Statistics and Its Interface Volume 14 (2021) 131–149

Lead time distribution for individuals with a screening history

Ruiqi Liu, Dongfeng Wu*†, and Shesh N. Rai

We derived the distribution of lead time for periodic screening in the future when an individual has a screening history with negative results. It is a mixture of a point mass at zero and a positive sub-PDF. The motivation comes from the reality that for people in older age, they may already have some screening exams for targeted cancer before and still look healthy and are asymptomatic at their current age. How to evaluate their future screening result is a challenge. We explored how the screening history would affect the lead time if one would be diagnosed with cancer in the future. Simulations were carried out on combinations of different initial screening age, current age, sensitivity, mean sojourn time, and screening schedule in the past and in the future. The method developed can be applied to periodic exams for any kind of chronic disease, such as cancer. We applied our new method of evaluating the lead time distribution for male and female heavy smokers using low-dose computed tomography in the National Lung Screening Trial.

1. INTRODUCTION

Cancer is a group of diseases involving abnormal cell growth with potential to invade or spread to other parts of the body. Most types of cancer can be described using staging; stage I means that the cancer is small and has not grown deeply into nearby tissues, while stage IV means the cancer has spread to other parts of the body. Cancer stage at diagnosis helps determine which treatment is available and corresponding survival time. In general, patients with early-stage cancer have better prognosis and higher survival rate than those with late-stage cancer. Specifically, the 5-year survival rate for patients with early-stage lung cancer is approximately 50%, while it is only about 5% for stage IV lung cancer patients [1].

As the primary technique for early detection, the goal of screening is to detect the disease earlier before any symptoms appear; so patients may receive earlier intervention and better treatment. Periodic screening is recommended for almost all kinds of cancers, such as breast, lung, colon, cervical cancer, etc. [2] Several major randomized controlled cancer screening studies have been carried out since the 1960s: the Health Insurance Plan of Greater New York Project [3]; the Mayo Lung Project [4]; the Johns Hopkins Lung Project [5]; the Minnesota Colon Cancer Control Study [6]; the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial [7] and the National Lung Screening Trial (NLST) [8, 9].

Early detection may mean more treatment choices and longer survivals for patients. However, since survival time is measured from the time of diagnosis, it could appear longer for screen-detected cases and screening may not truly contribute to overall survival. Lead time is the time interval between the time of early diagnosis using a screening exam and the time a clinical diagnosis would have been made without a screening. Therefore, to correctly estimate the survival time of screen-detected cases, it is critical to estimate the lead time first, and then subtract it from the overall survival. Hence, the lead time is an important factor when evaluating the effectiveness of a screening program [10].

A number of statistical methods were provided to estimate the mean and variance of lead time [11, 12, 13, 14]. Prorok [15] estimated the local lead time by focusing on the i-th screen-detected cases whose lead time is positive. However, he ignored the interval-incident cases, whose lead time is zero. Wu et al. [10] estimated the lead time for both screen-detected and interval-incident cases, where a person's lifetime is treated as a fixed value. This model was applied to the Mayo Lung Project data to estimate the lead time when human lifetime was assumed to be 80 years. Later, Wu et al. [16] extended the model to make it more practical by treating the lifetime as random, deriving its distribution from the actuarial life table of the US Social Security Administration [17]. The lead time for lung cancer screening using chest X-ray was estimated previously, when the lifetime was either fixed [18] or random [19]. Considering the advantage of low-dose helical computed tomography (LDCT) over traditional chest X-ray, Liu et al. [20] estimated the lead time distribution for both genders using LDCT when the lifetime is random in the National Lung Screening Trials.

All of the above methods were developed based on the assumption that an asymptomatic individual has not taken any screening exams at his/her current age, that is, there is no screening history. While in reality, participants aged 55 and older may already have had at least one (previous) screening exam in the past and look healthy right now. In this paper, we will introduce a lead time distribution model which can incorporate one's screening history and derive the

*Corresponding author.
†ORCID: 0000-0002-6764-0083.

lead time distribution when the lifetime is fixed and when it is random. And we will present our simulation results and applications using the NLST-LDCT data to estimate the lead time for lung cancer screening for subjects with any screening history.

2. METHOD

We will use the well-known disease progressive model, assuming that tumor develops through three states, denoted by $S_0 \to S_p \to S_c$ [21]. The state 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 the person has the disease that can be detected by a screening exam even though s/he is asymptomatic. S_c represents the clinical state, where clinical symptoms have presented.

We will use females' lung cancer as an example in the problem solving, while the result is equally valid for males and other kinds of cancer.

Assume a woman at current age t_{K_1} has gone through K_1 exams before, at her age $t_0 < t_1 < \cdots < t_{K_1-1}$, all test results were negative, and she plans to take K more exams at the future age $t_{K_1} < t_{K_1+1} < t_{K_1+2} < \cdots < t_{K_1+K-1}$. To derive the lead time distribution, we proceed as follows:

- We derive the lead time distribution when $K_1 = K = 1$, i.e. one previous screen and one future screen, the simplest case, with a fixed lifetime $T = t(> t_1)$; then we allow the lifetime T to be random.
- We derive the lead time distribution for any fixed positive integers K_1 and K, with a fixed lifetime $T = t(> t_{K_1+K-1})$; Finally, we allow T to be random; hence the number of future exams K is random as well.

Let $\beta(t)$ be the sensitivity of the exam at age t, i.e., the probability that the screening is positive given that the individual is in the preclinical state S_p at age t, and let $\beta_i = \beta(t_i)$. We let X be the time duration in the disease-free state S_0, with a pdf $w(x)$ (also called transition probability by other researchers); and let Y be the sojourn time (the length of time in the preclinical state S_p), with a pdf $q(y)$; and $Q(y) = \int_y^\infty q(x)dx$ be 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.

Let D be a binary random variable, representing true disease status: with $D = 1$ indicating "having cancer", and $D = 0$ indicating "no cancer" in one's whole lifetime. Let L represent the lead time for an individual. The distribution of the lead time is a mixture of a point mass at zero and a positive probability density, depending on whether the cancer was a clinical incident case, or was detected by screening.

2.1 Lead time distribution when $K_1 = K = 1$

Suppose an asymptomatic woman at current age t_1, had only one screening exam at age t_0 ($< t_1$), and it was negative. We define this event:

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

Let the (fixed) lifetime $T = t$, with $t > t_1$, if she will take another screening at her current age t_1, then the lead time distribution is composed of two parts:

(1)
$$P(L = 0|D = 1, H_1, T = t) = \frac{P(L = 0, D = 1, H_1|T = t)}{P(D = 1, H_1|T = t)},$$

(2) $$f_L(z|D = 1, H_1, T = t) = \frac{f_L(z, D = 1, H_1|T = t)}{P(D = 1, H_1|T = t)}.$$

We need to calculate three terms: $P(D = 1, H_1|T = t)$, $P(L = 0, D = 1, H_1|T = t)$ and $f_L(z, D = 1, H_1|T = t)$.

To calculate $P(D = 1, H_1|T = t)$, the probability of no lung cancer appears before/at age t_1 (with one screening at age t_0), but she would have lung cancer in (t_1, T); it could happen in three ways: i). she enters the preclinical state S_p in $(0, t_0)$, i.e., $X \in (0, t_0)$, but her cancer was not detected at t_0; ii). she enters S_p in (t_0, t_1), i.e., $X \in (t_0, t_1)$; or iii). she enters S_p in (t_1, T), i.e., $X \in (t_1, T)$. In all three cases, her sojourn time is longer enough, that the onset of her clinical state S_c is in (t_1, T). That is, $X + Y \in (t_1, T)$. Hence,

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

Similarly, to calculate $P(L = 0, D = 1, H_1|T = t)$, the probability of clinical incidence in (t_1, T) and with two exams at t_0 and t_1, (i.e., no early detection by screening). It could happen in three ways also, depending on when she enters the S_p. Hence,

$$\begin{aligned} & P(L = 0, D = 1, H_1|T = t) \\ &= (1 - \beta_0)(1 - \beta_1) \int_0^{t_0} w(x)[Q(t_1 - x) - Q(t - x)]dx \\ &+ (1 - \beta_1) \int_{t_0}^{t_1} w(x)[Q(t_1 - x) - Q(t - x)]dx \\ &+ \int_{t_1}^{t} w(x)[1 - Q(t - x)]dx. \end{aligned}$$

Finally, to obtain $f_L(z, D = 1, H_1|T = t)$, the probability density function of lead time when it is positive, it means that her lung cancer was detected at t_1, and if she were

untreated, cancer symptoms would appear at $(t_1 + z)$. Depending on the onset time of her preclinical state, it could happen in two ways: either she enters the preclinical state in $(0, t_0)$, or in (t_0, t_1); in either way, her sojourn time in the preclinical state would be exactly $(t_1 + z - x)$, where x is the onset time/age of her preclinical state. Hence, for $\forall z \in (0, t - t_1)$,

$$\begin{aligned} & f_L(z, D = 1, H_1 | T = t) \\ = \;& \beta_1 \Big[(1 - \beta_0) \int_0^{t_0} w(x) q(t_1 + z - x) dx \\ + \;& \int_{t_0}^{t_1} w(x) q(t_1 + z - x) dx \Big]. \end{aligned}$$

It is easy to check that:

$$\begin{aligned} & P(L = 0, D = 1, H_1 | T = t) \\ + \;& \int_0^{t - t_1} f_L(z, D = 1, H_1 | T = t) dz \\ = \;& P(D = 1, H_1 | T = t). \end{aligned}$$

Or, equivalently,

$$\begin{aligned} & P(L = 0 | D = 1, H_1, T = t) \\ & + \int_0^{t - t_1} f_L(z | D = 1, H_1, T = t) dz = 1. \end{aligned}$$

Hence, the derived mixture distribution of the lead time is a valid distribution. When her lifetime T is random and larger than her current age, i.e. $T > t_1$, the lead time distribution would be:

$$\begin{aligned} & P(L = 0 | D = 1, H_1, T > t_1) \\ (3) \quad = \;& \int_{t_1}^{\infty} P(L = 0 | D = 1, H_1, T = t) f_T(t | T > t_1) \, dt, \\ & f_L(z | D = 1, H_1, T > t_1) \\ (4) \quad = \;& \int_{t_1 + z}^{\infty} f_L(z | D = 1, H_1, T = t) f_T(t | T > t_1) \, dt, \\ & \qquad z \in (0, \infty). \end{aligned}$$

Where $f_T(t | T > t_1) = f_T(t) / P(T > t_1) = f_T(t) / [1 - F_T(t_1)]$ for $t > t_1$, is the conditional PDF of the lifetime. It is easy to verify that

$$\begin{aligned} & P(L = 0 | D = 1, H_1, T > t_1) \\ & + \int_0^{\infty} f_L(z | D = 1, H_1, T > t_1) dz \equiv 1. \end{aligned}$$

Therefore the mixture distribution of the lead time is valid.

2.2 Lead time distribution for any K_1 and K

Now we consider the general case for any positive integer K_1 and K. Suppose a woman has received K_1 screening exams, and she will continue with K screenings in the future. We define this event:

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

We first derive the distribution of lead time when human lifetime T is fixed, then we allow T to be random.

