observation is censored. For the sample paths given in Figure 3, the history of the two processes in the randomly changed time .6 cannot contain the information that the first will make a $2 \to 4$ transition, for this would be "seeing the future". On the other hand, the history must contain information on the second observation up to time .6, but the information $N_{121}(.6) = 0$ and $L_{121}(.6) = 0$ (in obvious notation) implies that observation 1 will make a $2 \to 4$ transition. Our modeling of these processes does allow us to let the observations' histories affect each other in the transformed time without losing the martingale structure of the M^*_{ij}'s, but from a practical point of view this would be absurd.

Forcing the semi-Markov processes into the multiplicative-intensity model allows us to apply the corresponding asymptotic theory (e.g. Aalen and Johansen, 1978) after suitable modifications. The most substantial change we make to such asymptotic theory arises because our integrals may be of the form $\int_0^t H^n(t, s)\, dM^n(s)$ where $\{M^n\}$ is a sequence of martingales correspoding to counting processes and $\{H^n\}$ is a sequence of random functions. The usual asymptotic theory assumes $H^n(t, s) = H^n(s)$ is $\mathscr{F}^n_s$-adapted, while in our situation

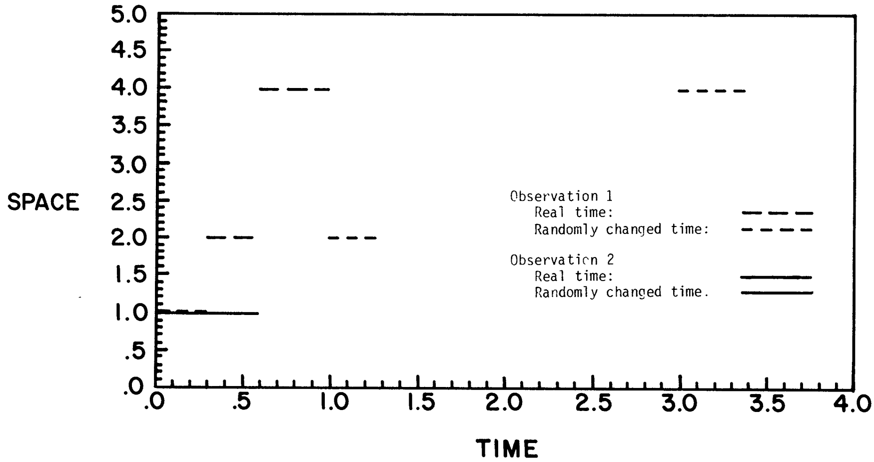

FIG. 3. *Sample paths for two observations, to show the importance of independence. Note: observation* 2 *is censored at time* 0.6.

it is not. Incidentally, it is this same reason that has precluded us from deriving small-sample properties of the estimator we present in the next section.

5. The PBRF and its large-sample properties. A reasonable model in cancer clinical trials assumes that each patient may either remain in an initial state, or progress, or respond and then possible relapse—see Figure 1. Temkin (1978) realized the use of such a model. She first pointed out that when two treatments are compared, the two most commonly used measures of a treatment's efficacy, probability of responding and time to progression or relapse, may yield contradictory information. A more cohesive measure of a treatment's ability, she argued, is the probability-of-being-in-response function (PBRF), viewed as a function of time. Assuming a Markov model, she developed an estimator of the PBRF in the presence of right-censored data and estimated its large-sample variance in the manner of Lagakos, Sommer and Zelen (1978), by formally assuming only a finite number of transition times. At the same time, Aalen and Johansen (1978) developed both small- and large-sample theory for the finite-state Markov model in a more rigorous framework; in particular, this theory can be applied to Temkin's estimator.

In this section, we present an estimator of the PBRF in the semi-Markov one-sample setting with censored data and derive its large-sample distributional properties. The semi-Markov PBRF at time t is

$$R(t) = \int_0^t F_1(s)a_{12}(s)F_2(t - s)\, ds,$$

where F_i, in non-standard notation, is the survival function corresponding to state i. Note that this equals its Markov counterpart,

$$\int_0^t F_1(s)a_{12}(s)(F_2(t)/F_2(s))\,ds,$$

whenever $a_{24}(t)$ is a constant function. (The case in which all the hazard functions are constants, the homogeneous Markov case, was investigated by Begg and Larson, 1982).

