LEAST SQUARES REGRESSION WITH CENSORED DATA

BY

RUPERT G. MILLER, JR.

TECHNICAL REPORT NO. 3
DECEMBER 12, 1974

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT 1 RO1 GM21215-01

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
LEAST SQUARES REGRESSION WITH CENSORED DATA

by

RUPERT G. MILLER, JR.

TECHNICAL REPORT NO. 3
DECEMBER 12, 1974

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT 1 R01 GM21215-01

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
LEAST SQUARES REGRESSION WITH CENSORED DATA

Rupert G. Miller, Jr.
Stanford University

1. Introduction.

The usual linear regression model for the $n$ pairs of variates $(x_i, y_i), i = 1, \ldots, n$, is

$$y_i = \alpha + \beta x_i + e_i,$$  \hspace{1cm} (1.1)

where the random variables $e_i$ are assumed to be independently, identically distributed according to the distribution $F$ which has mean $\mu = 0$ and finite variance $\sigma^2$. No assumption of normality is imposed on $F$ in this paper. Whether the $x_i$ should be viewed as a set of fixed values or in some cases as independent random variables distributed according to the distribution $G$ is left unspecified until it is necessary to do so.

The estimation of $\alpha$ and $\beta$ when the $y_i$ are subject to censoring is the topic of this paper. The primary type of censorship to be considered is single censoring on the right, i.e., the observable variable is not $y_i$ but

$$y_i^+ = \begin{cases} y_i & \text{if } y_i \leq c_i, \\
                          c_i & \text{if } y_i > c_i. \end{cases}$$  \hspace{1cm} (1.2)

$^+$Research supported by Public Health Service Grant 1 RO1 GM21215-01.
It is known whether each observation is censored or uncensored, and the presence or absence of a "+" on $y_i$ denotes this. When it is necessary to specify the structure of the $c_i$, it will be assumed that the $c_i$ are random variables independently distributed of each other and the $e_i$, but not always of the $x_i$.

This particular type of censoring is important in applications to medical statistics where $y_i$ is the survival time of a patient or animal under a set of experimental conditions. Censoring comes from the experiment being terminated or the data analyzed while some patients or animals are still alive, and also from some patients being lost to follow-up during the course of the study for reasons not related to the treatment.

The main idea for estimating $\alpha$ and $\beta$ is presented in Section 4. It utilizes the Kaplan-Meier (1958) product-limit estimator for a distribution function based on singly censored data. Extension of the idea to doubly censored data is outlined in Section 7, and the Appendix contains the mathematical details for various assertions in Section 4.

To put the proposals of this paper in perspective, the previously published results for the exponential distribution are summarized in Section 2, and Cox's (1972) conditional likelihood approach to a different but related nonparametric model is described in Section 3. Cox's approach has been receiving attention in the recent literature, and it represents a major competitor to the proposals in Section 4.

Section 5 gives another estimation procedure based on the Kaplan-Meier distribution estimator and Wald's (1940) method of estimating a functional relation. Section 6 contains a discussion of estimating $\alpha$
and \( \beta \) when the censoring mechanism has a more general structure than assumed in Sections 4 and 5. Finally, the techniques of Sections 4 and 5 are applied to the Stanford heart transplant data in Section 8 to illustrate their use.

2. Exponential Distribution Approach.

By far the easiest distribution to handle analytically when dealing with singly censored data is the exponential distribution. Feigl and Zelen (1965) introduced the survival model in which \( y_i \) has the exponential density \( \lambda_i \exp(-\lambda_i y), y > 0 \), with the mean survival time being related to the independent variable by

\[
\frac{1}{\lambda_i} = \alpha + \beta x_i .
\]  

(2.1)

Only uncensored data were considered by Feigl and Zelen, but Zippin and Armitage (1966) extended the analysis to include censoring caused by the termination of the study. Censoring patterns created by simultaneous and uniform entry of the patients into the study were examined in detail. More recently, Mantel and Myers (1971) have studied the problem of computational convergence in obtaining the maximum likelihood estimates for the multiple regression extension \( 1/\lambda_i = x_i^T \beta \) with \( x_i = (x_{i1}, \ldots, x_{ip}) \),

\[ \beta = (\beta_1, \ldots, \beta_p)^T . \]

For the uncensored case Feigl and Zelen (1965) also introduced the log-linear model in which

\[
\frac{1}{\lambda_i} = \alpha e^{\beta x_i} .
\]

(2.2)
This is the parametric ancestor to Cox's (1972) nonparametric model to be examined in the next section. Zippin and Lamborn (1969) gave the extension of this analysis to the censored case for simultaneous and uniform entry. In an adjunct report Lamborn (1969) established that a goodness-of-fit statistic for the model (2.2), or (2.1), does not have a $\chi^2$ distribution as conjectured but instead is distributed as a linear combination of squared unit normal random variables. Glasser (1967) was preceded by Feigl and Zelen (1965) and Zippin and Armitage (1966), but he also independently proposed the log-linear model $1/\lambda_{ij} = \alpha_{ij} e^{\beta x_{ij}}$ for an analysis of covariance with several populations.


In an important paper Cox (1972) proposed analyzing the model where the hazard rate (i.e., $\lambda(y; x) = dF(y; x)/(1-F(y; x))$) for survival with independent variable $x$ is

$$\lambda(y; x) = \lambda(y) e^{\beta x}.$$  \hfill (3.1)

This is equivalent to assuming a Lehmann-alternative family of distributions

$$1-F(y; x) = (1-F(y))^{\exp(\beta x)},$$  \hfill (3.2)

where

$$F(y) = F(y; 0) = 1 - e^{-\int_0^y \lambda(t) dt}.$$  \hfill (3.3)
The exponential term $e^{\beta x}$ in the exponent of (3.2) appears ungainly, but it has the definite advantage of being positive for all values of $x$ and $\beta$. Consequently, the hazard rate is always well-defined so that maximization of the likelihood cannot occur on a boundary.

Assuming the model (3.1), Cox advocated a conditional likelihood approach to estimating $\beta$. Namely, at each uncensored data point $y_1$ let $\mathcal{R}(y_1)$ be the set of observations still at risk at $y_1 - 0$, i.e.,

$$\mathcal{R}(y_1) = \{(x_j, y_j^{(+)}) : y_j^{(+)} \geq y_1\}. \quad (3.4)$$

Conditional on the set of points at risk $\mathcal{R}(y_1)$, the probability that the point $(x_i, y_i)$ is the one which occurs is

$$\frac{e^{\beta x_i}}{\sum_{(x_j, y_j^{(+)}) \in \mathcal{R}(y_1)} e^{\beta x_j}}. \quad (3.5)$$

The conditional likelihood $L(\beta)$ is defined to be the product of the terms (3.5) over all the uncensored points. Thus,

$$\ln L(\beta) = \beta \sum_{uc} x_i - \sum_{uc} \ln \left( \sum_{\mathcal{R}(y_1)} e^{\beta x_j} \right), \quad (3.6)$$

where the summation is over all the uncensored points $(x_i, y_i)$ in the sample. The derivative of (3.6) is

$$\frac{\partial}{\partial \beta} \ln L(\beta) = \sum_{uc} \left[ x_i - \frac{\sum_{\mathcal{R}(y_1)} x_j e^{\beta x_j}}{\sum_{\mathcal{R}(y_1)} e^{\beta x_j}} \right], \quad (3.7)$$

5
and the value of $\beta$ for which this equals zero is usually found through iteration on a computer utilizing the second derivative.

An estimate of the variance of the maximum likelihood estimator $\hat{\beta}$ can be obtained by substituting $\hat{\beta}$ into the inverse of

$$-\frac{d^2 \ln L(\beta)}{d\beta^2} = \sum_{uc} \left\{ \frac{\sum x_j^2 e^{\beta x_j}}{\mathcal{R}(y_{1i})} \beta x_j - \left[ \frac{\sum x_j e^{\beta x_j}}{\mathcal{R}(y_{1i})} \beta x_j \right]^2 \right\}. \tag{3.8}$$

Expression (3.8) is a sample estimate of the Fisher information whose reciprocal is the asymptotic variance.

The reader should note that the hazard rate function $\lambda(y)$ does not enter the likelihood in (3.6). Cox argued that since $\lambda(y)$ is unspecified in the model, it could be zero between uncensored $y_i$ and therefore the sample contains no information about $\beta$ except at the uncensored points. Once $\hat{\beta}$ has been obtained Cox's method returns to the problem of estimating $\lambda(y)$ at the uncensored points. These estimates are then combined in a product-limit fashion to produce an estimate of the distribution $F(y)$.

There was considerable discussion in the Journal of the Royal Statistical Society, Series B, following the Cox paper. In this discussion and a subsequent paper (1974) Breslow proposed a slightly different likelihood function based on the assumption that the hazard rate function $\lambda(y)$ is constant between uncensored observations. The likelihood function for $\beta$ differs from Cox when ties occur in the uncensored data, but the maximum likelihood estimate of $\beta$ is the same. The Breslow estimate for the distribution function $F$ is simpler than the Cox estimate.
Kalbfleisch and Prentice (1973) helped elucidate the Cox conditional likelihood approach by justifying it in terms of a marginal likelihood of rank statistics when no ties are present. With ties they derive a slightly different likelihood. Peto and Peto (1972) studied asymptotically efficient rank-invariant procedures which include the Cox procedure. Prentice (1973) applied Fraser's structural inference approach to the Cox model (3.1).

The Cox approach extends immediately to the case of multiple regression where it is assumed the hazard rate is

$$\lambda(y; \tilde{x}) = \lambda(y) e^{x'\tilde{\beta}}$$

(3.9)

with \( \tilde{x} = (x_1, \ldots, x_p) \), \( \tilde{\beta} = (\beta_1, \ldots, \beta_p)' \). Expression (3.7) becomes a set of \( p \) equations and (3.8) becomes a \( p \times p \) matrix of partial derivatives.

4. **Least Squares Approach.**

The backbone of the least squares approach is the product-limit estimator of a distribution function introduced by Kaplan and Meier (1958) to handle estimation for data singly censored on the right.

The product-limit estimator is the limit of the classical life-table approach obtained through increasingly finer subdivisions of the time scale. It is the nonparametric maximum likelihood estimator of the distribution function. Efron (1967) utilized the product-limit estimator in the two sample problem with censored data. Breslow and Crowley (1974) and Meier (1974) have extensively studied the distributional properties of this estimator.
In the single sample problem the product-limit estimator (hereafter referred to as the Kaplan-Meier estimator) of the right tail of a distribution function with right censored data \( y_1^{(+)}, \ldots, y_n^{(+)} \) is

\[
1 - F^{km}(y) = \prod_{y(i) \leq y} \left( \frac{n-i}{n-i+1} \right),
\]

(4.1)

where \( y_1^{(+)} \leq y_2^{(+)} \leq \ldots \leq y_{(n)}^{(+)} \) are the ordered observations (censored or uncensored). The product in (4.1) is just over the uncensored observations less than or equal to \( y \). If censored observations are tied with uncensored ones, the convention is to treat the uncensored observations as preceding (i.e., as being less than) the censored observations. If the largest observation \( y_{(n)}^{(+)} \) is censored, the product in (4.1) will not reach zero as \( y \to +\infty \). In this case \( F^{km}(y) \) is either considered undefined for \( y > y_{(n)}^{+} \) or is defined to be one for \( y \geq y_{(n)}^{+} \).

