Dynamic Scheduling of the Upcoming Exam in Cancer Screening

Dongfeng Wu[1] and Karen Kafadar[2]

[1]Department of Bioinformatics and Biostatistics
University of Louisville

[2]Department of Statistics
University of Virginia

ICHPS, January 12, 2018

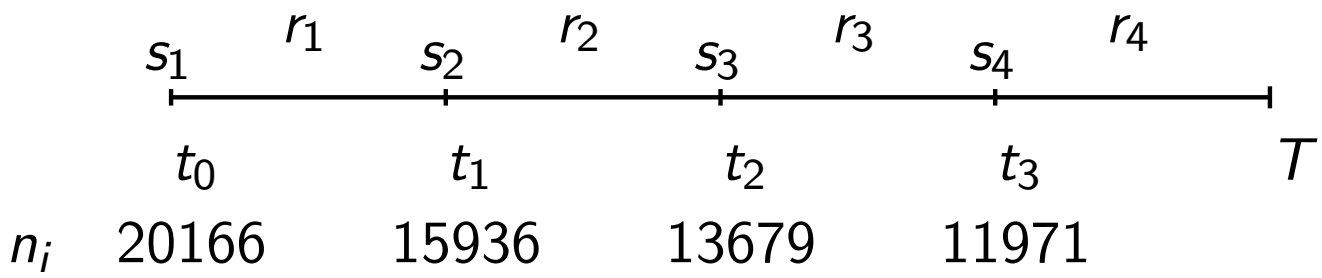
Motivation

- Efficient design of cancer screening is not available until now: at what age to start screening? and how frequently?
- For an individual who has gone through a few screening exams in the past and get negative results so far, when should s/he come back for the next exam?
- Some research has been done to study the problem of a fixed budget that allows only n exams in a fixed age interval, using some utility functions (Zelen 1993, Lee and Zelen 1998). The problem: utility functions are subjective, and the methods cannot be applied directly by diagnostic radiologists.
- We will use conditional probability of incidence before the next exam, to control the risk of incidence.
- The method can be applied to any kinds of screening; it will be applied to the women's breast cancer using the Health Insurance Plan for Greater New York (HIP) data.

The HIP data

- The Health Insurance Plan of Greater New York (HIP) study, which began at the end of 1963, was the first randomized clinical trial to examine mammogram screenings for breast cancer.
- Asymptomatic women without a history of breast cancer, with initial age: $40-64$, average age $= 51.2$, and with 15 years of follow-up.
- About 60,000 participants were equally randomized into 2 arms: Study and Control.
- Study group: mammogram + clinical exam in each screening, and 4 annual screenings.
- Control group: usual care without screening.

The HIP Study Group Data



- $t_0 < t_1 < \cdots < t_{k-1} < t_k$: 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.

The HIP Data

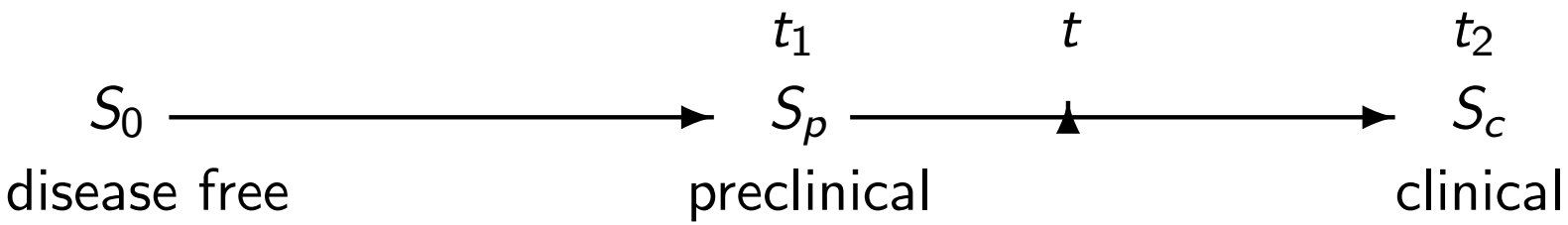
Table 1. HIP study group data: 4 annual screenings

Age	n_1	s_1	r_1	n_2	s_2	r_2	n_3	s_3	r_3	n_4	s_4	r_4
.....												
45	985	1	0	850	1	0	782	2	2	687	3	1
46	1014	3	0	887	2	1	833	1	0	744	0	0
47	933	2	1	808	2	0	747	0	2	658	1	0
48	978	1	0	849	0	0	777	0	0	683	2	0
49	915	1	0	817	0	0	765	1	0	677	2	1
.....												

Note: a characteristic of periodic cancer screening is that data were collected repeatedly.

The progressive model

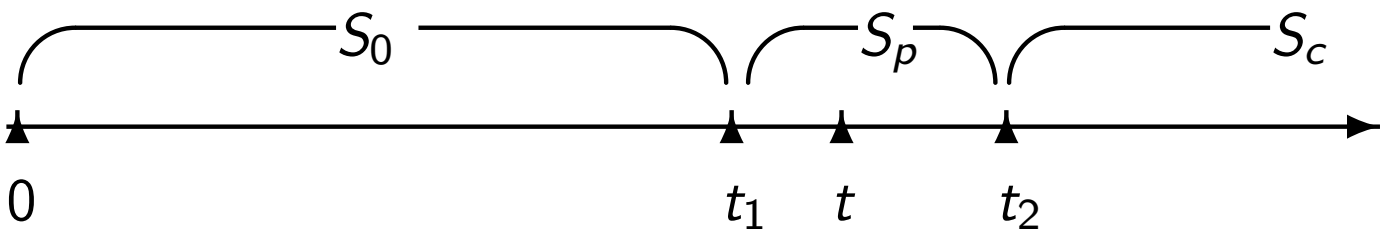
- The progressive disease model assumes that all clinical cancer will go through three states (Zelen and Feinleib in 1969):



