Long-Term Effects of Periodic Cancer Screening for Aged People with a Screening History

Dongfeng Wu*

Abstract

This research is an extension of the previous probability model for evaluating long-term outcomes due to regular screening. The previous model was for people without any screening history, while the current extension focus on old people who has a screening history and are superficially healthy so far. As before, People who take part in cancer screening are divided into four mutually exclusive groups: *True-early-detection, No-early-detection, Overdiagnosis*, and *Symptom-free-life*. For each case, we derive the probability formula with a screening history. Simulation studies using the HIP (Health Insurance Plan for Greater New York) breast cancer study's data provide estimates for these probabilities and corresponding credible intervals. These probabilities change with a person's age at study entry, screening history, future planned screening frequency, screening sensitivity, and other parameters. We also allow human lifetime to be subject to a competing risk of death from other causes. The model can provide policy makers with important information regarding the distribution of individuals participating in a screening program who eventually fall into one of the four groups. Finally the method is applicable to other kinds of screening as well.

Key Words: over-diagnosis, true-early-detection, no-early-detection, symptom-free-life, sensitivity, sojourn time, transition probability density

1. Introduction

We have developed a probability model for evaluating long term effects and over diagnosis for initially asymptomatic people who has no history of screening (Wu, et al 2014). However, most people in their 60s or older may have gone through at least one screening exam before. How to extend the original model to the aged group with a screening history is very challenging (Badgwell et al 2008): these people in their 70s or 80s maybe superficially healthy, but what is the long term prospect of taking screening exams? Is it still necessary to take screening exams at all? and if it is necessary, at what frequency, or how to plan it in the future? This research is targeting this critical problem.

We will use women's breast cancer as an example. We assume that a woman has taken a sequence of screening exams before, however, breast cancer has never been found, i.e., no history of breast cancer, and currently she is asymptomatic. If she plan to take screening exams in the future, then based on her future screening outcomes and her ultimate disease status, She will belong to one of four possible groups eventually:

- *Group 1*: **Symptom-free-life (SympF)**. A woman in Group 1 took part in screening exams, but breast cancer was never detected and ultimately she died of other causes.
- *Group 2*: **No-early-detection (NoED)**. A woman in Group 2 took part in screening exams, but disease manifested itself clinically and was not detected by scheduled screening exams.
- *Group 3*: **True-early-detection (TrueED)**. A woman in Group 3 was diagnosed with breast cancer at a scheduled screening exam and her clinical symptoms would have appeared before her death.

*Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, 485 E. Gray Street, Louisville, KY 40202

- *Group 4*: **Overdiagnosis (OverD)**. A woman in Group 4 was diagnosed with breast cancer at a scheduled screening exam but her clinical symptoms would NOT have appeared before her death.

The definitions of these four groups were given originally in Wu, et al 2014. And the probability of each group was derived when an individual has no history of screening and currently asymptomatic. In this project, we will expand upon the previous work, and assume an individual has a screening history.

The remainder of the paper is organized as follows. In Section 2 we propose a probability model and derive the probabilities for each of the four groups, treating the duration of human lifetime as a random variable, with the cause of death subject to other competing risks. In Section 3, we apply our method to the Health Insurance Plan for Greater New York (HIP) breast cancer screening data, and present some simulation results for different screening schedules. Policy makers may use these probabilities to help assess different screening strategies, such as changing screening frequencies, determining age to start screening, etc. We conclude with a discussion in Section 4.

2. Probability Formulation

Consider a currently asymptomatic individual with a screening history who has no cancer in the past. We assume the commonly followed disease progression model in which the disease develops through three states $S_0 \to S_p \to S_c$: S_0 refers to the disease-free state or the state in which the disease cannot be detected; S_p refers to the preclinical disease state, in which an asymptomatic individual unknowingly has disease that a screening exam can detect; and S_c refers to the disease state at which the disease manifests itself in clinical symptoms. Let $\beta(t)$ be the sensitivity at age t, the probability that the screening exam is positive given that the individual is in the preclinical state, and let $\beta_i = \beta(t_i)$. Let $w(t)$ be the probability density function (pdf) of time spending in the disease free state S_0, Let $q(x)$ be the probability density function of the sojourn time in S_p, and let $Q(z) = \int_z^\infty q(x)dx$ be the survivor function of the sojourn time. We assume that the sojourn time distribution does not depend on the age of entry into S_p, that is, sojourn time and time spend in disease-free state are independent. Throughout, the time variable t represents an individual's age at time of screening, and T represents a person's lifetime, a continuous random variable with a probability density function $f_T(t)$.

