

## Estimation of Sojourn Time and Transition Probability of Lung Cancer for Smokers using the PLCO Data

Dengzhi Wang<sup>1,2</sup>, Beth Levitt<sup>3</sup>, Tom Riley<sup>3</sup> and Dongfeng Wu<sup>1\*</sup>

<sup>1</sup>Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, USA

<sup>2</sup>Department of Neurological Surgery, University of Louisville, Louisville, KY, USA

<sup>3</sup>Information Management Services, Inc. Rockville, MD 20852, USA

### Abstract

**Objectives:** The goal of this study is to investigate time durations in the disease-free state and the preclinical state of lung cancer for male and female smokers, using lung cancer data from the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.

**Methods:** We applied a modified likelihood function to the lung cancer data, to obtain maximum likelihood estimate and make Bayesian inference of the transition probability from the disease-free to the preclinical state, and the sojourn time distribution. The data was stratified by age and gender for smokers in the periodic screening program. A scaled Beta distribution was used for the transition probability density function, and a Weibull distribution was used to model the sojourn time in the preclinical state.

**Results:** The epidemiological estimate of screening sensitivity is 0.649 for males and 0.68 for females. The transition probabilities are not the same for males and females: it is increasing monotonically to 80 years old for males; while it has a single maximum at age 72.5 for females. For male, the maximum likelihood estimate of mean sojourn time is 1.82 years, the Bayesian posterior mean and median sojourn time is 1.50 and 1.48 years, respectively. For female, the corresponding maximum likelihood estimate, posterior mean and median sojourn time are 1.84, 1.74 and 1.79 years respectively. The Bayesian mean lifetime risks for male and female smokers developing lung cancer are 12.0%, and 6.8%, respectively.

**Conclusion:** Our estimation showed that male smokers are more susceptible to lung cancer, because they have a higher lifetime risk and higher transition probability density than the same aged female smokers. Once they enter into the preclinical state, the male smokers have a shorter mean sojourn time than the female, meaning that they are quicker to develop clinical symptom of lung cancer.

**Keywords:** Lung cancer modeling; Screening sensitivity; Sojourn time; Transition probability; Epidemiological methods

### Introduction

Lung cancer is the leading cause of cancer death in the world. Based on the GLOBOCAN 2012 [1] estimates, there were about 1.825 million lung cancer incidence in 2012 in the world; and about 1.59 million deaths from lung cancer, of which 1.099 million for men, and 0.491 million for women. In the United States, based on the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results (SEER) program data, lung cancer is the second most common form of cancer, and the first leading cause of cancer death [2]. It was estimated that there were 224,390 new cases in 2016, which is around 13.3% of all new cancer cases; and there would be 158,080 lung cancer death in 2016, which is about 26.5% of the total number of cancer death [2]. Approximately 6.5% of men and women will be diagnosed with lung and bronchus cancer at some point during their lifetime, based on the SEER 2011-2013 data [2]. Lung cancer is more common in men than in women [2]. And smoking is widely recognized as the leading cause of lung cancer. About 80% of lung cancer deaths are directly resulted from smoking [3]. Despite the very serious prognosis of lung cancer, some people with earlier stage lung cancers are cured.

The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial is a multicenter randomized controlled trial (RCT) evaluating screening programs for the four kinds of cancer. The purpose is to determine whether each specific screening modality can reduce mortality from a specific cancer, e.g., PLCO-Lung is to check whether screening with chest X-ray can reduce mortality from lung cancer [4,5]. Secondary objectives of the PLCO are to assess screening sensitivity,

specificity, incidence, etc. It started in 1993 and ended enrollment in 2001; about 77,500 men and 77,500 women aged 55 to 74 who has no previous history of any PLCO cancer were enrolled in ten screening centers across the US. The PLCO data collection was completed in 2009; so the PLCO data are existing data. These data are available to the authors without participants' identifiers for the development of new statistical methods, and it was exempted from the IRB review by the rule of the NIH, since no human subjects were directly involved. Participants in the PLCO-Lung cancer screening were randomized to either study or control arm: people in the study arm were offered four annual chest X-rays, with a follow-up time up to 10 years; people in the control arm had usual care (no screening), and were followed for 13 years. There were 70,618 subjects that received at least one chest X-ray, with 70,560 subjects between age 55 and 74 at the first screen. Based on their gender and smoking status, participants in the study group can be separated into four cohorts: male smokers, male never-smokers, female smokers, and female never-smokers. This study will focus on the 4-annual chest X-ray (CXR) screening for lung cancer for male and

