Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–13

DOI: 10.1177/0962280212465499
smm.sagepub.com
SAGE

Estimation of sensitivity depending on sojourn time and time spent in preclinical state

Seongho Kim[1] and Dongfeng Wu[2]

Abstract

The probability model for periodic screening was extended to provide statistical inference for sensitivity depending on sojourn time, in which the sensitivity was modeled as a function of time spent in the preclinical state and the sojourn time. The likelihood function with the proposed sensitivity model was then evaluated with simulated data to check its reliability in terms of the mean estimation and the standard error. Simulation results showed that the maximum likelihood estimates of the proposed model have little bias and small standard errors. The extended probability model was further applied to the Johns Hopkins Lung Project data using both maximum likelihood estimation and Bayesian Markov chain Monte Carlo.

Keywords

Sensitivity, sojourn time, transition probability, cancer screening

1 Introduction

Early detection of cancer can greatly increase the chances for effective treatment, and one of the major components of early detection is screening. Screening is the use of tests across a healthy population to identify individuals who have the disease yet do not have symptoms. In general, the earlier the disease is being detected, the greater the odds of survival.

The disease progressive model was proposed by Zelen and Feinleib,[1] where the disease was assumed to progress through three states: $S_0 \rightarrow S_p \rightarrow S_c$, corresponding respectively to the disease-free state, the preclinical disease state when an asymptomatic individual unknowingly has the disease that a screening exam can detect, and the clinical state when the disease manifests itself in clinical symptoms. The difference $(t_2–t_1)$ is called sojourn time in the preclinical state if a person enters the preclinical state (S_p) at age t_1 and clinical symptoms present later at age t_2. If one is offered screening at time t within the interval (t_1, t_2) and the disease is diagnosed, then the length of the time $(t_2–t_1)$ is called lead time.

[1]Biostatistics Core, Karmanos Cancer Institute, School of Medicine, Wayne State University, Detroit, MI, USA
[2]Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, Louisville, KY, USA

Corresponding authors:
Seongho Kim, Wayne State University, Detroit, MI 48201, USA and Dongfeng Wu, University of Louisville, Louisville, KY 40202, USA.
Email: kimse@karmanos.org; dongfeng.wu@louisville.edu

The main objective in screening is to detect disease in the preclinical state. The screening model has three important parameters: the sensitivity of the screening modality, the sojourn time distribution, and the transition probability density from the disease-free to the preclinical state. The sensitivity is the conditional probability that a screening test is positive given that a person is in the preclinical stage. These three parameters are the building blocks for screening modeling, as all other parameters of interest can be expressed as a function of these three key parameters. Therefore, knowledge of the three key parameters, especially the sensitivity of the screening modality, is critical for evaluating the predictive performance of a screening program. The screening sensitivity may depend on a variety of factors, such as position, location and size of the tumor, experience of the radiologist, age, etc. Wu et al.[2] modeled the sensitivity as a function of age with the age-dependent transition probability density. They applied this model to the Health Insurance Plan of Greater New Yorker (HIP), a breast cancer screening study, and found that the sensitivity for mammogram increases as women's age increases, although the transition probability density from the disease-free to the preclinical state reaches its maximum around age 60. Wu et al.[3] applied the same model to a lung cancer screening study, the Mayo Lung Project (MLP) data, and found that the screening sensitivity of chest X-ray for lung cancer does not depend on age, though the transition probability density is not a monotonic function of age but has a peak around age 68. A careful study of the posterior estimates of the sensitivity and the sojourn time seems to suggest that they are negatively correlated.

Wu et al.[4] recently proposed a method to model the sensitivity as a function of age at diagnosis and the time in the preclinical state. In this model, the ratio of the time spent in the preclinical state to the sojourn time was multiplied by a logistic function of age. They found that the sensitivity was influenced more by the ratio of the time in the preclinical state to the sojourn time, than by age for breast cancer screening. This suggests that time spent in the preclinical state might overshadow the age effect in screening sensitivity. Moreover, Hori and Gambhir[5] suggested that the size of a tumor plays an important role in the screening exams. The rate of growth of a tumor may be different for a specific type of cancer, implying that the sensitivity may be varied according to the relative time spent in the preclinical state to the sojourn time. In other words, the sensitivity may be increasing rapidly right after entering the preclinical state before it reaches a fixed value. Combining the information, we treated the sensitivity as a function of the ratio of time spent in the preclinical state to the sojourn time in this study. Since the maximum sensitivity is usually less than 1 even in the clinical state, we further introduced another parameter to control the overall sensitivity.

To evaluate the likelihood function with the proposed sensitivity model, we carried out simulation studies using the maximum likelihood estimation. We then applied our model to the Johns Hopkins Lung Project data, using maximum likelihood estimation as well as Bayesian inference. All simulations were run by using the statistical software R (R Development Core Team, 2012), and the algorithms described in this study can be obtained upon request from the authors.

2 Methods

Let $t_0 < t_1 < \cdots < t_{K-1} < T$ represent K ordered screening exams starting at age t_0, where T is the fixed follow-up time after the last examination. The ith screening interval is (t_{i-1}, t_i), $i = 1, 2, \ldots, k$. The following notations are used: the ith annual screening exam occurs at the age $t_{i-1} = t_0 + i - 1$, for $i = 1, 2, \ldots, k$, and $t_{-1} = 0$; n_{i,t_0} is the total number of individuals examined at t_{i-1}; s_{i,t_0} is the number of cases diagnosed at t_{i-1}; and r_{i,t_0} is the number of interval cases within the interval (t_{i-1}, t_i).

