Biostatistics (2005), **6**, 4, *pp.* 604–614
doi:10.1093/biostatistics/kxi030
Advance Access publication on April 28, 2005

Robust modeling in screening studies: estimation of sensitivity and preclinical sojourn time distribution

YU SHEN*

Department of Biostatistics and Applied Mathematics,
M. D. Anderson Cancer Center, University of Texas, Houston, TX 77030, USA
yshen@mdanderson.org

MARVIN ZELEN

Department of Biostatistics, Harvard School of Public Health and the Dana-Farber Cancer Institute, Boston, MA 02115, USA

SUMMARY

In early-detection clinical trials, quantities such as the sensitivity of the screening modality and the preclinical duration of the disease are important to describe the natural history of the disease and its interaction with a screening program. Assume that the schedule of a screening program is periodic and that the sojourn time in the preclinical state has a piecewise density function. Modeling the preclinical sojourn time distribution as a piecewise density function results in robust estimation of the distribution function. Our aim is to estimate the piecewise density function and the examination sensitivity using both generalized least squares and maximum likelihood methods. We carried out extensive simulations to evaluate the performance of the methods of estimation. The different estimation methods provide complimentary tools to obtain the unknown parameters. The methods are applied to three breast cancer early-detection trials.

Keywords: Piecewise-constant density function; Preclinical duration; Screening clinical trials; Screening sensitivity; Weighted generalized least squares.

1. INTRODUCTION

This paper is motivated by the expanding early-detection programs in cancer to detect the disease and treat it at an early stage, so that the progression of the disease may be slowed or halted. This strategy has the potential of increasing cure rates or prolonging survival. Over the past decades, it has become a common practice to routinely screen for breast cancer, colon cancer, and other solid tumors. Accompanying the increase in efforts at early detection, questions have arisen concerning the reliable estimates of the natural history of the disease and sensitivity of the screening techniques. The natural history of breast cancer is often described by a progressive disease model, which consists of three states: disease-free state or a state in which the disease cannot be detected by the current screening technologies, the preclinical disease state, and a clinical state. During the preclinical state the disease is asymptomatic, having no signs or symptoms. The transition from the preclinical state to the clinical state occurs when the disease has signs/symptoms resulting in the diagnosis of disease by usual care. The disease is assumed

*To whom correspondence should be addressed.

to be progressive in that the natural history is described by transiting from the disease-free to the preclinical to the clinical state. The object of an early-detection program is to diagnose the disease in the preclinical state. The preclinical disease duration is defined as the time from the onset of detectable preclinical disease to the onset of clinical disease. Diagnosing the disease in the preclinical state depends on the sensitivity of the examinations; screening sensitivity is the probability that the disease is screen-detected given the disease in the preclinical state. Knowledge of the natural history of the disease can guide health policy makers in planning optimal and cost-effective public health screening programs to extend potential benefits to large populations. Although our investigations are motivated by breast cancer, the results are general and applicable to the early detection of any chronic disease which is progressive.

Directly estimating the preclinical disease duration and screening sensitivity is difficult because the nature of the disease process precludes exact observation of the onset of the preclinical disease state, and/or the onset of clinical disease when the disease is detected early by screening. Most statistical models estimate the sojourn time distribution using maximum likelihood methods. Explicitly, these models assume an absolutely continuous parametric model for the sojourn time distribution. A simplified parametric model has an advantage in that fewer parameters are estimated with relatively high efficiency (e.g. Day and Walter, 1984; Shen and Zelen, 1999; Parmigiani and Skates, 2001; Pinsky, 2001, among others). However, specifying a parametric model may have serious limitations, as the sojourn time is not observed directly. In the absence of prior knowledge, it is often difficult to justify a specific parametric distribution having a valid representation of the sojourn times (Louis *et al.*, 1978; Schwartz, 1978). Moreover, Etizoni and Shen (1997) observed that the estimate of the sensitivity relies heavily on a correct model assumption of the sojourn time. Estimation methods that make flexible model assumptions about the preclinical duration can lead to more robust estimates of the sojourn time distribution and screening sensitivity.

Louis *et al.* (1978) proposed a nonparametric procedure to describe the natural history of the disease process. But their approach has not been widely used due to its complexity and extensive computations. DeGruttola and Lagakos (1989) proposed a nonparametric method to estimate asymptomatic duration in AIDS. In contrast to cancer screening, by definition, the test for HIV seropositivity is effectively 100% sensitive after seroconversion. Seroconversion in AIDS is analogous to the time point at which cancer becomes first detectable by a screening examination. However, in cancer screening, we also need to estimate the unknown screening sensitivity. By generalizing the nonparametric method of DeGruttola and Lagakos (1989), Etzioni and Shen (1997) described a method to jointly estimate the nonparametric sojourn time distribution, the distribution of disease onset time, and the screening sensitivity. Their proposed model appears to be a useful tool for further exploration of parametric or semiparametric modeling. These methods use the EM algorithm which is computationally intensive and may yield a saddle point instead of a unique solution of the maximum likelihood function.

