When to initiate a screening exam if sensitivity is a function of sojourn time

Dongfeng Wu

Department of Bioinformatics and Biostatistics
School of Public Health and Information Sciences
The University of Louisville

SDSS, June 9, 2022

Motivation

- One major problem in cancer screening is the scheduling problem: When to initiate the exam? and for an asymptomatic individual who has gone through a few screening exams in the past and got negative results, when should s/he come back for the next exam?
- Wu (2022) used incidence probability to find the first screening time based on a person's current age and other parameters, by limiting the clinical incidence risk to a pre-selected small value. After this time interval is found, we can further estimate the lead time distribution and probability of overdiagnosis.

Motivation

- This project is an improvement of the original model, where screening sensitivity and sojourn time were assumed to be uncorrelated.
- The extension is based on the reality that the sensitivity is low when one just enters the preclinical state, and it is close to one at the end of the preclinical state. Therefore, the sensitivity is modeled as a function of the ratio of time one stayed in the preclinical state relative to the sojourn time.
- The method can be applied to any kind of screening. We will use the National Lung Screening Trial (NLST) computed tomography (CT) data for male and female heavy smokers as an example.

The NLST Study

- About 54,000 Male and Female heavy smokers were enrolled between 08/2002-04/2004. Data collection was finished by 12/2009.
- They were randomized to 2 arms: chest X-ray or low-dose spiral CT.
- Each arm underwent 3 annual screenings; more tumor cases were diagnosed in the CT arm than that in the chest X-ray.
- Initial screening age 55−74.

Table 1: The NLST Data - Overview

Group within Study	[a]total subj.	[b]Screen-diag. No.	[c]Interval No.
The NLST: Chest X-ray			
Overall	26226	279	177
male smokers	15500	165	107
female smokers	10726	114	70
The NLST: Spiral CT			
Overall	26452	649	60
male smokers	15621	384	44
female smokers	10831	265	16

[a] Total number of people who ever received chest X-ray for lung cancer.
[b] Total number of subjects diagnosed by regular screening.
[c] Total number of clinical incident cases between two regular screenings.

Definition

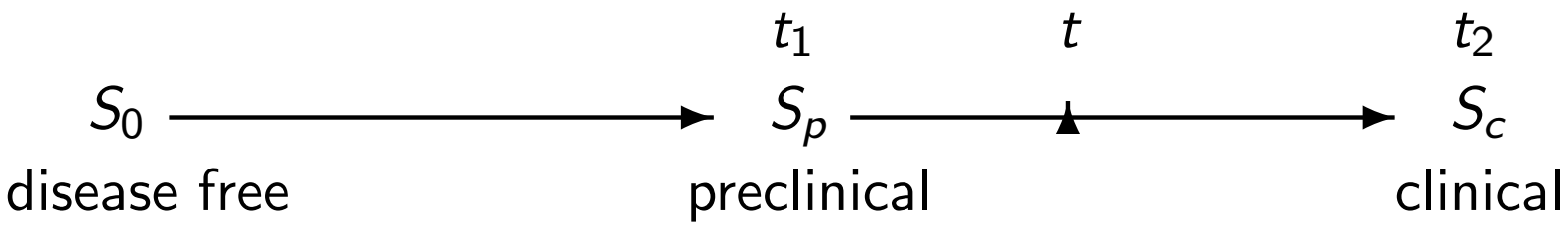
- Let $t_0 < t_1 < \cdots < t_{k-1} < t_k$: k ordered screening exam times.
- n_i: the number of individuals examined at t_{i-1}
- s_i: **screening detected cases** at the exam given at t_{i-1}
- r_i: **interval cases**, the number of cases found in the clinical state (S_c) within (t_{i-1}, t_i).
- (n_i, s_i, r_i): data stratified by initial **age** in the i-th interval.

Table 2: The NLST - CT group data

Age	n_1	s_1	r_1	n_2	s_2	r_2	n_3	s_3	r_3
.....									
60	1946	16	3	1847	13	1	1797	17	0
61	1786	18	0	1678	14	1	1659	11	3
62	1548	11	1	1452	8	2	1408	12	0
63	1427	14	1	1350	6	2	1320	11	0
64	1352	17	0	1287	18	72	1240	11	3
.....									

