

5-1-2008

When Sensitivity is a Function of Age and Time Spent in the Preclinical State in Periodic Cancer Screening

Dongfeng Wu

University of Louisville, dongfeng.wu@louisville.edu


Ricolindo L. Cariño

Mississippi State University, carino@cavs.msstate.edu

Xiaoqin Wu

Mississippi State University

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Wu, Dongfeng; Cariño, Ricolindo L.; and Wu, Xiaoqin (2008) "When Sensitivity is a Function of Age and Time Spent in the Preclinical State in Periodic Cancer Screening," *Journal of Modern Applied Statistical Methods*: Vol. 7: Iss. 1, Article 25.

Available at: <http://digitalcommons.wayne.edu/jmasm/vol7/iss1/25>

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized administrator of DigitalCommons@WayneState.

When Sensitivity is a Function of Age and Time Spent in the Preclinical State in Periodic Cancer Screening

Cover Page Footnote

This research was partially supported by the National Institutes of Health grant CA-115012.

When Sensitivity is a Function of Age and Time Spent in the Preclinical State in Periodic Cancer Screening

Dongfeng Wu
University of Louisville

Ricolindo L. Cariño Xiaoqin Wu
Mississippi State University

Probability models are extended for periodic cancer screening trials to model sensitivity when it is changing with an individual's age and time spent in the preclinical state. Wu et al. (2005) showed that sensitivity is monotone increasing with age, but intuitively, sensitivity is also a function of the time one has spent in the preclinical stage. This allows us to infer sensitivity at a late stage, just before symptoms manifest. We developed the probability model and applied Bayesian inference to the HIP study group data. The methodology we developed is also applicable to other kinds of chronic diseases.

Keywords: Periodic screening exam, breast cancer, sensitivity, sojourn time, transition probability, incidence.

Introduction

Early detection and treatment may be effective ways to decrease mortality rate from cancer. The primary technique for early detection is screening exams. According to a recent report of the National Institute of Health (NIH 2000), breast cancer is the most common form of cancer among women in the United States and the second leading cause of cancer deaths among women. In the past four decades, seven major randomized controlled breast cancer screening trials have been carried out in North America and Europe.

In a screening program, a large group of asymptomatic individuals are enrolled in the program to detect the presence of a specific disease. The natural history of the disease for an individual is assumed to follow a progressive stochastic model, which consists of three states,

denoted by $S_0 \rightarrow S_p \rightarrow S_c$, corresponding, 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 state when the disease manifests itself in clinical symptoms. The *sensitivity* is the probability that the screening exam is positive given that the individual is in the preclinical stage.

The *sojourn time* refers to the time beginning when the disease first develops until the manifestation of clinical symptoms, that is $(S_c - S_p)$. The *transition probability* into the preclinical stage is the probability density function of making 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, the experience of the radiologist, etc. For example, recent studies indicate that the sensitivity of mammography increases with age at diagnosis (Wu et al., 2005; Shapiro et al., 1988; Miller et al., 1992a, 1992b), 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).

Dongfeng Wu is an Associate Professor in the Department of Bioinformatics and Biostatistics. She is a member of the Editorial Board of *JMASM*. E-mail: dongfeng.wu@louisville.edu. Ricolindo L. Carino is on the faculty at the Center for Advanced Vehicular Systems. Xiaoqin Wu is an Assistant Professor in the Department of Mathematics, Computer and Information Sciences.

In Walter and Day (1983), it was found that sensitivity is negatively correlated with the sojourn time. Intuitively, when the tumor cell is just formed, the sensitivity is very small, while at the late stage, that is, the preclinical stage comes to an end and the clinical stage will start soon, the sensitivity is close to one. In Wu, et al. (2005), the sensitivity was modeled as a function of age, while transition probability is age-dependent as well. Previous result are extended by investigating changes in the sensitivity from simultaneous variation of age and time spent in the preclinical stage.

The Model

Consider a cohort of initially asymptomatic individuals who enroll in a screening program. The sensitivity is denoted by $\beta(t,s|T)$, where t is the individual's age at the screening exam, s is the time duration that one has been stayed in the preclinical state, and T represents the sojourn time in the preclinical state. Define $w(t)dt$ as the probability of a transition from S_0 to S_p during $(t, t+dt)$. Let $q(t)$ be the probability density function of the sojourn time in S_p . Let $Q(z) = \int_z^\infty q(x)dx$, that is, $Q(z)$ is the survivor function of the sojourn time in the preclinical state S_p . Throughout this paper, the time variable t represents the participating individual's age. If random variables T and S are the duration times in S_0 and S_p respectively, then an individual will enter the clinical state S_c at age $T+S$, the probability density function of $(T+S)$ is

$$I(t) = \int_0^t w(x)q(t-x)dx.$$

$I(t)$ is the observable incidence of clinical cases if no intervention exists.

