CENTER FOR COMPUTATIONAL MATHEMATICS REPORTS

University of Colorado at Denver
P.O. Box 173364, Campus Box 170
Denver, CO 80217-3364

Fax: (303) 556-8550
Phone: (303) 556-8442
http://www-math.cudenver.edu/

August 1996 UCD/CCM Report No. 85

An Estimate of the Variance of Estimators for Lead Time and Screening Benefit in Randomized Cancer Screening Trials

K. Kafadar, P.C. Prorok and P.J. Smith

An Estimate of the Variance of Estimators for Lead Time and Screening Benefit in Randomized Cancer Screening Trials

Karen Kafadar

University of Colorado - Denver
Department of Mathematics, Box 170
Denver, Colorado 80217-3364

Philip C. Prorok

National Cancer Institute
Division of Cancer Prevention and Control
Bethesda, Maryland 20892-7354

Paul J. Smith

University of Maryland
Department of Mathematics
College Park, Maryland 20742

ABSTRACT

Variance estimators are derived for estimators of the average lead time and average benefit time due to screening in a randomized screening trial via influence functions. The influence functions demonstrate that these estimators are asymptotically equivalent to the mean difference, between the study and control case groups, in the survival times. For estimating benefit time, the survival time is measured since start of study; for estimating lead time, the survival time is measured since time of diagnosis. Asymptotic variances of these estimators can be calculated in a straightforward manner from the influence functions, and these variances can be estimated from actual trial data. The performance of the variance estimates is assessed via a simulated screening trial. The situation involving censored data is also discussed.

Key words: Kaplan-Meier estimate, survival curve, influence function, asymptotic variance, censoring

1. Introduction.

Randomized screening trials provide an objective assessment of the benefit of screening for cancer and other chronic diseases, in terms of the real increase in survival time for cases detected among participants offered screening over that among participants not offered screening (Prorok and Connor 1986; Miller and Bulbrook 1982). *Lead time* is viewed as a bias when assessing the benefit of screening as an increase in survival time. However, estimates of both the average lead time and the average benefit time due to screening tests are important in their own right and have implications for setting screening

guidelines. For example, a randomized trial is currently underway to evaluate screening for cancers of the prostate, lung, colon, and ovary (Gohagen et al. 1994). The average benefit time and average lead time for these screening tests will be important for setting health care policy regarding their use and frequency.

A common model for disease progression consists of three phases: a disease-free phase, a preclinical or sojourn phase during which disease is present and detectable by screening but clinical symptoms are absent, and a clinical phase during which symptoms are present and a diagnosis is made (Morrison 1992). The difference in time between when disease is screen-detected (preclinical phase) and when it would have been diagnosed in the absence of screening is called *lead time*. Screening benefit usually is measured as a reduction in population mortality from the disease at a certain point in time following the start of the study (Morrison 1982; Prorok 1984, Shapiro et al. 1988). An alternative measure is the *average benefit time* (Gail 1975; Habbema et al. 1986; Kafadar and Prorok 1994), or the average length of time by which survival has been extended in the study arm (participants offered screening, sometimes denoted "screened arm") over the control arm (those not offered screening, sometimes denoted "unscreened arm"). Estimates of average lead time and average benefit time, together with estimates of their uncertainties, are the subject of this paper.

Typically, survival is measured as the difference between time of diagnosis and time of endpoint (cure, remission, or death) from the disease under investigation. This difference confounds benefit time (true extension of survival time) and lead time (time by which diagnosis is advanced as a result of earlier detection by screening). Various authors have proposed estimators of average lead time (Hutchison and Shapiro 1968, Zelen and Feinleib 1969, Prorok 1976, Morrison 1982, Day and Walter 1984), usually without discussion of their variances. Kafadar and Prorok (1994) show that, when the effects of benefit time and lead time on survival are additive, average benefit time and average lead time can be estimated from data from a randomized screening trial.

In this paper we derive the variances of these estimators by calculating the influence functions of the estimators viewed as bivariate functionals on the space of distribution functions. The influence function for a single sample was defined by Hampel (1974) and extended to two samples from populations having a common distributional form but differing in location (Lambert 1981, Hampel et al. 1986). The influence

function provides a tool for obtaining an estimate of the finite sample variance, because the asymptotic variance of the estimator is the integral of the squared influence function (Huber 1981, Hampel et al. 1986). We show that the influence function of the estimator for average benefit time is equivalent to that of the sample mean difference in case survival times, measured since start of study, between the two groups, while the influence function of the estimator for average lead time is equivalent to that of the sample mean difference in times from start of study to diagnosis of cases between the two groups. (Some authors call these quantities "program benefit time" and "program lead time," to be distinguished from "screened benefit time" and "screened lead time," to emphasize that the average applies to *all* individuals in the study arm, not simply those for whom disease was screen-detected.) Performance comparisons among the proposals in the literature for estimating average lead time is discussed in a separate paper (Kafadar and Prorok 1996a). Variances of the estimators for average benefit time and average lead time are derived in Section 2. The censored version is described briefly in Section 3. We demonstrate the success of these finite sample estimates of the variances for average lead time and average benefit time via a computer simulated screening trial in Section 4; concluding remarks are given in Section 5.

2. Estimators of average benefit time and average lead time based on uncensored data.

2.1 Notation and definitions.

Average benefit time, B, and average lead time, L, are estimated using survival times of histologically confirmed cases of the disease under investigation. We define the estimators of B and L using the notation:

X_i = time from start of study to endpoint for case i ($i = 1, ..., n$) in the screened group, cdf = F_S

$H_S = 1 - F_S$ = survival function of $\{X_i\}$

Y_j = time from start of study to endpoint for case j ($j = 1, ..., m$) in the control group, cdf = F_C

$H_C = 1 - F_C$ = survival function of $\{Y_j\}$

V_i = time from diagnosis to endpoint for case i ($i = 1, ..., n$) in the screened group, cdf = G_S

$R_S = 1 - G_S$ = survival function of $\{V_i\}$

U_j = time from diagnosis to endpoint for case j ($j = 1, ..., m$) in the control group, cdf = G_C

$R_C = 1 - G_C$ = survival function of $\{U_j\}$

The endpoint of interest may be cure, remission, or death from the disease under investigation. The last choice is the least ambiguous, but the methods described herein apply regardless of which endpoint is being considered. We assume that the cases from the two groups would have comparable disease characteristics in the absence of screening; this issue is discussed in §4.2

The estimator $\hat{B}$ of average benefit time uses only cases from the disease targeted by screening and hence is a function of the first two sets of random variables (X_i, Y_j), while the estimator $\hat{L}$ of average lead time uses all four sets of random variables (X_i, Y_j, V_i, U_j). If x_i, y_j, v_i, u_j denote the observed values of these random variables, respectively, then (cf. Kafadar and Prorok 1994)

$$\hat{B} = \sum_{i=1}^{n} [\, x_i - \hat{H}_C^{-1}(\hat{H}_S(x_i)) \,] / n \tag{1}$$

$$\hat{L} = \sum_{j=1}^{m} [\, \hat{R}_S^{-1}(\hat{R}_C(u_j)) - u_j \,] / m \;-\; \hat{B} \tag{2}$$

where "^" denotes an estimate of the corresponding quantity. Intuitively, $\hat{B}$ is the area between the two curves $\hat{H}_S$ and $\hat{H}_C$, and $\hat{L}$ is the area between the two curves $\hat{R}_C$ and $\hat{R}_S$ adjusted for $\hat{B}$ (Fig.1). Computationally, $\hat{B}$ is the average of all differences $(x_i - \hat{y}_i)$, where $\hat{y}_i = \hat{H}_C^{-1}(\hat{H}_S(x_i))$ is the value from $\hat{H}_C$ for which the survival probabilities $\hat{H}_C(\hat{y}_i)$ and $\hat{H}_S(x_i)$ are equal. (This interpretation of $\hat{B}$ does not require F_S or F_C to have finite first moment; e.g., if F_S and F_C are Cauchy distribution functions which differ only in location, then the area between F_S and F_C is finite, although their means are not.)