The progressive model

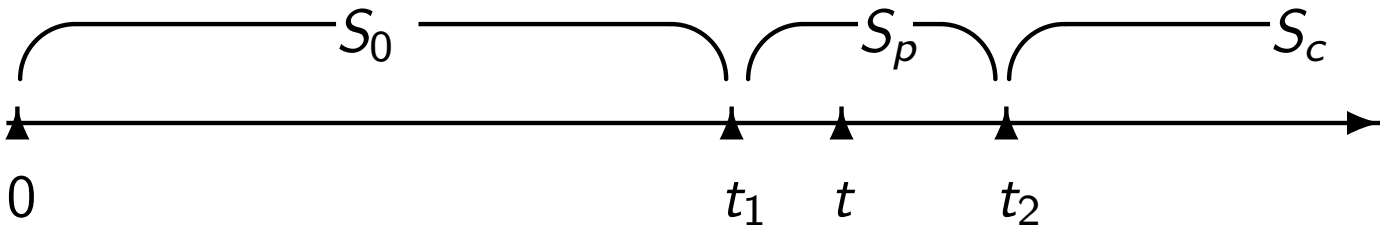
- The progressive disease model assumes that all clinical cancer will go through three states (Zelen and Feinleib in 1969):



- S_0 is the **disease-free** state or the state in which the disease can not be detected.
- S_p is the **preclinical** state, in which an asymptomatic individual unknowingly has the disease that a screening exam can detect.
- S_c is the **clinical** state at which the disease manifests itself in clinical symptoms.

Sojourn time, transition probability and lead time

- Illustration of disease progression and the lead time:

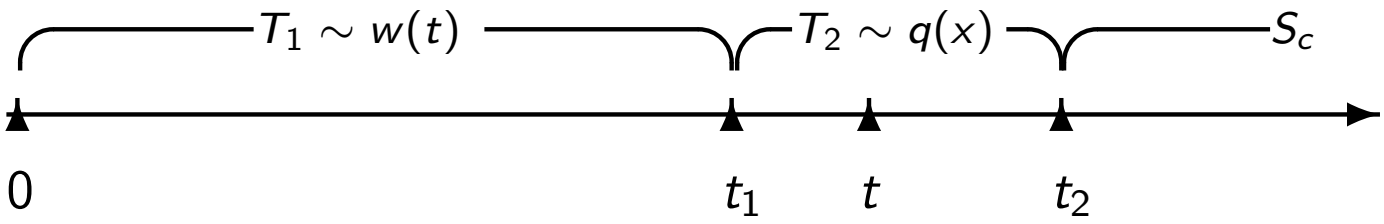


Let t_i represent a person's age.

- **sojourn time**: $(t_2 - t_1)$, the time duration in the preclinical state.
- **transition probability density**: measures the time duration in the disease free state, ie. the distribution of t_1.
- **lead time**: $(t_2 - t)$, the time interval between the diagnosis time t and the onset of S_c if not screened, ie. the length of time the diagnosis is advanced by screening.

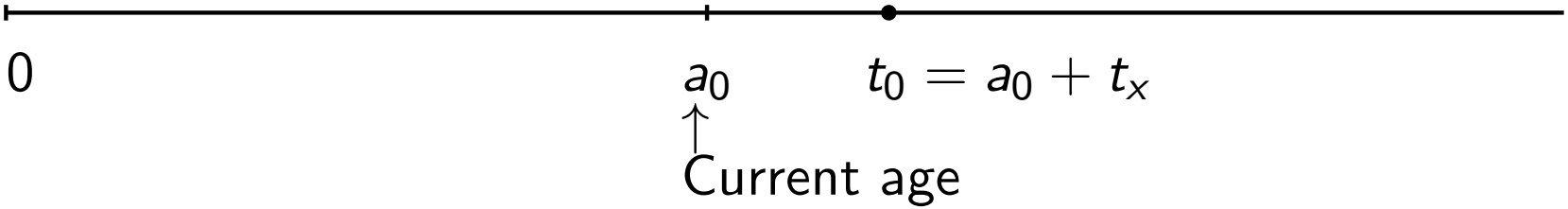
The three key parameters

- sensitivity $\beta = P(X = 1|D = 1) = \beta(s|T)$, where s is the length of time that one has stayed in S_P, and T is the total sojourn time in S_P, a random variable.



- $w(t)$: Probabilty Density Function (PDF) of the time spent in the disease-free state S_0.
- $q(x)$: PDF of the sojourn time (time duration in the preclinical state S_p).
- $Q(z) = Pr(T > z) = \int_z^\infty q(x)dx$, survival function of the sojourn time.
- The three key parameters: $\beta(\cdot)$, $w(\cdot)$, and $q(\cdot)$ (or $Q(\cdot)$). Any other term/probability can be expressed as a function of these three key parameters.

Define events



Suppose a woman at her current age a_0 is asymptomatic, and she has not taken any screening. Let $t_0 = a_0 + t_x$ be her first exam age. Define events:

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

The three mutually exclusive events (I_0, D_0, A_0) is a partition of the sample space:

$$I_0 \cup D_0 \cup A_0 = H_0.$$

Probability of incidence

Since most people won't have cancer, we are more concerned with those who are at risk before the first screening exam. The conditional probability of incidence before the first exam among *people at risk* (I_0 or D_0) is:

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

We need to calculate the probabilities: $P(I_0)$ and $P(D_0)$.