- 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:

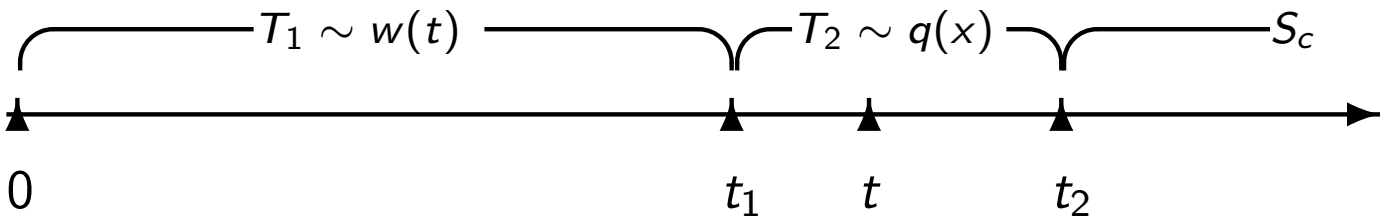


Let t_i represent a person's age.

- **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.

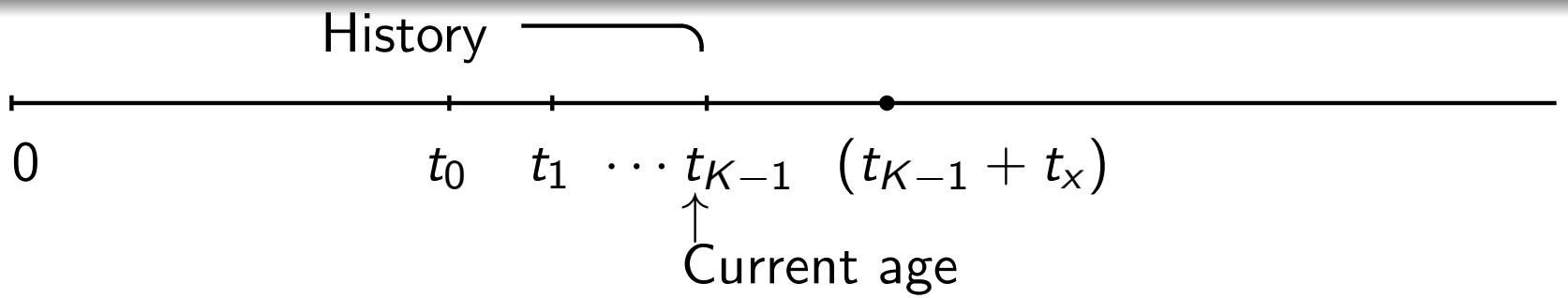
The three key parameters

- sensitivity at age t: $\beta(t) = P(X = 1|D = 1, t)$.



- $w(t)$: 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).
- $Q(z) = Pr(T > z) = \int_z^\infty q(x)dx$, survivor function of the sojourn time.
- The three key parameters: sensitivity $\beta(t)$, transition density $w(t)$, and sojourn time distribution $q(x)$. Any other term/probability is a function of the three key parameters.

Define events



Suppose one has taken $K =$ exams at ages $t_0 < t_1 < \cdots < t_{K-1}$ and is asymptomatic at current age t_{K-1}. Define events:

$$
\begin{aligned}
H_K &= \left\{ \begin{array}{l} \text{one is asymptomatic in } [0, t_{K-1}] \text{ after} \\ \text{taking } K \text{ exams at ages } t_0 < t_1 < \cdots < t_{K-1} \end{array} \right\}, \\
I_K &= \{ \text{one will be a clinical incident case first time in } (t_{K-1}, t_{K-1} + t_x)\}, \\
D_K &= \{ \text{one will be diagnosed at } (t_{K-1} + t_x) \text{ for the first time}\}. \\
A_K &= \{ \text{one will be asymptomatic in } (t_{K-1}, t_{K-1} + t_x]\},
\end{aligned}
$$

The mutually exclusive events (I_K, D_K, A_K) is a partition of the whole sample space:

$$I_K \cup D_K \cup A_K = \Omega.$$

Probability of incidence

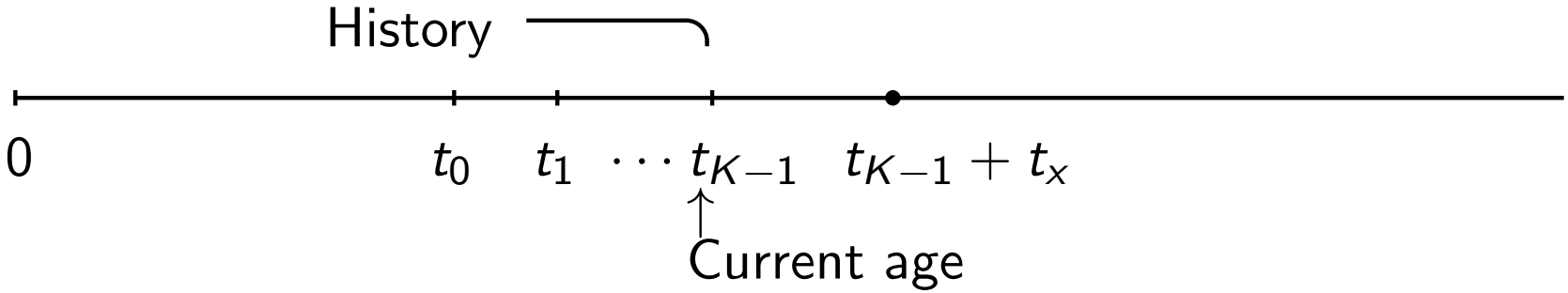
Let $t_K = t_{K-1} + t_x$, the probability of incidence before the next screening exam among *people at risk* (I_K or D_K) is:

$$P(I_K|I_K \cup D_K, H_K) = \frac{P(I_K|H_K)}{P(I_K \cup D_K|H_K)} = \frac{P(I_K|H_K)}{1 - P(A_K|H_K)} \quad (1)$$

$$= \frac{P(I_K \cap H_K)/P(H_K)}{1 - P(A_K \cap H_K)/P(H_K)} = \frac{P(I_K \cap H_K)}{P(H_K) - P(A_K \cap H_K)}$$

We need to calculate the probabilities: $P(H_K), P(I_K \cap H_K)$ and $P(A_K \cap H_K)$.

