Weighted Least Squares Estimation for Aalen's Additive Risk Model

FRED W. HUFFER and IAN W. MCKEAGUE*

Cox's proportional hazards model has so far been the most popular model for the regression analysis of censored survival data. However, the additive risk model of Aalen can provide a useful and biologically more plausible alternative. Aalen's model stipulates that the conditional hazard function for a subject, whose covariates are $Y = (Y_1, \ldots, Y_p)'$, has the form $h(t \mid Y) = Y'\alpha(t)$, where $\alpha = (\alpha_1, \ldots, \alpha_p)'$ is an unknown vector of hazard functions. This article discusses inference for α based on a weighted least squares (WLS) estimator of the vector of cumulative hazard functions. The asymptotic distribution of the WLS estimator is derived and used to obtain confidence intervals and bands for the cumulative hazard functions. Both a grouped data and a continuous data version of the estimator are examined. An extensive simulation study is carried out. The method is applied to grouped data on the incidence of cancer mortality among Japanese atomic bomb survivors.

KEY WORDS: Atomic bomb survivors; Grouped survival data; Martingale methods; MonteCarlo; Multivariate counting processes; Nonproportional hazards.

1. INTRODUCTION

In this article we study Aalen's (1980) additive risk model for the regression analysis of censored survival data. Let $h_i(t) = h_i(t \mid Y_i)$ denote the conditional hazard function at time t for subject i whose covariates are given by the p vector $Y_i = (Y_{i1}, \ldots, Y_{ip})'$. Aalen's model stipulates that

$$h_i(t) = \sum_{j=1}^{p} \alpha_j(t) Y_{ij} = Y_i'\alpha(t), \tag{1.1}$$

where $\alpha = (\alpha_1, \ldots, \alpha_p)'$ is an unknown vector of hazard functions.

The additive risk model provides a useful alternative to Cox's (1972) proportional hazards model. Since temporal effects are not assumed to be proportional for each covariate, Aalen's model is capable of providing detailed information concerning the temporal influence of each covariate not available from Cox's model. In studies of *excess* risk for instance, where the background risk and excess risk typically can have very different temporal forms, additive risk models seem to be biologically more plausible than proportional hazard models (see Buckley 1984). For example, the latent period for the risk of cancer following exposure to low doses of ionizing radiation can be better understood in terms of an additive risk model. Moreover, the use of the proportional hazards model when the true model is additive risk has been found by O'Neill (1986) to result in serious asymptotic bias.

Although parametric additive risk models have been widely applied in survival analysis (especially in epidemiology) for many years (see the references in Breslow 1986; Muirhead and Darby 1987), the nonparametric additive risk model (1.1) has only been studied recently (Aalen 1980, 1989; McKeague 1986, 1988a, b; Huffer and McKeague 1987; Mau 1986, 1988).

Aalen (1980) introduced estimators for the vector of cumulative hazard functions, $A(t) = \int_0^t \alpha(s)ds$, that use continuous data (containing the exact values of failure and censoring times). These estimators generalize the well-known Nelson–Aalen estimator, the natural estimator in the case of one covariate. One possible form of these estimators was motivated by a formal least squares principle. This estimator, defined precisely in subsection (2.2), is referred to here as *Aalen's least squares* estimator. Aalen observed that this estimator probably gives reasonable estimates, and he applied it to analysis of data from the Veterans Administration Lung Cancer Study Group.

The purpose of this article is to introduce a *weighted* least squares estimator for A. Two versions of this estimator will be examined: a grouped data version and a continuous data version. By *grouped* data we mean that only the person-years at risk and number of uncensored deaths over successive time intervals, for various levels of the covariates, are available. This type of data arises in studies involving the follow-up of large population groups over many years (see Breslow 1986). Even when the continuous data is available, the grouped data version of the estimator is useful because it is often desirable to compress the data by grouping, to facilitate computer implementation.

Note that, unlike the Cox model, inference for the additive risk model is complicated by the constraint that the hazard function (1.1) be nonnegative. This is not a weakness of the additive risk model itself—there is no constraint on any particular α_j, only on the overall hazard. It simply means that we should be careful that our preliminary estimate of the hazard (1.1), used in the computation of the weighted least squares estimator, is nonnegative.

In subsections 2.1 and 2.2 we describe the various es-

* Fred W. Huffer is Associate Professor and Ian W. McKeague is Associate Professor, Department of Statistics, Florida State University, Tallahassee, FL 32306-3033. Fred W. Huffer's research was supported by the Office of Naval Research under Grant N00014-86-K-0156. Ian W. McKeague's research was supported by the Army Research Office under Grant DAAL03-86-K-0094. The authors are grateful to the associate editor and a referee for their detailed comments. We also thank Richard Gill for a helpful discussion concerning the WLS estimator at the Oberwolfach meeting on martingale methods in statistics in December 1988, and David Hoel for telling us how to procure the data used in Section 4. These data were supplied by RERF, which is a private foundation that is funded equally by the Japanese Ministry of Health and Welfare and the U.S. Department of Energy through the U.S. National Academy of Sciences. The conclusions reached in this article are those of the authors and do not necessarily reflect the scientific judgment of RERF or its funding agencies.

Journal of the American Statistical Association
March 1991, Vol. 86, No. 413, Theory and Methods

timators. Confidence intervals and bands for the cumulative hazard functions are given in subsection 2.3, and tests for the presence of a covariate effect are discussed in subsections 2.4 and 2.5. The results of a simulation study are reported in Section 3. In Section 4 we apply the additive risk model to the analysis of grouped data on the incidence of cancer mortality among Japanese atomic bomb survivors. The asymptotic distribution of the weighted least squares estimator is derived in Section 5.

2. ESTIMATORS AND CONFIDENCE BANDS

Suppose that the ith individual's failure time U_i is an absolutely continuous random variable conditionally independent of a (right) censoring time V_i given the covariate vector Y_i. Let $X_i = \min(U_i, V_i)$ and $\delta_i = I(U_i \leq V_i)$ denote the time to the end-point event and the indicator for noncensorship respectively. The follow-up period is taken as $[0, T]$. Assume that the triples (X_i, δ_i, Y_i), for $i = 1, \ldots, n$, are iid, and the conditional hazard function $h_i(t)$ of U_i given Y_i satisfies the model (1.1).

2.1 The Grouped Data Case

Let $\mathcal{I}_1, \ldots, \mathcal{I}_d$ denote successive calendar periods of lengths $\ell_1, \ldots, \ell_d$ used in the grouping of the data. These intervals partition the follow-up period $[0, T]$. The standard approach to the estimation of α is to assume that it is constant within each interval. This yields a parametric (piecewise exponential) model with dp parameters that can be estimated using maximum likelihood. An auxiliary Poisson regression model is then invoked to calculate the maximum likelihood estimates (MLE's) via iteratively reweighted least squares or iterative proportional-fitting algorithms. This approach is justified since the true likelihood can be shown to be proportional to the likelihood for the Poisson regression model (see Laird and Olivier 1981).

When the data is grouped into sufficiently many calendar periods ($d \geq 8$, say), it is possible to use the parametric analysis as the basis for a nonparametric approach. The idea is to estimate α by regarding it as piecewise constant over each interval, but then *allow the partition to become finer* as the sample size increases, in the fashion of Grenander's (1981) method of sieves. Inference is carried out on the basis of an asymptotic result for an estimator $\hat{A}$ of A obtained by integrating a weighted least squares (WLS) estimator of α. To develop a satisfactory asymptotic theory, McKeague (1988b) found it necessary to slightly modify the standard approach by using a WLS estimator with *predictable* weights. The advantage of having predictable weights—depending only on events taking place over chronologically previous time intervals—is that martingale techniques become applicable.

We now proceed to define the various estimators. Let $T_{ir} = \int_{\mathcal{I}_r} I(X_i \geq t)\, dt$ be the total time that the ith individual is observed to be at risk in interval $\mathcal{I}_r$. Also, let $\delta_{ir} = \delta_i I(X_i \in \mathcal{I}_r)$ be the indicator that the ith individual undergoes an uncensored failure in $\mathcal{I}_r$. The ordinary least squares (OLS) estimator of α is defined by

$$\tilde{\alpha}(t) = \tilde{D}_r^{-1}\tilde{C}_r, \qquad \text{for } t \in \mathcal{I}_r,$$

where

$$\tilde{C}_r = \sum_{i=1}^{n} \delta_{ir} Y_i \quad \text{and} \quad \tilde{D}_r = \sum_{i=1}^{n} (Y_i Y_i')T_{ir}.$$

Here (and in the sequel), for any square matrix K, K^{-1} denotes the inverse of K if K is invertible, the zero matrix otherwise. Note that $\tilde{\alpha}$ is the standard OLS estimator based on grouped data consisting of the total time at risk and the number of uncensored failures in each interval $\mathcal{I}_1, \ldots, \mathcal{I}_d$ tabulated for all realized levels of the covariates.

The standard WLS estimate of α (coinciding with the MLE) is calculated using the following iteratively reweighted least squares algorithm. An initial WLS estimate is given by

$$\hat{\alpha}(t) = \hat{D}_r^{-1}\hat{C}_r, \qquad \text{for } t \in \mathcal{I}_r, \tag{2.1}$$

where

$$\hat{C}_r = \sum_{i=1}^{n} \delta_{ir} Y_i \hat{W}_{ir}, \qquad \hat{D}_r = \sum_{i=1}^{n} (Y_i Y_i')T_{ir}\hat{W}_{ir}, \tag{2.2}$$

and $\hat{W}_{ir} = 1/Y_i'\tilde{\alpha}(t)$ for $t \in \mathcal{I}_r$. The weight $\hat{W}_{ir}$ is an estimate of the "true" weight $1/h_i(t)$. An updated estimate of this quantity is obtained by using $\hat{\alpha}(t)$ in place of $\tilde{\alpha}(t)$ in $\hat{W}_{ir}$. This in turn leads to an updated WLS estimate by recalculating (2.2) and (2.1). The procedure is iterated, alternating between updating the weight and updating the WLS estimate.

The modification proposed here is to require the weights to be predictable. Instead of using the OLS estimator $\tilde{\alpha}$ to calculate the weight for the initial WLS estimate, we use a uniformly consistent and predictable estimator α^* of α. We take α^* to be a (piecewise constant) predictable version of the OLS estimator $\tilde{\alpha}$, defined on each interval $\mathcal{I}_r$ by smoothing $\tilde{\alpha}$ over the previous intervals $\mathcal{I}_1, \ldots, \mathcal{I}_{r-1}$. The weight is given by $\hat{W}_{ir} = 1/Y_i'\alpha^*(t)$ for $t \in \mathcal{I}_r$.

