The International Journal of Biostatistics

Volume 8, *Issue* 1 2012 *Article* 6

The Lead Time Distribution When Lifetime is Subject to Competing Risks in Cancer Screening

Dongfeng Wu, *University of Louisville*
Karen Kafadar, *Indiana University, Bloomington*
Gary L. Rosner, *Sidney Kimmel Comprehensive Cancer Center*
Lyle D. Broemeling, *Broemeling and Associates Inc.*

Recommended Citation:
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*: Vol. 8: Iss. 1, Article 6.
DOI: 10.1515/1557-4679.1363

The Lead Time Distribution When Lifetime is Subject to Competing Risks in Cancer Screening

Dongfeng Wu, Karen Kafadar, Gary L. Rosner, and Lyle D. Broemeling

Abstract

This paper extends the previous probability model for the distribution of lead time in periodic cancer screening exams, namely, in that the lifetime T is treated as a random variable, instead of a fixed value. Hence the number of screens for a given individual is a random variable as well. We use the actuarial life table from the Social Security Administration to obtain the lifetime distribution, and then use this information to project the lead time distribution for someone with a future screening schedule. Simulation studies using the HIP study group data provide estimates of the lead time under different screening frequencies. The projected lead time has two components: a point mass at zero (corresponding to interval cases detected between screening exams) and a continuous probability density. We present estimates of the projected lead time for participants in a breast cancer screening program. The model is more realistic and can inform optimal screening frequency. This study focuses on breast cancer screening, but is applicable to other kinds of cancer screening also.

KEYWORDS: lead time, sensitivity, sojourn time, transition probability density, lifetime, cancer screening

Author Notes: Corresponding author's email: dongfeng.wu@louisville.edu

1 Introduction

Breast cancer is the most common form of cancer among women in the U.S. and the second leading cause of cancer deaths among women. Breast cancer screening exams (mammography, clinical breast exam, breast self-exam) have been advocated since the 1970s, although opinions differ regarding the age to start screening and the screening frequencies. In November 2009, The U.S. Preventive Services Task Force (USPSTF 2009, Mandelblatt et al., 2009) announced its recommendation statement for breast cancer screening: biennial screening mammography for women aged 50 (versus 40) to 74 years.

The goal of screening is to detect malignant tumors early, which hopefully translates to early treatment and better prognosis. The difference between the time of diagnosis via a screening exam and the time of clinical disease onset without screening is called the lead time. Even in the absence of effective therapy, screening will appear to extend the survival time from diagnosis until death because of lead time. If one does not account for the lead time when analyzing the survival benefit due to screening, then one's inference is subject to lead-time bias. Hence accurate evaluation of the distribution of lead time is important.

We assume the commonly followed disease progressive model where the disease develops by progressing through three states, denoted by $S_0 \rightarrow S_p \rightarrow S_c$ (Zelen and Feinleib, 1969). The state S_0 refers to the disease-free state, where either the person does not have the disease, or the disease is in such an early stage that it cannot be detected by a screening exam. The preclinical disease state, S_p, is a state in which an asymptomatic individual unknowingly has disease that a screening exam can detect. The disease state S_c is a state at which the disease manifests itself with clinical symptoms. If a person enters the preclinical state (S_p) at age t_1 and becomes clinically incident (S_c) later at age $t_2 (> t_1)$, then $(t_2 - t_1)$ is the sojourn time in the preclinical state. If, however, this person undergoes a screening exam at time t within the time interval (t_1, t_2) and cancer is diagnosed, then the length of time $(t_2 - t)$ is the person's lead time.

Many researchers have proposed methods for inference on the lead time among participants in a screening program (Kafadar and Prorok, 1994, 1996, 2003; Kafadar et al., 1998, Walter and Day, 1983, Straatman et al., 1997), usually by providing formulas to estimate the mean and the variance of the lead time. Among these works, Prorok 1982 made a major contribution by deriving the conditional probability distribution of the lead time, given detection at the i-th screening exam. He applied his model in simulations to study the properties of the lead time, assuming different sojourn time distributions,

and found that when one increases the number of exams, keeping the between-exam interval fixed, the local lead time properties are relatively constant after 4 or 5 screening exams in his examples. The stabilization of the local lead time properties suggested a stopping rule for comparative studies, in that further screening exams may not yield more information about the benefit of screening versus no screening. However, he considered only screen-detected cases, ignoring interval cases for which the lead time is zero. His results apply to cases that are screen-detected at the i-th screening exam. He did not estimate the proportion of cases that are not detected by the periodic screening. This group of people does not derive any benefit from screening because their tumors were not diagnosed at scheduled exams; for these "clinical incident" cases, their lead time is zero. For policy purposes, it is important to estimate how large this proportion may be.

