UCLA
Department of Statistics Papers

Title
On Re-censoring for Censored Paired Data

Permalink
https://escholarship.org/uc/item/1rr8q7wb

Authors
J.J. Lee
K.C. Li
R.M. Elashoff

Publication Date
2011-10-24

On Recensoring for Censored Paired Data
Author(s): J. Jack Lee, Ker-Chau Li, Robert M. Elashoff
Source: *Journal of the American Statistical Association,* Vol. 88, No. 421 (Mar., 1993), pp. 104–118
Published by:
Stable URL: http://www.jstor.org/stable/2290703
Accessed: 18/05/2011 18:42

On Recensoring for Censored Paired Data

J. JACK LEE, KER-CHAU LI, and ROBERT M. ELASHOFF*

In comparing two treatment regimens, a matched-pair design is often used for nonhomogeneous experimental units. A standard method of analyzing the resulting data is to consider the difference within pairs, so that the nuisance parameters due to the heterogeneity between pairs can be eliminated and the parameter of interest—the treatment effect—can be estimated. This procedure becomes invalid when the data are subject to censoring, however. In addition, the available likelihood based methods for handling censored data may suffer from the inconsistency problem incurred by the increasing number of nuisance parameters. In this article, we propose a new approach for eliminating the nuisance parameters based on a crucial notion of "recensoring." Assuming that the censoring time for each observation is known, recensoring is a natural way to force a common lag between the two censoring times within each pair by resetting one of them to a smaller value. After recensoring, we may treat the new observed times as the survival times without censoring. Standard paired data analysis can be applied. M estimation is studied in detail. Root n consistency is established, and the asymptotic variance is obtained. Residual plots can be constructed in the usual way to check model assumptions. Application to a real data set is reported. Comparisons with the work of Holt and Prentice and Wei and Pee are made both in theory and in a simulation study.

KEY WORDS: Consistency; Location models; M estimation; Nuisance parameters; Residual plots.

1. INTRODUCTION

The outcome of medical research is often influenced by many factors. Even in a randomized controlled clinical trial, one needs to control prognostic factors or risk factors to detect the effectiveness of the treatment intervention (Armitage and Gehan 1974; Peto et al. 1976, 1977). The matched-pair design can be used in comparing two treatment regimens with correlated experimental units, as in twin studies (Cederlof, Epstein, Friberg, Hrubec, and Redford 1971; Hauge et al. 1968; Jablon, Neel, Gershawitz, and Atkinson 1967). The extraneous effects, which may or may not be known, can be greatly reduced by forming homogeneous pairs so that both members of a pair are equally likely to respond to the treatment. Hence the comparison of the treatment effect can be made more powerful and less influenced by extraneous effects.

We are interested in the time to an event (e.g., time to death, time to relapse of a disease) where censoring might occur. There are many tests available in the literature for handling censored paired data (Cheng 1984; Dabrowska 1989, 1990; Lachenbruch, Palta, and Woolson 1982; Mantel, Bohidar, and Ciminera 1977; Mantel and Ciminera 1979; Michalek and Mihalko 1983, 1984; O'Brien and Fleming 1987; Schluchter 1985; Wei 1980; Woolson and Lachenbruch 1980, 1981). But quite often our major interest is not only to know whether the null hypothesis of equal treatment effects can be rejected or not, but also to estimate the difference of treatment effects. This estimation problem under censoring is seldom addressed. The difficulty in applying the usual likelihood-based estimation schemes, such as in maximizing the joint likelihood, profile likelihood, or marginal likelihood, has two sides: the increasing number of nuisance

parameters and the presence of censoring. First, each pair carries a "pair effect," which may be different from pair to pair. We may consider one pair effect as one nuisance parameter. Consequently, the number of nuisance parameters increases at the same rate as the number of pairs. Although the general problem of estimating the treatment parameter with the presence of increasing number of nuisance parameters is an old one (Neyman and Scott 1948), our problem gets more complicated because of censoring. A general approach for dealing with nuisance parameters via semiparametric modeling can be found in Bickel, Klassen, Ritov, and Wellner (1990).

Holt and Prentice (1974) first noted the difficulty and presented two estimation methods. The semiparametric one based on the intrapair rank is consistent under the Cox's proportional hazard model, but the parametric one based on the "marginal likelihood" is not (see, for example, Gross and Huber 1987).

One possible way to fix the problem is to reduce the number of nuisance parameters. This can be achieved by either assuming a prior distribution for nuisance parameters (Wild 1983) or by forming a model for the nuisance parameters (pair effects) using covariate variables (Lee 1989). Both resolutions are consistent and efficient if the additional model assumptions hold. Yet for the Bayesian approach, only special cases such as gamma prior for Weibull survival are practically manageable (Wild 1983), and very little results on robustness against prior misspecification are available. On the other hand, modeling nuisance parameters with covariates is also hard to do without convincing knowledge on how the covariates affect the nuisance parameters.

Another approach is to invert a nonparametric test. Wei and Pee (1985) gave a test-based estimation method based on Wei (1980). The Wei test is a generalization of the Gehan–Wilcoxon test in the context of censored paired design. As pointed out by Tsiatis (1986), the test is not optimal in many cases. The method also assumes independent identically distributed censoring times for all pairs, an assumption that

may be questionable for heterogeneous samples. In addition, test-based estimation usually requires substantial computation.

In this article, we present a new method of estimation based on a concept of recensoring for eliminating nuisance effects subject to censoring. We consider a location model, as detailed in Section 2. We assume that censoring times are known for each pair. Then in Section 3, we bring up a favorable situation (3.4), wherein the difference between the two censoring times within each pair is equal to θ (the treatment effect). We find that if (3.4) holds, then the difference between the two observed times within each pair will be symmetric about θ. This finding motivates the idea of recensoring in Section 4.

Obviously, the censoring times in most data do not satisfy (3.4) to begin with. Recensoring offers a natural first step in forcing (3.4) to hold by readjusting one of the two censoring times within pairs to keep a common difference b (4.1). After recensoring, we can apply many standard estimation procedures to the new observed time differences for estimating the treatment effect, as if no censoring had occurred. Then we can compare the derived estimate $\hat\theta^{(b)}$ with b, the common difference in censoring times set earlier in the process of recensoring. The value of b must be chosen appropriately to yield a sensible estimate. The obvious suggestion is then to use the one that yields itself as the estimate: $b = \hat\theta^{(b)}$. We can achieve this either by iteration or by some global search methods.

In Section 5 we study the sampling property of our recensoring method when M estimation is used. Fisher consistency and root n consistency are obtained with the asymptotic variance calculated. The derivation is a modification of the standard M estimation theory to incorporate difficulties arising due to recensoring. In Section 6 we apply recensoring in conjunction with likelihood-based estimators and obtain large-sample properties.

We report asymptotic relative efficiencies in some simple situations in Section 7, and some Monte Carlo results, comparing two recensoring methods to three existing methods, in Section 8. In Section 9 we apply our methods to a real data set from a tumorigenesis experiment, emphasizing graphical aspects. Section 10 studies the hypothesis testing problem based on recensoring. We provide further discussion in Section 11, including issues of sensitivity and information loss caused by recensoring.

2. MODELS AND OBJECTIVES

Suppose that the true survival times T_{ji}^o for the n independent pairs in the study are

$$(T_{11}^o, T_{21}^o), (T_{12}^o, T_{22}^o), \ldots, (T_{1n}^o, T_{2n}^o).$$

We may take $j = 1$ as the treatment group and $j = 2$ as the control group. We consider the location effect model

$$T_{1i}^o = \theta + \mu_i + \varepsilon_{1i}$$

$$T_{2i}^o = \mu_i + \varepsilon_{2i}, \tag{2.1}$$

where θ is the parameter of main interest (the treatment effect), the μ_i's are the nuisance parameters (the pair effects),

and the $(\varepsilon_{1i}, \varepsilon_{2i})$'s are iid random errors. We assume that the joint density of $(\varepsilon_{1i}, \varepsilon_{2i})$ is diagonally symmetric; namely, $(\varepsilon_{1i}, \varepsilon_{2i})$ and $(\varepsilon_{2i}, \varepsilon_{1i})$ have the same distribution.

Instead of observing the survival time T_{ji}^o, we assume that it is subject to a right-censoring time K_{ji}. We can only observe $(T_{1i}, \delta_{1i}, T_{2i}, \delta_{2i})$, where $T_{ji} = \min(T_{ji}^o, K_{ji})$, and

$$\delta_{ji} = 1 \qquad \text{if } T_{ji}^o \le K_{ji}$$

$$= 0 \qquad \text{otherwise,}$$

$$\text{for} \quad i = 1, \ldots, n \quad \text{and} \quad j = 1, 2. \tag{2.2}$$

We also assume that the values of K_{ji}'s are given. A pair of observed times may fall into one of four cases: (a) $\delta_1 = \delta_2 = 1$, both true survival times observed; (b) $\delta_1 = 0$, $\delta_2 = 1$, T_2^o observed but not T_1^o; (c) $\delta_1 = 1$, $\delta_2 = 0$, T_1^o observed but not T_2^o; and (d) $\delta_1 = \delta_2 = 0$, none observed.

Our objective is to estimate the treatment effect in the presence of nuisance parameters (pair effects) and censoring. Point estimators and confidence intervals will be constructed, and sampling properties such as consistency, efficiency, length, and coverage rate of the confidence intervals will also be examined.

In real applications to have a location model (2.1), one might find it necessary to apply a suitable transformation on the data. Specifically, we might start with (X_{1i}^o, X_{2i}^o), $i = 1, \ldots, n$, and find a monotone function $h(\cdot)$ so that $h(X_{1i}^o) = T_{1i}^o$ and $h(X_{2i}^o) = T_{2i}^o$. For example, when taking $h(\cdot)$ as the logarithm transformation, we can convert a multiplicative model

$$X_{1i}^o = \phi B_i Y_{1i}$$

$$X_{2i}^o = B_i Y_{2i} \qquad \text{where} \quad Y_{ji} \text{ are iid} \tag{2.3}$$

into the location model (2.1). In general, although the transformation function $h(\cdot)$ also can be estimated from the data, it might be convenient to try a few candidates from, say, the family of Box–Cox power transformations. This can be accomplished more easily by some graphical displays for checking the symmetry in the errors assumed in the location model (2.1). Unfortunately, we find no such graphical tools available in others' works on censored data. With recensoring, however, we can construct residual plots in the usual way; see Section 9 for illustration.

Note that under the exponential transformation, the location model assumption leads to the accelerated life model instead of the proportional hazards model. For constant explanatory variables, Weibull distribution belongs to both the accelerated life model and the proportional hazard model (Cox and Oakes 1984).

The location model (2.1) assumed in this article is a common assumption imposed on many papers dealing with censored paired data (see, for example, Holt and Prentice 1974; Lachenbruch et al. 1982; Prentice 1978; Schluchter 1985; Wei and Pee 1985; and Woolson and Lachenbruch 1980). The model also contains commonly used log-linear models in survival analysis by taking a logarithm transformation of (2.3) when the survival distribution is exponential, Weibull, or lognormal. We shall also assume that the censoring time

is given. Typically, the censoring time is observed due to an earlier withdrawal from the study, loss to follow-up, or dying of competing risks. When the death or the event of interest happens, however, the censoring time is not directly observed. But a potential censoring time, such as the time from the subject entering the study to the termination of the study, is often available. Note that we do not need to assume an independently identical censoring distribution for all pairs. Our model can be applied when the censoring time varies greatly from pair to pair due to the heterogeneous nature of the paired data and their different entry times. In addition, as long as the censoring time is known, we do not need to assume that the survival time and censoring time are independent. In the next section we shall see how the censoring information can be used to eliminate the nuisance parameters.

3. ELIMINATION OF NUISANCE PARAMETERS

If T_{ji}^o's are observable, a natural way to get rid of the nuisance parameters μ_i is to take the difference in (2.1):

$$T_{1i}^o - T_{2i}^o = \theta + \varepsilon_{1i} - \varepsilon_{2i}.$$

We can rewrite this equation as

$$D_i^o = \theta + W_i, \qquad (3.1)$$

where D_i^o and W_i are the within-pair differences for (T_{1i}^o, T_{2i}^o) and (ε_{1i}, ε_{2i}). The difference D_i^o can be used to form a pivotal quantity for estimating θ. Without censoring, this is a standard location problem. θ can be estimated by many parametric, nonparametric, or robust methods, such as M estimation, L estimation, and R estimation methods (see, for example, Huber 1981).

