University of Louisville

ThinkIR: The University of Louisville's Institutional Repository

Electronic Theses and Dissertations

5-2013

Simulation study for the lead time in cancer screening when human lifetime is a competing risk.

Sarah K. Kendrick
University of Louisville

Follow this and additional works at: <https://ir.library.louisville.edu/etd>

Recommended Citation

Kendrick, Sarah K., "Simulation study for the lead time in cancer screening when human lifetime is a competing risk." (2013). *Electronic Theses and Dissertations*. Paper 736.
<https://doi.org/10.18297/etd/736>

This Master's Thesis is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact thinkir@louisville.edu.

SIMULATION STUDY FOR THE LEAD TIME IN CANCER SCREENING WHEN
HUMAN LIFETIME IS A COMPETING RISK

By
Sarah K. Kendrick
B.A., Wittenberg University, 2010

A Thesis
Submitted to the Faculty of the
University of Louisville School of Public Health and Information Sciences
in Partial Fulfillment of the Requirements
for the Degree of

Master of Science

Department of Biostatistics and Decision Science
University of Louisville
Louisville, KY

May 2013

SIMULATION STUDY FOR THE LEAD TIME IN CANCER SCREENING WHEN
HUMAN LIFETIME IS A COMPETING RISK

By

Sarah K. Kendrick
B.A., Wittenberg University, 2010

A Thesis Approved on

April 16, 2013

By the following Thesis Committee:

Dongfeng Wu, Thesis Director

Shesh Rai

Muriel Harris

ACKNOWLEDGEMENTS

First, I would like to thank my major advisor, Dr. Dongfeng Wu, for her unrelenting support, advice, and time throughout the past two years. I would also like to extend a big “Thank you” to my other committee members, Dr. Muriel Harris and Dr. Shesh Rai, for their continued help and support throughout this process. I would also like to extend my gratitude to my family, especially my parents, Leslie Townsend, Timothy Kendrick, Loren Townsend, and Amy Clore for their constant encouragement, advice, and cheerleading. Finally, I would like to thank the University of Louisville, specifically the faculty of the Bioinformatics and Biostatistics Department for allowing me the opportunity to pursue this degree and for supporting me during the past two years.

ABSTRACT

SIMULATION STUDY FOR THE LEAD TIME IN CANCER SCREENING WHEN
HUMAN LIFETIME IS A COMPETING RISK

Sarah K. Kendrick

April 16, 2013

PURPOSE: The purpose of this paper is to examine the lead time distribution in cancer screening trials when lifetime is a random variable in order to determine optimal initial age at screening and screening frequency. **METHODS:** Simulation was used in order to estimate the distribution of the lead time for a hypothetical individual with a future screening schedule. The lifetime distribution used comes from the Social Security Administration's actuarial life tables. The lead time distribution was then calculated based on two different sojourn time distributions (log-logistic and exponential) with four mean sojourn times (2, 5, 10, and 20 years), using three different initial screening ages, $t_0=40, 50, 60$, and four different screening frequencies, every six months, every year, every 1.5 years, and every two years for both males and females. **RESULTS:** Smaller time intervals between screenings yield a smaller probability of no benefit and a greater expected lead time.

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGMENTS.....	iii
ABSTRACT.....	iv
INTRODUCTION.....	1
METHODS.....	5
Lead Time Distribution when Lifetime is Fixed.....	5
The Lifetime Distribution.....	8
Lead Time Distribution when Lifetime is a Random Variable.....	11
RESULTS.....	13
DISCUSSION.....	20
REFERENCES.....	23
CURRICULUM VITAE.....	25

INTRODUCTION

Cancer has been sweeping the nation in recent history, and there has been a lot of research on cancer treatments and effective screening for cancer. Among men, prostate cancer is the most common cancer followed by lung cancer and colorectal cancer with lung cancer being the leading cause of cancer death in men followed by prostate cancer. For women, the most prevalent cancers are breast cancer, followed by lung cancer, and then colorectal cancer with lung cancer being the leading cause of cancer death, followed by breast cancer (Centers for Disease Control and Prevention, 2009).