Consider a cohort of women in the study group who are all aged t_0 at study entry, and a protocol for K ordered screening examinations occurring at ages $t_0 < t_1 < \dots < t_{K-1}$, where $t_i = t_0 + i$ for annual screening exams. 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) , $i=1, 2, \dots, K-1$. The i -th generation of individuals consists of those who

enter S_p during this interval. The 0 -th generation includes all who enter S_p before the initial screening exam, and we let $t_{-1} \equiv 0$.

For each screening exam, let n_{i,t_0} 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 that an individual will be diagnosed at the k -th scheduled exam (at which her age is $t_{k-1} = t_0 + k - 1$) given that she is already in the preclinical state. The probability that an individual in S_p is detected at the first scheduled exam (i.e. $k = 1$) at age t_0 is

$$D_{1,t_0} = \int_0^{t_0} w(x) \int_{t_0-x}^\infty q(t)\beta(t_0, t_0 - x | t) dt dx. \quad (1)$$

The double integral in equation (1) arises because she must have entered the preclinical state S_p before t_0 , remained in that state at least until t_0 , and with the sensitivity changing with the sojourn time and the time spent in the preclinical stage, as well as with age.

Consider an i -th generation individual who was diagnosed at the k -th screening exam ($1 \leq i < k$). There are two possibilities: one is that 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 onset of the preclinical state S_p ; the other possibility is that she entered S_p in the $(k-1)$ -th screening interval (t_{k-2}, t_{k-1}) . Hence, the probability is

$$\begin{aligned}
 D_{k,t_0} = & \sum_{i=0}^{k-2} \int_{t_{i-1}}^{t_i} w(x) \int_{t_{k-1}-x}^{\infty} q(t) \left(\prod_{j=i}^{k-2} [1 - \beta(t_j, t_j - x | t)] \right) \beta(t_{k-1}, t_{k-1} - x | t) dt dx \\
 & + \int_{t_{k-2}}^{t_{k-1}} w(x) \int_{t_{k-1}-x}^{\infty} q(t) \beta(t_{k-1}, t_{k-1} - x | t) dt dx, \\
 \text{for } & \forall k = 2, \dots, K.
 \end{aligned} \tag{2}$$

Let $I_{k,t_0}(t)dt$ be the probability that an individual enters the state S_c at a small age interval $(t, t + dt)$, where $t \in (t_{k-1}, t_k)$. If this woman was in the i -th generation, $i < k$, then she must have gone through her previous $(k-i)$ screening exams without being detected, and had a sojourn time $(t-x)$, where x is her onset age of the preclinical state S_p . Another possibility is that she may have entered S_p after the k -th exam and incident at age t . Hence for $\forall t \in (t_{k-1}, t_k)$,

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

Therefore, for any $k = 1, \dots, K$, the probability of being incident in the k -th screening interval (t_{k-1}, t_k) is

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

The likelihood function for this cohort of women is

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

The full likelihood for the study group across all ages is

$$\begin{aligned}
 L = & \prod_{t_0} \prod_{k=1}^K D_{k,t_0}^{S_{k,t_0}} I_{k,t_0}^{r_{k,t_0}} (1 - D_{k,t_0}^{S_{k,t_0}} - I_{k,t_0}^{r_{k,t_0}})^{n_{k,t_0} - S_{k,t_0} - r_{k,t_0}}
 \end{aligned} \tag{6}$$

The age effect and the time spent in the preclinical state were modeled in the sensitivity in the following way. The sensitivity is associated with age by a logistic link, and it is associated with the time spent in the preclinical state by a linear function. Let

$$\begin{aligned}
 \beta(t, s | T) = & \frac{1}{1 + \exp(-b_0 - b_1 * (t - \bar{t}))} \times \frac{s}{T},
 \end{aligned} \tag{7}$$

where t is an individual's age at the screening exam, s is the time one has spent in the preclinical state, T is the sojourn time in the preclinical state, and \bar{t} is the average age at entry in the whole study group. If $b_1 > 0$, $\beta(t)$ will be a monotone increasing function of age t .

The transition probability density function $w(t)$ is the instantaneous probability of a transition from S_0 to S_p . The integral $\int_0^{\infty} w(t) dt$ represents a lifetime risk for a healthy female to transit into 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 is 15.7%, which is less than a woman's lifetime risk of entering the preclinical disease state.

