Statistics and Its Interface Volume 15 (2022) 503–514

When to initiate cancer screening exam?

Dongfeng Wu

A probability method is developed to decide when to initiate cancer screening for asymptomatic individuals. The probability of incidence is a function of screening sensitivity, time duration in the disease-free state and sojourn time in the preclinical state; and it is monotonically increasing as time increases, given a person's current age. So a unique solution of the first screening time can be found by limiting this probability to a small value, such as 10% or 20%. That is, with 90% or 80% probability, one will not be a clinical incident case before the first exam. After this age is found, we can further estimate the lead time distribution and probability of over-diagnosis if one would be diagnosed with cancer at the first exam. Simulations were carried out under different scenarios; and the method was applied to two heavy smoker cohorts in the National Lung Screening Trial using low-dose computerized tomography. The method is applicable to other kinds of cancer screening. The predictive information can be used by physicians or individuals at risk to make informed decisions on when to initiate screening.

Keywords and phrases: Scheduling, Sensitivity, Sojourn time, Transition density, Lead time, Overdiagnosis.

1. INTRODUCTION

Early detection and effective treatments are critical to increase the cure rate and prolong survival of cancer patients. The primary technique for early detection is screening exam, with the goal that the disease may be found before symptoms are present. Although screening programs for different kinds of cancer have been carried out in the past six decades in North America, and the U.S. Preventive Services Task Force updates their recommendations regarding different kinds of cancer screening periodically [1], there are still many unanswered questions in the designing of screening. One major concern is: at what age should screening programs be initiated [3]? Suppose a superficially healthy person goes to a physician for regular health check up, should the physician give any advice on when to initiate screening for a certain kind of cancer based on the person's current age and other factors? We will develop a probability method to address this problem for asymptomatic individuals based on their current age.

There is almost no research in this area so far. There were research regarding optimal screening; however, these existing methods usually deal with how to schedule $(n+1)$ exams in a fixed age interval using a utility function [23, 24]. There are some other approaches to solve the scheduling of exams [8, 9], but all of them involved a utility function requiring specified costs or weights.

We will develop a new approach to handle the problem. We will not use any utility function, costs, nor weights, which is subjective. Instead, we will study the risk (probability) of incidence from one's current age, assuming one is asymptomatic and haven't been screened so far, the first screening time (or age) is chosen, such that the risk (or probability) of incidence is limited by some preselected small value, such as 10% or less. Therefore, with 90% or more chance, a person at risk would not become a clinically incidence case before the first screening exam if s/he follows this screening guideline. And for those who would be diagnosed with cancer at the first exam, we derive the lead time distribution and the probability of overdiagnosis. This provides predictive information regarding the initial screening age on a personal level. Policy makers or individuals can use this information to make informed decisions. We will use lung cancer as an example in this research, since it is the leading cause of cancer death in the United States, and counts about 22.4% of all cancer deaths [2]. The developed method can be applied to other kinds of screening as well.

The paper is organized as follows: in Section 2, we derive the probability of incidence given one is asymptomatic at current age a_0, and show that there is a unique solution of age t_0, given a fixed probability of incidence p; then we drive the distribution of the lead time and probability of overdiagnosis if one were diagnosed with cancer at the first screening at the future age t_0. Simulation results are presented in Section 3. We applied the new method to the National Lung Screening Trial (NLST) low-dose computed tomograph (CT) cohort data for male vs. female heavy smokers in Section 4; and we ended with a discussion in Section 5.

2. METHOD

We will use the well-known disease progressive model, which assumes that tumor develops through three states: $S_0 \to S_p \to S_c$ [22]. S_0 is the disease-free state, in which there is no disease or the disease is at a very early stage and cannot be detected by any screening exam. S_p refers to the preclinical state, in which a person has the disease that can be detected by a screening exam even though there is

no symptom. S_c represents the clinical state, where clinical symptoms have presented. We will use female lung cancer in the description and solution of the problem, while the procedure and derived formulas are equally valid for other kinds of cancer screening.

2.1 Probability of incidence and optimal first screening age

Suppose a woman at her current age a_0 is asymptomatic and has not taken any screening yet, should she start her first exam immediately, or wait for some time? We will develop a simple protocol to help with this decision making. The goal is to make sure that the probability of clinical incidence from now until her first screening is limited to a small value, such as 0.1 or 0.2. Suppose that her first screening

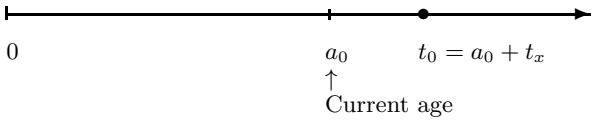

Figure 1. An individual's current age and the first exam.

time will happen at her age $t_0 = a_0 + t_x$, with $t_x > 0$, see Figure 1. We want to find the value t_x such that the probability of incidence is limited to a pre-selected value p. We start by defining a few events:

$$\begin{aligned} H_0 &= \{\text{One is asymptomatic in } [0, a_0], \text{and without any screening so far}\};\\ I_0 &= \{\text{One is asymptomatic in } [0, a_0], \text{and will be a clinical incident case in } (a_0, t_0)\};\\ D_0 &= \{\text{One is asymptomatic in } [0, a_0], \text{and will be diagnosed at age } t_0\};\\ A_0 &= \{\text{One is asymptomatic in } [0, t_0), \text{take the first screening at } t_0, \text{ and with a negative result}\}. \end{aligned}$$

The last three events are mutually exclusive, and they form a partition of the historic sample space H_0: $I_0 \cup D_0 \cup A_0 = H_0$. Let β be the sensitivity of the exam at age t_0, i.e., the probability that the screening is positive given that the individual is in the preclinical state S_p. We let X be the time duration in the disease-free state S_0, with a probability density function (PDF) $w(x)$, it is called transition probability or transition density in other papers; we let Y be the sojourn time, the length of time in the preclinical state S_p, with a PDF $q(y)$; and $Q(y) = \int_y^\infty q(x)dx$ is the survival function of the sojourn time Y. We assume that the sojourn time Y and the time duration in the disease-free state X are independent.

Now the probability of incidence among the people at risk can be calculated by:

$$P(I_0|I_0 \cup D_0) \tag{1}$$

$$= \frac{P(I_0)}{P(I_0 \cup D_0)} = \frac{P(I_0)}{P(I_0) + P(D_0)}$$

The numerator is the probability of incidence in (a_0, t_0), which could happen in two cases: (i) she enters the preclinical state at age $x \in (0, a_0)$ and her sojourn time is between $(a_0 - x)$ and $(t_0 - x)$, or (ii) she enters the preclinical state at age $x \in (a_0, t_0)$ and her sojourn time is less than $(t_0 - x)$. Hence,

$$\begin{aligned} P(I_0) &= \int_0^{a_0} w(x)[Q(a_0 - x) - Q(t_0 - x)]dx \\ &\quad + \int_{a_0}^{t_0} w(x)[1 - Q(t_0 - x)]dx. \end{aligned} \tag{2}$$

And $P(D_0)$ is the probability of detection at the first exam:

$$P(D_0) = \beta \int_0^{t_0} w(x)Q(t_0 - x)dx. \tag{3}$$

Since $P(I_0|I_0 \cup D_0)$ is monotonically increasing with the time interval t_x, and remember $t_0 = a_0 + t_x$, hence this probability is increasing with t_0. For any given p between 0 and 1, there is a unique solution t_0, such that

$$P(I_0|I_0 \cup D_0) = p.$$

We can use the binary search method to find the age t_0 for any given $p \in (0, 1)$.