When the lifetime $T = t (> t_{K_1 - 1})$ is fixed, similar to Equations (1) and (2), the distribution of lead time is:

$$\begin{aligned} (5) \qquad & P(L = 0 | D = 1, H_{K_1}, T = t) \\ = \;& \frac{P(L = 0, D = 1, H_{K_1} | T = t)}{P(D = 1, H_{K_1} | T = t)}, \\ (6) \qquad & f_L(z | D = 1, H_{K_1}, T = t) \\ = \;& \frac{f_L(z, D = 1, H_{K_1} | T = t)}{P(D = 1, H_{K_1} | T = t)}. \end{aligned}$$

We define $t_{-1} = 0$. To obtain $P(D = 1, H_{K_1} | T = t)$, the probability of no cancer detected before t_{K_1} (with a sequence of exams at age $t_0 < t_1 < \cdots < t_{K_1 - 1}$), but will have cancer in (t_{K_1}, t), this could happen in $(K_1 + 2)$ different ways: i). she enters the preclinical state S_p in $(t_{i-1}, t_i), i = 0, 1, \ldots, K_1 - 1$, and her cancer was not detected by the exams at and after t_i; ii). she enters S_p in $(t_{K_1 - 1}, t_{K_1})$; and iii). she enters S_p in (t_{K_1}, t). And in the above $(K_1 + 2)$ cases, her sojourn time is longer than $t_{K_1} - x$, but shorter than $t - x$, where $x \in (t_{i-1}, t_i)$ is the onset of S_p; Since these are mutually exclusive events, we simply add these probabilities:

$$\begin{aligned} & P(D = 1, H_{K_1} | T = t) \\ (7) \qquad = \;& \sum_{i=0}^{K_1 - 1} (1 - \beta_i) \cdots (1 - \beta_{K_1 - 1}) \\ & \times \int_{t_{i-1}}^{t_i} w(x) [Q(t_{K_1} - x) - Q(t - x)] \, dx \\ & + \int_{t_{K_1 - 1}}^{t_{K_1}} w(x) [Q(t_{K_1} - x) - Q(t - x)] \, dx \\ & + \int_{t_{K_1}}^{T} w(x) [1 - Q(t - x)] \, dx. \end{aligned}$$

To calculate $P(L = 0, D = 1, H_{K_1} | T = t)$, we define $t_{K_1 + K} = t$ (to simplify notation, but keep in mind that it is not a screening time), and we let $I_{K_1 + K, j}$ represent the probability of incidence in (t_{j-1}, t_j), $j = K_1 + 1, K_1 + 2, \ldots, K_1 + K$, then

$$\begin{aligned} & P(L = 0, D = 1, H_{K_1} | T = t) \\ (8) \qquad & = I_{K_1 + K, K_1 + 1} + I_{K_1 + K, K_1 + 2} + \cdots + I_{K_1 + K, K_1 + K}, \end{aligned}$$

Table 1. Values of input parameters in our simulation study

	Parameter	Settings	Value
Sensitivity	N/A	Sensitivity is a fixed value	0.7 or 0.9
Transition Probability	μ	The mode of the log-normal	4.4
	σ^2	distribution is set to be 70	0.16
Sojourn Time	λ	MST = 2	0.1963
		MST = 5	0.0314
		MST = 10	0.0079
	α	α is a fixed value	2.0

where

$$
\begin{aligned}
I_{K_1+K,\,j} = & \sum_{i=0}^{j-1}(1-\beta_i)\cdots(1-\beta_{j-1}) \\
& \times \int_{t_{i-1}}^{t_i} w(x)[Q(t_{j-1}-x)-Q(t_j-x)]\,dx \\
& + \int_{t_{j-1}}^{t_j} w(x)[1-Q(t_j-x)]\,dx, \\
& \text{for } j = K_1+1,\cdots,K_1+K.
\end{aligned} \tag{9}
$$

Finally, if $t-t_j < z \le t-t_{j-1}$, for $j = K_1+1,\ldots,K_1+K$, the PDF of lead time at z is:

$$
\begin{aligned}
& f_L(z, D=1, H_{K_1}|T=t) \\
& = \sum_{i=K_1}^{j-1} \beta_i \left\{ \sum_{r=0}^{i-1}(1-\beta_r)\cdots(1-\beta_{i-1}) \right. \\
& \times \int_{t_{r-1}}^{t_r} w(x)q(t_i+z-x)\,dx \\
& \left. + \int_{t_{i-1}}^{t_i} w(x)q(t_i+z-x)\,dx \right\}.
\end{aligned} \tag{10}
$$

We can prove that this mixed probability distribution is valid since

$$
\begin{aligned}
& P(L=0|D=1,H_{K_1},T=t) \\
& + \int_0^{T-t_{K_1}} f_L(z|D=1,H_{K_1},T=t)\,dz \equiv 1.
\end{aligned} \tag{11}
$$

When the lifetime T is random, the lead time distribution (with T greater than the current age t_{K_1}) can be obtained by

$$
\begin{aligned}
& P(L=0|D=1,H_{K_1},T>t_{K_1}) \\
& = \int_{t_{K_1}}^{\infty} P(L=0|D=1,H_{K_1},T=t) f_T(t|T>t_{K_1})\,dt,
\end{aligned} \tag{12}
$$

$$
\begin{aligned}
& f_L(z|D=1,H_{K_1},T>t_{K_1}) \\
& = \int_{t_{K_1}+z}^{\infty} f_L(z|D=1,H_{K_1},T=t) f_T(t|T>t_{K_1})\,dt, \\
& z \in (0,\infty),
\end{aligned} \tag{13}
$$

where

$$
f_T(t|T>t_{K_1}) = \begin{cases} \frac{f_T(t)}{P(T>t_{K_1})} = \frac{f_T(t)}{1-F_T(t_{K_1})}, & \text{if } t > t_{K_1}, \\ 0, & \text{otherwise.} \end{cases}
$$

And the lifetime distribution density $f_T(t|T>t_{K_1})$ can be obtained by using the US Social Security Administrations actuarial lifetable [16]. Again we can prove that

$$
\begin{aligned}
& P(L=0|D=1,H_{K_1},T>t_{K_1}) \\
& + \int_0^{\infty} f_L(z|D=1,H_{K_1},T>t_{K_1})dz \equiv 1.
\end{aligned}
$$

For a person at her current age t_{K_1}, if she plans to follow a future screening schedule, such as $t_{K_1} < t_{K_1+1} < \ldots$, then the number of screenings in the future $K=n$ if $t_{K_1+n-1} < T \le t_{K_1+n}$, therefore the future screening number $K = K(T)$ is random if the lifetime T is random. If the future screening exam is equally-spaced with a time interval Δ, then $K = K(T) = \lceil (T-t_{K_1})/\Delta \rceil$.

3. SIMULATION STUDY

Screening for breast, lung, colon and cervical cancers are recommended by the United States Preventive Services Task Force [2]. Different screening modalities/techniques are used to detect different cancer; and the screening sensitivity for these methods are very different. In addition, the speed of cancer grows and spreads also vary dramatically, which means, the sojourn times in the preclinical state are different. From the results in Section 2, the distribution of lead time is a function of the sensitivity, the sojourn time distribution and the transition probability density. We will explore the characteristics of the newly developed lead time distribution for different cancer and different screening schedules; and our simulations were done, using combinations of the following settings:

1. Three different initial screening ages: $t_0 = 56$, 60 and 64 years.
2. Two different screening sensitivities: $\beta = 0.7$ and 0.9.
3. Three different mean sojourn time: MST = 2, 5 and 10 years.

Table 2. A projection of the lead time distribution for individuals with screening history by current age and screening intervals, with MST = 2

(Δ_1, Δ_2)	$\beta = 0.7$				$\beta = 0.9$			
(years)	P_0	EL (s.d.)	Median	Mode	P_0	EL (s.d.)	Median	Mode
initial screening age $t_0 = 56$, current age $t_{K_1} = 60$								
(1,1)	20.10	1.14 (1.05)	1.25	0.85	10.32	1.33 (1.04)	1.35	0.95
(2,1)	20.02	1.13 (1.04)	1.25	0.75	10.05	1.32 (1.04)	1.35	0.95
(1,2)	41.26	0.76 (0.97)	1.15	0.35	28.00	0.95 (1.01)	1.15	0.45
(2,2)	40.69	0.76 (0.96)	1.15	0.35	27.08	0.96 (1.00)	1.15	0.45
initial screening age $t_0 = 56$, current age $t_{K_1} = 64$								
(1,1)	20.34	1.12 (1.04)	1.25	0.75	10.45	1.31 (1.03)	1.35	0.95
(2,1)	20.23	1.11 (1.03)	1.25	0.65	10.10	1.30 (1.03)	1.35	0.95
(1,2)	41.34	0.75 (0.96)	1.15	0.15	28.09	0.94 (1.00)	1.15	0.45
(2,2)	40.63	0.75 (0.96)	1.05	0.15	26.96	0.94 (0.99)	1.15	0.45
initial screening age $t_0 = 56$, current age $t_{K_1} = 68$								
(1,1)	20.69	1.09 (1.03)	1.25	0.65	10.63	1.29 (1.03)	1.25	0.85
(2,1)	20.53	1.08 (1.02)	1.25	0.65	10.19	1.27 (1.02)	1.25	0.85
(1,2)	41.46	0.74 (0.95)	1.05	0.15	28.21	0.92 (0.99)	1.15	0.45
(2,2)	40.58	0.74 (0.94)	1.05	0.15	26.81	0.93 (0.98)	1.05	0.15
initial screening age $t_0 = 56$, current age $t_{K_1} = 72$								
(1,1)	21.15	1.06 (1.01)	1.15	0.65	10.88	1.25 (1.01)	1.25	0.85
(2,1)	20.93	1.05 (1.01)	1.15	0.65	10.32	1.24 (1.01)	1.25	0.65
(1,2)	41.58	0.72 (0.93)	1.05	0.15	28.36	0.90 (0.97)	1.05	0.15
(2,2)	40.50	0.72 (0.93)	1.05	0.15	26.61	0.91 (0.97)	1.05	0.15
initial screening age $t_0 = 60$, current age $t_{K_1} = 64$								
(1,1)	20.34	1.12 (1.04)	1.25	0.75	10.45	1.31 (1.03)	1.35	0.95
(2,1)	20.23	1.11 (1.03)	1.25	0.65	10.10	1.30 (1.03)	1.35	0.95
(1,2)	41.34	0.75 (0.96)	1.15	0.15	28.09	0.94 (1.00)	1.15	0.45
(2,2)	40.63	0.75 (0.96)	1.05	0.15	26.96	0.94 (0.99)	1.15	0.45
initial screening age $t_0 = 60$, current age $t_{K_1} = 68$								
(1,1)	20.69	1.09 (1.03)	1.25	0.65	10.63	1.29 (1.02)	1.25	0.85
(2,1)	20.53	1.08 (1.02)	1.25	0.65	10.19	1.27 (1.02)	1.25	0.85
(1,2)	41.46	0.74 (0.95)	1.05	0.15	28.21	0.92 (0.99)	1.15	0.45
(2,2)	40.58	0.74 (0.94)	1.05	0.15	26.81	0.93 (0.98)	1.05	0.15
initial screening age $t_0 = 60$, current age $t_{K_1} = 72$								
(1,1)	21.15	1.06 (1.01)	1.15	0.65	10.88	1.25 (1.01)	1.25	0.85
(2,1)	20.93	1.05 (1.01)	1.15	0.65	10.32	1.24 (1.01)	1.25	0.65
(1,2)	41.58	0.72 (0.93)	1.05	0.15	28.36	0.90 (0.97)	1.05	0.15
(2,2)	40.50	0.72 (0.93)	1.05	0.15	26.61	0.91 (0.97)	1.05	0.15
initial screening age $t_0 = 60$, current age $t_{K_1} = 76$								
(1,1)	21.77	1.01 (0.99)	1.15	0.45	11.22	1.20 (1.00)	1.25	0.65
(2,1)	21.46	1.00 (0.98)	1.15	0.15	10.49	1.18 (0.99)	1.15	0.65
(1,2)	41.77	0.69 (0.91)	1.05	0.15	28.55	0.87 (0.95)	1.05	0.15
(2,2)	40.43	0.70 (0.91)	0.95	0.05	26.38	0.88 (0.94)	1.05	0.15
initial screening age $t_0 = 64$, current age $t_{K_1} = 68$								
(1,1)	20.69	1.09 (1.03)	1.25	0.65	10.63	1.29 (1.02)	1.25	0.85
(2,1)	20.53	1.08 (1.02)	1.25	0.65	10.19	1.27 (1.02)	1.25	0.85
(1,2)	41.46	0.74 (0.95)	1.05	0.15	28.21	0.92 (0.99)	1.15	0.45
(2,2)	40.58	0.74 (0.94)	1.05	0.15	26.81	0.93 (0.98)	1.05	0.15
initial screening age $t_0 = 64$, current age $t_{K_1} = 72$								
(1,1)	21.15	1.06 (1.01)	1.15	0.65	10.88	1.25 (1.01)	1.25	0.85
(2,1)	20.93	1.05 (1.01)	1.15	0.65	10.32	1.24 (1.01)	1.25	0.65
(1,2)	41.58	0.72 (0.93)	1.05	0.15	28.36	0.90 (0.97)	1.05	0.15
(2,2)	40.50	0.72 (0.93)	1.05	0.15	26.61	0.91 (0.97)	1.05	0.15
initial screening age $t_0 = 64$, current age $t_{K_1} = 76$								
(1,1)	21.77	1.01 (0.99)	1.15	0.45	11.22	1.20 (1.00)	1.25	0.65
(2,1)	21.46	1.00 (0.98)	1.15	0.15	10.49	1.18 (0.99)	1.15	0.65
(1,2)	41.77	0.69 (0.91)	1.05	0.15	28.55	0.87 (0.95)	1.05	0.15
(2,2)	40.43	0.70 (0.91)	0.95	0.05	26.38	0.88 (0.94)	1.05	0.15
initial screening age $t_0 = 64$, current age $t_{K_1} = 80$								
(1,1)	22.60	0.95 (0.96)	1.05	0.15	11.68	1.13 (0.97)	1.15	0.45
(2,1)	22.15	0.94 (0.95)	1.05	0.15	10.71	1.12 (0.96)	1.15	0.15
(1,2)	41.88	0.66 (0.88)	0.95	0.05	28.68	0.82 (0.92)	1.05	0.15
(2,2)	40.21	0.66 (0.88)	0.95	0.05	25.97	0.84 (0.92)	0.95	0.05