There is a need to develop approaches that require fewer model restrictions. It is well known that piecewise functions may provide good approximations for various distribution models with reasonable accuracy. In this paper, we propose a procedure for jointly estimating the sojourn time distribution and screening sensitivity within a flexible framework for the sojourn time distribution.

This research is motivated by breast cancer early-detection trials. Because of the limitation on the structure of observed data from screening trials, we can only observe the grouped data within prespecified intervals based on periodic screening programs. For example, in some of the randomized breast cancer early-detection trials, women were randomized to receive several periodic examinations or be in a usual care group. Data were collected for the number of screen-detected cases at each examination time and the number of clinical cases that surfaced between two scheduled examinations in the screening group. The periodic screening schedule suggests that we are unlikely to substantially improve a model where the sojourn time distribution is assumed to be a piecewise-constant density function over equally spaced screening intervals. Such a distribution serves as a suitable approximation for a variety of distributions.

In Section 2, we describe the underlying stochastic model and the structure of the observed data in a typical early-detection trial. Section 3 presents two types of estimation methods: the maximum likelihood estimation (MLE) and the generalized least squares (GLS) methods. In Section 4, we carry out an extensive simulation study to evaluate the proposed estimation methods. In Section 5, the methods are applied to data from three breast cancer randomized screening trials; i.e. the Health Insurance Plan of New York (HIP) (Shapiro *et al.*, 1988) and the two Canadian National Breast Screening Studies (CNBSS) (Miller *et al.*, 1992a,b). We conclude with a discussion in Section 6.

2. DATA STRUCTURE AND MODEL ASSUMPTIONS

Consider a cohort of initially asymptomatic individuals enrolled in a screening program who undergo periodic screening examinations for a specific disease. Based on a prior work of Zelen and Feinleib (1969), we use a model to characterize the disease progression through the three states: (1) disease-free or undetectable disease, S_0; (2) detectable preclinical disease, S_p; and (3) clinical disease, S_c. We assume that the natural history of the disease is progressive and that the transitions are $S_0 \rightarrow S_p \rightarrow S_c$.

Our goal is to jointly estimate the sensitivity of the screening examination, prevalence of the disease, and the preclinical disease duration from state S_p to S_c. In particular, we use a flexible structure to model the sojourn time distribution via a piecewise-constant probability density characterizing the transition from state S_p to S_c.

The data used to estimate the aforementioned quantities are generated by the observed periodic screening examination histories from the screening arm of a trial. Specifically, data consist of whether a subject is diagnosed in a scheduled screening examination or between the scheduled examinations and of the number of prior negative screening examinations. The cases diagnosed between examinations are also called interval cases. Let $t_0 < t_1 < \cdots < t_{K-1} < U$ denote K-ordered screening examination times, where U is the follow-up time past the last screening examination. Note that the kth screening examination is performed at t_{k-1}. Define the kth screening interval $[t_{k-1}, t_k)$ for $k = 1, \ldots, K$, where $t_K = U$. Let n_k be the total number of individuals examined at t_{k-1}, s_k be the number of cases detected at the examination given at t_{k-1}, and r_k be the number of interval cases diagnosed within the interval (t_{k-1}, t_k).

The outcome for each individual within a screening interval follows a trinomial distribution; i.e. screen-detected at t_{k-1}, interval case in (t_{k-1}, t_k), or neither. For each individual, the periodic screening history consists of longitudinal observations. At the kth screening interval $[t_{k-1}, t_k)$, for individuals in the at-risk set, let $y_{ik} = (y_{ik1}, y_{ik2})^T$ define the response variable for the ith individual. Define y_{ikj} to be an indicator variable such that $y_{ik1} = 1$ if the ith individual is detected to have the disease at the kth screening examination, $y_{ik2} = 1$ if the individual is an interval case in the kth interval, and $y_{ikj} = 0$ for $j = 1, 2$, otherwise.

3. ESTIMATION PROCEDURES

We begin by assuming that the sojourn time T has a piecewise-constant density function. Since most screening trials have equal intervals between screening examinations, our development will assume equal intervals between examinations. However, the model need not be restricted to equal intervals between examinations. Define $\delta = t_k - t_{k-1}$. When following a screening trial that offers K screening examinations with equal intervals, δ is the minimum admissible time interval for a piecewise-constant density function on K intervals.

