Published in final edited form as:

Biometrics. 2005 December ; 61(4): 1056–1063.

MLE and Bayesian Inference of Age-Dependent Sensitivity and Transition Probability in Periodic Screening

Dongfeng Wu^{1,*}, Gary L. Rosner², and Lyle Broemeling²

1Department of Mathematics and Statistics, Mississippi State University, Mississippi State, Mississippi 39762, U.S.A.

2Department of Biostatistics and Applied Mathematics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, U.S.A.

Summary

This article extends previous probability models for periodic breast cancer screening examinations. The specific aim is to provide statistical inference for age dependence of sensitivity and the transition probability from the disease free to the preclinical state. The setting is a periodic screening program in which a cohort of initially asymptomatic women undergo a sequence of breast cancer screening exams. We use age as a covariate in the estimation of screening sensitivity and the transition probability simultaneously, both from a frequentist point of view and within a Bayesian framework. We apply our method to the Health Insurance Plan of Greater New York study of female breast cancer and give age-dependent sensitivity and transition probability density estimates. The inferential methodology we develop is also applicable when analyzing studies of modalities for early detection of other types of progressive chronic diseases.

Keywords

Bayesian analysis; Breast cancer; Early detection; Markov chain Monte Carlo; Periodic screening; Sensitivity; Sojourn time; Transition probability

1. Introduction

A recent report by the National Institutes of Health (NIH, 2000) lists breast cancer as the most common form of cancer among women in the United States and the second leading cause of cancer deaths among women. Screening programs based on mammography and clinical breast examination have existed for a while, and the National Cancer Institute (NCI) and the American Cancer Society have made recommendations regarding the frequency of mammograms. Recently, a panel convened by the NCI caused a heated debate by choosing not to recommend that all women in their forties undergo annual mammographic exams (NIH, 1997). Thus, there is great interest in determining the relationship between the characteristics of breast cancer, aspects of a screening program that includes mammography, and a woman's age.

Some studies have suggested that a woman's age affects the sensitivity, the sojourn time distribution, and the transition probability into the preclinical stage (Dubin, 1979; Eddy, 1980; Walter and Day, 1983; Spratt, Greenberg, and Hauser, 1986; Lee and Zelen, 1998; Shen and Zelen, 1999). Sensitivity is the probability that the screening exam is positive, conditional on the individual being in the preclinical stage. By sojourn time, we mean the time from entry into the preclinical state until the manifestation of clinical symptoms. We consider the transition

*email: dwu@math.msstate.edu.

probability into the preclinical stage to be the p.d.f. of making the transition from the disease-free to the preclinical state.

Knowledge of the sensitivity of the screening modality is necessary for evaluating the predictive performance of a screening exam. The screening sensitivity may depend on a variety of factors, including age, position, location and size of the tumor, and the experience of the radiologist. Recent studies indicate that the sensitivity of mammography increases with age at diagnosis (Shapiro et al., 1988; Miller et al., 1992a,b), attributable to the fact that breast tissue tends to be more dense and fibrous in younger women and more soft and fatty in older women (Kerlikowske et al., 1996).

Although there is substantial evidence suggesting an association between screening sensitivity, the transition probability density into the preclinical state, and the age at diagnosis, few researchers have modeled such relationships or explored quantitatively the changing pattern of these parameters across different age groups. Baker (1998) and Baker et al. (2003) discuss evaluating periodic screening without modeling natural history or estimating sensitivity. Nevertheless, precise estimates of age-dependent sensitivity and the transition probability into the preclinical state should be of great value to those who set health-care policy, particularly regarding the initial age and frequency of periodic screening exams. Our goal is to gain a better understanding of the effect of age. The particular characteristics of interest are the screening modality's sensitivity and the disease transition probability. With limited data generally available from screening studies, however, estimation of these quantities is challenging.

We analyze data available from the Health Insurance Plan of Greater New York (HIP) study (Shapiro et al., 1988). The HIP study, which began at the end of 1963, was the first randomized clinical trial to examine regular screening exams that include mammography as a diagnostic screening test for breast cancer. The study randomized initially asymptomatic women aged 40 to 64 years without a history of breast cancer to two groups: the study group and the control group. Each group consisted of about 31,000 women. The screening program for the study group called for up to four annual breast cancer screening exams, with each screening exam including both a mammogram and a clinical breast exam. The control group received usual care. After 15 years of follow-up, more than 1100 (1171 vs. 1154) breast cancer cases were diagnosed in each group, with about 300 cases within the first 5 years of follow-up. In the study group, 132 cases were detected by screening exams.