We modeled the sensitivity to vary with the sojourn time and the time spent in the preclinical state by

$$\eta(t|S) = \left(\frac{t}{S}\right)^{\gamma}, \ \gamma \geq 0, \tag{1}$$

where S is the sojourn time, a random variable, and $t \in [0, S]$ is the time spent in the S_p. In this model, the maximum sensitivity is one when $t = S$. Since the sensitivity may not reach 100% even though the disease reaches its clinical state, we modified the model by adding another control parameter τ:

$$\eta(t|S) = \frac{1}{1+\tau}\left(\frac{t}{S}\right)^{\gamma}, \ \tau, \gamma \geq 0. \tag{2}$$

In equation (2), the overall sensitivity increases as τ goes to zero. The relationship of the sensitivity with respect to different τ and γ are displayed in Figure 1: the overall sensitivity when $\tau = 0$ is greater than that when $\tau > 0$; the sensitivity reaches its maximum at $t = S$; and the sensitivity increases as γ goes to zero for the same τ. Especially, when $\gamma > 1$, the sensitivity slowly increases

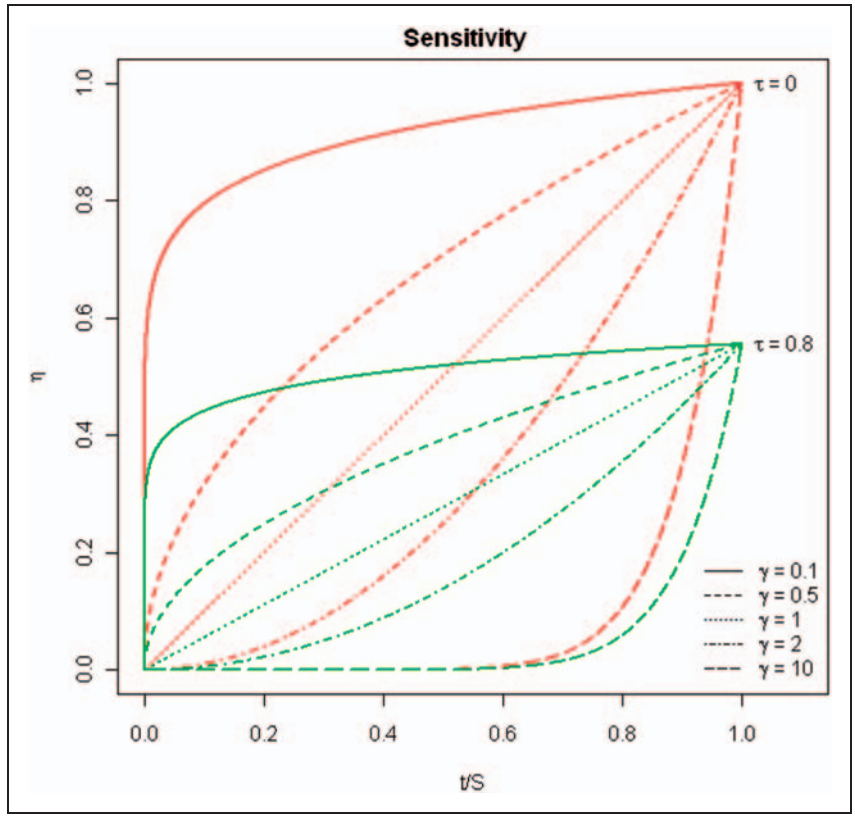

Figure 1. The behavior of the proposed sensitivity model with different τ and γ values.

from $t=0$, then sharply increases when t approaches S. On the other hand, when $\gamma < 1$, the sensitivity rapidly increases from $t=0$, then gradually flattens at $t=\mathrm{S}$.

Let D_{i,t_0} be the probability of an individual definitively diagnosed at the ith scheduled exam given at t_{i-1} and I_{i,t_0} the probability of an interval case in (t_{i-1},t_i). For $i=1,2,\ldots,K$, these two probabilities are

$$\begin{aligned} D_{i,t_0} = &\sum_{k=0}^{i-2} \int_{t_{k-1}}^{t_k} w(x) \int_{t_{i-1}-x}^{\infty} q(t) \left[\prod_{j=k}^{i-2} \{1-\eta(t_j-x|t)\} \right] \eta(t_{i-1}-x|t) dt dx \\ &+ \int_{t_{i-2}}^{t_{i-1}} w(x) \int_{t_{i-1}-x}^{\infty} q(t) \eta(t_{i-1}-x|t) dt dx, \end{aligned} \tag{3}$$

$$\begin{aligned} I_{i,t_0} = &\sum_{k=0}^{i-1} \int_{t_{k-1}}^{t_k} w(x) \int_{t_{i-1}-x}^{t_i-x} q(t) \left[\prod_{j=k}^{i-1} \{1-\eta(t_j-x|t)\} \right] dt dx \\ &+ \int_{t_{i-1}}^{t_i} w(x)[1-Q(t_i-x)] dx, \end{aligned} \tag{4}$$

and

$$D_{1,t_0} = \int_{0}^{t_0} w(x) \int_{t_0-x}^{\infty} q(t) \eta(t_0-x|t) dt dx. \tag{5}$$

