Biometrika (1999), **86,** 3, *pp.* 503–515

Printed in Great Britain

Parametric estimation procedures for screening programmes: Stable and nonstable disease models for multimodality case finding

BY YU SHEN

M. D. Anderson Cancer Center, The University of Texas, Houston, Texas 77030, U.S.A.
yushen@odin.mdacc.tmc.edu

AND MARVIN ZELEN

Department of Biostatistics, Harvard School of Public Health, and Dana Farber Cancer Institute, Boston, Massachusetts 02115, U.S.A.
zelen@jimmy.harvard.edu

SUMMARY

This paper develops methods for estimating the parameters associated with early detection programmes. Disease is considered to have three states: a disease-free state or a state in which the disease cannot be detected, a preclinical state and a clinical state. The natural history of the disease is assumed to be progressive. The parameters to be estimated are the sensitivity of one or two disease detection modalities and the characteristics of the preclinical sojourn time distribution under both the stable disease and nonstable disease models. The stable-disease model assumes that the incidence or prevalence of a disease is independent of age or chronological time, while the nonstable disease model allows these quantities to depend on time. With the nonstable disease model, the relevant parameters can be jointly estimated by a two-step iteration procedure from the likelihood function. For the stable disease model, the sensitivity and the parameters of the sojourn time distribution of the preclinical state can be obtained directly from a conditional likelihood function. Applications are made to recent clinical trials for the early detection of breast cancer.

Some key words: Conditional likelihood; Multiple screening examination; Nonstable disease model; Sensitivity; Sojourn time; Stable disease model.

1. INTRODUCTION

There is increasing interest in detecting chronic diseases early with the expectation that earlier diagnosis combined with therapy will result in more cures and longer survival. As a result many of these trials have been carried out in cancer sites, especially breast, colon and lung cancers. These trials generate data which can be used to estimate the sensitivity of the examinations, the sojourn time distribution of the preclinical state and other characteristics of the screening cohort. Estimates of these parameters are important in planning public health programmes to extend proven benefit to large populations as well as in designing future early detection trials.

Various investigations have been carried out to model the screening process and to

estimate the relevant quantities; see e.g. Hutchinson & Shapiro (1968), Zelen & Feinleib (1969), Prorok (1976a, b), Louis, Albert & Heghinian (1978), Walter & Day (1983), Day & Walter (1984), Parmigiani (1993, 1997), Baker & Chu (1990), Zelen (1993), Etzioni & Shen (1997), Hu & Zelen (1997) and Lee & Zelen (1998). However, these investigations typically focus on a single screening modality. Most assume a stable disease model. The stable disease model assumption can considerably simplify the procedure of estimating both the sojourn-time distribution and the sensitivity. In many situations, however, such an assumption is only an approximation or may not be valid.

Some screening clinical trials involve more than one screening modality. For example, the study participants in the Health Insurance Plan of Greater New York (HIP) study (Shapiro et al., 1988) received both a physical examination and a mammogram each year; likewise the trials performed by the Canadian National Breast Screening Studies (CNBS) program (Miller et al., 1992a, b) also used both the mammogram and physical examination in the study groups. It is important to estimate the individual sensitivities of the mammogram and physical examination. However, little attention has been given to statistical methods for estimating the sensitivity of each modality when screening programmes offer more than one modality.

In this paper, we consider more than one screening modality under both the stable and nonstable disease models. For the stable disease model we utilise the conditional likelihood by using a Poisson approximation to the full likelihood. The use of the conditional likelihood eases the computational complexities of solving for maximum likelihood estimates by eliminating a nuisance parameter. A recent article by Straatman, Peer & Verbeek (1997) has also proposed the use of the conditional likelihood to estimate the mean lead time and the sensitivity under the stable disease model. Their estimation procedure is concerned with a single modality. Under the nonstable disease model, we propose a two-step iterative procedure to estimate the sensitivities of the screening exams, sojourn time distribution and age-specific transition probabilities for entering the preclinical state.

This paper is organised as follows. The general data structure and notation are defined in § 2. Section 3 formulates the estimation procedures for the screening sensitivities and sojourn time distribution under the stable disease model for two screening modalities. In § 4, a two-step semiparametric estimation procedure is proposed for the nonstable disease model. Section 5 summarises the simulation results for various model assumptions. The methods are illustrated in § 6 using data from three breast cancer screening trials, one from the Health Insurance Plan study and two from the Canadian National Breast Screening studies. Section 7 concludes with a discussion.

2. Model assumptions and notation

Consider a cohort of initially asymptomatic individuals, who are enrolled in a screening programme to detect the presence of a specific disease. The natural history of the disease for an individual will be modelled as consisting of three states, denoted by S_0, S_p and S_c. State S_0 refers to the disease-free state or a state in which disease cannot be detected; S_p refers to the preclinical disease state; and S_c refers to the clinically diagnosed disease state. An underlying assumption is that the natural history of the disease is progressive; i.e. the transitions are $S_0 \rightarrow S_p \rightarrow S_c$.