Probability calculation



Let $t_{-1} = 0$, then

$$
\begin{aligned}
P(H_K) &= P(\text{ remained in } S_0 \text{ in } (0, t_{K-1})) \\
&+ P(\text{ entered and remained in } S_p \text{ and not being detected }) \\
&= 1 - \int_0^{t_{K-1}} w(x)dx \qquad (2) \\
&+ \sum_{i=0}^{K-1} (1-\beta_i) \cdots (1-\beta_{K-1}) \int_{t_{i-1}}^{t_i} w(x) Q(t_{K-1} - x)dx
\end{aligned}
$$

Probability calculation - cont.

$$P(I_K \cap H_K) = \sum_{i=0}^{K-1}(1-\beta_i)\cdots(1-\beta_{K-1})\int_{t_{i-1}}^{t_i} w(x)[Q(t_{K-1}-x)-Q(t_K-x)]dx + \int_{t_{K-1}}^{t_K} w(x)[1-Q(t_K-x)]dx, \quad (3)$$

And

$$P(A_K \cap H_K) = 1-\int_0^{t_K} w(x)dx + \sum_{j=0}^{K}(1-\beta_j)\cdots(1-\beta_K)\int_{t_{j-1}}^{t_j} w(x)Q(t_K-x)dx, \quad (4)$$

And in fact $P(A_K \cap H_K) = P(H_{K+1})$.

Find the optimal scheduling time

This probability of incidence, $P(I_K|I_K \cup D_K, H_K)$, is monotonically increasing as the upcoming screening time interval t_x increases. Therefore, for any pre-selected small value α, there exists a unique numerical solution t^*, that satisfies

$$P(I_K|I_K \cup D_K, H_K) = \alpha. \tag{5}$$

That is, with probability $(1-\alpha)$, she will NOT be a clinical cancer case before her next screening exam at her age $(t_{K-1} + t^*)$, where t_{K-1} is her current age.
One may choose α at a risk level that she is comfortable with, such as 0.05 or 0.10 (5% or 10%).

Lead time and overdiagnosis

After t^* is found (based on one's screenig history and other parameters), we can make more inferences if one were diagnosed with cancer at $(t_{K-1} + t^*)$:

- Calculate the conditional distribution of the lead time at $(t_{K-1} + t^*)$.
- Calculate the probability of overdiagnosis (and true-early-detection) at $(t_{K-1} + t^*)$.

These provide predictive information.

Lead time distribution

We let L be the lead time (the time diagnosis is advanced by screening), and let $t_K = t_{K-1} + t^*$, then the conditional probability density function (PDF) of the lead time given the event D_K (i.e., one will be diagnosed at t_K for the first time) is

$$f_L(z|D_K) = \frac{f_L(z, D_K)}{P(D_K)}, \qquad \text{for } z \in (0, \infty). \tag{6}$$

Where the denominator

$$\begin{aligned} P(D_K) &= \beta_K \left\{ \sum_{i=0}^{K-1} (1-\beta_i) \cdots (1-\beta_{K-1}) \int_{t_{i-1}}^{t_i} w(x) Q(t_K - x) dx \right. \\ &\quad \left. + \int_{t_{K-1}}^{t_K} w(x) Q(t_K - x) dx \right\}. \end{aligned} \tag{7}$$

Lead time distribution - cont.

And the numerator

$$f_L(z, D_K) = \sum_{i=0}^{K} f_L(z, D_K, \text{and she is from the } i\text{-th generation})$$

$$= \beta_K \left\{ \sum_{i=0}^{K-1} (1-\beta_i)\cdots(1-\beta_{K-1}) \int_{t_{i-1}}^{t_i} w(x) q(t_K + z - x) dx \right.$$

$$\left. + \int_{t_{K-1}}^{t_K} w(x) q(t_K + z - x) dx \right\}. \tag{8}$$

This is a valid probability density function (pdf), since

$$\int_0^{\infty} f_L(z|D_K) dz = 1. \tag{9}$$

Probability of overdiagnosis and true-early-detection

To calculate the probability, We first let the lifetime T to be a fixed value, then let it to be random. Given one would be diagnosed at $t_K = t_{k-1} + t^*$ and one's fixed lifetime $T = t(> t_K)$, the probability of overdiagnosis and true-early-detection can be derived:

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

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

where $P(D_K|T = t) = P(D_K)$ as in equation (7).

Probability of overD and trueED

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

$$= \beta_K \left\{ \sum_{i=0}^{K-1} (1-\beta_i) \cdots (1-\beta_{K-1}) \int_{t_{i-1}}^{t_i} w(x) Q(t-x) dx \right.$$

$$\left. + \int_{t_{K-1}}^{t_K} w(x) Q(t-x) dx \right\}. \quad (10)$$

$$P(\text{TrueED}, D_K | T = t)$$

$$= \beta_K \left\{ \sum_{i=0}^{K-1} (1-\beta_i) \cdots (1-\beta_{K-1}) \int_{t_{i-1}}^{t_i} w(x)[Q(t_K-x) - Q(t-x)] dx \right.$$

$$\left. + \int_{t_{K-1}}^{t_K} w(x)[Q(t_K-x) - Q(t-x)] dx \right\}. \quad (11)$$

And it can be verified that

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

Probability of Overdiagnosis - cont.