Probability formula

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

$$P(D_0) = P(X < t_0, X + Y > t_0, \beta = \beta(t_0 - X|Y)) = \int_0^{t_0} w(x) \int_{t_0 - x}^{\infty} q(t)\beta(t_0 - x|t)dtdx. \quad (2)$$

The optimal scheduling time

This probability of incidence, $P(I_0|I_0 \cup D_0)$, is *monotonically increasing* as the upcoming screening time interval t_x increases. Therefore, for any given $p \in (0, 1)$, there exists a unique numerical solution t_0, that satisfies

$$P(I_0|I_0 \cup D_0) = p. \tag{3}$$

That is, with probability $(1 - p)$, she will NOT be a clinical incidence case before her first exam at her age t_0. One may choose $p = 0.05$ or 0.10, or, any risk level that s/he is comfortable with.

Lead time and overdiagnosis

After t_0 is found, we can make inferences if one were diagnosed with cancer at t_0:

- Derive the lead time distribution at t_0.
- Derive the probability formula of overdiagnosis and true-early-detection at t_0.

These provide predictive information.

Lead time distribution

We let L be the lead time, the diagnosis time that is advanced by screening, then the probability density function (PDF) of the lead time given one will be diagnosed at t_0 for the first time is

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

Where the denominator $P(D_0)$ is the same as in equation (2); and

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

The validity of this probability density function can be verified by

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

Probability of overdiagnosis and true-early-detection

We first let the lifetime T be a fixed value, then let it be random. Given one would be diagnosed at t_0 and a fixed lifetime $T = t(> t_0)$, the probability of overdiagnosis and true-early-detection are:

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

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

where $P(D_0|T = t) = P(D_0)$ as in equation (2).

Probability of overD and trueED

$$P(\text{OverD}, D_0|T=t) = \int_0^{t_0} w(x) \int_{t-x}^{\infty} q(y)\beta(t_0-x|y)dydx.$$

$$P(\text{TrueED}, D_0|T=t) = \int_0^{t_0} w(x) \int_{t_0-x}^{t-x} q(y)\beta(t_0-x|y)dydx.$$

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

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

Overdiagnosis and true-early-detection: T is random

Now we allow human lifetime T to be random, Then,

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

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

where the conditional PDF of human lifetime $f_T(t|T > t_0) = \frac{f_T(t)}{1-F_T(t_0)}$, if $t > t_0$. It is derived from the actuarial life table: US Social Security Administration, http://www.ssa.gov/OACT/STATS/table4c6.html (Wu et al 2012). We can prove that

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

Simulation set up

- 4 probability of incidence before first screening: $p = 0.05, 0.10, 0.15, 0.20$;
- Three different screening sensitivities: $\beta_i(s|Y), i = 1, 2, 3$;
- Three different mean sojourn time (MST): 2, 5 and 10 years;
- Two different transition mode for $w(t)$: 65 and 69 years;
- Three different current age a_0: 55, 60 and 65 years.

Simulation set up - more details

- Sensitivity:

$$\beta(s|Y) = [1 + \exp(-b_0 - b_1 \cdot \frac{s}{Y})]^{-1}, \quad 0 \leq s \leq Y; \tag{6}$$

We chose $(b_0, b_1) = (0.85, 2.65), (1.40, 2.10), (2.20, 1.30)$.

- Sojourn time: Weibull

$$Q(x|\lambda, \alpha) = \exp(-\lambda x^{\alpha}), \quad \lambda > 0, \alpha > 0; \tag{7}$$

We chose $\alpha = 2.5, \lambda = 0.13109, 0.01326, 0.00234$, with correspond mean sojourn time 2, 5, and 10 years.

- Transition PDF: logNormal

$$w(t|\mu, \sigma^2) = \frac{0.3}{\sqrt{2\pi}\sigma t} \exp\left\{-(\log t - \mu)^2/(2\sigma^2)\right\}, \sigma > 0. \tag{8}$$

We chose $(\mu, \sigma^2) = (4.25, 0.02)$, which will have a mode around 69, to mimic lung cancer, and $(\mu, \sigma^2) = (4.35, 0.175)$, which will havea mode around 65, to mimic breast cancer.

Table 3: Optimal scheduling time t_0

when $(\mu, \sigma^2) = (4.25, 0.02)$, or mode of $w(t)$ around 69

	MST = 2 years								
	$a_0 = 55$			$a_0 = 60$			$a_0 = 65$		
p	[a] β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
0.05	55.11	55.11	55.12	60.10	60.10	60.11	65.10	65.10	65.10
0.10	55.23	55.24	55.25	60.21	60.22	60.23	65.20	65.21	65.22
0.15	55.37	55.39	55.40	60.34	60.36	60.37	65.32	65.33	65.34
0.20	55.54	55.56	55.58	60.49	60.51	60.53	65.45	65.47	65.49
	MST = 5 years								
	$a_0 = 55$			$a_0 = 60$			$a_0 = 65$		
p	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
0.05	55.39	55.41	55.43	60.32	60.34	60.35	65.28	65.29	65.30
0.10	55.86	55.91	55.96	60.69	60.73	60.76	65.59	65.61	65.63
0.15	56.44	56.52	56.60	61.13	61.18	61.23	65.94	65.97	66.01
0.20	57.15	57.27	57.39	61.63	61.71	61.78	66.34	66.39	66.44
	MST = 10 years								
	$a_0 = 55$			$a_0 = 60$			$a_0 = 65$		
p	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
0.05	56.52	56.63	56.74	61.05	61.11	61.17	65.78	65.81	65.85
0.10	58.59	58.84	59.09	62.32	62.46	62.59	66.66	66.74	66.82
0.15	61.03	61.38	61.72	63.81	64.02	64.23	67.66	67.79	67.91
0.20	63.50	63.89	64.26	65.47	65.74	65.99	68.77	68.94	69.10