2.2 Lead time distribution at the first exam

After we find the numerical solution $t_0 = t_0(p)$, we can derive the formula for the lead time at age t_0 if one were diagnosed with cancer at the first exam. The lead time is defined as the diagnosis time advanced by screening; in other words, the lead time is the time interval between the diagnosed cancer and the presence of clinical symptoms. Suppose one would be diagnosed with cancer at the initial exam at her future age t_0, let L represents the lead time, then the PDF of the lead time would be:

$$f_L(z|D_0) = \frac{f_L(z, D_0)}{P(D_0)}, \quad \text{for } z \in (0, \infty). \tag{4}$$

Where the numerator

$$f_L(z, D_0) = \beta \int_0^{t_0} w(x)q(t_0 + z - x)dx. \tag{5}$$

This is because she must have entered the preclinical state at some age $x \in (0, t_0)$, and her lead time is z; which means, if she were not screened, she would be a clinical incident case at age $(t_0 + z)$, therefore her sojourn time should be $(t_0 + z - x)$. The denominator $P(D_0)$ was given in equation (3). It is easy to prove that:

$$\int_0^\infty f_L(z|D_0)dz = 1. \tag{6}$$

Thus the derived formula is a valid PDF.

2.3 Probability of overdiagnosis at the first exam

We can find the probability of overdiagnosis (OverD) and true-early-detection (TrueED) if one were diagnosed with cancer at age t_0 for the first time. Given a fixed lifetime $T = t(> t_0)$, the probability of overdiagnosis and that of true-early-detection would be:

$$(7)\quad P(\text{OverD}|D_0, T=t) = \frac{P(\text{OverD}, D_0|T=t)}{P(D_0|T=t)},$$

$$(8)\quad P(\text{TrueED}|D_0, T=t) = \frac{P(\text{TrueED}, D_0|T=t)}{P(D_0|T=t)}.$$

Since $P(D_0|T=t) = P(D_0)$, we only need to find out the two numerators. To calculate $P(\text{OverD}, D_0|T=t)$, that is, one would be diagnosed with cancer at the initial exam at age t_0, but the symptom would not present until after her life time t; therefore, she must have entered the preclinical state at age $x \in (0, t_0)$, and her sojourn time is longer than $(t-x)$. Therefore,

$$(9)\quad P(\text{OverD}, D_0|T=t) = \beta \int_0^{t_0} w(x)Q(t-x)dx.$$

For the case of true-early-detection, her symptom would have appeared before her life time t, therefore, her sojourn time is between $(t_0 - x)$ and $(t-x)$:

(10)

$$P(\text{TrueED}, D_0|T=t) = \beta \int_0^{t_0} w(x)[Q(t_0-x) - Q(t-x)]dx.$$

And it is easy to verify that:

$$P(\text{OverD}, D_0|T=t) + P(\text{TrueED}, D_0|T=t) = P(D_0).$$

Hence,

$$P(\text{OverD}|D_0, T=t) + P(\text{TrueED}|D_0, T=t) = 1.$$

Now if we allow human lifetime T to be a random variable, then,

$$(11)\quad P(\text{OverD}|D_0, T>t_0) = \int_{t_0}^{\infty} P(\text{OverD}|D_0, T=t) f_T(t|T>t_0)dt,$$

$$(12)\quad P(\text{TrueED}|D_0, T>t_0) = \int_{t_0}^{\infty} P(\text{TrueED}|D_0, T=t) f_T(t|T>t_0)dt.$$

The conditional PDF of human lifetime $f_T(t|T>t_0) = \frac{f_T(t)}{1-F_T(t_0)}$, if $t \geq t_0$. And it is obtained by transforming the actuarial life table from the United States Social Security Administration[1]; for details, see Wu et al. 2012 [12]. We can prove that:

$$P(\text{OverD}|D_0, T>t_0) + P(\text{TrueED}|D_0, T>t_0) = 1.$$

[1] http://www.ssa.gov/OACT/STATS/table4c6.html, last access 11/19/2020.

3. SIMULATION STUDY

Since the probability of incidence $P(I_0|I_0 \cup D_0)$ is a function of one's current age a_0, the three key parameters (sensitivity β, sojourn time distribution $q(x)$ or $Q(x)$, transition density $w(t)$), and the first screening age t_0, we will find the optimal initial screening age t_0 based on the given factors. We will use the female lifetime distribution in the simulation; simulation results using the male lifetime distribution are similar and omit here. Specifically, we selected the following scenarios for the simulation:

1. Four values of the probability of incidence: $p = 0.05, 0.10, 0.15, 0.20$;
2. Three different screening sensitivities: $\beta = 0.80, 0.90, 0.95$;
3. Four different mean sojourn time (MST): MST = 1.5, 2.5, 5.0, 10.0 years;
4. Three different current age a_0: 45, 50, 55 years.

Based on the previous research in cancer screening, we use the parametric model of the transition density and the sojourn time [17]:

$$(13)\quad w(t|\mu, \sigma^2) = \frac{0.3}{\sqrt{2\pi}\sigma t} \exp\{-(\log t - \mu)^2/(2\sigma^2)\},$$

$$(14)\quad Q(x|\lambda, \alpha) = \exp(-\lambda x^{\alpha}), \quad \lambda > 0, \alpha > 0,$$

$$(15)\quad q(x|\lambda, \alpha) = \alpha\lambda x^{\alpha-1} Q(x|\lambda, \alpha).$$

For $w(t)$, the input parameters of μ and σ^2 were chosen, so that the mean/median/mode of the lung cancer transition age into the preclinical state was around 70 years old [17], that will give $\mu = 4.25$. Based on our previous research, the σ^2 has a mean value of 0.021 for males, 0.026 for females, and 0.022 for both genders [17], so we picked $\sigma^2 = 0.02$ in the simulation.

For $q(x)$, the parameters (λ, α) were chosen for four different mean sojourn times of 1.5, 2.5, 5.0, and 10.0 years (correspondlingly represents fast, moderate, and slow growing tumor) with: $\alpha = 3.0, \lambda = 0.21098, 0.04557, 0.00570, 0.00071$.