The OLS estimator itself is neither predictable nor uniformly consistent, hence the need for smoothing it over the past. Our modified WLS estimator could be updated in a similar way to the standard WLS estimator, but for simplicity we shall not do so here. The asymptotic theory remains the same whether further updating is done or not. Note that the WLS estimator can be evaluated from the grouped data. Integrating $\hat{\alpha}$, we obtain a WLS estimator of A:

$$\hat{A}(t) = \int_0^t \hat{\alpha}(s)\, ds.$$

Although a more sophisticated version of α^* is possible [such as a kernel estimator, cf. (2.3) in the continuous case], for most applications it seems adequate to take α^* to be the average of the OLS estimator over the previous n_S intervals, where n_S is chosen appropriately. This amounts to using a flat kernel to smooth the OLS estimator over a segment of the past. For the first n_S intervals we have set the weights $\hat{W}_{ir}$ equal to 1, which is equivalent to using the OLS estimator.

The OLS, rather than the WLS estimator of α, should be used in interval $\mathcal{I}_r$ if any of the weights $\hat{W}_{ir}$ involved in $\hat{D}_r$ 'blow up' or become negative; this will happen if all components of α^* are too close to zero or there are insufficient uncensored failures in the segment of n_S intervals over which the smoothing is carried out. In choosing n_S it is important to take into account the variance/bias tradeoff. If the length of the segment is too large, then the estimates of the weights might be biased; if the number of uncensored failures in the segment is too small, then estimates of the weights will have large variance.

2.2 The Continuous Data Case

The continuous data consist of the observations (X_i, δ_i, Y_i), for $i = 1, \ldots, n$. Our notation will parallel that of the grouped data case as much as possible. Aalen's OLS estimator is given by

$$\tilde{A}(t) = \sum_{\substack{\delta_i=1 \\ X_i \le t}} \tilde{D}_i^{-1} Y_i,$$

where the summation is over individuals i whose failure times are uncensored and less than or equal to t and $\tilde{D}_i$ is the $p \times p$ matrix

$$\tilde{D}_i = \sum_{X_k \ge X_i} Y_k Y_k',$$

where the summation is over individuals k at risk at time X_i.

Our WLS estimator of A is given by

$$\hat{A}(t) = \sum_{\substack{\delta_i=1 \\ X_i \le t}} \hat{D}_i^{-1} Y_i \hat{W}_{ii},$$

where

$$\hat{D}_i = \sum_{X_k \ge X_i} Y_k Y_k' \hat{W}_{ki}$$

and $\hat{W}_{ki}$ is an estimate of $1/h_k(t)$ evaluated at time $t = X_i$. We take $\hat{W}_{ki} = 1/Y_k'\alpha^*(X_i)$, where, as in the grouped data case, α^* is a predictable and uniformly consistent estimator of α. The estimator α^* is taken to be the *smoothed* OLS estimator

$$\alpha^*(t) = \frac{1}{b_n} \int_0^T K\left(\frac{t-s}{b_n}\right) d\tilde{A}(s), \qquad (2.3)$$

where K is a left-continuous kernel function of bounded variation, having integral 1, support $(\epsilon, 1]$ for some $0 < \epsilon < 1$, and $b_n > 0$ is a bandwidth parameter.

In practice, the WLS estimator is insensitive to the choice of kernel function, but sensitive to b_n. For simplicity, we take the kernel to be constant over $(\epsilon, 1]$, so $\alpha^*(t)$ can be easily calculated in terms of an increment of $\tilde{A}$ over a past segment of time. The length b_n of this segment plays an analogous role to n_S in the grouped data case. The increment in the WLS estimator should be replaced by the increment in the OLS estimator at time $t = X_i$ if any of the weights $\hat{W}_{ki}$ involved in $\hat{D}_i$ 'blow up' or become negative. In practice, b_n might be chosen in an adaptive fashion, letting it depend on the number of uncensored failures in the recent past.

2.3 Confidence Intervals and Bands

In this subsection we show how to construct confidence bands and confidence intervals for A_j based on the asymptotic result, given in Section 5, that $n^{1/2}(\hat{A} - A)$ converges in distribution to a p-variate Gaussian martingale with variance function $\int_0^{\cdot} \mathbf{V}^{-1} ds$, where $\mathbf{V}$ is a certain $p \times p$ matrix-valued function of time.

First consider the continuous case. Let Δ_i be the p-vector of jumps in the components of the WLS estimator at time X_i, that is, $\Delta_i = \hat{D}_i^{-1} Y_i \hat{W}_{ii}$, and define the $p \times p$ matrix-valued process

$$\hat{G}(t) = \sum_{\substack{\delta_i=1 \\ X_i \le t}} \Delta_i \Delta_i'$$

(cf. Aalen 1980, p. 6). We show in Section 5 that $n\hat{G}$ is a uniformly consistent estimator for the asymptotic variance function of $n^{1/2}(\hat{A} - A)$. The estimator $\hat{G}$ will be called WLS-1. A grouped data version is given by $\hat{G}(t) = \int_0^t \hat{g}(s)\, ds$, where

$$\hat{g}(t) = \hat{D}_r^{-1} \hat{Q}_r \hat{D}_r^{-1}, \quad \text{for} \quad t \in \mathcal{I}_r$$

and $\hat{Q}_r = \ell_r \sum_{i=1}^n Y_i Y_i' \hat{W}_{ir}^2 \delta_{ir}$.

An asymptotic $100(1 - \beta)\%$ confidence *interval* for $A_j(t)$, at fixed $t \in [0, T]$, is given by $\hat{A}_j(t) \pm z_{\beta/2} \hat{G}_{jj}(t)^{1/2}$, where $z_{\beta/2}$ is the upper $\beta/2$ quantile for the standard normal distribution.

Using a transformation to the Brownian bridge process B^0 (see Andersen and Borgan 1985, p. 114), we obtain the following asymptotic $100(1 - \beta)\%$ confidence *band* for A_j:

$$\hat{A}_j(t) \pm c_\beta \hat{G}_{jj}(T)^{1/2} \left(1 + \frac{\hat{G}_{jj}(t)}{\hat{G}_{jj}(T)}\right), \quad t \in [0, T], \qquad (2.4)$$

where c_β is the upper β quantile of the distribution of $\sup_{t \in [0,1/2]} |B^0(t)|$. Tables for c_β can be found in Hall and Wellner (1980).

There are two alternative estimators that, from an asymptotic point of view, might be used equally well in place of WLS-1. These estimators will be called WLS-2 and WLS-3, respectively. The WLS-2 estimator is given by

$$\hat{G}(t) = \int_0^t \hat{g}(s)\, ds, \qquad (2.5)$$

where, in the grouped case, $\hat{g}(t) = \ell_r \hat{D}_r^{-1}$ for $t \in \mathcal{I}_r$, and, in the continuous case, $\hat{g}(t) = \hat{D}_i^{-1}$ for $X_{i-1} < t \le X_i$.

The grouped WLS-3 estimator is given by (2.5), where

$$\hat{g}(t) = \ell_r \hat{D}_r^{-1} \hat{H}_r \hat{D}_r^{-1} \quad \text{and} \quad \hat{H}_r = \sum_{i=1}^n Y_i Y_i' \hat{W}_{ir}^2 T_{ir}(Y_i' \hat{\alpha}(t))$$

for $t \in \mathcal{I}_r$. The continuous WLS-3 estimator is given by

$$\hat{G}(t) = \sum_{\substack{\delta_i=1 \\ X_i \le t}} \hat{D}_i^{-1} \hat{H}_i \hat{D}_i^{-1},$$

where $\hat{H}_i = \sum_{X_k \ge X_i} Y_k Y_k' \hat{W}_{ki}^2 (Y_k' \Delta_i)$.

Recall that if any weight 'blows up' or becomes negative, then the jump in $\hat{A}$ is taken as the jump in the OLS estimate. When this happens, the increments in $\hat{G}$ should also be OLS-based. That is, using the same idea as in the WLS-3 estimate above, the jump in $\hat{G}$ at time X_i is taken as $\tilde{D}_i^{-1}\tilde{H}_i\tilde{D}_i^{-1}$, where $\tilde{H}_i = \Sigma_{X_k \geq X_i} Y_k Y_k'(Y_k'\tilde{\Delta}_i)$ and $\tilde{\Delta}_i$ is the jump in the OLS estimator at time X_i. There is a similar modification in the grouped data case. This is the procedure used in Sections 3 and 4.

Note that, unlike the confidence intervals, the confidence bands given above are dependent on the choice of T. As T increases the band widens at all points. It also becomes more unreliable since the effective sample size $n_T = \#\{i : X_i \geq T\}$ (the size of the risk set) for estimating the asymptotic variance at time T is getting smaller. In practice, we found it reasonable to set T so that n_T is at least 10% of the sample size. This is the case for the simulated data in Section 3 and for the atomic bomb survivor data in Section 4.

2.4 Testing for The Presence of a Covariate Effect

It is frequently of interest to find out whether a particular covariate has any effect on the overall hazard function, so we need a test of the null hypothesis

$$H_0 : \alpha_j(t) = 0, \qquad \text{for all} \quad t \in [0, T].$$

Aalen (1980, 1989) considered test statistics of the form $\int_0^T L_j(s) d\tilde{A}_j$, where L_j is a nonnegative predictable "weight" function. It would, of course, be possible to use the WLS estimator $\hat{A}_j$ in place of the OLS estimator $\tilde{A}_j$ here.

An alternative approach is to use a Kolmogorov–Smirnov type statistic based on $\hat{A}_j$. Employing the same transformation to the Brownian bridge used to obtain (2.4), it can be seen that under H_0 the process

$$\xi(\cdot) = \hat{A}_j(\cdot)\hat{G}_{jj}(T)^{1/2}(\hat{G}_{jj}(\cdot) + \hat{G}_{jj}(T))^{-1}$$

converges in distribution to the time-changed Brownian bridge $B^0(\tau(\cdot))$, where τ is an increasing continuous function on $[0, T]$ with $\tau(0) = 0$ and $\tau(T) = 1/2$. The test statistic $\sup_{t\in[0,T]}|\xi(t)|$ converges weakly to $\sup_{t\in[0,1/2]}|B^0(t)|$. Refer to tables in Hall and Wellner (1980) for its limiting distribution.