In the above equations, $w(t)$ is the probability density function of making a transition from S_0 to S_p at age t and is modeled as a sub-density of a log-normal distribution,

$$w(t) = \frac{0.2}{\sigma t \sqrt{2\pi}} \exp\left(-\frac{(\log\, t-\mu)^2}{2\sigma^2} \right), \tag{6}$$

where 20% was selected based on the previous analysis on Lung cancer screening[3]; $q(t)$ is the probability density function (pdf) of the sojourn time in S_p; and $Q(z) = \int_z^{\infty} q(x) dx$ is the survivor function of the sojourn time. The log-logistic distribution was used to model the sojourn time[2]:

$$q(x) = \frac{\kappa x^{\kappa-1} \rho^{\kappa}}{(1+x^{\kappa}\rho^{\kappa})^2}, x>0, \kappa>0, \rho>0, \tag{7}$$

where κ and ρ are positive scale and location parameters in the log-logistic family. The shape of the log-logistic distribution is similar to that of the log-normal distribution, while the former has a heavier tail. The median of the log-logistic distribution is $1/\rho$, the reciprocal of the location parameter. Its mode is $\frac{1}{\rho} \cdot (\frac{\kappa-1}{\kappa+1})^{\frac{1}{\kappa}}$ if $\kappa > 1$; and 0 otherwise. In other words, when $\kappa \leq 1$, the pdf is monotonically decreasing, and when $\kappa > 1$, it is unimodal. A convenience of using log-logistic distribution is that its survivor and hazard functions have simple analytical forms.

The parameters to be estimated are $\theta = (\tau, \gamma, \mu, \sigma^2, \kappa, \rho)$ with the proposed sensitivity model in equation (2). The likelihood function for the cohort aged t_0 at study entry is

$$L_{t_0}(\theta) = \prod_{j=1}^{K} D_{j,t_0}^{s_{j,t_0}} I_{j,t_0}^{r_{j,t_0}} \left(1 - D_{j,t_0} - I_{j,t_0}\right)^{n_{j,t_0} - s_{j,t_0} - r_{j,t_0}} \tag{8}$$

where K is the number of screenings. As a result, the whole likelihood function for the study group is the product of the age-specific contributions across all age groups

$$L(\theta) = \prod_{t_0} L_{t_0}(\theta). \tag{9}$$

3 Simulation results

Simulations were carried out to evaluate the reliability of the likelihood function with the proposed sensitivity model in terms of the accuracy of parameter estimation. First, we assumed that there were initially 100,000 individuals in each age group from age 40 to 70 and each individual took part in the annual screening for eight consecutive years. Next, screening data were generated using input values of θ, and then, using the likelihood function combined with the proposed sensitivity models, the maximum likelihood estimates (MLEs) of θ were computed using these simulated screening data. This procedure was repeated 250 times for each input values of θ, then the mean, the empirical standard error, and the bias of each parameter were calculated.

In all simulations, the true input values for $(\mu, \sigma^2, \kappa, \rho)$ were set to (4.33, 0.15, 2.21, 0.81). As for the parameters (τ, γ), we considered $\gamma = 0.5$, 1.0, and 2.0, and $\tau = 0$, 0.2 and 0.8, resulting in nine cases. Numerical solutions for MLE were found using the R package *nlminb*: the true values were used as initial guesses of the parameters, and iterative evaluation was continued until convergence was reached within an error tolerance of 10^{-8} for the likelihood function and the parameters.

Table 1 displays the results of the proposed sensitivity model: the mean, the empirical standard error, and the bias of the MLE for each parameter. The biases for (μ, σ^2, ρ) are quite small in all cases and their empirical standard errors are less than 0.03. The empirical standard error of κ is about 0.3. The empirical standard error and the bias of the MLE of γ become smaller as its true value becomes smaller. Namely, Cases 3, 6, and 9 have empirical standard errors of 0.39, 0.44, and 0.60 and biases of 0.18, 0.03, and 0.20, while Cases 1, 4, and 7 have empirical standard errors of 0.12, 0.17, and 0.22 and biases of −0.04, 0.02, and 0.20. Overall, the MLEs were consistent with the true input values of the parameters. The simulation study shows that the likelihood function and the modeling are reliable in the estimation of the parameters.

4 Application to the Johns Hopkins lung project data

We applied our method to the Johns Hopkins Lung Project. The model included six unknown parameters $\theta = (\tau, \gamma, \mu, \sigma^2, \kappa, \rho)$. The parameter estimation was performed using both the maximum likelihood estimation and Bayesian Markov Chain Monte Carlo (MCMC).

4.1 Johns Hopkins lung project data

The designs of the Johns Hopkins Lung Project (JHLP) can be found in the literature.[6] JHLP trials enrolled 10,386 men in the Baltimore metropolitan area between 1973 and 1978, aged 45 years and

Table 1. Simulation results of MLE of the proposed sensitivity model. The true value of the parameters $(\mu,\sigma^2,\kappa,\rho)$ is (4.33, 0.15, 2.21, 0.81) and the true value of (τ,γ) is given in the first column. The total number of simulation runs is 250 for each case. The SE stands for the empirical standard error