Denote the probability density function (pdf) of the sojourn time T and its survival function by $q(t)$ and $Q(t)$, respectively. Also define

$$q(t) = q_k/\delta, \quad t \in [(k-1)\delta, k\delta), \quad k = 1, \ldots, K,$$

where $\sum_{j=1}^{K} q_j = 1$ and t is internal time relative to an individual. Consequently,

$$Q(t) = \int_t^\infty q(x)\,\mathrm{d}x = q_k(k - t/\delta) + 1 - \sum_{j=1}^{k} q_j, \quad t \in [(k-1)\delta, k\delta),$$

where $k = 1, \ldots, K$. This structure ensures the identifiability of the probability in each admissible time interval of the sojourn time in S_p that overlaps with the screening intervals. In this setting, the hazard function is

$$h(t) = \frac{q_k}{q_k(k\delta - t) + (1 - \sum_{j=1}^{k} q_j)\delta}, \quad \text{for } t \in [(k-1)\delta, k\delta).$$

Let $\eta_k = h((k-1)\delta)$ be the hazard rate at the left end of the screening interval. We can derive a one-to-one relationship between the pdf and such a hazard rate for each interval. Suppose that the time interval is rescaled so that $\delta = 1$. Then the pdf of the sojourn time can be reduced to a generic form

$$q_k = \eta_k(1 - \eta_1), \ldots, (1 - \eta_{k-1}), \quad k = 1, \ldots, K.$$

The survival function, $Q(t)$, can be expressed in terms of hazard rates, η_k as $Q(t) = 1 - \eta_1 t$, for $t \in [0, \delta)$, and

$$Q(t) = \prod_{j=1}^{k}(1 - \eta_j\delta)[1 + \eta_k\{(k-1)\delta - t\}], \quad \text{for } t \in [(k-1)\delta, k\delta).$$

The linear constraints on the piecewise-constant density, $q_k \geqslant 0$, $\sum q_k \leqslant 1$, become, simply, nonnegative constraints on the hazard rates, $\eta_k > 0$, for $0 \leqslant k \leqslant K - 1$, and $\eta_K = 1$. Such a reparameterization typically eases the linear constraints in the computation procedure when solving the estimating equations.

3.1 *Likelihood-based approaches*

It is assumed that the natural history of the disease is described by a stable disease model with a constant screening sensitivity because it is a realistic assumption for the breast cancer screening trials that motivate this research. Under the stable disease model assumption, the proportion of preclinical and clinical cases in a given population remains essentially constant over chronologic time. The stable disease model may serve as a reasonable approximation to the observed data when the screening intervals in chronological time are short relative to the age distribution. Consequently, the disease incidence will be nearly constant over the screening schedule. This is the case for most breast cancer screening trials. Under the stable disease model, the probability that the disease is detected at the kth screening examination or is incident within the kth interval can be expressed in terms of screening sensitivity, β, prevalence, P, and the preclinical sojourn time distribution (Zelen, 1993; Shen and Zelen, 1999). The probability that the disease is detected at the kth screening examination when the sensitivity is β is $D_1(\theta) = P\beta$ and

$$D_k(\theta) = D_{k-1}(\theta) - P\beta^2(1 - \beta)^{k-2}Q_0((k-1)\delta), \quad \text{for } k = 2, \ldots, K,$$

where $Q_0(t) = \int_t^\infty Q(x)\,\mathrm{d}x/\mu$, $\mu = \int_0^\infty Q(t)\,\mathrm{d}t$ is the mean of sojourn time in S_p, and θ represents a vector of all unknown parameters to be estimated. Here P is the point prevalence, and $Q_0(t)$ is the tail probability of the forward recurrence time distribution. The total probability of being incident in the kth interval is $I_1(\theta) = P(\delta/\mu - \beta[1 - Q_0(\delta)])$ and

$$I_k(\theta) = I_{k-1}(\theta) - P\beta(1 - \beta)^{k-1}[Q_0((k-1)\delta) - Q_0(k\delta)], \quad \text{for } k = 2, \ldots, K.$$

We remark that $D_k(\theta)$, the probability of being detected at the kth examination given at time t_{k-1}, includes events in which the subject may not be in the preclinical state before t_{k-1} as well as the false-negative examinations prior to the kth examination. A similar interpretation holds for $I_k(\theta)$.