Table 1 provides the optimal initial screening age t_0^* using the method in section 2 and the binary search, for different p, the pre-selected probability of incidence before the first exam. It was carried out under different sensitivity β, mean sojourn time (MST), and current age a_0. This is how to read the table: look at the block under "MST = 2.5 years," the numbers under $a_0 = 45$ and $\beta = 0.95$ are 45.24, 45.53, 45.91 and 46.40. That is, if someone wants to have a 95% probability of no clinical incident before the first exam, then she should take the screening at age 45.24 (or after 3 months since her current age is 45); or if someone wants to have a 80% chance of not being a clinical case before the first exam, she can come back for screenging at age 46.40 (or

Table 1. Optimal initial screening age t_0^* (in years) found by binary search

p	MST = 1.5 years								
	$a_0 = 45$			$a_0 = 50$			$a_0 = 55$		
	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	45.09	45.10	45.11	50.08	50.09	50.10	55.07	55.08	55.09
0.10	45.19	45.22	45.23	50.17	50.19	50.21	55.16	55.18	55.19
0.15	45.32	45.36	45.38	50.28	50.31	50.33	55.25	55.28	55.30
0.20	45.46	45.53	45.56	50.40	50.45	50.48	55.36	55.41	55.43
p	MST = 2.5 years								
	$a_0 = 45$			$a_0 = 50$			$a_0 = 55$		
	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	45.20	45.22	45.24	50.16	50.18	50.19	55.14	55.16	55.17
0.10	45.44	45.50	45.53	50.35	50.40	50.42	55.30	55.34	55.36
0.15	45.74	45.85	45.91	50.57	50.65	50.69	55.48	55.54	55.58
0.20	46.13	46.31	46.40	50.84	50.96	51.03	55.69	55.79	55.84
p	MST = 5.0 years								
	$a_0 = 45$			$a_0 = 50$			$a_0 = 55$		
	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	45.84	45.96	46.03	50.54	50.62	50.65	55.40	55.45	55.47
0.10	47.17	47.55	47.76	51.25	51.44	51.54	55.87	55.99	56.05
0.15	49.16	49.88	50.24	52.19	52.55	52.73	56.45	56.65	56.76
0.20	51.52	52.44	52.88	53.42	53.98	54.26	57.15	57.46	57.62
p	MST = 10.0 years								
	$a_0 = 45$			$a_0 = 50$			$a_0 = 55$		
	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	52.13	52.92	53.28	53.46	53.98	54.23	56.84	57.10	57.23
0.10	57.37	58.24	58.64	57.63	58.44	58.82	59.20	59.76	60.03
0.15	60.84	61.75	62.17	60.94	61.82	62.24	61.69	62.44	62.80
0.20	63.56	64.49	64.92	63.60	64.53	64.96	64.02	64.87	65.27

after a year and 5 months). It shows that as the screening sensitivity increases from 0.8 to 0.95, the first scheduling age is slightly increasing if other factors were the same. However, the first screening age is dramatically increasing as the incidence probability p increases, or as the MST increases. The ideal first screening age t_0^* also changes with one's current age a_0, the time interval $(t_0^* - a_0)$ is decreasing as a_0 increases if other factors were the same. We want to point out that once the t_0^* is found, the lead time distribution $f_L(z|D_0)$ and the probability of overdiagnosis $P(\text{OverD}|D_0, T > t_0^*)$ don't directly depend on β, p and a_0; but they both depend on $t_0^*, w(t)$ and $Q(x)$. This is clearly shown by the formulas in subsections 2.2, 2.3, and is verified by the simulations.

Tables 2, 3 and 4 present the estimated mean, median, mode (i.e., central locations) and standard deviation (the spread) of the lead time at the optimal first screening time/age t_0^*, when the current age a_0 is 45, 50, or 55 years correspondingly. The results in the three tables show similar pattern: i) as the mean sojourn time increases, the mean, median and mode of the lead time increases as well. ii) the lead time distribution depends very little on the incidence probability p and the sensitivity β using the optimal scheduling time t_0^*. As the current age a_0 increases, the mean, median and mode of the lead time becomes smaller, although the standard deviation is almost the same.

The lead time PDF curves under different factors: p, β, a_0 and MST were plotted in Figure 2. The four panels showed the estimated lead time density when the optimal first screening age t_0^* was used, with three factors fixed and only the fourth factor was allowed to change. It shows that given t_0^*, the lead time distribution barely changes with the incidence probability p and sensitivity β. However, it changes quite a bit with one's current age a_0 and the mean sojourn time (MST): as the a_0 increases, the mean, median, and mode of the lead time slightly decrease; and as the MST increases, the central location of the lead time increases.

Table 5 provides the estimated probability of overdiagnosis (in percentage) using the optimal initial scheduling age t_0^*. That is, if one took the first screening exam at the age t_0^* provided by Table 1 and were diagnosed with cancer, then the probability of overdiganosis would be the result in Table 5. The probability of overdiagnosis is increaseing as the MST increases; the probability is increasing as p increases; and it is increasing as one's current age a_0 increases; but it barely changes with the sensitivity β. In general, when the MST is less than or equal to five years, the probability of overdiagnosis usually is less than 2%, which is negligible. In summery, the probability of overdiagnosis is very small overall, and the largest value of the probability of overdiagnosis is less than 10% when the MST is 10 years.

Table 2. Estimated mean, median, mode and standard deviation of the lead time at the optimal time t_0^* when $a_0 = 45$

	MST = 1.5 years		
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	0.94, 0.89, 0.73, 0.59	0.94, 0.89, 0.73, 0.59	0.94, 0.89, 0.73, 0.59
0.10	0.94, 0.89, 0.73, 0.59	0.94, 0.89, 0.73, 0.59	0.94, 0.89, 0.73, 0.59
0.15	0.94, 0.88, 0.72, 0.59	0.94, 0.88, 0.72, 0.59	0.94, 0.88, 0.72, 0.59
0.20	0.94, 0.88, 0.72, 0.59	0.94, 0.88, 0.72, 0.59	0.94, 0.88, 0.72, 0.59
	MST = 2.5 years		
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	1.66, 1.59, 1.45, 0.98	1.66, 1.59, 1.45, 0.98	1.66, 1.59, 1.45, 0.98
0.10	1.66, 1.59, 1.44, 0.98	1.66, 1.59, 1.44, 0.98	1.66, 1.59, 1.44, 0.98
0.15	1.65, 1.58, 1.43, 0.98	1.65, 1.58, 1.43, 0.98	1.65, 1.58, 1.43, 0.98
0.20	1.65, 1.57, 1.42, 0.98	1.65, 1.57, 1.41, 0.98	1.64, 1.57, 1.41, 0.98
	MST = 5.0 years		
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	3.71, 3.62, 3.52, 1.95	3.70, 3.62, 3.51, 1.95	3.70, 3.61, 3.51, 1.95
0.10	3.65, 3.56, 3.44, 1.96	3.63, 3.54, 3.42, 1.96	3.63, 3.53, 3.40, 1.96
0.15	3.57, 3.46, 3.32, 1.96	3.54, 3.43, 3.27, 1.96	3.52, 3.41, 3.25, 1.96
0.20	3.47, 3.35, 3.16, 1.96	3.43, 3.30, 3.10, 1.96	3.41, 3.28, 3.07, 1.96
	MST = 10.0 years		
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	7.91, 7.77, 7.61, 3.88	7.83, 7.69, 7.53, 3.88	7.80, 7.65, 7.48, 3.88
0.10	7.40, 7.21, 6.97, 3.91	7.31, 7.12, 6.85, 3.92	7.27, 7.07, 6.79, 3.92
0.15	7.05, 6.82, 6.45, 3.92	6.96, 6.71, 6.30, 3.92	6.92, 6.66, 6.23, 3.92
0.20	6.78, 6.49, 5.98, 3.92	6.68, 6.38, 5.80, 3.92	6.64, 6.33, 5.71, 3.92

Table 3. Estimated mean, median, mode and standard deviation of the lead time at the optimal time t_0^* when $a_0 = 50$