		τ	γ	μ	σ^2	κ	ρ
	True			4.33	0.15	2.21	0.81
Case 1	Mean ± SE	0.05 ± 0.10	0.46 ± 0.12	4.32 ± 0.02	0.15 ± 0.01	2.20 ± 0.27	0.82 ± 0.04
(0.00,0.50)[a]	Bias	0.05	–0.04	–0.01	0.00	0.01	0.01
Case 2	Mean ± SE	0.05 ± 0.09	1.05 ± 0.21	4.31 ± 0.03	0.14 ± 0.01	2.13 ± 0.23	0.82 ± 0.02
(0.00,1.00)	Bias	0.05	0.05	–0.02	–0.01	–0.08	0.01
Case 3	Mean ± SE	0.05 ± 0.11	2.11 ± 0.39	4.31 ± 0.03	0.14 ± 0.01	2.19 ± 0.35	0.80 ± 0.02
(0.00, 2.00)	Bias	0.05	0.18	–0.02	–0.01	–0.02	–0.01
Case 4	Mean ± SE	0.21 ± 0.18	0.52 ± 0.17	4.33 ± 0.03	0.15 ± 0.01	2.23 ± 0.37	0.81 ± 0.04
(0.20, 0.50)	Bias	0.01	0.02	0.00	0.00	0.02	0.00
Case 5	Mean ± SE	0.22 ± 0.19	1.07 ± 0.29	4.32 ± 0.03	0.15 ± 0.01	2.19 ± 0.33	0.81 ± 0.02
(0.20, 1.00)	Bias	0.02	0.07	–0.01	0.00	–0.02	0.00
Case 6	Mean ± SE	0.25 ± 0.24	2.03 ± 0.44	4.33 ± 0.04	0.15 ± 0.02	2.19 ± 0.39	0.80 ± 0.03
(0.20, 2.00)	Bias	0.05	0.03	0.00	0.00	–0.02	–0.01
Case 7	Mean ± SE	0.90 ± 0.38	0.52 ± 0.22	4.33 ± 0.04	0.15 ± 0.01	2.21 ± 0.47	0.81 ± 0.03
(0.80, 0.50)	Bias	0.10	0.02	0.00	0.00	0.00	0.00
Case 8	Mean ± SE	0.81 ± 0.46	1.10 ± 0.35	4.34 ± 0.05	0.15 ± 0.02	2.21 ± 0.51	0.81 ± 0.03
(0.80, 1.00)	Bias	0.01	0.10	0.01	0.00	0.00	0.00
Case 9	Mean ± SE	0.78 ± 0.50	2.20 ± 0.60	4.34 ± 0.05	0.16 ± 0.02	2.23 ± 0.52	0.81 ± 0.05
(0.80, 2.00)	Bias	–0.02	0.20	0.01	0.01	0.02	0.00

MLE: maximum likelihood estimate.
[a]The true value of (τ,γ).

older at enrollment, who smoked at least one pack of cigarettes per day (or who had smoked this much within 1 year of enrollment) and who had no prior history of respiratory cancer. Then all participants were randomized to two arms: chest X-ray only or a dual screen (chest X-ray and sputum cytology) group, resulting in 5160 men in the chest X-ray only arm and 5226 in the dual-screen arm. Participants in the chest X-ray group received chest X-ray screening test annually, for eight consecutive years. If any of the tests was positive, then the screen was considered positive and a definitive work-up exam, such as biopsy, was done. The data that we used in this study were the chest X-ray group, and it included the total number of participants in each screening exam, the number of detected and confirmed cancer cases in each screening exam, and the number of interval cases. These data were stratified by age at study entry from 45 to 88 years. However, we only used the data from age 45 to age 70 excluding age groups 47, 58, 62, 68, and 69, because these five age groups had few participants and might cause large bias in the estimation.

4.2 Maximum likelihood estimation

There was no analytical solution for the MLE, so we used the R package *nlminb* to find the MLE numerically. Based on the previous analysis,[2] we set up the ranges for the parameters for the proposed model as follows: $\mu \in [3.5, 5.4]$, $\sigma^2 \in [0, 1]$, $\kappa \in [1, 5]$, and $\rho \in [0, 2]$, and the parameters of the proposed sensitivity model were $\tau \in [0, 2]$ and $\gamma \in [0, 3]$.

After the MLE was calculated from the JHLP data, bootstrap method was used to estimate the empirical standard errors, the 95% confidence intervals, and the median of the MLEs. The summary

Table 2. The maximum likelihood estimates of the parameters for the JHLP data. The total number of bootstrap runs is 1000. The SE stands for the empirical standard error. The 95% confidence intervals of quantile are reported for each parameter along with its median using the parametric bootstrap

Parameter	Estimate	SE	2.5%	50%	97.5%
τ	0.0114	0.0385	0.0000	0.0001	0.1493
γ	0.1271	0.0373	0.0267	0.1277	0.1953
μ	4.4139	0.0184	4.3644	4.4188	4.4385
σ^2	0.0506	0.0036	0.0428	0.0511	0.0566
κ	1.8153	0.5935	1.0000	1.7337	3.2855
ρ	0.0389	0.0119	0.0153	0.0400	0.0610

JHLP: Johns Hopkins Lung Project.

of the estimation is reported in Table 2. The estimates $\hat{\mu}$, $\hat{\sigma}^2$, $\hat{\kappa}$, and $\hat{\rho}$ are 4.4139, 0.0506, 1.8153, and 0.0389, respectively. The estimate of γ is 0.1271. The estimate for τ is 0.0114, close to zero, with the maximum sensitivity of 0.9887 at $t = S$. Thus, the sensitivity of the JHLP data is similar to the curve of the sensitivity with $\tau = 0$ and $\gamma = 0.1$ in Figure 1. Of these estimates, $\hat{\kappa}$ has the largest empirical standard error (0.5935), and $\hat{\rho}$ has the smallest empirical standard error (0.0119).

The estimated sensitivity, the transition probability density, and the sojourn time using the MLE are plotted in Figure 2. The solid line is the estimated curve of the MLE, and the dotted lines are the quantiles of the 95% confidence intervals from bootstrap. The sensitivity is sharply increased right after an individual enters in the preclinical state, and, when $\frac{t}{S} \geq 0.19$, the sensitivity becomes more than 0.8. The transition density has the maximum at age 78.5, which is 0.0044. The mode of the sojourn time is 13.0 years, while its median is 25.7 years.

