The Annals of Statistics
1986, Vol. 14, No. 4, 1351–1365

# NONPARAMETRIC ESTIMATION OF A BIVARIATE SURVIVAL FUNCTION IN THE PRESENCE OF CENSORING[1]

By Wei-Yann Tsai, Sue Leurgans and John Crowley

*Brookhaven National Laboratory, Ohio State University, and The Fred Hutchinson Cancer Research Center and University of Washington*

A new family of estimators of a bivariate survival function based on censored vectors is obtained from a decomposition of the bivariate survival function. These estimators are uniformly consistent under bivariate censoring and are self-consistent under univariate censoring.

**1. Introduction.** Methods for analyzing univariate censored data have been studied by many people over the last few decades. Relatively little research has been devoted to the analysis of bivariate observations in the presence of censoring. In some studies, two times are observed because experimental units consist of pairs of components, and the lifetime of each component is recorded. Censoring occurs when the experimental unit is removed from the study before both components have been observed to fail. Examples include twin studies and matched pair studies. In other studies, two failure times are recorded for each individual or piece of equipment. Thus, in studies of chronic diseases, both recurrence times and death times are recorded, and their joint distribution needs to be estimated.

In this paper, we discuss nonparametric estimation of the bivariate survival function in the presence of censoring. In Section 3, we present a new class of estimators of the bivariate survival function. These estimators are based on a decomposition given in Section 2 of the bivariate survival function in terms of estimable functions. Sections 3 and 4 discuss properties of the estimators. The rest of this section summarizes previous work on this problem.

We first review estimation of a univariate survival distribution. In this case, let $T_i^0$, $i = 1, \dots, n$, be $n$ independent and identically distributed (iid) lifetimes with continuous survival function $S^0(t) = P(T_i^0 > t)$. Let $C_i$, $i = 1, \dots, n$, be an independent sample of $n$ censoring variables with survival function $H(t)$. It is not possible to observe both $T_i^0$ and $C_i$. Let $X \wedge Y$ denote $\min(X, Y)$ and $[A]$ denote the indicator of the event $A$. We observe $T_i$ and $D_i$, where $T_i = T_i^0 \wedge C_i$

Received December 1983; revised March 1986.

[1]Research supported in part by National Institute of Health Grants 5-R01-CA18332 at the University of Wisconsin-Madison and 2-R01-GM28314 at the Fred Hutchinson Cancer Research Center and in part by funds provided by the International Cancer Research Data Bank Program of the National Cancer Institute, NIH, under contract No. N01-CO-65341 (International Cancer Research Technology Transfer—ICRETT) and by the International Union Against Cancer. This manuscript has been authored under contract DE-AC-02-76CH00016 with the U.S. Department of Energy. Accordingly, the U.S. government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. government purposes.

*AMS* 1980 *subject classifications*. Primary 62G05; secondary 62P10.

*Key words and phrases*. Self-consistency, univariate and bivariate censored data, decomposition.

and $D_i = [T_i^0 \le C_i]$. Kaplan and Meier (1958) suggested

$$\hat{S}^0(t) = \prod_{T_i \le t} \left(1 - \frac{D_i}{\sum_j [T_j \ge T_i]}\right)$$

as an estimator of $S^0$. They showed that $\hat{S}^0$ is a nonparametric maximum likelihood estimator in the sense that $\hat{S}^0$ formally maximizes an expression that would be a likelihood function if a parameter were being estimated. Johansen (1978) showed that $\hat{S}^0$ is a generalized maximum likelihood estimator (GMLE) as defined by Kiefer and Wolfowitz (1956). Efron (1967) showed that $\hat{S}^0$ is the unique self-consistent estimator.

Several extensions of the product–limit estimator to bivariate times have been proposed. Before reviewing these, we extend the notation to bivariate times. The true pair of survival times will be denoted by $(T_1^0, T_2^0)$. The bivariate survival function of this vector is $S^0(t_1, t_2) = P(T_1^0 > t_1, T_2^0 > t_2)$. The pair of censoring times is $(C_1, C_2)$ and has bivariate survival function $H$. The observed vector is $(T_1, T_2, D_1, D_2)$, where $T_j = T_j^0 \wedge C_j$, $D_j = [T_j^0 \le C_j]$, $j = 1, 2$. When $P(C_1 = C_2) = 1$, the censoring will be referred to as univariate censoring. Otherwise, the censoring will be called bivariate censoring. (For discussion of these censoring mechanisms, see Leurgans, Tsai and Crowley (1982).)

Hanley and Parnes (1983) and Campbell (1981) studied maximum likelihood estimators for discrete data and extended self-consistency to this situation. Muñoz (1980) defined self-consistency for estimators of continuous bivariate distributions in the presence of univariate censoring. Unfortunately, the GMLE is not always unique. For example, if we observe one pair of uncensored times, say $(T_{11}, T_{21})$, and a data pair with first coordinate censored, say $(T_{12}, T_{22})$, where $T_{12} > T_{11}$, then any distribution that assigns probability $\frac{1}{2}$ to the point $(T_{11}, T_{21})$ and probability $\frac{1}{2}$ to the half-line from $(T_{12}, T_{22})$ to $(\infty, T_{22})$ is a GMLE and a self-consistent estimator.

A related nonuniqueness problem occurs whenever any of the pairs are censored in exactly one coordinate and the underlying bivariate distribution is jointly continuous. The continuity condition implies that with probability 1, none of the observed lifetimes is tied. The total mass given to a ray on which one of the one-component censored observations can lie is determined by the GML or the self-consistency conditions, but the distribution of that mass within such rays is not determined. Of course, some of the values of the survival functions will be changed if the distribution of probability measure within these rays is changed. Leurgans, Tsai and Crowley (1982) pointed out that some sequences of GMLE's of the survival function, such as the sequence in which all mass on all rays is concentrated at the finite endpoints of the ray and the sequence in which all mass on all rays is put near the infinite end of the ray, do not converge to the true survival function. A referee pointed out that GML does not require continuity of the conditional distributions of one component given the other. This freedom from assumptions was adequate in one dimension, but does not permit consistent estimation in more dimensions.