Current studies have focused primarily on the women in the HIP study group during the first 5 years of follow-up. We derive the appropriate likelihood by expanding the stochastic model proposed by Zelen (1993), Lee and Zelen (1998), and Shen and Zelen (1999). Separate link functions incorporate age as a covariate for the sensitivity and the transition probability density. We use both maximum likelihood (ML) and Bayesian posterior estimation to make inferences for all model parameters. The Bayesian method provides more information than the MLEs, such as posterior density curves, measures of uncertainty, and predictive distributions, however.

The remainder of the article is organized as follows. In Section 2, we derive the likelihood function and introduce the link functions. In Section 3, we present the MLEs and the Bayesian analysis of the HIP study. We conclude with a discussion in Section 4.

2. Methods

2.1 The Model

Following a commonly used model for the natural history of cancer, we assume that the disease develops by progressing through three states, denoted by $S_0 \rightarrow S_p \rightarrow S_c$. The three states

correspond, respectively, to the disease-free state; the preclinical disease state, in which an asymptomatic individual unknowingly has disease that the screening exam can detect; and the clinical disease state, when the disease manifests itself through clinical symptoms.

Consider a cohort of initially asymptomatic women who enroll in a screening program. The sensitivity is denoted by $\beta(t)$, where t is the individual's age at the exam. Define $w(t) dt$ as the probability of a transition from S_0 to S_p during $(t, t + dt)$. Let $q(\cdot)$ be the p.d.f. of the sojourn time in S_p . Finally, let $Q(z) = \int_z^\infty q(x) dx$ be the survivor function of the sojourn time in the preclinical state S_p . Throughout this article, the time t represents the participating individual's age.

Assume that the study design calls for K ordered screening examinations that occur for a specific individual at ages $t_0 < t_1 < \dots < t_{K-1}$. If exams are annual, $t_i = t_0 + i$. Define the i th screening interval as the time interval between the i th and the $(i + 1)$ th screening exams: (t_{i-1}, t_i) , $i = 1, 2, \dots, K - 1$. We define the i th generation of individuals as those women who enter S_p during the i th screening interval. The 0th generation consists of all women who entered S_p before the initial screening exam. We define $t_{-1} \equiv 0$.

We now derive the likelihood corresponding to the HIP study group during the first 5 years of follow-up. These data have been analyzed by many other researchers (Walter and Day, 1983; Zelen, 1993; Lee and Zelen, 1998; Shen and Zelen, 1999; Shen, Wu, and Zelen, 2001). We grouped the women according to their age at study entry, truncated to the age at the last birthday. There are 25 such subgroups, corresponding to initial ages between 40 and 64, inclusive. There were four annual screening exams. For the i th screening exam ($i = 1, \dots, K = 4$) and the cohort of women whose age at study entry was t_0 ($t_0 = 40, 41, \dots, 64$), we define n_{i,t_0} to be the total number of individuals in this cohort examined at the i th screening; s_{i,t_0} is the number of cases detected at the i th screening exam; and r_{i,t_0} is the number of cases diagnosed in the clinical state S_c within the interval (t_{i-1}, t_i) . The latter cases are called interval cases.

Let D_{k,t_0} be the probability of a diagnosis at the k th scheduled annual exam (at age $t_{k-1} = t_0 + k - 1$) for a woman whose age at study entry was t_0 , conditional on being in state S_p . The probability that an individual in S_p is detected at the first scheduled exam (i.e., $k = 1$) at age t_0 is

$$D_{1,t_0} = \beta(t_0) \int_0^{t_0} w(x) Q(t_0 - x) dx. \quad (1)$$

The integral in the equation arises because she must have entered the preclinical state S_p before t_0 and remained in that state at least until t_0 .