Table 3. A projection of the lead time distribution for individuals with screening history by current age and screening intervals, with MST = 5

(Δ_1, Δ_2)	$\beta = 0.7$				$\beta = 0.9$			
(years)	P_0	EL (s.d.)	Median	Mode	P_0	EL (s.d.)	Median	Mode
initial screening age $t_0 = 56$, current age $t_{K_1} = 60$								
(1,1)	6.47	3.44 (2.50)	3.35	2.45	2.77	3.72 (2.48)	3.45	2.75
(2,1)	6.44	3.42 (2.50)	3.35	2.45	2.55	3.71 (2.47)	3.45	2.75
(1,2)	17.21	2.80 (2.49)	2.95	1.85	8.81	3.21 (2.48)	3.15	2.25
(2,2)	16.73	2.80 (2.49)	2.95	1.85	8.09	3.23 (2.47)	3.15	2.25
initial screening age $t_0 = 56$, current age $t_{K_1} = 64$								
(1,1)	6.95	3.31 (2.46)	3.25	2.15	3.00	3.59 (2.44)	3.35	2.45
(2,1)	6.87	3.29 (2.46)	3.15	2.15	2.72	3.58 (2.44)	3.35	2.45
(1,2)	17.93	2.69 (2.44)	2.85	1.45	9.31	3.09 (2.44)	3.05	1.95
(2,2)	17.28	2.70 (2.44)	2.85	1.45	8.38	3.12 (2.43)	3.05	1.95
initial screening age $t_0 = 56$, current age $t_{K_1} = 68$								
(1,1)	7.64	3.14 (2.41)	3.05	1.95	3.34	3.41 (2.39)	3.15	1.95
(2,1)	7.51	3.12 (2.41)	3.05	1.95	2.95	3.41 (2.39)	3.15	1.95
(1,2)	18.91	2.54 (2.38)	2.75	1.45	10.01	2.93 (2.38)	2.85	1.95
(2,2)	18.05	2.56 (2.38)	2.75	1.15	8.79	2.97 (2.38)	2.85	1.95
initial screening age $t_0 = 56$, current age $t_{K_1} = 72$								
(1,1)	8.58	2.93 (2.33)	2.85	1.65	3.80	3.18 (2.32)	2.95	1.95
(2,1)	8.37	2.91 (2.33)	2.85	1.65	3.28	3.19 (2.32)	2.95	1.95
(1,2)	20.19	2.36 (2.29)	2.55	0.15	10.93	2.73 (2.30)	2.65	1.45
(2,2)	19.06	2.39 (2.29)	2.55	0.15	9.32	2.78 (2.30)	2.75	0.15
initial screening age $t_0 = 60$, current age $t_{K_1} = 64$								
(1,1)	6.96	3.31 (2.46)	3.25	2.15	3.00	3.59 (2.44)	3.35	2.45
(2,1)	6.90	3.28 (2.46)	3.15	2.15	2.72	3.58 (2.44)	3.35	2.45
(1,2)	17.93	2.69 (2.44)	2.85	1.45	9.31	3.09 (2.44)	3.05	1.95
(2,2)	17.29	2.69 (2.44)	2.85	1.45	8.38	3.12 (2.43)	3.05	1.95
initial screening age $t_0 = 60$, current age $t_{K_1} = 68$								
(1,1)	7.64	3.14 (2.41)	3.05	1.95	3.34	3.41 (2.39)	3.15	1.95
(2,1)	7.51	3.12 (2.41)	3.05	1.95	2.95	3.41 (2.39)	3.15	1.95
(1,2)	18.91	2.54 (2.38)	2.75	1.45	10.01	2.93 (2.38)	2.85	1.95
(2,2)	18.05	2.56 (2.38)	2.75	1.15	8.79	2.97 (2.38)	2.85	1.95
initial screening age $t_0 = 60$, current age $t_{K_1} = 72$								
(1,1)	8.58	2.93 (2.33)	2.85	1.65	3.80	3.18 (2.32)	2.95	1.95
(2,1)	8.37	2.91 (2.33)	2.85	1.65	3.28	3.19 (2.32)	2.95	1.95
(1,2)	20.19	2.36 (2.29)	2.55	0.15	10.93	2.73 (2.30)	2.65	1.45
(2,2)	19.06	2.39 (2.29)	2.55	0.15	9.32	2.78 (2.30)	2.75	0.15
initial screening age $t_0 = 60$, current age $t_{K_1} = 76$								
(1,1)	9.84	2.65 (2.23)	2.55	0.15	4.42	2.90 (2.22)	2.65	1.45
(2,1)	9.50	2.64 (2.23)	2.55	0.15	3.71	2.92 (2.23)	2.65	0.15
(1,2)	21.85	2.14 (2.17)	2.35	0.15	12.13	2.47 (2.19)	2.45	0.15
(2,2)	20.35	2.17 (2.18)	2.35	0.05	10.01	2.55 (2.20)	2.45	0.15
initial screening age $t_0 = 64$, current age $t_{K_1} = 68$								
(1,1)	7.65	3.14 (2.41)	3.05	1.95	3.34	3.41 (2.39)	3.15	1.95
(2,1)	7.54	3.11 (2.41)	3.05	1.95	2.95	3.41 (2.39)	3.15	1.95
(1,2)	18.91	2.54 (2.38)	2.75	1.45	10.01	2.93 (2.38)	2.85	1.95
(2,2)	18.06	2.55 (2.38)	2.75	1.15	8.78	2.97 (2.38)	2.85	1.95
initial screening age $t_0 = 64$, current age $t_{K_1} = 72$								
(1,1)	8.58	2.93 (2.33)	2.85	1.65	3.80	3.18 (2.32)	2.95	1.95
(2,1)	8.37	2.91 (2.33)	2.85	1.65	3.28	3.19 (2.32)	2.95	1.95
(1,2)	20.19	2.36 (2.29)	2.55	0.15	10.93	2.73 (2.30)	2.65	1.45
(2,2)	19.06	2.39 (2.29)	2.55	0.15	9.32	2.78 (2.30)	2.75	0.15
initial screening age $t_0 = 64$, current age $t_{K_1} = 76$								
(1,1)	9.84	2.65 (2.23)	2.55	0.15	4.42	2.90 (2.22)	2.65	1.45
(2,1)	9.50	2.64 (2.23)	2.55	0.15	3.71	2.92 (2.23)	2.65	0.15
(1,2)	21.85	2.14 (2.17)	2.35	0.15	12.13	2.47 (2.19)	2.45	0.15
(2,2)	20.35	2.17 (2.18)	2.35	0.05	10.01	2.55 (2.20)	2.45	0.15
initial screening age $t_0 = 64$, current age $t_{K_1} = 80$								
(1,1)	11.58	2.32 (2.08)	2.25	0.15	5.29	2.55 (2.08)	2.35	0.15
(2,1)	11.06	2.32 (2.09)	2.25	0.05	4.31	2.58 (2.09)	2.35	0.05
(1,2)	23.94	1.86 (2.01)	2.05	0.05	13.67	2.17 (2.04)	2.15	0.15
(2,2)	21.98	1.91 (2.02)	2.05	0.05	10.87	2.26 (2.05)	2.15	0.05

Table 4. A projection of the lead time distribution for individuals with screening history by current age and screening intervals, with MST = 10

(Δ_1, Δ_2)	$\beta = 0.7$				$\beta = 0.9$			
(years)	P_0	EL (s.d.)	Median	Mode	P_0	EL (s.d.)	Median	Mode
initial screening age $t_0 = 56$, current age $t_{K_1} = 60$								
(1,1)	3.43	6.32 (4.41)	5.95	3.95	1.47	6.60 (4.39)	6.05	4.25
(2,1)	3.38	6.32 (4.44)	5.95	3.95	1.24	6.66 (4.41)	6.15	4.25
(1,2)	9.19	5.58 (4.42)	5.45	2.85	4.48	6.07 (4.39)	5.75	3.85
(2,2)	8.64	5.63 (4.44)	5.45	2.85	3.78	6.17 (4.41)	5.75	3.85
initial screening age $t_0 = 56$, current age $t_{K_1} = 64$								
(1,1)	4.00	5.87 (4.23)	5.45	3.65	1.73	6.13 (4.21)	5.65	3.75
(2,1)	3.82	5.92 (4.26)	5.55	3.65	1.44	6.22 (4.24)	5.65	3.75
(1,2)	10.31	5.15 (4.23)	5.05	2.65	5.14	5.61 (4.21)	5.25	2.65
(2,2)	9.52	5.25 (4.26)	5.15	2.65	4.24	5.75 (4.23)	5.35	3.25
initial screening age $t_0 = 56$, current age $t_{K_1} = 68$								
(1,1)	4.82	5.31 (4.01)	4.95	2.95	2.11	5.57 (3.99)	5.05	2.95
(2,1)	4.55	5.38 (4.04)	5.05	2.95	1.72	5.67 (4.02)	5.15	3.15
(1,2)	11.79	4.64 (3.98)	4.55	0.15	6.04	5.07 (3.97)	4.75	1.95
(2,2)	10.74	4.76 (4.03)	4.65	0.15	4.84	5.23 (4.01)	4.85	2.65
initial screening age $t_0 = 56$, current age $t_{K_1} = 72$								
(1,1)	5.92	4.68 (3.72)	4.35	0.15	2.63	4.92 (3.70)	4.45	2.15
(2,1)	5.54	4.76 (3.77)	4.45	0.15	2.09	5.04 (3.75)	4.55	0.15
(1,2)	13.70	4.05 (3.68)	4.05	0.15	7.23	4.45 (3.68)	4.15	0.15
(2,2)	12.31	4.20 (3.73)	4.15	0.15	5.63	4.63 (3.73)	4.25	0.15
initial screening age $t_0 = 60$, current age $t_{K_1} = 64$								
(1,1)	4.00	5.86 (4.23)	5.45	3.65	1.73	6.13 (4.21)	5.65	3.75
(2,1)	3.93	5.87 (4.27)	5.45	3.65	1.44	6.21 (4.24)	5.65	3.75
(1,2)	10.27	5.15 (4.23)	5.05	2.65	5.14	5.61 (4.21)	5.25	2.65
(2,2)	9.55	5.21 (4.26)	5.05	2.65	4.21	5.74 (4.23)	5.35	3.25
initial screening age $t_0 = 60$, current age $t_{K_1} = 68$								
(1,1)	4.82	5.31 (4.01)	4.95	2.95	2.11	5.57 (3.99)	5.05	2.95
(2,1)	4.56	5.38 (4.05)	4.95	2.95	1.72	5.67 (4.02)	5.15	3.15
(1,2)	11.79	4.64 (3.98)	4.55	0.15	6.04	5.07 (3.97)	4.75	1.95
(2,2)	10.74	4.76 (4.03)	4.65	0.15	4.84	5.23 (4.01)	4.85	2.65
initial screening age $t_0 = 60$, current age $t_{K_1} = 72$								
(1,1)	5.92	4.68 (3.72)	4.35	0.15	2.63	4.92 (3.70)	4.45	2.15
(2,1)	5.54	4.76 (3.77)	4.45	0.15	2.09	5.04 (3.75)	4.55	0.15
(1,2)	13.70	4.05 (3.68)	4.05	0.15	7.23	4.45 (3.68)	4.15	0.15
(2,2)	12.31	4.20 (3.73)	4.15	0.15	5.63	4.63 (3.73)	4.25	0.15
initial screening age $t_0 = 60$, current age $t_{K_1} = 76$								
(1,1)	7.38	3.99 (3.37)	3.65	0.15	3.33	4.21 (3.35)	3.75	0.15
(2,1)	6.86	4.08 (3.43)	3.75	0.05	2.59	4.34 (3.41)	3.85	0.15
(1,2)	16.11	3.42 (3.31)	3.45	0.15	8.76	3.78 (3.32)	3.55	0.15
(2,2)	14.29	3.57 (3.38)	3.55	0.05	6.62	3.98 (3.38)	3.65	0.15
initial screening age $t_0 = 64$, current age $t_{K_1} = 68$								
(1,1)	4.82	5.31 (4.01)	4.95	2.95	2.11	5.57 (3.99)	5.05	2.95
(2,1)	4.70	5.32 (4.05)	4.95	0.15	1.71	5.66 (4.03)	5.15	3.15
(1,2)	11.73	4.64 (3.98)	4.55	0.15	6.04	5.07 (3.97)	4.75	1.95
(2,2)	10.76	4.72 (4.02)	4.55	0.15	4.80	5.22 (4.01)	4.85	0.15
initial screening age $t_0 = 64$, current age $t_{K_1} = 72$								
(1,1)	5.92	4.68 (3.72)	4.35	0.15	2.63	4.92 (3.70)	4.45	2.15
(2,1)	5.55	4.76 (3.77)	4.45	0.15	2.09	5.04 (3.75)	4.55	0.15
(1,2)	13.70	4.05 (3.68)	4.05	0.15	7.23	4.45 (3.68)	4.15	0.15
(2,2)	12.31	4.19 (3.73)	4.15	0.15	5.63	4.63 (3.73)	4.25	0.15
initial screening age $t_0 = 64$, current age $t_{K_1} = 76$								
(1,1)	7.38	3.99 (3.37)	3.65	0.15	3.33	4.21 (3.35)	3.75	0.15
(2,1)	6.86	4.08 (3.43)	3.75	0.05	2.59	4.34 (3.41)	3.85	0.15
(1,2)	16.11	3.42 (3.31)	3.45	0.15	8.76	3.78 (3.32)	3.55	0.15
(2,2)	14.30	3.57 (3.38)	3.55	0.05	6.62	3.98 (3.38)	3.65	0.15
initial screening age $t_0 = 64$, current age $t_{K_1} = 80$								
(1,1)	9.38	3.26 (2.97)	3.05	0.05	4.29	3.47 (2.96)	3.05	0.15
(2,1)	8.67	3.35 (3.03)	3.05	0.05	3.28	3.60 (3.02)	3.15	0.05
(1,2)	19.09	2.76 (2.89)	2.85	0.05	10.72	3.08 (2.91)	2.85	0.15
(2,2)	16.76	2.92 (2.97)	2.85	0.05	7.86	3.28 (2.99)	2.95	0.05