	MST = 1.5 years		
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	0.91, 0.85, 0.63, 0.58	0.91, 0.85, 0.63, 0.58	0.91, 0.85, 0.63, 0.58
0.10	0.91, 0.85, 0.63, 0.58	0.91, 0.85, 0.63, 0.58	0.91, 0.85, 0.63, 0.58
0.15	0.91, 0.85, 0.63, 0.58	0.91, 0.85, 0.63, 0.58	0.91, 0.85, 0.63, 0.58
0.20	0.91, 0.85, 0.63, 0.58	0.91, 0.85, 0.62, 0.58	0.91, 0.85, 0.62, 0.58
	MST = 2.5 years		
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	1.59, 1.51, 1.28, 0.98	1.59, 1.51, 1.28, 0.98	1.59, 1.51, 1.28, 0.98
0.10	1.59, 1.50, 1.28, 0.98	1.59, 1.50, 1.28, 0.98	1.59, 1.50, 1.28, 0.98
0.15	1.59, 1.50, 1.27, 0.98	1.59, 1.50, 1.27, 0.98	1.59, 1.50, 1.27, 0.98
0.20	1.58, 1.49, 1.26, 0.98	1.58, 1.49, 1.26, 0.98	1.58, 1.49, 1.25, 0.98
	MST = 5.0 years		
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	3.51, 3.39, 3.23, 1.96	3.51, 3.39, 3.22, 1.96	3.50, 3.39, 3.22, 1.96
0.10	3.48, 3.36, 3.18, 1.96	3.47, 3.35, 3.17, 1.96	3.47, 3.35, 3.16, 1.96
0.15	3.44, 3.32, 3.11, 1.96	3.43, 3.30, 3.09, 1.96	3.42, 3.29, 3.08, 1.96
0.20	3.39, 3.26, 3.03, 1.96	3.37, 3.23, 2.99, 1.96	3.36, 3.22, 2.97, 1.96
	MST = 10.0 years		
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	7.74, 7.59, 7.42, 3.87	7.69, 7.54, 7.36, 3.87	7.66, 7.51, 7.33, 3.88
0.10	7.33, 7.14, 6.89, 3.90	7.25, 7.05, 6.78, 3.90	7.21, 7.01, 6.72, 3.90
0.15	7.00, 6.76, 6.39, 3.90	6.91, 6.66, 6.25, 3.90	6.87, 6.61, 6.18, 3.90
0.20	6.74, 6.45, 5.93, 3.90	6.64, 6.34, 5.75, 3.90	6.60, 6.29, 5.67, 3.90

Table 4. Estimated mean, median, mode and standard deviation of the lead time at the optimal time t_0^* when $a_0 = 55$

	MST = 1.5 years		
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	0.89, 0.83, 0.53, 0.58	0.89, 0.83, 0.53, 0.58	0.89, 0.83, 0.53, 0.58
0.10	0.89, 0.82, 0.53, 0.58	0.89, 0.82, 0.53, 0.58	0.89, 0.82, 0.53, 0.58
0.15	0.89, 0.82, 0.53, 0.58	0.89, 0.82, 0.53, 0.58	0.89, 0.82, 0.53, 0.58
0.20	0.89, 0.82, 0.53, 0.58	0.89, 0.82, 0.52, 0.58	0.89, 0.82, 0.52, 0.58
	MST = 2.5 years		
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	1.53, 1.43, 1.10, 0.98	1.53, 1.43, 1.10, 0.98	1.53, 1.43, 1.10, 0.98
0.10	1.53, 1.43, 1.10, 0.98	1.53, 1.43, 1.10, 0.98	1.53, 1.43, 1.10, 0.98
0.15	1.53, 1.43, 1.09, 0.98	1.53, 1.43, 1.09, 0.98	1.53, 1.43, 1.09, 0.98
0.20	1.53, 1.43, 1.08, 0.97	1.53, 1.43, 1.08, 0.97	1.53, 1.42, 1.08, 0.97
	MST = 5.0 years		
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	3.32, 3.17, 2.88, 1.96	3.32, 3.17, 2.88, 1.96	3.31, 3.17, 2.87, 1.96
0.10	3.30, 3.15, 2.84, 1.96	3.30, 3.14, 2.83, 1.96	3.29, 3.14, 2.83, 1.96
0.15	3.28, 3.12, 2.80, 1.96	3.27, 3.11, 2.78, 1.96	3.27, 3.11, 2.77, 1.96
0.20	3.25, 3.09, 2.74, 1.96	3.24, 3.08, 2.72, 1.96	3.24, 3.07, 2.70, 1.96
	MST = 10.0 years		
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	7.41, 7.23, 7.00, 3.89	7.39, 7.20, 6.96, 3.89	7.37, 7.19, 6.95, 3.89
0.10	7.18, 6.96, 6.67, 3.90	7.12, 6.90, 6.58, 3.90	7.09, 6.87, 6.54, 3.90
0.15	6.93, 6.68, 6.27, 3.90	6.85, 6.59, 6.14, 3.90	6.82, 6.55, 6.08, 3.90
0.20	6.69, 6.40, 5.85, 3.90	6.61, 6.30, 5.68, 3.90	6.57, 6.25, 5.60, 3.90

Table 5. Estimated probability of overdiagnosis (in percentage) at the initial screening age t_0^*

p	MST = 1.5 years								
	$a_0 = 45$			$a_0 = 50$			$a_0 = 55$		
	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	0.19	0.19	0.19	0.30	0.30	0.30	0.43	0.43	0.43
0.10	0.19	0.20	0.20	0.30	0.30	0.30	0.43	0.43	0.43
0.15	0.20	0.20	0.20	0.30	0.30	0.30	0.43	0.43	0.43
0.20	0.20	0.20	0.20	0.30	0.30	0.30	0.43	0.43	0.43
p	MST = 2.5 years								
	$a_0 = 45$			$a_0 = 50$			$a_0 = 55$		
	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	0.36	0.36	0.36	0.54	0.54	0.54	0.76	0.76	0.76
0.10	0.37	0.37	0.37	0.55	0.55	0.55	0.76	0.76	0.76
0.15	0.38	0.38	0.38	0.56	0.56	0.56	0.77	0.77	0.77
0.20	0.39	0.39	0.40	0.57	0.57	0.58	0.78	0.78	0.79
p	MST = 5.0 years								
	$a_0 = 45$			$a_0 = 50$			$a_0 = 55$		
	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	0.96	0.97	0.98	1.35	1.35	1.36	1.79	1.80	1.80
0.10	1.07	1.09	1.11	1.41	1.43	1.43	1.84	1.85	1.86
0.15	1.23	1.29	1.32	1.49	1.52	1.54	1.90	1.92	1.93
0.20	1.43	1.51	1.55	1.60	1.65	1.68	1.98	2.00	2.02
p	MST = 10.0 years								
	$a_0 = 45$			$a_0 = 50$			$a_0 = 55$		
	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	4.15	4.33	4.43	4.44	4.57	4.64	5.39	5.50	5.54
0.10	5.62	5.94	6.08	5.66	5.96	6.09	6.26	6.47	6.62
0.15	7.01	7.46	7.71	7.00	7.43	7.68	7.37	7.78	7.96
0.20	8.47	9.04	9.29	8.42	8.99	9.24	8.72	9.19	9.50