But difficulty with censoring emerges when (T_{1i}^o, T_{2i}^o) is no longer observable. We can only observe (T_{1i}, δ_{1i}, T_{2i}, δ_{2i}) instead. Parallel to (2.1), we can write the observed time as

$$T_{1i} = \theta + \mu_i + \varepsilon_{1i}^*$$

$$T_{2i} = \mu_i + \varepsilon_{2i}^*, \qquad (3.2)$$

where $\varepsilon_{1i}^* = \min\{\varepsilon_{1i}, K_{1i} - \theta - \mu_i\}$ and $\varepsilon_{2i}^* = \min\{\varepsilon_{2i}, K_{2i} - \mu_i\}$. We may view ε_{ji}^* as the error ε_{ji} subject to a censoring time, which depends on the unknown parameters.

Now take the difference of the two equations in (3.2) to get

$$D_i = T_{1i} - T_{2i} = \theta + \varepsilon_{1i}^* - \varepsilon_{2i}^* = \theta + W_i^*. \qquad (3.3)$$

Unlike (3.1), this equation does not lead to an immediate solution for estimating the location parameter θ. In general the distribution of W_i^* still depends on the unknown parameters and is no longer symmetric about 0.

A critical observation, however, is that under the special condition of

$$K_{1i} - K_{2i} = \theta, \qquad (3.4)$$

$$W_i^* \text{ is symmetric about 0.} \qquad (3.5)$$

This is so because the censoring time is now the same for ε_{j1}^* and ε_{j2}^*. Although the distribution of W_i^* still depends on the pair effect, the symmetry of W_i^* alone is enough to ensure that θ is the symmetric center for the distribution of each observed difference D_i.

The following algebraic explanation shows how condition (3.4) leads to (3.5). Assume that $K_{1i} - K_{2i} = \theta + \tau$; then

$$T_{1i} - T_{2i} = \min(\varepsilon_{1i} + \mu_i + \theta, K_{1i}) - \min(\varepsilon_{2i} + \mu_i, K_{2i})$$

$$= \min(\varepsilon_{1i}, K_{1i} - \mu_i - \theta)$$

$$\quad - \min(\varepsilon_{2i} - \theta, K_{2i} - \mu_i - \theta)$$

$$= \min(\varepsilon_{1i}, K_{2i} - \mu_i + \tau)$$

$$\quad - \min(\varepsilon_{2i} - \theta, K_{2i} - \mu_i - \theta)$$

$$= \theta + \min(\varepsilon_{1i}, K_{2i} - \mu_i + \tau)$$

$$\quad - \min(\varepsilon_{2i}, K_{2i} - \mu_i).$$

Therefore, $T_{1i} - T_{2i}$ is symmetric about θ and $E(T_{1i} - T_{2i}) = \theta$ iff $\tau = 0$.

Thus under (3.4), most location estimators when applied to D_i's will estimate θ well. One important consequence is that we will have an equal amount of censoring in both coordinates if (3.4) holds. The expected numbers of cases (b) and (c) of observed time as defined in the sentence following (2.2) will be the same. The effect of the nuisance parameter is still balanced within each pair under censoring. Then the resulting intrapair difference is free from the influence of nuisance parameters.

The influence of the nuisance parameters and censoring on the pair difference may be better described by the following example. As in model (2.1), suppose that one pair of data (hence the subscript i will be dropped) is generated without the error term. When $\theta = 1$ and K_1 is fixed at 10, we vary the pair effect μ and K_2 to examine their effect on the pair difference $D = T_1 - T_2$.

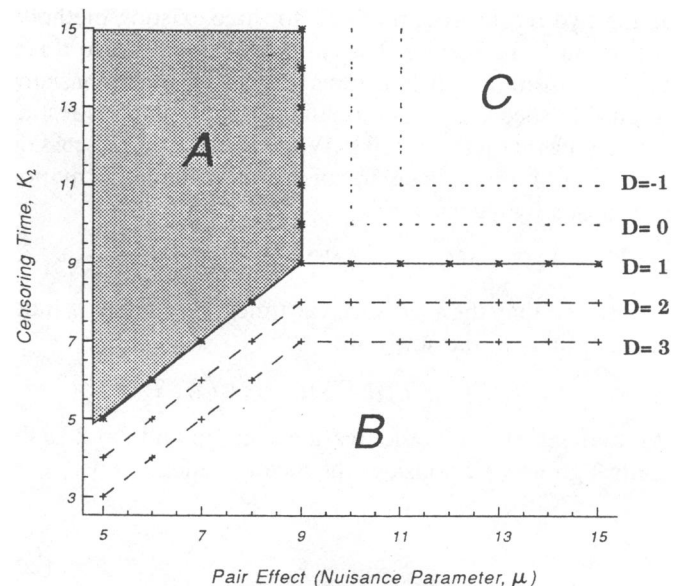The contour plot of D is displayed in Figure 1. As can be

Figure 1. Contour Plot of the Differences D Showing Confounding Between the Pair Effect and Censoring When $\theta = 1$ and $K_1 = 10$.

seen, the upper left corner (denoted as part A) indicates that both T_1^o and T_2^o are observed; it happens when $\mu < 9$ and $K_2 - \mu > 0$. In this case $D = T_1^o - T_2^o = 1$, which gives the correct estimate of θ. The rest of the plane can be divided into two parts: the lower part (part B) and the upper part (part C). In part B, T_2^o is always censored while T_1^o is censored when $\mu > 9$. In either case, $D > 1$ and the contour lines of D are drawn in Figure 1. Similarly, in part C, T_1^o is always censored but T_2^o is censored only when $K_2 < \mu$. Both cases have $D < 1$, with the contour lines drawn in Figure 1. One interesting observation is that when $K_2 = K_1 - \theta = 9$, the pair difference, D, is always equal to 1, which is θ regardless of the value of μ. Because we do not know μ, and in fact μ varies from pair to pair, it seems that setting $K_1 - K_2 = \theta$ is a sensible way to make the pair difference invariant of the pair effect, μ.

The favorable situation depicted by (3.4) rarely occurs in practice. But it helps us to motivate the idea of recensoring, a natural first step in forcing (3.4) to hold approximately.

4. RECENSORING

In this section we describe the idea of recensoring. To proceed, we first decompose (3.4) into two equations:

$$K_{1i} - K_{2i} = b, \quad \text{for} \quad i = 1, \ldots, n \qquad (4.1)$$

and

$$\theta = b. \qquad (4.2)$$

Equation (4.1) sets the difference of the censoring times within each pair to a constant b. Equation (4.2) further equates this constant to the unknown parameter θ.

For each b, (4.1) can be achieved by the following procedure.

Recensoring (RC) Step. First, reset the new censoring time $K_{1i}^{(b)}$ and $K_{2i}^{(b)}$ as follows:

(a) If $K_{1i} - K_{2i} < b$, let
$K_{1i}^{(b)} = K_{1i}$ and $K_{2i}^{(b)} = K_{1i} - b$;

(b) Otherwise, let
$K_{1i}^{(b)} = K_{2i} + b$ and $K_{2i}^{(b)} = K_{2i}$.

Then recensor the original data accordingly to obtain the new observed time $T_{ji}^{(b)} = \min\{T_{ji}, K_{ji}^{(b)}\}$ and the new censoring indicator

$$\delta_{ji}^{(b)} = 1 \qquad \text{if } \delta_{ji} = 1 \quad \text{and} \quad T_{ji} \le K_{ji}^{(b)}$$

$$= 0 \qquad \text{otherwise.}$$

In addition, we denote $D_i^{(b)}$ as the difference of $T_{1i}^{(b)}$ and $T_{2i}^{(b)}$.

After recensoring, we can proceed with any estimation procedure.

Estimation Step. Several estimation methods can be applied in estimating θ. For example, we can estimate θ from $D_i^{(b)}$ by M, L, or R estimates as if we have a standard location

problem in the no censoring case. Another possibility is to use the information of $D_i^{(b)}$ and $\delta_{ji}^{(b)}$ as in the likelihood-based estimation. We denote an estimator of θ as $\hat{\theta}^{(b)}$. The final estimate $\hat{\theta}$ for θ is the b that solves the equation

$$\hat{\theta}^{(b)} = b. \qquad (4.3)$$

In the next section we examine the sampling properties of our method under the framework of M estimation. The likelihood-based estimation is discussed in Section 6.

5. M ESTIMATION

Consider the use of M estimation in the estimation step described in the previous section. Without censoring, the standard M estimator $\hat{\theta}$ for the location parameter θ, based on an iid sample, $\{X_i, i = 1, \ldots, n\}$, can be obtained by solving

$$\sum_1^n \Psi(X_i - \theta) = 0$$

for some function $\Psi(\cdot)$.

When X_i's are iid, under suitable regularity conditions $\hat{\theta}$ can be shown to have an asymptotic normal distribution with mean θ and the variance given by

$$\frac{A}{B^2} = \frac{\int \Psi^2(x - \theta)\, dF(x)}{\left\{\dfrac{d}{d\theta} \int \Psi(x - \theta)\, dF(x)\right\}^2},$$

where $x \sim F(x)$.

Now applying M estimation, the estimation step in Section 4 is to solve

$$\sum_{i=1}^n \Psi(D_i^{(b)} - \theta) = 0 \qquad (5.1)$$

for finding $\hat{\theta}^{(b)}$. Thus from (4.3), we see that our final estimate of θ, $\hat{\theta}$, is the solution of the equation

$$\sum_{i=1}^n \Psi(D_i^{(b)} - b) = 0. \qquad (5.2)$$

Equation (5.2) is obviously more complicated than the standard M estimation equation. In Sections 5.1–5.3, we establish the Fisher consistency, the large-sample consistency, and the asymptotic normality of the estimate $\hat{\theta}$ obtained by solving (5.2). Section 5.4 discusses the numerical convergence property of an iterative method for solving (5.2). Before proceeding, it is useful to observe the following properties of $D_i^{(b)}$:

$D_i^{(b)}$ is nondecreasing and

$$D_i^{(b)} - b \text{ is nonincreasing in } b. \qquad (5.3)$$

The first property is obvious; to see that the second one holds, observe that for $b \ge K_{1i} - K_{2i}$, $D_i^{(b)} - b = \theta - b + \min\{\varepsilon_{1i}, K_{1i} - \theta - \mu_i\} - \min\{\varepsilon_{2i}, K_{1i} - b - \mu_i\}$, and for $b < K_{1i} - K_{2i}$, $D_i^{(b)} - b = \theta - b + \min\{\varepsilon_{1i}, K_{2i} + b - \theta - \mu_i\} - \min\{\varepsilon_{2i}, K_{2i} - \mu_i\}$.

5.1 Fisher Consistency

In this subsection we consider the population version of (5.2). We show that the solution of the equation

$$E \sum_{i=1}^{n} \Psi(D_i^{(b)} - b) = 0 \qquad (5.4)$$

is equal to θ under suitable regularity conditions on Ψ.

Theorem 5.1. Equation (5.4) has a unique solution $b = \theta$ under the following conditions: (C.1) $\Psi(\cdot)$ is antisymmetric about 0; (C.2) $\Psi(\cdot)$ is nondecreasing and $\Psi(t) = 0$ iff $t = 0$; (C.3) $|\Psi(t)| \le C_0 + C_1 |t|$ for some C_0, C_1; (C.4) $(\varepsilon_{1i}, \varepsilon_{2i})$ has a finite expectation for each i; and (C.5) The distribution of $(\varepsilon_{1i}, \varepsilon_{2i})$ is symmetric about the 45° line $\varepsilon_1 = \varepsilon_2$, and the support is the entire R^2.

Proof. First, observe that (C.3) and (C.4) imply that the expectation $E\Psi(D_i^{(b)} - b)$ exists for each i. Then, as discussed in Section 3, $D_i^{(\theta)} - \theta$ is symmetric about 0, implying that $E\Psi(D_i^{(\theta)} - \theta) = 0$ due to (C.1). Hence we see that θ is a solution of (5.4). It remains to show that θ is the only solution of (5.4). This is easy to verify due to (5.3), (C.2), and (C.5), which ensure that $\Psi(D_i^{(b)} - b)$ is nonincreasing in b and is decreasing in b with a positive probability.