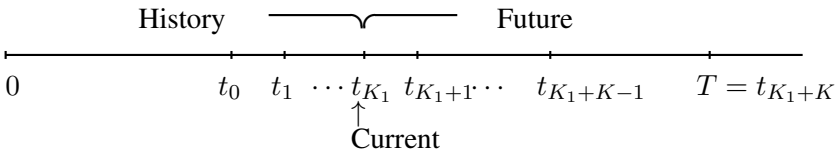

Figure 1: Illustration of the screening history and future schedule.

Assume that this individual has gone through a sequence of breast cancer screening at her ages $t_0 < t_1 < \cdots < t_{K_1-1}$, and no breast cancer has been found so far, and her current age is t_{K_1}. In the future, she plans to take a few more screenings at her ages $t_{K_1} < t_{K_1+1} < t_{K_1+2} < \cdots < t_{K_1+K-1}$, and her lifetime $T > t_{K_1+K-1}$, see Figure 1 for illustration.

To derive the probability of each of the four cases, we break down the problem into a few steps: 1). We derive the probability for the most simple case when $K_1 = K = 1$ with lifetime T fixed; 2). We allow the lifetime T to be a random variable when $K_1 = K = 1$;

3). We generalize the result for any fixed positive integers K_1 and K when lifetime T is fixed; 4). Finally we allow the lifetime T to be a random variable for any K_1, and hence the number of future screening exams K is a random variable as well.

2.1 Probability of Each Case When $K_1 = K = 1$

For a superficially healthy woman at her current age t_1, suppose she underwent only one screening exam at age $t_0 (< t_1)$ before, and no breast cancer was found, we will derive the probability of each case when there is only one screening exam (at her age t_1) during her lifetime, with a fixed lifetime $T = t$; then we extend to the case when her lifetime is a random variable, $T \sim f_T(t)$. See Figure 2 for illustration.

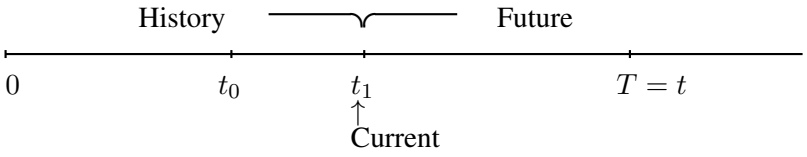

Figure 2: Screening history and future schedule when $K_1 = K = 1$.

First we define an event

$$H_1 = \left\{ \begin{array}{l} \text{A woman had a screening exam at age } t_0\text{, no BC was found,} \\ \text{and she is asymptomatic at current age } t_1 \end{array} \right\}.$$

To calculate the probability of H_1, that is, the conditional probability that *no* breast cancer was found before/at age t_1 when one's lifetime T exceeds t_1, this can arise as one of three mutually exclusive events: either (i) she remains in the disease-free state through age t_1, the probability of which is $1 - \int_0^{t_1} w(x)dx$; or (ii) she enters state S_p before t_0 but remains in S_p long enough that no symptom presents before t_1, and she was missed by the screening exam at t_0, the probability of which is $(1 - \beta_0) \int_0^{t_0} w(x)Q(t_1 - x)dx$; or (iii) she enters state S_p in (t_0, t_1), and no symptom presents before t_1, the proability of whihc is $\int_{t_0}^{t_1} w(x)Q(t_1 - x)dx$. Therefore the probability of H_1 is the total of the three probabilities (i)-(iii):

$$\begin{aligned} P(H_1|T \geq t_1) &= 1 - \int_0^{t_1} w(x)dx + (1 - \beta_0) \int_0^{t_0} w(x)Q(t_1 - x)dx \\ &+ \int_{t_0}^{t_1} w(x)Q(t_1 - x)dx. \end{aligned} \quad (1)$$

For a Group 1 case, a woman who never has detectable breast cancer during her lifetime $(0, t)$, it can follow one of 4 trajectories: (a) she has stayed in the disease-free state S_0 throughout her lifetime $(0, t)$; (b) she entered the preclinical state S_p before t_0, her cancer was not detected at t_0 nor at t_1, and her sojourn time was so long that no clinical symptom appeared before her death; (c) she entered the preclinical state S_p in (t_0, t_1), her cancer was not detected at t_1, and she had a long sojourn time, so that no symptom appeared before her death; (d) she entered the preclinical state S_p in (t_1, t) and had a long sojourn time, so that no symptom appeared before her death; Hence the conditional probability given her lifetime $T = t (> t_1)$ is:

$$\begin{aligned} & P(\textit{Case 1}\text{: SympF}, H_1|T = t) \\ = \; & 1 - \int_0^t w(x)dx + (1 - \beta_1)(1 - \beta_0) \int_0^{t_0} w(x)Q(t - x)dx \\ + \; & (1 - \beta_1) \int_{t_0}^{t_1} w(x)Q(t - x)dx + \int_{t_1}^t w(x)Q(t - x)dx. \end{aligned} \quad (2)$$