Let β be the sensitivity of a screening examination. Define $w(t)\,dt$ to be the probability of transition from S_0 to S_p during $(t, t+dt)$, and let $I(t)\,dt$ be the probability of transition from S_p to S_c within the interval $(t, t+dt)$. In addition, let $q(t)$ be the probability density

function of the sojourn time in S_p, and $Q(t) = \int_t^\infty q(x)\,dx$. We shall denote by θ the parameter vector associated with $Q(t)$.

Let $t_0 < t_1 < \ldots < t_{k-1} < T$ represent k ordered screening examination times, and let T denote the follow-up time past the time of the last examination. These quantities can refer to chronological time or age. Define the ith interval $[t_{i-1}, t_i)$ for $i = 1, 2, \ldots, k$, where $t_k = T$ and $\Delta_i = t_i - t_{i-1}$. Thus, the left-hand boundary is an examination time for each interval. We adopt the following notation: n_i is the number of individuals examined at t_{i-1}, s_i is the number of cases detected at the exam given at t_{i-1}, and r_i is the number of cases diagnosed in the clinical state (S_c) within (t_{i-1}, t_i). The latter cases are called interval cases. Note that the subscripts (n_i, s_i, r_i) refer to the ith interval.

When there are two screening modalities, three types of case can be identified by screening examinations during the course of the programme. For illustration consider the use of both a mammogram and a physical examination to detect breast cancer. Cases may be diagnosed by a mammogram only, modality 1, by a physical examination only, modality 2, and by both a mammogram and physical examination. At the ith screening examination, let s_{i1}, s_{i2} and s_{i3}, where $s_i = s_{i1} + s_{i2} + s_{i3}$, denote the number of cases detected by modality 1 only, by modality 2 only and by both modalities, respectively. Define α_1, α_2 and α_3 to be the probabilities of these mutually exclusive events. Then the overall sensitivity of the screening programme is $\beta = \alpha_1 + \alpha_2 + \alpha_3$. Define β_j to be the sensitivity of modality j. Then the individual sensitivities of mammography and physical examination are $\beta_1 = \alpha_1 + \alpha_3$ and $\beta_2 = \alpha_2 + \alpha_3$.

3. Stable disease model

In this section, we derive the likelihood function under the stable disease model. The examination times refer to chronological time. Under the stable disease model, it is assumed that the prevalence is constant over time. When the proportion of preclinical and clinical cases in a given population remains essentially constant over time, the corresponding transition probability function is also a constant, that is, $w(t) = w$ and $I(t) = I$, with respect to chronological time. Thus, the following well-known relation between incidence I, point prevalence P and the mean duration in S_p holds:

$$P = \mu I, \quad I = w,$$

where $\mu \equiv \mu(\theta) = \int_0^\infty Q(x)\,dx$ is the mean sojourn time in S_p. The stable disease model may serve as a suitable approximation to the observed phenomenon for relatively small changes in chronological time. While the prevalence and incidence functions may be age dependent, we can let P and I represent the average prevalence and incidence functions over the entire age distribution of a cohort.

Consider the ith screening interval $[t_{i-1}, t_i)$. Let $D_i(\beta)$ be the probability of an individual diagnosed at the ith scheduled examination given at t_{i-1}, and let $I_i(\beta)$ be the probability of an interval case occurring in the ith interval. The data obtained from studies with multiple screening modalities can be used to estimate the individual sensitivities. The full likelihood contribution from the ith screening interval is proportional to

$$L_i = \{D_i(\beta)\}^{s_i}\{I_i(\beta)\}^{r_i}\{1 - D_i(\beta) - I_i(\beta)\}^{n_i - s_i - r_i} \prod_{j=1}^{3} \left(\frac{\alpha_j}{\beta}\right)^{s_{ij}}, \tag{3·1}$$

for $i = 1, \ldots, k$. Since $D_i(\beta)$ and $I_i(\beta)$ are relatively small and n_i is large, we consider the limit where, as $n_i \to \infty$, $D_i(\beta) \to 0$ and $I_i(\beta) \to 0$, such that $n_i D_i(\beta) \to \rho_i$ and $n_i I_i(\beta) \to \lambda_i$.

Then the likelihood can be approximated by the product of two independent Poisson distributions, i.e.

$$L_i \simeq e^{-\lambda_i-\rho_i}(\rho_i+\lambda_i)^{s_i+r_i}\left(\frac{\rho_i}{\rho_i+\lambda_i}\right)^{s_i}\left(\frac{\lambda_i}{\rho_i+\lambda_i}\right)^{r_i}\prod_{j=1}^{3}\left(\frac{\alpha_j}{\beta}\right)^{s_{ij}}.$$