PERIODIC CANCER SCREENING

Table 1. Bayesian posterior estimates for the HIP data

| Parameters | Median | Mean | S.E. |
|------------|--------|-------|-------|
| b_0 | 1.730 | 1.708 | 0.984 |
| b_1 | 0.084 | 0.083 | 0.072 |
| μ | 4.384 | 4.392 | 0.065 |
| σ^2 | 0.235 | 0.253 | 0.095 |
| κ | 1.744 | 2.126 | 1.004 |
| ρ | 0.381 | 0.366 | 0.104 |

Table 2. Bayesian posterior estimates for the sensitivity β at the end of the preclinical state

| Age | Median | Mean | S.E. |
|-----|--------|-------|-------|
| 40 | 0.644 | 0.628 | 0.215 |
| 45 | 0.736 | 0.716 | 0.178 |
| 50 | 0.829 | 0.789 | 0.150 |
| 55 | 0.886 | 0.841 | 0.129 |
| 60 | 0.917 | 0.877 | 0.115 |
| 65 | 0.940 | 0.899 | 0.109 |

Table 3. Bayesian posterior estimates for the transition probability ${}^a w$.

| Age | Median | Mean | S.E. |
|-----|--------|-------|-------|
| 40 | 1.400 | 1.407 | 0.243 |
| 45 | 1.734 | 1.738 | 0.210 |
| 50 | 1.994 | 2.006 | 0.254 |
| 55 | 2.171 | 2.193 | 0.335 |
| 60 | 2.267 | 2.296 | 0.404 |
| 65 | 2.286 | 2.322 | 0.441 |

^a The unit is 10^{-3} .

Hence, 20% is chosen as a reasonable upper bound.

$$w(t) = \frac{0.2}{\sigma t \sqrt{2\pi}} \exp\left\{-\frac{(\log t - \mu)^2}{2\sigma^2}\right\} \quad (8)$$

which is the PDF of lognormal(μ , σ^2) multiplied by 20%. Here, $w(t)$ is a sub-density function, where μ and σ^2 are parameters to be estimated.

The log logistic distribution is adopted to model the sojourn time in the preclinical state,

$$q(x) = \frac{\kappa x^{\kappa-1} \rho^\kappa}{[1 + (x\rho)^\kappa]^2}, x > 0, \quad (9)$$

where x is the sojourn time, and κ and ρ are positive parameters representing the scale and location in the log logistic family. An advantage of this family over the exponential is that it has two parameters and is more robust in the tails.

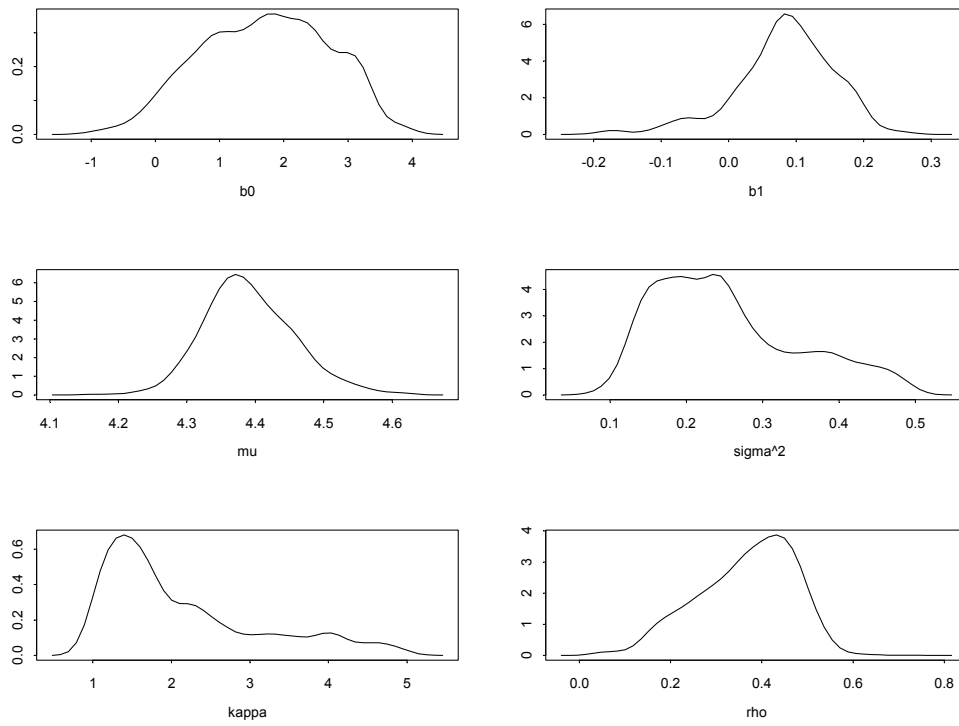


Figure 1: Density Curve for the posterior samples.