For a Group 2 case, a woman whose cancer became symptomatic and was found in (t_1, T), either (a) she entered S_p before t_0 and was not detected by the 2 exams, or (b) she entered the preclinical state in (t_0, t_1) and was missed by the screening exam at t_1, or (c) she entered the preclinical state after t_1. In all situations, her sojourn time in S_p was shorter than $(t - x)$, where x is her age entering S_p. Hence the conditional probability is

$$\begin{aligned} P(\textit{Case 2}\text{: NoED}, H_1|T = t) &= (1-\beta_1)(1-\beta_0)\int_0^{t_0} w(x)[Q(t_1-x) - Q(t-x)]dx \\ &+ (1-\beta_1)\int_{t_0}^{t_1} w(x)[Q(t_1-x) - Q(t-x)]dx \\ &+ \int_{t_1}^{t} w(x)[1 - Q(t-x)]dx. \end{aligned} \tag{3}$$

For a Group 3 case, a woman is truly detected early at t_1 by taking the scheduled exam, and her symptoms would have appeared before death. That is, she must have entered S_p at some age x before t_1, and her sojourn time would have been between $(t_1 - x)$ and $(T - x)$. Hence,

$$\begin{aligned} P(\textit{Case 3}\text{: TrueED}, H_1|T = t) &= \beta_1(1-\beta_0)\int_0^{t_0} w(x)[Q(t_1-x) - Q(t-x)]dx \\ &+ \beta_1\int_{t_0}^{t_1} w(x)[Q(t_1-x) - Q(t-x)]dx \end{aligned} \tag{4}$$

For a Group 4 case, the case of overdiagnosis, she is to be diagnosed at t_1 but her symptoms would not have appeared before death. That is, she must have entered S_p at some age $x(< t_1)$, but her sojourn time would have extended to beyond time $(T - x)$. Hence,

$$\begin{aligned} P(\textit{Case 4}\text{: OverD}, H_1|T = t) &= \beta_1(1-\beta_0)\int_0^{t_0} w(x)Q(t-x)dx \\ &+ \beta_1\int_{t_0}^{t_1} w(x)Q(t-x)dx. \end{aligned} \tag{5}$$

The probability of each case when the lifetime T is a random variable and $T \geq t_1$ can be obtained by

$$\begin{aligned} P(\text{Case } i, H_1|T \geq t_1) &= \int_{t_1}^{\infty} P(\text{Case } i, H_1|T = t) f_T(t|T \geq t_1)dt, \\ & \qquad i = 1, 2, 3, 4. \end{aligned} \tag{6}$$

where the conditional pdf $f_T(t|T \geq t_1)$ is

$$f_T(t|T \geq t_1) = \begin{cases} \frac{f_T(t)}{P(T>t_1)} = \frac{f_T(t)}{1-F_T(t_1)}, & \text{if } t \geq t_1 \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

By adding (2) to (5), one can verify that for any $t > t_1$,

$$\begin{aligned} \sum_{i=1}^{4} P(\text{Case } i, H_1|T = t) &= 1 - \int_0^{t_1} w(x)dx + (1-\beta_0)\int_0^{t_0} w(x)Q(t_1-x)dx \\ &+ \int_{t_0}^{t_1} w(x)Q(t_1-x)dx \\ &= P(H_1|T \geq t_1). \end{aligned} \tag{8}$$

Since the right hand side of (8) does not depend on t, we have

$$\begin{aligned}\sum_{i=1}^{4} P(\text{Case } i, H_1|T \geq t_1) &= \int_{t_0}^{\infty} [\sum_{i=1}^{4} P(\text{Case } i, H_1|T = t)] f_T(t|T \geq t_1)dt \\ &= P(H_1|T \geq t_1). \end{aligned} \quad (9)$$

This implies

$$\sum_{i=1}^{4} P(\text{Case } i|H_1, T \geq t_1) = \sum_{i=1}^{4} \frac{P(\text{Case } i, H_1|T \geq t_1)}{P(H_1|T \geq t_1)} = 1. \quad (10)$$

2.2 Probability of Outcomes for Any K_1 and K: Multiple Exams with a History

We generalize this idea to an individual with a history of any number of screenings, and we focus on modeling the impact of future screening schedule on the four possible outcomes: symptom-free-life, no-early-detection, true-early-detection and over-diagnosis. We assume that an initially asymptomatic individual has gone through K_1 screening exams so far, which occurred at her ages $t_0 < t_1 < \cdots < t_{K_1-1}$, see Figure 1. We let $t_{-1} = 0$, and her current age is $T = t_{K_1} (> t_{K_1-1})$. Define an event

$$H_{K_1} = \left\{ \begin{array}{l} \text{A woman had screening exams at her ages } t_0 < t_1 < \cdots < t_{K_1-1}, \\ \text{no breast cancer was found ,} \\ \text{and she is asymptomatic at her current age } t_{K_1} \end{array} \right\}.$$