For a complete development of this estimator the reader is referred to Kaplan and Meier (1958). Very briefly, the motivation for (4.1) is the following. Just before the occurrence of an uncensored point \( y_{(i)} \) there are \( n-i+1 \) points still to occur. This suggests that the conditional probability of the random variable exceeding \( y_{(i)}^{+} \) given that it is greater than \( y_{(i)} - 0 \) could be estimated by \( (n-i)/(n-i+1) \). The conditional probability that the random variable exceeds a censored point \( y_{(i)}^{+} + 0 \) given that it is greater than \( y_{(i)}^{+} - 0 \) could be estimated as one since the censored point gives no information on the hazard rate at \( y_{(i)}^{+} \). The unconditional probability of the random variable exceeding \( y \) is then the product of the conditional probabilities that it exceeds each uncensored point prior or equal to \( y \).
The Kaplan-Meier estimator of \( F \) changes only by jumps at the uncensored points. The size of the jump at each point is

\[
d^{\text{km}}(y(i)^+) = \begin{cases} 
\prod_{y(j) < y(i)} \left( \frac{n-j}{n-j+1} \right) \left( \frac{t(i)}{n-i+1} \right) & \text{if } y(i)^+ = y(i), \\
0 & \text{otherwise},
\end{cases}
\]  

(4.2)

where \( t(i) \) is the number of uncensored observations tied with \( y(i) \). If the largest observation is censored (i.e., if \( y(n)^+ = y(n) \)), then the convention of defining \( d^{\text{km}}(y) \) equal to one for \( y \geq y(n)^+ \) forces

\[
d^{\text{km}}(y(n)^+) = \prod_{y(j) < y(n)} \left( \frac{n-j}{n-j+1} \right). 
\]  

(4.3)

It will be notationally convenient at times to let \( w_i \) denote the weight assigned to \( y_i^{(+)} \) through the above process. Then,

\[
w_i = \frac{1}{t_i} d^{\text{km}}(y_i^{(+)}),
\]  

(4.4)

where the Kaplan-Meier jump for tied observations has been divided equally.

When no censoring is present, \( \hat{F}^{\text{km}} \) reduces to the usual sample distribution estimator \( \hat{F} \) which assigns weight \( 1/n \) to each observation. Under fairly general conditions \( \hat{F}^{\text{km}}(y) \) is an asymptotically consistent estimator of \( F(y) \) and is asymptotically normally distributed (cf., Kaplan and Meier (1958), Efron (1967), Breslow and Crowley (1974), Meier (1974)). As a function of \( y \), \( \hat{F}^{\text{km}} \) converges weakly to a Gaussian process (cf., Breslow and Crowley (1974), Meier (1974)).
The estimators of $\alpha$ and $\beta$ which are proposed in this section will be referred to as Kaplan-Meier least squares estimators and will be denoted by $\hat{\alpha}^{\text{km}}$ and $\hat{\beta}^{\text{km}}$. Their motivation stems from the definition of least squares estimates in the uncensored case. The usual least squares estimates ($\hat{\alpha}$ and $\hat{\beta}$) are those values of $a$ and $b$ which minimize the sum of squares $\sum (y_i - a - bx_i)^2$. Division by the constant $n$ does not affect this minimization so write

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - a - bx_i)^2 = \int z^2 \hat{F}^{a,b}(z), \quad (4.5)$$

where $\hat{F}^{a,b}$ is the sample distribution estimate based on $z_i = y_i - a - bx_i$, $i=1, \ldots, n$. Thus, the least squares estimates are those $a$ and $b$ which minimize the integral of squares in (4.5). Therefore, in the censored case why not choose $a$ and $b$ to minimize

$$\int z^2 \hat{F}^{\text{km}}^{a,b}(z), \quad (4.6)$$

where $\hat{F}^{\text{km}}^{a,b}$ is the Kaplan-Meier estimate of $F$ based on the censored sample $z_1^{(+)} = y_1^{(+)} - a - bx_1$, $i=1, \ldots, n$. For fixed $a$ and $b$ this integral is a weighted sum of squares

$$\sum_{i=1}^{n} w_i(a,b) (y_i^{(+)} - a - bx_i)^2, \quad (4.7)$$

where the weight $w_i(a,b)$ on $(y_i^{(+)} - a - bx_i)^2$ is that assigned to $z_i^{(+)} = y_1^{(+)} - a - bx_1$ by the Kaplan-Meier distribution estimator applied to $z_i^{(+)}$, $i=1, \ldots, n$.

In (4.7) the weight $w_i(a,b)$ on $(z_i^{+})^2$ corresponding to a censored $y_1^{+}$ is zero except when the largest $z_i^{(+)}$ is censored. Following the
usual convention, the remaining weight is assigned to $(a_+(n))^2$ as in (4.3). This treats $z_+(n)$ as an uncensored observation which has an unpleasant aspect. The remaining weight may be assigned to different $y_i^{(+)}$ as $b$ changes because which $y_i^{(+)}$ leads to the largest $z_i^{(+)}$ changes with $b$ (but not with $a$). However, for large $n$ the amount of weight being shifted is negligible.

Numerical computation of $\hat{\alpha}^{km}$ and $\hat{\beta}^{km}$ is not difficult with the aid of an electronic computer. Because of the memory storage necessary to retain the $y_i^{(+)}$, it will require a machine somewhat larger than most pocket or desk-top electronic calculators. There is no easy iterative scheme to locate the minimum of (4.7). When a censored $z_j^{+}$ shifts past an uncensored $z_i$, the weight $w_i(a, b)$ assigned to $z_i$ jumps in a discontinuous fashion. This causes the sum of squares in (4.7) to be discontinuous in $b$ even though it is locally convex between discontinuities. The minimum frequently occurs at a discontinuity so iterative schemes based on power series expansions are not useful. Instead the search for the minimum has to be conducted through a shrinking grid procedure. If the minimum is not actually attained because of a discontinuity, the infimum and the limiting $a, b$ will be declared the minimum and $\hat{\alpha}^{km}$, $\hat{\beta}^{km}$, respectively.

The numerical search for the minimum is aided by the fact that the search is only one-dimensional. For fixed $b$ the minimum with respect to $a$ can be computed directly without any iteration; it is

$$\hat{\alpha}^{km}_b = \frac{1}{\sum_{i=1}^{n} w_i(0, b) (y_i^{(+)} - bx_i)}.$$ 

(4.8)
The reason $\hat{\alpha}_b^{km}$ can be obtained directly is that for fixed $b$ any change in the variable $a$ shifts the $z_i^{(+)}$ rigidly without changing their order. If the $z_i^{(+)}$ remain in the same order, the weights are undisturbed so for fixed $b$ the minimum of (4.7) with respect to $a$ can be obtained by differentiation. The search for the overall minimum can therefore be conducted on the function

$$f(b) = \frac{1}{n} \sum_{i=1}^{n} w_i(b)(y_i^{(+)} - \hat{\alpha}_b^{km} - bx_i)^2,$$  \hspace{1cm} (4.9)

where $w_i(b) = w_i(0,b)$ henceforth.

The estimator $\hat{\alpha}_b^{km}$ is a familiar quantity. It is simply the estimator for the mean $u$ of $F$ proposed by Kaplan and Meier (1958) as applied to the variables $y_i^{(+)} - bx_i$. Symbolically, it can be expressed in the following form. Let $\hat{d}_b^{km}(y)$ denote the jump attached to $y_i^{(+)}$ from the Kaplan-Meier estimator applied to the $y_i^{(+)} - bx_i$; i.e.,

$$\hat{d}_b^{km}(y_i^{(+)}) = \hat{d}_b^{km}(y_i^{(+)} - bx_i).$$ \hspace{1cm} (4.10)

Then,

$$\hat{\alpha}_b^{km} = \int (y - bx)\hat{d}_b^{km}(y),$$  

$$= \bar{y}_b^{km} - b\bar{x}_b^{km}.$$ \hspace{1cm} (4.11)

The symbols $\bar{y}_b^{km}$ and $\bar{x}_b^{km}$ denote the sample mean estimators obtained through the Kaplan-Meier process. The reader should mentally note that in these means the weight attached to $y_i^{(+)}$ or $x_i$ is that assigned by the Kaplan-Meier process to $y_i^{(+)} - bx_i$ and not by the Kaplan-Meier
process applied to $y_i^{(+)}$ and $x_i$ separately. The $x_i$ corresponding to censored $y_i^{+}$ receive zero weight in $\frac{y_i^{km}}{x_i}$ (except if an $x_i$ corresponds to the largest $y_i^{(+)} - bx_i$ which is censored).

The numerical value of the weighted sum of squares for the minimizing $\hat{\alpha}^{km}$ and $\hat{\beta}^{km}$, namely,

$$
\frac{1}{L} \sum_{i=1}^{L} w_i(\hat{\beta}^{km}) (y_i^{(+)} - \hat{\alpha}^{km} - \hat{\beta}^{km} x_i)^2,
$$

(4.12)

provides an estimate of $\sigma^2$, the variance of the distribution $F$.

This estimate measures how tightly the points fit about the regression line, but it does not seem to play a direct role in estimating the variability of the estimators as it does in the uncensored case.

The concept of minimizing the integral (4.6) has intuitive appeal. The customary least squares estimators for uncensored data are included as a special case. Also, the Kaplan-Meier estimator for the mean $\mu$ of $F$ can be derived in this fashion. Unfortunately, beyond these two special cases, it appears necessary to impose an assumption on the form of the censoring mechanism to assure asymptotic consistency of the estimators. It is easy to produce examples in which $\hat{\alpha}^{km}$, $\hat{\beta}^{km}$ do not converge to the true $\alpha$, $\beta$ as $n \to \infty$ for disadvantageous censoring patterns, and one example is presented in the Appendix in Section A1. The two aforementioned special cases where this cannot happen are the only cases in which the estimates do not affect the values of the weights.