Now define $\hat{F}_1 = \exp(-\hat{A}_{12} - \hat{A}_{13})$ and $\hat{F}_2 = \exp(-\hat{A}_{24})$, where $\hat{A}_{ij}(t) = \int_0^t K_i Y_i\, dN_{ij}$, and $K_i(t) = I\{Y_i(t) > 0\}$. (We use the $\hat{F}_i$'s for the large-sample derivations because they exhibit the correct large-sample properties in more general settings than the one on which we concentrate here; however, in the actual one-sample setting, it may be more reasonable to use the corresponding Kaplan-Meier (1958) estimators. Such an interchange has no effect on the large-sample results.) Also define our estimator of $R(t)$,

$$\hat{R}(t) = \int_0^t \hat{F}_1(s^-)\hat{F}_2(t-s)K_1(s)(Y_1(s))^{-1}\,dN_{12}(s).$$

Note that although all these estimators are functions of n, the number of observations, we leave this relation implicit for readability.

THEOREM 4. *Let* $0 < c < 1$ *and assume for* $t \in [0, c]$

I) $Y_i(t)/n = p_i(t) + o_p(1)$ *for* $i = 1, 2$, *where the* p_i *are deterministic functions bounded away from* 0.

II) $nK_i(t)Y_i^{-1}(t)$ *is uniformly integrable in* (n, t, i).

III) $n^{1/2}\int_0^1 (1 - K_i)a_{ij}\, d\ell = o_p(1)$ *for all* (i, j).

Then $n^{1/2}(\hat{R} - R) \to_d Z$ *on* $[0, c]$, *where*

$$Z(t) = \int_0^t g_{12}(t,\cdot)\,dW_{12} - \int_0^t \left[\int_0^{t-s} F_1 F_{2;t} a_{12}\, d\ell\right] h_{24}(s)\,dW_{24}(s)$$

$$- \int_0^t \left[\int_s^t F_1 F_{2;t} a_{12}\, d\ell\right] (h_{12}(s)\,dW_{12}(s) + h_{13}(s)\,dW_{13}(s)),$$

with $F_{2;t}(s)$ *defined as* $F_2(t - s)$. *Further, the* W_{ij} *are independent Wiener processes, and*

$$g_{12}(t,s) = F_1(s)F_2(t-s)(a_{12}(s)/p_1(s))^{1/2}$$

$$h_{12}(s) = (a_{12}(s)/p_1(s))^{1/2}$$

$$h_{13}(s) = (a_{13}(s)/p_1(s))^{1/2}$$

$$h_{24}(s) = (a_{24}(s)/p_2(s))^{1/2}.$$

Note that we restrict our attention to $[0, c]$. In our setup we only observe $2 \to 4$ transitions in real time $[0, 1]$, so $Y_2(1) = 0$ a.e. $[P]$, for each n, assuming that each observation is in the initial state at time zero. Thus we must choose $c < 1$ so that assumption I of the theorem can hold.

COROLLARY 1. *If the conditions of Theorem 4 hold, then for* $0 \le u \le t \le c$, $\mathrm{Cov}(Z(u), Z(t))$ *is*

$$\int_0^u F_1^2 F_{2;t} F_{2;u} (a_{12}/p_1)\, d\ell$$
$$- \int_0^u F_1(s) F_{2;t}(s) \left[\int_s^u F_1 F_{2;u} a_{12}\, d\ell\right] (a_{12}(s)/p_1(s))\, ds$$
$$- \int_0^u F_1(s) F_{2;u}(s) \left[\int_s^t F_1 F_{2;t} a_{12}\, d\ell\right] (a_{12}(s)/p_1(s))\, ds$$
$$+ \int_0^u \left[\int_0^{t-s} F_1 F_{2;t} a_{12}\, d\ell\right]\left[\int_0^{u-s} F_1 F_{2;u} a_{12}\, d\ell\right] (a_{24}(s)/p_2(s))\, ds$$
$$+ \int_0^u \left[\int_s^t F_1 F_{2;t} a_{12}\, d\ell\right]\left[\int_s^u F_1 F_{2;u} a_{12}\, d\ell\right] \times \left(\frac{a_{12}(s)}{p_1(s)} + \frac{a_{13}(s)}{p_1(s)}\right) ds.$$

If in addition $n^2 K_i(t) Y_i^{-2}(t)$ *is uniformly integrable in* (n, t, i), *then an estimate of* $\mathrm{Cov}(Z(u), Z(t))$ *that is consistent uniformly in t is given by replacing* $F_1(s)$ *in the above by* $\hat{F}_1(s^-)$, $F_{2;t}(s)$ *by* $\hat{F}_2(t-s)$, $a_{ij}\, d\ell$ *by* $K_i Y_i^{-1}\, dN_{ij}$, *and* p_i *by* Y_i/n.