We first calculate $P(H_{K_1}|T \geq t_{K_1})$, the conditional probability that no breast cancer was found before/at age t_{K_1} given that her lifetime T exceeds t_{K_1}.There are $(K_1 + 2)$ mutually exclusive events for H_{K_1} to happen: (i) She never progressed out of the disease-free state S_0 throughout her lifetime, the probability of which is $1 - \int_0^{t_{K_1}} w(x)dx$; or (ii) shen entered state S_p in age interval $(t_{j-1}, t_j), j = 0, \ldots, K_1 - 1$, but remains in S_p long enough that no symptom presents before t_{K_1}, and she was missed by the following $(K_1 - j)$ exams; or (iii) she entered S_p after t_{K_1-1}, and with no symptoms before t_{K_1}. Thus the probability of H_{K_1} is the sum of these probabilities:

$$\begin{aligned} P(H_{K_1}|T \geq t_{K_1}) &= 1 - \int_0^{t_{K_1}} w(x)dx \\ &+ \sum_{j=0}^{K_1-1} (1 - \beta_j) \cdots (1 - \beta_{K_1-1}) \int_{t_{j-1}}^{t_j} w(x)Q(t_{K_1} - x)dx \\ &+ \int_{t_{K_1-1}}^{t_{K_1}} w(x)Q(t_{K_1} - x)dx. \end{aligned} \quad (11)$$

Now if she plans to undergo K screening exams in the future, occuring at her age $t_{K_1} < t_{K_1+1} < \cdots < t_{K_1+K-1}$, we first derive the conditonal probability for each outcomes when her life time T is fixed, then we will allow her lifetime to be a random variable.

Given that her lifetime is $T = t_{K_1+K} (> t_{K_1+K-1})$, a Group 1 case where clinical breast cancer never occurs in her lifetime, can arise as any one of $(K_1 + K + 2)$ mutually exclusive events: (a) she remained in the disease-free state S_0 throughout her lifetime, the probability of which is $1 - \int_0^{t_{K_1+K}} w(x)dx$. (b) she entered the preclinical state S_p when she was between ages t_{j-1} and $t_j, j = 0, \ldots, K_1 + K - 1$, and she was not detected by the following $(K_1 + K - j)$ exams, and she had a long sojourn time, so no clinical symptom

appeared before her death ($K_1 + K$ disjoint events). (c) she entered S_p after t_{K_1+K-1} and without clinical symptoms before her death. Then we add the probability together:

$$\begin{aligned} & P(\text{Case } 1, H_{K_1}|T = t_{K_1+K}) \\ = \; & 1 - \int_0^{t_{K_1+K}} w(x)dx + \int_{t_{K_1+K-1}}^{t_{K_1+K}} w(x)Q(t_{K_1+K} - x)dx \\ + \; & \sum_{j=0}^{K_1+K-1} (1-\beta_j)\cdots(1-\beta_{K_1+K-1}) \int_{t_{j-1}}^{t_j} w(x)Q(t_{K_1+K} - x)dx. \end{aligned} \quad (12)$$

For a Group 2 case, we calculate the probability of no early detection by defining $I_{K_1+K,j}$ as the probability of being an interval case in the interval $(t_{j-1}, t_j), j = K_1 + 1, K_1 + 2, \cdots, K_1 + K$, in a sequence of K screening exams. Thus

$$P(\text{Case } 2, H_{K_1}|T = t_{K_1+K}) = I_{K_1+K,K_1+1} + I_{K_1+K,K_1+2} + \cdots + I_{K_1+K,K_1+K}, \quad (13)$$

where

$$\begin{aligned} I_{K_1+K,j} \; = \; & \sum_{i=0}^{j-1} (1-\beta_i)\cdots(1-\beta_{j-1}) \int_{t_{i-1}}^{t_i} w(x)[Q(t_{j-1} - x) - Q(t_j - x)]dx \\ + \; & \int_{t_{j-1}}^{t_j} w(x)[1 - Q(t_j - x)]dx, \quad \text{for } \forall j = K_1 + 1, \cdots, K_1 + K. \end{aligned} \quad (14)$$