Now we allow human lifetime T to be random, and let $f_T(t|T > t_k)$ be the conditional PDF of the lifetime T, derived from the actuarial life table (US Social Security Administration, http://www.ssa.gov/OACT/STATS/table4c6.html), then

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

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

Where

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

Social Security

The Official Website of the U.S. Social Security Administration

Actuarial Life Table

Office of the Chief Actuary

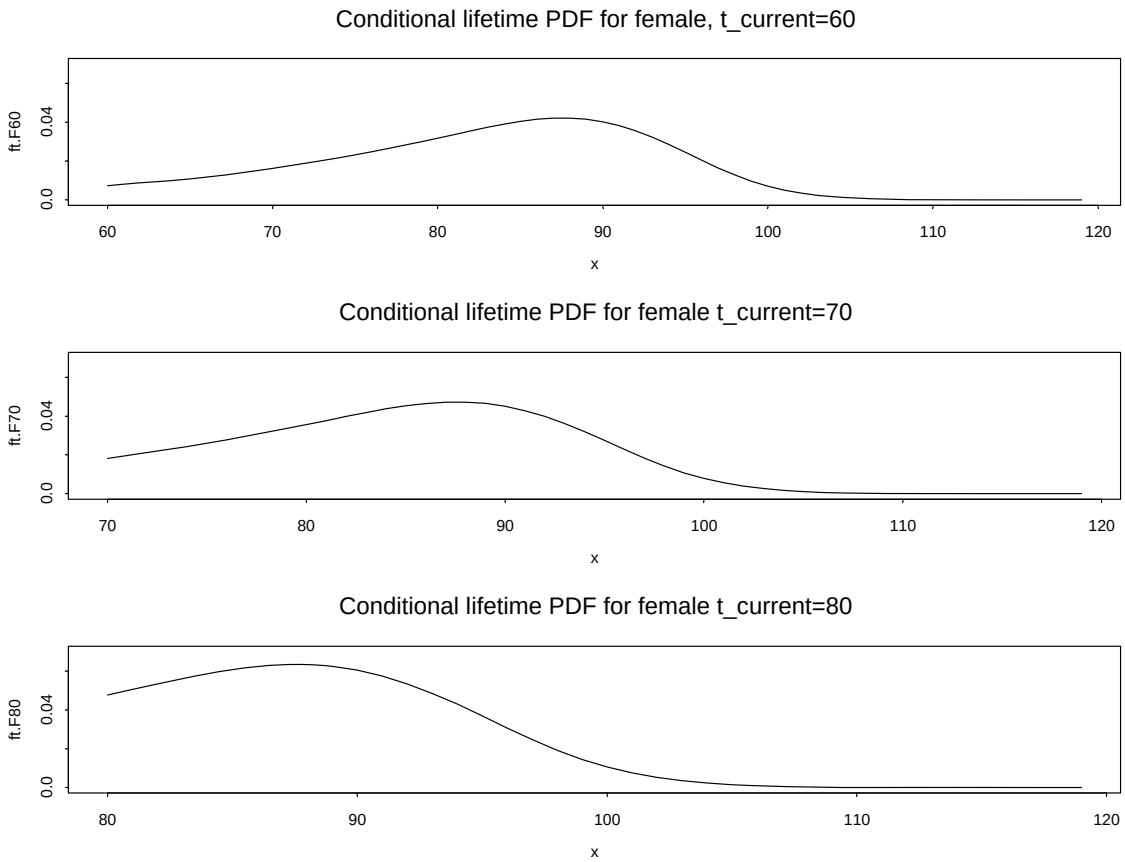
Life Tables

A period life table is based on the mortality experience of a population during a relatively short period of time. Here we present the 2007 period life table for the Social Security area population. For this table, the period life expectancy at a given age represents the average number of years of life remaining if a group of persons at that age were to experience the mortality rates for 2007 over the course of their remaining life.

Period Life Table, 2007

Exact age	Male			Female		
	Death probability [a]	Number of lives [b]	Life expectancy	Death probability [a]	Number of lives [b]	Life expectancy
0	0.007379	100,000	75.38	0.006096	100,000	80.43
1	0.000494	99,262	74.94	0.000434	99,390	79.92
2	0.000317	99,213	73.98	0.000256	99,347	78.95
3	0.000241	99,182	73.00	0.000192	99,322	77.97
4	0.000200	99,158	72.02	0.000148	99,303	76.99
5	0.000179	99,138	71.03	0.000136	99,288	76.00
6	0.000166	99,120	70.04	0.000128	99,275	75.01
7	0.000152	99,104	69.05	0.000122	99,262	74.02
8	0.000133	99,089	68.06	0.000115	99,250	73.03
9	0.000108	99,075	67.07	0.000106	99,238	72.04

Figure 1: The conditional PDF of lifetime for females when $t_K = 60, 70, 80$



Simulation set up

- Historic screening interval $\Delta = 1, 1.5, 2$ years, from $t_0 = 50$ to currrent age $t_{K-1} = 62$.
- Screening sensitivity $\beta = 0.7, 0.8, 0.9$.
- Sojourn time: Log logistic distribution

$$Q(x) = [1 + (x\rho)^{\kappa}]^{-1}, \quad \kappa > 0, \rho > 0.$$

  Let mean sojourn time be 2, 5, 10 and 15 years, i.e., $\kappa = 2.5, \rho = 0.661, 0.264, 0.132, 0.088$.
- Transition density

$$w(t|\mu, \sigma^2) = \frac{0.2}{\sqrt{2\pi}\sigma t} \exp\left\{-(\log t - \mu)^2/(2\sigma^2)\right\}, \sigma > 0.$$

  with mode around 60 years old, i.e. $\mu = 4.2, \sigma^2 = 0.1$.
- Probability (risk) of incidence before next screening: $\alpha = 0.05, 0.1, 0.15, 0.2$.