Campbell and Földes (1982) proposed two other estimators. One is based on the equation $S^0 = (t_1, t_2) = S^0(t_1, 0)P(T_2^0 > t_2 | T_1^0 > t_1)$. Each term in this product is estimated separately. The other is based on an estimator $\hat{R}(t_1, t_2)$ of the cumulative hazard function $R = -\ln S^0$, which is a line integral of the gradient of $-\ln S^0$. The survival function is then estimated by $\exp - \hat{R}(t_1, t_2)$. They show the estimator $\hat{R}$ is path dependent and point out another serious weakness: both of their estimators of $S^0(t_1, t_2)$ may fail to be survival functions. Campbell (1982) showed the weak convergence of these estimators. Because of these weaknesses, we propose a family of closed form estimators that is always survival functions, based on a decomposition formula presented in Section 2 and some smoothing techniques. Although the estimators we propose are kernel and bandwidth dependent, this arbitrariness does not seem to matter as much as the path dependence.

**2. Decomposition of bivariate survival functions.** Peterson (1977) introduced a decomposition of a univariate survival function in terms of identifiable survival and subsurvival functions. Theorem 2.1 below is a bivariate analogue of Peterson's decomposition. In Section 3 we present estimators of a bivariate survival function based on this decomposition.

Throughout the rest of this paper, we give formulas for $t_1 \geq t_2$. Definitions for $t_1 < t_2$ are obtained by reversing the coordinates.

We use two assumptions, (A1) and (A2), to derive the decomposition.

(A1) The vectors $(T_1^0, T_2^0)$ and $(C_1, C_2)$ are mutually independent.
(A2) The functions $S^0$ and $H$ are absolutely continuous with respect to Lebesgue measure on $R^2$.

Without some assumptions about the relationship between $(T_1^0, T_2^0)$ and $(C_1, C_2)$, $S^0$ is not identifiable. The assumption (A2) can be weakened, but is convenient for exposition.

The decomposition is expressed in terms of the following functions and sets:

$$\begin{aligned} S(x, y) &= P(T_1 > x, T_2 > y), \\ S_3^0(y) &= S^0(y, y), \\ S_2(x, y) &= P(T_1 > x, T_2 > y, D_2 = 1), \\ S_{12}(x, y) &= P(T_1 > x, T_2 > y, D_1 = 0, D_2 = 1), \\ S^0(x|y) &= P\left(T_1^0 > x | T_2^0 = y\right), \\ R(s, t) &= \{(x, y) | x > s \geq y > t\}, \\ \Delta(s, t) &= \{(x, y) | s \geq x \geq y > t\}. \end{aligned} \tag{2.1}$$

That is, $S$, $S_2$ and $S_{12}$ are the observable bivariate (sub)survival functions, $S^0(\cdot|\cdot)$ is the conditional survival function and $S_3$ is the probability that neither event has occurred.

Lemma 2.1 is a preliminary decomposition of $S^0(s, t)$. It is not a complete decomposition in terms of identifiable functions, because $S_3^0(y)$ and $S^0(x|y)$ cannot be estimated directly. However, since $S_3^0$ and $S^0(\cdot|y)$ are univariate survival functions, the univariate decomposition given in Lemma 2.2 below applies.

LEMMA 2.1. *Assume conditions* (A1) *and* (A2) *hold and let* $s > t > 0$ *be such that* $S(s, s) > 0$. *Then*

$$
\begin{aligned}
S^0(s,t) &= S_3^0(s) + \iint_{R(s,t)} S_3^0(y)/S(y,y)\,DS_2(x,y) \\
&\quad + \iint_{\Delta(s,t)} \frac{S_3^0(y)}{S(y,y)} \frac{S^0(s\mid y)}{S^0(x\mid y)} DS_{12}(x,y),
\end{aligned} \tag{2.2}
$$

*where* $0/0 = 0$, *the integrals are Riemann–Stieltjes integrals, well defined since both* $S_2$ *and* $S_{12}$ *have bounded variation, and* $D$ *is the differential operator.*

Lemma 2.1 is easier to appreciate if Figure 1 is examined. The probability $S^0$ assigned to the rectangle $R(s,t)$ can be split according to whether the first coordinate is censored before $s$ and written as the sum of $P\{(T_1^0, T_2^0) \in R(s,t), D_1 = 0, T_1 \le s\}$ and $P\{(T_1^0, T_2^0) \in R(s,t); T_1 > s\}$. The first probability is absolutely continuous with respect to the identifiable subdistribution $S_{12}$ on the triangle $\Delta(s,t)$, and is displayed as an integral against that measure in Lemma 2.1. Similarly, the second probability is written as an integral over $R(s,t)$ with respect to the identifiable subdistribution $S_2$. The factor $S_3^0(y)/S(y,y)$ is the reciprocal of the probability that $C_1 \wedge C_2$ is greater than $y$; the factor $S^0(s\mid y)/S^0(x\mid y)$ in the former integral is the conditional probability that $T_1^0$ is in the rectangle given that the second coordinate is $y$, which forces $(T_1^0, T_2^0)$ to be in $\Delta(s,t) \cup R(s,t)$.

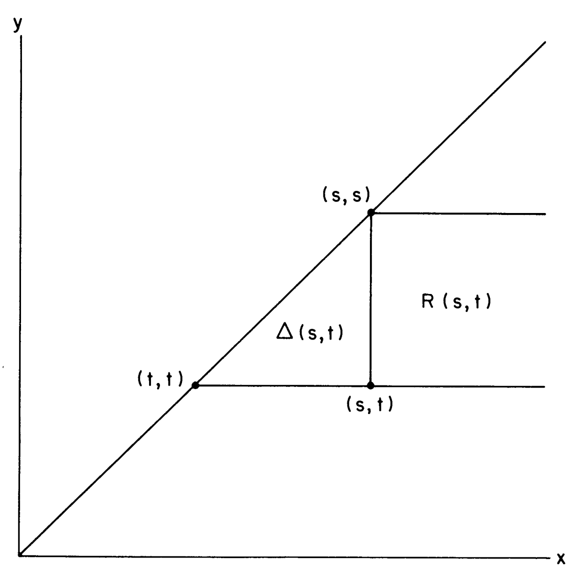


FIG. 1.