In the case of random censorship, let $L_x(c)$ be the distribution function for the censoring variables $c_i$ at the value $x$ of the
independent variable. If some variables \( y_i^{(+)} \) are not subject to censoring, the possibility \( L_x(+\infty) < 1 \) should be admitted, or if this is distasteful, \( L_x(c) \) can be thought of as having mass beyond the range of the dependent variable. If it is more appropriate to think of the censoring variables \( c_i \) as arbitrary constants, then \( L_x(c) \) should be interpreted as the limit of the empirical distribution of the sequence of censoring constants at \( x \). With this definition of \( L_x(c) \), a condition which guarantees the asymptotic consistency of \( \hat{\alpha}^{\text{km}}, \hat{\beta}^{\text{km}} \) is

\[
L_x(c) = L_x'(c + \beta(x' - x)).
\] (4.13)

In words, the censoring distribution should shift with the regression line just as the distribution of the dependent variable \( y \) does. If \( c_x^P \) is the \( p \)-percentile point of \( L_x(c) \) (i.e., \( L_x(c_x^P) = p \)), then another way of expressing (4.13) is

\[
c_x^P = c_0^P + \beta x.
\] (4.14)

This condition differs from the one needed in the next section where \( L_x(c) \) is required to be invariant under change in \( x \), i.e.,

\[
L_x(c) = L(c) \text{ for all } x.
\] (4.15)

Whether (4.13) or (4.15) is a realistic assumption depends, of course, on the particular censoring pattern encountered in your set of data. The Kaplan-Meier technique should not be too sensitive to departures from (4.13) except when the censoring about the regression line is heavier towards one end of the \( x \) range (cf., Section A1).
Some might consider asymptotic consistency to be a frivolous academic worry since for large \( n \) one can always change a procedure to make it consistent. However, a study of consistency often gives information on small sample bias, and this premise seems to be operating in the examples of Section 8.

A modification of the Kaplan-Meier least squares procedure is now introduced. The reasons for the modification are two-fold. (1) It is very difficult to analytically study the asymptotic properties of \( \hat{\alpha}^{km}, \hat{\beta}^{km} \). This is caused by the fact that the minimum of (4.9) can occur at a discontinuity point rather than at a point of continuity where expansions could be obtained. Occurrence of the minimum at a discontinuity is frequent in the data examples of Section 8. (2) Although the concept generalizes easily to multiple regression, it is computationally difficult to obtain the Kaplan-Meier least squares estimates with more than one regression variable. This too is caused by the occurrence of minima at discontinuities. A grid search procedure for location of the minimum is far more laborious in two or more dimensions.

The idea behind the modified estimators is to use an initial estimate of \( \beta \) to obtain the Kaplan-Meier weights and to then minimize the sum of squares weighted by these estimated weights.

Various possibilities are available for estimating \( \beta \) initially. The one which is recommended is to run a regression on just the uncensored points as though the usual regression model pertained to them. This gives the initial estimate
\[
\hat{\beta}^0 = \frac{\sum_{\text{uc}} y_i(x_i - \bar{x}_{\text{uc}})}{\sum_{\text{uc}} (x_i - \bar{x}_{\text{uc}})^2}.
\] (4.16)

The notation in (4.16) signifies summation over those pairs \((x_i, y_i^{(+)})\) where \(y_i^{(+) = y_i}\) is uncensored and \(\bar{x}_{\text{uc}}\) is the mean of the \(x_i\) for those pairs. Under the assumption (4.13) the initial estimator (4.16) is asymptotically consistent and normally distributed because the model (1.1) still holds except that the uncensored \(e_i\) are distributed according to

\[
\frac{dF(e)(1 - L(e))}{\int (1 - L(t))dF(t)}.
\] (4.17)

Since the \(e_i\) no longer have zero expectation the uncensored sample intercept is badly biased below \(x\), but the estimate of the slope is not distorted.

Another possibility is to run the regression on all the data treating each observation as uncensored whether it is or not. This and other possibilities were discarded in favor of (4.16) because the only difference between it and the final estimate is that the weights on the uncensored observations are different.

With the initial estimate \(\hat{\beta}^0\) the weights \(w_i(\hat{\beta}^0)\) are computed by applying the Kaplan-Meier process to \(y_i^{(+)} - \hat{\beta}^0 x_i, i = 1, \ldots, n\). The modified estimates (denoted by \(\hat{\alpha}^{km}, \hat{\beta}^{km}\)) are then those values which minimize the weighted sum of squares

\[
\sum_{1}^{n} w_i(\hat{\beta}^0) (y_i^{(+)} - a - bx_i)^2.
\] (4.18)
There is the problem of what to do with the unassigned weight if the largest \( y_i^{(+)} - a - bx_i \) is censored. The customary procedure, used previously in this paper, is to assign it to the largest censored \( y_i^{+} - a - bx_i \). When this was tried on the examples in Section 8, it was discovered that the estimate of \( \beta \) was unstable in a sense to be discussed shortly. The estimate of \( \beta \) obtained by normalizing the weights to sum to one behaved much better so this is the solution which is recommended. To be specific, let

\[
\hat{w}_i^*(\hat{\beta}^0) = \frac{w_i(\hat{\beta}^0)}{\sum_{j=1}^{n} w_j(\hat{\beta}^0)}. \tag{4.19}
\]

Then,

\[
\hat{\beta}_{km} = \frac{\sum_{uc} \hat{w}_i^*(\hat{\beta}^0) y_i (x_i - \bar{x}^*)}{\sum_{uc} \hat{w}_i^*(\hat{\beta}^0) (x_i - \bar{x}^*)^2}, \tag{4.20}
\]

where \( \bar{x}^* = \sum_{uc} \hat{w}_i^*(\hat{\beta}^0) x_i \). For the estimation of \( \alpha \), however, it seems better to assign the remaining weight to the largest \( y_i^{(+)} - \hat{\beta}_{km} x_i \) if it is uncensored so that \( \alpha \) will not be underestimated through removing too much mass from the upper tail. With this convention

\[
\hat{\alpha}_{km} = \frac{y_{km} - \hat{\beta}_{km} x_{km}}{\hat{\beta}^0} \tag{4.21}
\]

where \( \frac{y_{km}}{\hat{\beta}^0}, \frac{x_{km}}{\hat{\beta}^0} \) are computed with the \( \hat{w}_i(\hat{\beta}^0) \).

The obvious temptation is to now use \( \hat{\beta}_{km} \) as a new \( \hat{\beta}^0 \) and recompute (4.20) and (4.21). This can be repeated until \( \hat{\beta}_{km} \) converges. Unfortunately, convergence doesn't always occur. The sequence of \( \hat{\beta}_{km} \)
can settle down to oscillating back and forth between two values. This happened frequently in the examples when the unassigned weight was allocated to the largest censored observation. Furthermore, the two values were not as close as would be desirable. This is the result of the weights changing discontinuously and the remaining weight being assigned to different censored values. By using the normalized weights in (4.19) the same observations are weighted each time in (4.20), which greatly increases the likelihood of convergence. In a couple examples convergence did not occur with (4.20) because a small change in \( \hat{\beta}_{km} \) would shift a censored observation past an uncensored one and thereby change the weights. The two estimated values of \( \beta \) were in close proximity, however.

Since it seems sensible to iterate (4.20) and (4.21), it will be assumed that \( \hat{\beta}_{km} \) and \( \hat{\beta}_{km} \) refer to the estimates obtained through the iterative process.

In regression analysis the most common testing problem is to decide whether or not \( \beta = 0 \). At this moment little is known about the asymptotic behavior of \( \hat{\beta}_{km} \) so it is not possible to construct a test on it. A conjecture is that \( \hat{\beta}_{km} \) is asymptotically equivalent to \( \hat{\beta}_{km} \), but this requires substantiation. The modified estimator \( \hat{\beta}_{km} \) is easier to handle. Under assumption (4.13) it is asymptotically normally distributed with mean \( \beta \). If the true \( \beta \) were used in the computation of the weights in (4.19) (i.e., Kaplan-Meier applied to \( y_{1}^{(+)} - \beta x_{1} \)), the asymptotic variance would be
\[
\frac{1}{n} \int \frac{e^2}{1 - L(e)} \ dF(e), \tag{4.22}
\]

where

\[
\tau^2 = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 = \lim_{n \to \infty} \sum_{i=1}^{n} \hat{w}_i(\beta)(x_i - \overline{x})^2 \tag{4.23}
\]

and \( L(e) = L_{-\alpha/\beta}(e) \) is the common shifted censoring distribution.

The derivation of (4.22) is presented in the Appendix, Section A2, along with the motivation for estimating it by

\[
\sum_{uc} \left( \hat{\beta}_{km} \right)^2 (y_i - \hat{\theta}_{km} - \hat{\beta}_{km} x_i)^2 \sum_{uc} \hat{w}_i(\beta)(x_i - \overline{x}_*)^2
\]

where \( \overline{x}_* = \sum \hat{w}_i(\beta)(x_i) \).

If the heuristic arguments of Section A2 are indeed correct, then the asymptotic variance with \( \hat{\beta}_{km} \) in the weights of (4.19) may or may not equal (4.22). It seems to depend on whether the integral

\[
K = \int e^2 \tag{4.25}
\]

equals zero or not. In the nonzero case the variance (4.22) should be divided by \((1+K)^2\). If true, this says that for some censoring patterns the asymptotic variance for \( \hat{\beta}_{km} \) with weights based on \( \hat{\beta}_{km} \) can be smaller than the asymptotic variance for the (noncomputable) \( \hat{\beta}_{km} \) with weights based on the true \( \beta \). At the moment it is hard to see how to
estimate $K$ from the data. Hopefully, $K$ will not typically differ from zero by very much so it is recommended that (4.24) be used for evaluating the significance of $\hat{\beta}_{km}^{0}$ until further research elucidates the situation.

Brown, Hollander, and Korwar (1974) have alternative methods of testing for the presence or absence of regression with censored data. Their methods adapt Kendall's nonparametric correlation coefficient $\tau$ to accommodate censored data, but they do not give an estimate of the regression line. Assumption (4.15) is required for the Brown-Hollander-Korwar methods.