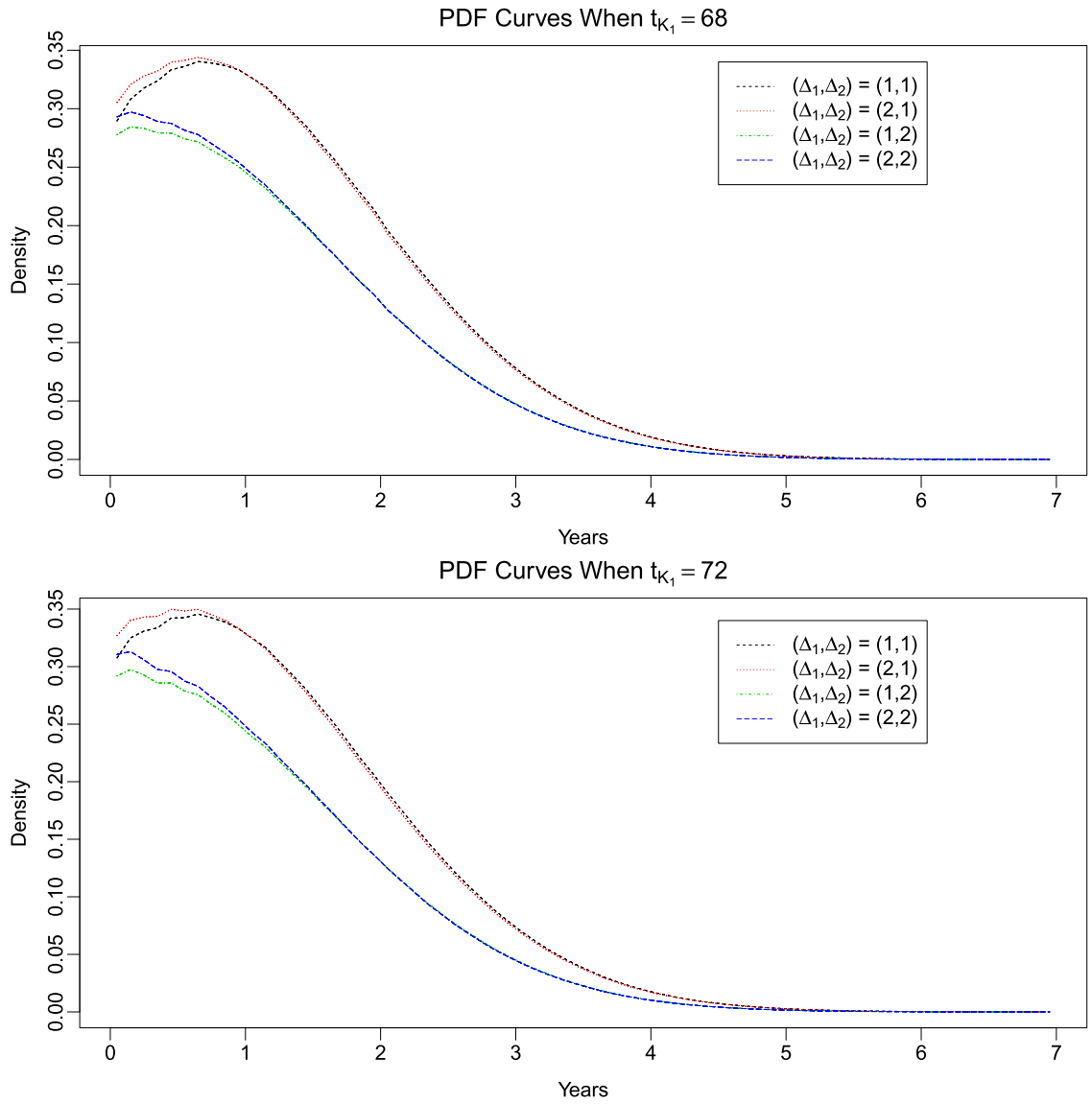

Figure 1. The PDF curves of the lead time for $t_{K_1} = 68$ and $t_{K_1} = 72$ with different t_0: 12 curves representing different screening schedules and different initial screening age t_0 are plotted for $t_{K_1} = 68$ (upper panel) and $t_{K_1} = 72$ (bottom panel), respectively. Curves with the same t_0 overlap each other, and only one curve for each t_0 shows. $\beta = 0.7$, MST = 2.

For a given initial screening age t_0, we examined the lead time distribution for four different current ages t_{K_1} with four-year intervals. That is, when $t_0 = 56$, we conducted simulations by setting $t_{K_1} = 60$, 64, 68 and 72; and when $t_0 = 60$, we let $t_{K_1} = 64$, 68, 72 and 76; and when $t_0 = 64$, we let $t_{K_1} = 68$, 72, 76 and 80. For each combination of initial screening ages, sensitivities, mean sojourn times and current ages, we considered four screening intervals in the past and in the future with $(\Delta_1, \Delta_2) = (1, 1)$, $(2, 1)$, $(1, 2)$ and $(2, 2)$. For example, $(\Delta_1, \Delta_2) = (1, 2)$ means that an individual received annual exams in the past and will take screening exams biennially in the future. We used the actuarial lifetime table for males in this simulation study. Since female results were similar, we omitted it here.

We use the same parametric model of the transition density and the sojourn time as in Liu et al. 2018 [20]:

$$w(t|\mu, \sigma^2) = \frac{0.3}{\sqrt{2\pi}\sigma t} \exp\{-(\log t - \mu)^2/(2\sigma^2)\}, \tag{14}$$

$$q(x|\lambda, \alpha) = \alpha\lambda x^{\alpha-1} \exp(-\lambda x^{\alpha}), \quad \lambda > 0, \alpha > 0, \tag{15}$$

$$Q(x|\lambda, \alpha) = \exp(-\lambda x^{\alpha}), \quad \lambda > 0, \alpha > 0. \tag{16}$$

The input parameters of μ and σ^2 were decided based on the mode of the log-normal distribution, here we let the mode be 70, as most lung cancer cases are diagnosed around

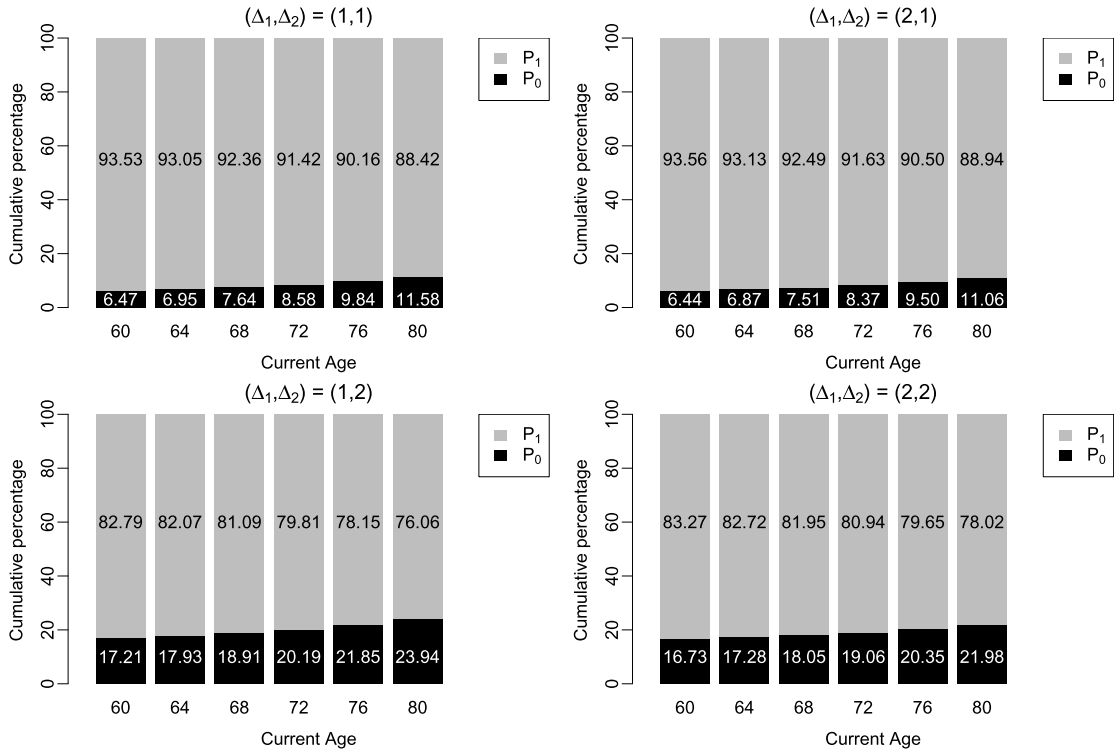

Figure 2. The bar plots of percentage changes for P_0 and P_1 with different t_{K_1}: Six bars representing different current ages are plotted for each of the four screening schedules, $(\Delta_1, \Delta_2) = (1, 1)$ (upper left panel), $(\Delta_1, \Delta_2) = (2, 1)$ (upper right panel), $(\Delta_1, \Delta_2) = (1, 2)$ (bottom left panel) and $(\Delta_1, \Delta_2) = (2, 2)$ (bottom right panel). $\beta = 0.7$, $MST = 5$, any $t_0 (< t_{K_1})$.

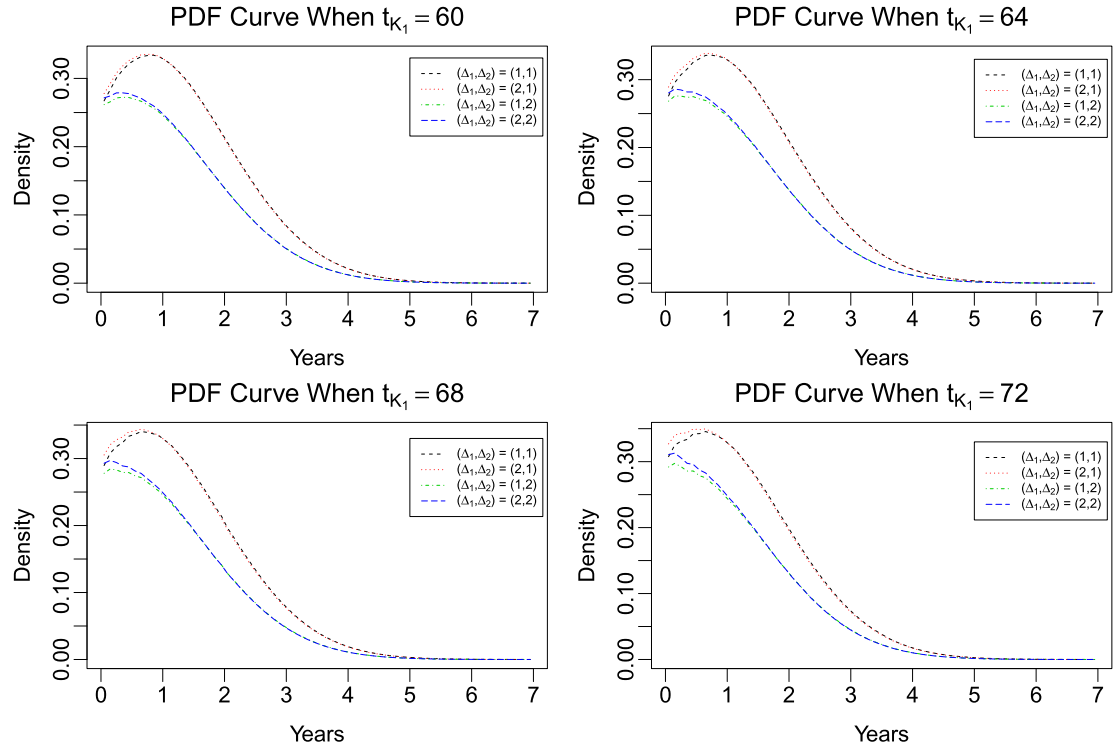

Figure 3. The sub-PDF curves of the lead time for $t_0 = 56$, $\beta = 0.7$, $MST = 2$.