Consider an i th generation individual who was detected at the k th screening exam ($1 \leq i < k$), which she underwent at age t_{k-1} . There are two cases: either she passed her previous $(k - i - 1)$ exams undetected and had a sojourn time of at least $(t_{k-1} - x)$, where $x \in (t_{i-1}, t_i)$ is her age at entry into state S_p ; or she entered S_p in the $(k - 1)$ th screening interval (t_{k-2}, t_{k-1}) . Hence the probability is

$$\begin{aligned}
D_{k, t_0} = & \beta(t_{k-1}) \left\{ \sum_{i=0}^{k-2} [1 - \beta(t_i)] \cdots \right. \\
& [1 - \beta(t_{k-2})] \int_{t_{i-1}}^{t_i} w(x) Q(t_{k-1} - x) dx \\
& \left. + \int_{t_{k-2}}^{t_{k-1}} w(x) Q(t_{k-1} - x) dx \right\}, \\
& \text{for all } k = 2, \dots, K.
\end{aligned} \tag{2}$$

Let $I_{k, t_0}(t) dt$ be the probability that an individual enters state S_c at age within $(t, t + dt)$, where $t \in (t_{k-1}, t_k)$. If this woman was in the i th generation, $i < k$, then she must have gone undetected in her $(k - i)$ previous screening exams and have a sojourn time $(t - x)$, where x is the entry age into state S_p . Alternatively, she may have entered S_p after the k th exam and developed clinical disease at time t . Hence,

$$\begin{aligned}
I_{k, t_0}(t) = & \sum_{i=0}^{k-1} [1 - \beta(t_i)] \cdots [1 - \beta(t_{k-1})] \int_{t_{i-1}}^{t_i} w(x) q(t - x) dx \\
& + \int_{t_{k-1}}^t w(x) q(t - x) dx, \quad \text{for } \forall t \in (t_{k-1}, t_k).
\end{aligned} \tag{3}$$

The probability of being incident in the k th screening interval (t_{k-1}, t_k) is

$$\begin{aligned}
I_{k, t_0} = & \int_{t_{k-1}}^{t_k} I_{k, t_0}(t) dt \\
= & \sum_{i=0}^{k-1} [1 - \beta(t_i)] \cdots [1 - \beta(t_{k-1})] \\
& \times \int_{t_{i-1}}^{t_i} w(x) [Q(t_{k-1} - x) - Q(t_k - x)] dx \\
& + \int_{t_{k-1}}^{t_k} w(x) [1 - Q(t_k - x)] dx, \\
& \text{for all } k = 1, \dots, K.
\end{aligned} \tag{4}$$

According to Shen and Zelen (1999), the likelihood for women aged t_0 at study entry is proportional to

$$\begin{aligned}
L(\cdot \mid t_0) = & \prod_{k=1}^K D_{k, t_0}^{s_{k, t_0}} I_{k, t_0}^{r_{k, t_0}} \\
& \times (1 - D_{k, t_0} - I_{k, t_0})^{n_{k, t_0} - s_{k, t_0} - r_{k, t_0}}.
\end{aligned} \tag{5}$$

The whole likelihood for the study group is the product of the age-specific contributions across all ages, given in (5) with $K = 4$,

$$L = \prod_{t_0=40}^{64} L(\cdot \mid t_0) = \prod_{t_0=40}^{64} \prod_{k=1}^4 L_k(\cdot \mid t_0).$$

2.2 The Age-Dependent Sensitivity and Transition Probability Density Functions

We extend previous models by formulating screening sensitivity and the transition p.d.f. from S_0 to S_p as functions of age in the likelihood.

As mentioned, sensitivity seems to increase with age (Shapiro et al., 1988; Miller et al., 1992a,b; Shen and Zelen, 1999), and it is limited to (0, 1). A mathematically convenient model for associating sensitivity $\beta(t)$ with age t is the logistic model,

$$\beta(t) = \frac{1}{1 + \exp(-b_0 - b_1 * (t - \bar{t}))},$$

where $\bar{t}$ is the average age at entry in the study regular group. If $b_1 > 0$, $\beta(t)$ will be a monotone increasing function of age t .

The transition density function $w(t)$ is the instantaneous probability of a transition from S_0 to S_p ; that is, $w(t) dt$ is the probability of entering S_p from S_0 during age interval $(t, t + dt)$. The integral $\int_0^\infty w(t) dt$ represents the lifetime risk of entering the preclinical state. According to the NCI's SEER database (Ries et al., 2002), a woman's lifetime risk of being diagnosed with breast cancer (invasive and in situ) is 15.7%, which is less than or equal to a woman's lifetime risk of entering the preclinical disease state. We chose 20% as a reasonable upper bound.