Each type of cancer has a different screening method(s) used for early detection which could lead to a better prognosis. Prostate cancer has two different screening methods, a digital rectal exam (DRE) and a prostate specific antigen test (PSA). However, the United States Preventative Services Task Force does not recommend getting the PSA test for men who are asymptomatic (U.S. Preventative Services Task Force, 2012). Screening options for breast cancer consist of self exams, mammograms, and clinical breast exams. Lumps and changes in size or shape of the breast(s) or underarm are warning signs for breast cancer, and are usually detected by self or clinical examination (examination by a nurse or doctor). Even though these two methods can be helpful in detecting breast cancer, they have not been shown to decrease one's risk of

death from breast cancer. However, getting regular mammograms, an X-ray of the breasts, has been shown to decrease an individual's risk of death due to breast cancer (Centers for Disease Control and Prevention, 2012). Screening options for lung cancer include chest X-rays, sputum cytology, and CT scans but there is debate over whether any of these actually help in decreasing deaths from lung cancer (Centers for Disease Control and Prevention, 2011). Colorectal cancer, liver cancer, and all other cancers have screening test(s) just as the cancers mentioned above. Screening is more beneficial in some cancers, such as breast cancer, than others, such as lung cancer.

The most common model for the progression of cancer is one occurring in three stages: S_0 , S_p , and S_c (ZELEN & FEINLEIB, 1969). S_0 is known as the disease-free state, S_p is the preclinical state in which the individual has developed the disease, and can be detected by screening, but has shown no clinical symptoms, and S_c is the clinical state in which the individual exhibits clinical symptoms. The difference in time between the age at which an individual enters the preclinical state and the age the individual begins to experience clinical symptoms ($t_c - t_p$ where $t_c > t_p$) is defined as the individual's sojourn time. If disease was detected by screening after the individual entered the preclinical state, but prior to experiencing any symptoms, the time difference between the age at detection by screening and the age at onset of symptoms is referred to as the individual's lead time.

The topic of lead time has been an area of much research over the past 20 to 30 years. Many studies have worked to determine the parameters, formulas, and distributions that most accurately estimate the important descriptive statistics for the lead time distribution, such as mean and variance. One of the largest contributions to the topic of

lead time was the work done by Prorok in 1982. Prorok's goal was to determine the optimal number of screenings required for effective and efficient evaluation in repetitive screening trials. He used the lead time properties of screen detected cases to develop a stopping rule for these kinds of studies. What he found was that the lead time tended to stabilize after four to five screenings when the screening frequency was held constant. This property showed that any screenings in excess of the fourth or fifth one may no longer provide any additional information. However, Prorok's design contained one major hole: he did not take interval cases, where lead time is zero, into account (Prorok, 1982). Wu et al. (2007) took the method proposed by Prorok a little further. They included both screen-detected cases, and interval cases in order to derive the probability distribution function for the lead time, and used Prorok's results in their model as a special case – where there are no interval cases. However, Wu et al. (2007) derived the lead time distribution while assuming that lifetime is a fixed value, which is unrealistic. Then, in 2012, Wu et al. extended their previous method to the case where lifetime is subject to competing risks, and is, therefore, a random variable.

In this paper we will extend the research done by Wu, Kafadar, Rosner, and Broemeling in 2012. We will look at the lead time distribution for several different screening situations in order to gain an understanding of how/if screening is beneficial for different cancers. We will look at three hypothetical cohorts of initially asymptomatic men and women. The three cohorts will be based on initial age at screening: 40, 50, and 60 years old. Within each cohort we will examine the effect of four different screening frequencies on the lead time distribution: $\Delta = 6$ months, one year, one and a half years, and two years. We will also look at two different distributions for the sojourn time:

log logistic and exponential. Our results will give insight into the benefits of screening for different types of cancers, as well as the ideal initial screening age and screening frequency.

METHODS

Lead Time Distribution when Lifetime is Fixed

This study used simulation data to determine the lead time distribution for a hypothetical cohort of initially asymptomatic individuals with no history of cancer. First we must define some variables and notation. Let t be the age of the individual at a screening exam, t_0 be the age at which the individual has their initial screening, T be the individual's lifetime, L indicate the lead time, D be a binary random variable using $D=1$ to indicate disease development and $D=0$ to indicate no disease, and K be the total number of screenings the individual will undergo in their lifetime. Now we let $\beta(t)$ be the sensitivity of the screening modality and define $w(t)dt$ as the probability that an individual will transition from the disease-free state(S_0) to the preclinical state (S_p) during the interval $(t, t+dt)$; $q(x)$ as probability distribution function of the sojourn time, and $Q(z) = \int_z^\infty q(x)dx$ as the survivor function for the sojourn time. From the Health Insurance Plan for Greater New York (HIP) study we know that $\beta(t)$ was characterized using a logistic model,

$$\beta(t) = \frac{1}{1 + \exp \{-b_0 - b_1(t - \bar{t})\}}$$

Where $\bar{t}$ =the average age at entry. The transition density function is that of a log normal (μ, σ^2) density function multiplied by 0.2, the upper limit of lifetime risk,