2.2 Interpretation in terms of functionals

Let $B(F_S, F_C)$ denote a functional defined on $\Omega \times \Omega$, where Ω is the space of probability distributions and $B(F_S, F_C)$ is defined by

$$B(F_S, F_C) = \int_0^{\infty} [\, x - H_C^{-1}(H_S(x)) \,]\, dF_S(x)\,. \tag{3}$$

The estimator $\hat{B}$ can be written as $\hat{B} = B(\hat{F}_S, \hat{F}_C)$, since a sample average over $\{x_i,\ i = 1, \ldots, n\}$ is equivalent to integration with respect to the empirical cdf $\hat{F}_S$. If the means of F_S and F_C exist and equal $\mu(F_S)$ and $\mu(F_C)$ respectively, then, by a change-of-variables argument,

$$B(F_S, F_C) = \mu(F_S) - \int_0^1 H_C^{-1}(z)\, dz \;=\; \mu(F_S) - \mu(F_C)\,.$$

In a completely analogous fashion, the lead time functional is defined by

$$L(R_S, R_C, F_S, F_C) = \int_0^\infty [\, R_S^{-1}(R_C(y)) - y \,]\, dG_C(y) \; - \; \{ \int_0^\infty [\, x - H_C^{-1}(H_S(x)) \,]\, dF_S(x) \}\,, \qquad (4)$$

which, when the means of F_S, F_C, G_S, and G_C exist and equal $\mu(F_S)$, $\mu(F_C)$, $\mu(G_S)$, and $\mu(G_C)$ respectively, becomes

$L(R_S, R_C, F_S, F_C) = [\mu(G_S) - \mu(G_C)] - [\mu(F_S) - \mu(F_C)] = [\mu(F_C) - \mu(G_C)] - [\mu(F_S) - \mu(G_S)]$

$=$ *average time to diagnosis among control cases − average time to diagnosis among study cases* .

We will see that the influence functions of both $B(F_S, F_C)$ and $L(R_S, R_C, F_S, F_C)$ are consistent with the influence functions of the differences in two means.

2.3 The influence function for the mean benefit time.

The univariate influence function describes the effect of an additional observation at x on the functional. In the same way that sensitivity is often measured by a rate of change from a nominal level (slope), the influence function can be regarded as the first derivative of the functional, evaluated at the nominal (uncontaminated) distribution in the direction of the contaminating distribution. In two dimensions, this concept can be generalized to measure the impact on $B(F_S, F_C)$ when F_S is replaced by a mixture of F_S and a point mass at x, and a F_C is replaced by a mixture of F_C and a point mass at y. Let $F_{S,\,r} = (1 - r)F_S + r\Delta_x = F_S + r(\Delta_x - F_S)$ and $F_{C,\,t} = (1 - t)F_C + t\Delta_y = F_C + t(\Delta_y - F_C)$ represent the contaminated distributions. A linearization of $B\,(F_S, F_C)$ contains two linear functionals involving analogues of the partial derivatives of a function of two variables with respect to the first and second variables. These are the influence functions

$$IF_S(\, x; B,\, F_S\,,\, F_C\,) = (\partial/\partial r)\, B(F_{S,\,r}; F_{C,\,t})\;|_{r=t=0}$$
$$IF_C(\, y; B,\, F_S\,,\, F_C\,) = (\partial/\partial t)\, B(F_{S,\,r}; F_{C,\,t})\;|_{r=t=0}$$

so the estimator $\hat{B} = B(\hat{F}_S; \hat{F}_C)$ can be linearized as

$$B(\hat{F}_S; \hat{F}_C) = B(F_S; F_C) + \int_0^\infty IF_S(x\,; B,\, F_S,\, F_C) d(\hat{F}_S - F_S)(x) \; + \; \int_0^\infty IF_C(y\,; B,\, F_S,\, F_C) d(\hat{F}_C - F_C)(y)$$
$$+ \; \textit{higher order terms}$$

(Reid 1981, Equation (1.6)). Because the observations from the study and control arms are independent, the asymptotic variance of $\hat{B}$ is

$$asy\ var(\hat{B}) = n^{-1}\int [\ IF_S(x; B, F_S, F_C)\ ]^2 dF_S(x) \ + \ m^{-1}\int [\ IF_C(y; B, F_S, F_C)\ ]^2 dF_C(y)\ .$$

A finite-sample estimate of the asymptotic variance is $\hat{Var}(\hat{B})$, given by

$$\hat{Var}(\hat{B}) = n^{-2}\sum_{i=1}^{n} IF_S^2(x_i; B, \hat{F}_S, \hat{F}_C) \ + \ m^{-2}\sum_{j=1}^{m} IF_C^2(y_j; B, \hat{F}_S, \hat{F}_C)\ . \tag{5}$$

In the next section we outline the procedure for obtaining the influence functions in (5).

2.4 Calculating the influence function for the mean benefit time

The influence function for $B(F_S, F_C)$ (3) leads to an estimate of the variance of $\hat{B} = B(\hat{F}_S, \hat{F}_C)$ via (5). Since the first term in (3), namely $\int_0^\infty x\ dF_S(x)$, is simply the mean function for F_S and does not involve F_C, it is easiest to derive $IF_C(y; B, F_S, F_C)$ when the means $\mu(F_S)$ and $\mu(F_C)$ are assumed to exist. As in the previous section, let $H_C = 1 - F_C$ and $H_S = 1 - F_S$; further, let $f_S(t)$ and $f_C(t)$ be the densities of F_S and F_C, respectively. Define $H_{C,\,t} = 1 - F_{C,\,t}$ and $H_{S,\,r} = 1 - F_{S,\,r}$ where $F_{C,\,t}$ and $F_{S,\,r}$ are the contaminated cumulative distribution functions defined in Section 3.3. Since

$$\begin{aligned}(\partial/\partial t)B(F_{S,\,r}, F_{C,\,t})\ |_{r=t=0} \ &= \ (\partial/\partial t)\int_0^\infty H_{C,\,t}^{-1}(H_{S,\,r}(y))\ dF_{S,\,r}(y)\ |_{r=t=0} \ = \ (\partial/\partial t)\int_0^1 H_{C,\,t}^{-1}(1-z)\ dz\ |_{r=t=0} \\ &= \int_0^1 (\partial/\partial t)\ H_{C,\,t}^{-1}(1-z)\ |_{r=t=0}\ dz\ ,\end{aligned}$$

the integrand can be obtained from straightforward calculus to yield

$$(\partial/\partial t)\ H_{C,\,t}^{-1}(1-z)\ |_{t=0} \ = \ -(\Delta_y - F_C)(H_C^{-1}(1-z))\,/\,f_C(H_C^{-1}(1-z))\ .$$

Therefore,

$$IF_C(y; B, F_S, F_C) = \int_0^1 (\Delta_y - F_C)(H_C^{-1}(1-z))\,/\,f_C(H_C^{-1}(1-z))\ dz = y - \mu(F_C)\ ,$$

which is the influence function for the functional which defines the mean (Hampel 1974).