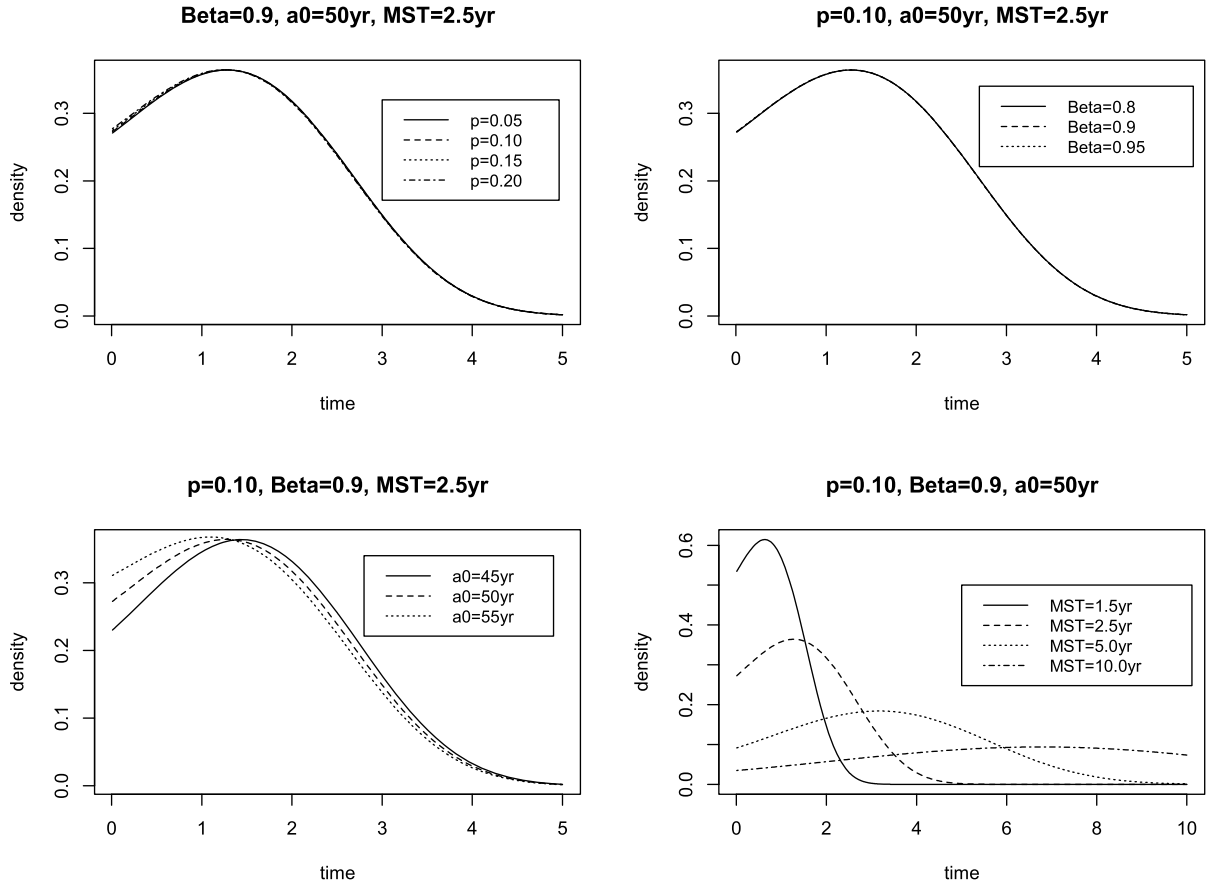

Figure 2. The PDF curves of the lead time under the four factors: fixed three factors and allow one to change.

For comparison purpose, we did some more simulations for the case of $a_0 = 0$. That is, assume that one was just born (i.e. current age is 0), all the other parameters are the same, and we tried two different values of σ^2: 0.02 or 0.03. Using methods in section 2 and the binary search, the found optimal first screening age are summarized in Table 9 in the Appendix. The estimated probability of overdiagnosis and estimated lead time statistics are summarized in Tables 10 and 11 in the Appendix. We omitted the table of the lead time when $\sigma^2 = 0.03$ since the result shows similar pattern as that in Table 11.

4. APPLICATION

The recently finished NLST study was designed to compare two different screening modalities for early detection of lung cancer among heavy smokers: low-dose computed tomography (CT) versus standard chest X-rays [18]. The NLST study enrolled approximately 54,000 male and female heavy smokers (current or former smokers with a smoking history of 30 or more pack-years, and at most 15 years since quitting if former smokers), aged 55 to 74 between August 2002 and April 2004. Participants were randomized into two intervention arms in equal proportions: low-dose CT or X-ray. Participants were offered three annual screening exams, with the first exam performed soon after study entry. Over 52,000 participants joined the NLST and were screened for lung cancer for the first time. In total, 15,537 men and 10,769 women were assigned to the CT arm, and 15,396 men and 10,634 women were assigned to the X-ray arm. If any of the screening results was abnormal, then the screen was considered positive and more diagnostic tests were conducted (i.e., biopsy) as a follow-up. Participants were followed with a median time of 6.5 years.

We now apply the method of scheduling to the NLST low-dose CT data for male and female heavy smokers; and after we found the scheduling time t_0^*, we will use it to estimate the lead time distribution and the probability of overdiagnosis and true-early detection.

From the two cohorts (male and female separately) in the NLST CT data, Liu et al. 2015 estimated the three key parameters: sensitivity $\beta(t)$, PDF of sojourn time $q(x)$, and transition density $w(t)$ [17]; where the sensitivity was modeled as a function of age $\beta(t|b_0, b_1) = \{1 + \exp(-b_0 - b_1(t - m))\}^{-1}$, and $w(t), q(x)$ and $Q(x)$ were the same as in equations (13)-(15). The unknown parameters were

Table 6. Estimated initial screening age t_0^* and its 95% HPD interval using the NLST low-dose CT data

MALE									
	$a_0 = 45$			$a_0 = 50$			$a_0 = 55$		
p	mean	s.e.	95% C.I.	mean	s.e.	95% C.I.	mean	s.e.	95% C.I.
0.05	45.10	0.013	(45.08, 45.12)	50.09	0.011	(50.07, 50.11)	55.09	0.014	(55.07, 55.12)
0.10	45.21	0.029	(45.16, 45.27)	50.19	0.025	(50.15, 50.24)	55.18	0.022	(55.14, 55.22)
0.15	45.35	0.050	(45.26, 45.44)	50.31	0.041	(50.24, 50.39)	55.28	0.036	(55.22, 55.35)
0.20	45.51	0.076	(45.38, 45.65)	50.45	0.061	(50.34, 50.57)	55.40	0.053	(55.31, 55.51)
FEMALE									
	$a_0 = 45$			$a_0 = 50$			$a_0 = 55$		
p	mean	s.e.	95% C.I.	mean	s.e.	95% C.I.	mean	s.e.	95% C.I.
0.05	45.11	0.018	(45.08, 45.14)	50.10	0.015	(50.07, 50.13)	55.09	0.014	(55.07, 55.12)
0.10	45.24	0.039	(45.17, 45.31)	50.22	0.033	(50.16, 50.28)	55.20	0.029	(55.15, 55.26)
0.15	45.39	0.066	(45.27, 45.51)	50.35	0.054	(50.24, 50.45)	55.32	0.048	(55.24, 55.42)
0.20	45.57	0.102	(45.39, 45.76)	50.50	0.081	(50.35, 50.65)	55.46	0.069	(55.34, 55.60)