5.2 Large-Sample Consistency

Fisher consistency often implies large-sample consistency with additional regularity conditions. We need a suitable asymptotic setting to regularize the sequences of the nuisance parameters K_{1i}, K_{2i}, and μ_i's.

Theorem 5.2. Under (C.1)–(C.5) and two more conditions, (C.6) each $(\varepsilon_{1i}, \varepsilon_{2i})$ has a bounded second moment and (C.7) there exist positive numbers C_2, C_3 such that $\#\{|K_{1i} - \mu_i| < C_2 \text{ and } |K_{2i} - \mu_i| < C_2\} \ge C_3 \cdot n$, the solution of (5.2) converges to θ in probability.

Proof. First, (C.6) and (C.3) imply that var$(1/n \sum_{i=1}^{n} \Psi(D_i^{(b)} - b))$ converge to 0 at the rate $1/n$. This implies that for any b,

$$\frac{1}{n} \sum_{i=1}^{n} \Psi(D_i^{(b)} - b) - E \frac{1}{n} \sum_{i=1}^{n} \Psi(D_i^{(b)} - b) \to 0$$

$$\text{in probability.} \quad (5.5)$$

On the other hand, (C.7), (C.5), and (C.2) imply that

$$\lim_{n \to \infty} \left| E \frac{1}{n} \sum_{i=1}^{n} \Psi(D_i^{(b)} - b) \right| > 0 \quad \text{for any} \quad b \ne \theta. \quad (5.6)$$

Due to the monotonicity of $\Psi(D_i^{(b)} - b)$ in b, (5.5) and (5.6) imply that in the solution of (5.2) $\hat{\theta}$ converges to θ in probability.

5.3 Asymptotic Variance

Following a routine argument in M estimation, we see that

$$\hat{\theta} \approx \theta + B_n^{-1} \sum_{i=1}^{n} \Psi(D_i^{(\theta)} - \theta),$$

and the asymptotic variance for $\hat{\theta}$ is A_n/B_n^2, where

$$A_n = \sum_{i=1}^{n} \text{var}(\Psi(D_i^{(\theta)} - \theta))$$

and

$$B_n = \left[\sum_{i=1}^{n} \frac{d}{db} E\Psi(D_i^{(b)} - b) \right]_{b=\theta}.$$

The asymptotic normality of $\hat{\theta}$ follows from the asymptotic normality of $1/\sqrt{n} \sum_{i=1}^{n} \Psi(D_i^{(\theta)} - \theta)$. The latter holds by the central limit theorem when we regard (μ_i, K_{1i}, K_{2i}), $i = 1, \ldots, n$ as independent realization from a common joint distribution so that $\Psi(D_i^{(\theta)} - \theta)$, $i = 1, \ldots, n$ are iid with mean 0 and finite variance. But weaker regularity conditions, not pursued here, can be established from the well-known Lindeberg–Feller condition.

Suppose that (C.8) $\Psi(\cdot)$ is differentiable and has continuous derivative $\Psi'(\cdot)$. Then, as outlined in Appendix A, we find that

$$B_n = \sum_{i:\theta < K_{1i} - K_{2i}} E[\Psi'(D_i^{(\theta)} - \theta)\delta_{1i}^{(\theta)}]$$

$$+ \sum_{i:\theta > K_{1i} - K_{2i}} E[\Psi'(D_i^{(\theta)} - \theta)\delta_{2i}^{(\theta)}].$$

Example 5.1: RC mean method (RMN). For the RC mean method, we take $\Psi(x) = x$. For this method, $\hat{\theta}^{(b)}$ is the sample mean of $D_i^{(b)}$, and from Appendix A,

$$-B_n = \sum_{\theta < K_{1i} - K_{2i}} P(\delta_{1i}^{(\theta)} = 1) + \sum_{\theta > K_{1i} - K_{2i}} P(\delta_{2i}^{(\theta)} = 1)$$

$$= \frac{1}{2} \sum_{i=1}^{n} [P(\delta_{1i}^{(\theta)} = 1) + P(\delta_{2i}^{(\theta)} = 1)],$$

where the last equality is due to the fact that $P\{\delta_{1i}^{(\theta)} = 1\} = P\{\delta_{2i}^{(\theta)} = 1\}$. Hence the asymptotic variance can be estimated by $\hat{A}_n/\hat{B}_n^2$, where

$$\hat{A}_n = \sum_{i=1}^{n} (D_i^{(\hat{\theta})} - \hat{\theta})^2$$

and

$$-\hat{B}_n = \frac{1}{2} \sum_{i=1}^{n} (\delta_{1i}^{(\hat{\theta})} + \delta_{2i}^{(\hat{\theta})}) = \frac{1}{2} \text{ (total number of uncensored observations after recensoring).}$$

For a given sample, the asymptotic variance can be approximated by the variance of $\hat{\theta}$ calculated from the data after being recensored at $\hat{\theta}$ divided by the square of the proportion not censored by recensoring. When there is no censoring, $B_{ni} = -1$, and it returns to the usual M estimation situation. With censoring, the asymptotic variance is the sample variance calculated from the recensored data, inflated by the squared reciprocal of the uncensored proportion.

Example 5.2: RC median method (RMD). When the median is used in the estimation step, we may take

$$\Psi(x) = -1 \qquad \text{if } x < 0$$
$$= 0 \qquad \text{if } x = 0$$
$$= 1 \qquad \text{if } x > 0.$$

Although Ψ is not differentiable and (C.8) does not hold, the asymptotic variance can be formed similarly. The term B_n is also given in Appendix A:

$$-B_n = \sum_{i:\theta < K_{1i}-K_{2i}} \left\{ 2 \int_{-\infty}^{K_{2i}} f_{2i}(t_2)\, dF_{2i}(t_2) \right.$$
$$+ [1 - F_{2i}(K_{2i})]f_{1i}(K_{1i}) \Big\}$$
$$+ \sum_{i:\theta > K_{1i}-K_{2i}} \left\{ 2 \int_{-\infty}^{K_{1i}} f_{1i}(t_{1i})\, dF_{1i}(t_1) \right.$$
$$+ [1 - F_{1i}(K_{1i})]f_{2i}(K_{2i}) \Big\} .$$

To ensure that θ is the only value for solving (5.4), we need an additional condition that the probability of doubly censoring, after recensoring correctly, is less than 50%. Thus for heavy censoring cases, RMD is not valid.

5.4 Convergence of the Iteration Process

One way of solving (5.2) is to iterate between the recensoring step and the estimation step. This process converges in general, unless the solution $\hat\theta$ is not unique. An outline of the proof follows. Suppose that the initial value $b = b_1$ is larger than the solution $\hat\theta$. It is enough to show that $b_1 > b_2 \geq \hat\theta$.

First, due to the monotonicity of $D_i^{(b)} - b$ [see (5.3)], we have

$$\sum_{i=1}^n \Psi(D_i^{(b_1)} - b_1) < \sum_{i=1}^n \Psi(D_i^{(\hat\theta)} - \hat\theta) = 0.$$

For the fixed b_1, the solution $\hat\theta^{(b_1)} = b_2$ of (5.1) must have $b_2 < b_1$ due to (C.2).

Next, we want to show that $b_2 \geq \hat\theta$. Because $D_i^{(b)}$ is nondecreasing in b, we have

$$\sum_{i=1}^n \Psi(D_i^{(b_1)} - \hat\theta) \geq \sum_{i=1}^n \Psi(D_i^{(\hat\theta)} - \hat\theta) = 0 = \sum_{i=1}^n \Psi(D_i^{(b_1)} - b_2).$$

Comparing the first and the last term in this expression, we see that $\hat\theta \leq b_2$, again due to (C.2).

Note that for the median estimate, as in the censor-free context, the solution may not be unique. This creates some problems later on in our simulation study. But to solve (5.2), we can always conduct a global search, such as the up–down method, which might not be difficult for simple estimates like the median.

6. RC LIKELIHOOD METHOD

In the previous section we showed the properties of the recensoring/M estimation methods where a suitable kernel function Ψ is applied to the difference of observed times within each pair. The idea of recensoring can also be applied to likelihood-based estimation methods. As can be seen later, the argument of the Ψ function of the likelihood methods involves the censoring indicators in addition to the differences. The Ψ function is also different for each of the four

cases defined after (2.2). In this section we discuss how the recensoring can be applied for the likelihood-based method.

Holt and Prentice (1974) proposed a likelihood-based estimator (denoted MAL) using the property that the difference of the log survival time is marginally sufficient for the parameter of interest in the location model. This property breaks down for censored cases, however, and the resulting estimate is no longer consistent (Dabrowska 1989). In this section we show how recensoring can restore the consistency for the likelihood-based estimators. We consider the exponential model as an example in constructing the likelihood, but the results in Section 6.3 as well as in Appendixes A and B can be applied to any survival distribution with suitable modifications.

6.1 No Censoring

When there is no censoring, $D_i = T_{1i} - T_{2i} = T_{1i}^o - T_{2i}^o = \theta + W_i$. The usual likelihood function is $L(\theta) = \prod_{i=1}^n f(D_i - \theta)$, where f is the density of W_i. For the case of the exponential survival, the likelihood function becomes

$$L_1(\theta) = \prod_{i=1}^n \frac{e^{D_i-\theta}}{(1 + e^{D_i-\theta})^2}. \tag{6.1}$$

Considering the multiplicative form of (2.3) for the moment, Holt and Prentice (1974) viewed $r_i = e^{D_i} = x_{1i}/x_{2i}$ as a marginal sufficient statistics for θ and considered the associated likelihood

$$L_2(\theta) = \prod_{i=1}^n \frac{e^{-\theta}}{(1 + r_i e^{-\theta})^2}, \tag{6.2}$$

which is of course proportional to $L_1(\theta)$. On the other hand, the likelihood based on the joint distribution of X_{1i} and X_{2i} is

$$L(\phi, \mathbf{B}) = \prod_{i=1}^n \frac{1}{\phi B_i^2} e^{-\left(\frac{X_{1i}}{\phi B_i} + \frac{X_{2i}}{B_i}\right)}.$$

The maximum likelihood estimator (MLE) of B_i is

$$\hat{B}_i = \frac{1}{2}\left(\frac{X_{1i}}{\phi} + X_{2i}\right).$$

Substituting these for B_i's and replacing ϕ with θ (note: $\theta = \log \phi$), we obtain the profile likelihood

$$L_3(\theta) = 4e^{-2} \prod_{i=1}^n \frac{e^{-\theta}}{(x_{1i}e^{-\theta} + x_{2i})^2}, \tag{6.3}$$

which is proportional to $L_2(\theta)$ and $L_1(\theta)$. This shows the equivalence between different likelihood approaches.

6.2 Censoring But Without Recensoring

Focusing on the observed ratio r_i, Holt and Prentice (1974) assumed the equal censoring time (i.e., $K_{1i} = K_{2i}$). Exploring the order relationship between the unobserved ratio r_i^o ($r_i^o = X_{1i}^o/X_{2i}^o$) and r_i, a likelihood function is constructed:

$$L(\theta) = \underbrace{\prod f(r_i)}_{(a)} \underbrace{\prod S(r_i)}_{(b)} \underbrace{\prod F(r_i)}_{(c)}.$$

Here f, S, and F are the density, the survival, and the cumulative distribution function of r_i^o and (a), (b), and (c)

refer to the first three cases of observed time as given after (2.2). For the exponential survival that we are interested in:

$$L(\theta) = \prod_{(a)} \frac{e^{-\theta}}{(1 + r_i e^{-\theta})^2} \prod_{(b)} \frac{1}{1 + r_i e^{-\theta}} \prod_{(c)} \left(1 - \frac{1}{1 + r_i e^{-\theta}}\right).$$

(6.4)

As explained later, the MLE of this likelihood function produces an inconsistent estimator unless $\theta = 0$ when $K_{1i} = K_{2i}$.

We take a closer look at the likelihood in (6.4). For case (b) [case (c) is similar], we see that for some constant γ,

$$\frac{x_1}{x_2} = \gamma \quad \text{and} \quad \delta_1 = 0, \quad \delta_2 = 1 \quad \text{implies} \quad \frac{x_1^o}{x_2^o} > \gamma.$$

But the density $f(x_1/x_2 = \gamma, \delta_1 = 0, \delta_2 = 1)$ is not proportional to $P((x_1^o/x_2^o) > \gamma)$. Hence (6.4) is not a true likelihood. The true likelihood should be based on

$$f\left(\frac{x_1}{x_2} = \gamma, \delta_1 = 0, \delta_2 = 1\right)$$

$$= \int_0^{K_2} P\left(\frac{x_1^o}{x_2^o} > \gamma \mid x_2^o = x_2\right) f(x_2^o = x_2) dx_2.$$

But this likelihood depends on pair effects, which cannot be eliminated.