A Group 3 case, true early detection, can arise as one of K disjoint events depending on her age at diagnosis by screening, namely, at $t_j, j = K_1, K_1 + 1, \cdots, K_1 + K - 1$. If she is diagnosed at t_j, then she must have entered the preclinical state S_p before t_j, and was not detected by the previous j exams, and her sojourn time must have been in the interval $(t_j - x, t_{K_1+K} - x)$, where x represent the onset time of the preclincal state. Therefore

$$\begin{aligned} & P(\text{Case } 3, H_{K_1}|T = t_{K_1+K}) \\ = \; & \sum_{j=K_1}^{K_1+K-1} \beta_j \left\{ \sum_{i=0}^{j-1} (1-\beta_i)\cdots(1-\beta_{j-1}) \int_{t_{i-1}}^{t_i} w(x)[Q(t_j - x) - Q(t_{K_1+K} - x)]dx \right. \\ & \left. + \int_{t_{j-1}}^{t_j} w(x)[Q(t_j - x) - Q(t_{K_1+K} - x)]dx \right\} \end{aligned} \quad (15)$$

A Group 4 case, overdiagnosis, also can arise as one of K disjoint events. She might have been diagnosed at the jth exam, but her sojourn time would have been longer than $t_{K_1+K} - x$, thus her symptoms did not appear before her death:

$$\begin{aligned} & P(\text{Case } 4, H_{K_1}|T = t_{K_1+K}) \\ = \; & \sum_{j=K_1}^{K_1+K-1} \beta_j \left\{ \sum_{i=0}^{j-1} (1-\beta_i)\cdots(1-\beta_{j-1}) \int_{t_{i-1}}^{t_i} w(x)Q(t_{K_1+K} - x)dx \right. \\ & \left. + \int_{t_{j-1}}^{t_j} w(x)Q(t_{K_1+K} - x)dx \right\}. \end{aligned} \quad (16)$$

It is proved that for any screening number $K_1 \geq 1$ and $K \geq 1$,

$$\sum_{i=1}^{4} P(\text{Case } i, H_{K_1}|T = t_{K_1+K}) = P(H_{K_1}|T \geq t_{K_1}) \quad (17)$$

For an individual currently at age t_{K_1}, her lifetime is not fixed but is a random variable, so it may not be realistic to consider the future number of exams K to be a fixed value. However, if she plans to follow a future screening schedule $t_{K_1} < t_{K_1+1} < \ldots$, then $K = n$ if $t_{K_1+n-1} < T < t_{K_1+n}$, the screening number $K = K(T)$ is a random variable, changing with the lifetime T. The probability of each case ($i = 1, 2, 3, 4$) when her lifetime T is longer than t_{K_1} can be obtained as the weighted average

$$P(\text{Case } i, H_{K_1}|T \geq t_{K_1}) = \int_{t_{K_1}}^{\infty} P(\text{Case } i, H_{K_1}|K = K(T), T = t) f_T(t|T \geq t_{K_1}) dt,$$

where $f_T(t|T \geq t_{K_1}) = \frac{f_T(t)}{1-F_T(t_{K_1})}$, if $t \geq t_{K_1}$. The probability inside the integration, $P(\text{Case i}, H_{K_1}|K = K(T), T = t)$, was derived in (12)-(16).

It is straight forward to verify by (17) that for any future screening schedule when the lifetime T is random,

$$\sum_{i=1}^{4} P(\text{Case } i|H_{K_1}, T \geq t_{K_1}) = 1. \tag{18}$$

3. A Projection of Long Term Outcomes Using the HIP Data

We applied our method to the Health Insurance Plan of the Greater New York (HIP) breast caner data (Shapiro et al. 1988), using lifetime distribution derived from the actuarial lifetable on the Social Security Administration (SSA) website.

3.1 Bayesian Estimate of the Probability

According to our results in Section 2, the probability of each case is a function of the three key parameters (screening sensitivity $\beta(t)$, the transition pdf $w(t)$, the survival function of the sojourn time $Q(x)$), a person's current age t_{K_1}, her screening history, her future screeing schedule, and human lifetime distribution.

The three key parameters: age-dependent sensitivity $\beta(t)$, the age-dependent transition probability $w(t)$, and the survival function of sojourn time $Q(x)$, were estimated from the HIP data in Wu, Rosner and Broemeling (2005). The parametric models for $\beta(t), w(t)$, and $Q(x)$ were

$$\beta(t) = \frac{1}{1 + \exp\{-b_0 - b_1(t-m)\}}, \tag{19}$$

$$w(t) = \frac{0.2}{\sqrt{2\pi}\sigma t} \exp\left\{-(\log t - \mu)^2/(2\sigma^2)\right\}, \tag{20}$$