Recall the definitions for n_k, s_k, r_k within the kth screening interval, and define $m_k = s_k + r_k$ to be the total number of cancers diagnosed in the kth screening interval. Note that n_{k+1} may not be equal to $(n_k - m_k)$ because of dropouts. We will consider $\{n_k, k = 1, \ldots, K-1\}$ as fixed quantities. Define $\mathbf{s} = (s_1, \ldots, s_{K-1})$, $\mathbf{r} = (r_1, \ldots, r_{K-1})$, $\mathbf{m} = (m_1, \ldots, m_{K-1})$, and $\mathbf{n} = (n_1, \ldots, n_{K-1})$. Then the joint likelihood for individuals from each screening interval conditional on $\mathbf{n}$ is

$$L(\mathbf{s}, \mathbf{r}|\mathbf{n}) = \prod_{k=1}^{K} \{D_k(\theta)\}^{s_k} \{I_k(\theta)\}^{r_k} \{1 - D_k(\theta) - I_k(\theta)\}^{n_k - m_k}. \tag{3.1}$$

The likelihood may also be expressed as a product of conditional likelihoods

$$L(\mathbf{s}, \mathbf{r}|\mathbf{n}) = L(\mathbf{m}|\mathbf{n}) L(\mathbf{s}|\mathbf{m}),$$

where

$$L(\mathbf{m}|\mathbf{n}) = \prod_{k=1}^{K} \{D_k(\theta) + I_k(\theta)\}^{m_k} \{1 - D_k(\theta) - I_k(\theta)\}^{n_k - m_k}$$

and

$$L(\mathbf{s}|\mathbf{m}) = \prod_{k=1}^{K} \left\{\frac{D_k(\theta)}{D_k(\theta) + I_k(\theta)}\right\}^{s_k} \left\{\frac{I_k(\theta)}{D_k(\theta) + I_k(\theta)}\right\}^{m_k - s_k}.$$

The quantity $L(\mathbf{s}|\mathbf{m})$ is the likelihood conditional on $\mathbf{m}$ and does not involve $\mathbf{n}$. Because of the low incidence of the disease in general, $n_k \gg m_k$. In this case, $L(\mathbf{m}|\mathbf{n})$ can be approximated by the likelihood of the Poisson distribution; i.e.

$$L(\mathbf{m}|\mathbf{n}) \cong \prod_{k=1}^{K} \mathrm{e}^{-n_k (D_k(\theta) + I_k(\theta))} \{D_k(\theta) + I_k(\theta)\}^{m_k}.$$

Hence, the full likelihood may be written as

$$L(\mathbf{s}, \mathbf{r}|\mathbf{n}) \cong \prod_{k=1}^{K} \mathrm{e}^{-n_k D_k(\theta)} D_k(\theta)^{s_k} \mathrm{e}^{-n_k I_k(\theta)} I_k(\theta)^{r_k}, \tag{3.2}$$

which is the same likelihood as the product of independent Poisson distributions. By solving the maximum likelihood equations, we can estimate the unknown parameters including the hazard rate $\{\eta_k\}$ over the admissible intervals, screening sensitivity, and prevalence of the disease. In view of (3.2), the estimation for θ can also be obtained by minimizing the chi-square objective function

$$C(\theta|\mathbf{n}) = \sum_{k=1}^{K} \left\{\frac{(s_k - n_k D_k(\theta))^2}{n_k D_k(\theta)} + \frac{(r_k - n_k I_k(\theta))^2}{n_k I_k(\theta)}\right\}, \tag{3.3}$$

with respect to θ.

3.2 *Generalized least squares*

An alternative estimation method is to use GLS. Note that periodically observed data from early-detection trials can be viewed from the perspective of longitudinal polytomous response data. The data are periodically observed over K screening intervals, as long as the subject is still at risk. Once a subject is diagnosed to have the disease, she is not at risk for the subsequent screening intervals. Note also that the probability that an individual is screen-detected at the kth examination (or an interval case) depends on her previous screening history, which is reflected from the probability functions D_k and I_k.

At the kth screening interval there will be n_k subjects entering the interval. Recall the definition of (s_k, r_k), we define $y_k = \sum_{i=1}^{n_k} y_{ik} = (s_k, r_k)^T$ and $\mu_k(\theta) = E(y_k) = (n_k D_k(\theta), n_k I_k(\theta))^T$. By our assumptions, the probability of being a screen- or interval-detected case is the same for all subjects within the same screening interval. Conditional on $\mathbf{n}$, (s_k, r_k) are random variables with

$$E(s_k) = n_k D_k(\theta), \quad E(r_k) = n_k I_k(\theta).$$

The weighted generalized least squares objective function has the form

$$\mathrm{GLS}(\theta|\Omega) = \sum_{k=1}^{K} (y_k - \mu_k(\theta))^T \Omega_k^{-1} (y_k - \mu_k(\theta)),$$