The modified Kaplan-Meier estimation procedure can easily be extended to the multiple regression problem. The initial estimator $\hat{\beta}^{0}$ can be the usual least square estimator $(\bar{X}^{T}\bar{X})^{-1}\bar{X}^{T}\bar{Y}$ applied to the uncensored observations. The modified estimator $\hat{\beta}_{km}^{0}$ is just a weighted combination $(\bar{X}^{T}\bar{W}(\hat{\beta}^{0})\bar{X})^{-1}\bar{X}^{T}\bar{W}(\hat{\beta}^{0})\bar{Y}$ of the uncensored observations where $\bar{W}(\hat{\beta}^{0})$ is the diagonal matrix with elements

$$w_{ii}(\hat{\beta}^{0}) = \frac{1}{t_{i}} f_{km}(y_{i} - x_{i}\hat{\beta}^{0}).$$ \hspace{1cm} (4.26)

5. Mean Difference Approach.

When the censoring distribution is invariant under change in $x$ (i.e., (4.15)), a different method of estimating $\alpha$ and $\beta$ based on the Kaplan-Meier distribution estimator can be proposed. It uses the idea of Wald (1940) for estimating a linear functional relation by splitting the sample into two groups on the basis of the independent variable.
Assume that the $c_i$ are independently, identically distributed according to $L$, and that the $x_i$ are independently, identically distributed according to $G$. For a selected value $x^0$ divide the sample into those pairs $(x_1, y_1^{(+)})$ with $x_i \leq x^0$ and those with $x_i > x^0$. Under the assumption of randomness of the $x_i$, the variables $y_i$ in each subsample are independently and identically distributed with respective means

$$
E(y_i | x_i \leq x^0) = \alpha + \beta E(x_i | x_i \leq x^0), \quad (5.1a)$$

$$
E(y_i | x_i > x^0) = \alpha + \beta E(x_i | x_i > x^0). \quad (5.1b)
$$

Under the assumption on the $c_i$, the Kaplan-Meier estimator can be applied to the $y_i^{(+)}$ in each subsample. Let

$$
\bar{y}_{1km} = \int y \text{ } d\hat{F}^{km}(y | x \leq x^0), \quad (5.2a)
$$

$$
\bar{y}_{2km} = \int y \text{ } d\hat{F}^{km}(y | x > x^0) \quad (5.2b)
$$

be the Kaplan-Meier mean estimators for the two subsamples. (For simplicity adopt the convention of assigning any unassigned weight to the largest censored observation.) Then, $\bar{y}_{1km}$ and $\bar{y}_{2km}$ are asymptotically consistent, normally distributed estimators of (5.1a) and (5.1b), respectively. Also, the ordinary sample means $\bar{x}_1$ and $\bar{x}_2$ from the two subsamples are asymptotically consistent, normally distributed estimators of $E(x_i | x_i \leq x^0)$ and $E(x_i | x_i > x^0)$, respectively. Therefore, combination of these estimators by fitting a straight line to $(\bar{x}_1, \bar{y}_{1km})$ and $(\bar{x}_2, \bar{y}_{2km})$ yields the asymptotically
consistent, normally distributed estimators
\[ \hat{\beta}^W = \frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1} \]  

and
\[ \hat{\eta}^W = \bar{y}_1^{km} - \hat{\beta}^W \bar{x}_1 \]
\[ = \bar{y}_2^{km} - \hat{\beta}^W \bar{x}_2. \]  

The variance of \( \bar{x}_1 \) given that \( n_1 \) values of \( x_i \) fall below \( x^0 \) is
\[ \frac{\tau_1^2}{n_1} = \frac{1}{n_1} \left\{ \left[ \int_{-\infty}^{x^0} x^2dG_1(x) \right. \right. \]
\[ \left. \left. - \left( \int_{-\infty}^{x^0} xdG_1(x) \right)^2 \right] \right\}, \]  

where
\[ c_1(x) = \begin{cases} 
G(x)/G(x^0) & x \leq x^0, \\
1 & x > x^0. 
\end{cases} \]  

The variance (5.5) is readily estimated from the first subsample \( x_{11}, \ldots, x_{1n_1} (\leq x^0) \) by
\[ \frac{\tau_1^2}{n_1} = \frac{1}{n_1(n_1-1)} \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2. \]  

Similar expressions apply for \( \bar{x}_2 \).
The asymptotic variance of \( y_{1m}^{k} \) as \( n_{1} \to +\infty \) is

\[
V_{1} = \frac{1}{n_{1}} \int \frac{A_{1}^{2}(y) \, dF_{1}(y)}{(1-L_{1}(y))(1-F_{1}(y))^{2}}, \quad (5.8)
\]

where

\[
A_{1}(y) = \int_{y}^{\infty} (1 - F_{1}(t)) \, dt. \quad (5.9)
\]

The variance (5.8) can be estimated from the first subsample \( y_{11}^{(\cdot)}, \ldots, y_{1n_{1}}^{(\cdot)} \) by

\[
\hat{\theta}_{1} = \sum_{uc} \frac{A_{1}^{2}(y_{11})}{(n_{1}-i)(n_{1}-i+1)}, \quad (5.10)
\]

where the summation is over the uncensored observations in the first subsample and

\[
\hat{A}_{1}(y) = \int_{y}^{\infty} (1 - \hat{F}_{1}^{km}(t)) \, dt. \quad (5.11)
\]

The integral in (5.11) is easily evaluated numerically since

\[
\hat{F}_{1}^{km}(y) = \hat{F}_{1}^{km}(y \mid x \leq x^{0}) \quad \text{is a step function}. \quad \text{(For details on (5.10) the reader is referred to Kaplan and Meier (1958).)} \)

A similar expression gives \( \hat{V}_{2} \).

From (5.7), (5.10), and the delta method, the asymptotic variance of (5.3) under the null hypothesis \( H_{0}: \beta = 0 \) is consistently estimated by

\[
\hat{\operatorname{Var}}(\hat{\beta}^{w}) = \frac{\left( \hat{\theta}_{1} + \hat{\theta}_{2} \right)}{(\bar{y}_{2} - \bar{y}_{1})^{2}} + \left( \frac{\hat{\tau}_{1}^{2}}{n_{1}} + \frac{\hat{\tau}_{2}^{2}}{n_{2}} \right) \frac{(\bar{y}_{2} - \bar{y}_{1})^{2}}{(\bar{y}_{2} - \bar{y}_{1})^{4}}. \quad (5.12)
\]
The square root of (5.12) can be used to evaluate the significance (from zero) of the slope $\hat{\beta}^W$.

The asymptotic variance of $\hat{\beta}^W$ for general $\beta$ is not currently known because it involves $\text{Cov}(\overline{y}_1^m, \overline{x}_1^i)$, $i = 1, 2$, and the form of these covariances has not been determined.

In the previous discussion $x^0$ was treated as a preassigned constant, but in practice this is rarely the case. The dividing point $x^0$ is typically selected to be the sample median so that nearly equal numbers of observations fall into each subsample. Although some of the previous probability statements are not precisely correct under this regimen, the distortion is not considered serious.

In connection with the original Wald problem a sequence of authors proposed and studied dividing the sample into three subsamples and using the two end samples to estimate $\alpha$ and $\beta$. For a summary of these papers the reader is referred to Kendall and Stuart (1961, pp. 404-5). This refinement could be invoked here as well, but its effect has not been explored theoretically or numerically.

Before the reader dismisses the mean difference estimators as being too crude for serious consideration, he/she should realize that the slope estimator $\hat{\beta}^W$ has one very definite advantage over the Kaplan-Meier least squares estimators. The assumption that $L_x(c)$ be invariant under change in $x$ is stronger than necessary. For the estimators to be consistent and the distributional statements to be correct it is sufficient that $L_x(c) = L_1(c)$ for $x \leq x^0$ and $L_x(c) = L_2(c)$ for $x > x^0$. The censoring mechanism need only be
invariant within the subsample from which a Kaplan-Meier mean is computed. Because of this the Wald estimator should be less sensitive than the Kaplan-Meier estimators to a systematic change in the censoring distributions from one end of the x range to the other. This would seem to be the case in two of the examples in Section 8, and the weaker requirement of local invariance is exploited further in the next section.

Extension of the mean difference approach to multiple regression raises questions linked to the next section. For example, with two independent variables one approach might be to divide each variable at its median and thereby produce four subsamples. But how should a plane be fitted to the four Kaplan-Meier subsample means?

6. General Censoring.

How should the regression analysis be handled when the censoring pattern apparently does not satisfy (4.13) or (4.15)? A hypothetical example (due to M. V. Johns) of this could arise in a competing risk model. Consider a model in which the length of survival after treatment decreases with the age of the patient at the time of treatment. If a patient could be lost to follow-up from death due to causes other than the one under study, then the censoring distributions might shift downward with increasing age but not at the same rate as the regression line for the particular cause of death under study.

If the independent variable x is controllable, then repeated observations could be made at a fixed set of values \( x_1, \ldots, x_k \).
With a sufficient number of observations the Kaplan-Meier mean estimator could be computed separately at each $x_i$. This would free the mean estimators from any assumptions on the relationship between the censoring mechanism and the independent variable. The Kaplan-Meier mean estimators could then be treated as dependent variables and be regressed against the $x_i$ in an unweighted or weighted analysis. However, since censoring tends to increase the number of observations needed to achieve a prescribed accuracy, this proposal may necessitate a sizeable experiment.

More often the independent variable is not controlled by the investigator, and the statistician must operate with whatever $x$ values he/she receives. It may be possible to cope with this situation by exploiting a local invariance to the censoring mechanism. If groups of neighboring $x$ values are created, then Kaplan-Meier mean estimators computed within the groups can be combined in an unweighted or weighted regression analysis. Distortion through change in the censoring pattern should be eliminated, and hopefully, any loss of accuracy in the regression analysis caused by the grouping should be minimal.

How is it best to balance group size against the number of groups? What rules should govern the creation of the groups? What conditions are needed to theoretically guarantee asymptotic consistency and normality? What about small sample behavior? These questions must be explored before any recommendations can be made about the general use of this type of estimator.
7. **Double Censoring.**

In the case of double censoring there is a lower (left) censoring variable $c_i$ and an upper (right) censoring variable $d_i$ for each observation $y_i$ where $c_i \leq d_i$. The observable variable is

$$y_i^{(\pm)} = \begin{cases} y_i^- = c_i & \text{if } y_i < c_i, \\ y_i & \text{if } c_i \leq y_i \leq d_i, \\ y_i^+ = d_i & \text{if } y_i > d_i, \end{cases}$$ (7.1)

and it is known whether the observation is left or right censored or is uncensored. As with single censoring it may be appropriate to consider the $(c_i, d_i)$ as random or in other situations as arbitrary constants.

Turnbull (1974) devised an iterative scheme to produce the self-consistent (cf., Efron (1967)), maximum likelihood estimator of the distribution function $F(y)$. This estimator could be used in the same fashion as the Kaplan-Meier product-limit estimator is used in Sections 4 and 5 to give estimates of $\alpha$ and $\beta$ in the model (1.1) under the censoring (7.1). More computation is required than for singly censored data, but since both analyses require a fair-sized electronic computer, the extra computation is insignificant once the necessary programming has been accomplished.
8. Heart Transplant Data.

To illustrate the application of the techniques of Sections 4 and 5 the survival data from the Stanford Heart Transplantation Program is selected. The search for the best analysis of this data was a motivating force behind this research.

Details on the Stanford heart transplant data have been presented elsewhere (Turnbull, Brown, and Hu (1974); Crowley and Hu (1974)). For this study 1 October 1967 is regarded as the start of the transplantation program at Stanford, and the cut-off date for this analysis was 1 April 1974. During this time interval 69 patients received heart transplants, and the lengths of their survival (in days) after transplantation are presented in column 2 of Table 1. Some patients in the program never received transplants, and they are not listed in Table 1. Twenty-four transplant patients (column 3, Table 1) were alive at the cut-off date so these constitute censored observations.

Four independent variables, which were felt a priori to possibly have an effect on survival, are:

i) T5 mismatch score,

ii) age (in years) of recipient at time of transplant,

iii) patient's waiting time (in days) from acceptance into the program to transplantation,

iv) calendar time (in days) from the start of the program (1 October 1967).

The values of these variables for each of the transplant patients are listed in columns 5-8 of Table 1.
The T5 mismatch score was developed by Dr. Charles Bieber at Stanford University. It measures the degree of dissimilarity between the donor and recipient tissue with respect to HL-A antigens, and it is therefore potentially related to the phenomenon of rejection of the donor heart by the recipient's immune mechanisms. A low score (< 1.0) represents a good match, and a high score (> 1.0) a poor match. In four cases the tissue typing was not performed so these values are missing in Table 1.