The above likelihood factors into the product of the marginal distribution of observing (s_i+r_i) diseased cases within the ith interval and the conditional likelihood of detecting s_i cases at the scheduled examination given a total number of diseased cases, $m_i=s_i+r_i$. The remaining term in the product is the likelihood of the individual modes of detection conditional on s_i cases detected at the scheduled examination. Hence, given the total number of cases m_i in the ith interval, the conditional likelihood of an individual being diagnosed in $[t_{i-1}, t_i)$ is proportional to

$$L_i(\alpha_1,\alpha_2,\alpha_3,\theta\,|\,m_i)=\frac{\rho_i^{s_i}\lambda_i^{m_i-s_i}}{(\lambda_i+\rho_i)^{m_i}}\prod_{j=1}^{3}\left(\frac{\alpha_j}{\beta}\right)^{s_{ij}}. \tag{3·2}$$

The functional forms for $D_i(\beta)$ and $I_i(\beta)$ depend on the specific model assumptions. Under the stable disease model, Zelen (1993) derived the following expressions for $D_i(\beta)$ and $I_i(\beta)$:

$$D_i(\beta)=\begin{cases}\beta P\{1-\beta\sum_{j=1}^{i-1}(1-\beta)^{i-j-1}Q_0(t_{i-1}-t_{j-1})\} & (i>1),\\ \beta P & (i=1),\end{cases} \tag{3·3}$$

$$I_i(\beta)=P\left[\frac{\Delta_i}{\mu}-\beta\sum_{j=0}^{i-1}(1-\beta)^{i-j-1}\{Q_0(t_{i-1}-t_j)-Q_0(t_i-t_j)\}\right], \tag{3·4}$$

where $Q_0(t)=\mu^{-1}\int_t^\infty Q(t)\,dt$. Note that

$$\frac{\rho_i}{\rho_i+\lambda_i}=\frac{D_i(\beta)}{D_i(\beta)+I_i(\beta)},$$

which is not a function of the prevalence, because the prevalence parameter P has been eliminated. Hence, $L(.\,|\,m_1,\ldots,m_k)=\Pi_{i=1}^{k}L_i(.\,|\,m_i)$ is dependent only on $\{\alpha_j\}$ and the parameter vector of the sojourn time distribution, θ.

With the observed data $\{(s_{i1}, s_{i2}, s_{i3}, r_i), 1\leqslant i\leqslant k\}$ for each screening interval, the conditional loglikelihood function is

$$\begin{aligned}&l(\alpha_1,\alpha_2,\alpha_3,\theta\,|\,m_1,\ldots,m_k)\\ &\quad=\sum_{i=1}^{k}\Bigg[r_i\log\{I_i(\beta)\}+s_i\log\{D_i(\beta)\}-(r_i+s_i)\log\{I_i(\beta)+D_i(\beta)\}\\ &\qquad+\sum_{j=1}^{3}s_{ij}\log(\alpha_j)-s_i\log(\beta)\Bigg].\end{aligned} \tag{3·5}$$

For a screening programme with only one screening modality, the conditional loglikelihood reduces to

$$l(\beta,\theta\,|\,m_1,\ldots,m_k)=\sum_{i=1}^{k}\Big[r_i\log\{I_i(\beta)\}+s_i\log\{D_i(\beta)\}-(r_i+s_i)\log\{I_i(\beta)+D_i(\beta)\}\Big]. \tag{3·6}$$

We can obtain the conditional maximum likelihood estimators from (3·6) under the constraints that $0 < \beta \leqslant 1$ and within the space for the parameter vector θ using a nonlinear minimisation method (Gay, 1983). With multiple screening modalities, parameter estimates from (3·5) have to satisfy an extra linear constraint, $0 < \sum \alpha_j = \beta \leqslant 1$. Such a constraint substantially complicates the computation procedure when solving for the estimators from (3·5). An alternative approach is initially to estimate the overall sensitivity β and θ based on the conditional loglikelihood of (3·6). We then replace β in (3·5) by $\hat{\beta}$ and impose the relationship of $\alpha_3 = \hat{\beta} - (\alpha_1 + \alpha_2)$, which leads to the solution $(\hat{\alpha}_1, \hat{\alpha}_2)$ from (3·5).

Data from a concurrent control group may also be used to augment the likelihood of the study group. One such control group is to have only a single examination at entry, t_0. Such a control group was used in the Canadian National Breast Screening study for the 40–49-year-old cohort; all women received an initial physical examination and have been followed for $T - t_0$ units of time. Then, the contribution to the loglikelihood is given by (3·6) with $k = 1$ and

$$I_1(\beta_2) = P\left[\frac{T - t_0}{\mu} - \beta_2\{1 - Q_0(T - t_0)\}\right], \quad D_1(\beta_2) = P\beta_2.$$

Another type of control group offers only a single exam modality following the same schedule as the screened group. This is exactly the control group used in the Canadian National Breast Screening group for entry ages 50–59; i.e. the control group received periodic physical examinations. The resulting likelihood is the same as (3·6) except that β_2 replaces β.

Different sampling plans, follow-up schedules and population dynamics result in various likelihood functions, which are generalisations of (3·3)–(3·4). For example, the power of β in equations (3·3) and (3·4) may be expressed by the actual number of negative screening examinations instead of $(i - j - 1)$, when an individual in the cohort missed some regular examinations or participated in the screening programme after t_0. Essentially, each individual may be treated separately in the likelihood function for irregular screening schedules.