Note that these five integrals may be rewritten as three integrals, each of which would involve only one of the a_{ij}/p_i, thus separating the covariance into components reflecting the variability in the estimation of the $1 \to 2$, $1 \to 3$, and $2 \to 4$ hazards.

As Temkin points out, the area under the curve R in $[0, t]$, $\int_0^t R\, d\ell$, is the average time spent in response in this interval. A slight generalization of this leads to

COROLLARY 2. *Let* w: $[0, 1] \to R$ *be bounded and measurable. Under the conditions of Theorem 4,* $n^{1/2} \int_0^t (\hat{R} - R)\, w\, d\ell \to_d Z_1(t)$ *as a process in t, where the Gaussian process* Z_1 *is* $Z_1(t) = \int_0^t Zw\, d\ell$. *The covariance of* $Z_1(u)$ *and* $Z_1(v)$, *for* $0 \le u \le v \le c$, *is*

$$\int_0^u \left[\int_0^u (F_1(s) - F_2(t-s) - I(s, t, t) w(t)\, dt)\right.$$
$$\times \left[\int_s^v (F_1(s) - F_2(t-s) - I(s, t, t)) w(t)\, dt\right] (a_{12}(s)/p_1(s))\, ds$$
$$+ \int_0^u \left[\int_s^u I(0, t-s, t)\, dt\right]\left[\int_s^v I(0, t-s, t)\, dt\right] (a_{24}(s)/p_2(s))\, ds$$
$$+ \int_0^u \left[\int_s^u I(s, t, t)\right]\left[\int_s^v I(s, t, t)\, dt\right] (a_{13}(s)/p_1(s))\, ds,$$

where $I(x, y, z) = \int_x^y F_1 F_{2;z} a_{12}\, d\ell$.

A suitable estimate of the above covariance can be determined along the lines of Corollary 1.

A graph of $t \to \int_0^t \hat{R}\, d\ell$ provides another way to examine the relative efficacies of therapies; however, for large t the graph may be misleading, for the same reason that the expected value of a random variable may be a poor measure of center for its corresponding distribution.

Under what situations will the conditions of Corollary 1 hold, so that we can set confidence intervals on $R(t)$, perform two-sample tests, etc.? We restrict ourselves to the one-sample random censorship model; then whether the above conditions are met simply depends on the type of censoring in effect.

COROLLARY 3. *If, using the notation of Section* 3, $X_{ik}^0(t) = I\{C_k \geq t\}$, *where the censoring random variables* $\{C_k\}$ *are* i.i.d., *are greater than* 1 *with positive probability, and are independent of the underlying transition times, then the conditions of Corollary* 1 *are satisfied whenever* $a_{12}(s) > 0$ *on a subinterval* A *of* $[0, 1 - c]$ *with* $\ell(A) > 0$.

Comparison of the estimators of the PBRF based on Markov and semi-Markov assumptions have been made under a homogeneous Markov model (Voelkel, 1980). These results will be summarized in a future report.

APPENDIX

PROOF OF THEOREM 3. That the M_{ijk}^* have the stated martingale properties follows from Doob (1953). To write M_{ijk}^* in terms of N_{ijk}^*, a_{ij}^*, and Y_{ik}^*, we first write

$$\begin{aligned} M_{ijk}^*(u) &= M_{ijk}^0(\Psi_k^{-1}(u)) \\ &= N_{ijk}^0(\Psi_k^{-1}(u)) - \int_0^{\Psi_k^{-1}} a_{ij}(s - T_{ik}) Y_{ik}^0(s)\, ds \\ &= N_{ijk}^0(\Psi_k^{-1}(u)) - \int_{\Psi_k(0)}^{\Psi_k \Psi_k^{-1}(u)} a_{ij}(\Psi_k^{-1}(s) - T_{ik}) \\ &\quad \times Y_{ik}^0(\Psi_k^{-1}(s))\, d\Psi_k^{-1}(s). \end{aligned}$$