[a] $\beta_i = \beta_i(s|T) = [1 + \exp(-b_0 - b_1 \cdot \frac{s}{Y})]^{-1}, 0 \leq s \leq T$; where the values of (b_0, b_1) equals to (0.85, 2.65), (1.40, 2.10), (2.20, 1.30) for $\beta_1, \beta_2, \beta_3$ respectively.

Table 4: Scheduling time t_0

when $(\mu, \sigma^2) = (4.35, 0.175)$, or mode of $w(t)$ around 65

MST = 2 years									
	$a_0 = 55$			$a_0 = 60$			$a_0 = 65$		
p	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
0.05	55.09	55.10	55.10	60.09	60.09	60.10	65.09	65.09	65.10
0.10	55.19	55.20	55.21	60.19	60.20	60.21	65.19	65.20	65.20
0.15	55.31	55.32	55.33	60.31	60.32	60.33	65.30	65.31	65.33
0.20	55.44	55.45	55.47	60.43	60.45	60.46	65.43	65.45	65.46
MST = 5 years									
	$a_0 = 55$			$a_0 = 60$			$a_0 = 65$		
p	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
0.05	55.24	55.25	55.26	60.24	60.24	60.25	65.23	65.24	65.25
0.10	55.51	55.53	55.55	60.50	60.52	60.54	65.49	65.51	65.52
0.15	55.81	55.84	55.88	60.79	60.82	60.85	65.78	65.81	65.83
0.20	56.16	56.20	56.24	61.12	61.17	61.21	66.10	66.14	66.18
MST = 10 years									
	$a_0 = 55$			$a_0 = 60$			$a_0 = 65$		
p	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
0.05	55.54	55.56	55.58	60.51	60.53	60.55	65.49	65.50	65.52
0.10	56.14	56.19	56.23	61.07	61.11	61.16	66.02	66.06	66.10
0.15	56.82	56.89	56.97	61.71	61.77	61.84	66.63	66.69	66.75
0.20	57.59	57.69	57.80	62.43	62.52	62.61	67.31	67.39	67.48

Summary of the simulated scheduling time

- The transition density $w(t)$ will affect the first screening time/age.
- Mean Sojourn Time (MST) plays an important role in the timing of the first exam. a longer MST (slow-growing cancer or low-risk people) means one can wait a long time to take the first exam.
- A higher probability of incidence p means a longer screening interval.
- The sensitivity functions β_i slightly affect the first screening time if all other conditions are the same.
- A person's age obviously plays a role in the scheduling: older people should come back for their first exam sooner than their younger counterparts