Table 7. Estimated mean, median, mode and standard deviation of lead time using NLST-CT data

MALE			
p	$a_0 = 45$	$a_0 = 50$	$a_0 = 55$
0.05	0.93, 0.83, 0.59, 0.65	0.91, 0.80, 0.50, 0.65	0.89, 0.78, 0.41, 0.64
0.10	0.93, 0.83, 0.59, 0.65	0.91, 0.80, 0.50, 0.65	0.89, 0.78, 0.41, 0.64
0.15	0.93, 0.83, 0.59, 0.65	0.91, 0.80, 0.50, 0.65	0.89, 0.78, 0.41, 0.64
0.20	0.93, 0.83, 0.59, 0.65	0.91, 0.80, 0.50, 0.65	0.89, 0.78, 0.40, 0.64
FEMALE			
p	$a_0 = 45$	$a_0 = 50$	$a_0 = 55$
0.05	1.02, 0.92, 0.70, 0.69	0.99, 0.89, 0.61, 0.69	0.97, 0.87, 0.51, 0.68
0.10	1.02, 0.92, 0.70, 0.69	0.99, 0.89, 0.60, 0.69	0.97, 0.86, 0.51, 0.68
0.15	1.01, 0.92, 0.69, 0.69	0.99, 0.89, 0.60, 0.69	0.97, 0.86, 0.50, 0.68
0.20	1.01, 0.92, 0.69, 0.69	0.99, 0.89, 0.60, 0.69	0.97, 0.86, 0.50, 0.68

$\theta = (b_0, b_1, \mu, \sigma^2, \lambda, \alpha)$. Using Markov Chain Monte Carlo (MCMC) with Gibbs sampler and a likelihood function, 130,000 samples were generated; after 30,000 burn-in steps and thinning every 200 iterations, a posterior sample of 500 from each chain was obtained. Running two initially over-dispersed chains provided 1000 Bayesian posterior samples (θ_j^*) for each gender. For more details, see Liu et al. 2015. We will use these 1000 posterior samples from each group in our simulation.

In this application, we designed hypothetical cohorts in the simulation: For each gender, there were three hypothetic cohorts according to the current age a_0: $a_0 = 45, 50, 55$. Then we used the 1000 posterior samples $\theta_j^*, j = 1, 2, \ldots, 1000$ from the MCMC of each gender to make Bayesian inference on optimal scheduling time/age.

Given the probability of incidence p, for each θ_j^*, using $P(I_0|I_0 \cup D_0, \theta_j^*) = p$, a scheduling age t_j^* ($j = 1, 2, \ldots, 1000$) can be found. We calculated the mean, the standard error (s.e.) and the 95% highest posterior density (HPD) interval (also called credible interval or C.I.) of the future screening age t_j^* (in years) and summarized the results for male and female heavy smokers using the NLST CT data in Table 6. The results show that the optimal first screening times are very close for the two genders under similar situations, that is, under same current age a_0 and same incidence probability p.

After the optimal first screening time was determined, we can further evaluate the lead time distribution and the probability of overdiagnosis. One lead time distribution can be obtained by using each pair of (θ_j^*, t_j^*), with $j = 1, 2, \ldots, 1000$. The posterior distribution of the lead time is the average:

$$f_L(z|NLST) \approx \frac{1}{1000} \sum_{j=1}^{1000} f_L(z|\theta_j^*).$$

We then calculate the mean, median, mode and standard deviation of the lead time using $f_L(z|NLST)$, the result is presented in Table 7. In general, females heavy smokers seem to have a slightly longer mean lead time than their male counterpart in similar conditions. The estimated lead time density curves under different current age a_0 and different incidence probability p were plotted using the NLST low-dose CT data in Figure 3. It showed that the lead time PDF curve didn't change much with the incidence probability p if the optimal scheduling time t_0^* were used. The density curves did change with the current age a_0: larger a_0 correspondes to a higher spike in the

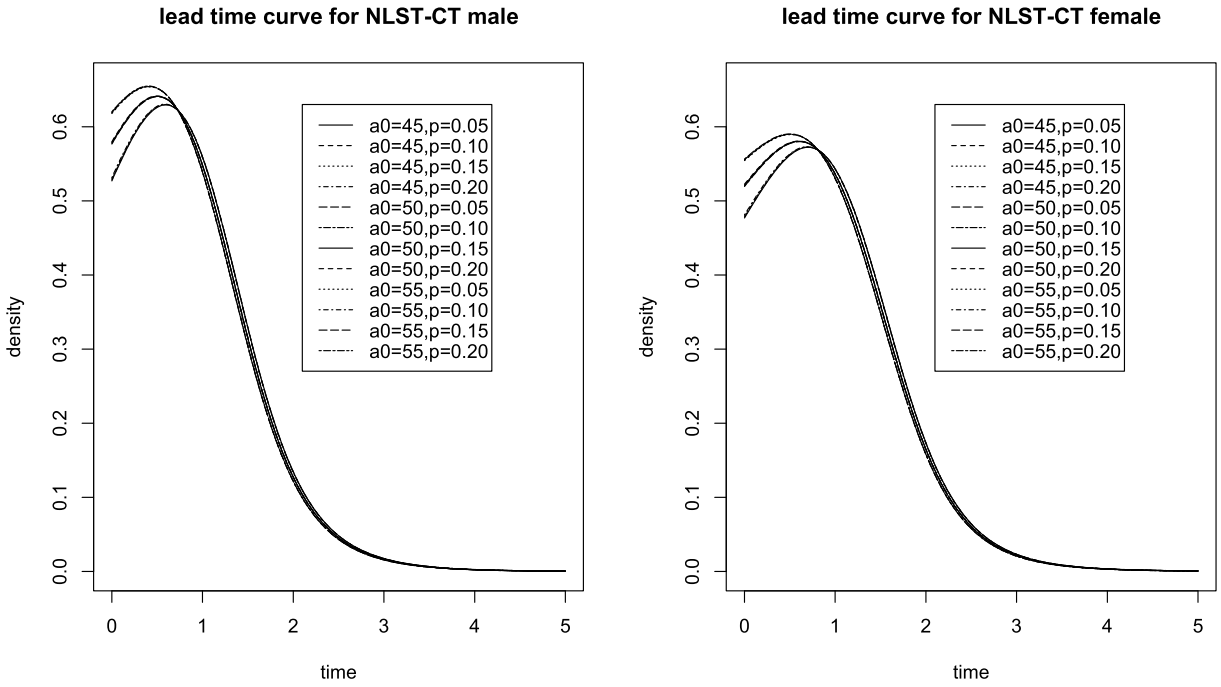

Figure 3. Lead time density curve.