Wu et al. (2007) derived the probability distribution of lead time for the whole diseased cohort, including both screen-detected cases and interval-incident cases, when the human lifetime was treated as a fixed value. The model includes Prorok's result as a special case (no interval cases). Hence, the model allows estimation of the proportion of patients whose lead time is zero and the proportion whose lead time is positive from the program. However, the lead time distribution was derived when the lifetime T was a fixed upper bound, which is not realistic.

In this paper, we first review briefly the method for deriving lead time distribution developed by Wu et al. 2007 (Section 2). We then extend the method by allowing the lifetime T, and the number of screens to be random variables. In Section 3, we derive the distribution of the lead time for the whole diseased cohort when the lifetime T is a random variable. We obtain the lifetime distribution by using the actuarial life table from the Social Security Administration in Section 4. Section 5 presents simulation results for different initial screening ages and different screening frequencies by applying the method to the HIP data using Bayesian inference. We conclude with a discussion in Section 6.

2 Existing method: the lead time distribution when T is fixed

Consider a cohort of initially asymptomatic individuals in a screening program. Throughout we use female breast cancer as an example, but the approach applies to screening for other diseases also. Let $\beta(t)$ be the sensitivity of the screening modality, where t is the individual's age at the exam. Define

$w(t)dt$ as the probability of a transition from S_0 to S_p during $(t, t+dt)$. Let $q(\cdot)$ be the probability density function (pdf) of the sojourn time in S_p, and let $Q(z) = \int_z^\infty q(x)dx$ be the survivor function of the sojourn time in the preclinical state S_p. Throughout this paper, the time variable t represents the participating individual's age; the random variable T represents human lifetime. We briefly review the probability model in Wu et al 2007, but with new notation, which allows a clearer presentation for the new model in Section 3.

Consider an initially asymptomatic woman with no history of breast cancer, and suppose she undergoes K screening exams at ages $t_0 < t_1 < \ldots < t_{K-1}$. Let $T = t_K$ denote her (fixed) lifetime, where $t_K > t_{K-1}$. Note that no screening takes place at t_K, but setting $t_K = T$ simplifies the formulae. We let $t_{-1} = 0$ and $\beta_i = \beta(t_i)$ to be the sensitivity at age t_i throughout the paper.

Let D be a binary random variable, with $D = 1$ indicating development of clinical disease and $D = 0$ indicating the absence of the clinical disease before death. Let L denote the lead time for an individual who develops symptomatic cancer. Because the lead time is 0 for individuals whose disease is not detected by the screening exam (e.g., interval cases where clinical symptoms arise between exams), the distribution of lead time L conditional on $T = t_K$, is a mixture of a point mass at 0 and the conditional pdf $f_L(z|D = 1, T = t_K)$.

When the lifetime $T = t_K$ is a fixed value, the distribution of lead time is:

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

$$f_L(z|D = 1, T = t_K) = \frac{f_L(z, D = 1|T = t_K)}{P(D = 1|T = t_K)}, \quad (2)$$

where

$$P(D = 1|T = t_K) = \int_{t_0}^{t_K} \int_0^t w(x)q(t - x)dxdt$$

$$= \int_0^{t_0} w(x)[Q(t_0 - x) - Q(t_K - x)]dx + \int_{t_0}^{t_K} w(x)[1 - Q(t_K - x)]dx. \quad (3)$$

and

$$P(L = 0, D = 1|T = t_K) = I_{K,1} + I_{K,2} + \cdots + I_{K,K}, \quad (4)$$

$$I_{K,j} = \sum_{i=0}^{j-1}(1 - \beta_i) \cdots (1 - \beta_{j-1}) \int_{t_{i-1}}^{t_i} w(x)[Q(t_{j-1} - x) - Q(t_j - x)]dx$$