In some patients the cause of death is clearly attributable to rejection of the donor heart. These are labelled with a "1" in column 4, Table 1. For others the primary cause of death (surgical, kidney failure, hepatitis, etc.) was not rejection of the new heart. These are labelled by a "0" in column 4. Some questionable cases were difficult to classify, but fortunately they were few in number.

Since the mismatch score is directed at the rejection phenomenon, the regression of survival on mismatch score was first analyzed with nonrejection related death being treated as censoring. This is the one example in this section where the censoring is not entirely due to the cut-off date. Since survival is the overall aim, the regression of survival on mismatch score was also analyzed with death from any cause as the dependent variable.

Several other mismatch scores were available for analysis, some of which are components of the T5 score. Preliminary study indicated that the T5 score should be the best predictor of survival so it is the one presented here for analysis.
Age of the recipient at transplant (or age at acceptance into the program) was felt to be potentially important to survival. The a priori guess would be that younger patients fair better. Both ages (acceptance or transplant) are so close together it is immaterial which is selected for analysis.

Waiting time from acceptance to transplant was also included for analysis. It could be argued that those patients with longer waiting times may be harder than those with shorter waiting times because a weak patient cannot survive a long wait and will die before receiving a transplant. On the other hand, a long wait may be debilitating to other organs in the patient's system and thereby lessen the chance for long survival.

The length of time from the start of the program to the particular patient's transplant is examined to see if there is any elongation of survival through improvements in the program. Better donor-recipient matching and better immunosuppressive techniques, for example, may have increased the patient's survival prospects.

Other independent variables could be investigated for their effect on survival, but, with one exception, these four are the main ones under consideration. The one exception is whether or not the patient had previous surgery. Statistical analysis indicates that patients with previous surgery survive longer, but a convincing medical explanation of this has not emerged. Since previous surgery is a dichotomous variable, analysis of its effect is more a two sample problem than a regression problem. It is not included for discussion.
here because the range of techniques for the two sample problem is much broader.

The logarithms (base 10) of the survival times were used in the regression analyses. A few very long term survivals make the use of a straight linear scale inappropriate. The log scale may have dampened the effect of the long term survivors a bit too much so a reanalysis might try a less severe transformation like the square root. Patient 38 is listed as having survived for zero days, but for taking logarithms this was changed to 1. Also, patient 60 is listed as having survived for 65 days. In fact, after 50 days he received a second transplant because the first transplanted heart was being rejected, and he survived an additional 15 days. In the regression analysis of rejection related death this patient was treated as having survived 50 days, and in the other analyses with death from any cause his survival time was taken to be 65 days. The four patients with missing mismatch scores were not included in the mismatch analysis.

The exponential approach (Section 2) has not been attempted on this data because Turnbull, Brown, and Hu (1974) did not find a good fit between the exponential model and an earlier set of data.

Crowley and Hu (1974) analyzed the data in Table 1 with the Cox approach (Section 3). The hazard rate approach can be expanded to include all patients admitted to the program whether or not they received a transplant by starting the analysis of each patient at the time of entry into the program. Crowley and Hu also analyzed this expanded data set.
Table 2 displays the estimates $\hat{\alpha}$, $\hat{\beta}$ for the Kaplan-Meier least squares ((4.9), (4.11)), the modified Kaplan-Meier least squares ((4.20), (4.21)), and the Wald mean difference ((5.3), (5.4)) approaches. The estimated asymptotic standard deviations of $\hat{\beta}$ for the modified Kaplan-Meier (4.24) and the Wald mean difference (5.12) estimators are also given. The final column exhibits the weighted sum of squares (4.12) for the unmodified and modified Kaplan-Meier least squares estimators.

Figures 1 to 5 are plots of the data for each of the four independent variables. The T5 mismatch score is plotted twice, once for rejection death and once for death from all causes. The Kaplan-Meier least squares lines and the Wald mean difference lines have been drawn in. In most cases the modified Kaplan-Meier line was very close to the unmodified line so only the unmodified line is displayed.

Some discussion of the results is in order. With one exception the conclusions drawn from the least squares approach agree pretty well with those from the Cox approach of Crowley and Hu (1974). The one exception is the age effect which Crowley and Hu found to be highly significant and dominant in size over the other effects. An extended discussion of this is reserved for last.

For rejection death the T5 mismatch score has a significant (one-tailed $P<.05$) negative slope by either the modified Kaplan-Meier or Wald methods. The slopes and intercepts are very similar by all three least squares methods, and the minimizing sums of squares for the two Kaplan-Meier procedures are essentially identical.
The modified Kaplan-Meier procedure did not converge, but the values are very close.

When death from all causes is considered, the effect of the T5 mismatch score is dampened. The slope is less negative and now differs from zero by approximately one standard deviation. The variance about the regression line has almost doubled. All three least squares estimates are in close agreement, and again the modified procedure did not converge. The dampening of the mismatch influence is reasonable because deaths from other causes should be evenly spread along the x axis if they are unrelated to rejection, and this will flatten the slope.

The agreement between the Kaplan-Meier and Wald estimators is less good for the waiting time to transplant. Kaplan-Meier finds no effect whatsoever whereas Wald indicates length of survival increases with waiting time. The Wald slope differs from zero by about one and a half standard deviations. In Figure 4 the Kaplan-Meier line seems more believable than the Wald line, and the Cox analysis of Crowley and Hu suggests that the waiting time has little effect. Partition of the x axis at the median waiting time does not fairly divide the x range, and it may be that a more equitable division would reduce the Wald slope. In this case the modified Kaplan-Meier least squares estimator converged. Brown, Hollander, and Korwar (1974), using their modifications of Kendall's \( \tau \) on an earlier data set, concluded waiting time was significantly positively correlated with survival.
Censored observations for calendar time occur only on the upper boundary (cf., Figure 5) because all the censoring is created by the cut-off date. Since the censoring pattern changes dramatically from one end of the \( x \) range to the other, the Kaplan-Meier least squares estimators are not to be trusted. An argument similar to the one given next for age would explain why the censoring pattern could produce a negative estimated slope when in fact the true slope is zero or positive. The Wald slope is more believable in this instance and agrees much better with the Cox analysis. Brown, Hollander, and Korwar found a small but significant nonparametric correlation between calendar time and survival on the earlier data set.

Attention now centers on the age effect where the different methods give very different answers. Crowley and Hu's (1974) conditional likelihood analysis of the Cox model found age to be highly significant (one-tailed \( P = .007 \)) with younger patients surviving longer. The modified Kaplan-Meier least squares procedure (cf., Table 2) gives a very small negative slope, which is highly nonsignificant (one-tailed \( P = .42 \)). The Wald mean difference method gives a larger negative slope, which is not quite significant (one-tailed \( P = .08 \)). On the earlier data set, which will not be discussed, Brown, Hollander, and Korwar (1974) found a small but significant (\( P = .02 \)) negative correlation.

The reader is invited to examine the data (Table 1 or Figure 3) and draw his/her own conclusion. I favor the Wald mean difference method as giving the most reasonable estimates and \( P \) value.
my opinion the censoring pattern has caused the conditional likelihood approach to exaggerate the $P$ value and has dampened the regression in the Kaplan-Meier least squares approach.

From inspection of Figure 3 I come away with the impression that the censoring pattern can be grossly described by Figure 6. Now postulate that there is a small negative regression. The lesson to be learned from the counterexample in Section A1 is that the Kaplan-Meier procedure will want to move the uncensored points in the lower left corner toward the center of the estimated distribution $\hat{F}_{km}$. The least squares procedure pulls the lower left uncensored points closer to the middle by lowering the regression line in the left region. Because the regression line is forced to pass through the middle of the right section or suffer a large increase in the sum of squares, the net effect is to lower the regression line in the left region by flattening the slope. Thus, the censoring pattern causes the Kaplan-Meier least squares procedure to underestimate the size and significance of the negative slope.

The Cox analysis compares those who actually died with those at risk just before each death. With increasing survival time the deaths occur predominantly in the higher age range (cf., Figures 3 and 6). The younger patients enter the conditional risk sets $\mathcal{R}(y_i - 0)$ until they are censored, but they rarely enter the numerator of the likelihood as a death. This is evidence in favor of age having an effect since the younger patients have not yet died, but I believe the Cox analysis emphasizes the nondeaths too much. If all the
censored patients were to die immediately, the classical regression line would have a slope which was barely negative ($\hat{\beta} = -0.004$) and very nonsignificant (one-tailed $P=.34$). This is an unfair comparison, but it suggests that the Cox analysis weights the surviving patients very heavily. A few more deaths in the younger ages may have a substantial effect on the Cox P value and the size of $\exp(\hat{\beta})$.

The Wald mean difference method is less affected than the Kaplan-Meier least squares by the change in the censoring distribution with age. It requires only that the censoring be invariant in the left sector and again in the right sector. However, because of the heavy censoring with the lower ages, the Kaplan-Meier mean survival for ages below the median is not as accurately measured as the mean for ages above the median.

The heavier censoring in the lower ages is the result of a trend in the Transplantation Program toward selecting younger patients in recent years. Since the censoring is caused by the cut-off date, a better picture of the age survival relationship should be available in the future, and this should settle the controversy.
APPENDIX


Let the underlying distribution F place probability 1/2 at y = -1 and +1. Let the independent variable x = 1 or 2 in the regression equation α + βx with α = 0, β = 1. Then,

\[ P(y = 0 | x = 1) = P(y = 2 | x = 1) = 1/2, \]

(Al.1)

\[ P(y = 1 | x = 2) = P(y = 3 | x = 2) = 1/2. \]

Assume that all observations are censored at y = 5/2; i.e., for x = 1 and 2

\[ L_x(c) = \begin{cases} 
0, & c < 5/2, \\
1, & c \geq 5/2.
\end{cases} \]  

(Al.2)

For convenience take an equal number of observations n/2 at each x.

With the true regression line subtracted from the observations, the Kaplan-Meier distribution estimator has the limiting (a.s.) values

\[ \lim_{n \to \infty} d_{0,1}^{\hat{F}_{km}}(z) = \begin{cases} 
\frac{1}{2} & \text{at } z = -1, \\
1 & \text{at } z = +1,
\end{cases} \]  

(Al.3)

which are the correct values. The limiting (a.s.) integral of squares (cf., (4.6)) is

\[ \lim_{n \to \infty} \int z^2 d_{0,1}^{\hat{F}_{km}}(z) = 1, \]  

(Al.4)
which is the true variance.