Table 2: Simulated optimal screening time t^*

MST = 2 years

	Beta = 0.7			Beta = 0.8			Beta = 0.9		
α \ [a]Δ	1.0	1.5	2.0	1.0	1.5	2.0	1.0	1.5	2.0
0.05	0.13	0.10	0.09	0.21	0.14	0.11	0.47	0.31	0.23
0.1	0.33	0.24	0.21	0.56	0.40	0.33	0.84	0.75	0.69
0.15	0.58	0.45	0.39	0.86	0.72	0.64	1.12	1.05	1.01
0.2	0.83	0.69	0.62	1.11	1.01	0.94	1.36	1.31	1.28

MST = 5 years

	Beta = 0.7			Beta = 0.8			Beta = 0.9		
α \ [a]Δ	1.0	1.5	2.0	1.0	1.5	2.0	1.0	1.5	2.0
0.05	0.87	0.53	0.38	1.31	0.98	0.69	1.67	1.52	1.35
0.1	1.65	1.28	1.00	2.07	1.84	1.61	2.42	2.31	2.21
0.15	2.25	1.94	1.67	2.67	2.48	2.30	3.03	2.95	2.86
0.2	2.78	2.52	2.28	3.22	3.06	2.91	3.60	3.53	3.46

MST = 10 years

	Beta = 0.7			Beta = 0.8			Beta = 0.9		
α \ [a]Δ	1.0	1.5	2.0	1.0	1.5	2.0	1.0	1.5	2.0
0.05	2.69	2.22	1.74	3.19	2.92	2.61	3.59	3.47	3.34
0.1	4.03	3.66	3.29	4.56	4.35	4.13	5.00	4.91	4.82
0.15	5.11	4.79	4.47	5.69	5.51	5.32	6.18	6.10	6.02
0.2	6.09	5.81	5.53	6.73	6.57	6.40	7.29	7.21	7.14

MST=15 years

	Beta = 0.7			Beta = 0.8			Beta = 0.9		
α \ [a]Δ	1.0	1.5	2.0	1.0	1.5	2.0	1.0	1.5	2.0
0.05	4.43	4.02	3.56	5.00	4.77	4.50	5.46	5.36	5.25
0.1	6.35	6.02	5.66	6.99	6.81	6.61	7.54	7.47	7.38
0.15	7.91	7.61	7.30	8.63	8.46	8.29	9.27	9.20	9.12
0.2	9.33	9.06	8.78	10.14	9.99	9.83	10.87	10.81	10.74

[a]Δ is the historical screening interval (in years) in the corresponding row.
α = probability of incidence before the next screening exam in the 1st column.
MST = mean sojourn time.

Summary of the simulated optimal scheduling time

- Sensitivity will affect the next screening time interval in a positive way: higher sensitivity means longer time interval to maintain the same incidence risk α.
- Mean Sojourn Time (MST) affects the next screening interval in a positive way: long MST (slow growing cancer or low risk people) means she can wait longer time to carry out the next screening.
- Screening history plays an important role: shorter screening interval in the past means longer screening interval for the upcoming test, and vise versa.
- Lower probability of incidence α, means shorter screening interval.

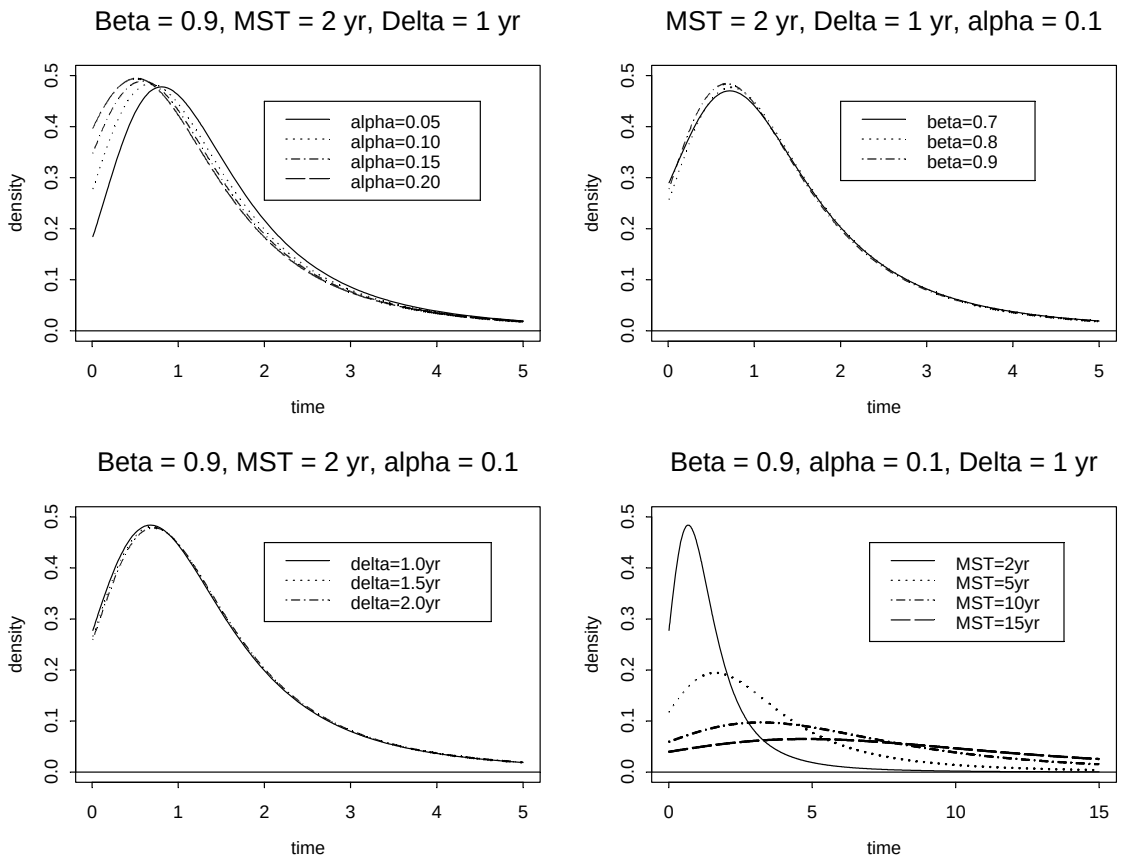


Figure 2: Lead time distribution using t^* under the four factors: α, β, Δ and MST.