$$+\int_{t_{j-1}}^{t_j} w(x)[1-Q(t_j-x)]dx, \qquad \text{for all } j=1,\cdots,K, \tag{5}$$

and where

$$f_L(z, D=1|T=t_k) = \beta_0 \int_0^{t_0} w(x)q(t_0+z-x)dx, \quad \text{if } t_K - t_1 < z \le t_K - t_0, \tag{6}$$

or

$$\begin{aligned} & f_L(z, D=1|T=t_K) \\ = & \sum_{i=1}^{j-1} \beta_i \left\{ \sum_{r=0}^{i-1} (1-\beta_r)\cdots(1-\beta_{i-1}) \int_{t_{r-1}}^{t_r} w(x)q(t_i+z-x)dx \right. \\ & \left. + \int_{t_{i-1}}^{t_i} w(x)q(t_i+z-x)dx \right\} + \beta_0 \int_0^{t_0} w(x)q(t_0+z-x)dx. \\ & \text{if } t_K - t_j < z \le t_K - t_{j-1}, j=2,3,\cdots,K. \end{aligned} \tag{7}$$

We note that we add $T=t_K$ in the conditional part in the above equations, to underline K screening exams in one's lifetime when $T=t_K$ is a fixed value.

The validity of this mixed probability distribution can be proved by confirming

$$P(L=0|D=1,T=t_K) + \int_0^{t_K-t_0} f_L(z|D=1,T=t_K)dz = 1. \tag{8}$$

3 The lead time distribution when T is a random variable

We now expand this model in Section 2 to the situation when the lifetime T is a random variable with a probability density function $f_T(t)$. Hence, an individual currently at age t_0 will have a variable number of screening exams $K=k(T)$, a function of her lifetime T. In fact, K is the largest integer such that $t_{K-1} < T$. The distribution of lead time when the lifetime T is greater than t_0 can be obtained by:

$$P(L=0|D=1,T\ge t_0) = \int_{t_0}^{\infty} P(L=0|D=1,T=t)f_T(t|T\ge t_0)dt, \tag{9}$$

$$f_L(z|D=1,T\ge t_0) = \int_{t_0+z}^{\infty} f_L(z|D=1,T=t)f_T(t|T\ge t_0)dt, \tag{10}$$

$$z \in (0, \infty),$$

where the conditional probability $P(L = 0|D = 1, T = t)$ and the conditional pdf $f_L(z|D = 1, T = t)$ were given in Equations (1)-(2) in Section 2; and the conditional pdf $f_T(t|T \geq t_0)$ is

$$f_T(t|T \geq t_0) = \begin{cases} \frac{f_T(t)}{P(T>t_0)} = \frac{f_T(t)}{1-F_T(t_0)}, & \text{if } t \geq t_0, \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

The lower bound for the integration in (10) should be $(t_0 + z)$ instead of t_0, because the lead time z should be less than $t - t_0$ for any fixed lifetime t, hence t should be larger than $t_0 + z$. Equations (9)-(11) yield a valid mixed probability distribution, because

$$\begin{aligned} & P(L = 0|D = 1, T \geq t_0) + \int_0^\infty f_L(z|D = 1, T \geq t_0)dz \\ = & \int_{t_0}^\infty P(L = 0|D = 1, T = t) f_T(t|T \geq t_0)dt \\ & + \int_0^\infty \int_{t_0+z}^\infty f_L(z|D = 1, T = t) f_T(t|T \geq t_0)dtdz \\ = & \int_{t_0}^\infty P(L = 0|D = 1, T = t) f_T(t|T \geq t_0)dt \\ & + \int_{t_0}^\infty \int_0^{t-t_0} f_L(z|D = 1, T = t) f_T(t|T \geq t_0)dzdt \\ = & \int_{t_0}^\infty f_T(t|T \geq t_0)dt = 1. \end{aligned}$$

4 The lifetime distribution $f_T(t|T \geq t_0)$

To obtain reliable information on the lifetime distribution, we used the actuarial life table from the United States Social Security Administration (SSA), which was published and updated regularly at:
`http://ssa.gov/OACT/STATS/table4c6.html`.