Since Ψ_k^{-1} has absolutely continuous sample paths, then $\Psi_k^{-1\prime}$ exists a.e.(ℓ) for all samples paths. In fact, let us mean by $\Psi_k^{-1\prime}$ the left-hand derivative of Ψ_k^{-1}. This has sample paths in $S[0, m']$, so we transform the integral above to get

$$M_{ijk}^*(u) = N_{ijk}^0(\Psi_k^{-1}(u)) - \int_0^u [a_{ij}(\Psi_k^{-1}(s) - T_{ik}) Y_{ik}^0(\Psi_k^{-1\prime}(s)) \Psi_k^{-1\prime}(s)]\, ds.$$

Note that the integrand is 0 for all values between u and $\Psi_k \Psi_k^{-1}(u)$, so the upper limit of integration is correct even though u and $\Psi_k \Psi_k^{-1}(u)$ may differ.

Define $N_{ijk}^*(u) = N_{ijk}^0(\Psi_k^{-1}(u))$. Since N_{ijk}^0 has at most one jump, which can only occur at T'_{ik}, and since the smallest u for which $\Psi_k^{-1}(u) = T'_{ik}$ is $u = i - 1 +$

$T'_{ik} - T_{ik}$, then

$$N^*_{ijk}(u) = I\{u \geq i - 1 + T'_{ik} - T_{ik}\}I\{N^0_{ijk}(1) = 1\}.$$

Let $Y^*_{ik}(s) = Y^0_{ik}(\Psi_k^{-1}(s))\Psi_k^{-1\prime}(s)$. Now $Y^0_{ik}(\Psi_k^{-1}(s))$ is non-zero only when $\Psi_k^{-1}(s) \in (T_{ik}, T'_{ik}]$, or when s is an element of

$$(i - 1, i - 1 + T'_{ik} - T_{ik}] \cup (i - 1 + T'_{ik} - T_{ik}, i] = (i - 1, i].$$

Also $\Psi_k^{-1\prime}(s) = 1$ if and only if $s \in \cup_i (i - 1, i - 1 + T'_{ik} - T_{ik}]$, where we interpret $(i - 1, i - 1 + T'_{ik} - T_{ik}]$ as $(i - 1, i - T_{ik}]$ if state i is entered but no $i \to j$ transition occurs, and as the null set if state i is never entered. Thus, $Y^*_{ik}(s)$ is non-zero only when $s \in (i - 1, i - 1 + T'_{ik} - T_{ik}]$ and on this interval equals $X^0_{ik}(s + T_{ik} - i + 1)$.

We now simplify $a_{ij}(\Psi_k^{-1}(s) - T_{ik})$, first noting that the term need be examined only for s such that $Y^*_{ik}(s)$ is non-zero. For such values of s, those in $(i - 1, i - 1 + T'_{ik} - T_{ik}]$, we have

$$a_{ij}(\Psi_k^{-1}(s) - T_{ik}) = a_{ij}(s + T_{ik} - i + 1 - T_{ik}) = a_{ij}(s - (i - 1)).$$

With this in mind, we define $a^*_{ij}(s) = a_{ij}(s - i + 1)$.

That the N^*_{ijk} are counting processes with intensities L^*_{ijk} is immediate. □

PROOF OF THEOREM 4. For real numbers and random variables, define o, O, o_p, and O_p in the usual way; for functions and stochastic processes define them to hold uniformly in t. To avoid writing dummy arguments in the integrals that follow, let $f_{;t}(s) = f(t - s)$ and $f_-(s) = f(s^-)$ for suitable functions or stochastic processes f. Define

$$\hat{R}_2(t) = \int_0^t \hat{F}_{1-}F_{2;t}K_1 Y_1^{-1}\, dN_{12}, \quad \hat{R}_{12}(t) = \int_0^t F_1 F_{2;t}K_1 Y_1^{-1}\, dN_{12}$$

$$R^+(t) = \int_0^t \hat{F}_{1-}\hat{F}_{2;t}K_1 a_{12}\, d\ell, \quad R_2^+(t) = \int_0^t \hat{F}_{1-}F_{2;t}K_1 a_{12}\, d\ell$$

$$R_{12}^+(t) = \int_0^t F_1 F_{2;t}K_1 a_{12}\, d\ell$$

and view $\hat{R} - R$ as a sum of five terms: $\hat{R} - R^+ - \hat{R}_2 + R_2^+$; $\hat{R}_2 - R_2^+$; $R^+ - R_2^+$; $R_2^+ - R_{12}^+$; and $R_{12}^+ - R$. The first and last of these terms are $o_p(n^{-1/2})$ under reasonable conditions.