Figure 1: Lead time density when mode of $w(t)$is 69

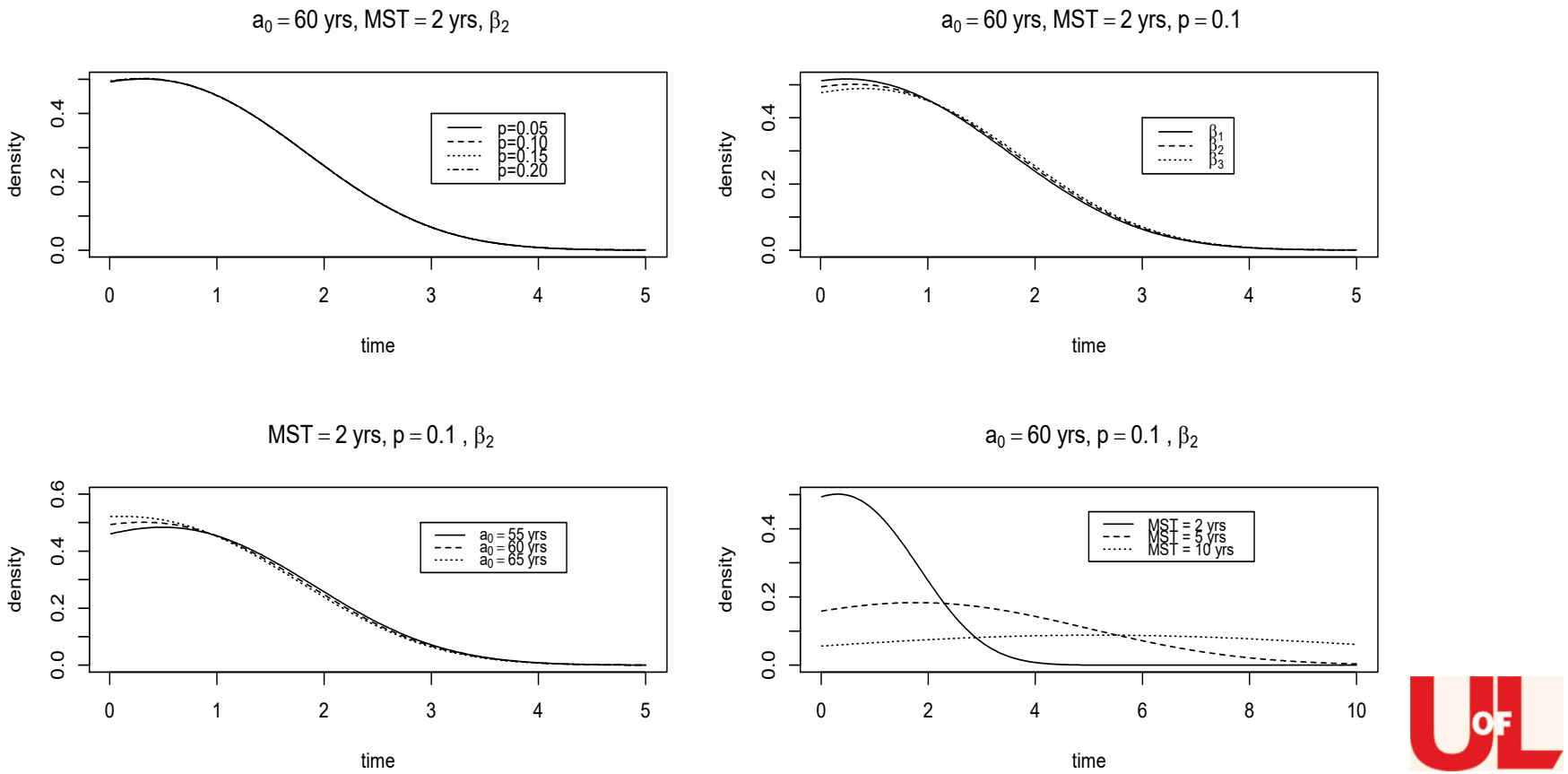
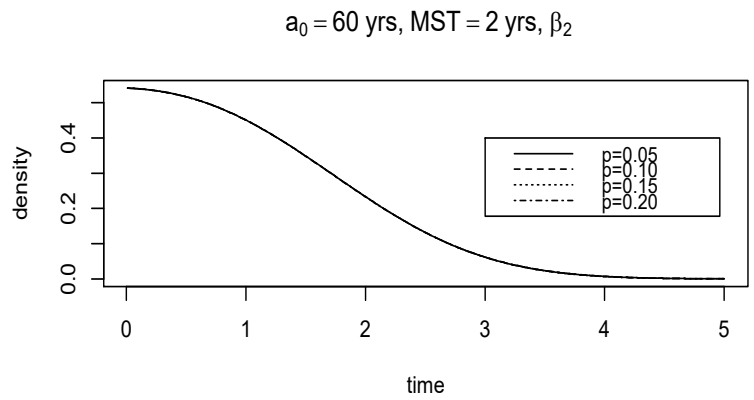
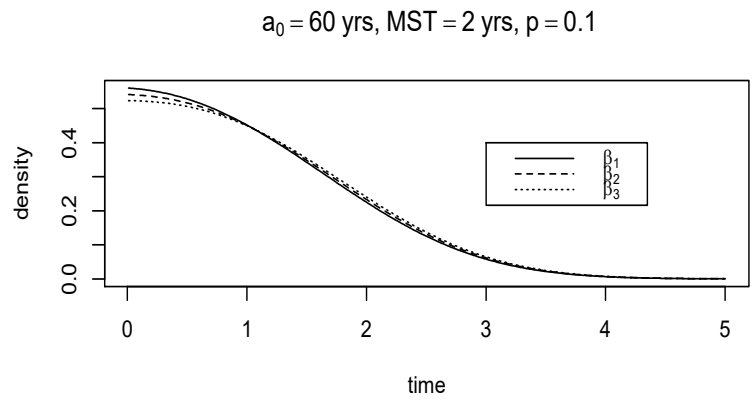
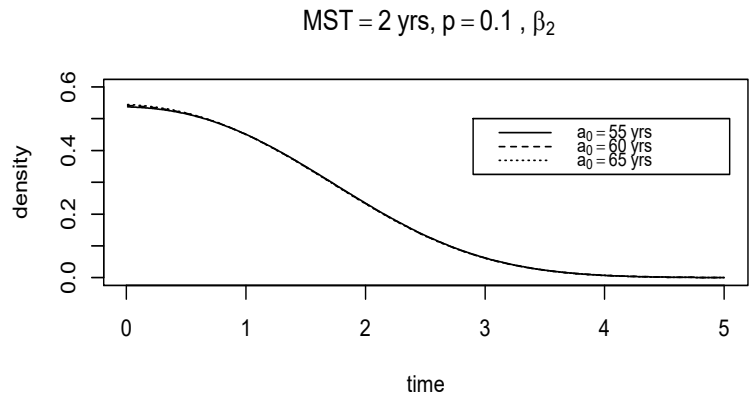
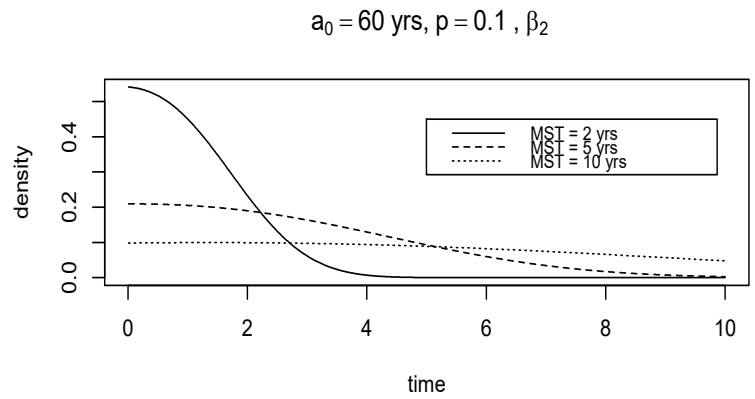