4. Nonstable disease model

In this section we formulate an estimation procedure when incidence and prevalence are age-dependent. Under the nonstable disease model, additional knowledge of the age-specific incidence of the disease is required to estimate the sensitivities and characteristics of the sojourn time in S_p. The extra quantity to be estimated under the nonstable disease model is the probability of the transition from S_0 to S_p, that is $w(t)$. In all that follows, t will refer to age. Lee & Zelen (1998) proposed a method for estimating $w(t)$ using data on age-specific incidence based on the equation

$$I(t) = \int_0^t w(x)q(t - x)\,dx, \tag{4·1}$$

which relates the point incidence function, the transition probability and the sojourn time probability density function in S_p. Incidence data are usually available in terms of equally spaced age groups. Thus, the $w(t)$ can only be estimated as a step function when there is no parametric assumption for w. For this purpose, consider m equally spaced age groups,

where the jth age group is defined by

$$A_j = \{t : x_{j-1} \leqslant t < x_j\}, \quad \delta = x_j - x_{j-1}, \quad x_0 = 0,$$

for $j = 1, 2, \ldots, m$. Define the transition probability and the interval incidence of the disease for the jth age group to be $w(t) = w_j$ for $t \in A_j$, and $I_j = \int_{x_{j-1}}^{x_j} I(t)\, dt$, respectively. Based on equation (4·1), w_j can be derived recursively by

$$I_1 = w_1[\delta - \mu\{1 - Q_0(\delta)\}]; \tag{4·2}$$

$$I_j = \mu\left(\sum_{i=1}^{j-1} w_i[Q_0\{(j-i-1)\delta\} - 2Q_0\{(j-i)\delta\} + Q_0\{(j-i+1)\delta\}] + w_j\left\{\frac{\delta}{\mu} - 1 + Q_0(\delta)\right\}\right). \tag{4·3}$$

The sojourn time distribution is assumed to be the same across different age groups. The strategy for estimating the relevant parameters including $\{w_j\}$ depends on having estimates of $\{I_j\}$.

One difficulty in estimation of the age-specific incidence of the disease from a screening programme is that, once a preclinical lesion is detected by a screening examination, it is treated and the natural history of the disease is altered. An ideal situation would be that a control group with no screening examinations can be used to estimate the underlying age-specific incidence. Some of the randomised early detection trials did utilise such a control group. However, many recent early detection trials also offered either an initial screening examination or annual physical examinations to the control group. In this case, we may use external resources such as the Surveillance, Epidemiology and End Results (SEER) registry, or other disease registry databases to estimate the underlying age-specific incidence. Estimation based on data from external resources may not be as satisfactory as having a concurrent randomised control group. Nevertheless it could provide comparable information.

Assume a screening cohort is divided into at most m groups by each participant's age at entry. For an individual in the jth group at entry, whose age is denoted by $t_0(j) \in A_j$, let $D_i(j)$ be the probability of being detected in S_p at the ith screening examination at $t_{i-1}(j)$, and let $I_i(t \mid j)$ be the probability that the individual is incident at age t after i screening examinations, where $t_{i-1}(j) < t < t_i(j)$, for $1 \leqslant i \leqslant k$ and $1 \leqslant j \leqslant m$. In order to ease the notation, we will drop the j in $t_i(j)$, but it is implicit that $t_0 = t_0(j)$ and $t_i = t_i(j)$. In particular, for a given diagnostic test with sensitivity β,

$$I_i(t \mid j) = (1-\beta)^i P_0(t_0) q_0(t - t_0) + \Phi(i-1) \sum_{l=1}^{i-1} (1-\beta)^{i-l} \int_{t_{l-1}}^{t_l} w(x) q(t-x)\, dx + \int_{t_{i-1}}^{t} w(x) q(t-x)\, dx,$$

where $\Phi(x) = 1$ if $x > 0$ and 0 otherwise, $q_0(t) = Q(t)/\mu$, $Q_0(t) = \int_t^\infty q_0(x)\, dx$, and

$$P_0(t_0) = \mu\left[\sum_{l=1}^{j-1} w_l\{Q_0(t_0 - x_l) - Q_0(t_0 - x_{l-1})\} + w_j\{1 - Q_0(t_0 - x_{j-1})\}\right]$$

is the prevalence of the jth age cohort. The probability of being incident in the screening interval (t_{i-1}, t_i) after i examinations is $I_i(j) = \int_{t_{i-1}}^{t_i} I_i(t \mid j)\, dt$. At the ith screening examin-

ation given at t_{i-1}, the probability of being detected in S_p is $D_1(j) = \beta P_0(t_0)$, for $i = 1$, and

$$D_i(j) = \beta \left\{ (1-\beta)^{i-1} P_0(t_0) Q_0(t_{i-1} - t_0) + \sum_{l=1}^{i-1} (1-\beta)^{i-l-1} \int_{t_{l-1}}^{t_l} w(x) Q(t_{i-1} - x)\, dx \right\}, \tag{4·4}$$

for $i > 1$.