LEMMA 1. *Assume that for* $t \in [0, c]$, *where* $c < 1$,

I) $E[K_i(t)Y_i(t)^{-1}] + E[1 - K_i(t)] = o(1)$ *for each* (i, t).

II) $Y_1(t)/n = p_1(t) + o_p(1)$ *for each* t, *and* p_1 *is bounded away from* 0.

III) $nK_1(t)Y_1(t)^{-1}$ *is uniformly integrable in* (n, t).

IV) $n^{1/2}\int_0^1 (1 - K_1)a_{12}\, d\ell = o_p(1)$.

Then $\hat{R} - R^+ - \hat{R}_2 + R_2^+$ *and* $R_{12}^+ - R$ *are both* $o_p(n^{-1/2})$.

PROOF. At t, $n^{1/2}(\hat{R} - R^+ - \hat{R}_2 + R_2^+)$ is $\int_0^t (\hat{F}_{2;t} - F_{2;t})n^{1/2}\hat{F}_{1-}K_1 Y_1^{-1}\, dM_{12}$ which equals $\int_0^t (A_t)\, dB$, say. But the submartingale inequality (Doob, 1953, page 317) and (I) ensure that $A_t(s)$ converges to 0 uniformly in s and t; in addition, the conditions imply that $B(s)$ converges weakly to a Gaussian process (Aalen and Johansen, 1978). By the continuous mapping theorem, $\int_0^t (A_t)\, dB$ converges weakly to a zero mean and variance process, and is therefore $o_p(1)$.

The second assertion clearly holds. □

What remains is to find the joint asymptotic distribution of $n^{1/2}(\hat{R}_2 - R_2^+, R^+ - R_2^+, R_2^+ - R_{12}^+)$. Note that $\hat{R}_2(t) - R_2^+(t) = \int_0^t \hat{F}_{1-}F_{2;t}K_1 Y_1^{-1}\, dM_{12}$, the integrand of $R^+(t) - R_2^+(t)$ contains the term $\hat{F}_{2;t}(s) - F_{2;t}(s) = \exp(-\hat{A}_{24}(t-s)) - \exp(-A_{24}(t-s))$, and the integrand of $R_2^+(t) - R_{12}^+(t)$ contains the term $\hat{F}_{1-}(s) - F_1(s) = \exp(-\hat{A}_{12}(s^-) - \hat{A}_{13}(s^-)) - \exp(-A_{12}(s) - A_{13}(s))$. This suggests examining the weak convergence of

$$\text{(A.1)} \quad Z_n(t) = \left(\int_0^t G_{12}\, dM_{12}, \int_0^t H_{12}\, dM_{12}, \int_0^t H_{13}\, dM_{13}, \int_0^t H_{24}\, dM_{24}\right),$$

where the G and H's are certain stochastic processes (depending on n, of course), and then, by applying the delta method and making some linear transformations, arriving at the weak-convergence result we seek.

First we state the following lemmas, the first of which is proven in Chung (1968, page 90), and the second of which is a Fubini theorem for stochastic integrals (proof omitted).