This life table is based on the information from all Social Security area populations, including all 50 states, DC, and surrounding islands of the U.S. until 2006. Due to the SSA's calculation method, there is a time lag of 4 years in the life table. We used the version that was reviewed and updated by SSA in April 2010 (the most recent version includes mortality rates until 2007, updated on April 5, 2011, but the changes are very slight). The period life table is based on population mortality; it provides the conditional probability of death within one year from age 0 to age 119, that is, $P(T < N + 1|T \geq N), N = 0, 1, 2, \ldots, 119$.

We denote this probability by $b_N = P(T < N+1|T \geq N)$, as given in the life table. Let $a_N = P(T \geq N+1|T \geq N) = 1 - b_N$. Using the conditional probability formula,

$$\begin{aligned} & P(T \geq N+2|T \geq N) = P(T \geq N+2, T \geq N+1|T \geq N) \\ = \;& P(T \geq N+1|T \geq N)P(T \geq N+2|T \geq N+1, T \geq N) \qquad (12) \\ = \;& a_N a_{N+1} \end{aligned}$$

By mathematical induction, for any integer age t_0,

$$\begin{aligned} P(T \geq t_0+N|T \geq t_0) &= \prod_{i=1}^{N} P(T \geq t_0+i|T \geq t_0+i-1) \\ &= \prod_{i=1}^{N} a_{t_0+i-1}, \quad \forall N = 1,2,\ldots,120-t_0. \qquad (13) \end{aligned}$$

Using a density approximation, we have

$$\begin{aligned} f_T(t = t_0+N|T \geq t_0) &= \lim_{\epsilon \to 0} \frac{P(t_0+N < T \leq t_0+N+\epsilon|T \geq t_0)}{\epsilon} \\ &\approx P(t_0+N < T \leq t_0+N+1|T \geq t_0) \qquad (14) \\ &= P(T \geq t_0+N|T \geq t_0) - P(T \geq t_0+N+1|T \geq t_0) \\ &= (1-a_{t_0+N}) \prod_{i=1}^{N} a_{t_0+i-1}. \qquad (15) \end{aligned}$$

Finally, for any real number $t \in (N, N+1)$ (here $N < 120$), we use a step function to approximate: $f_T(t|T \geq t_0) \approx f_T(N|T \geq t_0)$. We note that this approximation is a valid pdf, as

$$\begin{aligned} \sum_{N=0}^{120-t_0} f_T(t_0+N|T \geq t_0) &= \sum_{N=0}^{120-t_0} [P(T \geq t_0+N|T \geq t_0) \\ &\quad -P(T \geq t_0+N+1|T \geq t_0)] \\ &= P(T \geq t_0|T \geq t_0) = 1 \end{aligned}$$

The conditional density function of the lifetime T for females was plotted in Figure 1 for three initial ages at screening: $t_0 = 40, 50, 60$ (i.e. irrelevance of screening or any specific causes of death).