It is common in screening trials for the time between two consecutive screening examinations to be set at 1 or 2 years, whereas the incidence data are grouped over a longer age interval, e.g. 5 years. Thus, without loss of generality, it is reasonable to assume that the screening intervals are shorter than or equal to the age interval δ for grouped incidence. Hence, the integration term in $I_i(t \mid j)$ can be expressed as

$$\int_{t_{l-1}}^{t_l} w(x) q(t-x)\, dx = \begin{cases} w_l \int_{t_{l-1}}^{x_l} q(t-x)\, dx + w_{l+1} \int_{x_l}^{t_l} q(t-x)\, dx & \text{if } t_{l-1} \in A_l,\ t_l \in A_{l+1}, \\ w_l \int_{t_{l-1}}^{t_l} q(t-x)\, dx & \text{if } t_{l-1},\ t_l \in A_l. \end{cases}$$

Here A_l and w_l are also implicit functions of j, that is $A_l = A_l(j)$ and $w_l = w_l(j)$.

For the jth age group in a screening cohort, the observed data are recorded by $\{(n_i(j), s_i(j), r_i(j)), 1 \leqslant i \leqslant k\}$, which denotes the number of individuals examined at $t_{i-1}(j)$, the number of screening detected cases at $t_{i-1}(j)$ and the number of interval cases in the screening interval $(t_{i-1}(j), t_i(j))$. Given the observations in each screening interval within each age cohort, the full loglikelihood function for the screened group can be derived by the procedure similar to that under the stable disease model, that is,

$$l(\beta, \theta, w) = \sum_{j=1}^{m} \sum_{i=1}^{k} [r_i(j) \log\{I_i(j)\} + s_i(j) \log\{D_i(j)\} + \{n_i(j) - r_i(j) - s_i(j)\} \log\{1 - I_i(j) - D_i(j)\}], \tag{4·5}$$

where $w = (w_1, \ldots, w_m)$. For a screening study with more than one modality, the analogous likelihood can be obtained by generalising (4·5). Note that under the nonstable disease model the conditional likelihood function will not eliminate the nuisance parameter as in the stable disease model, because $w(t)$ is no longer a constant parameter that can cancel out. Thus, the full likelihood method may be preferable since the conditioning will generally result in a loss of information relative to that contained in the full likelihood. However, under some regularity conditions this loss can be minimum (Pereira & Lindley, 1987).

It is clear that the usual approach for obtaining maximum likelihood estimates from the full likelihood function (4·5) for the unknown parameters $\{\beta, \theta, w\}$ can be difficult. Our method uses a two-step iterative algorithm. We start by obtaining an initial value of $\theta^{(0)}$ from (3·6) assuming the stable disease model. This, together with the estimates of $\{I_j\}$, enables the initial estimation of $\{w^{(0)}\}$ to be achieved recursively from (4·2) and (4·3). Then updated values of (β, θ) can be obtained from the likelihood (4·5) by replacing $\{w\}$ by $\{\hat{w}^{(0)}\}$. Such a two-step procedure will be repeated until the estimated quantities converge. Distribution-free bootstrap or jackknife methods provide a basis for the estimation of standard deviations for the estimated parameters.

5. SIMULATIONS

A series of simulation studies were conducted to evaluate the proposed estimation methods for both the stable and nonstable disease models. One thousand simulations were carried out, each consisting of 50 000 or 100 000 screening individuals at initiation of the screening programme. Sojourn times were generated from an exponential distribution with a mean sojourn time of $\mu = 2{\cdot}5$. Screening sensitivity β was set at 0·8. For each simulation cohort, two, three or four equal screening intervals were used, and the length of screening interval was selected to be one year.

Under the stable disease model with a single modality, we assumed an underlying transition probability of $w = 4$ events per 1000 subjects per year. The first examination is carried out at $t_0 = 0$. Table 1 summarises the results. The consistency of the estimators obtained from the conditional likelihood is shown by their sample means and standard deviations. The proposed estimators are virtually unbiased for both sample size cohorts, though as would be expected the standard deviations for the larger cohort are smaller than those from the cohort with 50 000 initial samples. For comparison, we also use the full likelihood method to estimate the parameters of interest. Our simulation results indicate that the estimators from the two methods are comparable, while the estimators obtained from the full likelihood have smaller variability compared to those from the conditional likelihood. On the other hand, the average number of iterations in the estimation procedure for the conditional likelihood approach is substantially decreased compared with that for the full likelihood method, 10 steps rather than 18, when using the same initial values for the parameters.