6.3 Applying Recensoring

To get a consistent estimator, we may apply the recensoring step in Section 4 in conjunction with the likelihood kernel and call the method the recensoring maximum likelihood method (RML). Let us return to the log scale of the survival time again. The estimation step is to solve

$$\max_{\theta} \left\{ \prod_{(a)} g(D_i^{(b)} - \theta) \prod_{(b)} (1 - G(D_i^{(b)} - \theta)) \right.$$

$$\left. \times \prod_{(c)} G(D_i^{(b)} - \theta) \right\},$$

where $D_i^{(b)} = T_{1i}^{(b)} - T_{2i}^{(b)}$ and g, G are the density and the distribution function of $W_i = D_i^o - \theta$. The log-likelihood equation is

$$\sum_{(a)} \frac{\frac{d}{d\theta} g(D_i^{(b)} - \theta)}{g(D_i^{(b)} - \theta)} + \sum_{(b)} \frac{g(D_i^{(b)} - \theta)}{1 - G(D_i^{(b)} - \theta)}$$

$$+ \sum_{(c)} \frac{-g(D_i^{(b)} - \theta)}{G(D_i^{(b)} - \theta)} = 0. \quad (6.5)$$

For exponential survival, we have $g(w) = e^w/(1 + e^w)^2$, the logistic distribution. Substituting this term in (6.5), we obtain

$$\sum_{(a)} \left(\frac{2e^{D_i^{(b)} - \theta}}{1 + e^{D_i^{(b)} - \theta}} - 1\right) + \sum_{(b)} \frac{e^{D_i^{(b)} - \theta}}{1 + e^{D_i^{(b)} - \theta}}$$

$$- \sum_{(c)} \left[\frac{1}{1 + e^{D_i^{(b)} - \theta}}\right] = 0$$

or, equivalently,

$$\sum_{(a),(b)} \frac{e^{D_i^{(b)} - \theta}}{1 + e^{D_i^{(b)} - \theta}} - \sum_{(a),(c)} \frac{1}{1 + e^{D_i^{(b)} - \theta}} = 0. \quad (6.6)$$

Equation (6.6) takes a form similar to the M estimation (5.1). Let $\hat{\theta}^{(b)}$ be the solution of (6.6). We shall discuss Fisher consistency and other properties of $\hat{\theta}$, the estimate of θ obtained by solving (6.6) and (4.3). The arguments are similar to those in Section 5. Let

$$\Psi(D_i^{(b)} - \theta, \delta_{1i}^{(b)}, \delta_{2i}^{(b)})$$

$$= \frac{e^{D_i^{(b)} - \theta}}{1 + e^{D_i^{(b)} - \theta}} \delta_{2i}^{(b)} - \frac{1}{1 + e^{D_i^{(b)} - \theta}} \delta_{1i}^{(b)},$$

where $\delta_{1i}^{(b)}$ and $\delta_{2i}^{(b)}$ are the readjusted indicators of censoring defined in the recensoring step of Section 4.

The population version of (6.6) takes the form

$$\sum_{i=1}^{n} E\Psi(D_i^{(b)} - \theta, \delta_{1i}^{(b)}, \delta_{2i}^{(b)}) = 0.$$

Note that $\delta_{1i}^{(b)}$ is nondecreasing in b and $\delta_{2i}^{(b)}$ is nonincreasing in b.

As shown in Appendix B, we can directly verify that (1) $\theta = b$ is a solution, (2) for any fixed b, Ψ is decreasing in θ, and (3) with $\theta = b$, Ψ is decreasing in b. Therefore, similar arguments to those in Section 5 can be used to prove the Fisher consistency and large-sample consistency. But although we are unable to show that the iterative procedure always converges, at least we can find $\hat{\theta}$ by solving

$$\sum_{i=1}^{n} \Psi(D_i^{(b)} - b, \delta_{1i}^{(b)}, \delta_{2i}^{(b)}) = 0$$

directly, using the usual up–down method due to the monotonicity in b.

The asymptotic variance of $\hat{\theta}$ is given in Appendix A. Note that the true distribution for the difference of the log survival time need not be logistic. The asymptotic variance can be calculated for a given survival distribution and a suitable kernel.

7. ASYMPTOTIC RELATIVE EFFICIENCY (ARE)

The full efficiency comparison among different estimators may not be simple, because it depends on the complete specification of the survival distribution (ε_{ji} or W), the nuisance parameters (μ_i's), and the censoring times (K_{ji}'s). In the next section we evaluate these methods under different combinations of the choices on the survival distribution, pair effects, and censoring times via Monte Carlo studies. In this section the asymptotic variance is computed and compared only for simple cases. For convenience, we assume the exponential survival without the pair effect; namely, $B_i = 1$ in the multiplicative model (2.3) and Y_{ji} follows an exponential distribution with unit scale. The censoring time (K_{1i}, K_{2i}) is assumed to be (∞, ∞) when there is no censoring and ($K + \theta$,

K) when there is censoring. Four methods are considered here

Method 1: RMD, the recensoring median estimator (see Ex. 5.2 in Sec. 5.3)

Method 2: RMN, the recensoring mean estimator (see Ex. 5.1 in Sec. 5.3)

Method 3: RML, the recensoring MLE applying the recensoring algorithm described in Section 6.3

Method 4: UML, the usual MLE assuming there is no pair effect; that is, treating the estimation procedure as a two-sample problem, with no recensoring applied.

UML is our benchmark for this setting, because it assumes the knowledge of no block effects. The other three methods do not assume no block effects, which is more appropriate in analyzing the pair data. If block effects exist, UML will not be consistent but the other three methods will be. Also, note that although for a finite sample the MAL method proposed by Holt and Prentice (1974) and the RML method are different, they are asymptotically equivalent under (3.4).

7.1 No Censoring

The ARE in the no censored case is easy to calculate. The asymptotic variances for the four methods are $\sigma_1^2 = 4$, $\sigma_2^2 = \pi^2/3$, $\sigma_3^2 = 3$, and $\sigma_4^2 = 2$, and the ARE (i, j), where i denotes the row method and j denotes the column method, are listed in Table 1.

7.2 With Censoring

Let $c = \ln K$. The asymptotic variances for each of the four methods are

$$\sigma_1^2 = 1/B_1^2,$$

$$B_1 = \frac{1}{2}(1 - e^{-2c}),$$

$$\sigma_2^2 = \frac{A_2}{B_2^2},$$

$$A_2 = 2\{\alpha_2 - \alpha_1^2 + e^{-c}[(1 - e^{-c})(\ln c)^2 - 2\alpha_1 \ln c]\},$$

and

$$B_2 = 1 - e^{-c},$$

where

$$\alpha_1 = \int_{-\infty}^{\ln(c)} x e^{x - e^x} dx \quad \text{and} \quad \alpha_2 = \int_{-\infty}^{\ln(c)} x^2 e^{x - e^x} dx,$$

$$\sigma_3^2 = \frac{A_3}{B_3^2},$$

$$A_3 = \frac{1}{3}\left[1 + 2(4c^2 + 5c - 1)e^{-c} - (4c^2 + 7c - 1)e^{-2c} - 2c^2(4c + 9)\int_c^{2c} \frac{e^{-t}}{t} dt\right],$$

$$B_3 = -\frac{1}{3}\left[1 - 2(2c^2 + c + 1)e^{-c} + (2c^2 + 2c + 1)e^{-2c} + c(4c^2 + 6c + 3)\int_c^{2c} \frac{e^{-t}}{t} dt\right],$$

Table 1. Asymptotic Relative Efficiencies in the No Censoring Case With No Block Effect

	RMD	RMN	RML	UML
RMD	1.00	.82	.75	.50
RMN	1.22	1.00	.91	.61
RML	1.33	1.10	1.00	.67
UML	2.00	1.64	1.50	1.00

and

$$\sigma_4^2 = \frac{2}{1 - e^{-c}}.$$

The asymptotic variances are plotted in Figure 2 against the censoring rate for each coordinate. As expected, the UML has the smallest asymptotic variance, followed by the RML. The RMN is preferable to the RMD, except for the censoring rate in between .19 to .57, where the variance of the RMD is slightly smaller than that of the RMN.

We have demonstrated how the asymptotic variance can be computed for the recensoring methods when there is no block effect. When there is block effect, as should be the case for pair data, the formulas given in Appendix A can be applied to obtain the asymptotic variance.

8. MONTE CARLO STUDIES

To evaluate the finite sample properties of the recensoring methods and to compare with existing methods, we design the following Monte Carlo studies. We obtain the point estimates and the standard deviations of the estimates. In addition we construct confidence intervals with a 95% nominal coverage probability. All simulations are performed with
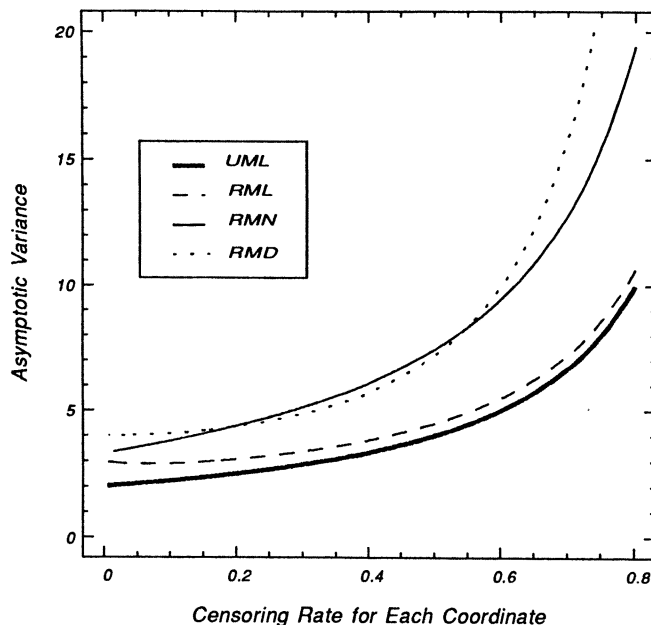

Figure 2. Asymptotic Variance of the Usual Maximum Likelihood Estimator (UML), the Recensoring Maximum Likelihood Estimator (RML), the Recensoring Mean Estimator (RMN), and the Recensoring Median Estimator (RMD) Under Censoring.

1,000 replicates. Thus for estimating the true coverage probability $1 - \alpha$, the standard error of the reported coverage probability is about 3% of $\sqrt{\alpha(1 - \alpha)}$. At the true coverage rate 95%, any interval that gives the coverage rate from 93.6% to 96.4% is statistically no different from 95%. All trials without any doubly observed pairs are considered as invalid trials. They are skipped and replaced with valid trials. The number of pairs in each group is chosen as $N = 60$ and $N = 300$. The effective sample size depends not only on N but also on the censoring rate. For example, $N = 300$ can be considered a "large" sample when there is no censoring. But with 80% censoring rate occurring in each group, we can observe only 4% (i.e., 12 pairs) doubly observed pairs and 64% (i.e., 192 pairs) doubly censored pairs. In this situation $N = 300$ can only be considered as a small or moderate sample size. Five methods are evaluated in the simulation: one conventional marginal likelihood method without recensoring, the MAL (Holt and Prentice 1974); one parametric and one nonparametric recensoring method, the RML and RMN; one nonparametric test-based method, the TWP (Wei and Pee 1985), and one method assuming the proportional hazard model, the PHM (Holt and Prentice 1974). The methods are implemented using five parameter sets. All parameter sets assume $\phi = .5$; that is, the treatment difference between two groups, θ, is equal to $-.693$. The RMD estimate is not included because of its nonuniqueness when heavy censoring occurs.

Parameter Set 1. An exponential survival time with the unit scale is assumed for the control group in the multiplicative model (2.3). We also assume to apply a common censoring time $(C_1, C_2) = (e^{K_1}, e^{K_2})$ to all pairs. C_1 and C_2 are chosen to yield a 0%, 20%, 40%, 60%, and 80% censoring rate for each component of a pair. The values of C_2 are ∞, 1.609, .916, .511, and .233. The value of C_1 is equal to ϕC_2. In parameter set 1 no pair effect is assumed, but the analysis was carried out without this knowledge.