Lee and Zelen (1998) estimated the transition probabilities $w(t)$, using age-specific incidence rates for the years 1988–1992 from the NCI's SEER program. The results were summarized in age intervals of 5 years. We plotted the step function based on their estimates of $w(t)$. The function appeared right skewed and could be fit very well by a log-normal distribution. Therefore, we chose

$$w(t) \Big| \mu, \sigma^2 = \frac{0.2}{\sqrt{2\pi\sigma t}} \exp \left\{ -(\log t - \mu)^2 / (2\sigma^2) \right\} t > 0,$$

which is the density of log-normal (μ, σ^2) multiplied by 20%.

Walter and Day (1983) modeled the sojourn time in the preclinical state with an exponential distribution. They estimated incidence, sensitivity, and sojourn time without any dependence on age. The exponential distribution has only one parameter and methods based on it are rather sensitive to modest departures in the tail (Cox and Oakes, 1984). We adopted the log-logistic distribution for the sojourn time. This distribution has analytic forms for the main functions (Cox and Oakes, 1984). The survivor function and the hazard function are, respectively,

$$Q(x) = \frac{1}{1 + (x\rho)^\kappa}, \quad h(x) = \frac{\kappa x^{\kappa-1} \rho^\kappa}{1 + (x\rho)^\kappa}, \quad \kappa > 0, \quad \rho > 0,$$

where x is the sojourn time in the state of S_p . The density is $q(x) = h(x)Q(x)$. This density is right skewed. An advantage of this family over others is the relatively simple form for $Q(x)$ and $h(x)$. The first moment can be calculated from $E(X) = (\pi / \rho\kappa) \csc(\pi / \kappa)$. For the r th moment to exist, $\kappa > r$ is needed. We assumed that the mean sojourn time does not change with age as have most researchers.

3. Application to the HIP Study

We now describe our analysis of the HIP study data based on the likelihood function derived in Section 2. Our model includes six unknown parameters $\theta = (b_0, b_1, \mu, \sigma^2, \kappa, \rho)$.

3.1 Likelihood Model Checking and the MLE

Although the six parameters in the likelihood have a domain of either $(-\infty, \infty)$ or $(0, \infty)$ theoretically, plausible values for them are limited to finite intervals in reality. We now discuss possible ranges for the parameters.

We know that the sensitivity is around 0.7 from the literature (Shapiro et al., 1988; Shen and Zelen, 1999). It should be at least 0.5 to make screening useful. That is, $0.5 \leq \beta(t) < 1$ for all $t = 40, 41, \dots, 64$. Solving for $t = 40$ and $t = 64$, using 0.5 and 0.99 as the lower and the upper bounds, gives $b_0 \in (0, 5)$ and $b_1 \in (-0.2, 0.2)$.

A log-normal r.v. has mean $\exp(\mu + \sigma^2/2)$. We put an upper limit of 70 on the mean, believing that the average age of a transition from S_0 to S_p should be younger than 70 years old. This gives one constraint: $\mu + \sigma^2/2 \leq 4.25$. On the other hand, the maximum value of a log-normal density is at $t^* = \exp(\mu - \sigma^2)$. The mode t^* should lie within the age interval (30, 70), leading to another constraint: $3.40 \leq \mu - \sigma^2 \leq 4.25$. Combining these gives $\mu \in (3.97, 4.25)$ and $\sigma^2 \in (0, 0.57)$. We chose intervals (3.5, 4.5) and (0, 1) for μ and σ^2 , respectively.

For the sojourn time's log-logistic distribution, the mean exists if and only if $\kappa > 1$. When the mean exists, its density curve is unimodal and reaches maximum at $x^* = (1/\rho)(\kappa - 1/\kappa + 1)^{1/\kappa}$. As κ increases, the density becomes more concentrated around x^* , and both the mean and the x^* approaches $1/\rho$. From the literature (Day and Walter, 1984; Lee and Zelen, 1998; Shen and Zelen, 1999), the mean sojourn time is around 2 years, varying from a few months to about 5 years. Setting $0.5 < 1/\rho < 10$ gives $0.1 < \rho < 2.0$. The scale parameter κ controls the spread of the density. With κ bigger than 5, there is almost no mass beyond 4 years. We chose $1 < \kappa < 5$.