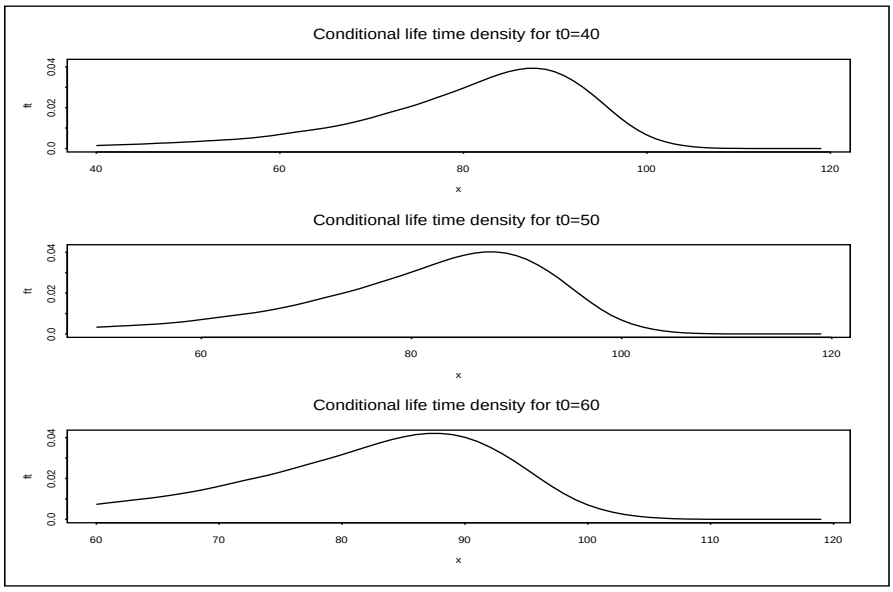

Figure 1: The conditional PDF of the lifetime for females derived from the life table when $t_0 = 40, 50, 60$.

5 Applications and simulations based on the HIP study

Using the results in Section 3, the lead time distribution is a function of the sensitivity $\beta(t)$, the transition probability density $w(t)$, the sojourn time distribution $q(x)$, a person's initial age at screening, and future screening frequency or screening schedule. Thus inference on the lead time distribution requires estimates of $\beta(t), w(t)$ and $q(x)$, which are available from the Health Insurance Plan for Greater New York (HIP) study using likelihoods and Bayesian inference (Wu et al 2005). The parametric model was:

$$\beta(t) = \frac{1}{1+\exp\{-b_0 - b_1(t-m)\}}, \tag{16}$$

$$w(t|\mu,\sigma^2) = \frac{0.2}{\sqrt{2\pi}\sigma t}\exp\left\{-(\log t-\mu)^2/(2\sigma^2)\right\}, \quad \sigma>0, \tag{17}$$

and

$$q(x) = \frac{\kappa x^{\kappa-1}\rho^{\kappa}}{[1+(x\rho)^{\kappa}]^2}, \qquad \kappa>0, \quad \rho>0, \tag{18}$$

where t represents age, m is the average age at study entry, and x is the sojourn time in the preclinical state S_p. The unknown parameters are $\theta = (b_0, b_1, \mu, \sigma^2, \kappa, \rho)$. Markov Chain Monte Carlo (MCMC) was used to draw posterior samples with noninformative priors, with 30,000 iterations, 10,000 burn-in steps, and then sampled every 20 steps. The final 2000 posterior samples for θ came from two parallel chains with overdispersed initial values (`http://www.tibs.org/biometrics` under Paper Information link).

We use the Bayesian posterior samples θ_i^* in the inference for the lead time. The posterior predictive distribution of the lead-time is:

$$\begin{aligned} f_L^{HIP}(l) &= \int f_L^{HIP}(l,\theta)d\theta \\ &= \int f_L^{HIP}(l|\theta) f_\Theta^{HIP}(\theta)d\theta \\ &\approx \frac{1}{n}\sum_{i=1}^{n} f_L^{HIP}(l|\theta_i^*), \end{aligned} \tag{19}$$

where θ_i^* is the posterior sample ($i = 1, \cdots, 2000$) and $f_L^{HIP}(l|\theta_i^*)$ is the mixture distribution defined by Equations (9)-(11).

We assumed that the sensitivity $\beta(t)$, the transition probability $w(t)$, and the sojourn time distribution $q(x)$ are the same today as it was when the HIP study was conducted. The method can be applied to any pre-planned screening schedule, for example, people can take screening annually between 50 and 60, then take it biennually between 60 and 80. For simplicity, we assume three cohorts of initially asymptomatic women, with initial screening age $t_0 =$ 40, 50, and 60. For each cohort, we examined various screening frequencies, with screening interval $\Delta =$12, 18, 24, and 30 months. The number of screenings $K = \lceil (T - t_0)/\Delta \rceil$ is a function of the lifetime T, therefore it is a random variable in the simulation. From Equation (19), the final distribution of the lead time is simply a weighted average of the different lengths of lifetimes.