\*Corresponding author: Dongfeng Wu, Information Management Services, Inc. Rockville, MD 20852, USA, Tel: (502)852-1888; E-mail: [dongfeng.wu@louisville.edu](mailto:dongfeng.wu@louisville.edu)

Received June 28, 2017; Accepted June 31, 2017; Published August 10, 2017

**Citation:** Wang D, Levitt B, Riley T, Wu D (2017) Estimation of Sojourn Time and Transition Probability of Lung Cancer for Smokers using the PLCO Data. J Biom Biostat 8: 360. doi: [10.4172/2155-6180.1000360](https://doi.org/10.4172/2155-6180.1000360)

**Copyright:** © 2017 Wang D, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

female smokers, stratified by age. The number of male smokers who participated the initial screening exam is 21,335, with the average age 62.7; and the number is 14,257 for female smokers, with the average age 62.1, correspondingly.

Based on the natural history of tumor growth, each cancer patients are assumed to experience three states: the disease-free state  $S_0$ , the preclinical state  $S_p$  in which an asymptomatic individual unknowingly has the disease that a screening exam can detect, and the clinical state  $S_c$  when the disease manifests itself in clinical symptoms. The progressive disease model was first used by Zelen and Feinleib [6], denoted by  $S_0 \rightarrow S_p \rightarrow S_c$  (Figure 1).

Transition probability is the probability density function of the time duration in the disease-free state  $S_0$ , and it provides important information on at what age people will move from the disease-free to the preclinical state. However, it is difficult to estimate without proper modeling. Sojourn time is the time duration in the preclinical state  $S_p$ . If a person enters the preclinical state ( $S_p$ ) at age  $t_p$ , and his (or her) clinical symptoms present later at age  $t_c$ , then  $Tp=(t_c-t_p)$  is the sojourn time in the preclinical state. The nature of data collection in a screening program makes it impossible to observe the onset of either  $S_p$  or  $S_c$ . Therefore, estimation of the sojourn time is difficult without proper modeling. Usually a person with a longer sojourn time means that it is easier to catch the disease by screening exams. If he (or she) is offered a screening exam at time  $t$  within the time interval  $(t_p, t_c)$  and cancer is diagnosed, then the length of the time  $L=(t_c-t)$  is the lead time (Figure 1). The screening sensitivity is the probability that the screening exam is positive, given that an individual is in the preclinical state  $S_p$ .

The screening sensitivity, the sojourn time distribution and the transition probability are the three key parameters in screening modeling, since all other estimations (such as the lead time distribution and probability of over-diagnosis) can be expressed as functions of the three key parameters. Therefore, accurate estimation of the three key parameters is important in cancer screening. Our goal is to provide accurate statistical inference for the distribution of sojourn time and the transition probability from the disease-free to the preclinical state for smokers using the PLCO-Lung cancer screening data, and we will use a new conditional likelihood function to achieve this.

**Methods**

We let  $\beta(t)$  be the screening sensitivity at age  $t$ ,  $q(x)$  be the probability density function (PDF) of the sojourn time, and  $w(t)$  be the PDF of the time duration in the disease-free state. Inspired by Wu et al. [7], a new conditional likelihood method for estimating sojourn time and transition probability density was developed and applied to the PLCO-Lung data for the two cohorts: male and female smokers. Data from each cohort includes the total number of participants at each screening exam  $n_{k,t_0}$ , the number of detected and confirmed cancer cases  $S_{k,t_0}$  at each screening exam, and the number of interval cases  $n_{k,t_0}$  between

two consecutive exams. These data were stratified by participants' age  $t_0$  at the study entry, which was from 55 to 74 (inclusive) in this study.