We carried out simulation studies to confirm that our model fitting routines and algorithms worked with this complicated likelihood. We found the resulting MLEs to be consistent with the input parameters.

3.2 MCMC and the Prior

We used Markov chain Monte Carlo (MCMC) to generate a random sample from the joint posterior distribution of the parameters for Bayesian inference. We partitioned the posterior simulation into four subchains, sampling the posterior for (b_0, b_1) , μ , σ^2 , (κ, ρ) separately. The Appendix provides more detail about our implementation of MCMC.

We chose a noninformative bivariate normal prior for (b_0, b_1) (i.e., mean vector (0, 0) and variance equal to 10^{10} times the identity matrix). We also gave μ a noninformative normal prior, namely, $N(0, 10^{10})$. The prior for σ^2 was uniform (0, 1), and κ was uniform (1, 5) a priori. The prior distribution for ρ was uniform (0, 2).

The two integrals $\int_a^b w(x) dx$ and $\int_a^b w(x)Q(t-x) dx$ in the likelihood (equations 1, 2, and 4) do not have an analytical form. We used the trapezoidal rule to evaluate them when calculating the likelihood.

We ran the MCMC for 30,000 steps, with a burn-in of 10,000 iterations. After the burn-in time, we sampled the posterior every 20 steps, giving 1000 posterior samples for the parameter vector θ . We ran two parallel chains, each with different starting values that were overdispersed with respect to the target distribution. Bayesian output analysis showed convergence. We pooled the 1000 saved posterior samples from each chain for the analysis.

3.3 Results

The ML and Bayesian posterior estimates of θ for the HIP data are shown in Table 1. We include bootstrap estimates of the standard errors of the MLEs. The posterior mean and median are very close to the MLEs, which we expected given our somewhat vague priors. The posterior standard errors are bigger than that of the bootstrap estimates, which is reasonable because Bayesian inference incorporates prior uncertainty. Also, incidence of breast cancer is small in each age group, so the bootstrap method might underestimate the variances.

Sensitivity appears to increase with age. The trend is shown in Figure 1, which shows age-specific posterior quantiles based on the posterior samples of the parameters. From ages 40 to 65 years, the posterior mean sensitivity increased from 0.603 to 0.875 and the posterior standard error dropped from 0.236 to 0.144. The posterior error of sensitivity was largest for women 40 years old, as one might expect since the sensitivity was closest to 0.5 for this age group. At the average age of 51.6 years, the posterior mean sensitivity was 0.779, with posterior standard error 0.186. For comparison, the estimated sensitivity for the HIP study in Shapiro et al. (1988) was 0.737, defined as the ratio of the screen-detected cases (132) divided by the total number of cancer cases (screen-detected and interval cases $132 + 47$) during the first 5 years of follow-up. This result is close to the posterior mean sensitivity we estimated when we fit a model without an age effect on sensitivity: posterior mean = 0.74 and 90% credible interval (0.43, 0.97).

The posterior median of the age-dependent transition probabilities from S_0 to S_p are also shown in Figure 1. The posterior medians ranged from 1.388×10^{-3} to 2.735×10^{-3} over the ages examined. The transition probabilities do not appear monotone with age; the function appears to peak around age 60 years.

Figure 2 shows the bootstrap quantiles for sensitivity and the transition probability based on the MLEs by age. Comparing Figures 1 and 2, we find that the 90% bootstrap confidence intervals for age-specific sensitivity and transition probabilities are narrower than the corresponding 90% Bayesian posterior intervals. The bootstrap plot is more suggestive of a trend in sensitivity, owing to smaller estimated standard errors. The bootstrap method might underestimate the variances, however, because there are only a few breast cancer cases in each age group.

We computed the posterior probability of a positive slope: $P(b_1 > 0 \mid \text{HIP data}) = 0.904$. A likelihood ratio test for the slope gives a P value of 0.092. Thus, there is some evidence for an age effect on sensitivity in the HIP data.

The posterior mean sojourn time was 1.88 years, and the posterior median was 1.51 years. The posterior standard deviation of the sojourn time was large, 1.65 years. The MLE for the mean sojourn time was 1.78 years.