4.3 Bayesian inference

We estimated the parameters using MCMC within Gibbs sampling in a Bayesian framework. The parameters τ, γ, σ^2, and ρ were log-transformed for computational efficiency. Non-informative normal distributions were chosen as prior distributions for τ, γ, and μ. Uniform priors were used for σ^2, κ, and ρ. The random-walk Metropolis algorithm with truncated normal kernel distribution was implemented (see Appendix).

We ran the MCMC for 3000 iterations, with a burn-in of the initial 1000 iterations, and thinning every 5th iteration afterwards. Figure 3 is the trace plots of the simulated Markov chains. The solid and dotted horizontal lines represent the mean and the median of each parameter, respectively.

Figure 4 displays the posterior density plots for each parameter. The solid and dotted vertical lines represent the posterior mean and median, respectively. The parameter γ seems to have two modes: one is close to zero and the other is close to 0.2, which may cause its posterior mean (0.1293) to be very close to its MLE (0.1271) in Table 2. Comparing the 95% credible intervals in Table 3 with the 95% confidence intervals in Table 2, all the credible intervals are overlapped with the confidence intervals except for the parameter μ. The MLE of μ is underestimated with a larger empirical standard error than that of its posterior mean. Overall, the estimates of the posterior distributions are generally consistent with the MLEs.

Figure 5 shows the estimated plots for the sensitivity, the transition probability density, and the sojourn time based on the posterior means of the parameters. The solid lines are the estimation, and

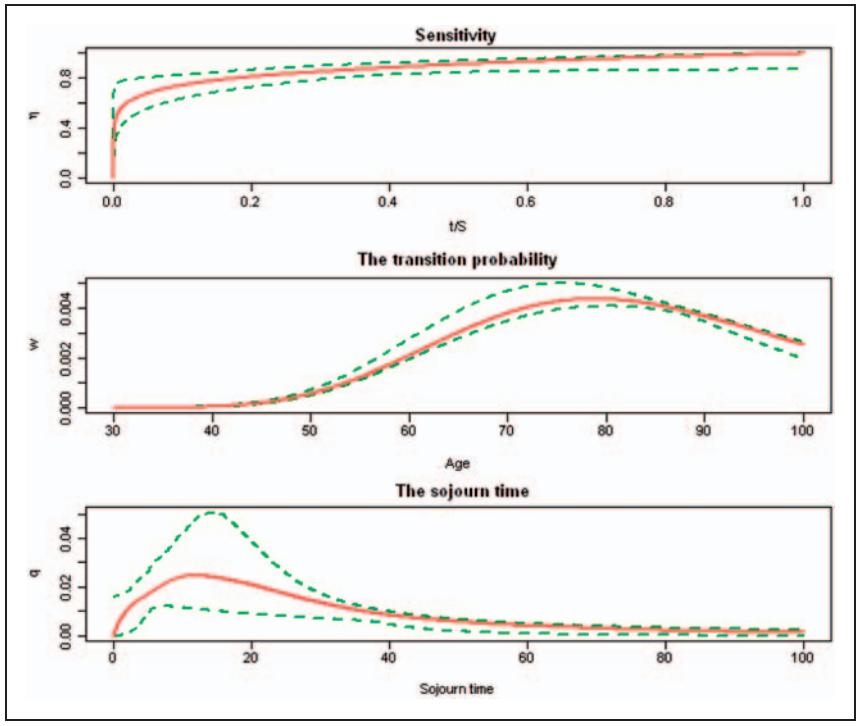

Figure 2. The prediction of the sensitivity, transition probability density, and sojourn time using maximum likelihood estimates. The red solid line is from the maximum likelihood estimates and the green dotted lines are 95% confidence curves of quantile using the parametric bootstrap.

the dotted lines are the 95% credible curves from the posteriors. The lower bound of the 95% credible interval of γ is 0.0001, causing the upper bound of the sensitivity to reach one in the very beginning of the time spent in S_p, as shown in the top plot of Figure 5. Since the posterior mean of γ is 0.1293, the sensitivity of the proposed model becomes more than 0.8 when $t/S > 0.21$. As for the transition density, the maximum occurs at age 73.8 (0.0051). The sojourn time has its mode at 15.78 years.

4.4 Goodness-of-fit test

Pearson's χ^2 goodness-of-fit test was performed to evaluate the modeling of the JHLP data by comparing the observed and the expected numbers of screen-diagnosed and interval cases. The expected numbers $\hat{s}_{t,t_0}$ and $\hat{r}_{t,t_0}$ of the screen-diagnosed cases and the interval cases were calculated by $\hat{s}_{t,t_0} = n_{t,t_0} \times \hat{D}_{t,t_0}$ and $\hat{r}_{t,t_0} = n_{t,t_0} \times \hat{I}_{t,t_0}$, where $\hat{D}_{t,t_0}$ and $\hat{I}_{t,t_0}$ are the probabilities calculated from the estimates $\hat{\theta}$. The χ^2 test statistic with the degrees of freedom (d.f.) $21 \times 8 \times 2 - |\theta|$ is

$$\sum_{t_0 \in A} \sum_{t=1}^{8} \left\{ \frac{\left(s_{t,t_0} - \hat{s}_{t,t_0}\right)^2}{\hat{s}_{t,t_0}} + \frac{\left(r_{t,t_0} - \hat{r}_{t,t_0}\right)^2}{\hat{r}_{t,t_0}} \right\} \tag{10}$$