Figure 2: Lead time density when mode of $w(t)$is 65









Summary of lead time distribution at the optimal t_0

- The distribution of lead time changes MORE with the mean sojourn time (MST); a longer MST means larger mean/median/mode, and larger standard deviation of the lead time.
- It changes slightly with the screening sensitivity β_i at the optimal scheduling time t_0.
- It changes slightly with one's current age a_0 when the mode of $w(t)$ is 69; and it barely changes with a_0 when the mode of $w(t)$ is 65.
- It barely changes with the incidence probability p.

Table 5: Probability of overdiagnosis (in %) at the t_0

when $(\mu, \sigma^2) = (4.25, 0.02)$, or mode of $w(t)$ around 69

	MST = 2 years								
	$a_0 = 55$			$a_0 = 60$			$a_0 = 65$		
p	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
0.05	0.595	0.610	0.623	0.820	0.840	0.858	1.148	1.176	1.208
0.10	0.598	0.613	0.627	0.824	0.845	0.863	1.161	1.190	1.216
0.15	0.602	0.617	0.635	0.829	0.850	0.868	1.169	1.198	1.224
0.20	0.611	0.626	0.640	0.834	0.860	0.879	1.178	1.208	1.234
	MST = 5 years								
	$a_0 = 55$			$a_0 = 60$			$a_0 = 65$		
p	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
0.05	1.864	1.919	1.956	2.481	2.539	2.590	3.431	3.514	3.587
0.10	1.924	1.981	2.019	2.527	2.603	2.655	3.508	3.619	3.695
0.15	1.988	2.046	2.098	2.606	2.667	2.737	3.624	3.711	3.813
0.20	2.070	2.128	2.181	2.687	2.769	2.825	3.727	3.818	3.926
	MST = 10 years								
	$a_0 = 55$			$a_0 = 60$			$a_0 = 65$		
p	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
0.05	5.535	5.686	5.819	7.186	7.398	7.541	9.769	10.080	10.291
0.10	6.192	6.444	6.640	7.790	8.026	8.238	10.399	10.732	11.037
0.15	7.186	7.488	7.830	8.590	8.915	9.216	11.160	11.521	11.938
0.20	8.362	8.797	9.217	9.566	10.009	10.366	12.076	12.562	12.921

Table 6: Probability of overdiagnosis (in %) at the t_0

when $(\mu, \sigma^2) = (4.35, 0.175)$, or mode of $w(t)$ around 65

MST = 2 years									
	$a_0 = 55$			$a_0 = 60$			$a_0 = 65$		
p	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
0.05	0.553	0.567	0.579	0.784	0.803	0.820	1.126	1.154	1.179
0.10	0.557	0.573	0.586	0.788	0.807	0.829	1.133	1.161	1.193
0.15	0.564	0.578	0.590	0.794	0.816	0.834	1.148	1.176	1.202
0.20	0.568	0.582	0.595	0.802	0.822	0.8840	1.157	1.186	1.212
MST = 5 years									
	$a_0 = 55$			$a_0 = 60$			$a_0 = 65$		
p	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
0.05	1.582	1.619	1.653	2.213	2.266	2.314	3.224	3.302	3.372
0.10	1.615	1.653	1.688	2.245	2.313	2.362	3.279	3.383	3.456
0.15	1.652	1.692	1.727	2.294	2.366	2.416	3.365	3.475	3.550
0.20	1.686	1.736	1.772	2.359	2.416	2.482	3.478	3.564	3.642
MST = 10 years									
	$a_0 = 55$			$a_0 = 60$			$a_0 = 65$		
p	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
0.05	3.960	4.061	4.150	5.571	5.716	5.845	8.168	8.443	8.638
0.10	4.123	4.229	4.347	5.786	5.974	6.112	8.586	8.816	9.083
0.15	4.320	4.431	4.560	6.089	6.252	6.441	9.028	9.274	9.562
0.20	4.525	4.672	4.808	6.420	6.638	6.841	9.577	9.843	10.151

Summary of over-diagnosis at the t_0

- Mean sojourn time (MST) plays the most important role in overdiagnosis: a longer MST means a larger probability of overdiagnosis.
- The probability of overdiagnosis increases faster as one's current age increases.
- It will slightly increase as the screening sensitivity β increases.
- When p increases from 0.05 to 0.20, the probability of over-diagnosis slightly increases.