Summary of the lead time distribution under the optimal t^*

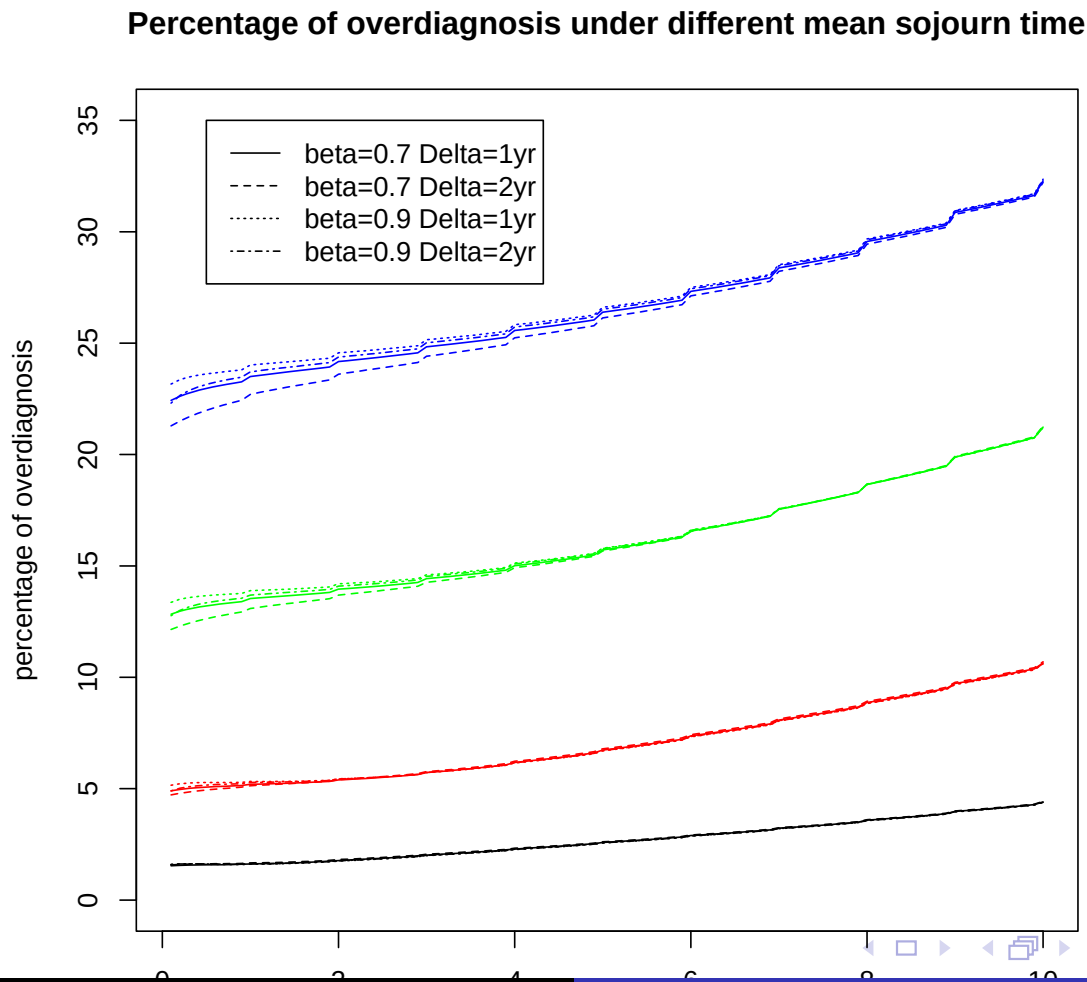
The density curve of the lead time changes MORE with the risk α and the mean sojurn time (MST), and it changes LESS with the screening sensitivity β and the past screening interval Δ if the optimal scheduling time t^* is adopted.

- Smaller α means larger mean/median/mode of the lead time, and smalller standard deviation of the lead time.
- Longer MST means larger mean/median/mode, and larger standard deviation of the lead time.