Consider, however, the incorrectly estimated line with \( a = 1/4, b = 3/4 \). Subtracting the incorrect line at \( x = 1 \) still makes it possible for the \( z = y - a - bx \) variable to take on the values \(-1\) or \(+1\), but at \( x = 2 \) the \( z \) variable can either equal \(-3/4\) or be censored at \(+3/4\). These combine to give the limiting (a.s.) values

\[
\lim_{n \to \infty} \frac{P^{km}}{1/4, 3/4}(z) = \begin{cases} 
1/4 & \text{at } z = -1, \\
1/4 & \text{at } z = -3/4, \\
1/2 & \text{at } z = +1.
\end{cases} \tag{A1.5}
\]

The probability distribution (A1.5) gives the limiting (a.s.) integral of squares

\[
\lim_{n \to \infty} \int z^2 \frac{P^{km}}{1/4, 3/4}(z) = \frac{57}{64}. \tag{A1.6}
\]

Since the value in (A1.6) is less than the value in (A1.4), the Kaplan-Meier least squares estimators will obviously not converge to the true regression parameters. The censoring at \( x = 2 \) allows the lower mass of uncensored observations to be moved in toward the center of the \( z \) distribution, and thereby create an incorrect distribution with smaller integral of squares. In fact, the mass could be drawn into the center even more by choosing \( b \) smaller than \( 3/4 \) in order to obtain an integral smaller than (A1.6).

Assume throughout this section that the $F$, $G$, and $L_x$ are sufficiently smooth and well-behaved collectively to permit the required integrals, derivatives, expansions, and limits to exist.

The definition of the modified Kaplan-Meier least squares estimates $\hat{\alpha}^{km}$, $\hat{\beta}^{km}$ is that pair of values $a, b$ which minimize

$$\int (y - a - bx)^2 \, d\hat{F}^{km}_{\beta^0}(y),$$  \hspace{1cm} (A2.1)

where $\hat{\beta}^0$ is an initial estimate of $\beta$ and $d\hat{F}^{km}_{\beta^0}(y_1^{(+)}$) is the weight attached to $y_1^{(+)}$ by the Kaplan-Meier distribution estimator applied to the $y^{(+)}_1 - \hat{\beta}^0 x_1$. The slope estimator can be symbolically written as

$$\hat{\beta}^{km} = \frac{\int y(x' - \bar{x}^{km}_{\beta^0}) \, d\hat{F}^{km}_{\beta^0}(y)}{\int (x - \bar{x}^{km}_{\beta^0})^2 \, d\hat{F}^{km}_{\beta^0}(y)},$$  \hspace{1cm} (A2.2)

where

$$\bar{x}^{km}_{\beta^0} = \int x \, d\hat{F}^{km}_{\beta^0}(y).$$  \hspace{1cm} (A2.3)

The reader should remember that the weight attached to $x_1$ in (A2.3) and to $(x_1 - \bar{x}^{km}_{\beta^0})^2$ in the denominator of (A2.2) is the weight assigned to the corresponding $y_1^{(+)}$ by the Kaplan-Meier method applied to the $y^{(+)}_1 - \hat{\beta}^0 x_1$. How any unassigned weight caused by a largest censored
observation is handled is unimportant to the asymptotic arguments of this section.

Substitution of the relation \( y = \alpha + \beta x + e \) into (A2.2) yields

\[
\hat{\beta}_{kn} = \beta + \frac{\int e(x - \frac{x_{kn}}{\hat{\beta}_0}) dF_{\hat{\beta}_0}(e)}{\int (x - \frac{x_{kn}}{\hat{\beta}_0})^2 dF_{\hat{\beta}_0}(e)},
\]

(A2.4)

where \( dF_{\hat{\beta}_0}(e) \) is the weight attached to \( e = y - \alpha - \beta x \).

Thus, study of the asymptotic distribution of \( \sqrt{n}(\hat{\beta}_{kn} - \beta) \) reduces to a study of the asymptotic distribution of

\[
\sqrt{n} \frac{\int e(x - \frac{x_{kn}}{\hat{\beta}_0}) dF_{\hat{\beta}_0}(e)}{\int (x - \frac{x_{kn}}{\hat{\beta}_0})^2 dF_{\hat{\beta}_0}(e)}.
\]

(A2.5)

If \( \hat{\beta}_0 \rightarrow \beta \) in probability and condition (4.13) of equal censoring about the true regression line holds, then

\[
\frac{x_{kn}}{\hat{\beta}_0} \overset{P}{\rightarrow} \eta = \int x dG(x),
\]

(A2.6)

where \( G \) is the probability distribution generating the \( x_i \) or it is the limiting empirical distribution. The unweighted mean \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \) also converges to \( \eta \) so replacement of \( \frac{x_{kn}}{\hat{\beta}_0} \) with \( \bar{x} \) in (A2.5) does not affect the asymptotic distribution.

Under \( \hat{\beta}_0 \rightarrow \beta \) in probability and (4.13),

\[
\int (x - \bar{x})^2 dF_{\hat{\beta}_0}(e) \overset{P}{\rightarrow} \tau^2 = \int (x - \eta)^2 dG(x).
\]

(A2.7)
Since the denominator in (A2.5) converges to a constant, attention can be focused on the asymptotic distribution of the numerator

\[ \sqrt{n} \iint e(x - \bar{x}) \, d\hat{F}^{km}(e). \]  
(A2.8)

Suppose for the moment that the true value \( \beta \) is used in computing the weights; i.e., \( \hat{\beta}^0 = \beta \). Then, (A2.8) can be rewritten as

\[ \sqrt{n} \iint e(x - \bar{x}) \, d\hat{F}^{km}(e), \]  
(A2.9)

where \( \hat{F}^{km}(e) \) is the Kaplan-Meier distribution estimator applied to the \( e^{(+)} = y^{(+)} - \alpha - \beta x \). Without any censoring the \( e_{i1} \) are independently, identically distributed according to \( F \) with mean \( \mu = 0 \) and variance \( \sigma^2 \). Under condition (4.13) the \( e_{i1}^{(+)} \) have no relationship with the \( x_{i1} \) except by linkage through a subscript. From the weak convergence of \( \sqrt{n}(\hat{F}^{km}(e) - F(e)) \) to a Gaussian process (cf., Breslow and Crowley (1974) and Meier (1974)) and the independence of \( e^{(+)} \) and \( x \), it follows that (A2.9) is asymptotically normally distributed with mean zero. The only question is what is its asymptotic variance.

By the conditional expectation property

\[ E \left[ \iint e(x - \bar{x}) \, d\hat{F}^{km}(e) \right]^2 \]

\[ = E \left[ \iint e(x - \bar{x}) \, d\hat{F}^{km}(e) \right] \left[ \iint f(x - \bar{x}) \, d\hat{F}^{km}(f) \right] \]  
(A2.10)

\[ = E \left\{ \iint e f \, d\hat{F}^{km}(e) \, d\hat{F}^{km}(f) E \left[ (x - \bar{x}) e | F^{km}(.) \right] \right\}, \]
where \((x - \overline{x})_e = (x_i - \overline{x})\) if \(dF_{km}(e) > 0\) and \(x_i\) is the value associated with \(e\), \((x - \overline{x})_e = 0\) otherwise, and similarly for \((x - \overline{x})_f\). By the independence of \(e^{(+)}\) and \(x\), the selection of \(x_i - \overline{x}\) for any uncensored \(e\) is purely random from the collection \(\{(x_j - \overline{x}), j = 1, \ldots, n\}\). Therefore, for \(e = f\),

\[
E \left[ (x - \overline{x})_e^2 \mid \hat{F}_{km}(\cdot) \right] = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2,
\]

(A2.11)

and for \(e \neq f\),

\[
E \left[ (x - \overline{x})_e (x - \overline{x})_f \mid \hat{F}_{km}(\cdot) \right] = \frac{1}{n(n-1)} \sum_{i \neq j} (x_i - \overline{x})(x_j - \overline{x}),
\]

\[
= - \frac{1}{n(n-1)} \sum_{i=1}^{n} (x_i - \overline{x})^2.
\]

(A2.12)

The covariance term (A2.12) is an order of \(n\) smaller than the variance (A2.11) so the off-diagonal terms in the double integral of (A2.10) do not play a role in the asymptotic variance. From (A2.10) - (A2.12) and the definition of \(\tau^2\) in (A2.7),

\[
\lim_{n \to \infty} nE \left[ \int e(x - \overline{x})dF_{km}(e) \right]^2
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \cdot nE \int e^2 (dF_{km}(e))^2,
\]

(A2.13)

\[
= \tau^2 \int e^2 \lim_{n \to \infty} nE(dF_{km}(e))^2.
\]

The asymptotic covariance (cf., Efron (1967); Breslow and Crowley (1974)) of \(\hat{F}_{km}(e)\) and \(\hat{F}_{km}(f)\) for \(e < f\) is
\[
\frac{(1 - F(e))(1 - F(f))}{n} \int e \frac{dF(t)}{(1 - F(t))^2(1 - L(t))} , \quad (A2.14)
\]

where \(L(e) = L_{\alpha/\beta}(e)\) is the common shifted censoring distribution.

Expansion of the right hand side in

\[
E(\hat{d}_{km}^2(e)) = E(\hat{d}_{km}^2(e + \Delta e) - \hat{d}_{km}^2(e))^2 \quad (A2.15)
\]

and application of (A2.14) gives

\[
\lim_{n \to \infty} n E(\hat{d}_{km}^2(e))^2 = \frac{dF(e)}{1 - L(e)} . \quad (A2.16)
\]

Expression (A2.16) in conjunction with (A2.13) shows that the asymptotic variance of (A2.9) is

\[
\tau^2 \int e^2 \frac{dF(e)}{1 - L(e)} . \quad (A2.17)
\]

If the true \(\beta\) could be used in computing the weights in the integrals of (A2.2), expression (A2.7) and (A2.17) show that the asymptotic variance of the slope estimator would be

\[
\frac{1}{n\tau^2} \int e^2 \frac{dF(e)}{1 - L(e)}. \quad (A2.18)
\]

An empirical approximation to this is used in (4.24). The approximation is constructed from
\[ i^2 = \sum_{i=1}^{n} w_i (\hat{\beta}_{\text{km}}^k (x_i - \bar{x})^2, \]
\[ e_i^2 = (y_i - \hat{\beta}_{\text{km}} - \hat{\beta}_{\text{km}}^k x_i)^2, \]
\[ \frac{dF_{\text{km}}^k (e_i)}{n(1 - L(e_i))} \approx (w_i (\hat{\beta}_{\text{km}}^k))^2. \]  

(A2.19)

Is it correct to assume that the initial estimator \( \hat{\beta}^0 \) has no effect on the asymptotic variance of \( \hat{\beta}_{\text{km}}^k \)? By (A2.7) the denominator is not affected, but what about the numerator? Write

\[ \sqrt{n} \int e(x - \bar{x}) dF_{\hat{\beta}^0}^k (e) = \sqrt{n} \int e(x - \bar{x}) dF_{\hat{\beta}}^k (e) \]
\[ + \sqrt{n} \int e(x - \bar{x}) (dF_{\hat{\beta}^0}^k (e) - dF_{\hat{\beta}}^k (e)). \]  

(A2.20)

If the second term on the right hand side converges in probability to zero, then \( \hat{\beta}^0 \) indeed has no effect on the asymptotic variance of \( \hat{\beta}_{\text{km}}^k \).