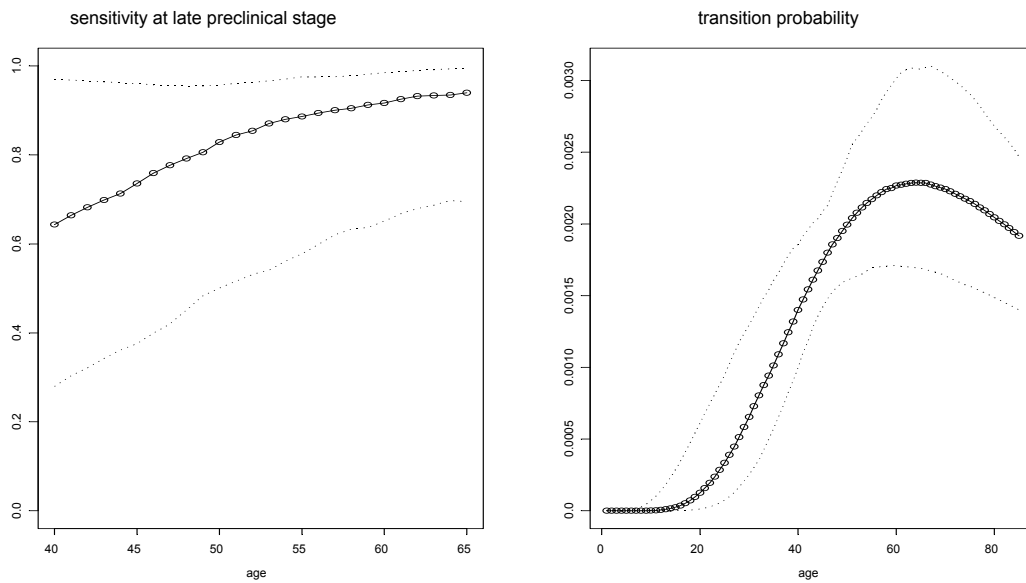


Figure 2: Posterior quantiles (5%, 50%, and 95%) of sensitivity and transition probability.

Another advantage of this family is that its relatively simple form is achieved for the survivor function and the hazard function. Its first moment can be calculated directly from

$$E(X) = \frac{\pi}{\rho\kappa} \operatorname{csc}\left(\frac{\pi}{\kappa}\right).$$

For the r -th moment to exist, $\kappa > r$ is required. For justifications on how these age effect functions are chosen, see Wu, et al. (2005).

Simulation Procedure and Results

The analysis of the HIP study data is now described based on the likelihood function and probability calculation derived above. In the proposed model, there are six unknown parameters, that is, $\theta = (b_0, b_1, \mu, \sigma^2, \kappa, \rho)$. Theoretically, the parameters have a domain of either $(-\infty, \infty)$ or $(0, \infty)$. The practical meaning of these parameters will limit them to a finite range. The range for each of them was identified as: $0 < b_0 < 5$, $-0.2 < b_1 < 0.2$, $3.5 < \mu < 4.5$, $0 < \sigma^2 < 1$, $0.1 < \rho < 2.0$, and $1 < \kappa < 5$. For justifications of these ranges, see Wu, et al. (2005).

Markov Chain Monte Carlo (MCMC) was used to generate a random sample from the joint posterior distribution of the parameters for Bayesian inference. The posterior simulation was partitioned into four sub-chains, sampling the posterior for $(b_0, b_1), \mu, \sigma^2, (\kappa, \rho)$ separately.

A noninformative bivariate normal prior for (b_0, b_1) was chosen, that is, a bivariate normal distribution with mean vector $(0, 0)$ and variance equal to 10^{10} times the identity matrix. A noninformative normal prior was chosen for μ , namely $N(0, 10^{10})$. The prior for σ^2 was uniform $(0, 1)$, and the prior distributions for κ and ρ were uniform $(1, 5)$ and uniform $(0, 2)$, respectively. The two-dimensional integrals in the likelihood function do not have an analytical form. The trapezoidal rule was used to evaluate them when calculating the likelihood.