When dealing with a covariate that is known to cause *excess* risk (i.e., α_j is nonnegative), then the one-sided statistic $\sup_{t\in[0,T]}\xi(t)$ would be more appropriate. Tables for its limiting distribution, that of $\sup_{t\in[0,1/2]}B^0(t)$, can also be found in Hall and Wellner (1980).

2.5 Assessment of Excess Risk

The cumulative hazard function A_j, corresponding to the effect of the jth covariate, relates to the chance of surviving exposure to age t in the absence of other causes of death. In practical applications (assessing the public health risk of a certain carcinogen for instance), however, it may be more important to estimate the integral of the excess hazard multiplied by the chance of surviving all causes of death to that age. When the excess hazard is small compared to background, then this integral is the approximate probability of death due to the carcinogen (per unit of exposure).

Let $\pi(t)$ be a (known) function representing a population-based estimate of the probability of survival to age t. We would like to estimate $A^\pi = \int_0^{\cdot} \pi\, dA$. The natural estimator is $\hat{A}^\pi = \int_0^{\cdot} \pi\, d\hat{A}$. It is routine to modify the above approach to deal with this estimator. For example, $n^{1/2}(\hat{A}^\pi - A^\pi)$ converges in distribution to the p-variate Gaussian martingale with predictable covariation process $\int_0^{\cdot} \pi^2\mathbf{V}^{-1}\, ds$.

3. SIMULATION RESULTS

In order to evaluate the performance of the confidence intervals and bands, a Monte-Carlo experiment was performed to see whether their asymptotic properties take effect under sample sizes, grouping, and censoring found in typical applications. We carried out simulations for both continuous and grouped data.

We used two different simulation models, each with $p = 2$ covariates:

1. $\alpha_1(t) = 1$, $\alpha_2(t) = t$, iid uniform covariates on the lattice $\{r/8,\ r = 1, \ldots, 8\}$.
2. $\alpha_1(t) = \alpha_2(t) = 1$, iid exponential (mean 1/2) covariates truncated at the 1% and 99% points.

For each model, the censoring time was independent of the failure time and exponentially distributed with parameter γ (mean $= 1/\gamma$), for various values of γ. The follow-up period was [0, 1]. The results are displayed in Tables 1 and 2.

Table 1 contains observed coverage probabilities for 95% confidence bands for Model 1, in which the cumulative hazard functions are $A_1(t) = t$ and $A_2(t) = t^2/2$. The censoring parameter γ was set to .3 and 1.5, amounting to 28% and 68% censoring prior to the end of follow-up. With $\gamma = .3$ ($\gamma = 1.5$) there were on average 33% (10%) surviving uncensored beyond the end of follow-up. In the grouped data case (Table 1), the intervals $\mathcal{I}_1, \ldots, \mathcal{I}_d$ were taken to be of equal length. The number of intervals d was set to 8, 16, and 64, with n_S taken as 1, 2, and 4 respectively. In the continuous data case (Table 1) the smoothing was carried out over time segments of length $b_n = 1/8$, matching the grouped data smoothing when $d = 8$ and 16.

From Table 1 it appears that the bands for the continuous data case have coverage probabilities close to their nominal value of .95, except for the WLS-2 method, where they are significantly less than .95. In the grouped data case, the WLS-2 values are close to .95, whereas the other methods appear to be conservative. Note that as d increases, the coverage probabilities decrease and get closer to .95 for both covariates under both light (28%) and heavy (68%) censoring, irrespective of sample size. This is to be expected since the grouped data estimators and bands are piecewise linear, so the Brownian bridge approximation tends to overestimate the probability of escape from a band, the effect becoming less pronounced as d increases.

Table 2, based on simulation Model 2 with $d = 8$, gives coverage probabilities for the pointwise confidence intervals and mean ratios (OLS/WLS) of the widths of the confidence intervals at the right endpoints of $\mathcal{I}_1, \ldots, \mathcal{I}_8$. The censoring parameter γ was set to .3, which gave 26% censoring prior to the end of follow-up.

For both continuous and grouped cases, the OLS, WLS-1 and WLS-3 confidence interval coverage probabilities appear to be close to their nominal value of .95, but WLS-2

Table 1. Observed Coverage Probabilities of 95% Confidence Bands; Model (I)

Continuous data case; $n = 1,000$, $b_n = 1/8$								
	28% Censoring[a]				68% Censoring[a]			
Cov	OLS	WLS-1	WLS-2	WLS-3	OLS	WLS-1	WLS-2	WLS-3
1	.9546	.9480	.9341	.9499	.9560	.9459	.9273	.9504
2	.9557	.9427	.9260	.9473	.9590	.9403	.9223	.9468

Grouped data case; $n = 1,000$									
		28% Censoring[a]				68% Censoring			
Cov	d	OLS	WLS-1	WLS-2	WLS-3	OLS	WLS-1	WLS-2	WLS-3
1	8	.9747	.9733	.9650	.9739	.9776	.9721	.9657	.9740
	16	.9689	.9685	.9595	.9696	.9716	.9658	.9592	.9670
	64	.9604	.9572	.9489	.9579	.9593	.9589	.9509	.9586
2	8	.9763	.9731	.9604	.9735	.9775	.9734	.9650	.9747
	16	.9711	.9672	.9556	.9686	.9727	.9646	.9565	.9669
	64	.9625	.9585	.9512	.9582	.9625	.9583	.9540	.9584

Grouped data case; $n = 4,000$									
		28% Censoring				68% Censoring			
Cov	d	OLS	WLS-1	WLS-2	WLS-3	OLS	WLS-1	WLS-2	WLS-3
1	8	.9750	.9749	.9668	.9756	.9783	.9780	.9699	.9784
	16	.9706	.9677	.9621	.9684	.9727	.9705	.9637	.9711
	64	.9607	.9597	.9535	.9603	.9621	.9589	.9522	.9594
2	8	.9762	.9778	.9654	.9779	.9809	.9788	.9675	.9795
	16	.9713	.9733	.9632	.9730	.9758	.9732	.9629	.9745
	64	.9632	.9647	.9576	.9645	.9656	.9599	.9539	.9623

NOTE: The data were generated using the uniform random number generator of Marsaglia, Zaman, and Tsang (1990). Runs marked [a] have the same initial seed. The number of samples in each run was 10,000.

is slightly less than .95. Thus although WLS-2 does better than the other WLS methods for the grouped data *bands*, it does worse for the confidence *intervals*. We found this defect of WLS-2 even more pronounced in Model 1. We suspect that WLS-2 is underestimating the variance. This counteracts the conservative effect of the Brownian bridge approximation for the bands, but there is no similar cancellation for the confidence intervals, so they turn out to have coverage probabilities less than .95.

The columns of ratios in Table 2 indicate that the WLS

Table 2. Observed Coverage Probabilities of 95% Confidence Intervals and Mean Ratios of CI Widths OLS/WLS; Model (II), 26% Censoring, $n = 1000$, Covariate 1

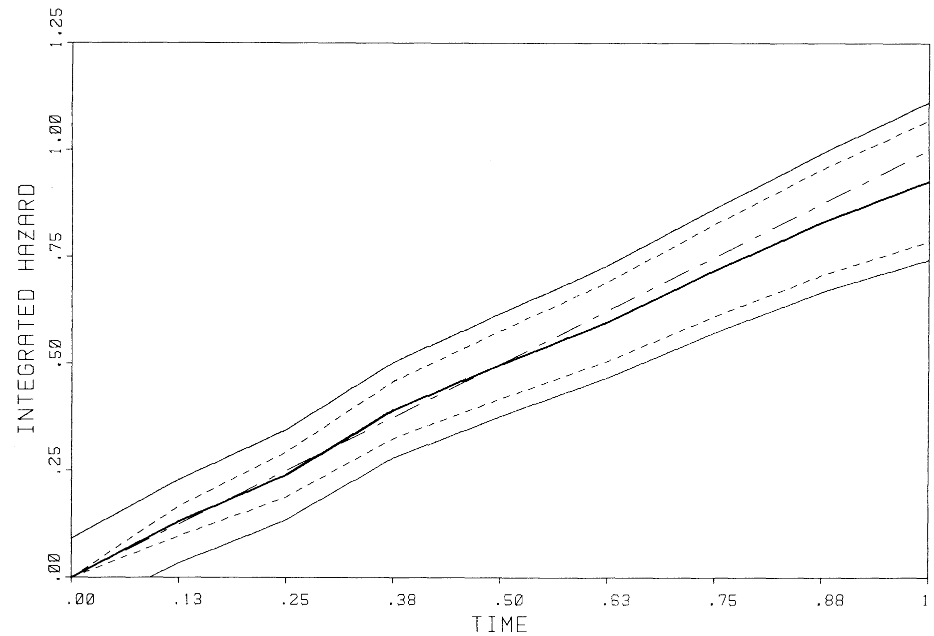
Continuous data case[a]; $b_n = 1/8$							
	Coverage probabilities				Ratios		
Int	OLS	WLS-1	WLS-2	WLS-3	WLS-1	WLS-2	WLS-3
1	.9477	.9477	.9477	.9477	1.0000	1.0000	1.0000
2	.9486	.9479	.9468	.9489	1.0620	1.0706	1.0613
3	.9483	.9485	.9441	.9487	1.0855	1.0987	1.0847
4	.9494	.9491	.9455	.9494	1.0984	1.1155	1.0977
5	.9518	.9485	.9460	.9504	1.1062	1.1268	1.1054
6	.9506	.9511	.9468	.9521	1.1117	1.1366	1.1104
7	.9509	.9469	.9415	.9488	1.1153	1.1450	1.1133
8	.9487	.9490	.9431	.9509	1.1164	1.1525	1.1137

Grouped data case[a]; $d = 8$, $n_S = 1$							
	Coverage probabilities				Ratios		
Int	OLS	WLS-1	WLS-2	WLS-3	WLS-1	WLS-2	WLS-3
1	.9489	.9489	.9489	.9489	1.0000	1.0000	1.0000
2	.9499	.9482	.9464	.9490	1.0624	1.0700	1.0620
3	.9474	.9490	.9481	.9496	1.0860	1.0940	1.0859
4	.9486	.9492	.9478	.9500	1.1092	1.0992	1.0990
5	.9509	.9490	.9467	.9499	1.1070	1.1180	1.1073
6	.9509	.9500	.9471	.9503	1.1129	1.1261	1.1129
7	.9516	.9486	.9454	.9497	1.1168	1.1318	1.1168
8	.9471	.9490	.9457	.9500	1.1188	1.1368	1.1189