**Proof.** We prove this lemma by rewriting each of the double integrals as integrals with respect to $DS^0(x, y)$. For the first integral, we interchange limits to show that this integral is equal to

$$-\int_t^s \frac{S_3^0(y)}{S(y, y)} D_y S_2(s, y).$$

Since the definitions and (A1) imply that $S(y, y) = S_3^0(y)H(y, y)$ and that $D_y S_2(s, y) = H(s, y) D_y S^0(s, y)$, this integral reduces to

$$-\int_t^s \frac{H(s, y)}{H(y, y)} D_y S^0(s, y) = \iint_{R(s, t)} \frac{H(s, y)}{H(y, y)} DS^0(x, y). \tag{2.3}$$

Similarly, the fact that $DS_{12}(x, y) = D_x H(x, y) D_y S^0(x, y)$ and $S^0(x|y) = D_y S^0(x, y)/D_y S^0(-\infty, y)$ implies that the second integral of (2.2) is

$$\int_{y=t}^s \int_{x=y}^s \frac{1}{H(y, y)} \frac{D_y S^0(s, y)}{D_y S^0(x, y)} D_x H(x, y) D_y S^0(x, y)$$

$$= -\int_{y=t}^s \frac{H(y, y) - H(s, y)}{H(y, y)} D_y S^0(s, y) \tag{2.4}$$

$$= \iint_{R(s, t)} \left[1 - \frac{H(s, y)}{H(y, y)}\right] DS^0(x, y),$$

where the last integral follows from the fact that $-D_y S^0(s, y)$ is the integral of $DS^0(x, y)$ over $x \in [s, \infty)$.

Therefore, the sum of the two integrals is the sum of (2.3) and (2.4) or

$$\iint_{R(s, t)} DS^0(x, y) = S^0(s, t) - S^0(s, s) = S^0(s, t) - S_3^0(s).$$

The lemma follows. □

Lemma 2.2 is Beran's (1981) extension of Peterson's (1977) decomposition to allow simultaneous discontinuities in death and censoring times. The statement of the decomposition in terms of the product integral emphasizes the similarity of the theoretical decomposition and the estimators of Section 3.

The product integral of a function $g$ is defined by

$$[\gamma(g)](t) = \lim_{\max_{1 \le k \le r} (u_k - u_{k-1}) \to 0} \prod_{i=1}^r \{1 - (g(u_i) - g(u_{i-1}))\},$$

where $0 = u_0 < u_1 < \cdots < u_r = t$.

For continuous functions $g$, $[\gamma(g)](t) = \exp(-g(t))$. If $g$ is an empirical cumulative hazard, $[\gamma(g)](t)$ is the corresponding product–limit or Kaplan–Meier estimator.

LEMMA 2.2. *If* $S_0(t) = P(T^0 > t)$, $S(t) = P(T^0 \wedge C > t)$, *and* $S_u(t) = P(C > T^0 > t)$, *where* $T^0$ *and* $C$ *are independent random variables, then*

$$S_0(t) = \left[\gamma\left(-\int_0^{\mu+} \frac{DS_u(x)}{S(x-)}\right)\right](t).$$

The function $S_3^0$ is the survival function of $T_3^0 = T_1^0 \wedge T_2^0$. Define the corresponding censoring time $C_3 = C_1 \wedge C_2$, the observed time $T_3 = T_3^0 \wedge C_3$, and the indicator $D_3 = [T_3^0 \le C_3]$. The survival function $S_3^0$ can be decomposed in terms of $S_3(t) = P(T_3 > t)$ and $S_{3u}(t) = P(T_3 > t, D_3 = 1)$. Since $T_3 = T_1 \wedge T_2$ and $D_3 = [T_1 > T_2]D_2 + [T_1 < T_2]D_1$, $(T_3, D_3)$ is a function of $(T_1, T_2, D_1, D_2)$ and $S_3$ and $S_{3u}$ can be estimated empirically.

We now state the decomposition.

THEOREM 2.1. *If the conditions* (A1) *and* (A2) *are met, then*

$$\begin{aligned} S^0(s,t) = {} & \left[\gamma\left(-\int_0^{\mu^+} \frac{DS_{3u}(z)}{S_3(z-)}\right)\right](s) \\ & + \iint_{R(s,t)} \left\{\left[\gamma\left(-\int_0^{\mu^+} \frac{DS_{3u}(z)}{S_3(z)}\right)\right](y-)\right\} \Big/ [S(y-,y-)]\, DS_2(x,y) \\ & + \iint_{\Delta(s,t)} \left\{\left[\gamma\left(-\int_0^{\mu+} \frac{DS_{3u}(z)}{S_3(z-)}\right)\right](y-)\right\} \Big/ [S(y-,y-)] \\ & \times \left\{\left[\gamma\left(-\int_0^{\mu+} \frac{D_z S_u(z|y^-)}{S(z-|y-)}\right)\right](s)\right\} \Big/ \\ & \left\{\left[\gamma\left(-\int_0^{\mu^+} \frac{D_z S_u(z|y-)}{S(z-|y-)}\right)\right](x-)\right\} DS_{12}(x,y), \end{aligned} \tag{2.5}$$

*where* $S_u(t|y) = P(T_1 > t, D_1 = 1|T_2 = y, D_2 = 1)$ *and* $S(t|y) = P(T_1 > t|T_2 = y, D_2 = 1)$.

PROOF. From the assumption (A1), $C_2$ is independent of $T_1^0$, we observe

$$\begin{aligned} S^0(t|y) &= P(T_1^0 > t|T_2^0 = y) \\ &= P(T_1^0 > t|T_2^0 = y, C_2 > y) \\ &= P(T_1^0 > t|T_2 = y, D_2 = 1). \end{aligned}$$

The theorem follows from applying Lemma 2.2 to $S_3$ and to $P(T_1^0 > t|T_2 = y, D_2 = 1)$ and substituting the resulting representations in Lemma 2.1. □

REMARK 2.1. Under the continuity assumptions of (A2), (2.5) reduces to (2.2). We prefer (2.5) because (A2) is not required for Lemma 2.1 or Theorem 2.1

and because (2.5) can be applied to empirical subsurvival functions to obtain estimators.

**3. Estimators of $S^0$.** Suppose the iid random vectors $\{(T_{1i}, T_{2i}, D_{1i}, D_{2i}), i = 1, \dots, n\}$ have the same distribution as the random vector $(T_1, T_2, D_1, D_2)$. In this section we develop an estimator of $S^0$ and establish its consistency.