$$Q(x) = \frac{1}{1 + (x\rho)^{\kappa}}, \quad \kappa > 0, \rho > 0, \tag{21}$$

where m is the average age of women at the study entry. The unknown parameters in this model are $\theta = (b_0, b_1, \alpha_1, \alpha_2, \kappa, \rho)$. The 0.2 in the $w(t)$ is the upper limit of making a transition from the disease-free state to the preclinical state. Using Markov Chain Monte Carlo (MCMC), 2000 Bayesian posterior samples (θ_j^*) were generated, for details, see Wu, et al 2005. Using the HIP data, the posterior predictive probability of each case can be estimated as

$$\begin{aligned} P(\text{Case } i|T \geq t_{K_1}, H_{K_1}, HIP) &= \int P(\text{Case } i, \theta|T \geq t_{K_1}, H_{K_1}, HIP) d\theta \\ &= \int P(\text{Case } i|T \geq t_{K_1}, H_{K_1}, \theta) f(\theta|HIP) d\theta \end{aligned}$$

$$\approx \quad \frac{1}{n}\sum_{j=1}^{n} P(\text{Case } i|T \geq t_{K_1}, H_{K_1}, \theta_j^*) \qquad (22)$$

where θ_j^* is the random sample drawn from the posterior distribution $f(\theta|HIP)$ and $n = 2000$ is the posterior sample size (Wu et al 2005).

3.2 Results

We applied (22) to the 2000 MCMC posterior samples, to conduct Bayesian inference on three hypothetical cohorts of asympomatic women currently at age 60, 70 and 80, assuming that they have started their first screening at age 50, with either annual screening or biennial screening until now ($\Delta_1 = 1$ or 2 years). For each group, we assumed that they are superficially healthy at their current age with no cancer ever found, and we have examined annual or biennial screening interval in the future ($\Delta_2 = 1, 2$ years). The number of screens $K = K(T) = \lceil (T - t_{K_1})/\Delta_2 \rceil$ is a function of the lifetime T, therefore it is a random variable in the simulation.

For the lifetime distribution, we used the conditional lifetime density for females at current age 60, 70, and 80 derived from the actuarial life table from the Social Security Administration (SSA) website. Details on how to derive the conditional PDF is given in Wu, et al (2012), and the corresponding PDF curves were ploted in Figure 3. The probabilities of each of the four cases $P(\text{Case } i|H_{K_1}, T \geq t_{K_1}, HIP)$ with standard errors were reported in Table 1.

The probability of "Overdiagnosis" for all age groups is very small, between 0.20% and 0.32% from Table 1. This probability decreases when a woman's current age increases, and it decreases as the screening time interval (Δ_2) increases. It depends more on the future screening interval Δ_2, and less on the historic screening interval Δ_1.

The probability of "True-early-detection" is higher for annual screening group in the future than that for biennial screening within each age group. That is, This probability decreases as the future screening time interval increases. The probability is also lower when current age increases from 60 to 80; Similar to the probability of over-diagnosis, the past screening interval causes little changes on the probability of "true-early-detection."

The probability of "No-early-detection" is between 0.45% and 2.76%. It increases as the future screening interval increases; and it decreases when current age increases.

The probability of "Symptom-free-life" is very high. It increases from about 93% to 98% when the current age increases from 60 to 80. It is comparatively stable: not changing much with the past or the future screening interval.

Most people are more concerned with the percentage of over-diagnosis among the screen-detected cases. The estimated probabilities (and its 95% HPD intervals) of "Over-diagnosis" and "True-early-detection" when it is a screen-detected case were listed in Table 2.

The percentage of "Overdiagnosis" among the screen-detected is about 7%, 10% and 15% for current age groups 60, 70, and 80 respectively, so this probability increases significantly with a person's current age. However, it doesn't change much with future or past screening intervals. The 95% highest probability density (HPD) interval for this percentage is pretty wide also. Since the probability of "true-early-detection" is one minus the probability of "over-diagnosis" conditional on that it is a screen-detected case, this probability is contrary to that of "over-diagnosis": it decreases with a person's current age, and remains stable with different future and past screening intervals. The length of the 95% HPD interval for these two probabilities (percentages) increases as a person's current age increases, showing large variation with an advanced age.

4. Discussion

This study provides a way to evaluate the long term outcomes of the screening program under different future screening schedules for old people with a screening history. We separated all asymptomatic participants in a screening program into four mutually exclusive groups: symptom-free-life, no-early-detection, true-early-detection, and over-diagnosis. This kind of analysis can provide policy makers with estimates of the probability of true-early-detection, overdiagnosis, and other outcomes that result from different future screening frequency. We used a Bayesian approach to incorporate uncertainty easily, and to calculate the variations and the credible invervals of the probability (or percentage).