This study is to accurately estimate the time durations in the disease-free state and the preclinical state, which will provide critical information for oncologists and clinicians. To achieve this, we first estimate the screening sensitivity  $\beta(t)$ . Based on the previous lung cancer screening data analysis [8,9] and input from lung cancer radiologists, sensitivity does not depend on age in lung cancer screening. Hence the sensitivity was estimated by the epidemiologic approach: using the total number of screen-detected cases divided by the sum of screen-detected cases and interval cases [10].

$$\beta(t) = \beta_0 = \sum_{t_0=55}^{74} \sum_{k=1}^k S_{k,t_0} / \left( \sum_{t_0=55}^{74} \sum_{k=1}^k S_{k,t_0} + \sum_{t_0=55}^{74} \sum_{k=1}^k r_{k,t_0} \right) \tag{1}$$

This provides  $\beta_0=0.649$  for male smokers, and  $\beta_0=0.680$  for female smokers, which would be used in the likelihood function for  $\beta(t)$ .

For each gender of the PLCO screening data, based on their initial age  $t_0$ , we developed a new conditional likelihood function  $L(\cdot | t_0)$ :

$$L = \prod_{t_0=55}^{74} L(\cdot | t_0) \tag{2}$$

$$L(\cdot | t_0) = \prod_{k=1}^k \left( \frac{D_{k,t_0}}{P_{0,t_0}} \right)^{S_{k,t_0}} \left( \frac{I_{k,t_0}}{P_{0,t_0}} \right)^{r_{k,t_0}} \left( \frac{P_{k,t_0}}{P_{0,t_0}} \right)^{n_{k,t_0} - S_{k,t_0} - r_{k,t_0} - n_{k+1,t_0}} \tag{3}$$

This likelihood function is different from the previous likelihood in Wu et al. [7], since it is conditional on the probability of no clinical cancer at or before the initial exam, which matches the enrollment criteria of the mass screening study. Here  $D_{k,t_0}$  is the probability that an individual will be diagnosed at the  $k$ -th scheduled exam, given that he is in the preclinical state  $S_p$ ; and  $I_{k,t_0}$  is the probability of being an incident case within the  $k$ -th screening interval  $(t_{k-1}, t_k)$ , with  $K=4$ , since there were four annual screening exams in the PLCO lung cancer study group. And  $P_{0,t_0}$  is the probability of no detected lung cancer before or at the initial exam with initial age  $t_0$ . The probabilities  $D_{k,t_0}$  and  $I_{k,t_0}$  were derived in Wu et al. 2005 [7]:

$$D_{k,t_0} = \beta(t_{k-1}) \left\{ \sum_{i=0}^{k-2} [1-\beta(t_i)] \dots [1-\beta(t_{k-2})] \int_{t_{i-1}}^{t_i} w(x) dx + \int_{t_{k-2}}^{t_{k-1}} w(x) Q(t_{k-1}-x) dx \right\} \tag{4}$$

$$I_{k,t_0} = \sum_{i=0}^{k-1} [1-\beta(t_i)] \dots [1-\beta(t_{i-1})] \int_{t_{i-1}}^{t_i} w(x) [Q(t_{i-1}-x) - Q(t_i-x)] dx + \int_{t_{k-1}}^{t_0} w(x) [1-Q(t_k-x)] dx \tag{5}$$

And

$$P_{0,t_0} = 1 - \int_0^{t_0} w(x) dx + \int_0^{t_0} w(x) Q(t_0-x) dx \tag{6}$$

$$n_{k+1,t_0}=0 \text{ and } k=1,2,3,4. \tag{7}$$

Where  $Q(x) = \int_x^\infty q(y) dy$  is the survivor function of the sojourn time in the preclinical state  $S_p$ .