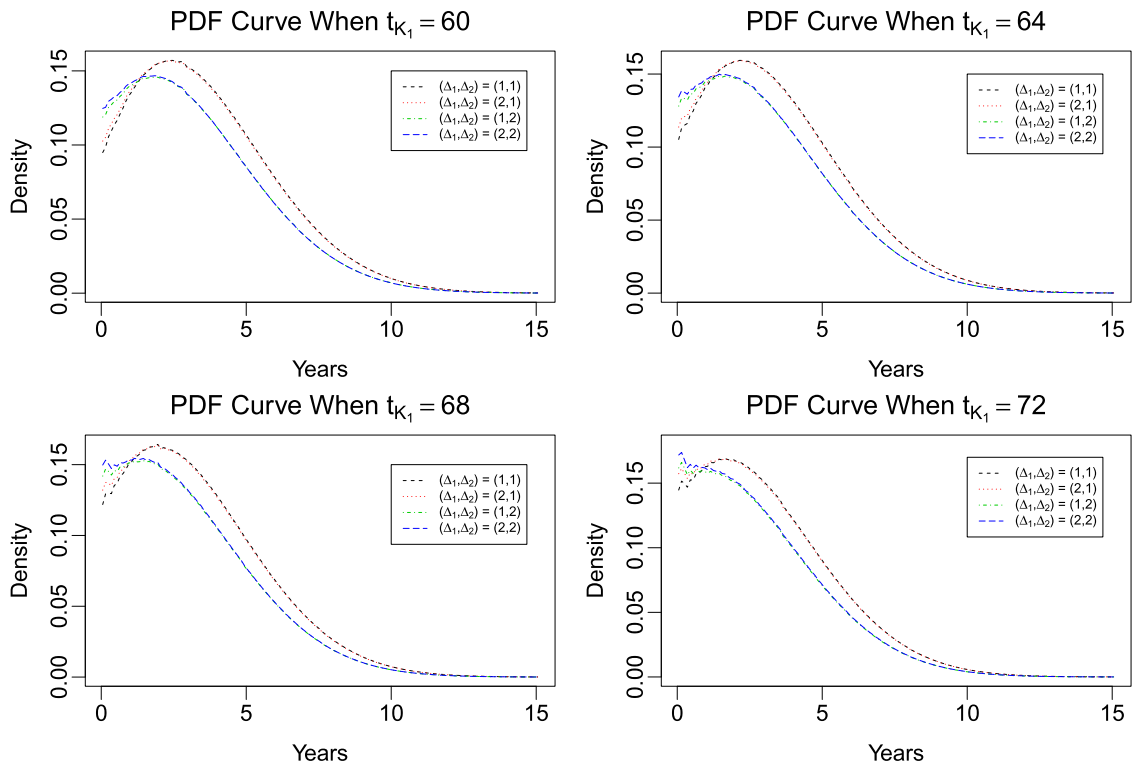

Figure 4. The sub-PDF curves of the lead time for $t_0 = 56$, $\beta = 0.7$, MST = 5.

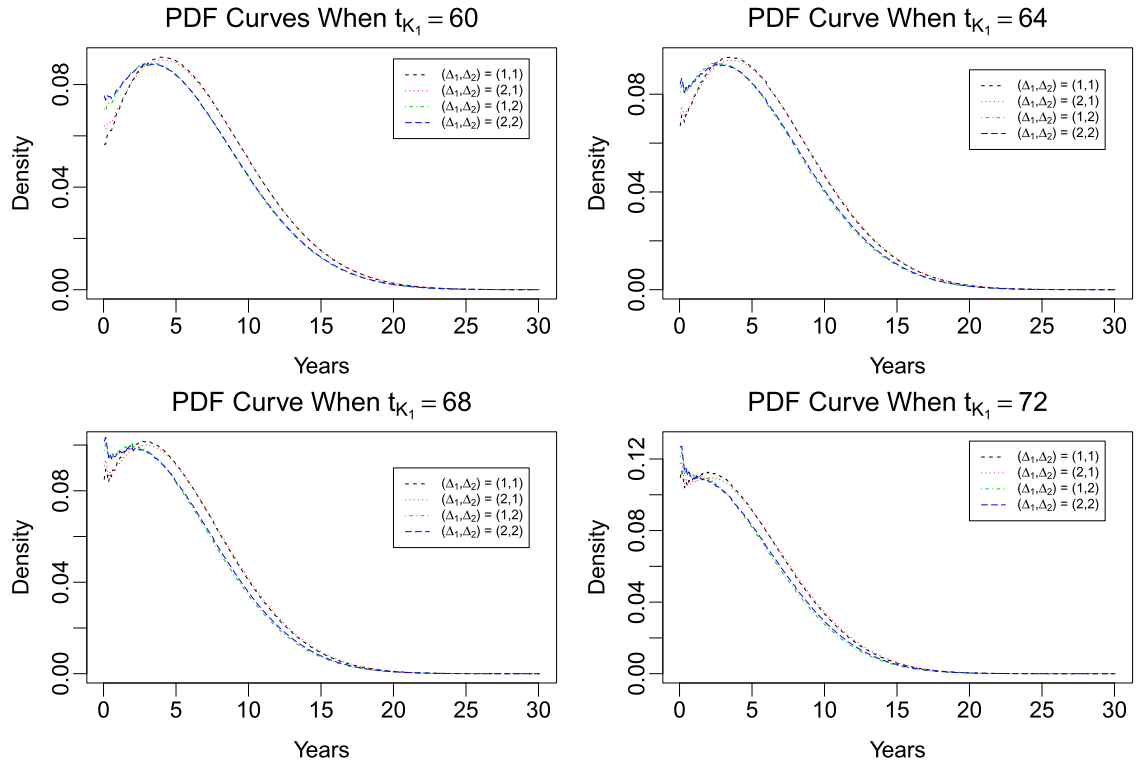

Figure 5. The sub-PDF curves of the lead time for $t_0 = 56$, $\beta = 0.7$, MST = 10.

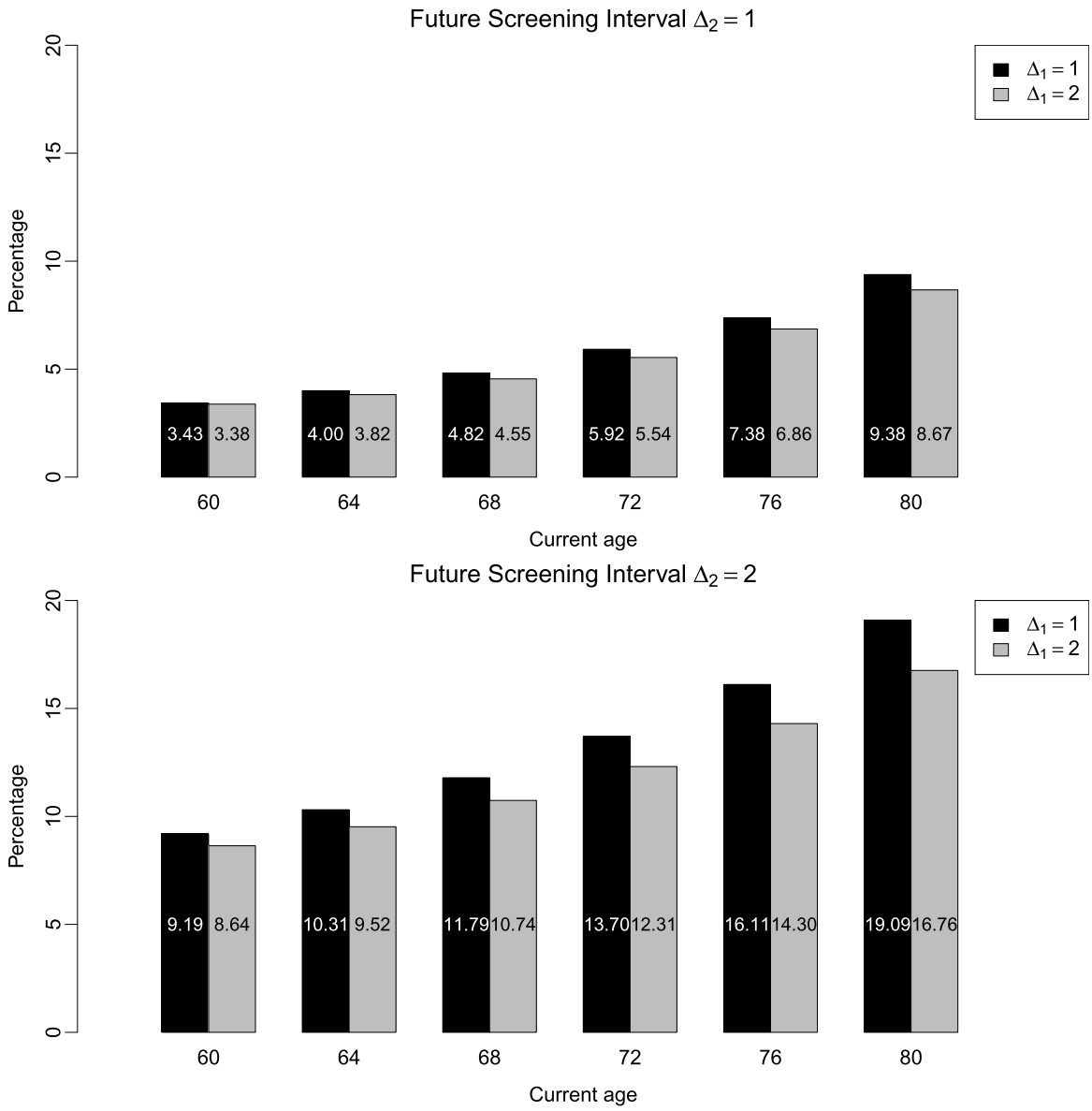

Figure 6. The bar plots of percentage changes for P_0 with different Δ_1 and the same Δ_2: Bars grouped by six different current ages are plotted for two future schedules, $\Delta_2 = 1$ (upper panel) and $\Delta_2 = 2$ (bottom panel). $\beta = 0.7$, MST = 10, any t_0.

that. For different mean sojourn time, we chose different values of λ. The values of the input parameters in the simulation are shown in Table 1.

Simulation results for MST = 2, 5 and 10 are shown in Tables 2–4, respectively. In each table, we report P_0, the probability that the lead time is zero, in percentage. This is also the possibility that one would not be detected early by screening. When the lead time is positive, we report the mean lead time EL, its standard deviation, the median and the mode of lead time in years. Larger lead time usually means that the person benefits more from the screening program since the treatments and interventions could be provided earlier.