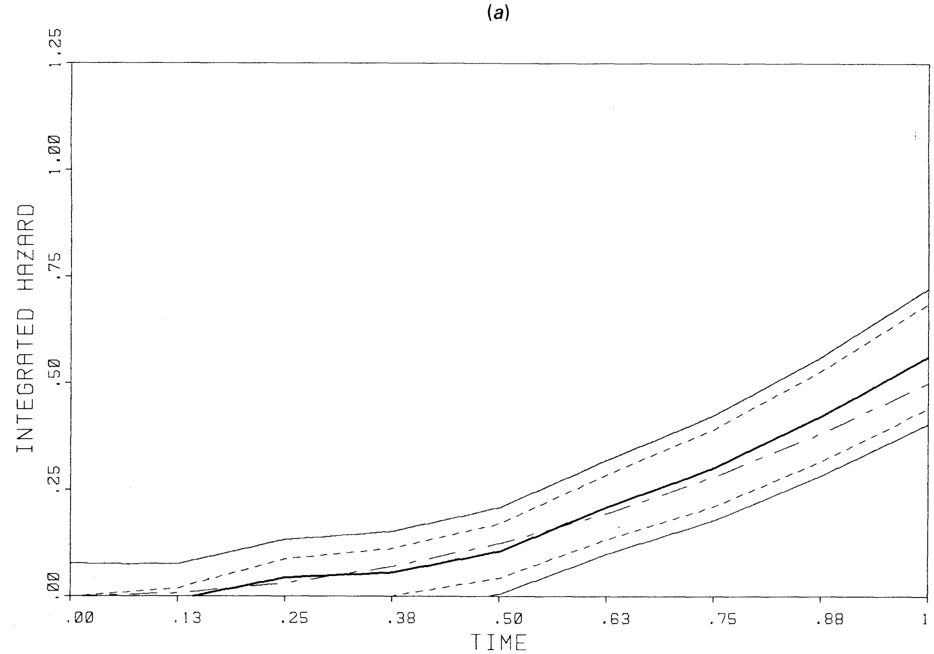
NOTE: See Table 1 Note.

methods give considerable improvement (up to 11% for WLS-1 and WLS-3, and up to 15% for WLS-2) over the OLS method for Model 2. For Model 1, however, we observed that the improvement was only about 1%–2%. It seems that the smaller α_2 in Model 1 makes it harder to estimate the weights successfully. Moreover, the WLS approach seems to show less improvement over OLS in situations where there is less dispersion in the covariates. Even when the "true" weights were used in Model 1, giving the theoretically best possible improvement of WLS over OLS, the improvement was only 4%–6% for covariate 1 and 5%–10% for covariate 2.

Figures 1–4 contain plots of the grouped data WLS estimator and its confidence intervals and bands under Model 1 for light (28%) and heavy (68%) censoring, $d = 8$ and 64. The WLS-1 method was used to estimate the variance. The sample size n was set to 2,000. The random number generator used the same initial seed for all runs, so that all runs have the same failure times (but censoring times were different for the light and heavy censoring).



(a)



(b)

Figure 1. Light censoring (28%), n = 2000, d = 8. (a) Covariate 1. (b) Covariate 2. Thick lines = WLS estimator; regular dashed lines = 95% pointwise confidence limits (WLS-1); thin lines = 95% confidence bands (WLS-1); irregular dashed lines = true cumulative hazard functions (Figs. 1–4).

As expected, the bands are wider under the heavier censoring; compare Figures 3 and 4 with Figures 1 and 2. It also appears that the estimator that uses $d = 8$ (in Figures 1 and 3) is a smoothed version of the estimator that uses $d = 64$ (in Figs. 2 and 4). Although the estimators are adequate in each case, notice that under heavy censoring the estimator that uses $d = 64$ (in Fig. 4) oscillates considerably when t is close to 1. This is due to the very small number of observed failures (averaging less than 5) for each of the intervals in this region. A more flexible procedure would be to merge adjoining intervals that contain very few observed failures. For contrast, in Figure 2 notice that the oscillation is negligible. This is because of the lower censoring rate, which gives an average of more than 10 observed failures per interval near $t = 1$.

4. APPLICATION TO ANALYSIS OF CANCER MORTALITY AMONG JAPANESE ATOMIC BOMB SURVIVORS

The Radiation Effects Research Foundation (RERF) in Hiroshima, Japan, has followed since 1950 a group of over 100,000 atomic bomb survivors. Data on these survivors is the primary source of information on the epidemiologic effects of ionizing radiation. The National Research Council (1980) report and the report by Kato and Schull (1982) con-

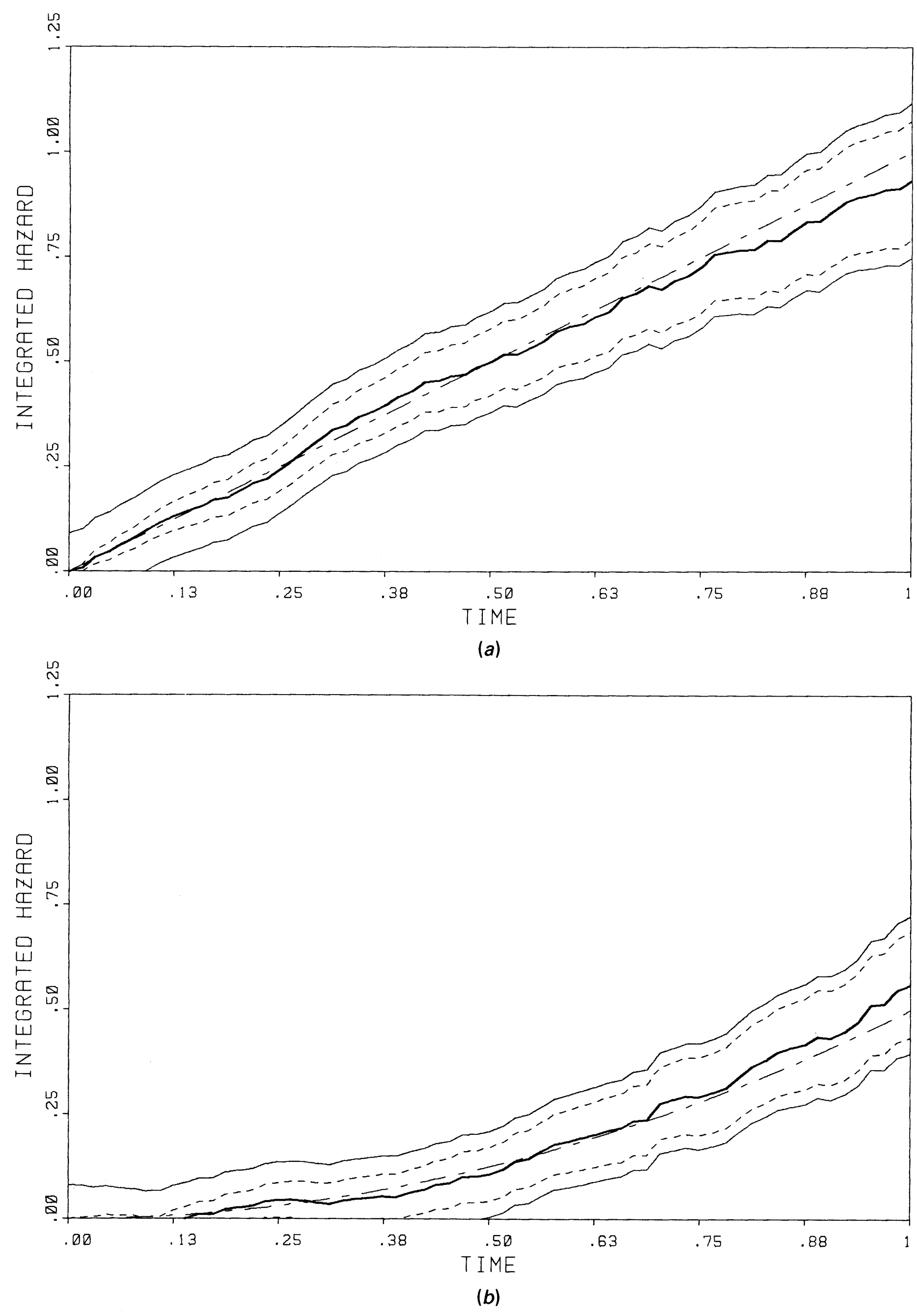


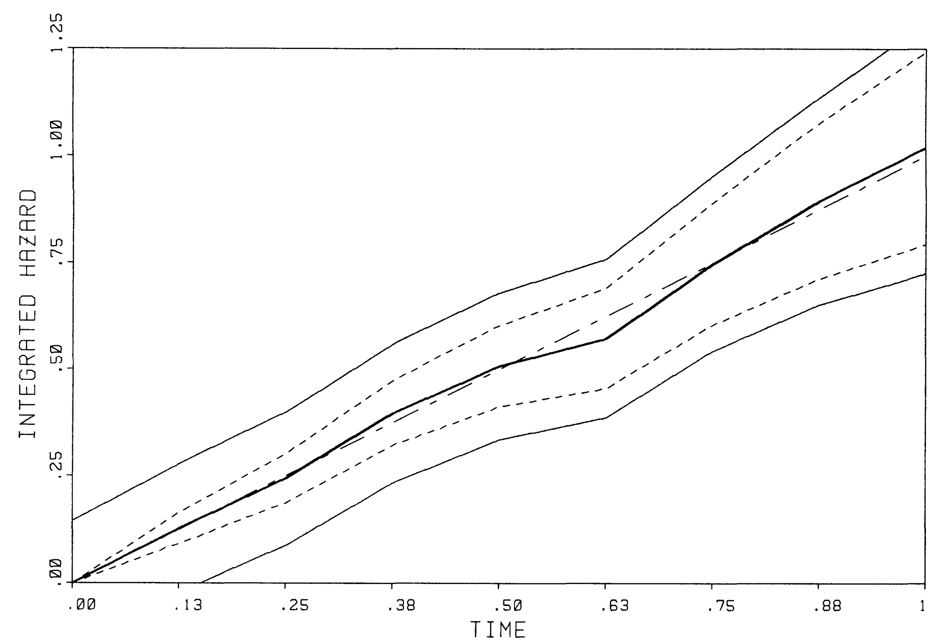
Figure 2. Light censoring (28%), $n = 2000$, $d = 64$. (a) Covariate 1. (b) Covariate 2. See Figure 1 for key.

tain detailed analyses of these data. As noted by Pierce and Preston (1984), however, a difficulty with the methods used in these reports is that they do not make explicit allowance for temporal variation in risks; they simply average the risk over the follow-up period.