3.4 Goodness-of-Fit Test

We applied Pearson's χ^2 goodness-of-fit test to evaluate the fit of the model to the HIP data. Let $\hat{s}_{i,t_0}$ and $\hat{r}_{i,t_0}$ be the expected number of screen-detected cases and interval cases when fitting our model. The values $\hat{s}_{i,t_0}$ and $\hat{r}_{i,t_0}$ are calculated from the MLE $\hat{\theta}$ by letting $\hat{s}_{i,t_0} = n_{i,t_0} \times \hat{D}_{i,t_0}$ and $\hat{r}_{i,t_0} = n_{i,t_0} \times \hat{I}_{i,t_0}$. The test statistic

$$\sum_{t_0=40}^{64} \sum_{i=1}^4 \left[\frac{(s_{i,t_0} - \hat{s}_{i,t_0})^2}{\hat{s}_{i,t_0}} + \frac{(r_{i,t_0} - \hat{r}_{i,t_0})^2}{\hat{r}_{i,t_0}} \right],$$

is approximately χ^2 distributed with d.f. = $4 \times 2 \times 25 - 6 = 194$. The test statistic is 193.85, with a P value of 0.49. Hence, the model is consistent with the HIP data.

As suggested by an associate editor, we further investigated the robustness of the result for other reasonable models. We fit a Weibull model for the age-dependent transition probability density. The resulting goodness-of-fit test P value was 0.48. The estimates were close to those with our original model. We also added a quadratic term of age to the logistic model for

sensitivity. The resulting χ^2 statistic was 191.28 on 193 degrees of freedom. The quadratic term is not statistically significant, based on a difference between χ^2 statistics of 2.57 on 1 d.f. ($P = 0.11$). We prefer the more parsimonious model.

We also changed the prior distribution for b_0 , b_1 , and μ . Instead of the normal prior, we considered a uniform prior over the reasonable ranges established in Section 3.1. The posterior estimates with both priors were very close to each other, except for a positive shift of the intercept b_0 for sensitivity. We think the difference relates to the fact that the normal prior was centered at 0, whereas the uniform prior put support only between 0 and 5. Based on the above analyses, we think our results for the HIP data are fairly robust to other reasonable model assumptions.

4. Discussion

We described a model for analyzing data arising from a study of periodic screening exams with sensitivity and the transition probability characterized as functions of age. Applying the model to data from the HIP study, we found an increasing relationship between age and sensitivity. The probability of a transition into the preclinical state was not monotone with age, peaking near age 60 years. We reported a fully Bayesian analysis, along with MLE. We prefer Bayesian inference. It provides more information than ML estimation, through posterior density curves, measures of uncertainty, and predictive distributions.

We estimated the standard errors of the MLEs with the bootstrap. The bootstrap standard errors were generally smaller than the Bayesian posterior standard errors. In particular, the plot of sensitivity as a function of age based on the MLEs was more suggestive of a trend in sensitivity with increasing age, because of the smaller standard errors. The bootstrap method might underestimate the variances, however, given the small number of breast cancer cases in each age group. The larger posterior variance may be appropriate, given the many sources of uncertainty.

As pointed out by a reviewer, the incidence of breast cancer in the population increases more steeply before age 50 years than after that. The change is thought to reflect age-related hormonal changes. It may be that the change we saw in the transition probabilities with age reflects a similar pattern, suggesting that the same physiologic changes alter the transition to S_p . Such a relationship is worth further investigation.

Shen and Zelen (1999) presented two models they called stable and nonstable. The stable model assumed constant transition probabilities across all ages (i.e., $w(t) \equiv w$). Their nonstable model considered $w(t)$ to be a step function of age, with $w(t)$ constant within each 5-year age group (40–44, 45–49, etc.). They assumed constant sensitivity and mean sojourn time across all ages in their models. They analyzed the HIP data and estimated sensitivity under the stable model to be 0.70 (standard error 0.20). The estimated sensitivity with the nonstable model was 0.72 (standard error 0.17). The nonstable disease model requires more assumptions, however. For example, a person's age at the initial and the last screening exams should fall into the same age group. Their stable estimate of the mean sojourn time was 2.5 years (standard error 1.2 years). With the nonstable model, the estimate was 2.2 years, with a standard error of 0.89 years. Our results seem closer to their nonstable results, except that the posterior standard error is larger, reflecting greater uncertainty.