Application to the NLST-CT Data

- The method that we derived are functions of the three key parameters: $\beta(\cdot), q(\cdot), w(\cdot)$, so we need to extract this information from the NLST-CT data first.
- Wu, Rai & Seow (2022) developed statistical inference procedures to estimate the sensitivity as a function of sojourn time and time in the preclinical state, using the NLST-CT data for male and female heavy smokers separately.
- We used the same parametric functions as in equations (6) to (8).
- The distribution of the life span $f_T(t)$ was derived from the period life table, Social Security Administration. http://www.ssa.gov/OACT/STATS/table4c6.html

Application to the NLST-CT Data - Details

- Let $\theta = (b_0, b_1, \mu, \sigma^2, \lambda, \alpha)$. Using the likelihood function and Markov Chain Monte Carlo (MCMC), 6000 iterations were generated, after 1000 burn-in, and thinning every 50 iterations, a posterior sample of 100 from each chain was obtained, we ran 8 parallel chains wiht over-dispersed starting values, and obtained 800 Bayesian posterior samples $\theta_j^*, j = 1, 2, \ldots, 800$ for each gender (Wu et al 2022).
- We conducted Bayesian inference using hypothetical cohorts with current age $a_0 = 50, 60, 70$; and incidence probability $p = 0.05, 0.10, 0.15, 0.20$.
- For each θ_j^*, and $P(I_0 | I_0 \cup D_0, H_0, \theta_j^*) = p$, a scheduling time $t_j^* (j = 1, 2, \ldots, 800)$ can be found.

Table 7: Posterior mean scheduling time t^* for NLST-CT

	MALE					
	$a_0 = 50$		$a_0 = 60$		$a_0 = 70$	
p	mean (s.e.)	95% C.I.	mean (s.e.)	95% C.I.	mean (s.e.)	95% C.I.
0.05	50.09 (0.012)	(50.06, 50.11)	60.08 (0.011)	(60.06, 60.10)	70.07 (0.010)	(70.05, 70.09)
0.10	50.19 (0.027)	(50.13, 50.24)	60.16 (0.023)	(60.12,,60.20)	70.14 (0.021)	(70.11, 70.19)
0.15	50.31 (0.045)	(50.22, 50.39)	60.26 (0.037)	(60.19, 60.33)	70.23 (0.033)	(70.17, 70.30)
0.20	50.44 (0.067)	(50.32, 50.58)	60.36 (0.053)	(60.27, 60.47)	70.33 (0.047)	(70.25, 70.42)
	FEMALE					
	$a_0 = 50$		$a_0 = 60$		$a_0 = 70$	
p	mean (s.e.)	95% C.I.	mean (s.e.)	95% C.I.	mean (s.e.)	95% C.I.
0.05	50.09 (0.016)	(50.06, 50.12)	60.08 (0.014)	(60.05, 60.10)	70.07 (0.012)	(70.05, 70.10)
0.10	50.20 (0.036)	(50.13, 50.26)	60.17 (0.029)	(60.12, 60.22)	70.15 (0.026)	(70.11, 70.20)
0.15	50.32 (0.059)	(50.21, 50.43)	60.27 (0.047)	(60.18, 60.36)	70.24 (0.041)	(70.17, 70.32)
0.20	50.46 (0.088)	(50.30, 50.62)	60.38 (0.067)	(60.26, 60.51)	70.34 (0.058)	(70.25, 70.46)

Table 8: Lead time summary at t^* using the NLST-CT

		MALE	
p	$a_0 = 50$	$a_0 = 60$	$a_0 = 70$
0.05	0.94, 0.75, 0.53, 0.63	0.89, 0.71, 0.28, 0.64	0.86, 0.68, 0.01, 0.63
0.10	0.94, 0.75, 0.53, 0.63	0.89, 0.71, 0.28, 0.64	0.86, 0.68, 0.01, 0.63
0.15	0.94, 0.75, 0.53, 0.63	0.89, 0.71, 0.28, 0.64	0.86, 0.68, 0.01, 0.63
0.20	0.94, 0.75, 0.52, 0.63	0.89, 0.71, 0.28, 0.64	0.86, 0.68, 0.01, 0.63
[a]95% C.I.	(0, 1.78)	(0, 1.78)	(0, 1.77)
		FEMALE	
p	$a_0 = 50$	$a_0 = 60$	$a_0 = 70$
0.05	0.92, 0.76, 0.68, 0.59	0.88, 0.72, 0.45, 0.59	0.85, 0.70, 0.01, 0.59
0.10	0.92, 0.76, 0.68, 0.59	0.88, 0.72, 0.44, 0.59	0.85, 0.70, 0.01, 0.59
0.15	0.92, 0.76, 0.67, 0.59	0.87, 0.72, 0.44, 0.59	0.85, 0.70, 0.01, 0.59
0.20	0.92, 0.76, 0.67, 0.59	0.87, 0.72, 0.44, 0.59	0.85, 0.70, 0.01, 0.59
95% C.I.	(0, 1.72)	(0, 1.71)	(0, 1.70)