Appropriate parametric functions for  $w(t)$  and  $q(x)$  were carefully

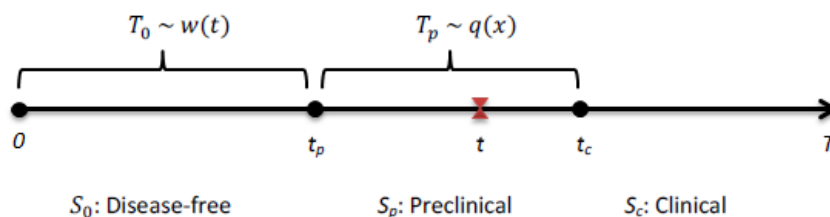


Figure 1: Illustration of disease progression and the lead time.

chosen. Instead of the log-normal distribution for  $w(t)$ , a scaled Beta distribution was used:

$$w(t|w_0, a, b) = \frac{w_0}{t_U - t_L} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{t-t_L}{t_U-t_L}\right)^{a-1} \left(1 - \frac{t-t_L}{t_U-t_L}\right)^{b-1} \quad (8)$$

Where  $t$  is the age at screening,  $a, b$  are the parameters in the Beta distribution,  $w_0$  is the lifetime risk of developing lung cancer at some point during one's lifetime for male or female smokers, a variable to be estimated. Based on the result from SEER, the age to make a transition from the disease free to the preclinical state is from 20 to 80 years old. Hence we let  $t_L=20, t_U=80$ , meaning that the transition from the disease-free state to the preclinical would happen in the age interval of (20, 80) if one develops clinical cancer. In this model,  $w_0, a$  and  $b$  are the parameters to be estimated.

We used the Weibull distribution to model the sojourn time in the preclinical state:

$$q(x|\alpha, \lambda) = \alpha \lambda x^{\alpha-1} \cdot e^{-\lambda x^\alpha}, Q(x) = e^{-\lambda x^\alpha} \quad (9)$$

where  $x$  is the sojourn time,  $\alpha$  and  $\lambda$  are positive parameters to be estimated.

In summary, as we mentioned earlier,  $w_0, a, b, \alpha$  and  $\lambda$  are the parameters to be estimated using the new likelihood function.

## Results

Both maximum likelihood estimates (MLE) and Bayesian posterior samples were used to make inferences for the five unknown parameters in the model, i.e.,  $\theta = (w_0, a, b, \alpha, \lambda)$ . Theoretically, the first parameter has a domain of (0, 1) and the last four have a domain of (0,  $\infty$ ). The practical meaning of these parameters will limit them to a finite range. The ranges were identified as:  $0.01 < w_0 < 0.99, 1.01 < a < 20, 0.5 < b < 10, 0.1 < \alpha < 5, 0.1 < \lambda < 2$ .

Markov Chain Monte Carlo (MCMC) was used to generate posterior random samples using non-informative priors and the joint posterior distribution of the parameters for a Bayesian inference. The posterior simulation was partitioned into 3 sub-chains, then Gibbs sampling was used to sample the posteriors for  $w_0, (a, b), (\alpha, \lambda)$  separately. Similar procedure in the Appendix from Wu et al. [7] was followed for this paper in the implementation of MCMC.

A non-informative priors for the parameters:  $a$  follows Uniform (1.01, 20), and  $b$  follows Uniform (0.5, 10). The prior distribution for  $\alpha$  was Uniform (0.1, 5), and the prior for  $\lambda$  was Uniform (0.1, 2). The prior for  $w_0$  was Uniform (0.01, 0.99). Each Markov Chain Monte Carlo simulation was run for 20,000 steps, with a burn-in of 5,000 steps. After the burn-in time, the posteriors were sampled every 100 steps, giving 150 posterior samples from each chain for the parameter vector  $\theta$ . Five chains were simulated, each with different starting values that are over dispersed with respect to the target distribution. Bayesian output analysis showed convergence. The 150 posterior samples from

each of the 5 chains were pooled for the analysis, giving a total of 750 posterior samples  $\theta_i^* = 1, 2, \dots, 750$ .