Table 1: *Results of* 1000 *simulations. Sample means and* (*standard deviations*) *of parameter estimators under stable disease model*; $\beta = 0{\cdot}8$ *and* $\mu = 2{\cdot}5$

No. of intervals	Conditional likelihood $\hat{\beta}$ (SD)	$\hat{\mu}$ (SD)	Full likelihood $\hat{\beta}$ (SD)	$\hat{\mu}$ (SD)
		$N = 50\,000$		
2	0·801 (0·127)	2·72 (1·16)	0·799 (0·092)	2·61 (0·64)
3	0·802 (0·121)	2·67 (1·01)	0·801 (0·086)	2·58 (0·52)
4	0·802 (0·114)	2·63 (0·77)	0·802 (0·082)	2·56 (0·47)
		$N = 100\,000$		
2	0·803 (0·121)	2·70 (0·64)	0·801 (0·088)	2·60 (0·59)
3	0·799 (0·091)	2·61 (0·65)	0·802 (0·060)	2·53 (0·33)
4	0·804 (0·068)	2·53 (0·34)	0·802 (0·047)	2·52 (0·24)

For the nonstable disease model, the underlying point incidence and cumulative incidence in the three consecutive screening intervals were generated by time-varying transition probabilities for both the screening and control groups. For simplicity, we assume a homogeneous age distribution for this screening cohort, and that the length of the screening interval and age interval are the same. The age-specific transition probabilities, w_i, take the values 0·02, 0·04 and 0·06 in three successive age intervals, respectively. The results of the simulation under the nonstable disease model are presented in Table 2. Overall, the results obtained from the conditional and full likelihood methods are broadly in agreement with each other, although the full likelihood method improves the precision when there is no loss of information relative to the conditional likelihood.

Table 2: *Results of* 1000 *simulations. Sample mean and standard deviation of parameter estimators under nonstable disease model with initial sample size* $N =$ 50 000

	β	μ	w_1	w_2	w_3
True value	0·80	2·50	0·02	0·04	0·06
Conditional likelihood					
Estimator	0·783	2·75	0·022	0·043	0·062
SD of estimator	0·162	0·99	0·006	0·011	0·014
Full likelihood					
Estimator	0·794	2·53	0·020	0·041	0·059
SD of estimator	0·052	0·14	0·001	0·002	0·002

6. Examples

We present two examples from published early detection randomised trials to illustrate the proposed methods. In the first of these, we apply our method to the breast cancer screening data of the Health Insurance Plan study (Shapiro et al., 1988). This study involved two screening modalities, mammogram and physical examination. For comparison, we fit the data for both the stable and nonstable disease models. The second example relates to the two Canadian National Breast Screening studies (Miller et al., 1992a, b). In the examples, the standard deviation for each estimated parameter is obtained based on 200 bootstrap samples.

Example 1. With the Health Insurance Plan study, we use data from the first four years of follow-up after the start of screening, which were published in Shapiro et al. (1988) and Shapiro (1977). Four screening examinations were given at yearly intervals. Based on Walter & Day's (1983) findings, the exponential distribution has been considered as the best-fitting distribution for the sojourn time. Under the stable disease model assumption the overall sensitivity of mammogram and physical examination, with standard deviation in parentheses, is estimated to be 0·70 (0·20), and the mean sojourn time is 2·5 (1·2) years compared to the Day–Walter estimators of 0·82 for the sensitivity and 1·7 years for the mean sojourn time. Furthermore, the individual sensitivities of mammogram and physical examinations and their standard deviations, in parentheses, are estimated to be 0·39 (0·11) and 0·47 (0·14) respectively. The finding that the sensitivity of mammograms is lower than that of physical examinations is consistent with the observation made in Shapiro et al. (1988).

Under the nonstable disease model, the age distribution within each screening interval for the observed breast cancer cases was obtained from the Health Insurance Plan database and is summarised in Table 3. We use the estimates of age-specific cumulative incidence from Lee & Zelen (1998), which are based on the Surveillance, Epidemiology, and End Results data. Note that there are five age groups and four screening intervals, and the length of the screening interval is one year, whereas the length δ of the age interval is 5 years. For simplicity, we assume that the age at entry and the age at the last screening examination fall into the same age group, since the yearly screening interval is relatively small compared to the age interval of five years. Table 4 summarises the results under the nonstable disease model using the methods in § 4. The estimated overall sensitivity of the

screening exam and the mean sojourn time, and their standard deviations, are 0·72 (0·17) and 2·2 (0·89) years. Note that the parameter estimates from the two different model assumptions are rather close to each other, but the standard deviations are reduced under the nonstable disease model. The estimated transition probability in each age group increases with age, except for the youngest age group of 40–44 years. This anomaly is well within the expected uncertainty.

Table 3: *Health Insurance Plan data. Age distribution for screening detected cases and interval cases. Data are from the Health Insurance Plan database of* 1989

	Age at entry (years)					Total number observed
	40–44	44–49	50–54	55–59	60+	
	Cases detected at examination (%)					
1st exam.	11·1	18·5	25·9	25·9	18·5	55
2nd exam.	9·7	19·4	22·6	32·3	16·1	32
3rd exam.	5·0	20·0	20·0	25·0	30·0	20
4th exam.	8·0	28·0	24·0	24·0	16·0	23
	Interval cases (%)					
(t_1, t_2)	8·0	28·0	24·0	24·0	16·0	13
(t_2, t_3)	23·1	23·1	30·8	15·4	7·7	8
(t_3, t_4)	18·2	36·4	18·2	27·3	0·0	10
(t_4, t_5)	26·9	19·2	11·5	15·4	26·9	10

Table 4: *Health Insurance Plan study. Estimated parameters and standard deviations assuming the nonstable disease model*

Parameter	β	μ	w_1	w_2	w_3	w_4	w_5
Estimate	0·72	2·20	42·4	41·0	50·6	59·8	75·6
SD of estimate	0·17	0·89	8·22	1·80	3·22	0·75	3·53

w_i is the rate per 10^5.
When calculating SD, restrict mean sojourn time between (0, 4).