Intuitively, the length of lead time will depend on the length of sojourn time. This is reflected in our simulation: a longer sojourn time will lead to a longer lead time. For example, the mean lead time is about 1 year if the mean sojourn time (MST) is 2 years; and it is about 2 to 3 years if the MST is 5 years, and it is about 3 to 7 years if the MST is 10 years. So they are positively correlated. Meanwhile, the probability of no-early-detection P_0 is negatively correlated with the MST: longer MST will cause smaller P_0, indicating that people with slow growing tumor will benefit more from screening. For example, if an individual started biennial screening exam ($\Delta_1 = 2$) at age 56, and will continue screening annually ($\Delta_2 = 1$) from her current age $t_{K_1} = 64$,

Table 5. A projection of the lead time distribution for male heavy smokers with initial screening age $t_0 = 56$

(Δ_1, Δ_2) (years)	P_0 (95% C.I.)	$1 - P_0$ (s.d.)	EL (s.d.)	Med/IQR
		current age $t_{K_1} = 60$		
(1,1)	11.83 (7.37, 17.93)	88.17 (2.66)	0.87 (0.69)	1.06
(2,1)	11.67 (7.28, 17.79)	88.33 (2.62)	0.86 (0.68)	0.94
(-,1)	11.65 (7.28, 17.76)	88.35 (2.61)	0.86 (0.68)	0.94
(1,2)	36.91 (28.21, 45.48)	63.09 (4.49)	0.54 (0.66)	0.83
(2,2)	36.30 (27.59, 44.99)	63.70 (4.53)	0.54 (0.66)	0.83
(-,2)	36.35 (27.67, 45.06)	63.65 (4.54)	0.55 (0.66)	0.83
		current age $t_{K_1} = 64$		
(1,1)	11.86 (6.93, 19.12)	88.14 (3.04)	0.86 (0.68)	0.94
(2,1)	11.62 (6.82, 18.83)	88.38 (3.00)	0.85 (0.68)	0.94
(-,1)	11.58 (6.82, 18.77)	88.42 (2.99)	0.85 (0.68)	0.94
(1,2)	36.60 (27.69, 45.84)	63.40 (4.72)	0.55 (0.66)	0.83
(2,2)	35.68 (26.67, 45.19)	64.32 (4.78)	0.54 (0.65)	0.83
(-,2)	35.68 (26.42, 45.11)	64.32 (4.83)	0.55 (0.66)	0.83
		current age $t_{K_1} = 68$		
(1,1)	12.00 (6.47, 20.89)	88.00 (3.62)	0.85 (0.68)	0.94
(2,1)	11.66 (6.42, 20.48)	88.34 (3.59)	0.84 (0.68)	0.94
(-,1)	11.61 (6.39, 20.48)	88.39 (3.58)	0.83 (0.67)	0.94
(1,2)	36.31 (27.09, 46.40)	63.69 (5.08)	0.54 (0.65)	0.83
(2,2)	35.01 (25.60, 45.55)	64.99 (5.18)	0.54 (0.65)	0.83
(-,2)	34.96 (25.23, 45.45)	65.04 (5.26)	0.54 (0.65)	0.94
		current age $t_{K_1} = 72$		
(1,1)	12.29 (6.07, 23.67)	87.71 (4.43)	0.83 (0.68)	0.94
(2,1)	11.84 (5.96, 23.35)	88.16 (4.43)	0.81 (0.67)	0.94
(-,1)	11.77 (5.96, 23.33)	88.23 (4.43)	0.81 (0.67)	1.06
(1,2)	36.04 (26.18, 47.83)	63.96 (5.60)	0.54 (0.65)	0.83
(2,2)	34.34 (24.37, 46.75)	65.66 (5.78)	0.54 (0.64)	0.94
(-,2)	34.22 (23.88, 46.78)	65.78 (5.89)	0.54 (0.64)	0.81

with screening sensitivity $\beta = 0.7$, then P_0 will be 20.23%, 6.87%, and 3.82% correspondingly under the different MST of 2, 5 and 10 years. Finally, screening sensitivity is negatively correlated with the P_0: higher sensitivity will lead to lower P_0. Comparing the results for $\beta = 0.7$ and 0.9 under the same conditions, it appears that the probability P_0 is almost doubled in case of the lower sensitivity. Therefore, higher sensitivity will contribute greatly to early detection.

By examining three different initial screening ages t_0, we want to see if this factor t_0 affects the lead time given a person looks healthy at current age, or if the screening history has any influence on the lead time given the same current age. We found that the lead time distribution tends to be the same for different t_0 if the current age t_{K_1} is the same. Simply look at the results of $t_{K_1} = 68$, we compare the results of $(t_0, t_{K_1}) = (56, 68)$, $(60, 68)$, and $(64, 68)$ in Tables 2–4, they are almost the same. It is also true when we compare the results of $(t_0, t_{K_1}) = (56, 72)$, $(60, 72)$, and $(64, 72)$. This indicates that the factor t_0 does not seem to affect the lead time distribution provided that the person still looks healthy at the current age t_{K_1}. Figure 1 shows the density plots of the lead time given the MST is 2 years and sensitivity $\beta = 0.7$, for $t_{K_1} = 68$ and $t_{K_1} = 72$, and under the 4 different combinations of (Δ_1, Δ_2). In this graph, each curve, in fact, represents the density of three different initial screening ages (i.e., $t_0 = 56$, 60 and 64), but since the three curves (of the three t_0) completely overlap each other, we can only observe one curve for each pair of (Δ_1, Δ_2) in each panel. However, the length of the past (Δ_1) and future screening interval (Δ_2) do affect the lead time density, with the future screening intervals Δ_2 cause more differences.

When the mean sojourn time (MST) is larger than 5 years, the probability of no-early-detection P_0 is slightly increasing with a participant's current age t_{K_1} given that all other factors are the same. This means that the younger participants (such as people in their 60s) may benefit slightly more from the screening. For example, in Table 3, the probability P_0 is 17.21% and the mean lead time is 2.80 yrs for an individual who started annual screening exam at age 56, and will begin screening biennially from current age $t_{K_1} = 60$, if the sensitivity $\beta = 0.7$. The probability P_0 slightly goes up to 20.19% and the mean lead time becomes 2.36 yrs when the individual's current age is 72. Figure 2 gives percentages of P_0 and $P_1 = 1 - P_0$, when $\beta = 0.7$, MST is 5 years, and t_{K_1} increases from 60 to 80 in a 4-year interval. We can see that P_0 increases as the current age t_{K_1} increases for all four screening schedules (Δ_1, Δ_2). Since the results are the same

Table 6. A projection of the lead time distribution for female heavy smokers with initial screening age $t_0 = 56$

(Δ_1, Δ_2) (years)	P_0 (95% C.I.)	$1 - P_0$ (s.d.)	EL (s.d.)	Med/IQR
		current age $t_{K_1} = 60$		
(1,1)	6.87 (3.94, 10.93)	93.13 (1.81)	1.06 (0.72)	1.17
(2,1)	6.78 (3.91, 10.73)	93.22 (1.77)	1.05 (0.72)	1.17
(-,1)	6.76 (3.91, 10.68)	93.24 (1.76)	1.05 (0.72)	1.17
(1,2)	28.69 (19.83, 38.26)	71.31 (4.64)	0.69 (0.73)	0.94
(2,2)	28.15 (19.39, 37.77)	71.85 (4.63)	0.69 (0.73)	0.94
(-,2)	28.26 (19.41, 37.85)	71.74 (4.65)	0.69 (0.73)	0.94
		current age $t_{K_1} = 64$		
(1,1)	6.84 (3.82, 10.98)	93.16 (1.90)	1.05 (0.72)	1.17
(2,1)	6.69 (3.74, 10.68)	93.31 (1.85)	1.04 (0.72)	1.17
(-,1)	6.67 (3.74, 10.60)	93.33 (1.84)	1.04 (0.72)	1.05
(1,2)	28.44 (19.38, 37.99)	71.56 (4.65)	0.69 (0.73)	0.94
(2,2)	27.63 (18.76, 37.18)	72.37 (4.65)	0.69 (0.72)	0.94
(-,2)	27.69 (18.71, 37.29)	72.31 (4.69)	0.69 (0.72)	0.94
		current age $t_{K_1} = 68$		
(1,1)	6.85 (3.70, 11.24)	93.15 (2.06)	1.04 (0.72)	1.17
(2,1)	6.64 (3.59, 10.99)	93.36 (2.00)	1.03 (0.72)	1.05
(-,1)	6.60 (3.59, 10.96)	93.40 (1.99)	1.02 (0.72)	1.17
(1,2)	28.20 (19.36, 37.59)	71.80 (4.68)	0.69 (0.72)	0.94
(2,2)	27.06 (18.24, 36.58)	72.94 (4.69)	0.69 (0.72)	0.94
(-,2)	27.06 (18.00, 36.61)	72.94 (4.77)	0.68 (0.72)	0.94
		current age $t_{K_1} = 72$		
(1,1)	6.92 (3.57, 12.24)	93.08 (2.31)	1.03 (0.72)	1.17
(2,1)	6.63 (3.39, 11.78)	93.37 (2.26)	1.00 (0.72)	1.06
(-,1)	6.59 (3.39, 11.63)	93.41 (2.25)	1.00 (0.71)	1.06
(1,2)	27.97 (19.11, 37.41)	72.03 (4.76)	0.69 (0.72)	0.94
(2,2)	26.45 (17.74, 35.99)	73.55 (4.78)	0.68 (0.71)	0.94
(-,2)	26.39 (17.20, 36.03)	73.61 (4.90)	0.68 (0.71)	0.75

for different t_0 (provided $t_0 < t_{K_1}$), we put results of all t_{K_1} together in the bar plots regardless of t_0.

For a given combination of the sensitivity, the initial and current age, we compare the results of different screening schedules (Δ_1, Δ_2). For the cases with the same Δ_2 (future screening interval) but different Δ_1 (historic screening interval), the lead time distribution tends to be very similar. For example, if we compare the results of $(\Delta_1, \Delta_2) = (1,1)$, $(1,2)$ and $(2,1)$. It is obvious that the results of $(\Delta_1, \Delta_2) = (1,1)$ are significantly different from the results of $(\Delta_1, \Delta_2) = (1,2)$, but the results for $(\Delta_1, \Delta_2) = (1,1)$ and $(\Delta_1, \Delta_2) = (2,1)$ are very close. We can also see this pattern from the PDF curves of lead time when MST = 2, 5 and 10, as shown in Figures 3–5, respectively. Since the results are similar, we only present curves of $t_0 = 56$ and $\beta = 0.7$ for different t_{K_1} and MST. The PDF curves for lead time with the same future screening interval Δ_2 almost overlap each other for different Δ_1 when given the same initial and current age, the sensitivity and the mean sojourn time.

We find a small trend that larger Δ_1 will result in smaller P_0 and larger mean lead time if the Δ_2 remains the same, and this trend is obvious when MST is larger. In Table 4, the probability P_0 is 9.19% and the mean lead time is 5.58 years for an individual whose initial screening age $t_0 = 56$ and current age $t_{K_1} = 60$ with screening schedules $(\Delta_1, \Delta_2) = (1,2)$ given $\beta = 0.7$. The probability P_0 decreases to 8.64% and the mean lead time increases to 5.63 years if the individual's past screening interval $\Delta_1 = 2$. We can see the trend more clearly in Figure 6, where smaller Δ_1 resulted in larger P_0, and the difference is getting larger as the current age increases. We didn't carry out more simulations to compare with the case when one's screening history doesn't exist (ie, previous model in Wu et al. [16]), however, based on the simulation results for different Δ_1 and different t_0, the influence of a person's screening history on his/her future lead time distribution seems negligible using our current model assumption.

4. APPLICATION

The recently finished NLST study is designed to compare two different screening modalities for early detection of lung cancer among heavy smokers: low-dose computed tomography (LDCT) versus standard chest X-rays [23]. 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: LDCT or X-ray. Partic-

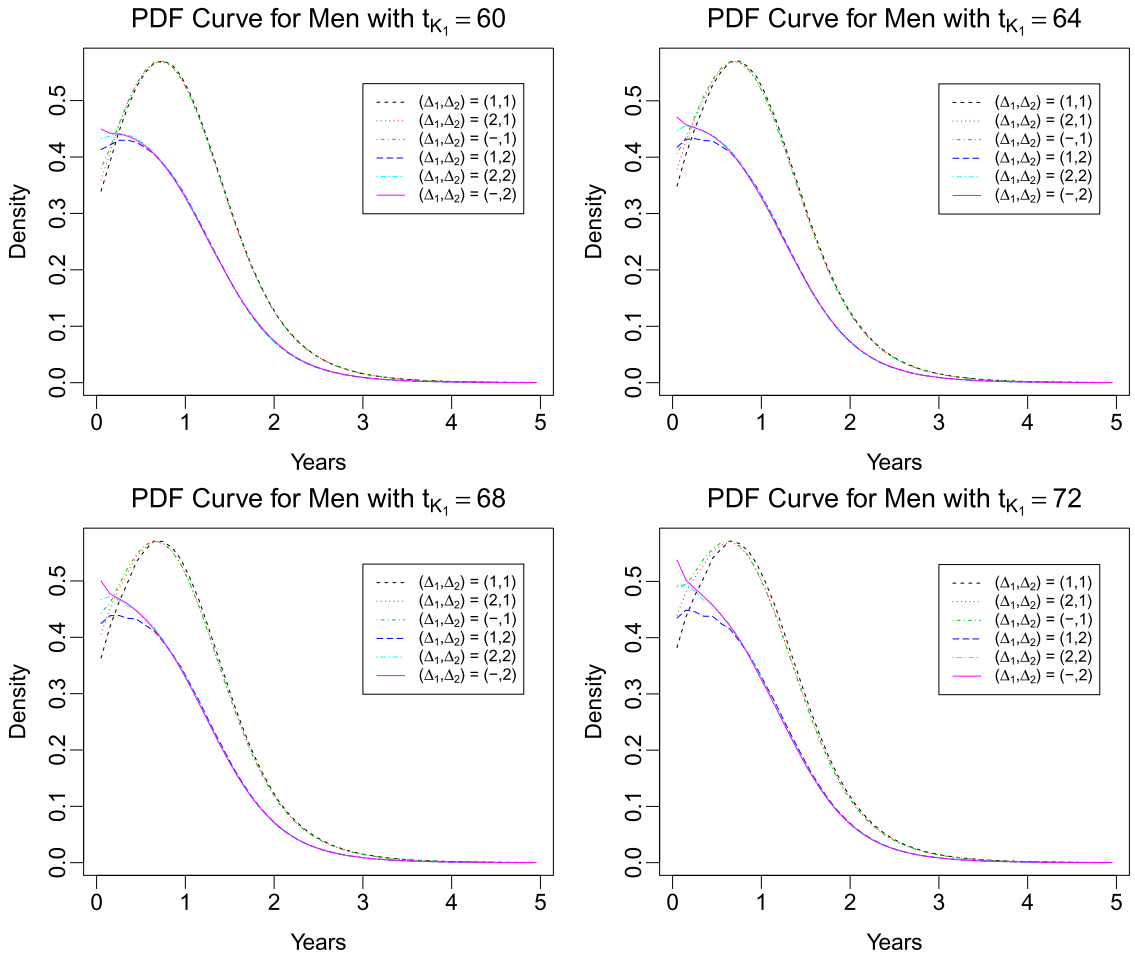

Figure 7. The sub-PDF curves of lead time for male heavy smokers when $t_0 = 56$: Six curves of different screening schedules for four current ages, $t_{K_1} = 60$ (upper left panel), $t_{K_1} = 64$ (upper right panel), $t_{K_1} = 68$ (bottom left panel) and $t_{K_1} = 72$ (bottom right panel).

ipants 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 LDCT 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. The NLST results seem to indicate that smokers screened by LDCT had a 20% lower chance of dying from lung cancer than those who were screened via chest X-rays.