By a similar argument, the influence function with respect to F_S leads to

$$IF_S(x; B, F_S, F_C) = (\partial/\partial t)\int_0^\infty [\ x - H_C^{-1}(H_{S,\,t}(x))\ ]\ dF_{S,\,t}(x)\ |_{t=0} = x - \mu(F_S)\ .$$

These derivations suggest replacing $IF_S^2(x_i; B, \hat{F}_S, \hat{F}_C)$ and $IF_C^2(y_j; B, \hat{F}_S, \hat{F}_C)$ in (5) with $(x_i - \bar{x})^2$ and $(y_j - \bar{y})^2$, respectively; instead, we use unbiased estimates of the sample variances for the following estimator of the variance of $\hat{B}$:

$$\hat{Var}(\hat{B}) = \sum_{i=1}^{n} (x_i - \bar{x})^2/[n\cdot(n-1)] + \sum_{i=1}^{m} (y_i - \bar{y})^2/[m\cdot(m-1)] = s_x^2/n + s_y^2/m\,. \qquad (6)$$

The influence function for $B(F_C, F_S)$ shows that $\hat{B}$ is asymptotically equivalent to $\bar{x} - \bar{y}$, the difference in the sample means of the times (since start of study) to death. We denote this estimator by $\bar{B} = \bar{x} - \bar{y}$.

2.5 Discussion

Although $\hat{B}$ is not calculated in this way, it is easy to see why it is asymptotically equivalent to the difference in the means of the two distributions $\hat{F}_S$ and $\hat{F}_C$. If x represents an observation (time between start of study and endpoint) in the study arm, then $y^* = H_C^{-1}(H_S(x))$ represents a survival time that might have been observed in the control arm at which the probability of survival, $H_C(y^*)$, would be exactly the same as that for observation x in the study arm, $H_S(x)$ [i.e., $H_C(y^*) = H_S(x)$]. If we had an extremely large number of cases in each arm, then, for any observed survival time in the study arm, say x_i close to x, with high probability, some observed y_j would be very close to $y^* \equiv H_C^{-1}(H_S(x))$, because the graph of $y_i^* \equiv \hat{H}_C^{-1}(\hat{H}_S(x_i))$ versus x_i is uniformly close to the graph of $H_C^{-1}(H_S(x))$ versus x (Feller 1966, p.37). The estimator $\hat{B}$ simply averages all differences $\{x_i - \hat{H}_C^{-1}(\hat{H}_S(x_i))\} \equiv \{x_i - y_i^*\}$ over the duration of the follow-up period of the study (cf. Fig.1).

Besides the fact that the definition of $\hat{B}$ does not require finite $\mu(F_S)$ and $\mu(F_C)$, an advantage of (1) over a simple difference in sample means is that individual differences $\{z_i\} = \{x_i - \hat{H}_C^{-1}(\hat{H}_S(x_i))\}$ may be useful for diagnostic purposes. In addition, one might consider a robust estimator of B from the $\{z_i\}$ using the trimmed mean or the biweight.

Finally, the influence function for the lead time estimator (2) can be derived in the same manner as above and is the same as the influence function for the difference in the mean times to diagnosis for the two groups. Since the times to diagnosis for cases in the study and control groups are $\{X_i - V_i,\ i=1,\ldots,\ n\}$ and $\{Y_j - U_j,\ j=1,\ldots,\ m\}$ respectively, the influence functions for L are simply

$$IF_S(x; L, F_S, F_C, G_S, G_C) = x - \mu_{X-V}$$
$$IF_C(y; L, F_S, F_C, G_S, G_C) = y - \mu_{Y-U}$$

where μ_{X-V} and μ_{Y-U} represent the means of the random variables $X-V$ and $Y-U$, respectively. Since the case groups are independent, the variance of $\hat{L}$ can be estimated as the sum of the sample variances of the two mean times to diagnosis:

$$\hat{Var}(\hat{L}) = s^2_{x-v}/n + s^2_{y-u}/m \,. \qquad (7)$$

Thus this estimator is very similar to that used by Shapiro et al. (1988) in the breast cancer screening trial of the Health Insurance Plan (HIP) in New York. We denote $\bar{L} = \overline{(X-V)} - \overline{(Y-U)}$. (The Shapiro estimator actually evaluates this difference in averages at "catch up" and one year prior to "catch up" (cf. §4.2), and averages the two results.)

3. Censored observations.

3.1 Notation and definitions.

Suppose that screening trial participants are followed until K years after the start of the study, so that the survival times of some cases are censored. The observable data are (X_i, V_i, D_{Si}), $i = 1,..., n$ in the study group and (Y_j, U_j, D_{Cj}), $j = 1,..., m$ in the control group. Here X_i and Y_j are the times from start of study until endpoint or time K, whichever occurs first, V_i and U_j are the times from diagnosis until endpoint or K, and D_{Si} and D_{Cj} are the indicators which take the value 1 if the actual time of the endpoint is observed and 0 otherwise.

The variables of interest are the uncensored times of endpoints, denoted X^o_i, V^o_i, Y^o_j, and U^o_j. Their cdf's are denoted F^o_S, G^o_S, F^o_C, and G^o_C, respectively. The corresponding survival functions are H^o_S, R^o_S, H^o_C, and R^o_C, The estimators of average benefit time and average lead time become

$$\hat{B} = \int_0^K [\, x - \hat{H}^{o\,-1}_C(\hat{H}^o_S(x)) \,]\, d\hat{F}_S{}^o(x)$$

$$\hat{L} = \int_0^K [\, \hat{R}^{o-1}_S(\hat{R}^o_C(u)) - u \,]\, d\hat{H}^o_C \;-\; \hat{B}$$

where $\hat{H}^o_S$, $\hat{H}^o_C$, $\hat{R}^o_S$, and $\hat{R}^o_C$ are the Kaplan-Meier estimators of the survival functions. We assume that K is large, so censorship can be regarded as noninformative. This issue is discussed in §4.3.

Note that $\hat{B}$ can be represented more simply as

$$\hat{B} = n^{-1}[\, \textstyle\sum{}' X_i + (n-n^*)K] \;-\; n^{-1}[\, \sum{}' \hat{H}^{o\,-1}_C(\, \hat{H}^o_S(X_i)\,) \;+\; (n-n^*)\hat{H}^{o\,-1}_C(\hat{H}^o_S(K))\,]$$

where the prime denotes summation over the uncensored observations in the screened group ($D_{Si} = 1$) and $n^* = \sum D_{Si}$ is the number of uncensored X_i's. An alternative estimator is

$$n^{-1}[\, \textstyle\sum{}' X_i + (n - n^*)K \,] \;-\; m^{-1}[\sum{}'' Y_j + (m - m^*)K \,] \;=\; \bar{X} - \bar{Y}$$

where $m^* = \sum D_{Cj}$ is the number of uncensored observations in the control group and the double prime denotes summation over these m^* uncensored observations. Arguments similar to those of Section 2 show that these two estimators are asymptotically equivalent.