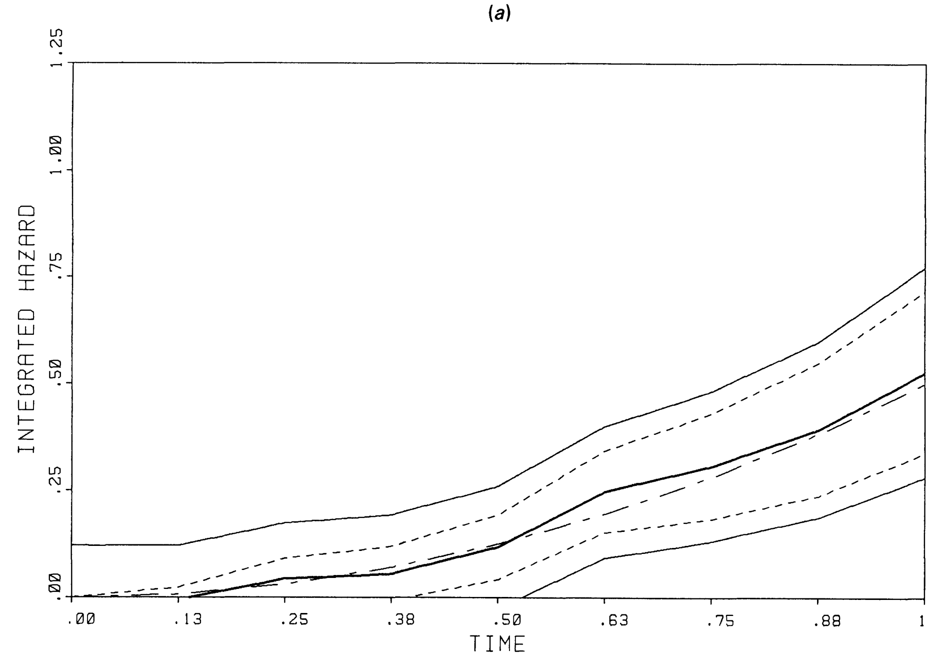
Pierce and Preston (1985) studied cancer mortality among the atomic bomb survivors using a parametric additive risk model. They found that background and excess rates of cancer mortality vary markedly with age at exposure and time since exposure.

We fitted the additive risk model (1.1) to the data used by Pierce and Preston (1985). We analyzed separately each of 4 cohorts defined by the age at exposure intervals 0–9, 10–19, 20–34, and 35–49 years of age at time of bomb. The time variable t is time since exposure, ranging from 5 to 37 years over the follow-up period, 1950–1982. The only censoring prior to the end of follow-up is that due to other (noncancer) causes of death. Summary information for each cohort is given in Table 3. Further description of the data can be found in Preston, Kato, Kopecky, and Fujita (1986).

Three covariates are used. For individual i they are: Y_{i1} = indicator (male), Y_{i2} = indicator (female), Y_{i3} = dose (in units of 100 rads). The hazard functions α_1 and α_2 are the background cancer mortality rates for males and females



(a)



(b)

Figure 3. Heavy censoring (68%), n = 2000, d = 8. (a) Covariate 1. (b) Covariate 2. See Figure 1 for key.

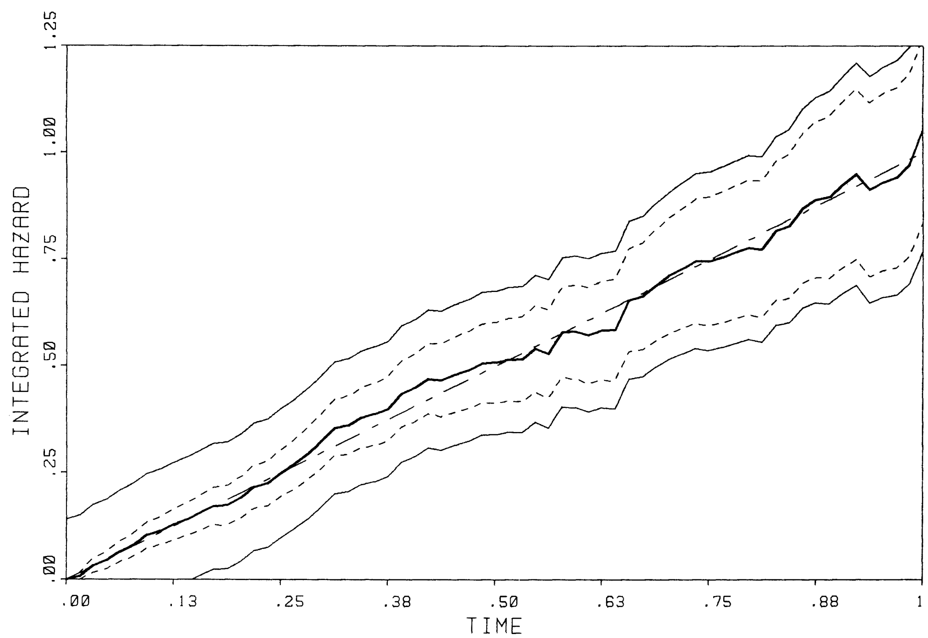
respectively. The third hazard function α_3 is the excess cancer mortality rate per 100 rads of radiation exposure.

In the data provided to us by the RERF, time since exposure is grouped into eight 4-year intervals: 5–9, ..., 33–37 years. It is natural to use these eight time intervals as the partition of the follow-up period in the evaluation of the WLS estimator. The Monte Carlo study in Section 3 indicates that satisfactory results can be obtained with as few as d = 8 intervals. As in Pierce and Preston (1985), the dose variable is taken as the midpoint of one of the six

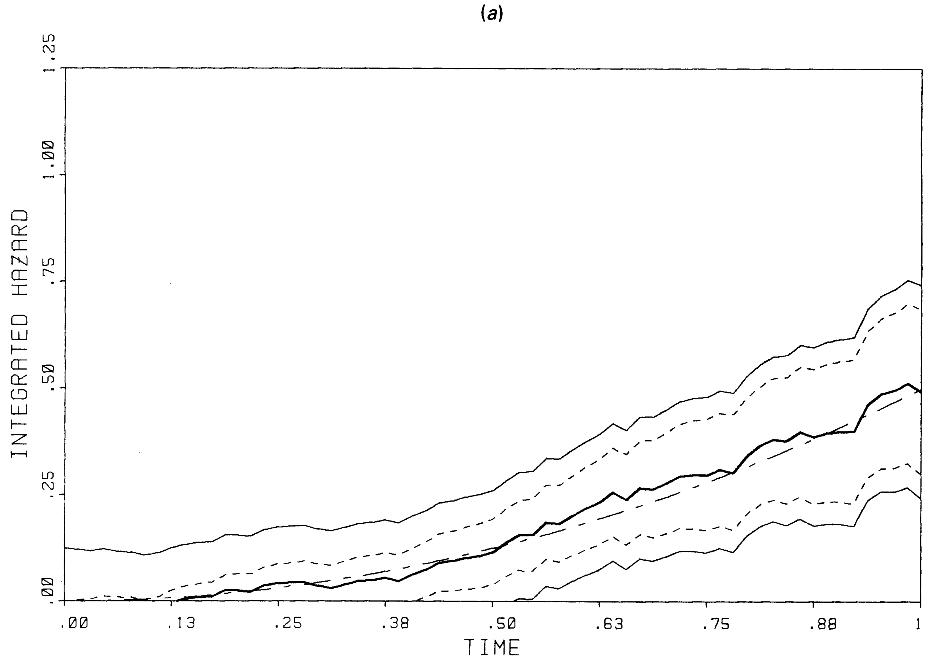
Table 3. Cohort Sizes, Summary Mortality Figures, and Censoring Prior to End of Follow-up

Age at exposure	Cohort size*	Deaths due to cancer excluding leukemia	Deaths from all causes	Censoring
0–9	18,416	93	728	87%
10–19	19,242	349	1,715	80%
20–34	17,694	949	3.075	69%
35–49	20,916	2,788	11,234	75%

*Approximate, having been estimated from the grouped data.



(*a*)



(*b*)

Figure 4. Heavy censoring (68%), n = 2000, d = 64. (a) Covariate 1. (b) Covariate 2. See Figure 1 for key.

dose groups: 0, 1–50, 50–100, 100–200, 200–300, >300 rads with dose = 400 for dose > 300. The data used here are based on the old dosimetry, T65DR. The analysis is limited to the epithelial cancer mortality, in which leukemia mortality is excluded. We use $n_S = 1$; that is, the weights are calculated using the OLS estimator for only the previous interval, and the variance estimator is WLS-1.