Lee and Zelen (1998) used SEER information (NCI, 1994) of breast cancer incidence to estimate $w(t)$, assuming the sojourn time is exponentially distributed with a mean of 4 years. Our estimated transition probabilities are larger than theirs, as might be expected, since we estimated lower sensitivities among younger women. This is the first time that the HIP data have been used directly to obtain estimates for the transition probability into S_p .

Chen et al. (2000) provided a method for estimating the mean sojourn time without using data from interval cases. They analyzed screening data from Taiwan. Their estimated mean sojourn time was 1.90 years, with a 95% CI (1.18, 4.86), which is compatible with our result. They assumed that the transition probabilities were exponentially distributed and 100% sensitivity, which they deemed unrealistic.

Straatman, Peer, and Verbeek (1997) estimated the mean lead time (in fact, the mean sojourn time) and sensitivity without estimating the incidence probability. Analyzing the Nijmegen data in 1975–1987, they estimated the mean sojourn time to be 1.3 years among women younger than 50 years old and 2.2 years in the age 50–65 group, both with large confidence intervals. Their sensitivity estimates were 0.92 and 1.00, respectively, for the two age groups.

Walter and Day (1983) did not include a dependence on age when they estimated incidence, sensitivity, and sojourn time from the HIP data. They modeled the sojourn times in the preclinical state as exponentially distributed. They found a high marginal correlation between sensitivity and the parameter for the sojourn time distribution. This high marginal correlation may reflect a common dependence on age. We are examining this further.

Some individuals have questioned the benefit of screening for breast cancer among women younger than 50 years. Based on the HIP study, we inferred lower sensitivity for women at younger ages than at older ages, although the posterior precision of the estimated sensitivity was also lower at the younger ages. The posterior credible intervals for sensitivity were rather large across all ages. Our estimates clearly relate to the state of things back in the 1960s, when the HIP study took place. Other large studies of breast cancer screening have been performed since this landmark study. Using all available data, more precise and informative inference could be obtained.

Acknowledgements

The authors would like to thank the editor, the associate editor, and an anonymous reviewer for their valuable comments and suggestions. This research was supported in part by National Institutes of Health grant CA-79466.

References

- Baker SG. Evaluating the age to begin periodic breast cancer screening using data from a few regularly scheduled screenings. *Biometrics* 1998;54:1569–1578. [PubMed: 9883553]
- Baker, S. G., Erwin, D., Kramer, B. S., and Prorok, P. C. (2003). Using observational data to estimate an upper bound on the reduction in cancer mortality due to periodic screening. *BMC Medical Research Methodology* Available at <http://www.biomedcentral.com/1471-2288/3/4>
- Chen THH, Kuo HS, Yen MF, Lai MS, Tabar L, Duffy SW. Estimation of sojourn time in chronic disease screening without data on interval cases. *Biometrics* 2000;56:167–172. [PubMed: 10783792]
- Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data* Boca Raton, Florida: Chapman & Hall/CRC.
- Day NE, Walter S. Simplified models of screening for chronic disease: Estimation procedures from mass screening programmes. *Biometrics* 1984;40:1–13. [PubMed: 6733223]
- Dubin N. Benefit of screening for breast cancer: Application of a probabilistic model to a breast cancer detection project. *Journal of Chronic Diseases* 1979;32:145–151. [PubMed: 447774]
- Eddy, D. M. (1980). *Screening for Cancer: Theory, Analysis and Design* Englewood Cliffs, New Jersey: Prentice Hall.
- Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. *Journal of the American Medical Association* 1996;276:33–38. [PubMed: 8667536]
- Lee SJ, Zelen M. Scheduling periodic examinations for the early detection of disease: Applications to breast cancer. *Journal of the American Statistical Association* 1998;93:1271–1281.

- Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. *Canadian Medical Association Journal* 1992a;147:1459–1476. [PubMed: 1423087]
- Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. *Canadian Medical Association Journal* 1992b;147:1477–1488. [PubMed: 1423088]
- National Cancer Institute. (1994). Surveillance, Epidemiology, and End Results Program. *Age-Specific Rates for Breast Cancer* Bethesda, Maryland: National Cancer Institute.
- National Institutes of Health. (1997) *Breast cancer screening for women ages 40–49* NIH Publication No. 103, January 21–23, 1997.
- National Institutes of Health. (2000). *What you need to know about breast cancer* NIH Publication No. 00-1556, December 12, 2000.
- Ries, L. A. G., Eisner, M. P., Kosary, C. L., Hankey, B. F., Miller, A. B., Clegg, L., and Edwards, B. K. (2002). SEER Cancer Statistics Review, 1973–1999. Bethesda, Maryland: National Cancer Institute. Available at <http://seer.cancer.gov/csr/19731999/>
- 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: Johns Hopkins University Press.
- Shen Y, Zelen M. Parametric estimation procedures for screening programmes: Stable and non-stable disease models for multimodality case finding. *Biometrika* 1999;86:503–515.
- Shen Y, Wu D, Zelen M. Testing the independence of two diagnostic tests. *Biometrics* 2001;57:1009–1017. [PubMed: 11764239]
- Spratt JS, Greenberg RA, Hauser LS. Geometry, growth rates, and duration of cancer and carcinoma in situ of the breast before detection by screening. *Cancer Research* 1986;46:970–974. [PubMed: 3940657]
- Straatman H, Peer PGM, Verbeek ALM. Estimating lead time and sensitivity in a screening program without estimating the incidence in the screened group. *Biometrics* 1997;53:217–229. [PubMed: 9147591]
- Tierney L. Markov chains for exploring posterior distributions. *Annals of Statistics* 1994;22:1701–1762.
- Walter SD, Day NE. Estimation of the duration of a preclinical disease state using screening data. *American Journal of Epidemiology* 1983;118:856–886.
- Zelen M. Optimal scheduling of examinations for the early detection of disease. *Biometrika* 1993;80:279–293.

Appendix

Posterior Simulation for the Parameters in the Bayesian Model

This Appendix summarizes our Bayesian posterior sampling procedure. We estimated the model by a combination of MCMC techniques applied successively to update subsets of the parameter vector, conditional on currently imputed values for the remaining parameters.

1. We start with an initial guess $\theta = (b_0, b_1, \mu, \sigma^2, \kappa, \rho)$.
2. Let i indicate the iterations in the MCMC. Update (b_{0i}, b_{1i}) by drawing a “candidate” $(b^*_0, b^*_1) \sim N(\mu_b, \sum_b)$, where $\mu_b = (b_{0i}, b_{1i})$ is the current imputed value, and $\sum_b$ is the probing variance. Evaluate

$$r = \min \left\{ 1, \frac{\pi(b^*_0, b^*_1) L(\theta_i^* | DATA)}{\pi(b_{0i}, b_{1i}) L(\theta_i | DATA)} \right\},$$

where π is the prior density function for μ_b . Here, θ_i^* is the currently imputed vector θ_i with (b_{0i}, b_{1i}) replaced by (b^*_0, b^*_1) , and $L(\cdot | DATA)$ is the likelihood function derived in Section 2. With probability r , replace (b_0, b_1) with new (b^*_0, b^*_1) ; otherwise, keep it unchanged. This is a general Metropolis step.

Repeat step 2 $n = 20$ times and take the last value as the updated value of (b_0, b_1) .

3. Update μ_i using similar steps.
4. Update σ_i^2 using similar steps.
5. Update (κ_i, ρ_i) using an independence chain (Tierney, 1994). Draw a candidate from the prior density, that is $(\kappa_i, \rho_i) \sim \text{uniform prior}$, with the acceptance probability

$$r = \min \left\{ 1, \frac{L(\theta_i^* | DATA)}{L(\theta_i | DATA)} \right\}.$$

6. Iterate Steps 2–5. After a burn-in period, we took the chain to have practically converged. We summarized posterior distributions of desired functions of parameters based on the MCMC simulated posterior samples taken after the burn-in period.

Figure 1.
Posterior quantiles (5%, 50%, 95%) of sensitivity and transition probability.

Figure 2.
Bootstrap quantiles (5%, 50%, 95%) of sensitivity and transition probability.

Table 1
ML and Bayesian posterior estimates for the HIP data

Parameters	MLE (SE)	Bayesian posterior estimate		
		Median	Mean	SE
b_0	1.122 (0.733)	1.581	1.676	1.338
b_1	0.081 (0.042)	0.084	0.085	0.078
μ	4.331 (0.048)	4.329	4.340	0.076
σ^2	0.145 (0.046)	0.172	0.190	0.076
κ	2.205 (0.732)	2.275	2.509	0.927
ρ	0.808 (0.181)	0.917	0.886	0.287