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

Finally, we will use the log logistic and the exponential distributions to define the sojourn time distribution. This yields the following survivor and hazard functions:

1. Log logistic

$$Q(x) = \frac{1}{1 + (x\rho)^\kappa} \quad \text{and} \quad h(x) = \frac{\kappa x^{\kappa-1} \rho^\kappa}{1 + (x\rho)^\kappa}, \quad \kappa > 0, \quad \rho > 0, \text{ and}$$

2. Exponential

$$Q(x) = e^{-\lambda x} \quad \text{and} \quad h(x) = \lambda.$$

We then combine these two functions to define the density function for the sojourn time: $q(x) = h(x)Q(x)$, where x is the sojourn time. The above equations contain the following unknown parameters, b_0 , b_1 , μ , σ^2 , κ , and ρ (Wu, Rosner, & Broemeling, 2005).

Each individual undergoes a set of K screening exams during their lifetime at the ages of $t_0 < t_1 < t_2 < \dots < t_{K-1}$. For the sake of simplicity we let $T = t_K$ even though there is no screening at age T . We now look at the distribution of lead time when lifetime is a fixed value. As we mentioned in the introduction, lead time consists of two parts: a point mass at zero, $P(L = 0|D = 1, T = t_K)$, indicating an interval case, and a continuous conditional probability distribution function, $f_L(z|D = 1, T = t_K)$.

We show that:

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

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

Where

$$P(D = 1|T = t_K)$$

$$= \int_{t_0}^{t_K} \int_0^t w(x)q(t-x)dxdt = \int_0^{t_0} w(x)[Q(t_0-x) - Q(t_K-x)]dx \\ + \int_{t_0}^{t_K} w(x)[1 - Q(t_K-x)]dx$$

And

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

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

And where

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

Or

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

We can prove the validity of this probability by confirming that

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

(Wu, Rosner, & Broemeling, 2007).

The Lifetime Distribution

In order to find the distribution of lifetime, we used the actuarial table provided by the Social Security Administration. This life table provides information on the probability of mortality, the conditional probability of death within one year ($P(T < N+1|T \geq N)$), and life expectancy for both men and women, $N=0, 1, 2, \dots, 119$. It is produced using data from all Social Security populations which is comprised of residents of the United States and the District of Columbia, civilian residents of U.S. territories, civilian Federal employees, individuals in the U.S. Armed Forces abroad and their dependents, and all other U.S. citizens (Social Security Administration, 2012).

To obtain the distribution for the lifetime, some derivations are needed first. If we let $b_N = P(T < N + 1|T \geq N)$ and let $a_N = 1 - b_N = P(T \geq N + 1|T \geq N)$, then using the conditional probability formula we find that,

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

Now we can use mathematical induction to apply this concept to any integer age, t_0 ,

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

Next, we use a density approximation for $f_T(t = t_0 + N|T \geq t_0)$,

$$\begin{aligned} f_T(t = t_0 + N|T \geq t_0) &= \lim_{\epsilon \rightarrow 0} P(t_0 + N < T \leq t_0 + N + \epsilon | T \geq t_0) \\ &\approx P(t_0 + N < T \leq t_0 + N + 1|T \geq t_0) \end{aligned}$$

$$= P(T \geq t_0 + N | T \geq t_0) - P(T \geq t_0 + N + 1 | T \geq t_0)$$

$$= (1 - a_{t_0+N}) \prod_{i=1}^N a_{t_0+i-1}.$$

In this case, where $N < 120$, we will use a step function to approximate $f_T(t | T \geq t_0)$ to be $\approx f_T(N | T \geq t_0)$ for any real number t in $[N, N+1]$. Finally, we can show the validity of this probability distribution approximation:

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

(Wu, Kafadar, Rosner, & Broemeling, 2012).

We now use these results to plot the conditional lifetime distributions for men and women with initial screening ages of 40, 50, and 60. We must note that this does not take screening frequency or cause of death into account.