This is an extension of our previous work on long term outcomes for people without a screening history (Wu, Kafadar, Rosner 2014). Comparing results for the 60-year-old age group, when there is no screening history, that is, when a 60-years-old woman takes her initial screening at 60 and who plans to take future screening annually, the probability for the 4 cases: symptom-free-life, no-early-detection, true-early-detection, and over-diagnosis, are 93.21%, 1.39%, 5.03%, 0.33% respectively. Which is similar to the results when a 60-years-old started her screening at 50, with either annual or biennial screening history, and plan to take screening annually in the future (see the first 2 rows in Table 1). Similarly, we can compare the age 60 cohort without a screening history and biennial screening in the future (Wu et al 2014 Table 2), with the age 60 cohort with an annual/biennial history, and biennial screening in the future, the results is very close too. When we focus on the screen-detected cases, the probability of over-diagnosis is only slightly higher for the age 60 cohort with a screening history. In summary, it seems that screening history didn't have much impact for the long term outcomes in the future if we know that they are asymptomatic and cancer-free at their current age.

We want to point out that from our simulation, the probability of over-diagnosis is increasing as people aging, so it seems wise to take fewer screening exams for the 80 years old. This is the first model to handle the situation of people with a screening history, and it provides a systematic approach to evaluate the long-term outcomes in regular screening. Our method is predictive: we use existing data to obtain information on three key parameters: screening sensitivity, sojourn time distribution, and transition probability, then use these parameters and our probability model to predict the probability of true-early-detection, no-early-detection, over-diagnosis and symptom-free-life for different age groups with different screening histories, and different future screening frequencies. we hope our model will help policy makers evaluate a screening program's long-term effects more appropriately.

REFERENCES

badgwell Badgwell B.D., Giordano S.H., Duan Z.Z., Fang S, Bedrosian I, Kuerer H.M., Singletary S.E., Hunt K.K., Nortobagyi G.N., and Babiera G.(2008). Mammography before diagnosis among women age 80 years and older with breast cancer. *Journal of Clinical Oncology* 26, 2482-2488.

National Institute of Health (2010). `http://seer.cancer.gov/csr/1975_2007/results_merged/topic_lifetime_risk.pdf`

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.* The Johns Hopkins University Press, Baltimore.

Social Security Administration (2010). Period Life Table. Actuarial Publications. `http://www.ssa.gov/OACT/STATS/table4c6.html`.

Wu D, Kafadar K, and Rosner GL. (2014). Inference of long term effects and over-diagnosis in periodic cancer screening. *Statistica Sinica*. 24, 815-831.

Wu, D., Kafadar, K, Rosner, G.L., and Broemeling, L.D.(2012). The lead time distribution when lifetime is subject to competing risks in cancer screening. *The International Journal of Biostatistics*. DOI: 10.1515/1557-4679.1363.

Wu, D., and Perez, A. (2011). A limited review of over-diagnosis methods and long term effects in breast cancer screening. *Oncology Review* 5, 143-147.

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* 61, 1056-1063.

Wu, D., Rosner, G.L., and Broemeling, L.D.(2007). Bayesian inference for the lead time in periodic cancer screening. *Biometrics* 63, 873-880.

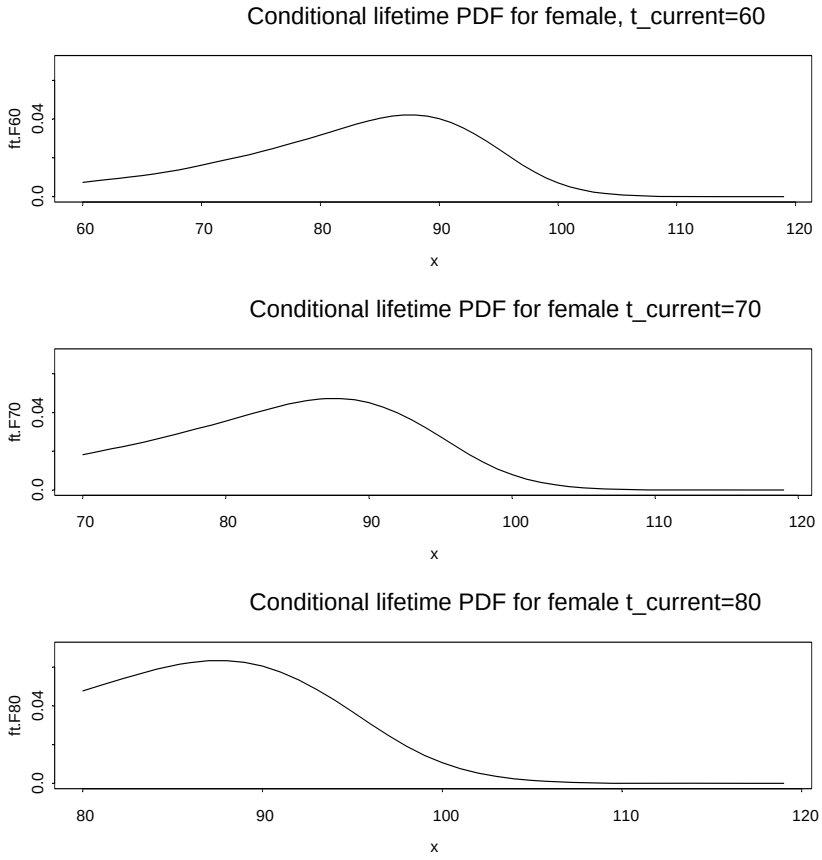

Figure 3: Conditional PDF of lifetime for females in the US at current age t_0.