Parameter Set 2. We use $C = C_1 = C_2 = \infty$, 1.609, .916, .511, and .223 for each case. In this setup, C_1 does not equal ϕC_2. Other parameter settings are the same as in set 1.

Parameter Set 3. This set is the same as in the simulation of Holt and Prentice (1974). The fixed common censoring times $C = C_1 = C_2 = \infty$, 25, and 5 are used. The block effects B_i's in (2.3) are set by $B_i = a + b*\text{Mod}(i, 60)$, with $a = 5$ and $b = .5$. ($\text{Mod}(i, 60)$ is the remainder of i divided by 60.) When $C = 25$, the probability for cases (a), (b), (c), and (d) is (.39, .11, .34, .16). This changes to (.05, .10, .21, .64) for $C = 5$.

Parameter Set 4. The survival distributions are assumed to be Weibull with shape 2 for both the control and the treatment groups. The scale parameter for the control group is set at 1, and a block effect similar to parameter set 3 is assumed with $a = 0$ and $b = .5$. The fixed common censoring times $C = C_1 = C_2 = \infty$, 7.7, and 3.6 are used to yield 0%, 30%, and 60% of doubly censored pairs.

Parameter Set 5. Lognormal distributions are assumed for the control and the treatment groups with both mean and variance as 1 in the control group. Again, similar block effect as in parameter set 3 is assumed with $a = 0$ and $b = .05$. The fixed common censoring times $C = C_1 = C_2 = \infty$, 9.2, and 3.4 are used to yield 0%, 30%, and 60% of doubly censored pairs.

The results from the simulation study are presented in Tables 2–6. Cases are sorted according to the probability of censoring (PC) for Table 2 and common censoring time (C) for Tables 3–6. The first line of each case gives the point estimator for the three estimators considered, with standard deviations from 1,000 trials in parentheses. The next line gives the coverage probability of a 95% confidence interval. For the MAL, RML, RMN, and PHM, the confidence intervals are constructed symmetric about the point estimates

Table 2. Simulation Results for Parameter Set 1

PC	MAL	RML	RMN	TWP	PHM
N = 60					
0	−.70 (.22)	−.70 (.22)	−.71 (.23)	−.70 (.21)	−.71 (.27)
	.95 (.84)	.95 (.76)	.95 (.91)	.97 (.91)	.97 (1.09)
.2	−.70 (.28)	−.69 (.23)	−.68 (.30)	−.70 (.23)	−.71 (.29)
	.95 (1.04)	.96 (.83)	.95 (1.12)	.97 (.98)	.97 (1.14)
.4	−.71 (.35)	−.70 (.26)	−.68 (.36)	−.70 (.27)	−.73 (.32)
	.96 (1.42)	.96 (.92)	.95 (1.37)	.97 (1.14)	.97 (1.27)
.6	−.72 (.47)	−.72 (.33)	−.67 (.48)	−.71 (.37)	−.74 (.39)
	.96 (1.88)	.95 (1.27)	.96 (1.80)	.97 (1.50)	.96 (1.52)
.8	−.74 (.73)	−.73 (.52)	−.65 (.74)	−.73 (.50)	−.76 (.54)
	.95 (2.81)	.93 (1.83)	.95 (2.73)	.97 (2.54)	.97 (2.15)
N = 300					
0	−.70 (.10)	−.70 (.10)	−.70 (.11)	−.70 (.09)	−.70 (.12)
	.96 (.40)	.96 (.39)	.95 (.41)	.97 (.38)	.96 (.48)
.2	−.70 (.12)	−.70 (.10)	−.70 (.12)	−.70 (.09)	−.70 (.13)
	.96 (.50)	.96 (.40)	.96 (.48)	.97 (.40)	.95 (.51)
.4	−.70 (.15)	−.71 (.12)	−.70 (.15)	−.70 (.11)	−.70 (.14)
	.96 (.64)	.95 (.44)	.96 (.58)	.96 (.45)	.97 (.56)
.6	−.71 (.20)	−.70 (.14)	−.70 (.18)	−.70 (.14)	−.70 (.16)
	.96 (.83)	.95 (.54)	.97 (.74)	.97 (.56)	.96 (.66)
.8	−.71 (.30)	−.70 (.21)	−.69 (.28)	−.70 (.20)	−.71 (.23)
	.96 (1.33)	.94 (.80)	.97 (1.10)	.96 (.82)	.96 (.91)

NOTE: This parameter set assumes exponential survival without block effect, censoring times $C_1 = \phi C_2$, $\theta_0 = -.693$, and $1 - 2\alpha = .95$.
PC refers to the probability of censoring for each member of the pair.

Table 3. Simulation Results for Parameter Set 2

C	MAL	RML	RMN	TWP	PHM
N = 60					
∞	−.70 (.22)	−.70 (.22)	−.71 (.23)	−.70 (.21)	−.71 (.27)
	.95 (.84)	.95 (.76)	.95 (.91)	.97 (.91)	.97 (1.09)
1.609	−.81 (.25)	−.68 (.22)	−.72 (.27)	−.70 (.22)	−.71 (.27)
	.94 (.95)	.96 (.96)	.95 (1.06)	.97 (.95)	.97 (1.09)
.916	−.92 (.31)	−.70 (.25)	−.73 (.32)	−.71 (.25)	−.71 (.28)
	.90 (1.17)	.95 (.95)	.95 (1.25)	.97 (1.07)	.97 (1.12)
.511	−1.04 (.40)	−.74 (.31)	−.75 (.42)	−.72 (.32)	−.72 (.31)
	.87 (1.49)	.93 (1.08)	.95 (1.59)	.97 (1.32)	.96 (1.24)
.223	−1.18 (.61)	−.79 (.51)	−.78 (.59)	−.76 (.57)	−.74 (.41)
	.90 (2.38)	.90 (2.07)	.95 (2.30)	.97 (2.12)	.96 (1.60)
N = 300					
∞	−.70 (.10)	−.70 (.10)	−.70 (.10)	−.70 (.09)	−.70 (.12)
	.96 (.40)	.96 (.39)	.96 (.41)	.97 (.38)	.96 (.48)
1.609	−.80 (.11)	−.70 (.10)	−.70 (.12)	−.70 (.09)	−.70 (.12)
	.87 (.46)	.95 (.41)	.96 (.47)	.97 (.39)	.95 (.48)
.916	−.92 (.14)	−.71 (.11)	−.71 (.14)	−.70 (.10)	−.70 (.12)
	.65 (.56)	.95 (.44)	.96 (.56)	.96 (.44)	.95 (.50)
.511	−1.03 (.17)	−.72 (.14)	−.71 (.17)	−.70 (.13)	−.70 (.13)
	.55 (.73)	.94 (.50)	.96 (.70)	.97 (.53)	.96 (.54)
.223	−1.13 (.25)	−.74 (.20)	−.72 (.25)	−.71 (.18)	−.71 (.17)
	.63 (1.07)	.95 (.77)	.95 (1.00)	.96 (.74)	.96 (.69)

NOTE: This parameter set assumes exponential survival without block effect, censoring time $C_1 = C_2 = C$, $\theta_0 = -.693$, and $1 - 2\alpha = .95$.

under the asymptotic normal assumption. For the TWP, however, we assume the asymptotic normality of the test statistics and invert the test. The confidence interval obtained for the TWP may not be symmetric about the point estimates. The average length of the confidence interval is given in parentheses following the coverage rate.

As can be seen in Table 2, all five methods work well for point estimation, except for $N = 60$ with PC = .8 for parameter set 1. The MAL works because the censoring times are chosen luckily so that (3.4) holds even without recensoring. The standard deviation of the RML and TWP are smaller than that of the MAL, RMN, and PHM. The RMN is slightly more efficient than the PHM when there is no censoring, is about as efficient as the PHM on slight to moderate censoring, and is less efficient than the PHM on heavy censoring. All five methods also give the desired 95% coverage rate. The RML has the shortest length of the confidence interval in both censored and no censored cases. On the other hand, the PHM gives longest interval when the censoring proba-

bility is less than or equal to 20% for each member of the pair. The MAL yields the longest interval when the censoring rate is moderate to heavy. For all censoring rates, the lengths of the TWP confidence intervals are about the same as those of the RML when $N = 300$ but are consistently wider than those of the RML when $N = 60$. The TWP gives tighter intervals compared to the RMN, except when the sample size is small ($N = 60$) with little censoring.

Parameter set 2 is identical to set 1 except that the censoring condition (3.4) does not hold. Table 3 shows that the MAL fails except in the no censoring case. All other four methods still work in general. The point estimators are slightly off for $N = 60$ with $C = .511$ and .223. The coverage rate for the RML is lower for $N = 60$ and $C = .223$.

The general patterns of behaviors for the RML, RMN, TWP, and PHM are similar to what we have seen in parameter set 1, except that when the censoring rate is extremely high ($C = .223$, 80% censoring in the control group), the PHM outperforms all other methods. This is due to the fact

Table 4. Simulation Results for Parameter Set 3

C	MAL	RML	RMN	TWP	PHM
N = 60					
∞	−.70 (.22)	−.70 (.22)	−.69 (.24)	−.70 (.21)	−.71 (.27)
	.95 (.84)	.95 (.75)	.94 (.92)	.97 (.93)	.97 (1.09)
25	−.95 (.34)	−.71 (.28)	−.70 (.36)	−.71 (.27)	−.72 (.30)
	.90 (1.25)	.93 (1.00)	.93 (1.36)	.97 (1.18)	.96 (1.19)
5	−1.20 (.73)	−.78 (.47)	−.61 (.65)	−.77 (.52)	−.75 (.50)
	.90 (2.62)	.93 (1.69)	.94 (2.66)	.97 (2.70)	.98 (1.90)
N = 300					
∞	−.70 (.10)	−.70 (.10)	−.70 (.11)	−.70 (.09)	−.70 (.12)
	.96 (.40)	.96 (.38)	.95 (.41)	.96 (.39)	.96 (.48)
25	−.94 (.15)	−.71 (.12)	−.70 (.15)	−.71 (.11)	−.70 (.13)
	.64 (.62)	.94 (.48)	.96 (.61)	.96 (.48)	.96 (.53)
5	−1.15 (.31)	−.74 (.23)	−.71 (.30)	−.72 (.22)	−.71 (.20)
	.71 (1.29)	.94 (.82)	.94 (1.19)	.95 (.89)	.95 (.81)

NOTE: This parameter set assumes exponential survival with block effect, censoring time $C_1 = C_2 = C$, $\theta_0 = -.693$, and $1 - 2\alpha = .95$.

Table 5. Simulation Results for Parameter Set 4

C	RML	RMN	TWP	PHM
N = 60				
∞	−.70 (.11)	−.70 (.11)	−.70 (.11)	−.72 (.17)
	.96 (.38)	.95 (.45)	.98 (.51)	.97 (.65)
7.7	−.72 (.17)	−.71 (.19)	−.71 (.18)	−.73 (.20)
	.95 (.54)	.95 (.75)	.97 (.80)	.97 (.79)
3.6	−.76 (.25)	−.72 (.28)	−.72 (.30)	−.75 (.28)
	.94 (.81)	.93 (1.04)	.97 (1.24)	.98 (1.09)
N = 300				
∞	−.70 (.05)	−.70 (.05)	−.70 (.05)	−.70 (.07)
	.96 (.20)	.96 (.20)	.95 (.21)	.95 (.28)
7.7	−.70 (.07)	−.70 (.09)	−.70 (.08)	−.70 (.09)
	.96 (.32)	.95 (.34)	.96 (.32)	.96 (.34)
3.6	−.71 (.11)	−.70 (.12)	−.70 (.12)	−.71 (.11)
	.95 (.43)	.94 (.48)	.94 (.46)	.96 (.46)

NOTE: This parameter set assumes Weibull survival with block effect and censoring time $C_1 = C_2 = C$.

that equal censoring is the optimal design for PHM, whereas $C_1 = \phi C_2$ is optimal for the recensoring methods.

For parameter set 3 (Table 4), the RML corrects the inconsistency problem of the MAL. In addition, the RML is also more efficient than the RMN, TWP, and PHM for all censoring at $N = 60$. When $N = 300$, the performance of the TWP is close to that of the RML for all censoring, and the performance of the PHM is close to that of the RML and TWP on high censoring ($C = 5$).