Natural unbiased estimators of the (sub)survival functions in (2.5) are defined below in terms of $T_{1i}, T_{2i}, D_{1i}, D_{2i}, T_{3i} = T_{1i} \wedge T_{2i}$ and $D_{3i} = [T_{1i} > T_{2i}]D_{2i} + [T_{1i} < T_{2i}]D_{1i}$:

$$S^e(x, y) = \frac{1}{n} \sum_i [T_{1i} > x, T_{2i} > y],$$

$$S^e_{12}(x, y) = \frac{1}{n} \sum_i [T_{1i} > x, T_{2i} > y, D_{1i} = 0, D_{2i} = 1],$$

$$S^e_2(x, y) = \frac{1}{n} \sum_i [T_{1i} > x, T_{2i} > y, D_{2i} = 1],$$

$$S^e_3(x) = \frac{1}{n} \sum_i [T_{3i} > x],$$

and

$$S^e_{3u}(x) = \frac{1}{n} \sum_i [T_{3i} > x, D_{3i} = 1].$$

Substituting $S^e_3$ and $S^e_{3u}$ into Lemma 2.2, we have the Kaplan–Meier estimator for $S^0_3(t)$:

$$\hat{S}^0_3(t) = \left[\gamma\left(-\int_0^{\mu^+} \frac{DS^e_{3u}(x)}{S^e_3(x-)}\right)\right](t). \tag{3.1}$$

The functions $S(x|y)$ and $S_u(x|y)$ are the conditional probabilities given $T_{2i} = y$ and $D_{2i} = 1$. Since the assumption of absolute continuity implies that there is (a.s.) at most one $T_{2i}$ that is equal to $y$ with $D_{2i} = 1$, $S(x|y)$ and $S_u(x|y)$ cannot be estimated stably without smoothing. To estimate a conditional survival function given $y$, we apply the nonnegative weight $W_{ni}(y)$ to $(T_{1i}, T_{2i})$. The weight $W_{ni}(y)$ depends on the data through the distance between $y$ and $T_{2i}$ and through the second components $\{(T_{2j}, D_{2j}), j = 1, \dots, n\}$. With the assumption that $\sum_i W_{ni}(y) D_{2i} = 1$, the following estimators are discrete subsurvival functions:

$$\begin{aligned} \hat{S}(x|y) &= \sum_i W_{ni}(y)[T_{1i} > x, D_{2i} = 1], \\ \hat{S}_u(x|y) &= \sum_i W_{ni}(y)[T_{1i} > x, D_{1i} = D_{2i} = 1]. \end{aligned} \tag{3.2}$$

This class of estimators has several attractive features. One feature is that the use of Theorem 3.1 to construct the estimators makes it a natural extension of the product–limit estimator to two dimensions. In Section 4 we show that these estimators are self-consistent under univariate censoring. Another advantage is that, once the weights have been computed, no iteration is required.

Substituting the estimators $\hat{S}(x|y)$ and $\hat{S}_u(x|y)$ into the equation yields the following natural estimator for $\hat{S}^0(x|y)$:

$$\hat{S}^0(x|y) = \left[\gamma\left(-\int_0^{\mu^+} \frac{D_t\hat{S}_u(t|y)}{\hat{S}(t|y)}\right)\right](x).$$

If $\hat{S}(x|y) > 0$, and if the jump points of $\hat{S}_u(\cdot|y)$ are $x_1, \dots, x_m$, then

$$\hat{S}^0(x|y) = \prod_{x_i \le x}\left[1 - \frac{\hat{S}_u(x_i^-|y) - \hat{S}_u(x_i|y)}{\hat{S}(x_i^-|y)}\right].$$

Substituting $\hat{S}_3^0$, $S_2^e$, $S_{12}^e$, $S^e$ and $\hat{S}^0(x|y)$ into (2.2), we obtain the following estimator of $S^0$:

$$\begin{aligned}\hat{S}^0(s,t) &= \hat{S}_3^0(s) + \iint_{R(s,t)} \hat{S}_3^0(y^-)/S^e(y^-,y^-)\,DS_2^e(x,y) \\ &\quad + \iint_{\Delta(s,t)} \hat{S}_3^0(y^-)\hat{S}^0(s|y^-)/(S^e(y^-,y^-)\hat{S}^0(x^-|y^-))\,DS_{12}^e(x,y).\end{aligned} \tag{3.3}$$

If we rewrite the double integral in the above equation as a double sum, we can easily see that the estimator is a step function. Furthermore, $\hat{S}^0(\infty,\infty) = 0$, $\hat{S}^0(0,0) = 1$ and the probability mass at the point $(s,t)$ is

$$\begin{aligned}&\hat{S}^0(s,t) + \hat{S}^0(s^-,t^-) - \hat{S}^0(s^-,t) - \hat{S}^0(s,t^-) \\ &= \frac{1}{n}\left[\frac{\hat{S}_3^0(t^-)}{S^e(t^-,t^-)}\right]\sum_{i=1}^n D_{1i}D_{2i}[T_{1i} = s, T_{2i} = t] \\ &\quad + \frac{1}{n}\left[\frac{\hat{S}_3^0(t^-)}{S^e(t^-,t^-)}\right][\hat{S}^0(s^-|t^-) - \hat{S}^0(s|t^-)] \\ &\quad \times \sum_{i=1}^n \frac{(1-D_{1i})D_{2i}}{\hat{S}^0(T_{1i}^-|t^-)}[t \le T_{1i} < s, T_{2i} = t] \\ &\ge 0.\end{aligned}$$

Therefore, $\hat{S}^0$ corresponds to a genuine probability measure.

We devote the rest of this section to a discussion of the consistency of this class of estimators. First we discuss the conditions on the weight functions that imply consistent estimation of the conditional survival functions. Lemma 3.1 asserts that these conditions imply consistent estimation of $S^0(\cdot|y)$. Theorem 3.1 gives conditions for the consistency of $\hat{S}^0$.

Choice of an estimator within this class requires the specification of the weight functions $W_{ni}(y)$. We confine our investigation to a demonstration that the weight functions can be chosen to provide a consistent sequence of estimators. Of course, important issues remain in the selection of weight functions for use with specific sample sizes.