where

$$\Omega_k(\theta) \equiv \mathrm{Cov}(y_k) = \begin{pmatrix} n_k D_k(\theta)(1 - D_k(\theta)) & -n_k D_k(\theta) I_k(\theta) \\ -n_k D_k(\theta) I_k(\theta) & n_k I_k(\theta)(1 - I_k(\theta)) \end{pmatrix}$$

is the variance–covariance function for y_k. When Ω_k is known, θ can be solved by minimizing the above weighted GLS objective function. The probabilities of the outcomes are functions of the sojourn time distribution, screening sensitivity, and disease prevalence.

Because of the low incidence of breast cancer for a general screening cohort, the probabilities of the disease being screen-detected or being an interval case in a screening interval, $D_k(\theta)$ and $I_k(\theta)$, are very small. The off-diagonal terms in Ω_k are nearly zero and may be neglected. The GLS objective function can then be approximately simplified to

$$\mathrm{GLS}(\theta|\Omega) \cong \sum_{k=1}^{K} \left\{ \frac{(s_k - n_k D_k(\theta))^2}{n_k D_k(\theta)(1 - D_k(\theta))} + \frac{(r_k - n_k I_k(\theta))^2}{n_k I_k(\theta)(1 - I_k(\theta))} \right\}.$$

A further simplification to the denominators of the above function by letting the terms D_k^2 and I_k^2 be zero leads to the objective function

$$\mathrm{GLS}(\theta|\Omega) \cong \sum_{k=1}^{K} \left\{ \frac{(s_k - n_k D_k(\theta))^2}{n_k D_k(\theta)} + \frac{(r_k - n_k I_k(\theta))^2}{n_k I_k(\theta)} \right\}. \tag{3.4}$$

Note that this alternative formulation leads to the same objective function as (3.3), which utilized a Poisson approximation to the likelihood function.

The complete knowledge for the components of Ω_k is not available, since D_k and I_k are functions of the unknown parameter θ. A natural approach is to substitute for the variance–covariance matrix Ω_k by its consistent estimator, $\hat{\Omega}_k$ (Carroll *et al.*, 1988). Note that the variance–covariance matrix consists only of functions of $D_k(\theta)$ and $I_k(\theta)$, and

$$\hat{D}_k(\theta) = \frac{s_k}{n_k} \quad \text{and} \quad \hat{I}_k(\theta) = \frac{r_k}{n_k}$$

are consistent estimators of $D_k(\theta)$ and $I_k(\theta)$. Minimizing the objective function (3.4) may be eased by using $\hat{D}_k$ and $\hat{I}_k$ instead of D_k and I_k in the weight function; i.e.

$$\mathrm{GLS}(\theta|\hat{\Omega}) \cong \sum_{k=1}^{K} \left\{ \frac{(s_k - n_k D_k(\theta))^2}{s_k} + \frac{(r_k - n_k I_k(\theta))^2}{r_k} \right\}. \tag{3.5}$$

Thus, we have three ways of estimating θ: (1) maximize the likelihood function, (2) maximize the Poisson approximation to the likelihood function, and (3) minimize the GLS utilizing consistent estimates of the weights.

4. SIMULATIONS

We conducted a series of simulation studies to investigate the performance of the various estimation approaches with the flexible piecewise model specification. We carried out 2000 simulations for all scenarios, each consisting of 50 000 or 100 000 screening individuals at the initiation of the screening program (see Tables 1 and 2). Sojourn times were generated from three different piecewise-constant densities. The scenarios specify the piecewise-constant probabilities over three 1-year intervals as (0.1, 0.4, 0.5), (0.1, 0.3, 0.6), and (0.05, 0.25, 0.7). The corresponding piecewise hazard rates for the three scenarios are (0.1, 0.444, 1), (0.1, 0.333, 1), and (0.1, 0.263, 1), respectively. The mean sojourn time for each scenario is 1.9, 2, and 2.15 years, respectively. Screening sensitivity was set at $\beta = 80\%$. For each simulation cohort, we used three equal screening intervals with a length of 1 year. The stable disease model was assumed with an underlying annual incidence of disease of 0.01.