3.2 Influence functions and variance estimators.

The influence functions for the Kaplan-Meier estimate and related functionals have been computed by Reid (1981). In particular, her results imply that if observations are censored at time K, then the mean functional based on the Kaplan-Meier estimate $\hat{H}_S^o$, namely $\hat{\mu}_S = \int_0^K x\, d\hat{F}_S^o$, has asymptotic variance

$$asy\ var(\hat{\mu}_S) = n^{-1} \int_0^K [H_S^o(s)]^{-2} \left[\int_s^T H_S^o(t) dt \right]^2 dF_S^o(s)$$

for the special case of censoring at time K. Using arguments similar to those of §3.4, the asymptotic variance of $\hat{B}$ is

$$asy\ var(\hat{B}) = asy\ var(\bar{B}) = n^{-1} \int_0^K [H_S^o(s)]^{-2} \left[\int_s^K H_S^o(t)\, dt \right]^2 dF_S^o(s)$$
$$+ m^{-1} \int_0^K [H_C^o(s)]^{-2} \left[\int_s^K H_C^o(t)\, dt \right]^2 dF_C^o(s)\,.$$

By substituting Kaplan-Meier estimates for the true survival functions, one obtains

$$\hat{Var}(\hat{B}) = \hat{Var}(\bar{B}) = s_x^2/n + s_y^2/m$$

where now s_x^2 and s_y^2 are the sample variances of the censored data X_i and Y_j.

Estimating $Var(\hat{L})$ is more complicated because $\hat{L}$ depends on estimates of $E(V_i^o)$ and $E(U_j^o)$, and the $\{V_i^o\}$ and $\{U_j^o\}$ are subject to random informative censoring. However, as argued in §4.3, when K is large, the censorship is light and nearly independent of the survival times of interest. Under these circumstances, $\hat{L}$ and $\bar{L} = (\overline{X-V} - \overline{Y-U})$ are approximately equal, and their variances may be estimated by

$$s_{x-v}^2/n + s_{y-u}^2/m\,,$$

just as in the uncensored case, but using sample variances of censored data.

3.3 Censorship and bias.

The usual theory of the Kaplan-Meier estimator is based on the assumption that censorship is independent of the lifetimes of interest. In this situation, the times from start of study until endpoint are

subject to deterministic type I censoring, so the assumption holds. However, the times from diagnosis until endpoint are censored as follows:

$$V_i = \min \{ V_i^o , \; K - (X_i^o - V_i^o) \}$$
$$D_{Si} = 1 \quad \text{if } X_i^o > K \;\; \textit{or equivalently} \;\; V_i^0 > K - (X_i^o - V_i^o)$$
$$= 0 \quad \textit{otherwise} .$$

Since $(X_i^o - V_i^o)$ is the time from start of study to time of diagnosis, censoring of V_i^o is related to time of diagnosis; i.e., the earlier diagnosis in the screened arm may be censored because the study was halted at time K. If screening is indeed beneficial, this earlier diagnosis would lead to extended life. Thus, in the screened arm, a censored versus an uncensored V_i^o could be related to the effect of screening.

To avoid the censoring bias, we assume that K is at least as large as the time at which case groups are deemed "comparable". This issue is discussed in Section 4.2.

4. Demonstration.

To evaluate the performance of the variance estimators, we simulate a randomized screening trial for the uncensored case only.

4.1 Simulation design.

A three-phase disease natural history model (healthy, preclinical, clinical disease) is simulated with parameters that are reasonably consistent with a breast cancer screening trial such as that described in Shapiro et al. (1988). Cases are assumed to arrive according to a Poisson process with rate $\lambda = 0.001$. This value was chosen because the annual incidence rate of female breast cancer is the U.S. is approximately one per thousand women in the age groups currently targeted for screening. We assume there are 20,000 persons in each arm (study and control). The pseudo-random variates P and D, corresponding to durations of preclinical disease and clinical disease respectively, are generated according to a bivariate gamma distribution (Chen, Prorok, and Graf 1984). Three members of this family are:

(a) $\mu_P = 2, \sigma_P^2 = 4, \mu_D = 5, \sigma_D^2 = 25, corr(P , D) = 0.0.$

(b) $\mu_P = 2, \sigma_P^2 = 1, \mu_D = 4, \sigma_D^2 = 4, corr(P , D) = 0.3;$

(c) $\mu_P = 2, \sigma_P^2 = 1, \mu_D = 2, \sigma_D^2 = 1, corr(P , D) = 0.9.$

Members of this family for which the correlation is zero, as in (a), have marginal exponential densities. In the control arm, the times of diagnosis and endpoint occur at the end of the preclinical and clinical phases, respectively. In the study arm, screens occur annually at times $\{t_j = j, j = 0, 1, ..., 5\}$ with test sensitivity $\beta = 0.80$ (i.e., 20% chance of a false negative result). If the i^{th} case is detected at a screen, then: (1) the lead time for the case, l_i, is calculated as the time difference between the screen and the end of the preclinical duration in the absence of screening, and (2) a random benefit time for the case, b_i, is generated as a function of the case's preclinical duration. If the case from the study arm is an interval case (detected between screens), both the lead time and the benefit time are 0. (We assume that trial participation heightens awareness of health matters equally among individuals in both arms of the trial.) Cases can begin as early as $t_0 - \mu_P - 2\sigma_P$ but cases whose preclinical durations end before t_0 (i.e., clinical disease has surfaced before the initial screen at time t_0) are excluded from the trial. Cases in both arms continue to be ascertained until $K = 20$ years following start of study.

Other choices for the distributions of P and D, $\{t_j\}$, β, and K can be selected. In addition, both the test sensitivity β and the benefit times b_i may depend on additional simulated variates such as tumor grade or age at start of study. A flow chart of the process for screen-detected cases and the algorithm for generating a random benefit time appear in Kafadar and Prorok (1996a).

4.2 Numbers of cases used in estimators

Note that, because the durations P and D are stochastic, the numbers of cases observed in the two groups at time t, namely $N_S(t)$ and $N_C(t)$, are random variates; for a Poisson process with rate λ, we expect $E(N_C) = 20{,}000\lambda t$ cases in the control arm. (The formula for the expected number of cases observed in the study arm by time t can be calculated from the formulas in the appendix of Kafadar and Prorok 1994). The $N_S(t)$ cases in the study arm form a mixture of cases which are screen-detected, interval (detected between screens), and post-screened (detected after screening ended). As t increases, many of the $N_S(t)$ cases will have zero lead time (as occurs for interval cases and cases diagnosed after screening ends), so it may be difficult to detect a non-zero average lead time or non-zero average benefit time in a comparison of all $N_S(t)$ cases with all $N_C(t)$ cases. Conversely, if t is too small, then the $N_S(t)$ cases in the study arm form a length-biased sample; i.e., the presence of long sojourn times results in a greater likelihood of detecting the

case by screening. (This has been noted in Aron and Prorok 1986, Connor and Prorok 1994, Etzioni and Self 1995, Etzioni et al. 1995.)