We drop the MAL on parameter sets 4 and 5 because we know that it will not work well in the case of $\phi = .5$ and $C_1 = C_2$. For parameter set 4, similar results to those for parameter set 3 can also be found in Table 5. When $N = 60$ and no censoring is done, the length of the confidence interval is the smallest for the RML, followed by the RMN, the TWP, and then the PHM. The RMN yields a shorter interval than the TWP when $N = 60$; but when $N = 300$, both methods give comparable results.

Last, we evaluate these four methods on the lognormal distribution. Table 6 shows that the PHM failed because the lognormal distribution is not a proportional hazard model on which the method is constructed. Among the three remaining methods, all give the correct point estimates and the desired coverage probability, except for a somewhat lower coverage rate for the RMN when $N = 60$, $C = 3.4$. When $N = 60$, the length of the confidence interval for the TWP is wider than that for the RMN and the RML; when $N = 300$, the TWP gives a similar length as the RMN. The RML still has the best performance in general, however.

9. APPLICATION TO THE DATA ANALYSIS FROM TUMORIGENESIS EXPERIMENT

Mantel, Bohidar, and Ciminera (1977) considered a litter-matched tumorigenesis experiment. For each block of size three, rats were randomly assigned into one drug-treated group and two control groups. The weeks to tumor appearance were recorded to evaluate whether the drug had any effect on tumorigenesis. All rats were sacrificed at the end of 104 weeks. Table 7 gives the time to tumor appearance for the drug-treated group and control group 1 in the female rats, which had also been analyzed by Wei and Pee (1985).

In this particular example the rats either developed tumors, died of other reasons (treated as censored), or were still alive at 104 weeks without tumors (also treated as censored). Because all rats were to be sacrificed at the end of 104 weeks, we use this number as the potential censoring time for those uncensored cases. Note that the censoring rate is very high in the data set. There are 24/50 (48%) doubly censored pairs, 18/50 (36%) when the drug-treated group was observed but the control group is censored, 4/50 (8%) when the control group is observed but the drug-treated group is censored, and only 4/50 (8%) doubly observed.

After taking the logarithm transformation on the data and applying the estimation methods, the results of the point estimator and its standard error (in parentheses) are RMN: −.159 (.142); RML: −.275 (.268); and MAL: −1.082 (.405). The corresponding 95% confidence intervals are RMN: (−.437, .119); RML: (−.800, .250); and MAL: (−1.876, −.288). Wei and Pee (1985) had $\hat{\theta} = -.195$ and two versions of the 95% confidence interval as (−.844, −.010) and (−.916, −.020). Figures 3 and 4 give $\hat{\theta}(b)$ as a function of b and shows the solution $\hat{\theta}$ for the RMN and RML. The RMN, RML, and Wei and Pee methods indicate that the drug may reduce the time to tumor by 15–24% (percent survival time reduced = $(1 - \phi) \times 100\% = 1 - \exp(\theta)$). This is rather different from the 66% derived from the MAL. Wei and Pee's confidence intervals are highly asymmetric and barely exclude 0. The box plot of the residual $t_1^{(b)} - t_2^{(b)} - b$ is given in Figure 5 for b from −1 to .4. It is clear that the residual distribution in either extreme case, $b = -1$ or .4, is not symmetric about 0. Yet it is more symmetric when b is between −.2 and −.1. Discussion of this data set is continued in Section 10.

10. TESTING HYPOTHESIS AND INVERTING THE TEST UNDER RECENSORING

The application of recensoring is not restricted to the estimation problem. In this section we discuss the hypothesis testing problem. By inverting the test, one can construct confidence intervals as usual.

Consider the two-sided test of H_0: $\theta = \theta_0$ against H_1: $\theta \neq \theta_0$. Recensor the data as in Section 4 with $b = \theta_0$. Then, under the null hypothesis, $D_i^{(\theta_0)}$ is symmetric about θ_0.

Table 6. Simulation Results for Parameter Set 5

C	RML	RMN	TWP	PHM
N = 60				
∞	−.71 (.19)	−.71 (.19)	−.71 (.20)	−.82 (.28)
	.94 (.79)	.94 (.71)	.96 (.82)	.96 (1.11)
9.2	−.70 (.23)	−.71 (.25)	−.71 (.24)	−.91 (.36)
	.93 (.80)	.92 (.93)	.96 (1.02)	.94 (1.37)
3.4	−.69 (.33)	−.72 (.34)	−.72 (.34)	−1.07 (.52)
	.95 (1.10)	.90 (1.27)	.97 (1.44)	.94 (1.91)
N = 300				
∞	−.69 (.08)	−.69 (.08)	−.70 (.08)	−.79 (.13)
	.96 (.29)	.97 (.32)	.96 (.34)	.89 (.49)
9.2	−.69 (.10)	−.70 (.10)	−.70 (.10)	−.89 (.15)
	.94 (.36)	.96 (.43)	.96 (.42)	.77 (.60)
3.4	−.69 (.14)	−.69 (.14)	−.70 (.13)	−1.01 (.21)
	.98 (.53)	.95 (.58)	.96 (.55)	.69 (.81)

NOTE: This parameter set assumes lognormal survival with block effect and censoring time $C_1 = C_2 = C$.

Table 7. Time to Tumorigenesis in Weeks for 50 Female Rats (Drug-Treated Group, Control Group)

(101+, 49),	(88, 96),	(89, 91+),	(85+, 72+),	(104+, 104+)
(89, 104+),	(76+, 87+),	(94, 104+),	(72, 95+),	(49+, 83+)
(104+, 102+),	(104, 94+),	(91+, 70+),	(104+, 63+),	(104+, 104+)
(78+, 104+),	(103, 73),	(104+, 104+),	(73, 104+),	(89, 104+)
(104+, 104+),	(96, 104+),	(39, 45+),	(104+, 104+),	(104+, 83+)
(104+, 81),	(102, 104+),	(104+, 101),	(92, 104+),	(88+, 79+)
(77, 97+),	(82+, 77+),	(103, 69+),	(81+, 104+),	(87+, 104+)
(86, 55),	(80, 104+),	(76+, 84),	(104+, 98+),	(103, 91+)
(89+, 104+),	(70, 104+),	(93+, 104+),	(67, 104+),	(104+, 104+)
(34, 104+),	(45, 79+),	(80, 81),	(55+, 104+),	(104+, 104+)

Therefore, we can apply tests for symmetry, such as the sign test or signed-rank test. The ties, $D_i^{(\theta_0)} = \theta_0$, occurring for doubly censored pairs will be ignored.

Let SIGN = number of cases for which $D_i^{(\theta_0)} > \theta_0$ and let N_0 = number of cases for which $D_i^{(\theta_0)} \neq \theta_0$. Then, conditional on N_0, SIGN is a binomial random variable with $n = N_0$ and $p = .5$ under H_0. The two-sided sign test can be found in the usual way by rejecting H_0 if SIGN is either too large or too small. The signed-rank test can be formed in the same way.

As usual, inverting a test leads to a procedure of constructing confidence intervals. The resulting 95% confidence intervals for the tumorigenesis experiment data of Section 8 based on signed test and signed rank test are $(-.56, -.02)$ and $(-.87, .08)$. It is seen that the sign test-based interval barely excludes 0 (similar to Wei and Pee's result), while the signed-rank test interval includes 0. Because this is a highly censored data set without strong treatment effect, we have grave reservations in concluding the existence of treatment effect.

11. CONCLUSION

Paired-data analysis is one of the basic topics covered in almost all elementary statistics textbooks. The method taught there is unanimous; namely, converting the paired-data problem into a one-sample problem by taking the difference of the two observations in each pair so that heterogeneity due to distinct characteristics for different pairs can be eliminated. Here we assume that the definition of treatment effect is given in terms of location parameters, possibly after a suitable scale transformation. When paired data are subject to censoring times, however, this approach has apparently been abandoned for the obvious reason that the observed differences within pairs no longer reflect the treatment effect unbiasedly. In the literature, alternative methods have been sought under rather different rationales.

We have a different attitude toward the traditional method. We enjoy its simplicity and believe in its rationale. To save these merits, we introduce the idea of recensoring for absorbing the destructive impact from censoring. Recensoring is motivated by the discovery that the difference in observed times within each pair can still be unbiased in estimating the treatment effect if we were to have a favorable situation in which the difference in the two censoring times within each pair is equal to the treatment effect. Such a favorable situation rarely exists automatically, but we can try to create it by first forcing the censoring times within each pair to take a common difference. This can be done easily by resetting one of
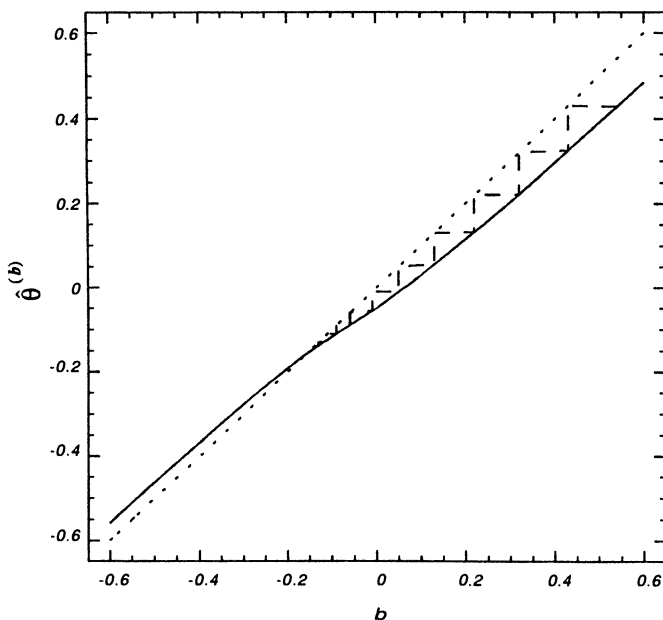

Figure 3. The Convergence of the Recensoring Mean Method (Initial Value = .54, $\hat{\theta}$ = −.159).
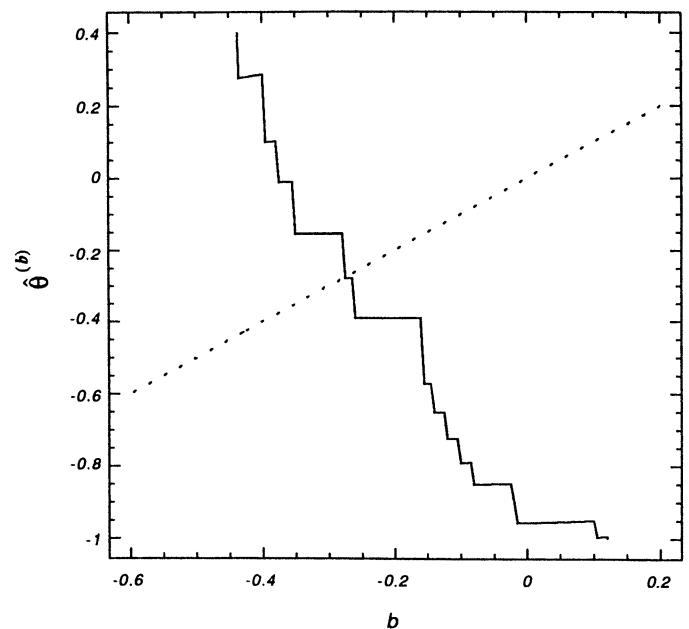

Figure 4. The Solution of the Recensoring Maximum Likelihood Method ($\hat{\theta}$ = −.275).
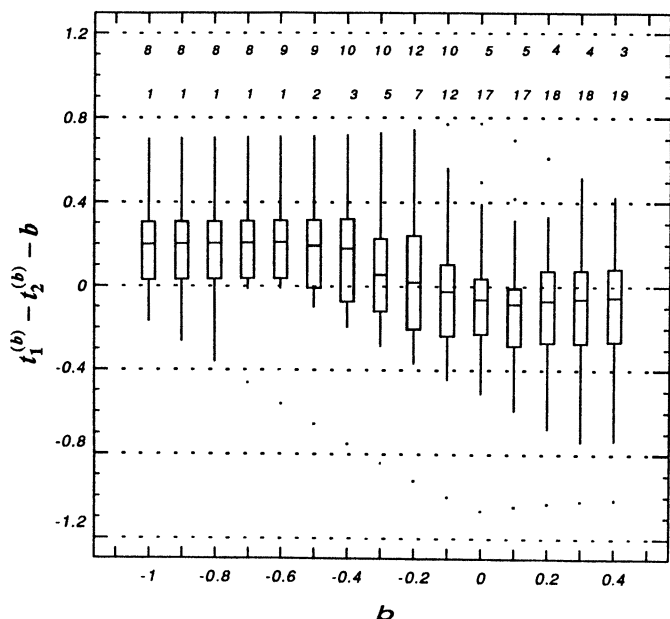
Figure 5. Box Plot of $t_1^{(b)} - t_2^{(b)} - b$ Versus b for Nondoubly Censored Pairs. The number of positive and negative $t_1^{(b)} - t_2^{(b)} - b$ pairs are given above the corresponding box plot. The top figure is the number of positive pairs, and the bottom figure is the number of negative pairs.

the two censoring times in each pair to a smaller value. We use these new censoring times to recensor the existing data. After recensoring, we can resort to the traditional method of paired-data analysis. The remaining question involves the choice of the common difference for the new censoring times. We resolve this question by using the one that yields itself as the estimate of the treatment effect after recensoring.