We apply the new lead time method to the NLST-LDCT group data, using a parametric model for the sensitivity, $\beta(t) = \{1+\exp[-b_0 - b_1(t-\bar{t})]\}^{-1}$ and the same parametric models for the $w(t), q(x)$ and $Q(x)$ as in Equations (14)–(16). The six unknown parameters $\theta = (b_0, b_1, \mu, \sigma^2, \lambda, \alpha)$ were estimated using Markov Chain Monte Carlo (MCMC) and a likelihood function based on the NLST-LDCT data [22]: two Markov Chains were simulated with overdispersed initial values. Each chain ran 130,000 iterations, with 30,000 burn-in steps; after the burn-in, the posteriors were sampled every 200 steps, providing 500 posteriors θ^* from each chain. We then used the pooled 1,000 posterior samples $\theta_j^*, j = 1, \ldots, 1000$ (from the two chains) to make inference on the lead time for hypothetic cohorts. The posterior predictive distribution of the lead time is a weighted average at each θ_j^*: $f_L(l|NLST) \approx \frac{1}{1000}\sum_{j=1}^{n} f_L(l|\theta_j^*)$.

For each gender, we assumed that there were four cohorts of initially asymptomatic individuals, with current age $t_{K_1} = 60$, 64, 68 and 72, respectively. Then within each cohort, we examined six different screening schedules $(\Delta_1, \Delta_2) = (1,1)$, $(2,1)$, $(-,1)$, $(1,2)$, $(2,2)$, and $(-,2)$, The symbol $\Delta_1 = -$ means that the individuals have no screening history; this is added for comparison purpose. We assumed that the initial screening age $t_0 = 56$ for all cases (except those without any screening history), since our simu-

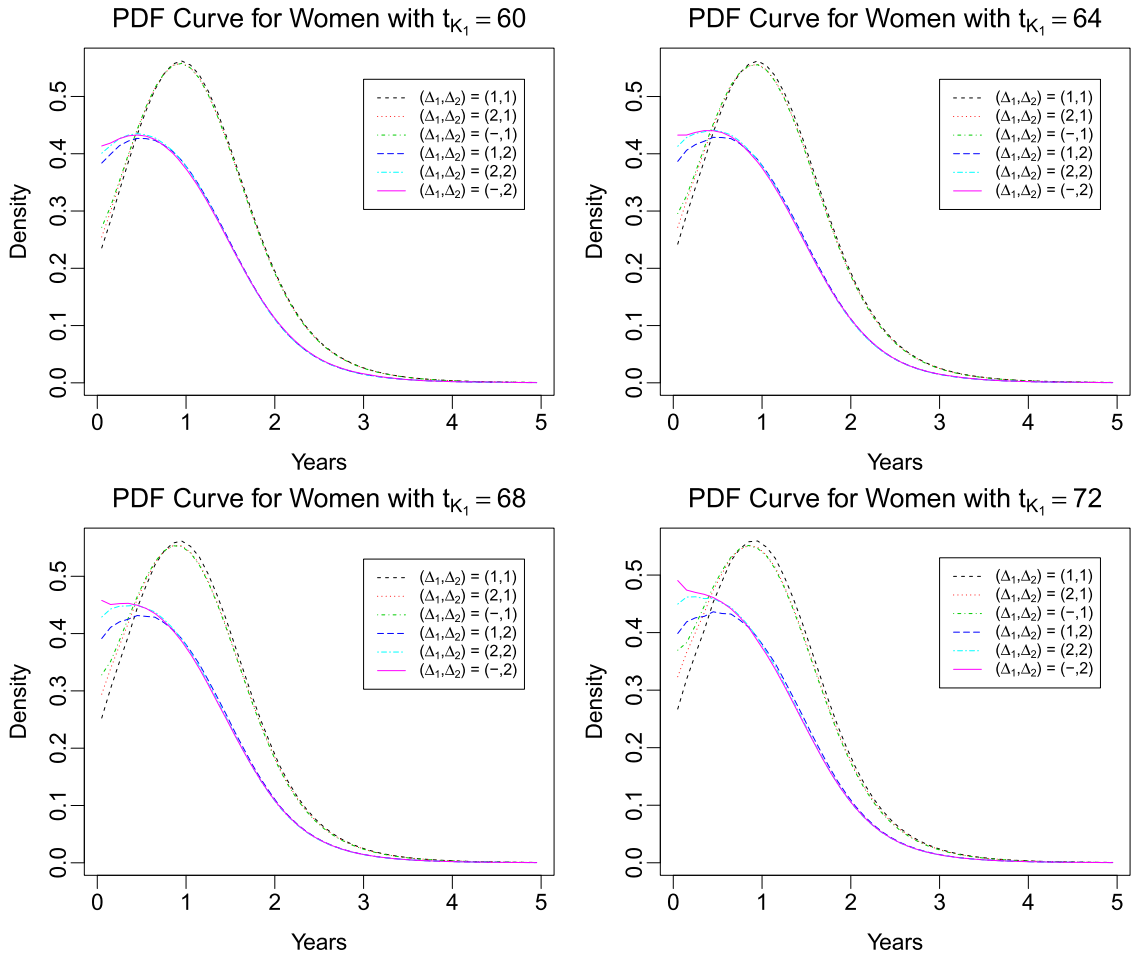

Figure 8. The sub-PDF curves of lead time for female heavy smokers when $t_0 = 56$: Six curves of different screening schedules for four current ages, $t_{K_1} = 60$ (upper left panel), $t_{K_1} = 64$ (upper right panel), $t_{K_1} = 68$ (bottom left panel) and $t_{K_1} = 72$ (bottom right panel).

lation study showed that the initial screening age has no significant influence on the lead time distribution. We present simulation results of these 24 scenarios (4 ages times 6 schedules) for both genders.

Table 5 and Table 6 present the Bayesian predictive inference for the probability of no-early-detection (P_0), the probability of early-detection ($1 - P_0$) and the mean lead time (in years) for male and female heavy smokers, respectively. The probabilities, with its 95% C.I., and the corresponding standard deviation are reported as percentages. The mean and the standard deviation are reported in years; we also reported the ratio of the median over the interquartile range (Med/IQR).

For both male and female heavy smokers, we can see that the results of $(\Delta_1, \Delta_2) = (1, 1)$, $(2, 1)$, and $(-, 1)$ are very similar, and the results of $(\Delta_1, \Delta_2) = (1, 2)$, $(2, 2)$, and $(-, 2)$ are similar, which are compatible with our expectations. That is, the future screening schedule plays a far more important role than the historic screening schedule regarding the lead time distribution. For example, in Table 5, the probability P_0 is 11.83% and the mean lead time is 0.87 years for screening schedules $(\Delta_1, \Delta_2) = (1, 1)$, and P_0 is 11.67% and the mean lead time is 0.86 for $(\Delta_1, \Delta_2) = (2, 1)$ given the person's current age $t_{K_1} = 60$. The probability P_0 is 11.65% and the mean lead time is 0.86 years if a male heavy smoker has no screening history, but plans to take annual exam starting at age 60. We also present the sub-PDF curves of lead time (the lead time distribution is a mixture of a point mass at 0 and a positive sub-PDF) for male and female heavy smokers in Figure 7 and 8, respectively. In each figure, four panels represent four different current ages ($t_{K_1} = 60$, 64, 68 and 72). In each panel, the six curves are the corresponding lead time densities for six different screening intervals: $(\Delta_1, \Delta_2) = (1, 1)$, (2,1), (-,1), (1,2), (2,2) and (-,2). The sub-PDF curves of lead time for the same future screening interval Δ_2 almost overlap with each other given the same current age, and they are almost the same if the lead time is larger than one year.

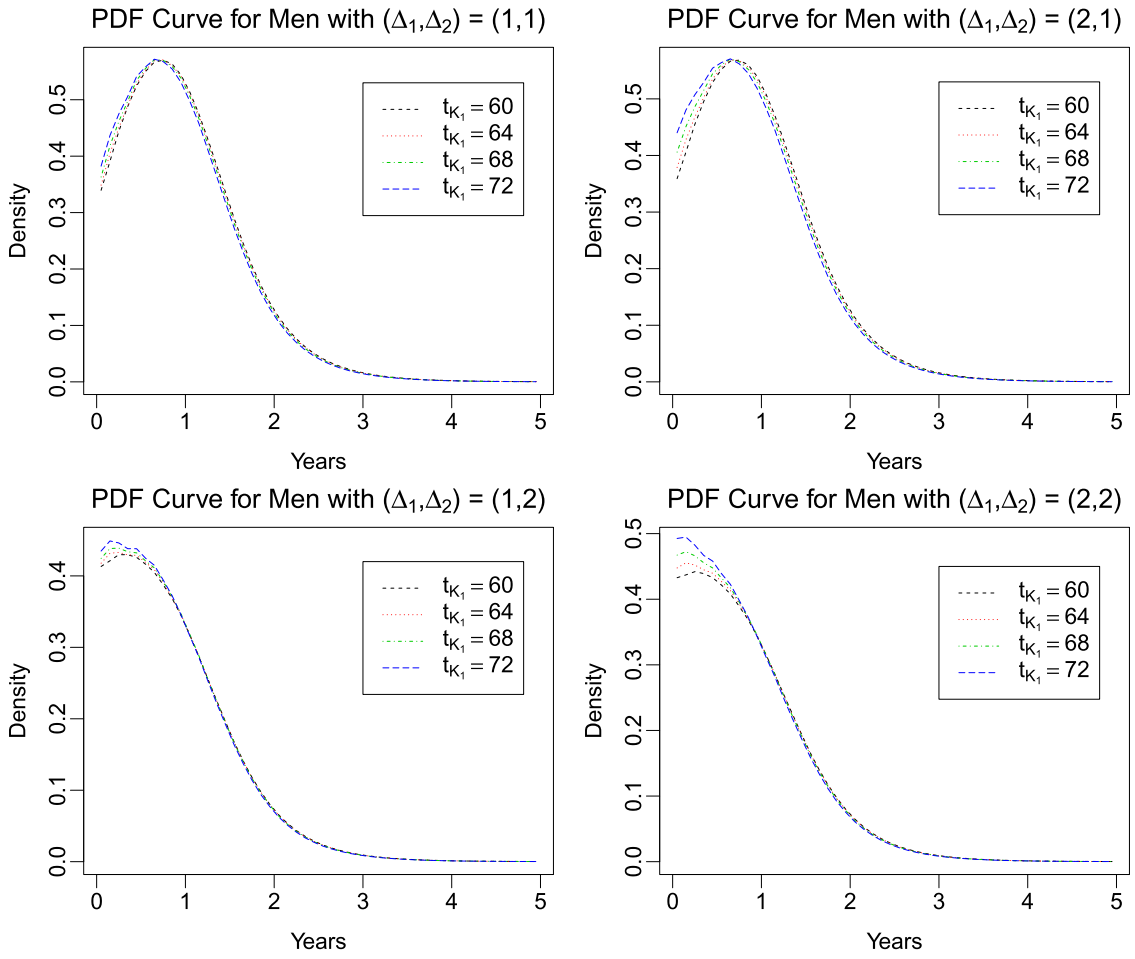

Figure 9. The lead time sub-PDF curves for male heavy smokers when $t_0 = 56$: Four curves representing different current ages of four screening schedules, $(\Delta_1, \Delta_2) = (1, 1)$ (upper left panel), $(\Delta_1, \Delta_2) = (2, 1)$ (upper right panel), $(\Delta_1, \Delta_2) = (1, 2)$ (bottom left panel) and $(\Delta_1, \Delta_2) = (2, 2)$ (bottom right panel).