LEMMA 2. *Let X_n be a sequence of uniformly integrable random variables which converges in probability to a constant d. Then $E[X_n] = d + o(1)$.*

LEMMA 3. *For each $t \in [0, 1]$, let $h(\cdot, \cdot, t)$: $[0, 1]^2 \to (-\infty, \infty)$ be square integrable with respect to Lebesgue measure on $[0, 1]^2$. If W is a Wiener process on $[0, 1]$, then*

$$\int_0^t \int_0^t h(u, s, t)\, dW(u)\, ds = \int_0^t \int_0^t h(u, s, t)\, ds\, dW(u).$$

The processes we associate with (A.1) are

$$G_{12}(t, s) = n^{1/2}\hat{F}_1(s-)F_2(t-s)K_1(s)Y_1^{-1}(s),$$

$$H_{12} = n^{1/2}K_1 Y_1^{-1},\ H_{13} = n^{1/2}K_1 Y_1^{-1}, \text{ and } H_{24} = n^{1/2}K_2 Y_2^{-1}.$$

Then $Z_n = n^{1/2}(\hat{R}_2 - R_2^+, \hat{A}_{12} - A_{12}^+, \hat{A}_{13} - A_{13}^+, \hat{A}_{24} - A_{24}^+)$, where $A_{ij}^+(t) = \int_0^t K_i a_{ij}\, d\ell$.

In what follows, let $p_1(t)$ and $p_2(t)$ be two functions that are bounded away from 0 on $[0, c]$, where $c < 1$.

LEMMA 4. *Assume that for $t \in [0, c]$*

I. $Y_i(t)/n = p_i(t) + o_p(1)$ for $i = 1, 2$.

II. $nK_i(t)Y_i(t)^{-1}$ *is uniformly integrable in* (n, t, i).

Then on $[0, c]$

$$Z_n(t) \to_d \left(\int_0^t g_{12}(t, \cdot)\, dW_{12}, \int_0^t h_{12}\, dW_{12}, \int_0^t h_{13}\, dW_{13}, \int_0^t h_{24}\, dW_{24}\right),$$

as a process in t, where the W_{ij}, g_{12}, h_{12}, h_{13} *and* h_{24} *are defined in Theorem* 4.

PROOF. The result would be an application of Aalen and Johansen's (1978) Theorem 4.1 except that $G_{12}(t, s)$ is a function of t as well as s and that two of the integrals in Z_n are taken with respect to the same martingale. To circumvent these problems, consider instead the weak convergence of $Z_n'(t) = (\int_0^t H(\mathbf{a}, \mathbf{u}, \mathbf{b}, \mathbf{v}, s)\, dM_{12}(s), \int_0^t H_{13}\, dM_{13}, \int_0^t H_{24}\, dM_{24})$, where $H(\mathbf{a}, \mathbf{u}, \mathbf{b}, \mathbf{v}, s) = \sum a_i I\{s \le u_i\} G_{12}(u_i, s) + \sum b_j I\{s \le v_j\} H_{12}(s)$, the a_i and b_i are arbitrary real numbers, the u_i and v_j are in $[0, 1]$, and the sums are finite. Then the conditions of Aalen and Johansen's theorem are satisfied for Z_n', and, as a sequence of processes,

$$Z_n'(t) \to_d \left(\sum a_i \int_0^{u_i \wedge t} g_{12}(u_i, \cdot)\, dW_{12} \right.$$
$$\left. + \sum b_j \int_0^{v_j \wedge t} h_{12}\, dW_{12}, \int_0^t h_{13}\, dW_{13} \int_0^t h_{24}\, dW_{24}\right).$$

By the Cramér-Wold argument, the finite-dimensional distributions of Z_n converge weakly to the finite-dimensional distributions asserted by the theorem. But Z_n is also tight. Its last three coordinates clearly are because, by the above, they converge weakly to $(\int h_{12}\, dW_{12}, \int h_{13}\, dW_{13}, \int h_{24}\, dW_{24})$. Now let $G_{12}'(s) = G_{12}(t, s)/F_{2;t}(s)$, $M_{12}'(t) = \int_0^t G_{12}'\, dM_{12}$. By the change of variables formula, the first coordinate is $F_2(0)M_{12}'(t) - F_2(t)M_{12}'(0) - \int_0^t M_{12}'\, dF_{2;t}$. The conditions above insure that M_{12}' converges weakly, so the continuous mapping theorem implies that the first coordinate also converges weakly, and is hence tight. □

LEMMA 5. *If the conditions of Lemma* 4 *hold and if* $n^{1/2} \int_0^1 (1 - K_i) a_{ij}\, d\ell = o_p(1)$ *for each* (i, j), *then, as processes in* $t \in [0, c]$,

$$n^{1/2}(\tilde{R}_2(t) - R_2^+(t), \tilde{F}_1(t) - F_1(t), \tilde{F}_2(t) - F_2(t))$$
$$\to_d \left(\int_0^t g_{12}(t, \cdot)\, dW_{12}, -\tilde{F}_1(t) \int_0^t (h_{12}\, dW_{12} + h_{13}\, dW_{13}), \right.$$
$$\left. - F_2(t) \int_0^t h_{24}\, dW_{24}\right).$$