As we have rigorously shown, recensoring has successfully removed the ill effect due to censoring for many estimation procedures, including M estimation in the location model and the likelihood-based estimate of Holt and Prentice (1974). Asymptotic results for the derived estimates have been obtained. In addition, our approach preserves another important feature of the traditional method: graphics. As is well known, one standard assumption on the one-sample problem converted from the differences within pairs is that these differences follow a symmetric distribution (with the symmetric center being the treatment effect). Many plots are available for visually assessing symmetry in the data. These graphical procedures are equally vulnerable to the ill effect of censoring. But with recensoring, all of them can still be used for checking symmetry.

To evoke recensoring, the censoring times must be given for all experimental units. Although censoring times may not be observed directly for the uncensored cases, "potential censoring times" are usually available. Almost all studies have a definite study period. For example, laboratory experiments generally have a predefined observation period, and clinical trials are typically set up for certain years of accrual plus follow-up. Such information can be used to determine the censoring time for each uncensored case.

One referee raised the issue of the sensitivity of the various estimates to the censoring times in the event that they cannot be accurately determined. We have one result that is easy to

verify. For the sample median method, each estimate $\hat{\theta}^{(b)}$ (5.1) does not depend on the true censoring times for the noncensored cases. Thus the derived estimate is completely insensitive to the undetermined censoring times. But, before jumping to any conclusion, we must realize that our sensitivity issue has two sides. Although we want our procedure to be as robust as possible under the false censoring times, we also want it to respond well to the correct censoring times. For example, both Holt and Prentice's intrapair rank method and Wei and Pee's method are insensitive to the undetermined censoring times—in fact, they do not need it at all. However, the former assumes a common censoring time for the two experimental units in each pair, and the latter depends on an iid assumption about the censoring times. These methods are not designed to incorporate the empirical information on the censoring times for the uncensored cases. The induced insensitivity is not necessarily a virtue if their assumptions meet challenges from the well-informed censoring times. Apparently, an in-depth study on this important issue is warranted.

APPENDIX A: THE DERIVATION OF ASYMPTOTIC VARIANCE OF THE RECENSORING ESTIMATORS

As described in Section 5.3, the asymptotic variance of $\hat{\theta}$ can be expressed as A_n / B_n^2.

We now present a method of deriving

$$B_n = \sum_{i=1}^{n} \left\{ \frac{d}{db} E_\theta \Psi(D_i^{(b)} - b, \delta_{1i}^{(b)}, \delta_{2i}^{(b)}) \right\}_{b=\theta} = \sum_{i=1}^{n} B_{ni}, \quad (A.1)$$

where

$$\Psi(D_i^{(b)} - b, \delta_{1i}^{(b)}, \delta_{2i}^{(b)}) = \Psi_1(D_i^{(b)} - b)\delta_{1i}^{(b)}\delta_{2i}^{(b)}$$
$$+ \Psi_2(D_i^{(b)} - b)(1 - \delta_{1i}^{(b)})\delta_{2i}^{(b)}$$
$$+ \Psi_3(D_i^{(b)} - b)(1 - \delta_{2i}^{(b)})\delta_{1i}^{(b)} \quad (A.2)$$

for some functions Ψ_1, Ψ_2, and Ψ_3. For the RMN and RMD methods, $\Psi_1 = \Psi_2 = \Psi_3$. For the RML, Ψ_1, Ψ_2, and Ψ_3 are the derivatives of the log density, survival, and cumulative density of D_i^o.

We shall compute

$$\frac{d}{db} E_\theta \Psi(D_i^{(b)} - b)$$

$$= \lim_{\Delta b \to 0} \frac{E_\theta \Psi(D_i^{(b+\Delta b)} - (b + \Delta b)) - E_\theta \Psi(D_i^{(b)} - b)}{\Delta b}.$$

Let $F_i(t_1, t_2)$, $F_{1i}(\cdot)$, and $F_{2i}(\cdot)$ denote the distributions of (T_{1i}^o, T_{2i}^o), T_{1i}^o, and T_{2i}^o. We consider the case $K_{1i} > K_{2i} + b$ only (the other case can be derived similarly), so that $K_{1i}^{(b)} = K_{2i} + b$ and $K_{2i}^{(b)} = K_{2i}$. Clearly, by definition we have

$$E\Psi(D_i^{(b+\Delta b)} - (b + \Delta b)) - E\Psi(D_i^{(b)} - b)$$

$$= \int_{-\infty}^{K_{2i}} \int_{-\infty}^{K_{2i}+(b+\Delta b)} \Psi_1(t_1 - t_2 - (b + \Delta b)) \, dF_i(t_1, t_2)$$

$$- \int_{-\infty}^{K_{2i}} \int_{-\infty}^{K_{2i}+b} \Psi_1(t_1 - t_2 - b) \, dF_i(t_1, t_2)$$

$$+ \int_{-\infty}^{K_{2i}} \int_{K_{2i}+(b+\Delta b)}^{\infty} \Psi_2(K_{2i} - t_2) \, dF_i(t_1, t_2)$$

$$- \int_{-\infty}^{K_{2i}} \int_{K_{2i}+b}^{\infty} \Psi_2(K_{2i} - t_2) \, dF_i(t_1, t_2)$$

$$+ \int_{K_{2i}}^{\infty} \int_{-\infty}^{K_{2i}+(b+\Delta b)} \Psi_3(t_1 - K_{2i} - (b + \Delta b)) \, dF_i(t_1, t_2)$$

$$- \int_{K_{2i}}^{\infty} \int_{-\infty}^{K_{2i}+b} \Psi_3(t_1 - K_{2i} - b) \, dF_i(t_1, t_2). \quad (A.3)$$

Assuming the differentiability of Ψ_i, we can approximate this expression by

$$\int_{-\infty}^{K_{2i}} \int_{-\infty}^{K_{2i}+b} (-\Delta b)\Psi_1'(t_1 - t_2 - b) \, dF_i(t_1, t_2)$$

$$+ \int_{-\infty}^{K_{2i}} \int_{K_{2i}+b}^{K_{2i}+b+\Delta b} \Psi_1(t_1 - t_2 - (b + \Delta b)) \, dF_i(t_1, t_2)$$

$$- \int_{-\infty}^{K_{2i}} \int_{K_{2i}+b}^{K_{2i}+b+\Delta b} \Psi_2(K_{2i} - t_2) \, dF_i(t_1, t_2)$$

$$+ \int_{K_{2i}}^{\infty} \int_{-\infty}^{K_{2i}+b} (-\Delta b)\Psi_3'(t_1 - K_{2i} - b) \, dF_i(t_1, t_2)$$

$$+ \int_{K_{2i}}^{\infty} \int_{K_{2i}+b}^{K_{2i}+b+\Delta b} \Psi_3(t_1 - K_{2i} - (b + \Delta b)) \, dF_i(t_1, t_2)$$

$$\cong -\Delta b \left\{ \int_{-\infty}^{K_{2i}} \int_{-\infty}^{K_{2i}+b} \Psi_1'(t_1 - t_2 - b) \, dF_i(t_1, t_2) \right.$$

$$+ f_{1i}(K_{2i} + b) \int_{-\infty}^{K_{2i}} [\Psi_2(K_{2i} - t_2) - \Psi_1(K_{2i} - t_2)] \, dF_{2i}(t_2)$$

$$+ (1 - F_{2i}(K_{2i})) \left[\int_{-\infty}^{K_{2i}+b} \Psi_3'(t_1 - K_{2i} - b) \, dF_{1i}(t_1) \right.$$

$$\left. \left. - \Psi_3(0) f_{1i}(K_{2i} + b) \right] \right\}.$$

Therefore, when $\theta < K_{1i} - K_{2i}$, we have

$$-B_{ni} = \int_{-\infty}^{K_{2i}} \int_{-\infty}^{K_{2i}+\theta} \Psi_1'(t_1 - t_2 - \theta) \, dF_i(t_1, t_2) + f_{1i}(K_{2i} + \theta)$$

$$\times \int_{-\infty}^{K_{2i}} [\Psi_2(K_{2i} - t_2) - \Psi_1(K_{2i} - t_2)] \, dF_{2i}(t_2)$$

$$+ (1 - F_{2i}(K_{2i})) \left[\int_{-\infty}^{K_{2i}+\theta} \Psi_3'(t_1 - K_{2i} - \theta) \, dF_{1i}(t_1) \right.$$

$$\left. - \Psi_3(0) f_{1i}(K_{2i} + \theta) \right]. \tag{A.4}$$

Similarly, when $\theta > K_{1i} - K_{2i}$, we have

$$-B_{ni} = \int_{-\infty}^{K_{1i}-\theta} \int_{-\infty}^{K_{1i}} \Psi_1'(t_1 - t_2 - \theta) \, dF_i(t_1, t_2) + f_{2i}(K_{1i} - \theta)$$

$$\times \int_{-\infty}^{K_{1i}} [\Psi_1(t_1 - K_{1i}) - \Psi_3(t_1 - K_{1i})] \, dF_{1i}(t_1)$$

$$+ (1 - F_{1i}(K_{1i})) \left\{ \int_{-\infty}^{K_{1i}-\theta} \Psi_2'(K_{1i} - t_2 - \theta) \, dF_{2i}(t_2) \right.$$

$$\left. + \Psi_2(0) f_{2i}(K_{1i} - \theta) \right\}. \tag{A.5}$$

Note that in the special case of $\Psi_1 = \Psi_2 = \Psi_3$ and $\Psi_i(0) = 0$, this is reduced to

$$-B_{ni} = \left[\int_{-\infty}^{K_{2i}} \int_{-\infty}^{K_{2i}+\theta} \Psi'(t_1 - t_2 - \theta) \, dF_i(t_1, t_2) \right.$$

$$\left. + (1 - F_{2i}(K_{2i})) \int_{-\infty}^{K_{2i}+\theta} \Psi'(t_1 - K_{2i} - \theta) \, dF_{1i}(t_1) \right],$$

$$\text{for} \quad \theta < K_{1i} - K_{2i}$$

and

$$-B_{ni} = \int_{-\infty}^{K_{1i}-\theta} \int_{-\infty}^{K_{1i}} \Psi'(t_1 - t_2 - \theta) \, dF_i(t_1, t_2)$$

$$+ (1 - F_{1i}(K_{1i})) \int_{-\infty}^{K_{1i}-\theta} \Psi'(K_{1i} - t_2 - \theta) \, dF_{2i}(t_2),$$

$$\text{for} \quad \theta > K_{1i} - K_{2i}.$$

Therefore, we obtain

$$-B_{ni} = E\Psi'(D_i^{(\theta)} - \theta)\delta_{1i}^{(\theta)} + E\Psi'(D_i^{(\theta)} - \theta)\delta_{2i}^{(\theta)}. \tag{A.6}$$

Note that although the independence assumption between T_{1i}^o and T_{2i}^o is used in this derivation, (A.6) holds without it.