Figure 1. Conditional lifetime distributions for initial screening ages of 40, 50, and 60.

Lead Time Distribution When Lifetime is a Random Variable

Using the ideas and formulas when lifetime was a fixed number, we can determine the distribution of the lead time when lifetime is a random variable by implementing a few modifications. When lifetime is a random variable, the number of screenings an individual will receive in their lifetime is also variable and is a function of their lifetime, $K=k(T)$ where K is the largest integer that satisfies $t_{K-1} < T$. To obtain the distribution of the lead time when lifetime is a random variable we define the following,

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

$$f_L(z|D = 1, T \geq t_0) = \int_{t_0+z}^{\infty} f_L(z|D = 1, T = t) f_T(t|T \geq t_0) dt \quad z \in (0, \infty)$$

The lower bound in the above equation is (t_0+z) instead of t_0 since the lead time should be less than $t-t_0$, therefore, t should be greater than t_0+z . The equations $P(L = 0|D = t, T = t)$ and $f_L(z|D = 1, T = t)$ come from the situation where lifetime is a fixed value and the conditional distribution of the lifetime is,

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

We can show that the equations above are a valid mixed probability distribution:

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

(Wu et al., 2012).

RESULTS

We can apply the above methods to any screening schedule. For our example we will consider three different initial screening ages for both men and women of $t_0=40, 50$, and 60 . For each initial age we will examine the effects of four different screening frequencies, $\Delta=0.5, 1.0, 1.5$, and 2.0 years. Finally, for each combination of initial screening age and screening frequency we will consider the case where the sojourn time has a log-logistic distribution and the case where it has an exponential distribution. The following tables give the probability of no benefit (lead time equal to zero), for each initial screening age, screening frequency, and sojourn time distribution combination as well as the estimate for the probability of benefit, the expected lead time, and the median lead time (for females only since the results are very similar for males and females).

Table 1

A projection of the lead time distribution when mean sojourn time = 2 years

Δ^a	Log-Logistic				Exponential			
	P_0^b	$1-P_0$	EL ^c	Med ^d	P_0^b	$1-P_0$	EL ^c	Med ^d
Initial Screening Age $t_0=40$								
6 Months	28.40	71.60	0.533	2.9	32.26	67.74	0.745	2.9
12 Months	44.94	55.06	0.453	2.7	43.02	56.98	0.684	2.7
18 Months	52.87	47.13	0.380	2.8	51.58	48.42	0.627	2.8
24 Months	63.68	36.32	0.324	2.7	58.59	41.41	0.578	2.7
Initial Screening Age $t_0=50$								
6 Months	8.34	91.66	1.42	1.9	14.23	85.77	1.69	1.9
12 Months	24.10	75.90	4.20	1.8	24.87	75.13	1.52	1.8
18 Months	39.97	60.03	1.01	1.8	33.54	66.46	1.55	1.8
24 Months	52.71	47.29	0.861	1.7	40.76	59.24	1.42	1.7
Initial Screening Age $t_0=60$								
6 Months	5.40	94.6	1.92	1.7	11.32	88.68	2.41	1.9
12 Months	19.56	80.44	1.63	1.6	20.80	79.2	2.21	1.9
18 Months	34.32	65.68	1.37	1.6	28.77	71.23	2.02	1.9
24 Months	46.32	53.68	1.18	1.6	35.51	64.49	1.86	1.9

^a Time between screenings ($t_i - t_{i-1}$)

^b $P_0 = P(L=0|D=1)$ = Probability of no early detection (%)

^c The mean lead time is in years

^d The median when $L > 0$ (for non-interval cases)

Table 2

A projection of the lead time distribution when mean sojourn time = 5 years

Δ	Log-Logistic				Exponential			
	P_0	$1-P_0$	EL	Med	P_0	$1-P_0$	EL	Med
Initial Screening Age $t_0=40$								
6 Months	18.15	81.85	1.44	4.3	23.55	76.45	1.22	4.7
12 Months	20.48	79.52	1.42	4.1	29.40	70.60	1.20	4.5
18 Months	23.06	76.94	1.38	3.9	34.25	65.75	1.18	4.4
24 Months	26.18	73.82	1.35	3.8	38.50	61.5	1.16	4.2
Initial Screening Age $t_0=50$								
6 Months	7.35	92.65	3.31	3.1	7.47	92.53	2.79	3.0
12 Months	20.87	79.13	3.25	2.8	13.25	86.75	2.73	2.9
18 Months	34.43	65.57	3.16	2.6	18.32	81.68	2.66	2.9
24 Months	45.31	54.69	3.07	2.4	22.91	77.09	2.59	2.8
Initial Screening Age $t_0=60$								
6 Months	0.304	99.696	4.14	2.7	4.79	95.21	3.60	2.4
12 Months	1.43	98.57	4.08	2.6	9.20	90.8	3.52	2.4
18 Months	3.42	96.58	3.99	2.5	13.29	86.71	3.43	2.4
24 Months	6.15	93.85	3.89	2.4	17.08	82.92	3.36	2.4