Example 2. The Canadian National Breast Screening study group conducted two randomised controlled studies to evaluate the efficacy of the screening examination combining annual mammogram with physical examination for breast cancer. The first study was restricted to women aged 40–49 years at study entry and the second study registered women aged 50–59. It may be reasonable to use the stable disease model in each of these trials since the age range of each trial is restricted.

The observed cases for this first four screening intervals were obtained from Miller et al. (1992a, b). The estimators and their standard deviations are summarised in Table 5, based on the estimation approach proposed in § 3. In both age groups, the sensitivity of mammograms is higher than that for physical examinations. In the 50–59-year-old study group, the sensitivity of mammograms is about 10% higher relative to the younger groups, but the sensitivity of physical examinations is about 33% lower. Note that the standard deviations are much larger for the younger group. The estimated mean sojourn time is

substantial larger for the 50–59 year age group relative to the younger group, i.e. 3·8 years as opposed to 2·1 years. It is worth mentioning that we used data from the three screening intervals rather than from the four screening intervals for the 40–49 age group because the irregular trend at the fourth screening examination causes the conditional maximum likelihood estimator to be non-unique. In addition, in order to obtain more reliable variance estimates, we restricted the parameter space for the mean sojourn time to the interval (0, 4) years in the bootstrap estimation procedure.

Table 5: *Canadian National Breast Screening studies. Estimated parameters and standard deviations assuming the stable disease model*

	Mean soj. time (years)	Sensitivities Mam.	Phys. exam.	Overall	Mean soj. time (years)	Sensitivities Mam.	Phys. exam.	Overall
	Age group 40–49 (MP and UC)				Age group 50–59 (MP and PE)			
Parameter estimate	2·08	0·57	0·56	0·86	3·77	0·62	0·37	0·78
SD of estimate	1·50	0·15	0·14	0·22	1·28	0·10	0·07	0·13

MP, study group with annual mammograph and physical examination; UC, control group with only initial physical examination; PE, control group receiving annual physical examination.
Mean soj. time, mean sojourn time; mam., mammogram; phys. exam., physical examination.

7. Discussion

Brookmeyer & Day (1987) also considered conditional likelihood methods in the context of a case-control screening study. In fact, the concepts of conditional likelihood have been introduced and notably applied in the proportional hazard model of Cox (1975) and in 2×2 contingency tables; the method has the advantage of eliminating the nuisance parameters (Basu, 1977; Cox & Reid, 1987; Cox, 1988). In the context of screening trials, the conditional probability results in elimination of the prevalence P from the working likelihood and simplifies the computation under the stable disease model. Intuitively, the conditional likelihood function, based on only interval and prevalent cases in the whole cohort, carries most of the information needed for estimation. Moreover, the incidence of cancer in a screening cohort is rather low, and therefore inclusion of all non-diseased or non-detected cases in the analysis is not computationally efficient (Etzioni & Shen, 1997). On the other hand, the conditional likelihood is not influenced by non-compliance and makes use of only those who have been diagnosed with disease. Further study might be of interest to investigate the relative merits of full and conditional likelihoods in different model settings.

The stable disease model may serve as a good approximation to the nonstable disease model, when the transition probabilities can be approximated by a step function in which the length of the interval is equal to or larger than the time interval between the first and last scheduled examinations, that is $\delta \geqslant t_{k-1} - t_0$. Therefore, if $w(t)$ is constant in $[t_0, t_k)$, the conditional likelihood will eliminate $w(t)$ or equivalently the prevalence factor. The calculation proceeds by breaking the data into different age cohorts. Within each age cohort, the contribution of the ith interval to the conditional likelihood is the same as that for the stable disease model. Since the conditional likelihood kernel is that of a binomial distribution, combining cohorts for the same interval is the same as combining several binomial distributions with a common success probability. Therefore, combining

the cohort groups results in the same conditional likelihood as initially conditioning on each interval without taking account of the cohort groups.

It is worth pointing out that the maximum number of parameters which can be estimated is equal to the number of independent observations from the screening data. Hence in the study group with a single modality there will be k independent observations for k intervals, whereas the number of independent observations for two detection modalities is $2k$ in the conditional likelihood framework. For example, under the stable disease model, the number of parameters to be estimated is 2 for single modality, or 3 for two modalities, assuming an exponential sojourn time distribution having one parameter. In both cases it is required to have $k \geqslant 2$. Under the nonstable disease model, we have one transition probability parameter for each interval, resulting in k parameters plus one for each detection modality plus one parameter for the exponential distribution. Therefore, the number of parameters to be estimated is $(k+2)$ or $(k+3)$, and it is necessary to have a control group which supplies k additional observations so that the restriction of $2k \geqslant (k+2)$, for single modality, or $3k \geqslant (k+3)$, two modalities, should be satisfied.