Aron and Prorok (1986) suggest a methodology for obtaining comparable cases from both arms to minimize the effects of both length biased sampling and dilution bias: we select all cases which occur up to the "catch up time," which is defined as the time, say C_{AP}, at which the cumulative number of cases observed in the control group first equals or exceeds that in the study group. Kafadar and Prorok (1996b) propose a modified catchup rule which depends on an estimate of the mean sojourn time:

$$C_\mu = T + \hat{\mu} + \hat{\mu}^{1/2}$$

where T denotes the time since start of study of the final screen and $\hat{\mu}$ is an estimate of the mean sojourn time. (In roughly 27-30% of the simulation runs discussed in the next section, the Aron-Prorok "catchup time" never occurs, so $C_{AP} \equiv K$ for such runs; this percentage drops to 0-3% if C_μ is used to define comparable case groups.)

In the next section, we will compare estimates of average lead time and average benefit time, and their variance estimates, based on these two rules for defining comparable case groups: one set of estimates is based on the first $N_S(C_\mu)$ and $N_C(C_\mu)$ cases in the study and control arms at time C_μ, respectively, and one set of estimates is based on the first $N_S(C_{AP})$ and $N_C(C_{AP})$ cases in the study and control arms at time C_{AP}, respectively.

4.3 Estimates of average benefit time and variance estimates.

Based on the results in Section 3, two forms for the estimator of average benefit time are investigated. The first estimator is a finite sample implementation of (3) (or Eqn (1), where $n = N_S(C_\mu)$):

$$\hat{B}(C_\mu) = \sum_{i=1}^{N_S(C_\mu)} [\, x_i - \hat{H}_C^{-1}(\hat{H}_S(x_i))\,] / N_S(C_\mu)\,, \tag{9}$$

where $\hat{H}_C$ and $\hat{H}_S$ are Kaplan-Meier estimated survival functions since start of study in the control and study case groups, respectively, at time C_μ. $\hat{B}(C_{AP})$ is defined similarly. The second estimator is a simple difference in the mean times to endpoint for the two case groups:

$$\bar{B}(C_\mu) = \sum_{i=1}^{N_S(C_\mu)} x_i / N_S(C_\mu) - \sum_{j=1}^{N_C(C_\mu)} y_j / N_C(C_\mu) = \bar{x} - \bar{y}\,; \tag{10}$$

$\bar{B}(C_{AP})$ is defined likewise. The variances of $\hat{B}(C_\mu)$ and $\bar{B}(C_\mu)$ are estimated as

$$\hat{Var}(\hat{B}) = s_x^2 / N_S(C_\mu) + s_y^2 / N_C(C_\mu) \tag{11}$$

where s_x^2, s_y^2 are the usual sample variances of samples $\{x_i\}$ and $\{y_j\}$, respectively.

4.4 Estimates of average lead time and variance estimates.

In a similar fashion, two estimates of average lead time are computed for each simulated trial. The first estimate, $\hat{L}$, is a finite sample implementation of (4) (or Eqn (2), where $m = N_C(C_\mu)$):

$$\hat{L} = \sum_{j=1}^{N_C} [\, \hat{R}_S^{-1}(\hat{R}_C(u_j)) - u_j \,] / N_C(C_\mu) - \hat{B}(C_\mu) \tag{12}$$

where $\hat{R}_S$ and $\hat{R}_C$ are Kaplan-Meier curves based on survival since time of diagnosis for the two case groups. The second estimator, $\bar{L}$, is asymptotically equivalent to (12):

$$\bar{L} = \sum_{j=1}^{N_C(C_\mu)} u_j / N_C(C_\mu) - \sum_{i=1}^{N_S(C_\mu)} v_i / N_S(C_\mu) - \bar{B} \;=\; (\bar{v} - \bar{u}) - (\bar{x} - \bar{y}) \;=\; (\bar{y} - \bar{u}) - (\bar{x} - \bar{v}) \tag{13}$$

where $\{u_j\}$ and $\{v_i\}$ are the observed survival times since diagnosis for the control and screened groups, respectively. The variance of either $\hat{L}$ or $\bar{L}$ is estimated by:

$$\hat{Var}(\hat{L}) = \hat{Var}(\bar{L}) = (s_y^2 + s_u^2)/N_C(C_\mu) - 2Cov(\bar{y}, \bar{u}) + (s_x^2 + s_v^2)/N_S(C_\mu) - 2Cov(\bar{x}, \bar{v})\,. \tag{14}$$

The estimators $\hat{L}(C_{AP})$, $\bar{L}(C_{AP})$, and $\hat{Var}(\hat{L}(C_{AP}))$ are defined by replacing C_μ by C_{AP}.

Notice that all of the above equations for the standard errors are designed to reflect variability only within a single trial, conditional on observing $N_S(C_\mu)$ and $N_C(C_\mu)$ cases in the two groups; e.g., $Var(\bar{Y}) = Var(Y)/N_C(C_\mu) \approx s_y^2/N_C(C_\mu)$. In reality, $Var(\bar{Y}) = E(Var(\bar{Y}\,|\,N_C(C_\mu))) + Var(E(\bar{Y}\,|\,N_C(C_\mu))) = E(\sigma_y^2/N_C(C_\mu)) = \sigma_y^2 \cdot E(1/N_C(C_\mu))$, since the second term is the variance of a constant. When the time between case arrivals is exponential, the number of cases in the control group at the end of the study, K, is a Poisson variate with mean $\nu = 20{,}000\lambda K$; conditional on $N_C(K) > 0$, the expectation of its reciprocal is exactly $\nu^{-1} - e^{-\nu}/(1 - e^{-\nu}) \approx \nu^{-1}$ (when $\nu > 11$, the relative error in this approximation is less than 0.01%.) Since C_μ is a random time, the numbers of cases used in the estimators are also random. Thus, an assessment of the variability in these estimators across all trials is not easy to calculate analytically, even in the simple situation of Poisson case arrivals. However, in a simulation, the accuracy of these estimators may be evaluated within a trial and compared with the actual variance in a given trial as well as that across tri-

als. We will see in the next section that the variability in C_μ is much less than that in C_{AP}, so that the variance estimates are in fact quite close to the actual variances in the simulation.

Finally, notice that the form of estimators makes it tempting to consider an estimate of variance based on the differences being summed. Such estimates were considered but were found to be much smaller than the actual variance. Thus they will not be considered further.