Table 1 summarizes the Bayesian predictive inference for the lead time. The time interval Δ between screens was 12, 18, 24 and 30 months with initial ages 40, 50, and 60. The probability that the lead time is zero and the corresponding 95% C.I., and the probability that the lead time is positive, and the corresponding standard errors are reported as percentages in Table 1. The mean lead time and its standard error were reported in years. Since the lead time distribution is very skewed, we report also the median and the fourth-spread as more sensible summaries of location and spread. The median of the lead time (when it is positive) is about 0.85 years for all 12 situations; the first quartile of the lead time ranges from 0.35 to 0.45 years, and the third quartile ranges from 1.65 to 1.85 years. The density curves for the lead time

Table 1: A projection of the lead time distribution using posterior samples from the HIP study data

Δ^a	$P_0^b(C.I.)$	1-P_0 (s.e.)	EL^c (s.e.)	Med/IQR^d
	Age at initial screen $t_0 = 40$			
12 mo.	26.29 (16.48, 39.68)	73.71 (5.99)	1.04(1.69)	0.71
18 mo.	39.25 (27.10, 50.72)	60.75 (5.96)	0.87(1.63)	0.61
24 mo.	49.04 (36.19, 59.00)	50.96 (5.95)	0.75(1.58)	0.61
30 mo.	56.34 (42.66, 65.99)	43.66 (5.87)	0.66(1.52)	0.57
	Age at initial screen $t_0 = 50$			
12 mo.	23.76 (12.39, 41.42)	76.24 (7.07)	1.07(1.70)	0.71
18 mo.	36.51 (22.19, 52.25)	63.49 (7.35)	0.91(1.65)	0.61
24 mo.	46.32 (29.37, 59.98)	53.68 (7.35)	0.79(1.60)	0.61
30 mo.	53.71 (35.75, 65.76)	46.29 (7.18)	0.70(1.55)	0.57
	Age at initial screen $t_0 = 60$			
12 mo.	21.98 (8.22, 45.49)	78.02 (8.47)	1.08(1.68)	0.65
18 mo.	34.28 (16.78, 55.15)	65.72 (9.06)	0.92(1.63)	0.65
24 mo.	43.91 (22.88, 61.45)	56.09 (9.08)	0.81(1.59)	0.61
30 mo.	51.21 (29.42, 66.50)	48.79 (8.84)	0.72(1.55)	0.61

[a] $\Delta = t_i - t_{i-1}$ is the time interval between screens.
[b] $P_0 = P(L = 0|D = 1)$ is the probability of "no-early-detection" Columns 2 and 3 are in percentages.
[c]the mean and the standard deviation(S.E.) in column 4 are in years, for the mixture distribution.
[d]the median over the fourth-spread (interquartile range) when the lead time is greater than zero.
This is a simulated projection. the number of screens K is a random variable, changing with the lifetime T.

are shown in Figure 2 for different screening intervals only when $t_0 = 50$, as the density curves when the initial screening age is 40 or 60 are similar.

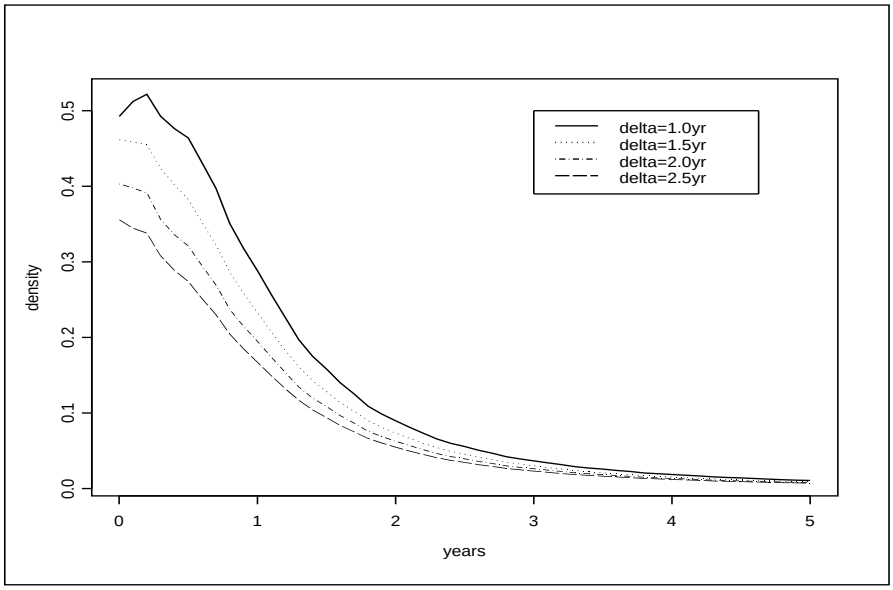

Figure 2: Density Curve for the Lead Time in the HIP Study Using Bayesian Inference with $t_0 = 50$.