Table 3

A projection of the lead time distribution when mean sojourn time = 10 years

Δ	Log-Logistic				Exponential			
	P_0	$1-P_0$	EL	Med	P_0	$1-P_0$	EL	Med
Initial Screening Age $t_0=40$								
6 Months	15.54	84.46	1.76	5.9	19.29	80.71	1.25	6.3
12 Months	16.94	83.06	1.81	5.6	23.21	76.79	1.28	6.1
18 Months	17.84	82.16	1.85	5.4	26.40	73.60	1.29	6.0
24 Months	18.95	81.05	1.89	5.2	29.22	70.78	1.31	5.9
Initial Screening Age $t_0=50$								
6 Months	2.52	97.48	3.70	4.2	5.04	94.96	2.75	3.7
12 Months	3.80	96.2	3.78	4.0	8.87	91.13	2.76	3.6
18 Months	5.29	94.71	3.83	3.8	12.27	87.73	2.76	3.5
24 Months	6.89	93.11	3.88	3.6	15.40	84.6	2.77	3.5
Initial Screening Age $t_0=60$								
6 Months	0.066	99.934	4.75	3.6	2.58	97.42	3.76	2.9
12 Months	0.315	99.685	4.81	3.4	5.04	94.96	3.75	2.9
18 Months	0.779	99.221	4.85	3.3	7.39	92.61	3.72	2.9
24 Months	1.47	98.53	4.90	3.2	9.64	90.36	3.71	2.8

Table 4

A projection of the lead time distribution when mean sojourn time = 20 years

Δ	Log-Logistic				Exponential			
	P_0	$1-P_0$	EL	Med	P_0	$1-P_0$	EL	Med
Initial Screening Age $t_0=40$								
6 Months	12.53	87.47	4.32	8.1	16.44	83.56	1.14	7.1
12 Months	13.73	86.27	4.53	7.7	19.37	80.63	1.18	6.9
18 Months	14.47	85.53	4.71	7.4	21.70	78.30	1.21	6.7
24 Months	15.05	84.95	4.91	7.1	23.74	76.26	1.26	6.5
Initial Screening Age $t_0=50$								
6 Months	1.87	98.13	6.86	5.9	3.81	96.19	2.42	4.4
12 Months	3.05	96.95	7.11	5.6	6.63	93.37	2.47	4.2
18 Months	4.07	95.93	7.31	5.3	9.12	90.88	2.50	4.1
24 Months	5.05	94.95	7.52	5.1	11.43	88.57	2.54	4.1
Initial Screening Age $t_0=60$								
6 Months	0.017	99.983	8.60	4.3	1.47	98.53	3.55	3.3
12 Months	0.078	99.922	8.72	4.3	3.09	96.91	3.56	3.3
18 Months	0.193	99.807	7.80	4.2	4.57	95.43	3.57	3.3
24 Months	0.366	99.634	8.93	4.1	6.01	93.99	3.58	3.3

The following figures give the density curves for the lead time for the different screening intervals, sojourn times, and sojourn distributions but only for females when $t_0 = 50$ since we get similar results for men and the other initial screening ages.

Figure 2. Lead time probability distributions when sojourn time distribution is log-logistic

Figure 3. Lead time probability distributions when sojourn time distribution is exponential

These results show that for a mean sojourn time (MST) of two years and a log-logistic sojourn time distribution, an individual that begins screening at age 40 and receives screenings twice a year has a 28.4% chance that their cancer will not be detected by early screening. This value increases to 44.94% and 63.68% when screening frequency is decreased to once a year and twice a year, respectively. We see a similar trend for initial screening ages of 50 years, 8.34%, 24.10%, and 52.71%, and 60 years,

5.40%, 19.56%, and 46.32%, for $\Delta = 0.5$, 1.0, and 2.0 years respectively. We find that the probability an individual will not experience detection by screening increases as screening frequency decreases when we hold MST and age constant. This probability decreases as age increases when holding screening frequency and MST constant, and decreases as MST increases with age and screening frequency held constant.