For a fixed uncensored \( e_i \) but random \( e_j^{(+)} \) for \( j \neq i \), it is easy to see from (4.2) and the consistency of the Kaplan-Meier estimator that with the identification \( e_i = e^{(k)} \)
\[ ndF_{\hat{\beta}}^k (e_i) = \left( \prod_{e_j^{(k)}}^{n-k+1} \left( \frac{n}{n-j+1} \right) \right) \]
\[ \rightarrow (1 - F(e_i)) \frac{1}{(1 - F(e_i))(1 - L(e_i))} \]  

(A2.21)

\[ = \frac{1}{1 - L(e_i)}. \]
Suppose now that $\hat{\beta}^0$ misses $\beta$ by a fixed amount $\Delta \beta$; i.e., $\hat{\beta}^0 = \beta + \Delta \beta$. Each $e_i^{(+)}$ is shifted to $f_i^{(+)} = e_i^{(+)} - \Delta \beta x_i$, and the error and censoring distributions governing $f_i^{(+)}$ are $F$ and $L$ shifted by $\Delta \beta x_i$. The Kaplan-Meier estimator is therefore applied to observations coming from a mixture of different error and censoring distributions where the mixture is controlled by $G(x)$, the probability or empirical distribution of the independent variable. For a hypothetical jump at the point $e$ the limit of the Kaplan-Meier weight on $e$ under the shifting from $e_i^{(+)}$ to $f_i^{(+)}$ is given by

$$\lim_{\Delta \beta \to 0} n d F_{\beta + \Delta \beta}^{km}(e) = \frac{1}{1-L(e)} + \Delta \beta \eta \frac{d}{de} \left( \frac{1}{1-L(e)} \right)$$

(A2.22)

for $\Delta \beta$ small. The parameter $\eta$ in (A2.22) is the limiting mean of the $x$ variable, and $\Delta \beta \eta$ measures the average shift in the mixed population of observations. Since an uncensored observation $e_i$ is shifted to $e_i - \Delta \beta x_i$, the limit of the weight assigned to it by the Kaplan-Meier estimator applied to the mixed population created by $\hat{\beta}^0 = \beta + \Delta \beta$ is given by

$$\lim_{\Delta \beta \to 0} n d F_{\beta^0}^{km}(e_i) = \frac{1}{1-L(e_i - \Delta \beta x_i)} + \Delta \beta \eta \frac{L'(e_i - \Delta \beta x_i)}{(1-L(e_i - \Delta \beta x_i))^2}$$

(A2.23)

Combination of (A2.21) with (A2.23) gives

$$n(d F_{\beta^0}^{km}(e_i) - d F_{\beta}^{km}(e_i)) \to -\Delta \beta (x_i - \eta) \frac{L'(e_i)}{(1-L(e_i))^2} + o(\Delta \beta).$$

(A2.24)
A more complete justification of (A2.22) could be given by showing that

\[
\prod_{f(j) \leq e} \left( \frac{n-j}{n-j+1} \right)^{f_{j}^{k_{m}}} = 1 - \frac{\tilde{p}_{\beta+\Delta\beta}(e)}{f_{\beta+\Delta\beta}(e)}
\]

(A2.25)

\[
P \exp \left\{ - \int e \frac{\int dF(t+\Delta\beta x)(1-L(t+\Delta\beta x))dG(x)}{\int f(t+\Delta\beta x)(1-L(t+\Delta\beta x))dG(x)} \right\}
\]

and

\[
\frac{n}{n-k+1} P \int \left[ (1-F(e+\Delta\beta x))(1-G(e+\Delta\beta x))dG(x) \right]^{-1},
\]

(A2.26)

where the inner integrals in (A2.25) are with respect to \( x \). A power series expansion of the product of (A2.25) and (A2.26) yields (A2.22).

Substitution of (A2.24) (with \( \bar{x} \) replacing \( \eta \) and \( o(\Delta\beta) \) being ignored) into the second term on the right hand side of (A2.20) suggests that for large \( n \)

\[
\sqrt{n} \int e(x - \bar{x}) (d_{\beta}^{k_{m}}(e) - d_{\beta}^{k_{m}}(e))
\]

\[
\approx - \sqrt{n} \Delta\beta \int e(x - \bar{x})^{2} \frac{L'(e)}{1-L(e)} dF^{k_{m}}(e),
\]

(A2.27)

where the relation \( d_{\beta}^{k_{m}}(e) \approx [n(1-L(e))]^{-1} \) has been used in the right hand side. Since \( x \) and \( e \) are independent, (A2.27) seems to say that the variation due to \( \bar{\beta}^{0} \) does not vanish asymptotically unless

\[
K = \int e \frac{L'(e)}{1-L(e)} dF(e)
\]

(A2.28)

equals zero.
Maybe the approximation (A2.27) can be exploited further. With the identification \( \Delta \beta = \beta^0 - \beta \) and with averaging out the \( (x - \bar{x})^2 \), (A2.27) becomes

\[
\sqrt{n} \int e(x - \bar{x}) (d\beta^0_0(e) - d\beta_0(e)) \text{d}x
\]

(A2.29)

\[
\approx - \sqrt{n}(\beta^0 - \beta) \tau^2 K.
\]

Through the iteration in the modified Kaplan-Meier procedure it may be legitimate to make the approximation

\[
\sqrt{n}(\beta^0 - \beta) = \sqrt{n}(\hat{\beta}^{km} - \beta).
\]

(A2.30)

Then, division of (A2.20) by \( \tau^2 \) and utilization of (A2.4) and (A2.29) produces

\[
\sqrt{n}(\hat{\beta}^{km} - \beta) \approx \sqrt{n}(\beta^{km} - \beta) - K\sqrt{n}(\hat{\beta}^{km} - \beta),
\]

(A2.31)

where \( \hat{\beta}^{km} \) is the (noncomputable) Kaplan-Meier estimate with the true \( \beta \) used in the calculation of the weights. Rearrangement of (A2.31) gives finally

\[
\sqrt{n}(\hat{\beta}^{km} - \beta) \approx \sqrt{n}(\beta^{km} - \beta) (1 + K)^{-1}.
\]

(A2.32)

If true, (A2.32) leads to surprising but not totally impossible conclusions. Since \( K \) can be positive or negative depending on the relation of \( L \) to \( F \), the asymptotic variance of the modified Kaplan-Meier estimator can be greater \( (K < 0) \) or less \( (K > 0) \) than the asymptotic variance if the true parameter were used in the calculation of the Kaplan-Meier weights.
Clearly more rigor needs to be applied at various steps in the above argument to determine if it is indeed valid under conditions on F, G, and L. Monte Carlo work might also lend credence to (A2.32) or cast doubt on it. If (A2.32) holds up under further scrutiny, then the size of K and its effect on inferences about \( \beta \) will have to be assessed for empirically reasonable choices of F, G, and L. If the effect is substantial, then methods of estimating K must be explored.
ACKNOWLEDGEMENTS

I thank Byron W. Brown, Bradley Efron, Lincoln Moses, and especially Vernon Johns for extended conversations on censoring and regression analysis. My thanks also go to Jerry Halpern who developed the programs to run the Kaplan-Meier regression procedures, to Marie Hu who did some related programming and computation, to Maria Jedd who drew the figures, and to Martha Thomas who typed the manuscript.
REFERENCES


Table 1. Stanford heart transplant data.