These results suggest that a woman who begins annual screening (i.e., $\Delta = 12$ months) when she is 50 years old and develops breast cancer sometime during her life has a 23.76% chance that she will not be detected early by the regular screening exams. This probability of no early detection from the screening program increases to 46.32% if the exams are biennial. For a woman with initial screening age at 40 [respectively, 60], the probability of no early detection with annual screens will be 26.29% [respectively 21.98% for age 60]. The probability of no early detection is monotonically increasing when the screening interval increases within the same age group. This probability is mononically decreasing as the initial age increases for the same screening interval. The difference between the initial age 50 and 60 is smaller than that corresponding difference between the initial age 40 and 50 groups.

Mean lead time decreases as the screening time interval increases in Table 1; i.e., more frequent screening exams result in longer lead times. The increase in the mean lead time is due partly to the smaller point mass for zero lead time when screening exams are closer together. The standard deviation of

the lead time decreases as the time between screening exams increases. Table 1 also reveals that the standard deviation for the lead time is larger than the mean lead time. However, since the density of the lead time is very skewed, the median and fourth-spread are more meaningful estimates of the location and the spread; hence, we also consider the ratio of median/fourth-spread. This ratio decreases as the screening interval increases, and lies between 0.57 and 0.71 for different cases in Table 1.

The mode of the lead time is less than 0.1 year (or 1 month) when the screening interval is 2 or 2.5 years for all three groups; when the screening interval is 1.5 years, the mode is slightly above 0.1 years (or 1.2 months) for all three age groups; when the screening interval is one year, the mode is around 0.24 years (or 2.8 months) for all three age groups.

We reported the 95% credible intervals for the probability of "no early detection" in Table 1. For example, for the 50-year-old group who received annual screens, the probability of no-early-detection could be as low as 12.39%, or as high as 41.42%. The standard deviation of the probability of early-detection is also reported in Table 1 (5% - 9%).

We ran a simulation according to the recommandation of the U.S. Preventive Services Task Force, that is, biennial screening for women aged 50 to 74. The result does not bode well for screening. The probability of no-early-detection (the interval case) is 56.36% (standard error $\approx$ 5.24%), or effectively about 50-50 odds. In another simulation with annual screening from age 50 to 60, followed by biennial screening from 60 to 80 (and no further screens), the probability of zero lead time is 42.93%; it is very similar to the case of starting age at 50 or 60 in Table 1, with biennial screenings. Other statistics are also similar, such as the mean and median lead time, and the ratio (median/ fourth spread).

6 Discussion

We derived the probability distribution for the lead time in a periodic screening scenario when the lifetime is subject to a competing risk. The distribution of the lead time is a mixture of a point mass at zero and a piecewise continuous density function. The mixture formulation considers all diseased cases among participants in a screening program, that is, those diagnosed by screening and those interval cases.

Our model extends previous work on the lead time, where human lifetime was fixed at an upper bound, such as 80 years old. The new screening

model is more realistic. When comparing the results, the probability of no-early-detection for the 50-year-old group is very close to the results in the old model for different screening intervals with a fixed upper bound of the lifetime. The old model was simulated using the initial age $t_0 = 50$ and the fixed upper bound of age $T = 80$. However, the mean and the standard deviation of the lead time are both much smaller in the new model's simulations. The mode of the lead time is also much smaller. The median of the lead time (when it is positive) is fairly stable, about 0.85 years for all cases, suggesting that screening intervals greater than one year will have little effect on early detection. The result also shows that about a quarter of those who were detected early will have a lead time less than 0.4 years (the first quartile), and about a half of them will have a lead time between 0.4 and 1.7 years.

Since the estimates of the sensitivity, the transition probability, and the sojourn time distribution that we used in the simulation were from the HIP data, which is an old study, the results might not reflect current conditions and measuring equipment. We hope to obtain more accurate estimates of these key parameters.