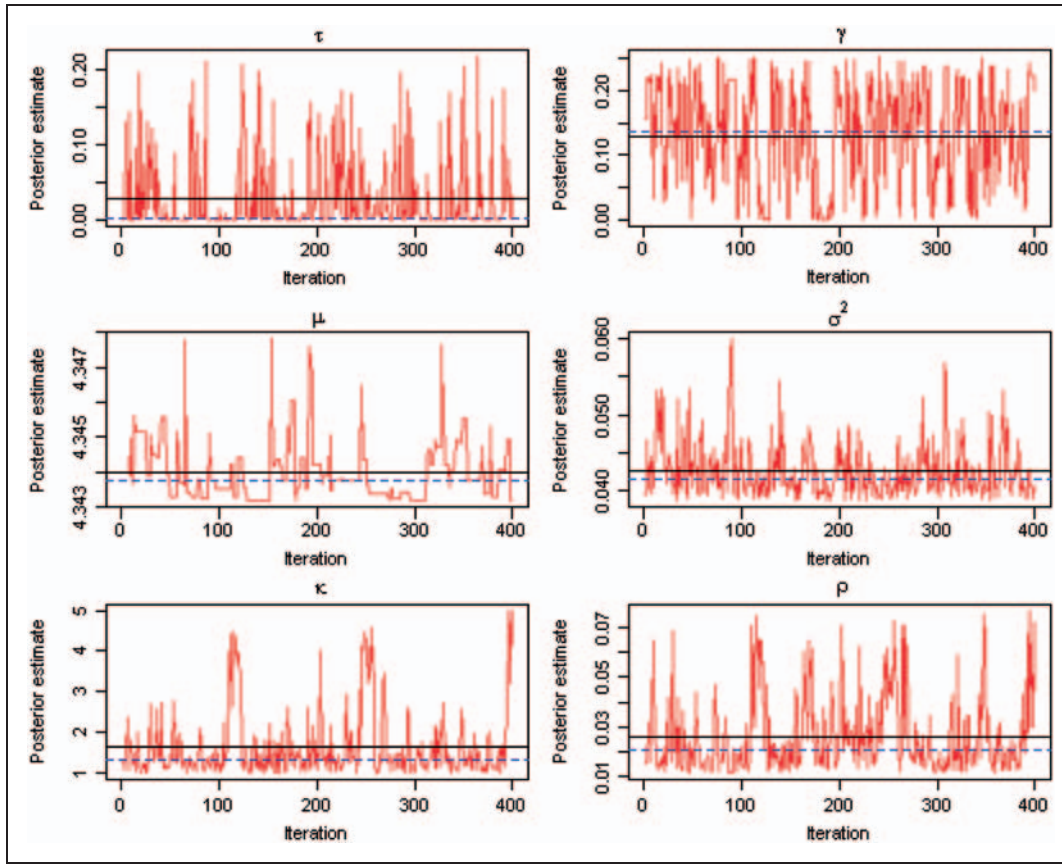

Figure 3. The trace plots of the parameters using Bayesian Markov chain Monte Carlo.

where $A = \{t|45 \leq t \leq 70\}\backslash\{47, 58, 62, 68, 69\}$ and $|\theta|$ is the number of parameters. Since there were six parameters, the d.f. of the proposed model was 330. The MLE has a p value of 0.3598 with the statistic of 338.63, while the posterior mean has a p value of 0.9541 with the statistic of 289.95, showing that the proposed model is consistent with the data.

5 Discussion

We proposed a new sensitivity model that depends on the time spent in the preclinical state and the sojourn time, then evaluated the model by simulations and the JHLP data analysis.

The estimate of the rate of the sensitivity $\hat{\gamma}$ for the proposed model in equation (2) is 0.1293, less than one for the JHLP data (see Table 3). As a result, the sensitivity rapidly increases in the beginning and gradually reaches one at the end of the preclinical state, which is similar to the sensitivity curve with $\tau = 0$ and $\gamma = 0.1$ in Figure 1. Since the preclinical state is the disease state when the size of tumor becomes large enough to be identified by screening, the sensitivity is high in the beginning of the preclinical state, demonstrating that the proposed sensitivity model is capable of describing the changes of the sensitivity according to the time spent in the preclinical state.