4.5 Simulation.

We use a simulation to assess the performance of the estimators of average benefit time [(9), (10)] and average lead time [(12), (13)] as well as the success of the corresponding estimates of their variances [(11), (14)] in characterizing the variability. Because simulation results involve many variables, we use the following notation to describe the output, illustrated with estimator $\hat{L}(C_\mu)$: (in all cases, the subscript k denotes the simulation run, $k = 1, ..., 500$):

n_k = Number of cases detected in the study group at time C_μ

m_k^L = average of the n_k actual lead times for the n_k cases

L_k = value of $\hat{L}$ in simulation run k

Using this notation, we calculate the following performance measures (all summations are over k from 1 to 500):

(1) *ave* $\{m_k^L\} = \sum m_k^L/500 \equiv \overline{m^L}$ = actual average lead time across all 500 simulations

(2) *standard error* $(\overline{m^L}) = [\sum (m_k^L - \overline{m^L})^2/ 500 \cdot 499\ ]^{1/2} = SE(\overline{m^L})$

(3) *ave* $\{L_k\} = \sum L_k/500 \equiv \bar{L}$ = estimated average lead time using $\hat{L}$

(4) *standard error* $(\bar{L}) = [\sum (L_k - \bar{L})^2/ 500 \cdot 499\ ]^{1/2} = SE(Lbar)$

(5) *ave* $\{L_k - m_k^L\} = \sum (L_k - m_k^L) / 500$ = average bias in $\hat{L} \equiv bias(\hat{L})$

(6) *standard error (average bias)* $= [\sum (L_k - m_k^L - bias(\hat{L}))^2/ 500 \cdot 499\ ]^{1/2} = SE(bias(\hat{L}))$

(7) *ave* $\{\hat{Var}_k(\hat{L})\} = \sum \hat{Var}_k(\hat{L}) / 500 \equiv \overline{\hat{Var}(\hat{L})}$ = average variance in average lead time

(8) *standard error (average estimated variance of L)* $= \{\sum [\hat{Var}_k(\hat{L}) - \overline{\hat{Var}(\hat{L})}]^2/500 \cdot 499\}^{1/2} = SE(\overline{\hat{Var}})$

(9) *empirical variance of* $\hat{L} = \sum (L_k - \bar{L})^2 / 499 = s^2(\hat{L})$

(10) *standard error of empirical variance of* $\hat{L} = s^2(\hat{L}) \cdot (2/499)^{1/2}$

These same quantities are computed for $\hat{B}(C_\mu)$ as well as for $\hat{L}(C_{AP})$, $\bar{L}(C_\mu)$, $\bar{L}(C_{AP})$ and $\hat{B}(C_{AP})$, $\bar{B}(C_\mu)$, $\bar{B}(C_{AP})$. Notice that m_k^L and m_k^B are the target parameters for the estimators $\hat{L}$ and $\hat{B}$, and $s^2(\hat{L})$ and $s^2(\hat{B})$ are the targets for $\hat{Var}_k(\hat{L})$ and $\hat{Var}_k(\hat{B})$.

4.6 Simulation results.

The results are shown in Table 1; the line numbers in this table correspond to the ten quantities described in the previous section. The three situations corresponding to three bivariate gamma densities are shown in panels (a), (b), (c), comparing side by side the results when comparable case groups are determined via C_{AP} or by C_μ.

In terms of relative error (bias/average), both estimators of average lead time perform similarly, as do the two estimators of average benefit time. (Studies with fewer cases suggest a slight advantage of $\hat{L}$ over $\bar{L}$ and of $\hat{B}$ over $\bar{B}$, not shown here.) Generally, the bias in the estimators is reduced using C_μ to determine comparable case groups; this is particularly true with situations (b) and (c) for estimating L, where C_{AP} incurs a bias of 17-23% but only 0.3-3.2% using C_μ. Reductions in bias occur for estimating B as well, though not as dramatic (5.6-6.5% using C_{AP}; 0.9-1.8% using C_μ). The magnitude of the bias is relatively unchanged in the zero-correlation situation (a).

The variance estimators of $\hat{L}$ and $\hat{B}$, are, on average over the 500 trials, generally about the same or even a bit higher than the empirical variance of the actual 500 trial average lead times in the simulation [compare lines (7) and (9)]. An exception occurs for the zero-correlation situation (a) using C_{AP}, where variance estimates are 4.3%-6.5% too low; the estimates are much closer to their targets when using C_μ. However, the C_μ rule does not bring variance estimates any closer to the simulated empirical variances for situation (c) where the correlation between the preclinical and clinical durations is extremely high (0.9). Even then, the estimates are only about 3% low for $Var(\hat{L})$, but they are 13-14% too low for $Var(\hat{B})$.

In terms of average absolute error in $\hat{L}$ and $\hat{B}$, the C_μ rule always resulted in lower error, sometimes by as much as a factor of 2. The variance of C_{AP} is often as much as 6 times the variance of C_μ for the

same sampling situation, and nominal 95% confidence intervals for $\hat{L}$ and $\hat{B}$ using two estimated standard errors are more nearly attained using C_μ. The number of trials out of 500 whose so-constructed "95% confidence intervals" fail to cover the true L and B, either because they are too high or too low, should in principle be about 12.5 (±3.5); the actual numbers of "failed" trials is shown at the bottom of the table, where it can be seen that coverage is far better using C_μ than C_{AP}.

5. Summary.

We have derived the variances for estimators of average benefit time B and average lead time L in a randomized screening trial and evaluated their performance using a simulation study. Two variance estimators, $\hat{Var}(\hat{B})$ and $\hat{Var}(\hat{L})$, are shown via simulation to capture the within-trial variability [(11) and (17), respectively] in the estimators of average lead time and average benefit time. The simulation also demonstrates successful performance of C_μ to determine a time point at which case groups from the two arms of the trial may be deemed comparable for assessing average screening benefit and average lead time. Moreover, $\hat{B}$ and $\hat{L}$ are successful for estimating these two quantities. The corresponding formulae for the censored case are derived and are appropriate to use when intermediate analysis may be desired. For studies whose outcomes of all participants are known, the uncensored formulas identified here are appropriate and provide satisfactory estimates of both types of variability.

In view of the fact that these variance estimates do not account for the expected variation in the time at which case groups become comparable (i.e., not subject to length bias), their performance for estimating the variances in $\hat{B}$ and $\hat{L}$ is surprisingly good. Further work on this issue across a broader range of sampling situations is being conducted and will be reported elsewhere.

References

[1] Aron, J.L. and Prorok, P.C. (1986) An analysis of the mortality effect in a breast cancer screening study, *International Journal of Epidemiology* **15**, 36-43.

[2] Chen, J., Prorok, P.C., and Graff, K.M. (1983), An age dependent stochastic model of periodic screening: Length bias at a prevalence screen, *Mathematical Biosciences 65,* 93-123.

[3] Connor, R.J. and Prorok, P.C. (1994), Issues in the mortality analyses of randomized controlled trials of cancer screening, *Controlled Clinical Trials* 15, 81-99.

[4] Feller, W.F. (1966), *Introduction to Probability Theory and Its Applications, Vol. II,* Wiley, New York.

[5] Etzioni, R. and Self, S.D. (1995), On the catch-up time method for analyzing cancer screening trials, *Biometrics* **51**, 31-43.

[6] Etzioni, R., Self, S.D., Connor, R.J., Prorok, P.C. (1995), *Controlled Clinical Trials*

[7] Gohagen, J.K., Prorok, P.C., Kramer, B.S., Cornett, J.E. (1994), Prostate cancer screening in the prostate, lung, colorectal, and ovarian cancer screening trial of the National Cancer Institute, *Journal of Urology* **152**, 1905-1909.

[8] Habbema et al. (1986), *Journal of the National Cancer Institute*

[9] Hampel, F.R. (1974), The influence curve and its role in robust estimation. *Journal of the American Statistical Association 69,* 383-393.

[10] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A. (1986), *Robust Statistics: The Approach Based on Influence Functions,* Wiley, New York.