Table 3: Estimated probability of overdiagnosis (in percentage) using the t^*

[a] $MST = 2$ **yrs**									
	$\beta = 0.7$			$\beta = 0.8$			$\beta = 0.9$		
[b] $\alpha \backslash \Delta$	1.0	1.5	2.0	1.0	1.5	2.0	1.0	1.5	2.0
0.05	1.55	1.55	1.57	1.58	1.57	1.57	1.62	1.62	1.63
0.10	1.58	1.58	1.60	1.60	1.60	1.62	1.61	1.61	1.62
0.15	1.59	1.60	1.62	1.61	1.61	1.63	1.62	1.62	1.63
0.20	1.59	1.62	1.64	1.62	1.63	1.64	1.64	1.64	1.64
$MST = 5$ **yrs**									
	$\beta = 0.7$			$\beta = 0.8$			$\beta = 0.9$		
$\alpha \backslash \Delta$	1.0	1.5	2.0	1.0	1.5	2.0	1.0	1.5	2.0
0.05	5.17	5.01	4.90	5.29	5.21	5.11	5.38	5.34	5.31
0.10	5.31	5.21	5.13	5.43	5.37	5.31	5.53	5.50	5.48
0.15	5.46	5.38	5.31	5.60	5.54	5.49	5.73	5.71	5.68
0.20	5.63	5.55	5.49	5.80	5.75	5.70	5.96	5.93	5.91
$MST = 10$ **yrs**									
	$\beta = 0.7$			$\beta = 0.8$			$\beta = 0.9$		
$\alpha \backslash \Delta$	1.0	1.5	2.0	1.0	1.5	2.0	1.0	1.5	2.0
0.05	14.27	13.91	13.54	14.61	14.41	14.19	14.88	14.79	14.70
0.10	15.03	14.72	14.44	15.43	15.26	15.08	15.80	15.71	15.63
0.15	15.80	15.52	15.25	16.31	16.14	15.97	16.76	16.69	16.62
0.20	16.64	16.37	16.11	17.27	17.10	16.94	17.86	17.77	17.70
$MST = 15$ **yrs**									
	$\beta = 0.7$			$\beta = 0.8$			$\beta = 0.9$		
$\alpha \backslash \Delta$	1.0	1.5	2.0	1.0	1.5	2.0	1.0	1.5	2.0
0.05	25.90	25.41	24.86	26.51	26.22	25.90	27.00	26.87	26.74
0.10	27.67	27.24	26.77	28.45	28.19	27.92	29.12	29.02	28.90
0.15	29.45	29.01	28.57	30.42	30.16	29.90	31.32	31.21	31.09
0.20	31.31	30.90	30.47	32.53	32.29	32.03	33.67	33.56	33.44

Figure 3: Percentage of overdiagnosis vs. next screening time t



Summary of the over-diagnosis under the t^*

- Mean sojourn time (MST) plays the most important role in overdiagnosis: longer MST means larger probability of overdiagnosis.
- Probability of overdiagnosis will slightly increase as the screening sensitivity β increases.
- For the same fixed MST, the probability of overdiagnosis won't change much as the risk α increases.
- We plot the probability of overdiagnosis versus future scheduling time in Figure 3. It shows that the probability of overdiagnosis is monotonic increasing as t^* increases, and it won't change much with sensitivity β and past screening interval Δ for a fixed MST.

Application to the HIP Data

- All methods that we derived before, are functions of the three key parameters: $\beta(t), q(x), w(t)$, so we need to extract these information from the HIP data first.
- Wu, Rosner & Broemeling (2005) developed statistical inference procedures to estimate the sojourn time q(x), the age-dependent sensitivity $\beta(t)$, and the age-dependent transition density $w(t)$ using parametric link and likelihood function.
- Sensitivity: $\beta(t) = [1 + \exp(-b_0 - b_1 * (t - m))]^{-1}$.
- Transition density $S_0 \to S_p$: 0.2 * lognormal pdf

$$w(t|\mu, \sigma^2) = \frac{0.2}{\sqrt{2\pi}\sigma t} \exp\left\{-(\log t - \mu)^2/(2\sigma^2)\right\}, \sigma > 0.$$

- Sojourn time: Log logistic distribution

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

Application to the HIP Data

Let HIP = HIP data, and $\theta = (b_0, b_1, \mu, \sigma^2, \kappa, \rho)$. Using the likelihood function and Markov Chain Monte Carlo (MCMC), 2000 Bayesian posterior samples (θ_j^*) were generated; for details, see Wu et al. (2005).
Using each θ_j^*, and $P(I_K | I_K \cup D_K, H_K, \theta_j^*) = \alpha$, we conducted Bayesian inference on hypothetical cohorts of asymptomatic women with current age $t_{K-1} = 62$ (or 72), assuming that they have started their first screening at age $t_0 = 50$ (or 60), and with different screening intervals $\Delta = 1, 2$ and 3 years in the 12 years.

Table 4: Optimal scheduling time using the HIP data

α	$\Delta = 1.0$	$\Delta = 2.0$	$\Delta = 3.0$
	$t_0 = 50, t_{K-1} = 62$		
0.05	0.30 (0.19)	0.23 (0.15)	0.22 (0.15)
0.10	0.56 (0.31)	0.45 (0.19)	0.43 (0.19)
0.15	0.78 (0.42)	0.67 (0.24)	0.64 (0.21)
0.20	0.99 (0.52)	0.88 (0.32)	0.85 (0.25)
	$t_0 = 60, t_{K-1} = 72$		
0.05	0.42 (0.40)	0.32 (0.25)	0.30 (0.20)
0.10	0.68 (0.55)	0.59 (0.40)	0.55 (0.31)
0.15	0.91 (0.67)	0.82 (0.53)	0.78 (0.43)
0.20	1.11 (0.79)	1.04 (0.65)	1.00 (0.55)

Table 5: Estimated mean/median/mode/standard deviation of the lead time at t^* using HIP data

α	$\Delta = 1.0$	$\Delta = 2.0$	$\Delta = 3.0$
	$t_0 = 50, t_{K-1} = 62$		
0.05	1.84,0.97,0.49(3.42)	1.88,0.95,0.47(3.61)	1.92,0.95,0.46(3.77)
0.10	1.80,0.92,0.42(3.42)	1.85,0.93,0.41(3.58)	1.89,0.93,0.41(3.72)
0.15	1.78,0.89,0.35(3.43)	1.83,0.90,0.35(3.58)	1.87,0.91,0.35(3.70)
0.20	1.77,0.87,0.30(3.46)	1.82,0.88,0.30(3.59)	1.85,0.89,0.30(3.70)
	$t_0 = 60, t_{K-1} = 72$		
0.05	1.83,0.96,0.48(3.38)	1.87,0.96,0.47(3.53)	1.90,0.96,0.46(3.67)
0.10	1.78,0.91,0.40(3.39)	1.83,0.92,0.40(3.52)	1.86,0.93,0.40(3.63)
0.15	1.77,0.88,0.34(3.41)	1.81,0.89,0.34(3.52)	1.84,0.90,0.34(3.62)
0.20	1.76,0.86,0.28(3.44)	1.79,0.87,0.29(3.54)	1.82,0.88,0.29(3.63)