The simulation results show that the estimators for prevalence, P, and sensitivity, β, have small bias and small standard errors. The estimators for the hazard rates are virtually consistent as shown in Table 2, but the fluctuations are large, especially for the cohort size of 50 000 in Table 1. The maximum likelihood estimators are obtained from the two likelihood functions (3.1) and (3.2), as indicated in Tables 1 and 2. They are comparable with the weighted GLS estimators as expected. With a cohort size of 50 000, the MLEs are more likely to converge to the boundary. Although the cohort sizes seem large, the actual information for estimation is from the individuals who are designated and diagnosed to have the disease over the follow-up period. This is approximately 1% of the cohort size.

In addition, we carried out a small simulation study to check the robustness of the piecewise-constant density model compared with a specific parametric model for the sojourn time distribution. The setup

Table 1. *Results from* 2000 *simulations with cohort size* 50 000

Estimation approach	Estimated parameters (SE)			
True values	$P = 0.0190$	$\beta = 0.8$	$\eta_1 = 0.1$	$\eta_2 = 0.444$
GLS	0.019 (0.001)	0.801 (0.041)	0.095 (0.074)	0.410 (0.330)
MLE	0.019 (0.001)	0.801 (0.040)	0.097 (0.075)	0.412 (0.336)
MLE (Poisson)	0.019 (0.001)	0.799 (0.041)	0.095 (0.074)	0.411 (0.332)
True values	$P = 0.020$	$\beta = 0.8$	$\eta_1 = 0.1$	$\eta_2 = 0.333$
GLS	0.020 (0.001)	0.799 (0.038)	0.091 (0.066)	0.332 (0.297)
MLE	0.020 (0.001)	0.799 (0.038)	0.092 (0.067)	0.336 (0.305)
MLE (Poisson)	0.020 (0.001)	0.797 (0.037)	0.089 (0.066)	0.338 (0.303)
True values	$P = 0.022$	$\beta = 0.8$	$\eta_1 = 0.05$	$\eta_2 = 0.263$
GLS	0.022 (0.001)	0.801 (0.030)	0.050 (0.046)	0.241 (0.229)
MLE	0.022 (0.001)	0.801 (0.030)	0.050 (0.046)	0.253 (0.229)
MLE (Poisson)	0.021 (0.001)	0.803 (0.031)	0.051 (0.046)	0.248 (0.232)

Table 2. *Results from* 2000 *simulations, cohort size* 100 000

Estimation approach	Estimated parameters (SE)			
True values	$P = 0.019$	$\beta = 0.8$	$\eta_1 = 0.1$	$\eta_2 = 0.444$
GLS	0.019 (0.001)	0.800 (0.034)	0.096 (0.066)	0.419 (0.294)
MLE	0.019 (0.001)	0.799 (0.033)	0.094 (0.066)	0.424 (0.295)
MLE (Poisson)	0.019 (0.001)	0.799 (0.033)	0.095 (0.066)	0.416 (0.289)
True values	$P = 0.020$	$\beta = 0.8$	$\eta_1 = 0.1$	$\eta_2 = 0.333$
GLS	0.020 (0.001)	0.798 (0.030)	0.091 (0.059)	0.339 (0.269)
MLE	0.020 (0.001)	0.797 (0.030)	0.089 (0.059)	0.350 (0.270)
MLE (Poisson)	0.020 (0.001)	0.798 (0.030)	0.090 (0.060)	0.341 (0.275)
True values	$P = 0.022$	$\beta = 0.8$	$\eta_1 = 0.05$	$\eta_2 = 0.263$
GLS	0.022 (0.001)	0.801 (0.023)	0.048 (0.041)	0.254 (0.201)
MLE	0.022 (0.001)	0.800 (0.023)	0.048 (0.042)	0.262 (0.205)
MLE (Poisson)	0.022 (0.001)	0.801 (0.024)	0.049 (0.042)	0.258 (0.208)

Table 3. *Comparisons of estimates: under true and assumed models with* $\beta = 0.8$ *(values in parentheses are standard errors)*

True sojourn times	Estimation based on the assumed models				
	Exponential		Piecewise-constant		
	β	μ	β	η_1	η_2
Exponential	0.810	1.193	0.797	0.617	0.577
$\mu = 1.2$ ($\eta_1 = \eta_2 = 0.6$)	(0.047)	(0.086)	(0.067)	(0.090)	(0.588)
Piecewise-constant	0.987	2.031	0.801	0.049	0.264
$\eta_1 = 0.05$, $\eta_2 = 0.263$	(0.007)	(0.043)	(0.026)	(0.046)	(0.211)