<table>
<thead>
<tr>
<th>ID</th>
<th>survival after tx</th>
<th>dead=1</th>
<th>alive=0</th>
<th>reject=1</th>
<th>no=0</th>
<th>mismatch T5 score</th>
<th>age at tx</th>
<th>waiting time</th>
<th>calendar time</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1.11</td>
<td>54.3</td>
<td>0</td>
<td>97</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.66</td>
<td>40.4</td>
<td>35</td>
<td>214</td>
</tr>
<tr>
<td>7</td>
<td>624</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.32</td>
<td>51.0</td>
<td>50</td>
<td>335</td>
</tr>
<tr>
<td>10</td>
<td>46</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.61</td>
<td>1.89</td>
<td>42.5</td>
<td>11</td>
<td>326</td>
</tr>
<tr>
<td>11</td>
<td>127</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.36</td>
<td>48.0</td>
<td>48.0</td>
<td>24</td>
<td>344</td>
</tr>
<tr>
<td>13</td>
<td>64</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8.78</td>
<td>54.4</td>
<td>16</td>
<td>370</td>
</tr>
<tr>
<td>14</td>
<td>1350</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6.12</td>
<td>54.1</td>
<td>16</td>
<td>391</td>
</tr>
<tr>
<td>16</td>
<td>280</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2.05</td>
<td>49.5</td>
<td>56.9</td>
<td>19</td>
<td>416</td>
</tr>
<tr>
<td>18</td>
<td>23</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2.76</td>
<td>55.3</td>
<td>55.3</td>
<td>17</td>
<td>503</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.13</td>
<td>43.4</td>
<td>43.4</td>
<td>7</td>
<td>496</td>
</tr>
<tr>
<td>21</td>
<td>1024</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.47</td>
<td>54.2</td>
<td>42.8</td>
<td>11</td>
<td>545</td>
</tr>
<tr>
<td>22</td>
<td>39</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.38</td>
<td>58.4</td>
<td>2</td>
<td>550</td>
</tr>
<tr>
<td>23</td>
<td>73</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.96</td>
<td>52.0</td>
<td>82</td>
<td>654</td>
</tr>
<tr>
<td>24</td>
<td>156</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.62</td>
<td>33.3</td>
<td>24</td>
<td>599</td>
</tr>
<tr>
<td>25</td>
<td>1775</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.06</td>
<td>54.2</td>
<td>40.0</td>
<td>15</td>
<td>685</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.58</td>
<td>64.5</td>
<td>64.5</td>
<td>16</td>
<td>714</td>
</tr>
<tr>
<td>30</td>
<td>83</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.89</td>
<td>49.0</td>
<td>49.0</td>
<td>0</td>
<td>838</td>
</tr>
<tr>
<td>32</td>
<td>60</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.38</td>
<td>40.6</td>
<td>40.6</td>
<td>15</td>
<td>825</td>
</tr>
<tr>
<td>33</td>
<td>1536</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2.09</td>
<td>61.5</td>
<td>61.5</td>
<td>18</td>
<td>955</td>
</tr>
<tr>
<td>34</td>
<td>1549</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.87</td>
<td>41.5</td>
<td>41.5</td>
<td>4</td>
<td>951</td>
</tr>
<tr>
<td>36</td>
<td>54</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.87</td>
<td>50.5</td>
<td>50.5</td>
<td>1</td>
<td>963</td>
</tr>
<tr>
<td>37</td>
<td>47</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.75</td>
<td>48.6</td>
<td>48.6</td>
<td>40</td>
<td>1007</td>
</tr>
<tr>
<td>38</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.98</td>
<td>45.5</td>
<td>45.5</td>
<td>57</td>
<td>1110</td>
</tr>
<tr>
<td>39</td>
<td>51</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.0</td>
<td>36.2</td>
<td>36.2</td>
<td>0</td>
<td>1192</td>
</tr>
<tr>
<td>40</td>
<td>1367</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.81</td>
<td>48.6</td>
<td>48.6</td>
<td>1</td>
<td>1198</td>
</tr>
<tr>
<td>41</td>
<td>1264</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1.38</td>
<td>47.2</td>
<td>47.2</td>
<td>20</td>
<td>1240</td>
</tr>
<tr>
<td>45</td>
<td>44</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.35</td>
<td>36.8</td>
<td>36.8</td>
<td>35</td>
<td>1268</td>
</tr>
<tr>
<td>46</td>
<td>994</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1.55</td>
<td>46.1</td>
<td>46.1</td>
<td>82</td>
<td>1315</td>
</tr>
<tr>
<td>47</td>
<td>51</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.08</td>
<td>48.8</td>
<td>48.8</td>
<td>31</td>
<td>1301</td>
</tr>
<tr>
<td>49</td>
<td>1106</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.08</td>
<td>47.5</td>
<td>47.5</td>
<td>40</td>
<td>1410</td>
</tr>
<tr>
<td>50</td>
<td>897</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.08</td>
<td>47.5</td>
<td>47.5</td>
<td>40</td>
<td>1417</td>
</tr>
<tr>
<td>51</td>
<td>253</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.51</td>
<td>52.5</td>
<td>52.5</td>
<td>9</td>
<td>1417</td>
</tr>
<tr>
<td>53</td>
<td>147</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.98</td>
<td>38.9</td>
<td>38.9</td>
<td>66</td>
<td>1499</td>
</tr>
<tr>
<td>55</td>
<td>51</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.82</td>
<td>48.1</td>
<td>48.1</td>
<td>20</td>
<td>1473</td>
</tr>
<tr>
<td>56</td>
<td>875</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.19</td>
<td>41.6</td>
<td>41.6</td>
<td>77</td>
<td>1536</td>
</tr>
<tr>
<td>58</td>
<td>32</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.66</td>
<td>49.1</td>
<td>49.1</td>
<td>2</td>
<td>1511</td>
</tr>
<tr>
<td>59</td>
<td>838</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.93</td>
<td>32.7</td>
<td>32.7</td>
<td>26</td>
<td>1559</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.12</td>
<td>48.9</td>
<td>48.9</td>
<td>32</td>
<td>1616</td>
</tr>
<tr>
<td>63</td>
<td>815</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.12</td>
<td>51.3</td>
<td>51.3</td>
<td>11</td>
<td>1629</td>
</tr>
<tr>
<td>64</td>
<td>551</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1.02</td>
<td>19.7</td>
<td>19.7</td>
<td>56</td>
<td>1691</td>
</tr>
<tr>
<td>65</td>
<td>66</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.68</td>
<td>45.2</td>
<td>45.2</td>
<td>2</td>
<td>1652</td>
</tr>
<tr>
<td>67</td>
<td>288</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.20</td>
<td>48.0</td>
<td>48.0</td>
<td>9</td>
<td>1714</td>
</tr>
<tr>
<td>68</td>
<td>65</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.68</td>
<td>53.0</td>
<td>53.0</td>
<td>4</td>
<td>1725</td>
</tr>
<tr>
<td>69</td>
<td>660</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.97</td>
<td>47.5</td>
<td>47.5</td>
<td>30</td>
<td>1785</td>
</tr>
<tr>
<td>70</td>
<td>25</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.46</td>
<td>26.7</td>
<td>26.7</td>
<td>3</td>
<td>1782</td>
</tr>
<tr>
<td>71</td>
<td>589</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2.16</td>
<td>56.4</td>
<td>56.4</td>
<td>26</td>
<td>1833</td>
</tr>
<tr>
<td>72</td>
<td>592</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.61</td>
<td>29.2</td>
<td>29.2</td>
<td>4</td>
<td>1818</td>
</tr>
<tr>
<td>73</td>
<td>63</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.61</td>
<td>29.2</td>
<td>29.2</td>
<td>4</td>
<td>1818</td>
</tr>
<tr>
<td>74</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.61</td>
<td>29.2</td>
<td>29.2</td>
<td>4</td>
<td>1818</td>
</tr>
</tbody>
</table>
Table 1. Stanford heart transplant data (continued).

<table>
<thead>
<tr>
<th>ID</th>
<th>survival after tx</th>
<th>dead=1 alive=0</th>
<th>reject=1 no=0</th>
<th>mismatch</th>
<th>T5 score</th>
<th>age at tx</th>
<th>waiting time</th>
<th>calendar time</th>
</tr>
</thead>
<tbody>
<tr>
<td>76</td>
<td>499</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.70</td>
<td>52.2</td>
<td>45</td>
<td>1875</td>
</tr>
<tr>
<td>78</td>
<td>305</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.81</td>
<td>49.3</td>
<td>209</td>
<td>2069</td>
</tr>
<tr>
<td>79</td>
<td>29</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.08</td>
<td>54.0</td>
<td>66</td>
<td>1953</td>
</tr>
<tr>
<td>80</td>
<td>456</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.41</td>
<td>46.5</td>
<td>25</td>
<td>1918</td>
</tr>
<tr>
<td>81</td>
<td>439</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.94</td>
<td>52.9</td>
<td>5</td>
<td>1935</td>
</tr>
<tr>
<td>83</td>
<td>48</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3.05</td>
<td>53.4</td>
<td>31</td>
<td>1973</td>
</tr>
<tr>
<td>84</td>
<td>297</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.60</td>
<td>42.8</td>
<td>36</td>
<td>1984</td>
</tr>
<tr>
<td>86</td>
<td>389</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.44</td>
<td>48.9</td>
<td>7</td>
<td>1985</td>
</tr>
<tr>
<td>87</td>
<td>50</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2.25</td>
<td>46.4</td>
<td>59</td>
<td>2057</td>
</tr>
<tr>
<td>88</td>
<td>339</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.68</td>
<td>54.4</td>
<td>30</td>
<td>2035</td>
</tr>
<tr>
<td>89</td>
<td>68</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.33</td>
<td>51.4</td>
<td>138</td>
<td>2151</td>
</tr>
<tr>
<td>90</td>
<td>26</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.82</td>
<td>52.5</td>
<td>159</td>
<td>2173</td>
</tr>
<tr>
<td>92</td>
<td>30</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.16</td>
<td>45.8</td>
<td>309</td>
<td>2344</td>
</tr>
<tr>
<td>93</td>
<td>237</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.33</td>
<td>47.8</td>
<td>27</td>
<td>2137</td>
</tr>
<tr>
<td>94</td>
<td>161</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.20</td>
<td>43.8</td>
<td>3</td>
<td>2178</td>
</tr>
<tr>
<td>95</td>
<td>14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>167</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.46</td>
<td>26.7</td>
<td>12</td>
<td>2207</td>
</tr>
<tr>
<td>97</td>
<td>110</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.78</td>
<td>23.7</td>
<td>20</td>
<td>2264</td>
</tr>
<tr>
<td>98</td>
<td>13</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.77</td>
<td>28.9</td>
<td>95</td>
<td>2361</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.67</td>
<td>35.2</td>
<td>37</td>
<td>2373</td>
</tr>
</tbody>
</table>
Table 2. Regression Estimates for Heart Transplant Data.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\hat{a}$</th>
<th>$\hat{b}$</th>
<th>SD($\hat{b}$)</th>
<th>$\hat{\theta}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T5 mismatch score - rejection death</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaplan-Meier least squares</td>
<td>3.036</td>
<td>-.394</td>
<td>--</td>
<td>.377</td>
</tr>
<tr>
<td>Modified K-M least squares</td>
<td>3.120</td>
<td>-.452</td>
<td>.236</td>
<td>.389</td>
</tr>
<tr>
<td>Wald mean difference (median = 1.08)</td>
<td>3.127</td>
<td>-.491</td>
<td>.183</td>
<td>--</td>
</tr>
<tr>
<td><strong>T5 mismatch score - all death</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaplan-Meier least squares</td>
<td>2.504</td>
<td>-.180</td>
<td>--</td>
<td>.685</td>
</tr>
<tr>
<td>Modified K-M least squares</td>
<td>2.551</td>
<td>-.218</td>
<td>.206</td>
<td>.689</td>
</tr>
<tr>
<td>Wald mean difference (median = 1.08)</td>
<td>2.584</td>
<td>-.253</td>
<td>.215</td>
<td>--</td>
</tr>
<tr>
<td><strong>Age at transplant</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaplan-Meier least squares</td>
<td>2.537</td>
<td>-.058</td>
<td>--</td>
<td>.672</td>
</tr>
<tr>
<td>Modified K-M least squares</td>
<td>2.111</td>
<td>-.036</td>
<td>.165</td>
<td>.721</td>
</tr>
<tr>
<td>Wald mean difference (median = 48.0)</td>
<td>3.331</td>
<td>-.228</td>
<td>$10^{-1}$</td>
<td>$10^{-1}$</td>
</tr>
<tr>
<td><strong>Waiting time to transplant</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaplan-Meier least squares</td>
<td>2.272</td>
<td>.007</td>
<td>--</td>
<td>.699</td>
</tr>
<tr>
<td>Modified K-M least squares</td>
<td>2.275</td>
<td>-.000</td>
<td>.378</td>
<td>.700</td>
</tr>
<tr>
<td>Wald mean difference (median = 25)</td>
<td>2.058</td>
<td>.575</td>
<td>$10^{-2}$</td>
<td>$10^{-2}$</td>
</tr>
<tr>
<td><strong>Calendar time to transplant</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaplan-Meier least squares</td>
<td>2.559</td>
<td>-.255</td>
<td>--</td>
<td>.675</td>
</tr>
<tr>
<td>Modified K-M least squares</td>
<td>2.403</td>
<td>-.115</td>
<td>.192</td>
<td>.684</td>
</tr>
<tr>
<td>Wald mean difference (median = 1473)</td>
<td>1.998</td>
<td>.179</td>
<td>$10^{-3}$</td>
<td>$10^{-3}$</td>
</tr>
</tbody>
</table>
Figure 1. T5 Mismatch Score vs. Survival.  + = alive or nonrejection death,  • = rejection death;  --- Kaplan-Meier least squares line,  ------ Wald mean difference line.
Figure 2. T5 Mismatch Score vs. Survival. + = alive, • = dead; --- Kaplan-Meier least squares line, ---- Wald mean difference line.
Figure 3. Age vs. Survival. + = alive, • = dead; —— Kaplan-Meier least squares line, —— Wald mean difference line.
Figure 4. Waiting Time to Transplant vs. Survival. + = alive, • = dead;
--- Kaplan-Meier least squares line, --- Wald mean difference line.
Figure 5. Calendar Time to Transplant vs. Survival. + = alive, • = dead; ___________________ Kaplan-Meier least squares line, ___________________ Wald mean difference line.
Figure 6. Censoring Pattern in Age vs. Survival.