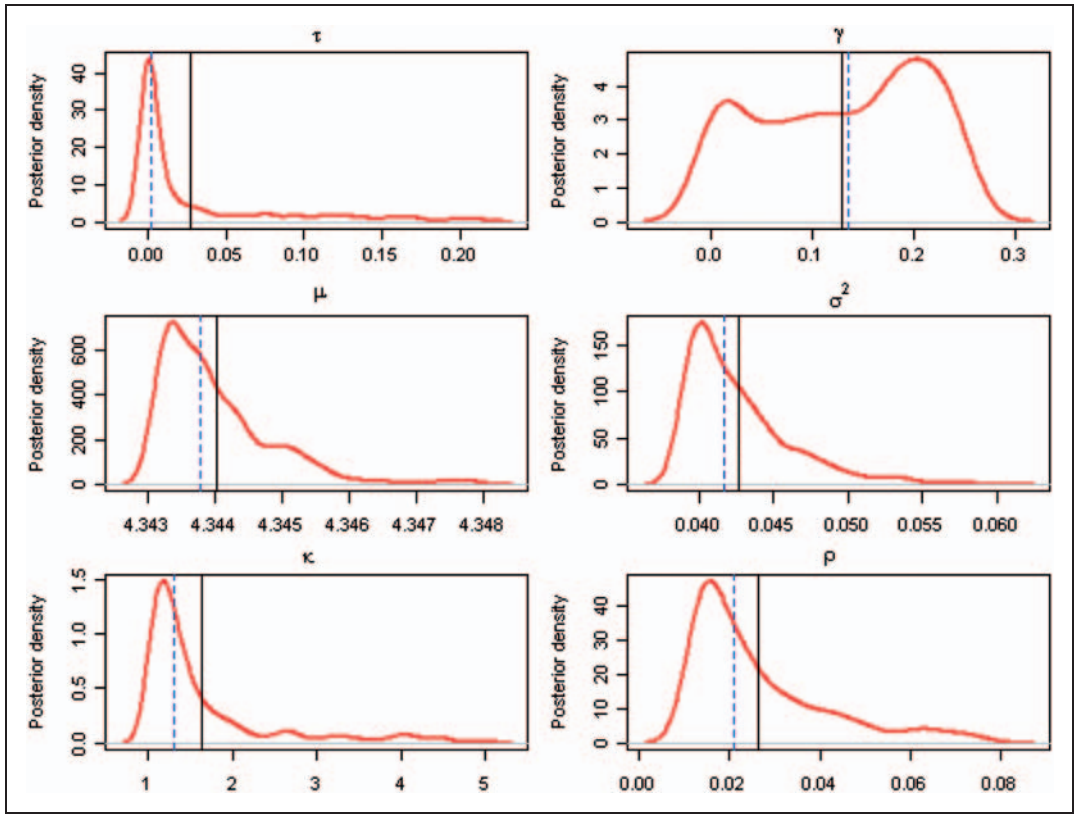

Figure 4. The posterior density plots of the parameters using Bayesian Markov chain Monte Carlo.

Table 3. The posterior estimates of the parameters for the JHLP data. The chain size is 400. The SE stands for the empirical standard error

Parameter	Mean	SE	2.5%	50%	97.5%
τ	0.0276	0.0497	0.0000	0.0019	0.1721
γ	0.1293	0.0806	0.0001	0.1360	0.2475
μ	4.3440	0.0008	4.3432	4.3438	4.3460
σ^2	0.0426	0.0036	0.0389	0.0416	0.0524
κ	1.6278	0.8242	1.0123	1.3064	4.2588
ρ	0.0263	0.0150	0.0119	0.0208	0.0671

JHLP: Johns Hopkins Lung Project.

The mode of the sojourn time is 15.78 years for the proposed sensitivity model, estimated from their posterior distributions. Wu et al.[3] analyzed the MLP data using the age-independent sensitivity model. For the MLP data, the posterior sensitivity had the posterior mean and median of 0.89 and 0.91, respectively, and the 95% highest posterior density (HPD) interval was (0.72, 0.98), which

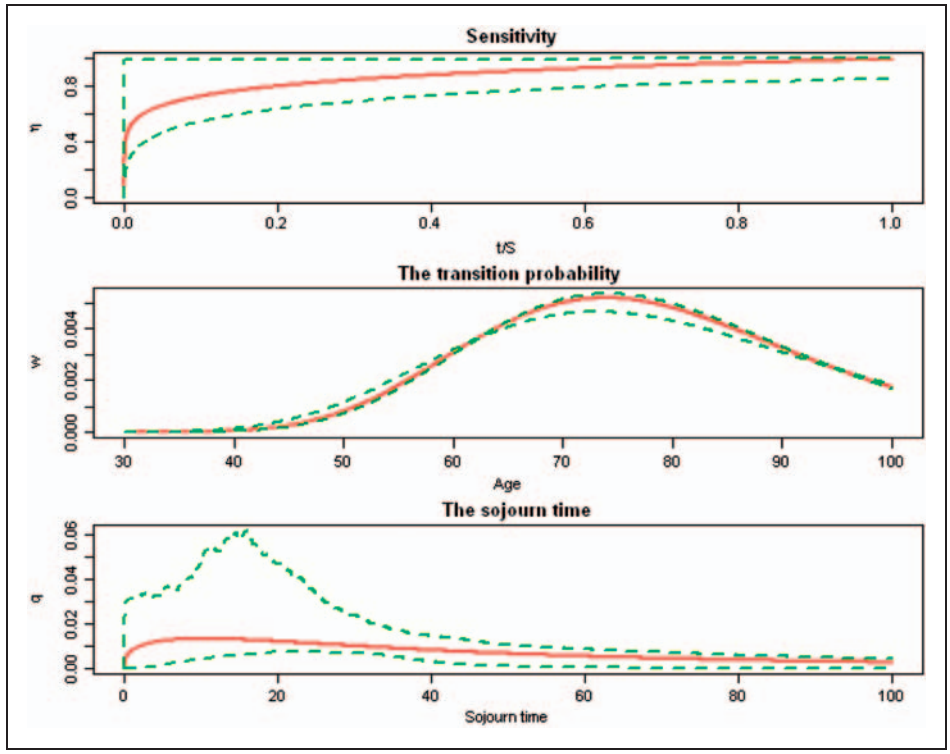

Figure 5. The prediction of the sensitivity, transition probability density, and sojourn time using Bayesian Markov chain Monte Carlo. The red solid line is from the posterior median and the green dotted lines are 95% credible curves.

are smaller than the estimated modes of the sensitivity from the JHLP data. This might be because their estimate of ρ ($\hat{\rho} = 1.040$) is much larger than our estimate ($\hat{\rho} = 0.0263$). A possible explanation is the change of the sensitivity model. When the sensitivity and the sojourn time are treated as independent in the model, the estimates of these two are usually negatively correlated.[8] In the MLP study, the sensitivity was modeled to be independent of age and the sojourn time, which may cause the overestimation of the sensitivity and the underestimation of the sojourn time. In the new proposed model, the sensitivity depends on the sojourn time as well as the time spent in the preclinical state. Thus, the estimated sensitivity ranges from zero to one, influencing the sojourn time to have a longer tail (see Figures 2 and 5). We hope the new model could remove the correlation between the estimated sensitivity and the estimated sojourn time.