regarding the prevalence and cohort size is similar to that in Table 1. In particular, we generated a cohort of 50 000 screening participants receiving three annual examinations with an exponential sojourn time distribution having a mean of 1.2 years, or equivalently with a constant hazard rate of 0.6. We fitted the data with both the piecewise-constant density and parametric models. The results are summarized in Table 3. All the estimators are consistent. As expected, the variation of the parameters from the piecewise model is larger compared with the estimators obtained from the correctly specified parametric model. Alternatively, we also generated a data set ($n = 50\,000$) with the sojourn time distribution following a piecewise step function. We fitted the data assuming an exponential sojourn time model. Unlike the robust piecewise-constant density estimates, the parametric estimators and the screening sensitivity are subject to a severe bias resulting from model misspecification. For example, the estimator for sensitivity was 0.987 compared to the true value of 0.80; the estimated mean sojourn time was 2.03 years (or equivalent to a constant hazard rate of 0.5) compared to the true value of 2.15 years for the mean. It is clear that a misspecified model may result in estimates having considerable bias.

5. Examples

We applied the estimation methods to three published breast cancer early-detection trials for illustration. The HIP trial was initiated in the 1960s. Approximately 62 000 women, 40–64 years of age at entry, were randomly allocated to a control or screening study group. Women in the study group were invited for an initial screening, and then offered three additional annual screening examinations which included

independent clinical breast examinations plus mammograms. To apply the estimation methods to the HIP trial, we used the observed data from the first 3 years of follow-up after the initiation. The estimated annual incidence of breast cancer from the control group is about two cases per thousand women. Assuming a piecewise-constant density for sojourn time distribution, the overall sensitivity of the screening examination is estimated to be 0.71 with a standard error of 0.13 (the standard error was estimated from 300 bootstrap simulations). The estimated mean sojourn time was 2 years with 20% probability that the duration in S_p is less than 1 year using the weighted GLS estimate. The estimators are close to those estimators obtained from the model with an exponential sojourn time distribution assumption: the screening sensitivity of the mammogram with a clinical breast examination was estimated to be 0.70 (0.20), and the estimate of the mean sojourn time was 2.5 (1.2) years (Shen and Zelen, 1999).

We also applied the estimation procedures to the two CNBSS. The first trial recruited women aged 40–49 years at study entry, and the second study was restricted to women aged 50–59 years at the entry. A total of 25 214 women in the study arm for ages 40–49 years were offered four or five annual mammogram and clinical breast examinations. In the second study, a total of 19 711 women in the study arm aged 50–59 years were offered annual mammogram and clinical breast examinations. The screening interval was 1 year for the both trials. With the flexible model assumption for the sojourn time distribution, we estimated the overall sensitivity and their standard errors of the screening examinations for the younger and older groups to be 0.84 (0.16) and 0.74 (0.08), respectively, using data including the initial prevalent round in the screening arm. The corresponding estimates for the mean sojourn times are 2 and 3.3 years, respectively, for the two cohorts. Moreover, for the younger women the probability that their preclinical duration in S_p is less than 1 year is 37% (95% CI: 0%, 82%) whereas such a probability for the older cohort is 2.2% (95% CI: 0%, 36%).

6. DISCUSSION

Most models which describe the natural history of cancer and the properties of the screening examinations are based on parametric assumptions for the sojourn time distribution in the preclinical phase of the disease. In contrast, our approach allows a more flexible model for the sojourn time distribution, while still allowing the joint estimates of the sojourn time distribution, screening sensitivity, and disease prevalence. Moreover, the linear constraints on the piecewise-constant density function can be converted to rectangular constraints on the hazard rates. The computation demands are eased when the hazard function is estimated instead of the cumulative probability function of the preclinical duration of the disease. Our simulations also indicate that a piecewise assumption of the pdf for the preclinical duration of the disease leads to a robust estimate for screening sensitivity.

In addition to the maximum likelihood method, we have considered the GLS method to estimate the parameters of interest. For the periodic screening data, it may be more appropriate to use the term 'transition model' to describe the screening data, since the probabilities of D_k and I_k at the kth screening interval are conditional on the past screening history. Using the estimated weights from the observed data directly may improve the efficiency of the GLS estimators compared to the ordinary least square estimators. The consistency of the estimates is well known for settings using the weighted GLS. In view of the fact that terms such as $D_k(\theta)$ and $I_k(\theta)$ are close to zero, we have derived an asymptotic equivalence between the likelihood function with its chi-square approximation and weighted GLS objective function.