Table 8. Estimated mean, standard error and 95% C.I. for probability of overdiagnosis at the first exam for the NLST-CT data (in percentage)

p	MALE								
	$a_0 = 45$			$a_0 = 50$			$a_0 = 55$		
	mean	s.e.	95% C.I.	mean	s.e.	95% C.I.	mean	s.e.	95% C.I.
0.05	0.304	0.059	(0.214, 0.423)	0.472	0.091	(0.336, 0.657)	0.712	0.135	(0.490, 0.967)
0.10	0.306	0.059	(0.214, 0.426)	0.475	0.092	(0.324, 0.650)	0.715	0.136	(0.492, 0.973)
0.15	0.308	0.061	(0.215, 0.432)	0.479	0.094	(0.326, 0.658)	0.720	0.138	(0.493, 0.980)
0.20	0.313	0.063	(0.217, 0.440)	0.484	0.096	(0.328, 0.668)	0.725	0.140	(0.496, 0.991)
	FEMALE								
	$a_0 = 45$			$a_0 = 50$			$a_0 = 55$		
	mean	s.e.	95% C.I.	mean	s.e.	95% C.I.	mean	s.e.	95% C.I.
0.05	0.212	0.042	(0.139, 0.291)	0.326	0.064	(0.216, 0.448)	0.467	0.089	(0.322, 0.649)
0.10	0.214	0.043	(0.141, 0.296)	0.329	0.065	(0.217, 0.452)	0.470	0.091	(0.323, 0.653)
0.15	0.216	0.044	(0.142, 0.300)	0.331	0.066	(0.218, 0.458)	0.473	0.092	(0.324, 0.659)
0.20	0.219	0.046	(0.147, 0.311)	0.335	0.068	(0.221, 0.467)	0.477	0.094	(0.325, 0.667)

density curve, which translates to a slighly smaller mode value.

Finally, we use each pair $(\theta_j^*, t_j^*), j = 1, 2, \ldots, 1000$, to estimate the probability of overdiagnosis. And we can calculate the posterior mean, standard error and 95% HPD interval of the probability (or percentage) of overdiagnosis. Correspondingly, the probability of true-early-detection is just 1 minus the probability of overdiagnosis. The posterior mean, the standard error (s.e.), and the 95% HPD interval (credible interval or C.I.) are listed in Table 8. The probability of overdiagnosis is very low at the first screening for heavy smokers using the parameters derived from the NLST-CT data ($< 1\%$). This risk of overdiagnosis slightly increases with one's current age for both genders. And it is slightly higher for male heavy smokers than their female counterpart. The probability of overdiagnosis slightly increases when p increases. Since the maximum probability of overdiagnosis was less than 1% for both genders in our simulation, overdiagnosis is not an issue at the first screening exam using low-dose CT for heavy smokers.

5. DISCUSSION

We developed a new method to determine when an asymptomatic person should take the first screening exam, given one's current age. The incidence probability was used to control the risk of clinical incidence before one's first exam, and the optimal screening time was found by limiting this probability to a pre-selected small value. Extensive simulations were done to examine the relationship of the optimal first screening time/age with other factors, such as current age, screening sensitivity, and sojourn time.

In the simulation study, we found that the time interval between one's current age and the first screening time slightly increases with the screening sensitivity if other factors were the same; it also increases as the incidence probability increases. If one were diagnosed with cancer at the first screening exam, the lead time barely changes with the incidence probability and the sensitivity; however, the mean, median and mode of lead time slightly decreases as one's current age increases; and the lead time is positively correlated to the mean sojourn time: longer mean sojourn time means longer mean lead time. Using the calculated first screening age, the probability of overdiagnosis is positively correlated with the mean sojourn time, the incidence probability, and one's current age; and it barely changes with the sensitivity, especially when the MST is less than 2 years.

We estimated the optimal first screening time for male and female heavy smokers using the NLST low dose CT arm data. The result is compatible with the simulation. The probability of overdiagnosis is negligible at the first screening. Based on our research, overdiagnosis is more related to a person's life time. Since the first screening time happens at a comparatively younger age, the possibility of overdiagnosis is very small.

We want to point out that the estimated optimal first screening time is a function of the three key parameters: screening sensitivity, sojourn time in the preclinical state and transition density into the preclinical state, as many other terms are. Therefore, accuracy of the proposed method depends on accurate estimation of the three key parameters. And these three key parameters uniquely determine the process of periodic screening. In summary, this project is the first study to work on the optimal screening time for an asymptomatic person for a certain kind of cancer. More improvement can be done to refine this method. We are working on optimal scheduling problem based on one's screening history and other parameters, and we hope to provide some guidelines regarding the timing of future screening exam.

APPENDIX

Simulation results when current age $a_0 = 0$ are in Tables 9 to 11 in the Appendix.

Table 9. Simulated optimal initial screening age t_0^* (in years) with $\sigma^2 = 0.02$ and 0.03

	$\sigma^2 = 0.02$			$\sigma^2 = 0.03$		
	MST = 1.5 years					
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	21.82	22.41	22.69	17.46	17.97	18.21
0.10	25.90	26.63	26.97	21.02	21.67	21.97
0.15	28.92	29.76	30.15	23.72	24.48	24.84
0.20	31.50	32.43	32.87	26.06	26.92	27.33
	MST = 2.5 years					
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	28.67	29.36	29.68	23.51	24.13	24.43
0.10	33.37	34.19	34.58	27.82	28.59	28.95
0.15	36.74	37.66	38.09	30.99	31.87	32.28
0.20	39.54	40.55	41.02	33.68	34.66	35.12
	MST = 5.0 years					
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	39.62	40.41	40.77	33.78	34.54	34.90
0.10	44.85	45.73	46.15	38.94	39.83	40.25
0.15	48.42	49.37	49.81	42.57	43.56	44.02
0.20	51.28	52.28	52.74	45.55	46.61	47.11
	MST = 10.0 years					
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	52.01	52.83	53.21	46.29	47.15	47.55
0.10	57.35	58.23	58.63	51.98	52.93	53.38
0.15	60.84	61.75	62.17	55.81	56.82	57.30
0.20	63.55	64.49	64.92	58.85	59.92	60.41

Table 10. Probability of overdiagnosis (in percentage) at the initial screening age t_0^* with $\sigma^2 = 0.02$ and 0.03

	$\sigma^2 = 0.02$			$\sigma^2 = 0.03$		
	MST = 1.5 years					
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	0.054	0.055	0.056	0.034	0.037	0.037
0.10	0.062	0.063	0.064	0.048	0.050	0.051
0.15	0.069	0.072	0.073	0.054	0.055	0.056
0.20	0.078	0.082	0.084	0.058	0.060	0.060
	MST = 2.5 years					
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	0.129	0.133	0.135	0.100	0.102	0.102
0.10	0.161	0.168	0.172	0.116	0.119	0.122
0.15	0.195	0.205	0.211	0.134	0.139	0.142
0.20	0.231	0.248	0.258	0.152	0.161	0.166
	MST = 5.0 years					
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	0.592	0.628	0.645	0.381	0.398	0.407
0.10	0.890	0.952	0.986	0.524	0.558	0.575
0.15	1.166	1.247	1.284	0.688	0.745	0.775
0.20	1.413	1.500	1.539	0.872	0.948	0.988
	MST = 10.0 years					
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	4.128	4.310	4.414	2.689	2.848	2.919
0.10	5.619	5.936	6.072	3.804	4.015	4.125
0.15	7.008	7.459	7.711	4.743	5.038	5.201
0.20	8.462	9.043	9.291	5.721	6.142	6.368

Table 11. Estimated mean, median, mode and standard deviation of lead time at t_0^* when $a_0 = 0$ and $\sigma^2 = 0.02$