PROOF. The condition added in this corollary insures that we may replace A_{ij}^+ in Z_n by A_{ij}. Two applications of the delta method yield the desired result after the second and third processes of Z_n have been summed. □

LEMMA 6. *If the conditions of Lemma 5 hold, then*

$$n^{1/2}(\tilde{R}_2(t) - R_2^+(t), R^+(t) - R_2^+(t), R_2^+(t) - R_{12}^+(t))$$
$$\to_d \left(\int_0^t g_{12}(t, \cdot)\ dW_{12}, - \int_0^t F_1(s)F_2(t-s)\left[\int_0^{t-s} h_{24}\, dW_{24}\right] \times a_{12}(s)\ ds, - \int_0^t F_1(s)X(s)S_2(t-s)a_{12}(s)\ ds \right)$$

as processes in t, where $X(s) = \int_0^s h_{12}\, dW_{12} + \int_0^s h_{13}\, dW_{13}$.

PROOF. The conclusion follows by noting that the K_i terms become 1 in the limit, by summing the second and third processes of Z_n, by applying the delta method several times, (eg, $\hat{A}_{24} - A_{24}$ to $\hat{F}_2 - F_2$), by noting that $\hat{F}_{1-} = F_1 + o_p(1)$, and by employing some continuous maps (eg, $\hat{F}_2 - F_2$ to $R_2^+ - R_{12}^+$). □

The main assertion of Theorem 4 now follows: the sum of the three terms of Lemma 6 is, by Lemma 1, asymptotically equivalent to $n^{1/2}(\hat{R} - R)$, noting that (I) of Lemma 1 is implied by (I) and (II) of this theorem via Lemma 2. That the order of integration may be reversed follows from Lemma 3. □

PROOF OF COROLLARY 1. The first part of the corollary follows directly from Theorem 4, so we only need to show the convergence of estimated terms to their counterparts in Corollary 1. By using the submartingale inequality and Lemma 2 of the Appendix, conditions (I) and (II) insure that estimates of the terms in the brackets converge uniformly to the correct terms. It thus suffices to show that $\sup_t | \int_0^t K_i Y_i^{-2} n\, dN_{ij} - \int_0^t (a_{ij}/p_i)\, d\ell\, | = o_p(1)$. This difference can be written as

$$\int_0^t K_i Y_i^{-2} n\, dM_{ij} + \int_0^t (K_i Y_i^{-1} n - p_i^{-1}) a_{ij}\, d\ell.$$

By Theorem 2, the variance of the first term is less than $\int_0^1 E[K_i Y_i^{-3} n^2] a_{ij}\, d\ell$. By using (I) and (II) with Lemma 2 and then applying the Dominated Convergence Theorem, this bound on the variance converges to 0, so the first term is $o_p(1)$ uniformly in t. That the second term is also $o_p(1)$ uniformly in t follows from (I), (II), and Lemma 2. □

PROOF OF COROLLARY 3. It sufficies to show that (I) of Theorem 4 and the added condition of Corollary 1 are satisfied, since these imply (II) and (III) of Theorem 4. Let G be the survival function associated with each C_k. Then the distribution of $Y_1(t)$ is binomial with parameters n and $F_1(t)G(t)$, so $Y_1(t)/n = F_1(t)G(t) + o_p(1)$ and, by assumption, $F_1(t)G(t)$ is bounded away from 0 for $0 \leq t \leq 1$; thus, $p_1(t)$ is identified. The distribution of $Y_2(t)$ is also binomial, with parameters n and $\int_0^{1-t} F_1(s)a_{12}(s)F_2(t)G(t+s)\, ds$. By the condition made on a_{12}, this integral is bounded away from 0 for all t in $[0, c]$, so $Y_2(t)/n = p_2(t) + o_p(1)$. Thus (I) of Theorem 4 is satisfied.