The MLE and Bayesian posterior estimates of 8 for the PLCO data are shown in Table 1, for both male and female smokers. The posterior mean and median are close to the MLEs, especially for the female group. For the male group, the largest difference is in the estimation of the parameter  $\alpha$  for the sojourn time distribution: the MLE is less than 1 (0.970), while the posterior mean and median are 1.852 and 1.389 correspondingly. This causes the different shape of the sojourn time distribution near zero, and a large difference in the mean sojourn time (MST) estimate, compared with the result from their female counterpart.

Another issue for the male cohort is that the MLE of the transition density parameter  $b$  is less than 1 (0.903), while the Bayesian posterior mean and median are greater than 1 (1.056 and 1.014, respectively). Even though the values are close, this causes different trend for the transition density curve when it is approaching 80 years old (see first graph in Figure 2). Since our study was focus on the age interval between 55 and 74, the results from these two methods are pretty matched.

The estimated probability density curve  $w(t)$  based on the MLE and the posterior median (with 95% confidence band) are plotted in Figure 2. The posterior median transition probability varies from  $1.24 \times 10^{-3}$  to  $6.04 \times 10^{-3}$  for males aged 55–74. This means, in every 1000 people, there will be 1.24–6.04 people making a transition from the disease-free state to the preclinical state lung cancer per year, depending on their age, whereas these numbers are 0.97–3.22 per 1000 for females. The transition probability is not a monotone function of age for female, with a single maximum at age 72.5; whereas for male, it tends to increase all the way to 80 years old. Female smokers have a much lower transition probability compared with the male smokers to enter into the preclinical state. This is also reflected on the much lower estimated  $w_0$  for females (Bayesian median 0.066) than for males (Bayesian median 0.117), because  $w_0$  indicates the lifetime risk over all ages for lung cancer.