From the graphs of the lead time distributions based on the log-logistic sojourn time distribution, we see that the mode, mean, and median are monotonically increasing as MST increases. We find the same trend for the mean and median when the sojourn time distribution is exponential, however, the mode when the sojourn time distribution is exponential is relatively constant across mean sojourn times (right around one year). We also find that when the MST is two or five years the mean decreases as screening frequency decreases, but find the opposite effect when the MST is ten or twenty years. Also, mean increases as age increases with MST and screening frequency being held constant. The trends we see in the median are that it increases as age increases (frequency and MST held constant) and decreases as frequency decreases (MST and age held constant).

DISCUSSION

We took the model presented by Wu, Kafadar, Roner, and Broemling, 2012 and derived the lead time distribution for different sets of periodic cancer screening when lifetime was subjected to competing causes (it was a random variable). This model extends Wu et al.'s previous work to include more simulations that can apply to cancers other than breast cancer.

We examined the lead time distribution for two different sojourn time distributions: log-logistic and exponential, four different mean sojourn times: 2 years, 5 years, 10 years, and 20 years, and four different screening frequencies: twice a year, once a year, once every year and a half, and every two years, for both men and women. We found that the outcomes were very similar for both men and women, and therefore it wasn't necessary to present the results for both genders.

Gathering data from all of these simulations allows us to extend our results to several prevalent cancers such as breast cancer, prostate cancer, lung cancer, and colorectal cancer which are some of the most common cancers amongst men and/or women today. One study examined HIP study population, a sample from Edinburgh, and a sample from Canada in order to determine the mean sojourn times for each group. They assumed an exponentially distributed sojourn time and found the mean sojourn times to

be 2.5, 4.3, and 1.9 (ages 40-49) and 3.1 (ages 50-59), respectively (Shen & Zelen). Lung cancer and colorectal cancer were found to have similar mean sojourn times as breast cancer when assuming an exponentially distributed sojourn time. Lung cancer showed mean sojourn times between 1.38 and 3.86 years (Chien & Chen); proximal colorectal cancer gave mean sojourn times of 3.86 years for individuals aged 45 to 54 years, 3.78 for 55 to 64 year olds, and 2.70 for 65-74 year olds; and distal colorectal cancer was found to have mean sojourn times of 3.35 for individuals aged 45 to 54 years, 2.24 for 55 to 64 year olds, and 2.10 for 65 to 74 year olds (Zheng & Rutter). A doctor could use our results where MST is two or five years and the sojourn time distribution is exponential to develop an efficient screening program for individuals aged 40 to 60 years old who are at risk for breast cancer, lung cancer, or colorectal cancer. Another study, assuming an exponentially distributed sojourn time as well, found the mean sojourn time for prostate cancer to be 11.3 years for men aged 50 to 59 years and 12.6 years for men aged 60 to 69 years (Pashayan et al.). In this case, our results where MST is equal to either 10 or 20 years and the sojourn time distribution is exponential, could be used to come up with effective screening programs for males at risk for prostate cancer.

The Greater New York Health Insurance Plan study was conducted over 50 years ago. Therefore, since our estimates for the transition probability distribution, sojourn time distribution, and sensitivity distribution were adopted from the HIP study, our results may not be an accurate reflection of current conditions. Also, our simulations only take age and gender into account as a covariate. In the future, we hope to not only get more accurate estimates for the aforementioned parameters and distributions, but also to include some other possible covariates to make our results applicable to more subgroups.

Finally, our model assumes that sensitivity and sojourn time are independent of each other even though there is evidence that suggests this may not be the case. In future work, we may want to examine this relationship more closely.