Figures 5–8 indicate our estimates and 95% confidence intervals and bands for the background and excess cumulative cancer mortality rates as functions of time since exposure. The estimates of Pierce and Preston are also plotted. All estimates are given in units of deaths per 1,000 persons at risk. Pierce and Preston's parametric model for the cancer mortality rate $\lambda(t)$ is given by

$$\lambda(t) = \exp\{\nu_{0e} + \nu_{1s} + \nu_{2s}\log(e + t)\} + \beta d \exp\{\rho_{0e} + \rho_1 \log(t)\}, \qquad (4.1)$$

where e = midpoint of age at exposure interval, s = sex, and d = dose. The first and second terms in (4.1) represent the background and excess cancer mortality rates which Pierce and Preston fitted by maximum likelihood. We have

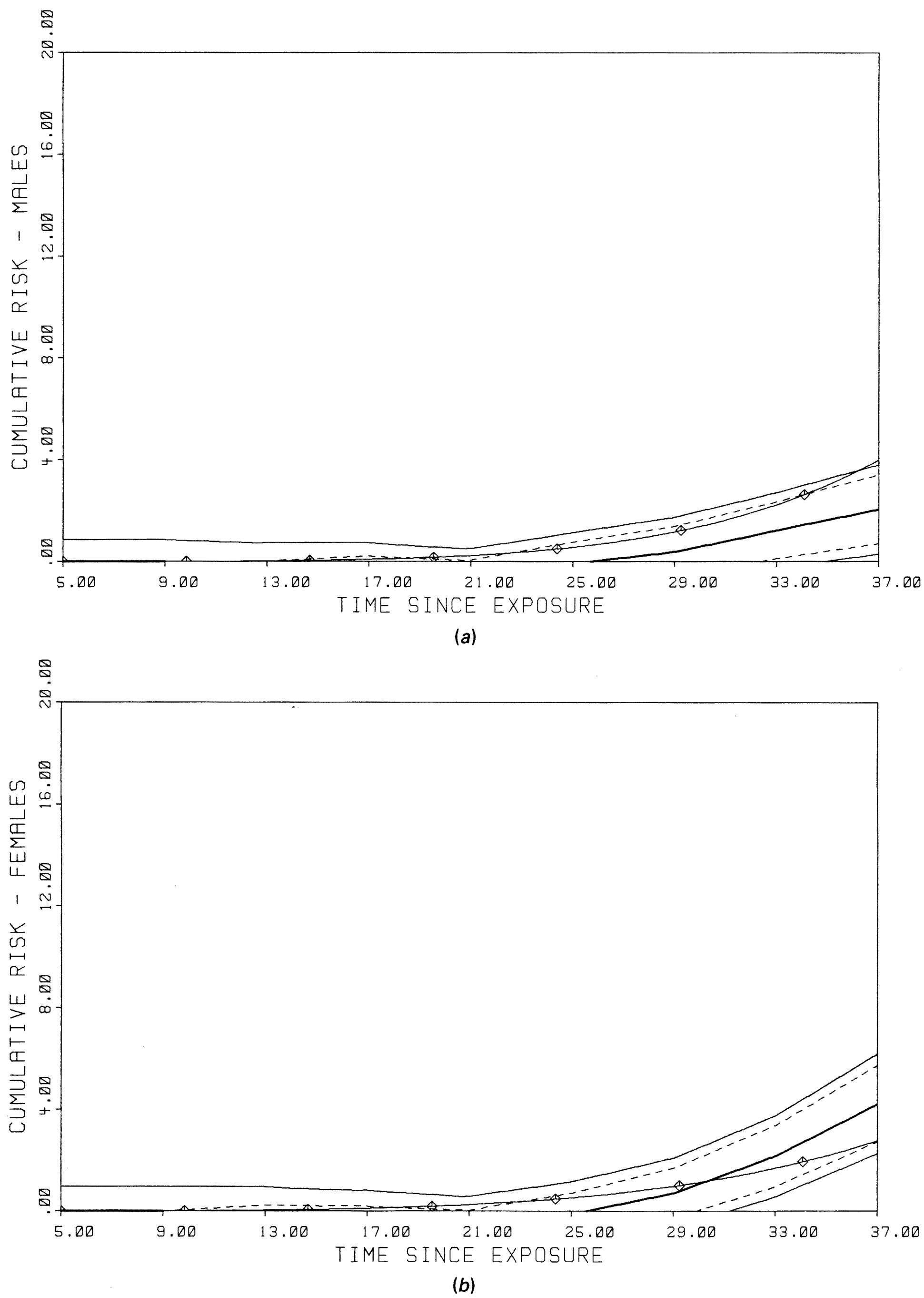
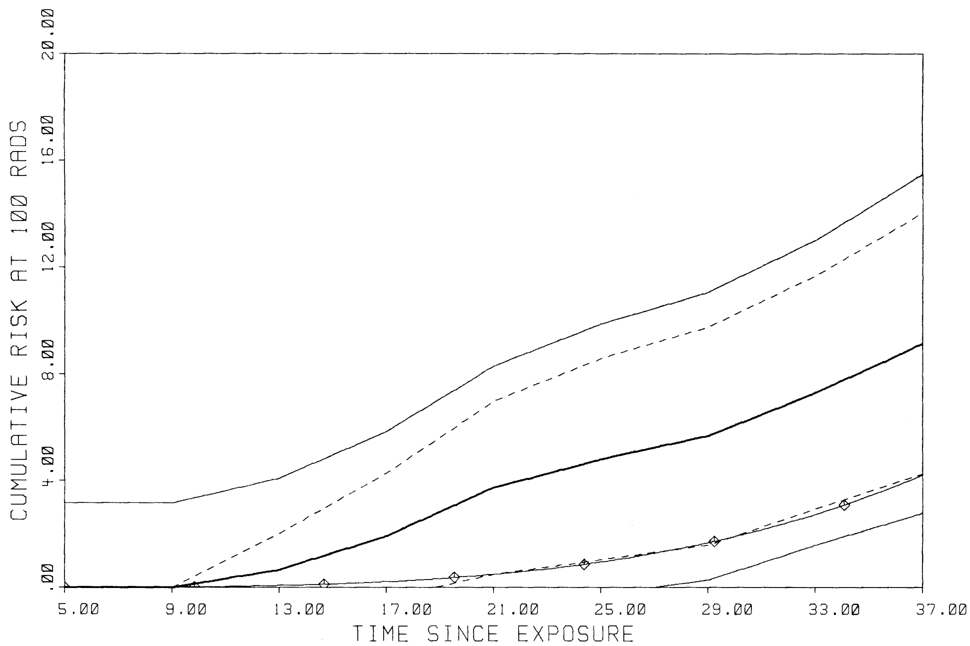


Figure 5. (continued)



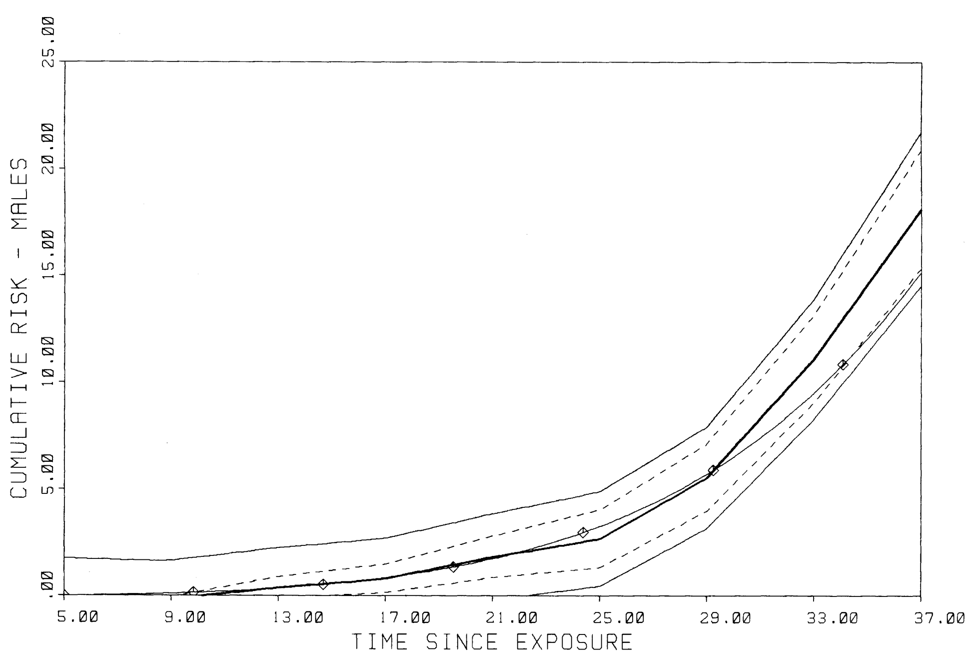
(c)

Figure 5. 0–9 Years of Age at Exposure. (a) Background for males. (b) Background for females. (c) Excess. Thick lines = WLS estimator; regular dashed lines = 95% pointwise confidence limits (WLS-1); thin lines = 95% confidence bands (WLS-1); lines marked with boxes = estimates of Pierce and Preston (Figs. 5–8).

integrated these to provide a comparison with our WLS estimates.

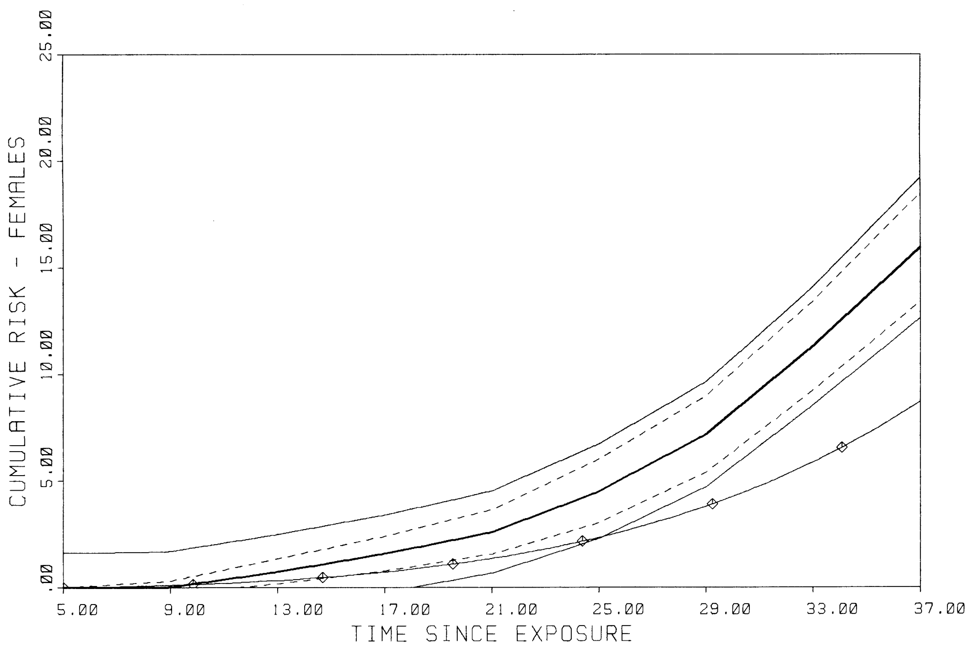
From Figures 5–8 we see that there is a significant dose effect (since the band for dose does not contain the zero function) for all cohorts. This is despite the fact that the confidence bands with $d = 8$ are conservative according to our simulation results in Section 3. In interpreting Figures 5–8, bear in mind that the vertical scales are different for each cohort. Despite appearances, the dose effect bands have roughly the same width for each cohort.