[11] Huber, P. (1981), *Robust Statistics,* Wiley, New York.

[12] Kafadar, K. and Prorok, P.C. (1994), A data-analytic approach for estimating lead time and screening benefit based on survival curves in randomized cancer screening trials, *Statistics in Medicine* 13, 569-586.

[13] Kafadar, K. and Prorok, P.C. (1996a), Computer simulation of randomized screening trials to compare methods of estimating lead time and benefit time. To appear in *Computational Statistics and Data Analysis.*

[14] Kafadar, K. and Prorok, P.C. (1996b), Estimating the Difference in Location Parameters of Two Survival Curves, with Applications to Cancer Screening. To appear in *Journal of Statistical Planning and Inference.*

[15] Lambert, D. (1981), Influence functions for testing, *Journal of the American Statistical Association 75,* 649-657.

[16] Miller, A.B. and Bulbrook, R.D. (1982), Screening, detection, and diagnosis of breast cancer, *The Lancet,* 15 May 1982, 1109-1110.

[17] Morrison, A. (1982), The effects of early treatments, lead time, and length bias on the mortality experienced by cases detected by screening, *International Journal of Epidemiology 11,* 261-267.

[18] Morrison, A. (1992), *Screening for Chronic Diseases,* 2nd ed., Oxford University Press, Oxford.

[19] Prorok, P.C. (1984), Evaluation of screening programs for the early detection of cancer, Chapter 7 in *Statistical Methods for Cancer Studies,* R.G. Cornell, Ed., Marcel Dekker, New York, 267-328.

[20] Prorok, P.C. and Connor, R.J. (1986), Screening for the early detection of cancer. *Cancer Investigation 4(3),* 225-238.

[21] Reid, N. (1981), Influence functions for censored data, *Annals of Statistics 9,* 78-92.

[22] Shapiro, S., Venet, W., Strax, P., Venet, L. (1988), *Periodic Screening for Breast Cancer: The Health Insurance Plan Project and its Sequelae, 1963-1986,* Johns Hopkins University Press: Baltimore.

[23] Statistical Sciences Inc. (1991), *S-Plus User's Manual, Version 3.0.* Seattle, Washington.

Table 1

Estimates of Average Benefit Time and Average Lead Time and their Variances
Three sampling situations (§4.2) and two rules for determining comparable case groups(C_{AP}, C_{μ})

(a) $\mu_P = 2, \sigma_P^2 = 1, \mu_D = 4, \sigma_D^2 = 4, \text{corr}(P, D) = 0.3$

		C_{AP}			C_{μ}	
	True L	$\hat{L}$	$\bar{L}$	**True L**	$\hat{L}$	$\bar{L}$
(1) True average	**1.3762**			**1.4279**		
(2) Standard error of (1)	0.0075			0.0212		
(3) Average of estimator		1.6128	1.6262		1.3820	1.3914
(4) Standard error of (3)		0.0346	0.0349		0.0209	0.0208
(5) Average bias		0.2366	0.2500		-0.0458	-0.0365
(6) Standard error of (5)		0.0213	0.0212		0.0184	0.0184
(7) Average sample variance	**0.6633**			**0.2790**		
(8) Standard error of (7)	0.0360			0.0184		
(9) Actual empirical variance		0.5991	0.6112		0.2184	0.2161
(10) Standard error of (9)		0.0379	0.0387		0.0138	0.0137
No. intervals too high/low		18,52	18,51		9,15	10,14

		C_{AP}			C_{μ}	
	True B	$\hat{B}$	$\bar{B}$	**True B**	$\hat{B}$	$\bar{B}$
(1) True average	**3.7680**			**3.8788**		
(2) Standard error of (1)	0.0148			0.0067		
(3) Average of estimator		3.5579	3.5549		3.9486	3.9489
(4) Standard error of (3)		0.0360	0.0360		0.0274	0.0274
(5) Average bias		-0.2101	-0.2131		0.0698	0.0701
(6) Standard error of (5)		0.0365	0.0367		0.0257	0.0257
(7) Average sample variance	**0.7093**			**0.3874**		
(8) Standard error of (7)	0.0294			0.0151		
(9) Actual empirical variance		0.6482	0.6495		0.3755	0.3755
(10) Standard error of (9)		0.0410	0.0411		0.0238	0.0238
No. intervals too high/low		15,15	14,13		10,4	10,4

(b) $\mu_P = 2$, $\sigma_P^2 = 4$, $\mu_D = 5$, $\sigma_D^2 = 25$, corr(P, D) = 0.0

		C_{AP}			C_μ	
	True L	$\hat{L}$	$\bar{L}$	**True L**	$\hat{L}$	$\bar{L}$
(1) True average	**1.9661**			**1.9987**		
(2) Standard error of (1)	0.0099			0.0085		
(3) Average of estimator		2.2918	2.2783		1.6678	1.6455
(4) Standard error of (3)		0.0485	0.0486		0.0236	0.0239
(5) Average bias		0.3257	0.3122		-0.3310	-0.3532
(6) Standard error of (5)		0.0473	0.0474		0.0248	0.0252
(7) Average sample variance	**1.1020**			**0.2674**		
(8) Standard error of (7)	0.0497			0.0124		
(9) Actual empirical variance		1.1758	1.1803		0.2792	0.2863
(10) Standard error of (9)		0.0744	0.0747		0.0177	0.0181
No. intervals too high/low		19,70	20,71	3,80	3,85	

		C_{AP}			C_μ	
	True B	$\hat{B}$	$\bar{B}$	**True B**	$\hat{B}$	$\bar{B}$
(1) True average	**3.8770**			**3.9300**		
(2) Standard error of (1)	0.0120			0.0081		
(3) Average of estimator		3.5375	3.5353		4.2936	4.2819
(4) Standard error of (3)		0.0672	0.0679		0.0484	0.0489
(5) Average bias		-0.3395	-0.3513		0.3636	0.3519
(6) Standard error of (5)		0.0679	0.0687		0.0470	0.0475
(7) Average sample variance	**2.1583**			**1.1965**		
(8) Standard error of (7)	0.0598			0.0164		
(9) Actual empirical variance		1.1758	1.1803		1.1712	1.1947
(10) Standard error of (9)		0.0744	0.0747		0.0742	0.0756
No. intervals too high/low		12,25	13,25	20,6	21,6	

(c) $\mu_P = 2$, $\sigma_P^2 = 1$, $\mu_D = 2$, $\sigma_D^2 = 1$, corr(P, D) = 0.9

		C_{AP}			C_μ	
	True L	$\hat{L}$	$\bar{L}$	**True L**	$\hat{L}$	$\bar{L}$
(1) True average	**1.3500**			**1.3983**		
(2) Standard error of (1)	0.0073			0.0041		
(3) Average of estimator		1.6357	1.6635		1.3695	1.3936
(4) Standard error of (3)		0.0362	0.0367		0.0234	0.0234
(5) Average bias		0.2857	0.3135		-0.0288	-0.0048
(6) Standard error of (5)		0.0340	0.0345		0.0239	0.0239
(7) Average sample variance	**0.6789**			**0.2656**		
(8) Standard error of (7)	0.0358			0.0162		
(9) Actual empirical variance		0.6553	0.6729		0.2742	0.2735
(10) Standard error of (9)		0.0415	0.0426		0.0174	0.0173
No. intervals too high/low		18,56	20,55	12,26	14,24	