Now let the family of random variables $\{X_n(t)\}$, $t \in [0, c]$, $n = 1, 2, \cdots$ have distributions $X_n(t) \sim$ Binomial $(n, p(t))$ where $\inf_{t\in[0,c]} p(t) = p_0 > 0$. To prove the added condition of Corollary 1 holds, it is enough to show that $E[n^3 I\{X_n(t) > 0\} X_n^{-3}(t)]$ is bounded uniformly in (n, t) (eg, Neveu, 1965, page 54). That this expected value is indeed uniformly bounded follows by an extension of an idea, used by Aalen (1976), which shows that a uniform bound is $p_0^{-3} + 115p_0^{-4}$. □

Acknowledgment. The authors wish to thank Tom Kurtz for numerous helpful suggestions. We are also grateful to the referee for a very careful reading of the original version.

REFERENCES

AALEN, O. (1975). Statistical inference for a family of counting processes. Ph.D. dissertation, University of California, Berkeley.

AALEN, O. (1976). Nonparametric inference in connection with multiple decrement models. *Scandinavian J. Statist.* **3** 15–27.

AALEN, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Math. Statist.* **6** 701–726.

AALEN, O. (1980). A model for nonparametric regression analysis of counting processes. *Springer Lecture Notes in Statistics* **2** 1–25.

AALEN, O. and JOHANSEN, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian J. Statist.* **5** 141–150.

ANDERSEN, P. K., BORGAN, Ø. GILL, R. D., and KEIDING, N. (1982). Linear nonparametric inference for comparison of counting processes, with applications to censored survival data. *Int. Statist. Rev.* **50** 219–258.

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

BEGG, C. and LARSON, M. (1982). A study of the use of the probability of being in response function as a summary of tumor response data. *Biometrics* **38** 59–66.

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

BRESLOW, N. and CROWLEY, J. (1974). A large-sample study of the life table and product limit estimates under random censorship. *Ann. Math. Statist.* **2** 437–453.

BRÉMAUD, P. and JACOD, J. (1977). Processus ponctuels et martingales: resultats récent sur le modélisation et filtrage. *Adv. Appl. Probability* **9** 362–416.

CHUNG, K. L. (1968). *A Course in Probability Theory.* Harcourt, New York.

COX, D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

DOLÉANS-DADÉ, C. and MEYER, P. (1970). Intégrales Stochastiques par Rapport aux Martingales Locales. *Séminaire de Probabilitiés IV Lecture Notes in Mathematics* **124** 77–80. Springer, New York.

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

FLEMING, T. R. and HARRINGTON, D. P. (1979). A class of hypothesis tests for one and two sample censored survival data. *Comm. in Statist., Theory and Methods* **A10(8)** 763–794.

FLEMING, T. R. O'FALLON, J. R, O'BRIEN, P. C. and HARRINGTON, D. P. (1980). Modified Komogorov-Smirnov test procedures with application to arbitrarily right-censored data. *Biometrics* **36** 607–625.

GEHAN, E. (1965). A generalized Wilcoxon test for comparing arbitrarily single censored samples. *Biometrika* **52** 203–223.

GILL, R. D. (1980). Nonparametric estimation based on censored observations of a Markov renewal process. *Z. Wahrsch. verw. Gebiete* **53** 97–116.

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

LAGAKOS, S. W., SOMMER, C. J. and ZELEN, M. (1978). Semi-Markov models for partially censored data. *Biometrika* **65** 311–317.

MANTEL, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports* **50** 163–170.

NELSON, W. (1969). Hazard plotting for incomplete failure data. *J. Qual. Tech.* **1** 27–52.

NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability.* Holden, San Francisco.

TEMKIN, N. (1978). An analysis for transient states with application to tumor shrinkage. *Biometrics* **34** 571–580.

VOELKEL, J. G. (1980). Multivariate counting processes and the probability of being in response function. Ph.D. thesis, Department of Statistics, University of Wisconsin-Madison.

ALLIED CHEMICAL CORPORATION
SYRACUSE RESEARCH LABORATORY
MILTON AVENUE
P.O. BOX 6
SOLVAY, NEW YORK 13209

EPIDEMIOLOGY AND BIOSTATISTICS ZD-08
THE FRED HUTCHINSON CANCER RESEARCH CENTER
1124 COLUMBIA STREET
SEATTLE, WASHINGTON 98104