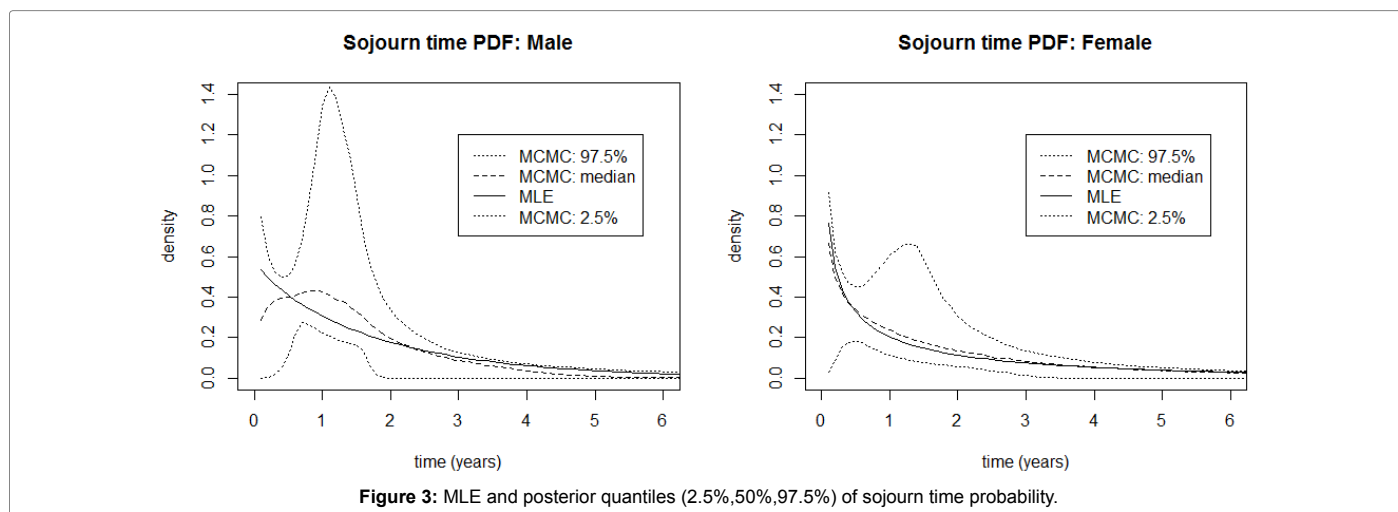
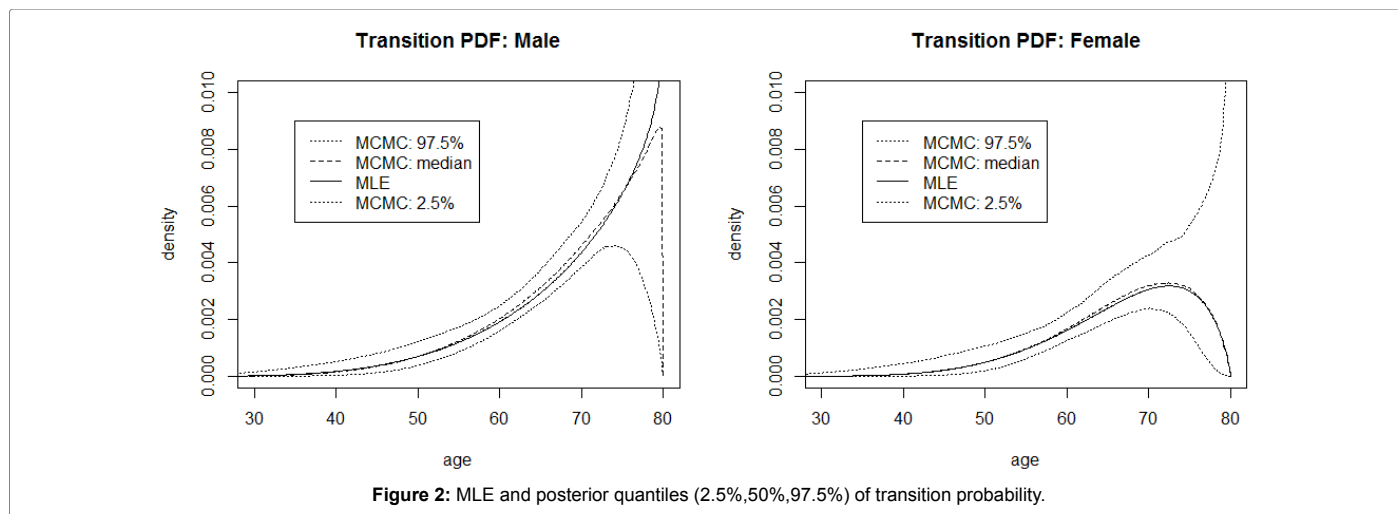
The sojourn time probability distribution  $q(x)$  can be seen from Figure 3. It is clear that the probability densities are concentrated within 2 years for both genders. The posterior mean sojourn time (MST) is 1.50 years for male, with a posterior median of 1.48 years, and the 95% highest posterior density (HPD) interval (1.06, 2.05). The posterior MST for female is 1.74 years, with a posterior median of 1.79 years, and the 95% highest posterior density (HPD) interval (1.10, 2.25). The MST from MLE are 1.82 and 1.84 years, for male and female respectively. So the MST for female seems longer than the MST for male, by either MLE or Bayesian estimate, meaning that females may have a longer sojourn time in the preclinical state.