		C_{AP}			C_μ	
	True B	$\hat{B}$	$\bar{B}$	**True B**	$\hat{B}$	$\bar{B}$
(1) True average	**3.7315**			**3.8326**		
(2) Standard error of (1)	0.0140			0.0069		
(3) Average of estimator		3.4900	3.4909		3.8667	3.8731
(4) Standard error of (3)		0.0356	0.0357		0.0270	0.0268
(5) Average bias		-0.2415	-0.2406		0.0341	0.0405
(6) Standard error of (5)		0.0368	0.0369		0.0248	0.0246
(7) Average sample variance	**0.6408**			**0.3111**		
(8) Standard error of (7)	0.0275			0.0123		
(9) Actual empirical variance		0.6354	0.6381		0.3637	0.3591
(10) Standard error of (9)		0.0402	0.0404		0.0230	0.0227
No. intervals too high/low		15,20	14,20	15,10	15,10	

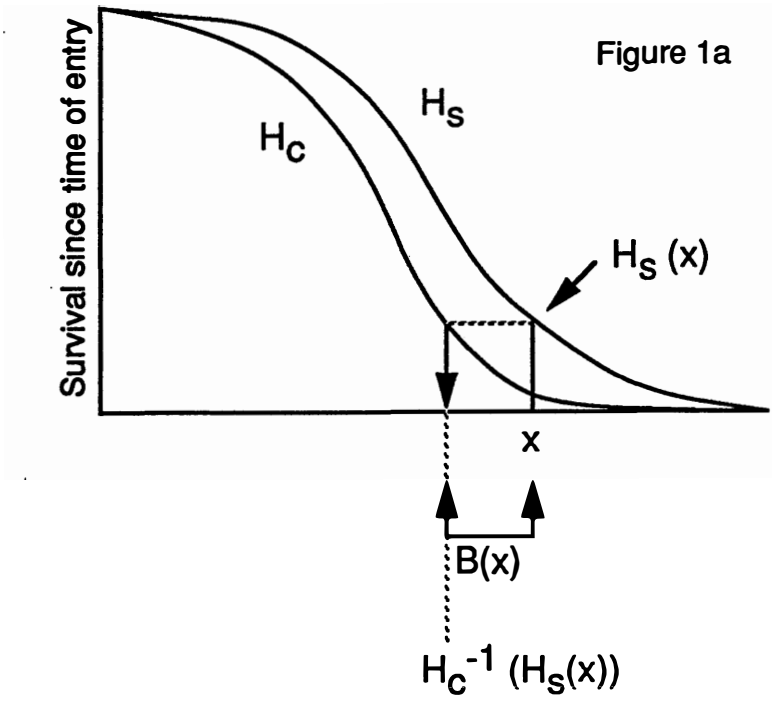

Figure 1a

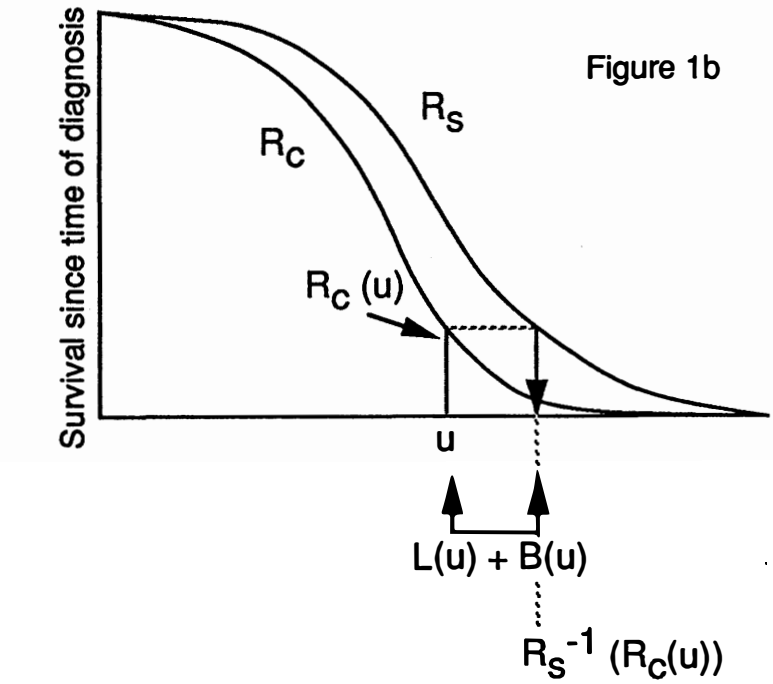

Figure 1b

CENTER FOR COMPUTATIONAL MATHEMATICS REPORTS

University of Colorado at Denver
P.O. Box 173364, Campus Box 170
Denver, CO 80217-3364

Fax: (303) 556-8550
Phone: (303) 556-8442
http://www-math.cudenver.edu/

69. P. Vaněk and S. Ghosal, "A New Technique for Construction of Image Pyramids."
70. L.P. Franca and A. Russo, "Mass Lumping Emanating from Residual-Free Bubbles."
71. L.P. Franca and A. Russo, "Unlocking with Residual-Free Bubbles."
72. C. Liu, Z. Liu and G. Xiong, "Direct Numerical Simulation for the Receptivity and the Whole Process of Transition Around 2-D Airfoils."
73. R. Tezaur, P. Vaněk and M. Brezina, "Two-Level Method for Solids on Unstructured Meshes."
74. T.F. Russell, D.W. Dean, T.H. Illangasekare, R. Mapa and J. Garcia, "Upscaling of Dispersivity in Modeling of Solute Transport: Mathematical Theory and Laboratory Experiments."
75. T.F. Russell, R.W. Healy, R.G. Striegl, G.L. Hutchinson and G.P. Livingston, "Analytical Solution for the Problem of 1-Dimensional Diffusion into a Static Chamber."
76. T.F. Russell and R.V. Trujillo, "The Finite Volume Element Method For Elliptic and Parabolic Equations."
77. R.W. Healy and T.F. Russell, "Solution of the Advection-Dispersion Equation in Two Dimensions by a Finite-Volume Eulerian-Lagrangian Localized Adjoint Method."
78. Z. Cai, R.R. Parashkevov, T.F. Russell and X. Ye, "Domain Decomposition for a Mixed Finite Element Method in Three Dimensions."
79. S.E. Payne, T. Pentilla and G.F. Royle, "Building a Cyclic q-Clan."
80. K.D. Jamison and W.A. Lodwick, "Minimizing Unconstrained Fuzzy Functions."
81. F. Brezzi, L.P. Franca, T.J.R. Hughes and A. Russo, "$b = \int g$."
82. L.P. Franca, C. Farhat, M. Lesoinne and A. Russo, "Unusual Stabilized Finite Element Methods and Residual-Free-Bubbles."
83. F. Brezzi, L.P. Franca, T.J.R. Hughes and A. Russo, "Stabilization Techniques and Subgrid Scales Capturing."
84. J. Mandel, R. Tezaur and C. Farhat, "A Scalable Substructuring Method by Lagrange Multipliers for Plate Bending Problems."