On the whole, our estimates are in agreement with those of Pierce and Preston (1985). We also observe that the relative risk (dose effect vs. background) decreases sharply with age at exposure. Our estimates for the dose effect are

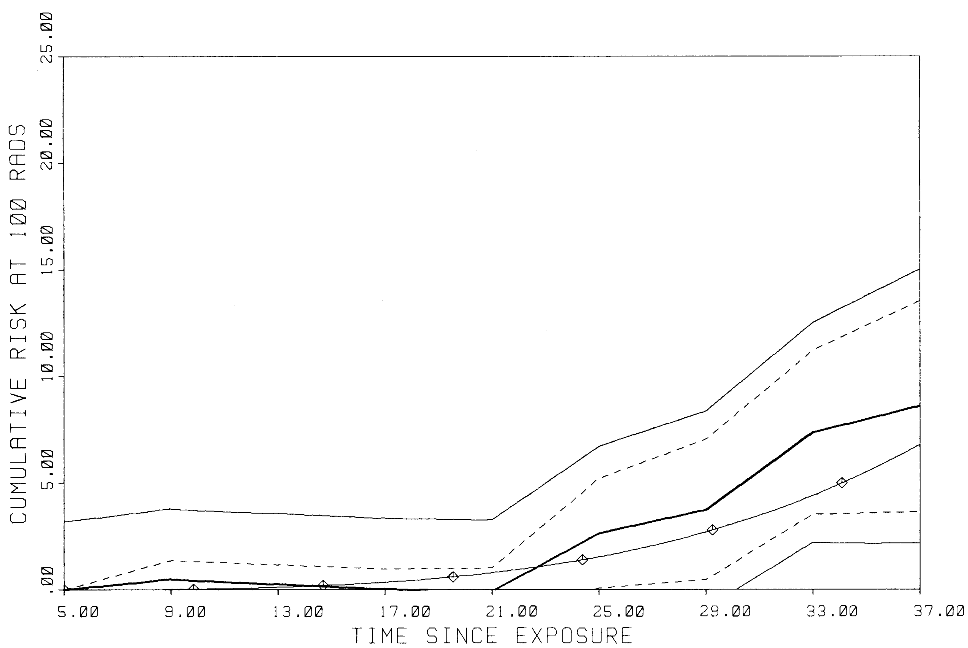


(a)

Figure 6. (continued)



(b)



(c)

Figure 6. 10–19 Years of Age at Exposure. (a) Background for males. (b) Background for females. (c) Excess. See Figure 5 for key.

somewhat higher than Pierce and Preston's, but their estimates are still within our 95% bands. There is a significant difference between our estimates and Pierce and Preston's for the female background mortality rate in all but the 0–9 years of age cohort, however.

5. ASYMPTOTIC THEORY

First consider the continuous case. Define the multivariate counting process $\mathbf{N}(t) = (N_1(t), \ldots, N_n(t))'$, where $N_i(t) = I(X_i \le t, \delta_i = 1)$ is the indicator of an uncensored failure for subject i prior to time t. This process has intensity $\lambda(t) = \mathbf{Y}(t)\alpha(t)$, where $\mathbf{Y}(t)$ is the $n \times p$ matrix with ith row $Y_i' I(X_i \ge t)$. That is to say, $\mathbf{M}(t) = \mathbf{N}(t) - \int_0^t \lambda(s)\, ds$ is an n-variate martingale with respect to the filtration $\mathcal{F}_t = \sigma(\mathbf{N}(s), \mathbf{Y}(s), 0 \le s \le t)$; see Aalen (1980).

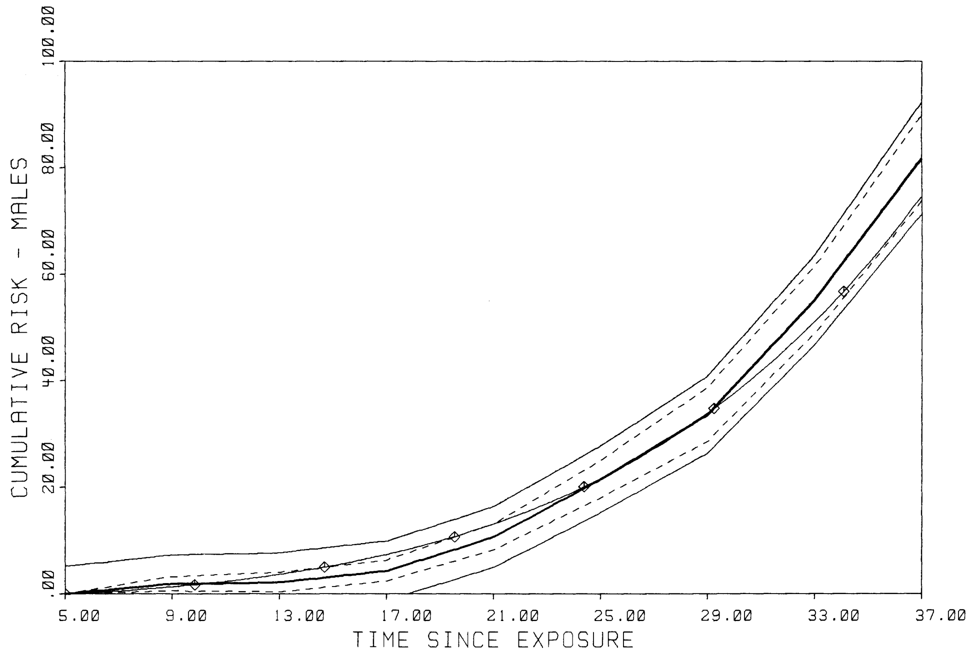
Aalen's least squares estimator is given by

$$\tilde{A}(t) = \int_0^t (\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\, d\mathbf{N}.$$

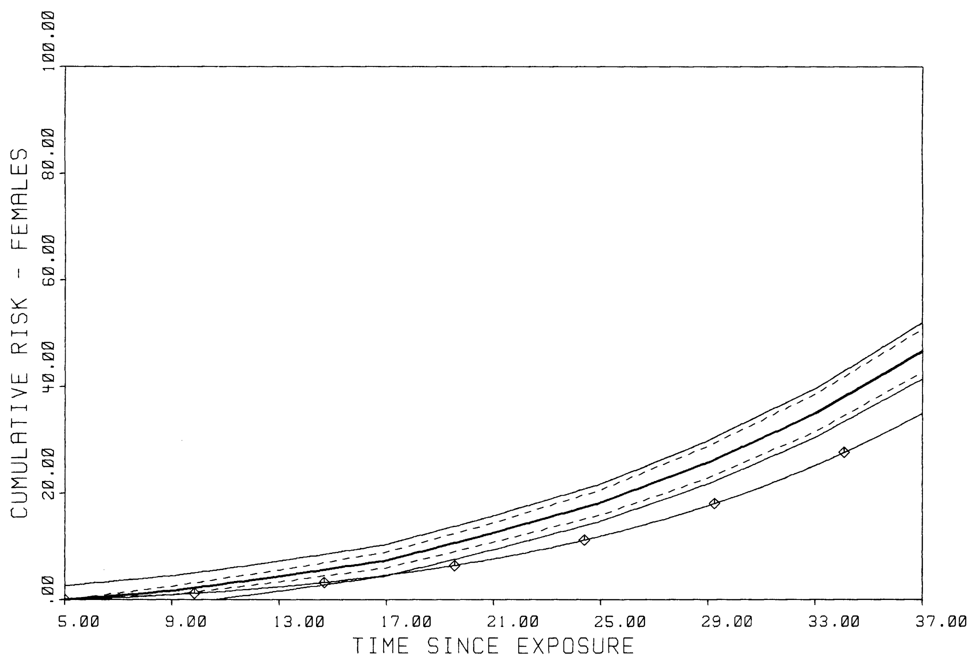
Assume that the covariates Y_{ij} are bounded. By a time reversed version of Ranga Rao's (1962) law of large numbers, the $p \times p$ matrices $n^{-1}\mathbf{Y}'\mathbf{Y}$, $n^{-1}\mathbf{Y}'(\operatorname{diag} \mathbf{Y}\alpha)\mathbf{Y}$ converge uniformly on $[0, T]$, in probability, to deterministic limits. Here diag $\mathbf{Y}\alpha$ is the diagonal matrix having diagonal $\mathbf{Y}\alpha$. Assume that the limit of $n^{-1}\mathbf{Y}'\mathbf{Y}$ is nonsingular and continuous (or, more generally, has smallest eigenvalue bounded away from zero) on $[0, T]$. Then $\tilde{A} - A$ coincides in probability, as $n \to \infty$, with the p-variate martingale $\tilde{\mathbf{M}} = \int_0^{\cdot} (\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\, d\mathbf{M}$. Now, $\tilde{\mathbf{M}}$ has predictable covariation process

$$\langle \tilde{\mathbf{M}} \rangle_t = \int_0^t (\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'(\operatorname{diag} \mathbf{Y}\alpha)\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1}\, ds,$$

so that $\langle n^{1/2}\tilde{\mathbf{M}} \rangle$ converges pointwise in probability to a continuous deterministic function. The maximum jump of $n^{1/2}\tilde{\mathbf{M}}$ is $O_P(n^{-1/2})$. Thus Rebolledo's (1980) martingale central limit theorem implies that $n^{1/2}\tilde{\mathbf{M}}$ [and hence $n^{1/2}(\tilde{A} - A)$] converges in distribution to a p-variate continuous Gaussian martingale. In particular, if $b_n \to 0$ and $nb_n \to \infty$, then the smoothed least squares estimator given by (2.3) is

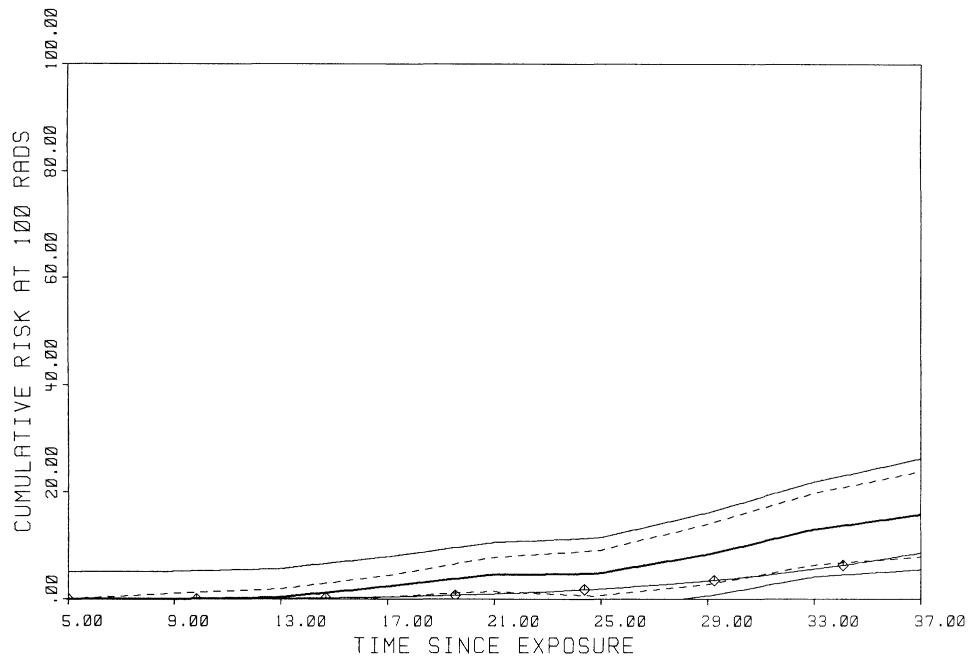


(a)