The MCMC was run for 4,800 steps, with a burn-in of 3,200 iterations. After the burn-in time, the posteriors were sampled every 20 steps, giving 80 posterior samples for the

parameter vector θ . Sixteen chains were simulated, each with different starting values that are over dispersed with respect to the target distribution. The two-dimensional integrals in the likelihood function are very time consuming. The chains were simulated in parallel on a Linux cluster, taking 192 hours to complete. The 80 posterior samples from each of the 16 chains were pooled for the analysis, giving a total of 1280 posterior samples.

The Bayesian posterior estimates of θ for the HIP study data are shown in Table 1. Sensitivity at the end of the preclinical state appears to increase with age. If $s \rightarrow T$, that is, if the time spent in the preclinical state s converges to the sojourn time T , then the sensitivity in the late stage of the preclinical state can be estimated. This trend is obvious from the quantile plot of the saved posterior samples of the parameters in Figure 1. In the HIP data, the posterior mean sensitivity increases from 0.628 to 0.899 from ages 40 to 65 years, and the posterior standard error drops from 0.215 to 0.109. In fact, the posterior error of sensitivity was monotone decreasing as age increases.

The age-dependent transition probability is itself a sub-pdf from our model construction. The posterior density curve of the transition probability could be seen from Figure 1. The posterior mean of the transition probability varies from 1.407×10^{-3} to 2.322×10^{-3} . The transition probability is not a monotone function of age; it has a single maximum at age 64. If the posterior means for the parameters κ and ρ were used, then the posterior mean sojourn time is 4.06 years, with a mode of 1.70 years.

Discussion

The previous model in Wu, et al. (2005) is extended by adding one element in the sensitivity function, namely, the time spent in the preclinical state, with the sojourn time. Walter and Day (1983) found that the sensitivity and the sojourn time distribution were negatively correlated for a given incidence. Our modeling should be able to handle that problem, as our sensitivity was defined as a function of the sojourn time as well, and it was reciprocal to the sojourn time. More complicated models

could be explored, but the computation time will be greatly increased, and simulation would take too long to be finished.

The result was compared with Wu, et al. (2005). It was found that the late stage sensitivity was slightly higher, with a slightly smaller standard error. It was also found that the transition probability was changing less across different ages, with a slightly smaller standard error. However, the mean sojourn time is much longer than the previous result.

Acknowledgements

This research was partially supported by the National Institutes of Health grant CA-115012.

References

- Baker, S. G. (1998). Evaluating the age to begin periodic breast cancer screening using data from a few regularly scheduled screenings. *Biometrics*, *54*, 1569-1578.
- Chen, T. H. H., Kuo, H. S., Yen, M. F., Lai, M. S., Tabar, L., & Duffy, S. W. (2000). Estimation of sojourn time in chronic disease screening without data on interval cases. *Biometrics*, *56*, 167-172.
- Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. NY: Chapman & Hall/CRC.
- Day, N. E., & Walter, S. (1984). Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. *Biometrics*, *40*, 1-13.
- Eddy, D. M. (1980). *Screening for cancer: Theory, analysis and design*. Englewood Cliffs, NJ: Prentice Hall.
- Kerlikowske, K., Grady, D., Barclay, J., Sickles, E. A., & Ernster, V. (1996). Effect of age, breast density, and family history on the sensitivity of first screening mammography. *Journal of the American Medical Association*, *276*, 33-38.
- Lee, S. J. & Zelen, M. (1998). Scheduling periodic examinations for the early detection of disease: Applications to breast cancer. *Journal of the American Statistical Association*, *93*, 1271-1281.
- 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. *Canadian Medical Association Journal*, *147*(10), 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. *Canadian Medical Association Journal*, *147*(10), 1477-88.
- National Institute of Health (2000). *NIH Publication No. 00-1556*, 12/12/2000.
- 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: The Johns Hopkins University Press.
- Shen, Y., Wu, D., & Zelen, M. (2001). Testing the independence of two diagnostic tests. *Biometrics*, *57*, 1009-1017.
- 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*, 217-229.
- Walter, S. D. & Day, N. E. (1983). Estimation of the duration of a preclinical disease state using screening data. *American Journal of Epidemiology*, *118*, 856-86.
- Wu, D., Rosner, G. L., and Broemeling, L. D. (2005). MLE and Bayesian inference of age-dependent sensitivity and transition probability in periodic screening. *Biometrics*. Vol.61, No.4, 1056-1063.
- Zelen, M. (1993). Optimal scheduling of examinations for the early detection of disease. *Biometrika*, *80*, 279-93.