One trade-off for a more flexible model is the possible requirement for larger sample sizes in order to obtain well-behaved estimators. The estimators are more likely to converge to the boundaries when the number of parameters increases. In our model, the number of parameters increases when the number of screening intervals increases. It is true for both the MLE and GLS methods. This is also a problem as noted by DeGruttola and Lagakos (1989) and Etzioni and Shen (1997). In fact, the problem of converging

to boundaries (or multiple local maxima) is common in high-dimensional optimization. One primary motivation for proposing the GLS method is to offer an alternative procedure when the maximum likelihood estimators converge to the boundary. Based on our exploratory simulations, among the maximum likelihood estimators which converge to the boundary, there are about 10–15% cases in which the weighted GLS estimators do not converge to the boundary. Finally, our simulations show that misspecified models could result in fairly biased estimators, whereas specifying a piecewise-constant sojourn time distribution leads to robust estimation. The application of the proposed method focuses on breast cancer screening trials, but the general statistical methods are applicable to periodic screening trials of other chronic diseases.

In the current work, the parameters are estimated under the stable disease model assumption. This assumption makes the problem of estimation more amenable. In practice, the stable disease model may serve as a reasonable approximation to the actual model for relatively short screening horizons as in the HIP and CNBSS trials. These randomized breast early-detection trials had screening intervals of 1 year and took place within the first 3–4 years of the initiation of the study. During the first few years of the trial, the cohort incidence and prevalence do not change very much. In one of our earlier papers (Shen and Zelen, 1999), we investigated the nonstable disease model with a parametric sojourn time distribution for the HIP trial. The parameters of the model were very close to the estimators which assumed a stable disease model.

While acknowledging the possibility of age dependence on the screening sensitivity and sojourn time distribution, estimates in the current paper can be regarded to represent the average screening sensitivity and sojourn time distributions over the entire age range. One way to incorporate age information into the estimation, while also allowing the model flexibility, is to perform subset analysis within each age cohort. However, small sample sizes in the subset analysis can lead to unreliable estimates of the relevant quantities, especially under the flexible model assumption in this work. In another investigation, we have modeled the age effect on screening sensitivity and sojourn time distribution in breast cancer screening using a more structured model (Cong *et al.*, 2005).

Acknowledgments

This research was partially supported by the National Institutes of Health and U01 CA-88270. The authors would like to thank the editor and referees for their constructive comments.

References

Carroll, R. J., Wu, C. F. J. and Ruppert, D. (1988). The effect of estimating weights in weighted least squares. *Journal of the American Statistical Association* **83**, 1045–1054.

Cong, X., Shen, Y. and Miller, A. B. (2005). Estimation of age-specific sensitivity and sojourn time in breast cancer screening studies. *Statistics in Medicine* (in press).

Day, N. E. and Walter, S. (1984). Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. *Biometrics* **40**, 1–13.

DeGruttola, V. G. and Lagakos, S. W. (1989). Analysis of double censored survival data, with application to AIDS. *Biometrics* **45,** 1–11.

Etzioni, R. and Shen, Y. (1997). Estimating asymptomatic duration in cancer: the AIDS connection. *Statistics in Medicine* **16**, 627–644.

Louis, T. A., Albert, A. and Heghinian, S. (1978). Screening for the early detection of cancer—III. Estimation of disease natural history. *Mathematical Biosciences* **40**, 111–144.

MILLER, A. B., BAINES, C. J., TO, T. AND WALL, C. (1992a). Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. *Canadian Medical Association Journal* **147**, 1459–1476.

MILLER, A. B., BAINES, C. J., TO, T. AND WALL, C. (1992b). Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. *Canadian Medical Association Journal* **147**, 1477–1488.

PARMIGIANI, G. AND SKATES, S. (2001). Estimating distribution of age of the onset of detectable asymptomatic cancer. *Mathematical and Computer Modelling* **33**, 1347–1360.

PINSKY, P. F. (2001). Estimation and prediction for cancer screening models using deconvolution and smoothing. *Biometrics* **57**, 389–395.

SCHWARTZ, M. (1978). An analysis of the benefits of serial screening for breast cancer based upon a mathematical model of the disease. *Cancer* **41**, 1550–1564.

SHAPIRO, S., VENET, W., STRAX, P. AND VENET, L. (1988). *Periodic Screening for Breast Cancer: The Health Insurance Plan Project and its Sequelae, 1963–1986*. Baltimore, MD: The Johns Hopkins University Press.

SHEN, Y. AND ZELEN, M. (1999). Parametric estimation procedures for screening programmes: stable and nonstable disease models for multimodality case finding. *Biometrika* **86**, 503–515.

ZELEN, M. (1993). Optimal scheduling of examinations for the early detection of disease. *Biometrika* **80**, 279–293.

ZELEN, M. AND FEINLEIB, M. (1969). On the theory of screening for chronic disease. *Biometrika* **56**, 601–614.

[*Received December 16, 2004; revised April 21, 2005; accepted for publication April 26, 2005*]