	MST = 1.5 years		
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	1.20, 1.19, 1.17, 0.58	1.19, 1.18, 1.16, 0.58	1.19, 1.17, 1.15, 0.58
0.10	1.14, 1.12, 1.09, 0.59	1.13, 1.10, 1.08, 0.59	1.12, 1.10, 1.07, 0.59
0.15	1.10, 1.07, 1.03, 0.59	1.08, 1.06, 1.02, 0.59	1.08, 1.05, 1.01, 0.59
0.20	1.06, 1.03, 0.98, 0.59	1.05, 1.02, 0.97, 0.59	1.05, 1.01, 0.96, 0.59
	MST = 2.5 years		
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	2.00, 1.97, 1.94, 0.97	1.98, 1.96, 1.92, 0.97	1.98, 1.95, 1.91, 0.97
0.10	1.89, 1.85, 1.81, 0.98	1.87, 1.83, 1.79, 0.98	1.86, 1.82, 1.78, 0.98
0.15	1.82, 1.77, 1.71, 0.98	1.80, 1.75, 1.69, 0.98	1.79, 1.74, 1.67, 0.98
0.20	1.76, 1.71, 1.63, 0.98	1.74, 1.69, 1.60, 0.98	1.73, 1.68, 1.58, 0.98
	MST = 5.0 years		
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	3.98, 3.92, 3.86, 1.93	3.95, 3.89, 3.82, 1.93	3.93, 3.87, 3.80, 1.93
0.10	3.75, 3.67, 3.58, 1.95	3.71, 3.63, 3.52, 1.95	3.70, 3.61, 3.50, 1.95
0.15	3.60, 3.50, 3.36, 1.96	3.56, 3.45, 3.30, 1.96	3.54, 3.43, 3.27, 1.96
0.20	3.48, 3.36, 3.18, 1.96	3.44, 3.31, 3.11, 1.96	3.42, 3.29, 3.07, 1.96
	MST = 10.0 years		
p	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$
0.05	7.92, 7.78, 7.63, 3.88	7.84, 7.70, 7.54, 3.88	7.81, 7.66, 7.49, 3.89
0.10	7.40, 7.22, 6.97, 3.91	7.31, 7.12, 6.85, 3.92	7.27, 7.07, 6.79, 3.92
0.15	7.05, 6.82, 6.45, 3.92	6.96, 6.71, 6.30, 3.92	6.92, 6.66, 6.23, 3.92
0.20	6.78, 6.50, 5.98, 3.92	6.68, 6.38, 5.80, 3.92	6.64, 6.33, 5.71, 3.92

ACKNOWLEDGEMENT

The author would like to thank the two anonymous referees, the AE, and the Editor for their timely review and helpful comments. This research was partially supported by NIH/NCI 1R15CA242482 (Wu).

Received 5 March 2021

REFERENCES

[1] U.S. Preventive Services Task Force. 2020. Lung Cancer Screening Draft Recommendation Statement. July 07, 2020. https://www.uspreventiveservicestaskforce.org/uspstf/draft-recommendation/lung-cancer-screening-2020. Last accessed on 8/23/2020.

[2] National Cancer Institute Surveillance, Epidemiology, and End Results Program. Cancer Stat Facts: Lung and Bronchus Cancer. http://seer.cancer.gov/statfacts/html/lungb.html. Last accessed on 8/23/2020.

[3] 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.

[4] Berlin, N. I., Buncher, C. R., Fontana, R. S., Frost, J. K. and Melamed, M. R. (1984). The National Cancer Institute cooperative early lung cancer detection program: results of the initial screen (prevalence). *American Review of Respiratory Disease*. **130**(4) 545–549.

[5] Mandel, J.S., Bond, J.H., Church, T.R., et al. (1993). Reducing Mortality from colorectal cancer by screening for fecal occult blood: Minnesota colon cancer control study. *New England Journal of Medicine*. **328**. 1365–1371.

[6] Wang, D., Levitt, B., Riley, T. and Wu, D. (2017). Estimation of sojourn time and transition probability of lung cancer for smokers using the PLCO data. *Journal of Biometrics and Biostatistics*. **8:360**. doi: 10.4172/2155-6180.1000360.

[7] The National Lung Screening Trial Research Team (2011). Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine* **365** 395–409.

[8] Parmigiani, G. (1997). Timing medical examinations via intensity functions. *Biometrika*, **84**, 803–816. MR1625035

[9] Parmigiani, G., Skates, S. and Zelen, M. (2002). Modeling and optimization in early detection programs with a single exam. *Biometrics*, **58**, 30–36. MR1891040

[10] Shieh, Y. and Bohnenkamp, M. (2017). Low-dose CT scan for lung cancer screening: clinical and coding considerations. *Chest* **152**(1) 204–209.

[11] Wu, D., Rosner, G. L. and Broemeling, L. D. (2007). Bayesian inference for the lead time in periodic cancer screening. *Biometrics* **63** 873–880. MR2395806

[12] 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*. **Article 8**. MR2923280

[13] The United States Social Security Administration Actuarial Life Table, https://www.ssa.gov/oact/STATS/table4c6.html, [Accessed: 2016-10-10].

[14] Wu, D., Erwin, D. and Rosner, G. L. (2011). Sojourn time and lead time projection in lung cancer screening. *Lung Cancer*. **72**(3) 322–326.

[15] Jang, H., Kim, S. and Wu, D. (2013). Bayesian lead time estimation for the Johns Hopkins Lung Project data. *Journal of Epidemiology and Global Health* **3**(3) 157–163.

[16] Liu, R., Pérez, A. and Wu, D. (2018). Estimation of lead time via low-dose CT in the National Lung Screening Trial. *Journal of Healthcare Informatics Research*. **2**(4) 353–366.

[17] Liu, R., Levitt, B., Riley, T. and Wu, D. (2015). Bayesian estimation of the three key parameters in CT for the National Lung Screening Trial data. *Journal of Biometrics & Biostatistics* **06**(05) 1.

[18] The National Lung Screening Trial Research Team (2011). The National Lung Screening Trial: Overview and Study Design. *Radiology* **258**(1) 243–253.

[19] 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**(4) 1056–1063. MR2216199

[20] Smith, R. A., Andrews, K. S., Brooks, D., Fedewa, S. A., Manassaram-Baptiste. D., Saslow, D., Brawley, O. W. and Wender, R. C. (2018). Cancer screening in the United States, 2018: a review of current American Cancer Society guidelines and current issues in cancer screening *CA: A Cancer Journal for Clinicians* **68** 297–316.

[21] Wu, D., Kafadar, K., and Rai, S. N. (2018). Inference of long-term screening outcomes for individuals with screening histories. *Statistics and Public Policy* **5**(1) 1–10.

[22] Zelen, M. and Feinleib, M. (1969). On the Theory of Screening for Chronic Diseases. *Biometrika* **56**(3) 601–614. MR0258224

[23] Zelen, M. (1993). Optimal scheduling of examinations for the early detection of disease. *Biometrika*. **80** 279–93. MR1243504

[24] Lee, S.J. and 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.

Dongfeng Wu
Department of Bioinformatics and Biostatistics
University of Louisville
E-mail address: dongfeng.wu@louisville.edu