Here we focus on kernel weights. These weights are constructed by selecting a nonnegative function $k(\cdot)$ of bounded variation on the real line and a sequence of positive bandwidths $\{h(n), n \ge 1\}$ converging to zero. The probability weights are then

$$W_{n,i}(y) = k((T_{2i} - y)/h(n))D_{2i} \Big/ \Big( \sum_j k((T_{2j} - y)/h(n))D_{2j} \Big). \tag{3.4}$$

That is, we give positive weight only to those observations with an observed failure in the second component near $y$. Theorem A.1 of the Appendix (also see Theorem 4.3 of Tsai (1982)) shows that if the true conditional survival functions $S(x|y)$ are uniformly continuous in $x$ and in $y$ and if bandwidths converge slowly enough that $\Sigma_{n=1}^{\infty}\exp(-rh^2(n)) < \infty$ for every positive $r$, then

$$\begin{aligned} &\sup_{0 \le x,\, y \le M} |\hat{S}(x|y) - S(x|y)| \to 0, \quad \text{a.s. and} \\ &\sup_{0 \le x,\, y \le M} |\hat{S}_u(x|y) - S_u(x|y)| \to 0, \quad \text{a.s. as } n \to \infty, \end{aligned} \tag{3.5}$$

where $M$ is such that $-H(t, y)[\partial S^2(t, y)/\partial y] > 0$ for every $\max(t, y) \le M$.

In our proofs, we do not assume that kernel weights have been used. Nearest neighbor weights can also be used. Beran (1981), in related work on nonparametric regression in the presence of censoring, discusses these possibilities.

LEMMA 3.1. *If* (A1) *holds, then*

$$S^0(x|y) = S(x|y) - \int_0^x \frac{S^0(x|y)}{S^0(z|y)} D_z S_c(z^-|y) \qquad \textit{for } x \ge y \ge 0, \tag{3.6}$$

*where*

$$S_c(x|y) = S(x|y) - S_u(x|y) = P(T_1 > t, D_1 = 0|T_2 = y, D_2 = 1).$$

PROOF. Assumption (A1) implies

$$\begin{aligned} S^0(x|y) &= E\big([T_1^0 > x]|T_2^0 = y\big) \\ &= E\big(E\big([T_1^0 > x]|T_2^0 = y, T_1, D_1\big)|T_2^0 = y\big) \\ &= -\int_0^{\infty} E\big([T_1^0 > x]|T_2^0 = y, T_1 = z, D_1 = 1\big) \\ &\qquad \times D_z P\big(T_1 > z^-, D_1 = 1|T_2^0 = y\big) \\ &\quad - \int_0^{\infty} E\big([T_1^0 > x]|T_2^0 = y, T_1 = z, D_1 = 0\big) \\ &\qquad \times D_z P\big(T_1 > z^-, D_1 = 0|T_2^0 = y\big) \\ &= P(T_1 > x, D_1 = 1|T_2 = y, D_2 = 1) \\ &\quad - \left\{\int_0^x + \int_x^{\infty}\right\} E\big([T_1^0 > x]|T_2^0 = y, T_1 = z, D_1 = 0\big) \\ &\qquad \times D_z P\big(T_1 > z^-, D_1 = 0|T_2 = y, D_2 = 1\big) \\ &= S_u(x|y) + S_c(x|y) - \int_0^x \frac{S^0(x|y)}{S^0(z|y)} D_z S_c(z^-|y). \qquad \square \end{aligned}$$

The following assumption is required by Lemma 3.2 below.

(A3) The probability weights have been chosen so that (3.5) holds.

LEMMA 3.2. *Assume* (A1), (A2) *and* (A3) *hold. If the constant* $M$ *satisfies* $S(x, y) > 0$ *for* $x, y < M$, *then*

$$\sup_{0\le x\le y\le M} |\hat{S}^0(x|y) - S^0(x|y)| \to 0 \quad a.s. \text{ as } n \to \infty.$$

PROOF. Consider the functional equation

$$\begin{aligned}&[T(S^0(\cdot|\cdot), S(\cdot|\cdot), S_u(\cdot|\cdot))](x, y)\\ &\quad = S^0(x|y) - S(x|y) - \int_0^x \frac{S^0(x|y)}{S^0(z|y)} D_z(S(z^-|y) - S_u(z^-|y)),\end{aligned}$$

where for any fixed $y$, $S^0(\cdot|y)$ and $S(\cdot|y)$ are survival functions and $S_u(\cdot|y)$ is a conditional subsurvival function. Theorem 2.4 of Tsai (1986) shows that the unique solution of

$$[T(\psi(S(\cdot|\cdot), S_u(\cdot|\cdot)), S(\cdot|\cdot), S_u(\cdot|\cdot))](x, y) = 0$$

for any fixed $y = y_0$ is

$$[\psi(S(\cdot|\cdot), S_u(\cdot|\cdot))](x, y_0) = \left[\gamma\left(\int_0^{\mu^+} \frac{D_t S_u(t|y_0)}{S(t-|y_0)}\right)\right](x), \quad \text{if } S(x|y_0) > 0.$$

Therefore, for $x, y < M$,

$$[\psi(\hat{S}(\cdot|\cdot), \hat{S}_u(\cdot|\cdot))](x, y) = \left[\gamma\left(\int_0^{\mu^+} \frac{D_t\hat{S}_u(t|y)}{\hat{S}(t^-|y)}\right)\right](x) = \hat{S}^0(x|y)$$

is the unique solution of $T(\hat{S}^0(\cdot|\cdot)\hat{S}(\cdot|\cdot), \hat{S}_u(\cdot|\cdot)) = 0$. Similarly, $S^0(x|y)$ is the unique solution of $T(S^0(\cdot|\cdot), S(\cdot|\cdot), S_u(\cdot|\cdot)) = 0$ by Lemma 3.1. For any conditional survival function $S^*(x|y)$ satisfying

$$\sup_{0\le x,\, y\le M} |S^*(x|y) - S^0(x|y)| > \varepsilon,$$

we have $T(S^*(\cdot|\cdot), S(\cdot|\cdot), S_u(\cdot|\cdot)) \neq 0$. The assumption (A3) implies that

$$\begin{aligned}\sup_{0\le x,\, y\le M} &|[T(S^*(\cdot|\cdot), \hat{S}(\cdot|\cdot), \hat{S}_u(\cdot|\cdot))](x, y)\\ &- [T(S^*(\cdot|\cdot), S(\cdot|\cdot), S_u(\cdot|\cdot))](x, y)| \to 0 \quad \text{a.s. as } n \to \infty.\end{aligned}$$