As suggested by an anonymous reviewer, we tried to calculate the MLE and the posterior distributions for the MLP data using the new proposed method. However, in the MLP study, the participants took 19 screenings, with 4 months apart, making the calculation of the likelihood function very slow. In addition, the probability D_{i,t_0} and I_{i,t_0} have changed dramatically from our previous models, due to the changes in the sensitivity, with double layers of integration in the new approach. For these reasons, the MLE and the MCMC ran long time without convergence, so we could not make a direct comparison of these two data sets using exactly the same method. This is a limitation of using this new approach: too many screenings make it unlikely to get an estimate in a reasonable time frame. One way to overcome the computational burden and to speed up the

convergence is to rely on parallel computation. In general, the parallel computation is more efficient when a lot of independent calculations are required. We will consider the use of parallel computation in the future.

The estimated sojourn time is much longer than other existing method. Most literature estimated the sojourn time to be about 2 years. In the MLP study, the posterior mean of the sojourn time was 2.24 years, with a posterior median of 2.20 years for male heavy smokers. The 95% HPD interval for the mean sojourn time was (1.57, 3.35) years. In this work, the sojourn time is estimated to be around 15 years. There might be some medical evidence to show that for heavy smokers, the sojourn time is longer. For example, lung cancer may occur long after the smoker quitted smoking. Another contribution of the developed sensitivity model is to suggest a new way to evaluate the sensitivity: to improve the sensitivity is equivalent to improve the sensitivity right after a person enters the preclinical state, which is not to improve the overall sensitivity. Therefore, the proposed sensitivity model might provide policy makers a more accurate assessment of the predictive performance of a screening exam.

Acknowledgement

The authors thank the editor and the anonymous reviewers for their constructive suggestions to improve the manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Zelen M and Feinleib M. On the theory of screening for chronic diseases. *Biometrika* 1969; **56**: 601–614.
2. Wu D, Rosner GL and Broemeling LD. MLE and Bayesian inference of age-dependent sensitivity and transition probability periodic screening. *Biometrics* 2005; **61**: 1056–1063.
3. Wu D, Erwin D and Rosner GL. Sojourn time and lead time projection in lung cancer screening. *Lung Cancer* 2011; **61**: 322–326.
4. Wu D, Carino RL and Wu X. When sensitivity is a function of age and time spent in the preclinical state in periodic cancer screening. *J Modern Appl Stat Meth* 2008; **7**: 297–303.
5. Hori SS and Gambhir SS. Mathematical model identifies blood biomarker-based early cancer detection strategies and limitations. *Sci Translat Med* 2011; **3**: 109–116.
6. Berlin NI, Buncher CR, Fontana RS, et al. The National Cancer Institute cooperative early lung cancer detection program results of the initial screen (prevalence). Early lung cancer detection: introduction. *Am Rev Respir Dis* 1984; **130**: 545–549.
7. Gilks WR, Richardson S and Spiegelhalter DJ. *Markov chain Monte Carlo in practice*. Boca Raton, FL: Chapman & Hall/CRC, 1998.
8. Walter SD and Day NE. Estimation of the duration of a preclinical disease state using screening data. *Am J Epidemiol* 1983; **118**: 856–886.

Appendix

Bayesian Markov chain Monte Carlo within Gibbs sampling

Let $\theta = (\theta_i)_{i=1}^m$ be $(\tau, \gamma, \mu, \sigma^2, \kappa, \rho)$, where $m = 6$, and denote the full conditional posterior distribution of θ_i as $[\theta_i|\cdot]$, that is, the posterior distribution of θ_i given all other quantities in the model. Our model is nonlinear so that it leads to a form for $[\theta_i|\cdot]$ that is non-standard and is known only up to a normalizing constant. For this reason, we updated the parameters via Metropolis-Hastings-within-Gibbs steps and chose random-walk Metropolis for updating the θ_i parameters, which is a natural choice of Metropolis-Hastings.[7] In order to sample from the full conditional distributions, we used a truncated normal distribution as a proposal distribution centered at the

current value of θ_i, with variance ψ, where the variance ψ and the lower and upper truncation points were given by the parametric bootstrap runs. We describe the method for sampling θ_i^k from $[\theta_i|\cdot]$ at the kth iteration:

1. Sample a point y^k from $N_{(a_i,b_i)}(\theta_i^{k-1}, \psi)$, where θ_i^{k-1} is the sample at the $(k-1)$th iteration and (a_i, b_i) is the lower and upper truncation points for the parameter θ_i;
2. Sample a point u from a uniform distribution $Unif(0, 1)$;
3. If $u \leq \min\left\{1, \frac{[\theta_i=y^{k-1}|\cdot]}{[\theta_i=y^l|\cdot]}\right\}$ accept y^k as θ_i^k; else set θ_i^{k-1} equal to θ_i^k.