Like in the simulation study, we also find a similar trend that larger Δ_1 will result in smaller P_0 if Δ_2 is the same. In Table 6, the probability P_0 is 6.84% for current age $t_{K_1} = 64$ with screening schedules $(\Delta_1, \Delta_2) = (1, 1)$, and it drops to 6.69% if the individual's past screening interval $\Delta_1 = 2$.

For all heavy smokers, it is obvious that the probability P_0 increases and the mean lead time decreases as the future screening interval Δ_2 increases within the same age group. Across the age groups, the probability P_0 and the mean lead time does not seem to have significant differences. To illustrate, let us see the lead time density curves of both genders in Figure 9 and 10. Four panels represent four different screening schedules, and four curves represent four current ages within each panel. In each panel, the curves do not differentiate too much except in the very beginning.

In addition, the projected lead time is significantly affected by gender. Comparing to female heavy smokers, male heavy smokers usually have a larger P_0 and a shorter mean lead time, given the same age and the same screening schedule. It seems that male heavy smokers will have a smaller chance to be detected early by LDCT screening than their female counterparts.

5. DISCUSSION

We derived the lead time distribution for individuals with a screening history when human lifetime is random. Simulation study was done to investigate the effect of a person's screening history on the lead time distribution in the future. We estimated the projected lead time for male and female heavy smokers, regarding different ages, past and future screening schedules using the NLST low dose CT arm data. Although our simulations were carried out for equally-spaced screening intervals, the method can be applied to any screening scenarios.

In the simulation study, we found that the lead time and the sojourn time are positively correlated: people with a longer sojourn time (or slowly growing tumors) will benefit

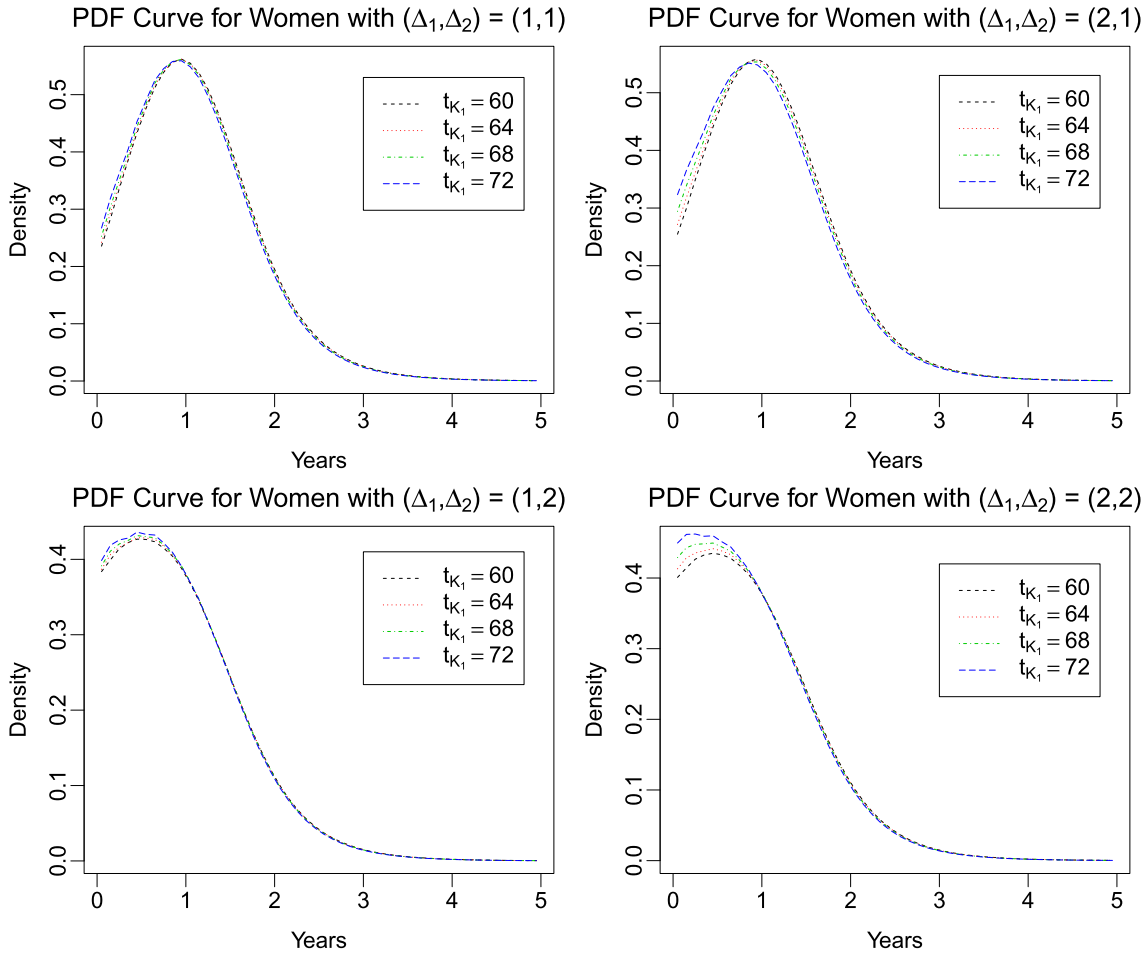

Figure 10. The lead time sub-PDF curves for female heavy smokers when $t_0 = 56$*: Four curves representing different current ages of four screening schedules,* $(\Delta_1, \Delta_2) = (1, 1)$ *(upper left panel),* $(\Delta_1, \Delta_2) = (2, 1)$ *(upper right panel),* $(\Delta_1, \Delta_2) = (1, 2)$ *(bottom left panel) and* $(\Delta_1, \Delta_2) = (2, 2)$ *(bottom right panel).*

more from screening, in terms of a longer lead time and a higher probability of early detection (i.e., smaller P_0). The lead time distribution depends more on one's current age t_{K_1}, rather than on one's initial exam age t_0. The probability of no-early-detection P_0 slightly increases with one's current age, if all other conditions are the same. There is a small trend that larger screening interval Δ_1 in the past will result in a smaller P_0 (probability that the lead time is 0), or, in another words, a slightly larger chance of early-detection and a longer mean lead time if the individual's future screening interval is fixed (see Figure 6 and Table 4). In the NLST application, we found a similar trend. For a given current age, the length of screening history does not really affect the lead time distribution much. This indicates that one's current age is more important to the lead time distribution than the initial screening age, since the person still looks healthy at his/her current age.

The influence of a person's screening history on his/her future lead time distribution is not significant. This may be due to our model assumption. We assumed that the sojourn time Y and the time duration in the preclinical state X are independent; we also assumed that the sensitivity does not depend on the time one stays in the preclinical state. While in reality, the sensitivity is low when one just enters the preclinical state, and it is close to one when one is at the end of the preclincal state (i.e., one is close to the onset of the clinical state). These assumptions were made to simplify the derivation procedure of the lead time. We will consider relaxing on these assumptions in the next step, in which case, the screening history may play some role on one's future lead time when more realistic (and complicated) models are applied. Another possible reason that one's screening history seems to have little effect on future lead time, may be due to the fact that, we are predicting the future lead time given two conditions: a). one's screening history, and b). one looks healthy and is asymptomatic at the current age. The second condition may carry much more information/message than the first, and thus may overshadow the impact of the first

one. This may explain why the lead time distribution is more related to one's current situation. More research are needed to verify this point.

We want to point out that the estimated lead time distribution was robust under diffent parametric models. We have done extensive simulations on the behavior of the three key parameters $\beta(t), w(t), q(x)$, using different parametric models [24]. It turns out that these three key parameter won't change much under different parametric models and under the current model assumption. Since the lead time is a function of these three, it won't change much either under the current model assumption.

In summary, this project is an extension on the lead time distribution given one's screening history with negative exam results. We realize that the modeling approach is just one way of dealing with the problem. Other models and approaches may be possible. The lead time distribution is an important indicator of how effective a screening schedule and screening modality are. However, early detection may cause higher percentage of over-diagnosis, that is, cancer cases that would never have caused clinical symptoms in one's lifetime, and the individual would die of other causes [25]. Too frequent screening exam may not be necessary, especially for people with low risk. Too many exams is not only a waste of resource, but also may cause unnecessary stress and anxiety if one gets a false positive result. We are working on optimal scheduling problem based on one's historical screening history and other parameters, and we hope to provide some guidelines regarding the timing of future screening exam.

ACKNOWLEDGEMENT

We want to thank the two anonymous referees and the AE for their constructive comments and suggestions.

Received 19 September 2019

REFERENCES

[1] SEER Fast Stats Results, NIH., http://seer.cancer.gov/statfacts/html/lungb.html, [Accessed: 2017-10-15].

[2] The United States Preventive Services Task Force, https://www.uspreventiveservicestaskforce.org/BrowseRec/Index, [Accessed: 2016-10-10].

[3] Shapiro, S., Venet, W., Strax, P. and Venet, L. (1988). *Periodic Screening for Breast Cancer: The Health Insurance Plan Project and Its Sequelae 1963–1986*. Baltimore: Johns Hopkins University Press.

[4] Fontana, Robert S. et al. (1975). The Mayo Lung Project for Early Detection and Localization of Bronchogenic Carcinoma: A Status Report. *CHEST* **67**(5) 511–522.

[5] Berlin, Nathaniel I., Buncher, C. Ralph, Fontana, Robert S., Frost, John K. and Melamed, Myron 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.

[6] 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.

[7] Wang D., Levitt B., Riley T., 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.

[8] 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.

[9] Shieh Y. and Bohnenkamp M. (2017). Low-dose CT Scan for Lung Cancer Screening: Clinical and Coding Considerations. *CHEST* **152**(1) 204–209.

[10] Wu, Dongfeng, Rosner, Gary L. and Broemeling, Lyle D. (2007). Bayesian Inference for the Lead Time in Periodic Cancer Screening. *Biometrics* **63** 873–880. MR2395806

[11] Kafadar K. and Prorok Philip C. (1994). A Data-analytic Approach for Estimating Lead Time and Screening Benefit Based on Survival Curves in Randomized Cancer Screening Trials. *Statistics in Medicine* **13**(5-7) 569–586.

[12] Kafadar K. and Prorok Philip C. (1996). Computer Simulation of Randomized Cancer Screening Trials to Compare Methods of Estimating Lead Time and Benefit Time. *Computational Statistics & Data Analysis* **23**(2) 263–291.

[13] Kafadar K., Prorok Philip C. and Smith Paul J. (1999). An Estimate of the Variance of Estimators for Lead Time and Screening Benefit Time in Randomized Cancer Screening Trials. *Biometrical Journal* **40**(7) 801–821.

[14] Kafadar K. and Prorok Philip C. (2003). Alternative Definitions of Comparable Case Groups and Estimates of Lead Time and Benefit Time in Randomized Cancer Screening Trials. *Statistics in Medicine* **22**(1) 83–111.

[15] Prorok P. C. (1982). Bounded Recurrence Times and Lead Time in the Design of A Repetitive Screening Program. *Journal of Applied Probability* **19**(1) 10–19. MR0644415

[16] Wu, Dongfeng, Kafadar, Karen, Rosner, Gary L. and Broemeling, Lyle D. (2012). The Lead Time Distribution When Lifetime is Subject to Competing Risks in Cancer Screening. *The International Journal of Biostatistics* **8**. MR2923280

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

[18] 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.

[19] 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.

[20] Liu, Ruiqi, Pérez, Adriana and Wu, Dongfeng (2018). Estimation of Lead Time via Low-Dose CT in the National Lung Screening Trial. *Journal of Healthcare Informatics Research* **2**(4) 353–366.

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

[22] 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.

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

[24] Wu, Dongfeng, Rosner, Gary L. and Broemeling, Lyle D. (2005). MLE and Bayesian Inference of Age-Dependent Sensitivity and Transition Probability in Periodic Screening. *Biometrics* **61** 1056–1063. MR2216199

[25] 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.

Ruiqi Liu
Department of Bioinformatics and Biostatistics
University of Louisville
USA
E-mail address: ruiqi.liu@louisville.edu

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

Shesh N. Rai
Department of Bioinformatics and Biostatistics
University of Louisville
USA
E-mail address: shesh.rai@louisville.edu