Therefore, for almost all realizations, there exists an $m$ such that for all $n > m$, $[T(S^*(\cdot|\cdot), \hat{S}(\cdot|\cdot), \hat{S}_u(\cdot|\cdot))](x, y) \neq 0$, for some $(x, y)$. Thus the unique solution $\hat{S}^0(x|y)$ of $[T(\hat{S}^0(\cdot|\cdot), \hat{S}(\cdot|\cdot), \hat{S}_u(\cdot|\cdot))](x, y) = 0$ must satisfy $\sup_{0\le x,\, y\le M}|\hat{S}(x|y) - S^0(x|y)| < \varepsilon$. Letting $\varepsilon$ decrease to zero, we see that

$$\sup_{0\le x,\, y\le M} |\hat{S}^0(x|y) - S^0(x|y)| \to 0 \quad \text{a.s. as } n \to \infty. \qquad \square$$

The uniform consistency of $\hat{S}_3^0$, which is needed to derive the consistency of $\hat{S}^0$, is the one-dimensional result for the Kaplan–Meier estimator of Földes and Rejtö (1981).

THEOREM 3.1. *Assume* (A1) *through* (A3) *hold. If the constant M is such that* $S(x|y) > 0$, *for* $x,\ y < M$ *and* $S_3(M) > 0$, *then*

$$\sup_{0 \le x \le y \le M} |\hat{S}^0(x, y) - S^0(x, y)| \to 0 \quad a.s.\ as\ n \to \infty.$$

PROOF. The uniform consistency of the empirical survival function and of product limit estimator and Lemma 3.1 imply that $\hat{S}_3^0(y)/S^e(y, y)$, $S_2^e(x, y)$, $\hat{S}_3^0(y)\hat{S}^0(x|y)/(S^e(y, y)\hat{S}^0(x|y))$, and $S_{12}^e(x, y)$ converge uniformly to $S_3^0(y)/S(y, y)$, $S_2(x, y)$, $S_3^0(y)S^0(x|y)/(S(y, y)S(x|y))$ and $S_{12}(x, y)$, respectively. The theorem follows from the bivariate generalization of Lemma 6 of Aalen (1976). □

**4. Self-consistency of $\hat{S}^0$ under univariate censoring.** Self-consistency was defined for estimators of survival functions for univariate observations by Efron (1967) and extended to bivariate observations by Hanley and Parnes (1983), Campbell and Földes (1982) and Muñoz (1980). An estimator $\hat{S}^0$ of $S^0$ is said to be self-consistent if

$$\begin{aligned} n\hat{S}^0(t_1, t_2) = \sum_{i=1}^{n} \Bigg\{ & [T_{1i} > t_1, T_{2i} > t_2] \\ & + [D_{1i} = 0, D_{2i} = 1, T_{1i} \le t, T_{2i} > t_2] \\ & \times \left( \frac{\hat{S}^0(t_1, T_{2i}^-) - \hat{S}^0(t_1, T_{2i})}{\hat{S}^0(T_{1i}, T_{2i}^-) - \hat{S}^0(T_{1i}, T_{2i})} \right) \\ & + [D_{1i} = 1, D_{2i} = 0, T_{1i} > t, T_{2i} \le t_2] \\ & \times \left( \frac{\hat{S}^0(T_{1i}^-, t_2) - \hat{S}^0(T_{1i}, t_2)}{\hat{S}^0(T_{1i}^-, T_{2i}) - \hat{S}^0(T_{1i}, T_{2i})} \right) \\ & + [D_{1i} = D_{2i} = 0, T_{1i} \le t_1 \text{ or } T_{2i} \le t_2] \\ & \times \left( \frac{\hat{S}^0(\max(t_1, T_{1i}), \max(t_2, T_{2i}))}{\hat{S}^0(T_{1i}, T_{2i})} \right) \Bigg\}. \end{aligned}$$

Proposition 4.1 from Muñoz (1980) characterizes self-consistency under univariate censoring in terms of $m(t) = \hat{S}_3^0(t^-) - \hat{S}_3^0(t)$, the probability mass function of the product–limit estimator of the univariate censoring distribution. We state Muñoz's characterization and use it to show that $\hat{S}^0$ is self-consistent.

PROPOSITION 4.1. *Under univariate censoring, if the* $2n$ *observed times* $\{T_{ki}, 1 \le i \le n, k = 1, 2\}$ *are distinct, then* $\tilde{S}^0$ *is a self-consistent estimator of* $S^0$ *if and only if condition* (iv) *holds and, for each* $i$, *the appropriate condition of*

(i) *through* (iii) *holds*:

$$\text{(i)}\quad \tilde{S}^0(T_{1i}, T_{2i}) + \tilde{S}^0(T_{1i}^-, T_{2i}^-) - \tilde{S}^0(T_{1i}^-, T_{2i}) - \tilde{S}^0(T_{1i}, T_{2i}^-) = m(T_{1i} \wedge T_{2i}), \quad \text{if } D_{1i} = D_{2i} = 1,$$

$$\text{(ii)}\quad \tilde{S}^0(T_{1i}^-, T_{2i}^-) - \tilde{S}^0(T_{1i}^-, T_{2i}) = m(T_{2i}) \quad \text{if } D_{1i} = 0,\ D_{2i} = 1,$$

$$\text{(iii)}\quad \tilde{S}^0(T_{1i}^-, T_{2i}^-) - \tilde{S}^0(T_{1i}, T_{2i}^-) = m(T_{1i}) \quad \text{if } D_{1i} = 1,\ D_{2i} = 0,$$

$$\text{(iv)}\quad \tilde{S}^0(t, t) = \sum_{j=1}^{n} \left[D_{3j} = 1, T_{1j} \wedge T_{2j} \ge t\right] m(T_{1j} \wedge T_{2j}).$$

THEOREM 4.1. *Under univariate censoring, if the* $2n$ *observed times* $\{T_{ki},\ 1 \le i \le n,\ k = 1, 2\}$ *are distinct, then* $\hat{S}^0$ *is a self-consistent estimator of* $S^0$.