Table 6: Estimated mean probability of overdiagnosis with 95% CI(in percentage) at t^* using HIP data

α	$\Delta = 1.0$	$\Delta = 2.0$	$\Delta = 3.0$
	$t_0 = 50, t_{K-1} = 62$		
0.05	2.25(0.57,10.97)	2.38(0.56,12.47)	2.52(0.56,13.80)
0.10	2.26(0.53,11.35)	2.37(0.54,12.61)	2.49(0.54,13.65)
0.15	2.29(0.52,11.37)	2.39(0.52,12.84)	2.50(0.52,13.67)
0.20	2.34(0.51,11.54)	2.43(0.51,12.91)	2.52(0.51,13.80)
	$t_0 = 60, t_{K-1} = 72$		
0.05	4.65(1.40,19.98)	4.76(1.40,20.48)	4.89(1.40,21.45)
0.10	4.68(1.33,20.37)	4.78(1.33,21.39)	4.89(1.33,21.42)
0.15	4.76(1.30,20.76)	4.84(1.31,21.86)	4.94(1.31,21.73)
0.20	4.87(1.30,21.49)	4.93(1.30,22.07)	5.02(1.31,22.57)

Summary

- The next screening time is slightly decreasing if the past screening interval increases from 1 year to 3 years; it follows the same pattern as in our simulation study.
- The mean/median/mode of the lead time is decreasing as the risk α increases.
- The mean/median/mode of the lead time is increasing as the past screening interval Δ increases if other factors are the same.
- The probability of overdiagnosis is low in the HIP study (2-5%), although it slightly increasing with age, past screening interval Δ.

Conclusion

- We developed a probability method to dynamically schedule one's upcoming screening exam, based on one's past screen history, risk tolerance α, screening sensitivity, sojourn time distribution, etc.
- The method can handle any screening history $t_0 < t_1 < ...t_{K-1}$, including unequally-spaced screening intervals.
- Simulations show that longer screening interval in the past means shorter interval for the upcoming test.
- We provide predictive information on the lead time and overdiagnosis if one were diagnosed with cancer at the future time. This may be the first step towards 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. The important point is to recognize that screening has outcomes & consequences that one should consider, especially for policy purposes.

Thank You!!

Any questions or comments??

The actuarial life table of SSA

- The distribution of the lifetime $f_T(t)$ was derived from the actuarial life table, Social Security Administration. http://www.ssa.gov/OACT/STATS/table4c6.html
- The period life table is based on population mortality from all Social Security area, including 50 states, DC, and surrounding islands of the US.
- It provides the conditional probability of death within one year given one's current age $P(T < N+1 | T \geq N)$, from age $N = 0$ to 119.
- Let $a_N = P(T \geq N+1 | T \geq N) = 1 - P(T < N+1 | T \geq N)$.

Transfer the life table to a valid pdf

- For any integer age t_0,

$$P(T \geq t_0 + N | T \geq t_0) = \prod_{i=1}^{N} a_{t_0+i-1}, \quad \forall N = 1, 2, \ldots, 120 - t_0.$$

- Using a density approximation, we have

$$f_T(t_0 + N | T \geq t_0) = \lim_{\epsilon \to 0} \frac{P(t_0 + N < T \leq t_0 + N + \epsilon | T \geq t_0)}{\epsilon}$$

$$\approx \quad P(t_0 + N < T \leq t_0 + N + 1 | T \geq t_0) = (1 - a_{t_0+N}) \prod_{i=1}^{N} a_{t_0+i-1}.$$

- Finally, for any $t \in (N, N+1)$ $(N < 120)$, we use a step function to approximate: $f_T(t | T \geq t_0) \approx f_T(N | T \geq t_0)$.
- It is a valid pdf because $\sum_{N=0}^{120-t_0} f_T(t_0 + N | T \geq t_0) = 1$.

Likelihood function and probability of detection at kth exam

$$L = \prod_{t_0=40}^{64} \prod_{k=1}^{4} D_{k,t_0}^{s_{k,t_0}} I_{k,t_0}^{r_{k,t_0}} (1 - D_{k,t_0} - I_{k,t_0})^{n_{k,t_0} - s_{k,t_0} - r_{k,t_0}}$$

$$D_{1,t_0} = \beta(t_0) \int_0^{t_0} w(x) Q(t_0 - x) dx.$$

$$D_{k,t_0} = \beta(t_{k-1}) \left\{ \sum_{i=0}^{k-2} [1 - \beta(t_i)] \cdots [1 - \beta(t_{k-2})] \int_{t_{i-1}}^{t_i} w(x) Q(t_{k-1} - x) dx \right.$$
$$\left. + \int_{t_{k-2}}^{t_{k-1}} w(x) Q(t_{k-1} - x) dx \right\}, \qquad \text{for } \forall k = 2, \cdots, K.$$

Probability of incidence within the kth interval

$$
\begin{aligned}
I_{k,t_0} &= \sum_{i=0}^{k-1} [1-\beta(t_i)] \cdots [1-\beta(t_{k-1})] \\
&\quad \times \int_{t_{i-1}}^{t_i} w(x)[Q(t_{k-1}-x) - Q(t_k - x)]dx \\
&\quad + \int_{t_{k-1}}^{t_k} w(x)[1 - Q(t_k - x)]dx, \\
&\qquad\qquad \text{for } \forall k = 1, \cdots, K.
\end{aligned}
$$