[a]the 95% C.I. is the 95% highest probability density (HPD) interval using the Bayesian empirical method. Since the lead time curve for different p are almost the same, we list the largest interval for different p if there is a small discrepancy.

Table 9: Probability of overdiagnosis at t^* using NLST-CT

	MALE					
	$a_0 = 50$		$a_0 = 60$		$a_0 = 70$	
p	mean (s.e.)	95% C.I.	mean (s.e.)	95% C.I.	mean (s.e.)	95% C.I.
0.05	0.464 (0.107)	(0.270, 0.659)	0.990 (0.224)	(0.578, 1.391)	1.892 (0.428)	(1.176, 2.737)
0.10	0.467 (0.108)	(0.276, 0.669)	0.995 (0.226)	(0.579, 1.405)	1.902 (0.432)	(1.179, 2.754)
0.15	0.470 (0.110)	(0.265, 0.665)	1.000 (0.228)	(0.580, 1.413)	1.912 (0.438)	(1.181, 2.773)
0.20	0.476 (0.112)	(0.273, 0.678)	1.007 (0.232)	(0.582, 1.423)	1.925 (0.444)	(1.186, 2.794)
	FEMALE					
	$a_0 = 50$		$a_0 = 60$		$a_0 = 70$	
p	mean (s.e.)	95% C.I.	mean (s.e.)	95% C.I.	mean (s.e.)	95% C.I.
0.05	0.284 (0.069)	(0.169, 0.414)	0.579 (0.137)	(0.349, 0.836)	1.248 (0.297)	(0.751, 1.803)
0.10	0.286 (0.070)	(0.169, 0.417)	0.582 (0.138)	(0.350, 0.841)	1.256 (0.300)	(0.753, 1.821)
0.15	0.288 (0.071)	(0.170, 0.423)	0.585 (0.140)	(0.350, 0.846)	1.265 (0.305)	(0.754, 1.844)
0.20	0.291 (0.073)	(0.171, 0.430)	0.589 (0.142)	(0.348, 0.854)	1.276 (0.311)	(0.757, 1.861)

Note: report the posterior mean probability and 95% HPD credible interval in percentage.

Summary of NLST-CT application

- The scheduling time t^* is very close for both genders in heavy smokers under similar conditions.
- The older heavy smokers should come back earlier for the first exam.
- The lead time changes with one's current age for both genders; And female heavy smokers usually have a slightly shorter mean lead time than their male counterparts at the first exam.
- The probability of overdiagnosis is very low at the first exam for heavy smokers. It slightly increases with one's current age for both genders; it is slightly higher for male heavy smokers.
- The probability of overdiagnosis slightly increases when p increases. However, the maximum probability of overdiagnosis was less than 2% for both genders. In summary, it is not a big issue to use low-dose CT in lung cancer screening.

Conclusion

- This is an extension of the original probability method to dynamically schedule one's first screening exam, based on one's current age, risk tolerance, and other parameters.
- The major improvement is that the sensitivity is a function of the ratio of time one spent in the preclinical state relative to the total sojourn time.
- The method can provide predictive information on the lead time and overdiagnosis if one were diagnosed with cancer in the future time. This may be the first step towards a personalized screening schedule in the near future.
- The modeling approach is just one way of thinking about the problem. Other models and approaches are possible. A key point is to recognize that screening has outcomes & consequences that one should consider, especially for policy purposes.

References and acknowledgement

- Wu D, Rai SN and Seow A. (2022). Estimation of preclinical state onset age and sojourn time for heavy smokers in lung cancer. *Statistics and Its Interface*. Vol. 15, No. 3, 349-358. NIHMS ID: 1734205.
- Wu D, Kafadar K, Rosner GL, and Broemeling LD. (2012). The lead time distribution when lifetime is subject to competing risks in cancer screening. *The International Journal of Biostatistics* **6** DOI: 10.1515/1557-4679.1363.
- Wu D. (2022). When to initiate cancer screening exam? *Statistics and Its Interface*. Vol. 15, No. 4, 503-514.
- Wu D. (2022). When to initiate a screening exam if sensitivity is a function of sojourn time? Draft.

This research was partially supported by NIH/NCI 1R15CA242482.