## Discussion and Conclusion

We applied a new likelihood function to the PLCO data and

Parameters	Male Smokers				Female Smokers			
	Bayesian posterior estimate				Bayesian posterior estimate			
	MLE	Mean	Median	SE	MLE	Mean	Median	SE
$w_0$	0.115	0.12	0.117	0.022	0.062	0.068	0.066	0.015
$a$	4.381	4.843	4.8	1.327	6.163	6.522	6.19	2.203
$b$	0.903	1.056	1.014	0.367	1.738	1.843	1.769	0.692
$\alpha$	0.97	1.852	1.389	1.129	0.623	0.862	0.745	0.525
$\lambda$	0.547	0.501	0.507	0.105	0.592	0.56	0.56	0.115
MST (years)	1.817	1.503	1.477	0.284	1.842	1.744	1.789	0.298

Table 1: MLE and Bayesian posterior estimates for the PLCO data.



obtained the maximum likelihood estimate and Bayesian estimate of the key parameters in lung cancer for smokers. We used epidemiological method to estimate the sensitivity for the study, and the sensitivity is 0.649 for males, and 0.68 for females.

The NCI's Cooperative Early Lung Cancer Group conducted an important study regarding the sensitivity, specificity, and predictive values of chest X-ray (CXR) in the early detection of lung carcinoma in 1984. The NCI trials demonstrated that the sensitivity of CXR is from 0.54-0.84, with an average at 0.69 [11]. Our simple epidemiological estimate of the sensitivity is compatible with their result. Jang et al. [12] studied Johns Hopkins Lung Project (JHLP) data with CXR and got the estimated sensitivity as 0.568. Kim et al. [13] studied the efficacy of dual lung cancer screening by CXR and sputum cytology using JHLP data, the study showed that the screening procedure with X-ray only has improved from 79.93% to 85.34% when the screening exams were combined with cytology. Ten Haaf et al. [14] used individual-level data from the National Lung Screening Trial (NLST) and PLCO trial to estimate the screening sensitivity for different stage of lung cancer. According to their results, except for the IV stage, the sensitivities of CXR at the earlier stage (IA-IIB) are below 50% for the non-small cell carcinoma, but the sensitivity could reach 97.31% for CXR to detect small cell carcinoma at stage IV.

For smokers in the PLCO-Lung study, the MLE of the mean sojourn time (MST) is about 1.82 years for males, and 1.50 years using Bayesian posterior mean, with a 95% Highest Posterior Density (HPD) credible interval of (1.06, 2.05) years. For females, the MLE of the MST is about 1.84 years, and 1.74 years by Bayesian posterior mean, with a 95% HPD credible interval of (1.10, 2.25) years. For the Mayo Lung Project study [15], of which the study design is similar to this study, the MST was 2.2 years for male smokers. Liu [8] studied NLST for lung cancer with CT scan, they estimated the mean sojourn time was 1.44 years for males and 1.62 years for females. By using The Lung Cancer Screening Program at the Memorial Sloan-Kettering Cancer Center (MSKC-LCSP) data, Chen et al. [9] had a MST about 3.35 years for male smokers. Chien et al. [16] summarized several MST estimates from different low dose spiral CT, ranging from 1.38– 3.86 years. Our MST estimates (1.48~1.84) are within this range. Whereas ten Haaf et al. [14] estimated a higher MST for both genders: between 3.09-5.32 years for males, and 3.35-6.01 years for females, depending on the type of carcinoma.

The transition probability from the disease-free to the preclinical state increases all the way to age 80 for male smokers, while it has a peak around age 72.5 for females. We compared this result with the SEER database. The "SEER Cancer Stat Fact Sheets" [2] shows that the



probability of developing lung cancer has a single maximum between age 65 and 74 for both genders. Our female results agree with that fact, but the male results do not. The transition density from NLST [8] is a sub-density with a unimodal around age 70 for both genders.