PROOF. The theorem will be proved by showing that $\hat{S}^0$ satisfies the four characterizing equations of Propositions 4.1. Since $\hat{S}_3^0(\cdot)$ is the product limit estimator of $S_3^0$, Efron's (1967) equation (7.9) implies that

$$m(T_{3i}) = D_{3i}\hat{S}_3^0(T_{3i}^-)/(nS^e(T_{3i}^-, T_{3i}^-)).$$

For cases (i) to (iii), $D_{3i} = 1$.

(i) In the absence of ties, the mass at a point $(T_{1i}, T_{2i})$ with $T_{1i} \ge T_{2i}$ reduces to $D_{1i}D_{2i}\hat{S}_3^0(T_{2i}^-)/(nS^e(T_{2i}^-, T_{2i}^-)) = m(T_{3i})$.

(ii) In the absence of ties, no mass on the ray form $(T_{1i}, T_{2i})$ to $(T_{1i}, \infty)$ is contributed by the integral over $\Delta(s, t)$. Since $D_{1i} = 0$ implies $m(T_{1i}) = 0$, the mass on the ray comes only from the integral over $R(s, t)$, and is $D_{2i}\hat{S}_3^0(T_{2i}^-)/(nS^e(T_{2i}^-, T_{2i}^-) = m(T_{2i})$.

(iii) This case follows from (ii) by symmetry.

(iv) Since $\hat{S}^0(T_{3i}^-, T_{3i}^-) - \hat{S}^0(T_{3i}, T_{3i}) = m(T_{3i})D_{3i}$, therefore, $\hat{S}^0(t, t) = \sum_{i=1}^{n}[T_{3i} \ge t, D_{3i} = 1]m(T_{3i})$. □

Although the $\hat{S}^0$ is a self-consistent estimator of $S^0$ under univariate censoring, $\hat{S}^0$ is not self-consistent under bivariate censoring. For example, suppose we have four data points (5., 5., 1, 1) (4., 4., 1, 1), (3., 5., 0, 1) and (5., 3., 1, 0). The self consistent estimator of $S^0$ will put probability mass $\frac{1}{4}$ at the point (4., 4.) and probability mass $\frac{3}{4}$ at the point (5., 5.), but $\hat{S}^0$ will put probability mass $\frac{1}{2}$ at the points (4., 4.) and (5., 5.).

**Acknowledgment.** We thank the referees for their constructive comments.

## APPENDIX

LEMMA A.1. *Let* $\{h(n) = n \ge 1\}$ *be a sequence of positive real numbers such that* $h(n) \to 0$ *and* $\sum_{n=1}^{\infty}\exp(\gamma n h^2(n)) < \infty$ *for every* $\gamma < 0$. *Let* $k(u)$ *be a* *density function of bounded variation on the real line. Set*

$$G_1(x, y) = \int_x^\infty \int_x^\infty \int_y^\infty f^0(t, y) h(u, v)\, dv\, du\, dt,$$

$$G_2(x, y) = \int_x^\infty \int_t^\infty \int_y^\infty f^0(t, y) h(u, v)\, dv\, du\, dt,$$

$$\hat{G}_1(x, y) = (nh(n))^{-1} \sum_i k((T_{2i} - y)/h(n))[T_{1i} > x, D_{2i} = 1],$$

$$\hat{G}_2(x, y) = (nh(n))^{-1} \sum_i k((T_{2i} - y)/h(n))\left[T_{1i} > x, D_{1i} = D_{2_i} = 1\right],$$

$$f^0(x, y) = \frac{\partial^2 S^0(x, y)}{\partial x\, \partial y} \quad \text{and} \quad h(x, y) = \frac{\partial^2 H(x, y)}{\partial x\, \partial y}.$$

*If $G_i(x, y)$, $i = 1, 2$, are uniformly continuous for $0 \le x, y \le \infty$, then*

$$\sup_{0 \le x, y < \infty} |\hat{G}_i(x, y) - G_i(x, y)| \to 0 \quad a.s. \text{ as } n \to \infty \text{ for } i = 1, 2.$$

PROOF. The proof is similar to that of Theorem 1 of Nadaraja (1965) where uniform consistency is established for univariate kernel density estimators.

We abbreviate $\sup_{0 \le x, y < \infty}$ by sup. Integrating the double integral by parts, we find

$$\begin{aligned}
A_n &= \sup|\hat{G}_1(x, y) - E(\hat{G}_1(x, y))| \\
&= \sup\left|\frac{1}{h(n)} \iint k\left(\frac{t - y}{h(n)}\right)[s > x]\, DS_2^e(s, t)\right. \\
&\qquad \left. - \frac{1}{h(n)} \iint k\left(\frac{t - y}{h(n)}\right)[s > x]\, DS_2(s, t)\right| \\
&= \sup\left|\frac{1}{h(n)} \iint (S_2^e(s, t) - S_2(s, t))\, D_t k\left(\frac{t - y}{h(n)}\right) D_s[s > x]\right| \\
&\le \sup \frac{1}{h(n)} \iint \left|S_2^e(s, t) - S_2(s, t)\right| \left|D_t k\left(\frac{t - y}{h(n)}\right)\right| \left|D_s[s > x]\right| \\
&\le \sup|S_2^e(x, y) - S_2(x, y)|\nu/h(n),
\end{aligned}$$

where $\nu = \int |Dk|$ is the variation of $k$. Theorem 1 of Kiefer and Wolfowitz (1956) implies that for any $\varepsilon$,

$$\begin{aligned}
P(A_n > \varepsilon) &\le P\{\sup|S_2^e(x, y) - S_2(x, y)| > \varepsilon h(n)/\nu\} \\
&\le C \exp - \varepsilon_1 n h^2(n),
\end{aligned}$$

where $\varepsilon_1 = (\varepsilon/\nu)^2$ and $0 < C < \infty$. The Borel–Cantelli lemma and the assumption that $k(\cdot)$ has bounded variation imply that $A_n \to 0$ a.s. as $n \to \infty$. It remains to show that

$$\sup|E(\hat{G}_1(x, y)) - G_1(x, y)| \to 0 \quad \text{a.s. as } n \to \infty.$$

Let $g(x, y) = (\partial^2 S_2(x, y))/\partial x\,\partial y$. Since $G_1(x, y) = (-\partial S_2(x, y))/\partial y$, and $G_1(\infty, y) = 0$, we have

$$G_1(x, y) = \int [s > x] g(s, y)\, ds = \int \frac{1}{h(n)} k\left(\frac{t}{h(n)}\right) \int [s > x] g(s, y)\, ds\, dt.$$