By incorporating the possibility of death from a cause other than breast cancer, the inference can be extended to provide information on the effect of a screening program on a woman's risk of dying from breast cancer or even her risk of a diagnosis of breast cancer during her lifetime. One can use our model to infer future outcome measures that relate to different screening schedules and different initial ages of periodic screening, such as the possibility of being detected early from various screening programs. One can also use our model to evaluate and compare the characteristics of different possible screening programs. For example, the model can provide answers to questions, such as those concerning future outcomes of periodic screening exams for a woman in her 50s or 60s, and these outcomes change as the frequency of screening exams changes (e.g., screening every 12 or 24 months). Other questions of interest are: What is the probability that a woman's cancer will be detected early if she has cancer? How do changes in the screening program affect the lead time distribution?

As suggested by an anonymous referee, other factors, such as race, receptor status, etc. may affect outcomes for the lead time. Since the distribution of lead time is a function of the screening sensitivity, the transition density and the sojourn time distribution, and these three parameters can be estimated from screening data of specific sub-groups. We plan to collect data on some specific sub-groups and explore the effect of risk group in the future. Further work to improve the model includes the case when the sensitivity and the sojourn time are correlated. Our current model assumes that these two

key parameters are independent, but in fact there is strong evidence to suggest that these two key parameters may be correlated. For example, the sensitivity might be positively related to the time a person has remained in the preclinical state (the longer an individual stays in the preclinical state, the higher the sensitivity will be). We plan to investigate such extension in future work.

REFERENCES

Kafadar K., Prorok P.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* 1994; 13, 569-586.

Kafadar K., Prorok P.C. (1996): Computer simulation of randomized cancer screening trials to compare methods of estimating lead time and benefit time. *Computational Statistics and Data Analysis* 1996; 23, 263-291.

Kafadar K., Prorok P.C. (2003): Alternative definitions of comparable case groups and estimates of lead time and benefit time in randomized cancer screening trials. *Statistics in Medicine* 2003; 22, 83-111.

Kafadar K., Prorok P.C., Smith P.J. (1998): An estimate of the variance of estimators for lead time and screening benefit time in randomized cancer screening trials. *Biometrical Journal* 1998; 40, 801-821.

Mandelblatt J.S., Cronin K.A., Bailey S., Berry D.A., de Koning H.J., Draisma G., Huang H., Lee S.J., Munsell M., Plevritis S.K., Ravdin P., Schechter C.B., Sigal B., Stoto M.A., Stout N.K., van Ravesteyn N.T., Venier J., Zelen M., Feuer E.J. for the Breast Cancer Working Group of the Cancer Intervention and Surveillance Modeling Network.(2009): Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Ann Intern Med* 2009; 151(10): 738-747.

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

Prorok P.C. (1982) Bounded recurrence times and lead time in the design of a repetitive screening program. *J. Appl. Prob.* 1982; 19, 10-19.

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

Straatman H., Peer P.G.M., Verbeek A.L.M. (1997): Estimating lead time and sensitivity in a screening program without estimating the incidence in the screened group. *Biometrics* 1997; 53, 217-229.

U.S. Preventive Services Task Force (2009): Screening for Breast Cancer Recommendation Statement. 2009. `http://www.ahrq.gov/clinic/uspstf09/breastcancer/brcanrs.htm`.

Walter S.D., Day N.E. (1983): Estimation of the duration of a preclinical disease state using screening data. *American Journal of Epidemiology* 1983; 118, 856-86.

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

Wu D., Rosner G.L., Broemeling L.D. (2005): MLE and Bayesian inference of age-dependent sensitivity and transition probability in periodic screening. *Biometrics* 2005; 61, 1056-1063.

Xu J., Prorok P.C. (1995): Non-parametric estimation of the post-lead-time survival distribution of screen-detected cancer cases.*Statistics in Medicine* 1995; 14, 2715-2725.

Xu J., Fagerstrom R.M., Prorok P.C. (1999): Estimation of post-lead-time survival under dependence between lead-time and post-lead-time survival. *Statistics in Medicine* 1999; 18, 155-162.

Zelen M, Feinleib M.(1969): On the theory of screening for chronic diseases. *Biometrika*, 56, 601-614.