The parameter estimator of β is sensitive to the model misspecification for the sojourn time distribution, based on simulations not presented here because of limitations on space. Therefore, some nonparametric estimation procedures, for example as in Etzioni & Shen (1997), may be helpful to suggest a suitable parametric model for the sojourn time distribution. Families with flexible parametric distributions, such as the generalised odds-ratio model, which includes both Weibull and log-odds ratio distributions, may be appropriate for this purpose.

ACKNOWLEDGEMENT

We are grateful to the editor and two referees for their constructive comments; also to Dr Phil Prorok, National Cancer Institute, who made the Health Insurance Plan database available to us. This research was supported in part by grants from the National Institutes of Health.

REFERENCES

BAKER, S. G. & CHU, K. C. (1990). Evaluating screening for the early detection and treatment of cancer without using a randomized control group. *J. Am. Statist. Assoc.* **85**, 321–7.

BASU, D. (1977). On the elimination of nuisance parameters. *J. Am. Statist. Assoc.* **72**, 355–66.

BROOKMEYER, R. & DAY, N. E. (1987). Two-stage models for the analysis of cancer screening data. *Biometrics* **43**, 657–69.

COX, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–76.

COX, D. R. (1988). Some aspects of conditional and asymptotic inference: A review. *Sankhyā* **50** A, 314–37.

COX, D. R. & REID, N. (1987). Parameter orthogonality and approximate conditional inference (with Discussion). *J. R. Statist. Soc.* B **49**, 1–39.

DAY, N. E. & WALTER, S. D. (1984). Simplified models of screening for chronic disease: Estimation procedures from mass screening programmes. *Biometrics* **40**, 1–13.

ETZIONI, R. & SHEN, Y. (1997). Estimating asymptotic duration in cancer: The AIDS connection. *Statist. Med.* **16**, 627–44.

GAY, D. M. (1983). A trust-region approach to linearly constrained optimization. In *Lecture Notes in Mathematics: Numerical Analysis*, Ed. A. Dold and B. Eckmann, pp. 72–105. Berlin: Springer.

HU, P. & ZELEN, M. (1997). Planning clinical trials to evaluate early detection programmes. *Biometrika* **84**, 817–30.

HUTCHINSON, G. B. & SHAPIRO, S. (1968). Lead time gained by diagnostic screening for breast cancer. *J. Nat. Cancer Inst.* **41**, 665–81.

LEE, S. & ZELEN, M. (1998). Scheduling periodic examinations for the early detection of disease: Applications to breast cancer. *J. Am. Statist. Assoc.* **93**, 1271–81.

Louis, T. A., Albert, A. & Heghinian, S. (1978). Screening for the early detection of cancer — III. Estimation of disease natural history. *Math. Biosci.* **40**, 111–44.
Miller, A. B., Baines, C. J., To, T. & Wall, C. (1992a). Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. *Can. Med. Assoc. J.* **147**, 1459–76.
Miller, A. B., Baines, C. J., To, T. & Wall, C. (1992b). Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. *Can. Med. Assoc. J.* **147**, 1477–88.
Parmigiani, G. (1993). On optimal screening ages. *J. Am. Statist. Assoc.* **88**, 622–8.
Parmigiani, G. (1997). Timing medical examinations via intensity functions. *Biometrika* **84**, 803–16.
Pereira, C. A. & Lindley, D. V. (1987). Examples questioning the use of partial likelihood. *Statistician* **36**, 15–20.
Prorok, P. C. (1976a). The theory of periodic screening. I: Lead time and proportion detected. *Adv. Appl. Prob.* **8**, 127–43.
Prorok, P. C. (1976b). The theory of periodic screening. II: Doubly bounded recurrence times and mean lead time and detection probability estimation. *Adv. Appl. Prob.* **8**, 460–76.
Shapiro, S. (1977). Evidence of screening for breast cancer from a randomised trial. *Cancer* **39**, 2772–82.
Shapiro, S., Venet, W., Strax, P. & Venet, L. (1988). *Periodic Screening for Breast Cancer. The Health Insurance Plan Project and its Sequelae, 1963–1986*. Baltimore, MD: Johns Hopkins University Press.
Straatman, H., Peer, P. G. M. & Verbeek, A. L. M. (1997). Estimating lead time and sensitivity in a screening program without estimating the incidence in the screened group. *Biometrics* **53**, 217–29.
Walter, S. D. & Day, N. E. (1983). Estimation of the duration of a pre-clinical disease state using screening data. *Am. J. Epidem.* **118**, 856–86.
Zelen, M. (1993). Optimal scheduling of examination for the early detection of disease. *Biometrika* **80**, 279–93.
Zelen, M. & Feinleib, M. (1969). On the theory of screening for chronic diseases. *Biometrika* **56**, 601–14.

[*Received December* 1997. *Revised August* 1998]