Therefore, for $\delta > 0$,

$$\begin{aligned}
&\sup |E\hat{G}_1(x, y) - G_1(x, y)| \\
&\quad \le \sup \left\{ \int_{|t| \le \delta} + \int_{|t| > \delta} \right\} \left\{ \frac{t}{h(n)} k\left(\frac{t}{h(n)}\right) \right\} \\
&\qquad \times \left| \int [s > x] \{ g(s, y - t) - g(s, y) \}\, ds \right| dt \\
&\quad \le \sup \sup_{|t| \le \delta} \left| \int [s > x] \{ g(s, y - t) - g(s, y) \}\, ds \right| \qquad \text{(A.1)} \\
&\qquad + 2M \int_{|t| > \delta/h(n)} k(t)\, dt. \\
&\quad = \sup \sup_{|t| \le \delta} |G(x, y - t) - G(x, y)| + 2M \int_{|t| > \delta/h(n)} k(t)\, dt,
\end{aligned}$$

where $M = \sup G_1(x, y) < \infty$. Let $\varepsilon$ be an arbitrarily small positive number. Because $G_1$ is uniformly continuous, we can make the first term of the right-hand side of (A.1) less than $\varepsilon/2$ by choosing $\delta$ sufficiently small. Having so chosen $\delta$, we can then choose $n$ so large that $\delta/h(n)$ is large enough so that

$$\int_{|t| > \delta/h(n)} k(t)\, dt < \varepsilon/2M.$$

Thus (A.1) implies that $\sup |E\hat{G}_1(x, y) - G_1(x, y)| < \varepsilon$. Therefore,

$$|\hat{G}_1(x, y) - G_1(x, y))| \to 0 \quad \text{a.s. as } n \to \infty.$$

The uniform consistency of $\hat{G}_2$ can be established similarly. □

THEOREM A.1. *Assume the conditions of Lemma* A.1 *hold. If the constant* $M$ *satisfies* $G_2(0, y) > 0$ *for* $y \le M$, *then*

(a) $$\sup_{0 \le x,\, y \le M} |\hat{S}(x|y) - S(x|y)| \to 0 \quad a.s.\ as\ n \to \infty,$$

(b) $$\sup_{0 \le x,\, y < M} |\hat{S}_u(x|y) - S_u(x|y)| \to 0 \quad a.s.\ as\ n \to \infty,$$

*where* $\hat{S}(x|y)$ *and* $\hat{S}_u(x|y)$ *are defined in* (3.2) *and* $W_{n,i}(y)$ *is defined in* (3.4).

PROOF. The proof is based on Lemma A.1 and the fact

$$\hat{S}(x|y) = \frac{\hat{G}_1(x, y)}{\hat{G}_1(0, y)} \quad \text{and} \quad \hat{S}_u(x|y) = \frac{\hat{G}_2(x, y)}{\hat{G}_2(0, y)}. \qquad \square$$

## REFERENCES

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

BERAN, R. (1981). Nonparametric regression with randomly censored survival data. Unpublished manuscript, Univ. of California, Berkeley.

CAMPBELL, G. (1982). Nonparametric bivariate estimation with randomly censored data. *Biometrika* **68** 417–422.

CAMPBELL, G. (1982). Asymptotic properties of several nonparametric multivariate distribution function estimators under random censorship. In *Survival Analysis* (J. Crowley and R. A. Johnson, eds.) 243–256. IMS, Hayward, Calif.

CAMPBELL, G. and FÖLDES, B. (1982). Large-sample properties of nonparametric bivariate estimators with censored data. Proc. Conf. Nonparametric Statistical Inference, Budapest. *Colloq. Math. Soc. János Bolyai* **32** 103–121.

EFRON, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **4** 831–853. Univ. of California Press.

FÖLDES, A. and REJTÖ, L. (1981). Strong uniform consistency for nonparametric survival curve estimators from randomly censored data. *Ann. Statist.* **9** 122–129.

HANLEY, J. A. and PARNES, M. N. (1983). Nonparametric estimation of a multivariate distribution in the presence of censoring. *Biometrics* **39** 129–139.

JOHANSEN, S. (1978). The product limit estimator of maximum likelihood estimator. *Scand. J. Statist.* **5** 195–199.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

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

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906.

KIEFER, J. and WOLFOWITZ, J. (1958). On the deviation of the empiric distribution function of vector chance variables. *Trans. Amer. Math. Soc.* **87** 173–186.

LEURGANS, S., TSAI, W.-Y. and CROWLEY, J. (1982). Freund's bivariate exponential distribution and censoring. In *Survival Analysis* (J. Crowley and R. A. Johnson, eds.) 230–242. IMS, Hayward, Calif.

MUÑOZ, A. (1980). Nonparametric estimation from censored bivariate observations. Technical Report No. 60, Division of Biostatistics, Stanford Univ.

NADARAJA, E. A. (1965). On non-parametric estimates of density functions and regression curves. *Theory Probab. Appl.* **10** 186–190.

PETERSON, A. V. (1977). Expressing the Kaplan–Meier estimator as a function of empirical subsurvival functions. *J. Amer. Statist. Assoc.* **72** 854–858.

STONE, C. J. (1977). Consistent nonparametric regression (with discussion). *Ann. Statist.* **5** 595–645.

TSAI, W. Y. (1982). Bivariate survival time and censoring. Ph.D. dissertation, Univ. of Wisconsin-Madison.

TSAI, W. Y. (1986). Estimation of survival curves from dependent censorship models via a generalized self-consistent property with nonparametric Bayesian estimation application. *Ann. Statist.* **14** 238–249.

WEI-YANN TSAI
APPLIED MATHEMATICS DEPARTMENT
BROOKHAVEN NATIONAL LABORATORIES
UPTON, NEW YORK 11973

SUE LEURGANS
DEPARTMENT OF STATISTICS
OHIO STATE UNIVERSITY
1958 NEIL AVENUE
COLUMBUS, OHIO 43210

JOHN CROWLEY
THE FRED HUTCHINSON CANCER RESEARCH CENTER
1124 COLUMBIA STREET
SEATTLE, WASHINGTON 98104