(b)

Figure 7. (continued)



(c)

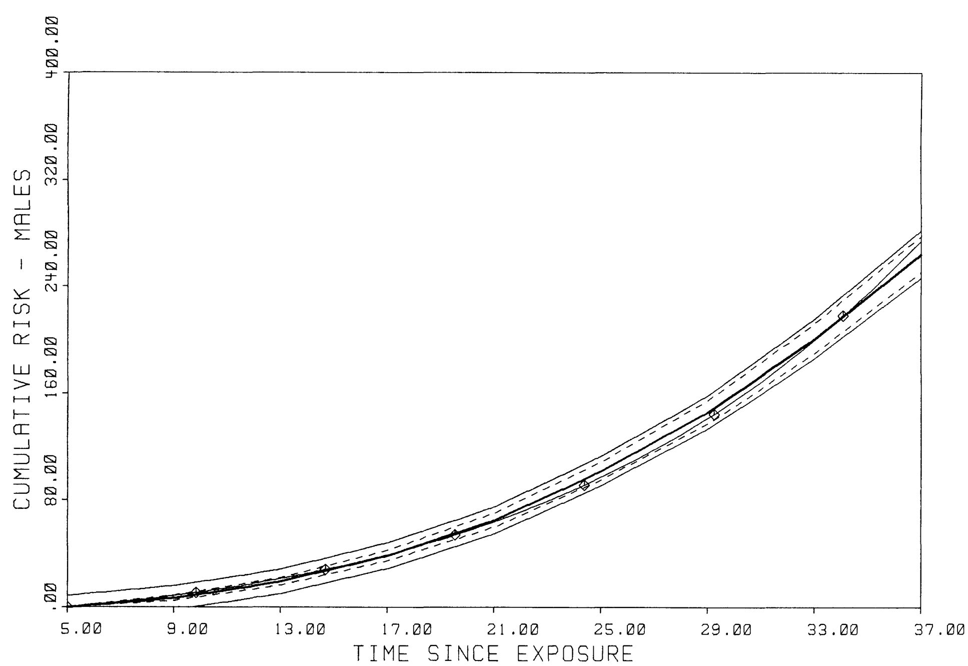
Figure 7. 20–34 Years of Age at Exposure. (a) Background for males. (b) Background for females. (c) Excess. See Figure 5 for key.

uniformly consistent over $[b_n, T]$; see the proof of Theorem 4.1.2 of Ramlau-Hansen (1983).

Next consider the WLS estimator

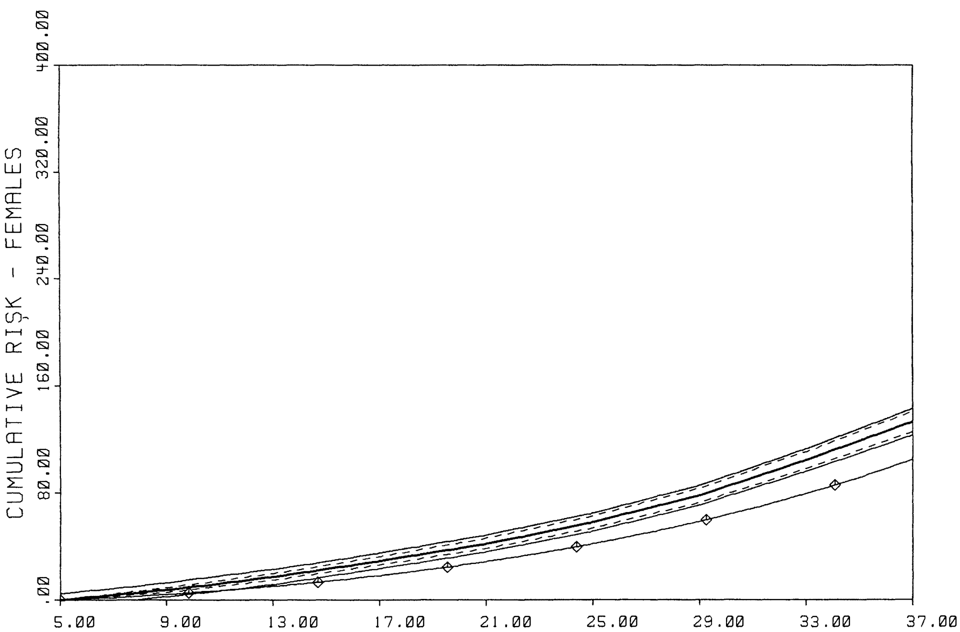
$$\hat{A}(t) = \int_0^t (\mathbf{Y}'\hat{\mathbf{W}}\mathbf{Y})^{-1}\mathbf{Y}'\hat{\mathbf{W}}\, d\mathbf{N},$$

where $\hat{\mathbf{W}} = (\text{diag } \mathbf{Y}\alpha^*)^{-1}$ and α^* is the kernel estimator (2.3). In practice, we use the WLS estimator on $[b_n, T]$ and the OLS estimator on $[0, b_n]$. Note that, since $b_n \to 0$, this modification has no effect asymptotically. Assume that the intensity $\lambda_i(t)$ of N_i is uniformly bounded away from zero unless $\mathbf{Y}_{ij}(t) = 0$ for all $j = 1, \ldots, p$. Then $\hat{\mathbf{W}}$ is a uniformly consistent estimator of the "true" weight $\mathbf{W} = (\text{diag } \mathbf{Y}\alpha)^{-1}$ over $[b_n, T]$. To simplify the notation, assume from

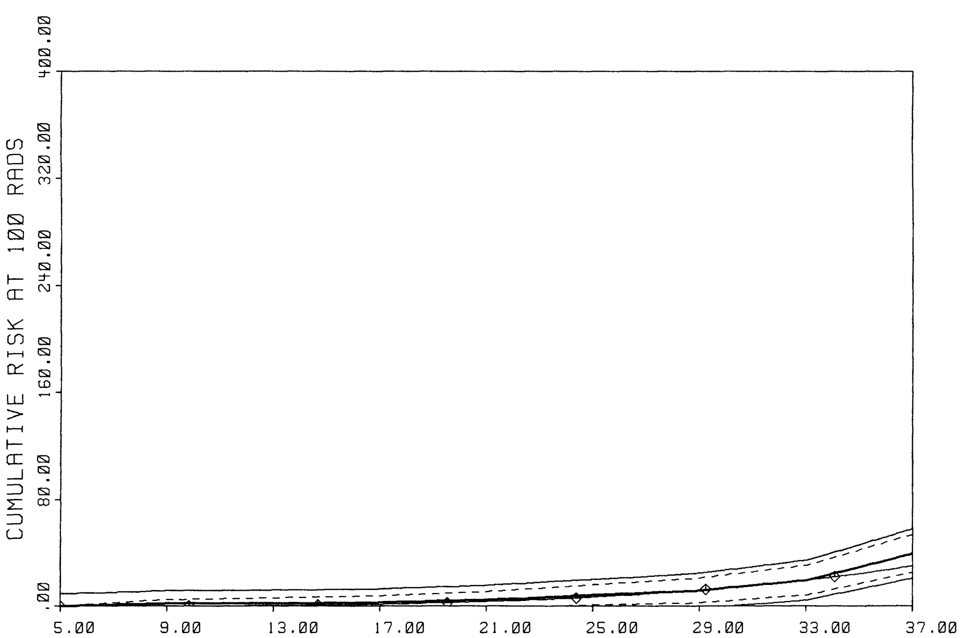


(a)

Figure 8. (continued)



(b)



(c)

Figure 8. 35–49 Years of Age at Exposure. (a) Background for males. (b) Background for females. (c) Excess. See Figure 5 for key.

now on that this is so over the whole of $[0, T]$. Using Ranga Rao's (1962) law of large numbers again, we have that $n^{-1}\mathbf{Y}'\mathbf{W}\mathbf{Y}$ converges uniformly in probability on $[0, T]$ to a deterministic matrix function $\mathbf{V}$. By uniform consistency of $\hat{\mathbf{W}}$ and boundedness of the covariates, $n^{-1}\mathbf{Y}'\hat{\mathbf{W}}\mathbf{Y}$ also converges uniformly in probability on $[0, T]$ to $\mathbf{V}$. Assume that $\mathbf{V}$ is nonsingular and smooth on $[0, T]$. Then $\hat{A} - A$ coincides in probability, as $n \to \infty$, with $\hat{\mathbf{M}} = \int_0^{\cdot}(\mathbf{Y}'\hat{\mathbf{W}}\mathbf{Y})^{-1}\mathbf{Y}'\hat{\mathbf{W}}\, d\mathbf{M}$, which is a p-variate martingale, since $\hat{\mathbf{W}}$ was arranged to be predictable. Moreover,

$$\langle\hat{\mathbf{M}}\rangle_t = \int_0^t (\mathbf{Y}'\hat{\mathbf{W}}\mathbf{Y})^{-1}\mathbf{Y}'\hat{\mathbf{W}}(\operatorname{diag}\mathbf{Y}\alpha)\hat{\mathbf{W}}\mathbf{Y}(\mathbf{Y}'\hat{\mathbf{W}}\mathbf{Y})^{-1}\, ds,$$

so that $\langle n^{1/2}\hat{\mathbf{M}}\rangle_t$ converges in probability to $\int_0^t \mathbf{V}^{-1}\, ds$. Thus Rebolledo's (1980) martingale central limit theorem implies that $n^{1/2}(\hat{A} - A)$ converges in distribution to a p-variate Gaussian martingale with predictable covariation process