Table 1: A projection of breast cancer screening outcomes using the HIP data

$(\Delta_1, \Delta_2)^a$	P[b](SympF)	P(NoED)	P(TrueED)	P(OverD)
Initial screen age $t_0 = 50$**, current age** $t_{K_1} = 60$				
(1 yr, 1 yr)	93.62(1.07)	1.34(0.63)	4.69(1.04)	0.32(0.22)
(2 yr, 1 yr)	93.52(1.06)	1.34(0.62)	4.78(1.04)	0.32(0.22)
(1 yr, 2 yr)	93.69(1.08)	2.75(0.89)	3.28(0.71)	0.25(0.19)
(2 yr, 2 yr)	93.59(1.06)	2.76(0.88)	3.36(0.72)	0.25(0.20)
Initial screen age $t_0 = 50$**, current age** $t_{K_1} = 70$				
(1 yr, 1 yr)	95.84(0.69)	0.84(0.48)	3.00(0.72)	0.30(0.19)
(2 yr, 1 yr)	95.73(0.68)	0.85(0.48)	3.09(0.72)	0.31(0.20)
(1 yr, 2 yr)	95.91(0.69)	1.73(0.62)	2.11(0.50)	0.23(0.17)
(2 yr, 2 yr)	95.81(0.68)	1.74(0.62)	2.20(0.52)	0.24(0.18)
Initial screen age $t_0 = 50$**, current age** $t_{K_1} = 80$				
(1 yr, 1 yr)	97.74(0.34)	0.45(0.31)	1.55(0.40)	0.27(0.16)
(2 yr, 1 yr)	97.64(0.34)	0.45(0.31)	1.63(0.41)	0.28(0.17)
(1 yr, 2 yr)	97.80(0.34)	0.90(0.35)	1.09(0.28)	0.20(0.14)
(2 yr, 2 yr)	97.70(0.34)	0.91(0.35)	1.17(0.31)	0.21(0.15)

[a] Δ_1, Δ_2 are scr. interval in history and in future correspondingly.
[b] The mean probability (with standard error) are in percentage.

Table 2: Estimated probability of true-early-detection and over-diagnosis in screen-detected cases (with 95% credible interval)

(Δ_1, Δ_2)	P(TrueED$\|D^e$)	P(OverD$\|D$)
Initial screen age $t_0 = 50$**, current age** $t_{K_1} = 60$		
(1 yr, 1 yr)	93.15 (74.55, 97.87)	6.85 (2.13, 25.45)
(2 yr, 1 yr)	93.24 (74.84, 97.88)	6.76 (2.12, 25.16)
(1 yr, 2 yr)	92.81 (73.18, 97.94)	7.19 (2.06, 26.82)
(2 yr, 2 yr)	92.93 (73.44, 97.96)	7.07 (2.04, 26.56)
Initial screen age $t_0 = 50$**, current age** $t_{K_1} = 70$		
(1 yr, 1 yr)	90.32 (67.38, 96.70)	9.68 (3.30, 32.62)
(2 yr, 1 yr)	90.49 (67.98, 96.73)	9.51 (3.27, 32.02)
(1 yr, 2 yr)	90.00 (66.09, 96.80)	10.00 (3.20, 33.91)
(2 yr, 2 yr)	90.21 (67.18, 96.84)	9.79 (3.16, 32.82)
Initial screen age $t_0 = 50$**, current age** $t_{K_1} = 80$		
(1 yr, 1 yr)	84.89 (56.66, 94.22)	15.11 (5.78, 43.34)
(2 yr, 1 yr)	85.16 (57.19, 94.27)	14.84 (5.73, 42.81)
(1 yr, 2 yr)	84.56 (55.40, 94.36)	15.44 (5.64, 44.60)
(2 yr, 2 yr)	84.92 (55.76, 94.38)	15.08 (5.62, 44.24)

[d] The estimated conditional probability was calculated as $p_i^*/(p_3^*+p_4^*), i = 3, 4$, for each of the 2000 posterior samples, then averaged. It is a percentage.
[e] The event $ScrD = \{$Screen-detected case$\}$