REFERENCES

- Centers for Disease Control and Prevention. (2009, January 8, 2013). United States Cancer Statistics (USCS). Retrieved March 12, 2013, from <http://www.cdc.gov/Features/CancerStatistics/>
- Centers for Disease Control and Prevention. (2011, December 15, 2011). Lung Cancer Screening. Retrieved March 12, 2013, from http://www.cdc.gov/cancer/lung/basic_info/screening.htm
- Centers for Disease Control and Prevention. (2012). Breast Cancer. Retrieved February 28, 2013, from <http://www.cdc.gov/cancer/breast/.htm>
- Chien, C. R., & Chen, T. H. Mean sojourn time and effectiveness of mortality reduction for lung cancer screening with computed tomography. *International Journal of Cancer*, 122(11), 2594-2599.
- Pashayan, N., Duffy, S. W., Pharoah, P., Greenberg, D., Donovan, J., Martin, R. M., . . . Neal, D. E. Mean sojourn time, overdiagnosis, and reduction in advanced stage prostate cancer due to screening with PSA: implications of sojourn time on screening. *British Journal of Cancer*, 100(7), 1198-1204.
- Prorok, Philip C. (1982). Bounded Recurrence Times and Lead Time in the Design of a Repetitive Screening Program. *Journal of Applied Probability*, 19(1), 10-19. doi: 10.2307/3213911
- Shen, Y., & Zelen, M. Screening sensitivity and sojourn time from breast cancer early detection clinical trials: mammograms and physical examinations. *Journal of Clinical Oncology*, 19(15), 3490-3499.

- Social Security Administration. (2012, 2012). Actuarial Life Table. Retrieved February 28, 2013, from <http://www.socialsecurity.gov/OACT/STATS/table4c6.html>
- U.S. Preventative Services Task Force. (2012). Screening for Prostate Cancer. Retrieved March 12, 2013, from <http://www.uspreventiveservicestaskforce.org/prostatecancerscreening/prostatecancerfact.pdf>
- Wu, D., Kafadar, K., Rosner, G. L., & Broemeling, L. D. (2012). The lead time distribution when lifetime is subject to competing risks in cancer screening. *Int J Biostat*, 8(1). doi: 10.1515/1557-4679.1363
- Wu, D., Rosner, G. L., & Broemeling, L. (2005). MLE and Bayesian inference of age-dependent sensitivity and transition probability in periodic screening. *Biometrics*, 61(4), 1056-1063. doi: 10.1111/j.1541-0420.2005.00361.x
- Wu, D., Rosner, G. L., & Broemeling, L. D. (2007). Bayesian inference for the lead time in periodic cancer screening. *Biometrics*, 63(3), 873-880.
- ZELEN, M., & FEINLEIB, M. (1969). On the theory of screening for chronic diseases. *Biometrika*, 56(3), 601-614. doi: 10.1093/biomet/56.3.601
- Zheng, W., & Rutter, C. M. Estimated mean sojourn time associated with hemoccult SENSa for detection of proximal and distal colorectal cancer. *Cancer Epidemiology, Biomarkers & Prevention*, 21(10), 1722-1730.

CURRICULUM VITAE

Sarah Kendrick

909 Cannons Lane Louisville, KY 40207

502-259-8766 skkend01@louisville.edu

Education

Bachelor of Arts in Mathematics; Minor in Chemistry, May 2010

Wittenberg University, OH

Masters of Science in Biostatistics, May 2013

University of Louisville, KY

Honors

Member, Alpha Lambda Delta Honor Society 2006-2007

Recipient, Departmental Scholarship Spring 2012

Member, Deans List, Wittenberg University and University of Louisville

Member, Kentucky Public Health Association 2012-Present

Work Experience

Research Assistant, University of Louisville, Louisville KY, January 2012-present

Tutor, University of Louisville School of Public Health and Information Sciences,
Louisville, KY Fall 2011

Personal Assistant, Street Moda, Louisville, KY July 2011-January 2012

Customer Service Attendant, Kohl's Department Store, Jeffersonville, IN August 2010-
July 2011

Work Study, Wittenberg University, Springfield, OH February 2008-May 2010

Nanny, Prospect, KY, 2007-2010 Summers

Retail Associate, Quest Outdoors, Louisville, KY July 2005-December 2009

Volunteer Work

Family Allergy and Asthma; intern; Louisville, KY; summer 2009

Community Hospital Pediatric Unit; Springfield, OH; 2008

Mercy Hospital Pharmacy; Springfield, OH; 2007

Big Brothers Big Sisters of Kentuckiana, 2004-2006

Skills

Technical: Proficient in Microsoft Word, Excel, and PowerPoint; Extensive experience using Internet and Email including Microsoft Internet Explorer, Google Chrome, and Mozilla Firefox; proficient in statistical software such as SAS, R, and S-Plus.

Communication: Proficient in oral and written communication.