Example 1: RMN. The RMN method takes the form of the identity kernel and can be considered as a special that has $\Psi_1 = \Psi_2 = \Psi_3$ and $\Psi(0) = 0$. By (A.6), when $\theta < K_{1i} - K_{2i}$,

$$-B_{ni} = \int_{-\infty}^{K_{2i}} \int_{-\infty}^{K_{2i}+\theta} dF_i(t_1, t_2) + (1 - F_{2i}(K_{2i})) \int_{-\infty}^{K_{2i}+\theta} dF_{1i}(t_1)$$

$$= F_{2i}(K_{2i})F_{1i}(K_{2i} + \theta) + (1 - F_{2i}(K_{2i}))F_{1i}(K_{2i} + \theta)$$

$$= F_{1i}(K_{2i} + \theta)$$

$$= \Pr(T_{1i} \text{ uncensored after recensoring}).$$

Example 2: RMD. Similar to the RMN method, the RMD method also has $\Psi_1 = \Psi_2 = \Psi_3$ and $\Psi(0) = 0$. The additional complexity occurs due to the nondifferentiability of Ψ at 0. But this can be handled directly in the first approximation of our general derivation. From the definition of the median kernel, terms 3, 4, 5, and 6 in (A.3) can be directly computed. The first and the second term in (A.3) can be obtained by fixing t_2 at a certain constant, breaking the inner integral into the positive and negative parts, and then integrating over t_2. The final result is

$$-B_{ni} = 2 \int_{-\infty}^{K_{2i}} f_{2i}(t_2) \, dF_{2i}(t_2) + (1 - F_{2i}(K_{2i})) f_{1i}(K_{1i})$$

$$\text{for} \quad K_{1i} - K_{2i} > \theta.$$

By symmetry, the result of $K_{1i} - K_{2i} < \theta$ can also be obtained easily.

Example 3: RML. For the RML method with the logistic kernel as discussed in Section 6.3, we take

$$\Psi_d(D_i^{(\theta)} - \theta) = \frac{2e^{D_i^{(\theta)}-\theta}}{1 + e^{D_i^{(\theta)}-\theta}} - 1 \quad \text{for} \quad d = 1$$

$$= \frac{e^{D_i^{(\theta)}-\theta}}{1 + e^{D_i^{(\theta)}-\theta}} \quad \text{for} \quad d = 2$$

$$= \frac{e^{D_i^{(\theta)}-\theta}}{1 + e^{D_i^{(\theta)}-\theta}} - 1 \quad \text{for} \quad d = 3.$$

This leads to

$$-B_{ni} = \int_{-\infty}^{K_{2i}} \int_{-\infty}^{K_{2i}+\theta} \frac{2e^{t_1-t_2-\theta}}{(1 + e^{t_1-t_2-\theta})^2} \, dF_i(t_1, t_2) + f_{1i}(K_{2i} + \theta)$$

$$\times \int_{-\infty}^{K_{2i}} \left(\frac{1}{1 + e^{K_{2i}-t_2}} \right) dF_{2i}(t_2) + (1 - F_{2i}(K_{2i}))$$

$$\times \left\{ \int_{-\infty}^{K_{2i}+\theta} \frac{e^{t_1-K_{2i}-\theta}}{(1 + e^{t_1-K_{2i}-\theta})^2} \, dF_{1i}(t_1) + \frac{1}{2} f_{1i}(K_{2i} + \theta) \right\}$$

$$\text{for} \quad \theta < K_{1i} - K_{2i}$$

and

$$-B_{ni} = \int_{-\infty}^{K_{1i}-\theta} \int_{-\infty}^{K_{1i}} \frac{2e^{t_1-t_2-\theta}}{(1 + e^{t_1-t_2-\theta})^2} \, dF_i(t_1, t_2) + f_{2i}(K_{1i} - \theta)$$

$$\times \int_{-\infty}^{K_{1i}} \frac{e^{t_1-K_{1i}}}{1 + e^{t_1-K_{1i}}} \, dF_{1i}(t_1) + (1 - F_{2i}(K_{1i}))$$

$$\times \left\{ \int_{-\infty}^{K_{1i}-\theta} \frac{e^{K_{1i}-t_2-\theta}}{(1 + e^{K_{1i}-t_2-\theta})^2} \, dF_{2i}(t_2) + \frac{1}{2} f_{2i}(K_{1i} - \theta) \right\}$$

$$\text{for} \quad \theta > K_{1i} - K_{2i}.$$

Note that the second term of $-B_{ni}$ involves the density function of t_1 (or t_2), which contains the pair effects.

APPENDIX B: PROPERTIES OF THE LIKELIHOOD KERNEL FOR THE RECENSORING MAXIMUM LIKELIHOOD METHOD

First, we show that, assuming $g(\cdot)$ is symmetric about 0,

$$E\Psi(D_i^{(b)} - \theta, \delta_{1i}^{(b)}, \delta_{2i}^{(b)}) = E\left\{ \sum_{(a)} \frac{\frac{d}{d\theta} g(D_i^{(b)} - \theta)}{g(D_i^{(b)} - \theta)} + \sum_{(b)} \frac{g(D_i^{(b)} - \theta)}{1 - G(D_i^{(b)} - \theta)} + \sum_{(c)} \frac{-g(D_i^{(b)} - \theta)}{G(D_i^{(b)} - \theta)} \right\}$$

is antisymmetric at $b = \theta$.

First, note that the first summation term is antisymmetric because

$$\frac{\frac{d}{d\theta} g(x - \theta)}{g(x - \theta)} = \frac{-\frac{d}{d\theta} g(-(x - \theta))}{g(-(x - \theta))}.$$

The numerators and the denominators are equal on both sides due to the symmetry of $g(\cdot)$. Second, parts (b) and (c) are antisymmetric about 0 when $b = \theta$. Let $u = D_i^{(\theta)} - \theta$; then if $u > 0$, we have the case in part (b). On the other hand, when $u < 0$, the case will contribute in part (c). Note that the negative of the negative in the argument of part (b) is

$$\frac{-g(-u)}{1 - G(-u)} = \frac{-g(u)}{G(u)} = \text{part (c)}$$

also by the symmetry of $g(\cdot)$. Because the probability of being in case (b) and case (c) is equal when $b = \theta$, parts (b) and (c) are antisymmetric about 0.

Second, we show that Ψ is a monotone function for the logistic kernel. Let

$$\Psi(u) = \sum_{(a)} \frac{-\dot{g}(u)}{g(u)} + \sum_{(b)} \frac{g(u)}{1 - G(u)} + \sum_{(c)} \frac{-g(u)}{G(u)}.$$

Then

$$\dot{\Psi}(u) = \sum_{(a)} \frac{-\ddot{g}(u)g(u) + \dot{g}(u)^2}{g(u)^2} + \sum_{(b)} \frac{\dot{g}(u)(1 - G(u)) + g(u)^2}{(1 - G(u))^2}$$
$$+ \sum_{(c)} \frac{-\dot{g}(u)G(u) + g(u)^2}{G(u)^2}.$$

For the logistic kernel,

$$g(u) = \frac{e^u}{(1 + e^u)^2}, \qquad G(u) = \frac{e^u}{1 + e^u}$$

and

$$\dot{g}(u) = \frac{e^u(1 - e^u)}{(1 + e^u)^3}, \qquad \ddot{g} = \frac{e^u(1 - 2e^u)^2}{(1 + e^u)^4};$$

$$\text{part (a)} = \frac{2e^u}{(1 + e^u)^2} > 0,$$

and

$$\text{part (b)} = \frac{e^u}{(1 + e^u)^2} = \text{part (c)} > 0.$$

The monotonicity may not hold for other distributions in general, however.

[*Received February 1991. Revised September 1991.*]

REFERENCES

Armitage, P., and Gehan, E. A. (1974), "Statistical Methods for the Identification and Use of Prognostic Factors," *International Journal of Cancer*, 13, 16–36.

Bickel, P., Klassen, C., Ritov, Y., and Wellner, J. (1990), *Efficient and Adaptive Estimation in Non- and Semi-Parametric Models*, Baltimore: Johns Hopkins University Press.

Cederlof, R., Epstein, F. H., Friberg, L., Hrubec, Z., and Redford, E. (1971), "Twin Registries in the Study of Chronic Disease," *Acta Medica Scandinavica Supplementum*, 535, 9–40.

Cheng, K. F. (1984), "Asymptotically Nonparametric Tests With Censored Data," *Communications in Statistics, Part A—Theory and Methods*, 13, 1453–1470.

Cox, D. R., and Oakes, D. (1984), *Analysis of Survival Data*, New York: Chapman and Hall.

Dabrowska, D. M. (1989), "Rank Tests for Matched-Pair Experiments With Censored Data," *Journal of Multivariate Analysis*, 28, 88–114.

—— (1990), "Signed-Rank Tests for Censored Matched Pairs," *Journal of the American Statistical Association*, 85, 478–485.

Gross, S. T., and Huber, C. (1987), "Matched-Pair Experiments: Cox and Maximum Likelihood Estimation," *Scandinavian Journal of Statistics*, 14, 27–41.

Hauge, M., Harvard, B., Fischer, M., Gotlieb-Jensen, K., Juel-Nielsen, N., Raebild, I., Shapiro, R., and Videbech, T. (1968), "The Danish Twin Register," *Acta Geneticae Medicae et Gemellologiae*, 17, 315–332.

Holt, J. D., and Prentice, R. L. (1974), "Survival Analysis in Twin Studies and Matched-Pair Experiments," *Biometrika*, 61, 17–30.

Huber, P. J. (1981), *Robust Statistics*, New York: John Wiley.

Jablon, S., Neel, J. V., Gershawitz, H., and Atkinson, G. F. (1967), "The NAS–NRC Twin Panel: Method of Construction of the Panel, Zygosity Diagnosis, and Proposed Use," *American Journal of Human Genetics*, 19, 133–161.

Lachenbruch, P. A., Palta, M., and Woolson, R. F. (1982), "Analysis of Matched-Pairs Studies With Censored Data," *Communication in Statistics, Part A—Theory and Methods*, 11, 549–565.

Lee, J. J. (1989), "Estimation of the Location Parameter for Censored Paired Data," unpublished Ph.D. dissertation, University of California, Los Angeles, Dept. of Biostatistics.

Mantel, N., Bohidar, N. R., and Ciminera, J. L. (1977), "Mantel–Haenszel Analyses of Litter-Matched Time-to-Response Data, With Modification for Recovery of Interlitter Information," *Cancer Research*, 37, 3863–3868.

Mantel, N., and Ciminera, J. (1979), "Use of Log-Rank Scores in the Analysis of Litter-Matched Data on Time to Tumor Appearance," *Cancer Research*, 39, 4308–4315.

Michalek, J. E., and Mihalko, D. (1983), "On the Use of Log-Rank Scores in the Analysis of Litter-Matched Data on Time to Tumor Appearance," *Statistics in Medicine*, 2, 314–321.

—— (1984), "Linear Rank Procedures on Litter-Matched Data," *Biometrics*, 40, 487–491.

Neyman, J., and Scott, E. L. (1948), "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1–32.

O'Brien, P. C., and Fleming, T. R. (1987), "A Paired Prentice–Wilcoxon Test for Censored Paired Data," *Biometrics*, 43, 169–180.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J., Smith, P. G. (1976), "Design and Analysis of Randomized Clinical Trials Requiring Prolonged Observation of Each Patient, I: Introduction and Design," *British Journal of Cancer*, 34, 585–612.

—— (1977), "Design and Analysis of Randomized Clinical Trials Requiring Prolonged Observation of Each Patient, II: Analysis and Examples," *British Journal of Cancer*, 35, 1–39.

Prentice, R. L. (1978), "Linear Rank Tests With Right-Censored Data," *Biometrika*, 65, 167–179.

Reid, N. (1981), "Influence Functions for Censored Data," *The Annals of Statistics*, 9, 78–92.

Schluchter, M. D. (1985), "An Aligned Rank Test for Censored Data From Randomized Block Designs," *Biometrika*, 72, 609–618.

Tsiatis, A. A. (1986), "Two-Sample Linear Rank Tests for Survival Analysis," in *Proceedings of the Biopharmaceutical Section, American Statistical Association*, pp. 219–238.

Wei, L. J. (1980), "A Generalized Gehan and Gilbert Test for Paired Observations That are Subject to Arbitrary Right Censorship," *Journal of the American Statistical Association*, 75, 634–637.

Wei, L. J., and Pee, D. (1985), "Distribution-Free Methods of Estimating Location Difference With Censored Paired Data," *Journal of the American Statistical Association*, 80, 405–410.

Wild, C. J. (1983), "Failure Time Models With Matched Data," *Biometrika*, 70, 633–641.

Woolson, R. F., and Lachenbruch, P. A. (1980), "Rank Tests for Censored Matched Pairs," *Biometrika*, 67, 597–606.

—— (1981), "Rank Tests for Randomized Block Designs," *Biometrika*, 68, 427–435.