Lung cancer is more common in men than in women. Overall, the chance that a man will develop lung cancer in his lifetime is about 7.19% (1 in 14); for a woman, the risk is about 6.04% (1 in 17) [17]. These numbers include both smokers and non-smokers. The risk is higher for smokers, and lower for non-smokers. Our estimated posterior mean of  $w_0$  was 11.95% for male smokers, and 6.82% for female smokers, which are reasonable, because they are both higher than the corresponding values for the general population. This is the first time that the lifetime risk was treated as a variable in the model. The risk for male smokers has increased 66.2% comparing with the general male population (from 7.19% to 11.95%); and the risk for female smokers has increased 12.9% comparing with the general female population (from 6.04% to 6.82%). These indicate that the risk of developing lung cancer is much higher for male smokers than for female smokers. Villeneuve and Mao [18] studied lifetime probability of developing lung cancer, by smoking status for Canadian people. They found that 172/1,000 of male current smokers will eventually develop lung cancer; this probability among female current smokers was 116/1,000. Our estimated  $w_0$  for both genders are lower than their result.

Our estimation showed that male smokers are more susceptible to lung cancer, because male smokers have a higher lifetime risk and higher transition probability than their female counterpart. Once they enter into the preclinical state, the male smokers seem to have a shorter mean sojourn time than the females, meaning that their tumors seem quickly to develop into the clinical disease state. The key parameters obtained from this study are also important, because other interesting terms, such as the lead time distribution, the percentage of over-diagnosis, etc., are functions of these key parameters, and our future work on estimating long term outcomes will use the estimated values of the parameters from this paper.

#### Acknowledgements

We authors thank the National Cancer Institute (NCI) for access to the NCI's data collected by the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.

#### References

1. GLOBOCAN (2012) Lung Cancer Estimated Incidence, Mortality and Prevalence Worldwide.
2. SEER (2016) Cancer Stat Facts Lung and Bronchus Cancer.
3. American Cancer Society (2017) Lung Cancer Risk Factors.
4. Prorok PC (2000) Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials* 6: 273-309.
5. Gohagan JK, Prorok PC, Hayes RB, Kramer BS (2000) The Prostate, Lung Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history *Control Clin Trials* 6: 251-272.
6. Zelen M, Feinleib M (1969) On the theory of screening for chronic diseases. *Biometrika* 3: 601-614.
7. Wu D, Rosner GL, Broemeling L (2005) MLE and Bayesian inference of age-dependent sensitivity and transition probability in periodic screening. *Biometrics* 4: 1056-1063.
8. Liu R (2015) Bayesian Estimation of the Three Key Parameters in CT for the National Lung Screening Trial Data. *Journal of Biometrics & Biostatistics*.
9. Chen YT, Erwin D, Wu D (2014) Over-diagnosis in Lung Cancer Screening using the MSKC-LCSP Data. *Journal of Biometrics & Biostatistics*.
10. Walter SD, Day NE (1983) Estimation of the duration of a pre-clinical disease state using screening data. *Am J Epidemiol* 6: 865-866.
11. Gavelli G, Giampalma E (2000) Sensitivity and specificity of chest X-ray screening for lung cancer review article. *Cancer* 11: 2453-2456.
12. Jang H, Kim S, Wu D (2013) Bayesian lead time estimation for the Johns Hopkins Lung Project data. *J Epidemiol Glob Health* 3: 157-163.
13. Kim S, Erwin D, Wu D (2012) Efficacy of Dual Lung Cancer Screening by Chest X-Ray and Sputum Cytology Using Johns Hopkins Lung Project Data. *Journal of Biometrics & Biostatistics*.
14. Ten Haaf K, Van Rosmalen J, De Koning HJ (2015) Lung cancer detectability by test, histology, stage, and gender estimates from the NLST and the PLCO trials. *Cancer Epidemiol Biomarkers Prev* 1: 154-161.
15. Wu D, Erwin D, Rosner GL (2011) Sojourn time and lead time projection in lung cancer screening. *Lung Cancer* 3: 322-326.
16. Chien CR, Chen TH (2008) Mean sojourn time and effectiveness of mortality reduction for lung cancer screening with computed tomography. *Int J Cancer* 11: 2594-2599.
17. NIH (2015) SEER Cancer Statistics Review.
18. Villeneuve PJ, Mao Y (1994) Lifetime probability of developing lung cancer, by smoking status, Canada. *Can J Public Health* 6: 385-388.