Introduction

Lung cancer, the leading cause of cancer death for both men and women, occurs in the lungs and claims more lives each year than do breast, colon, and prostate and ovarian cancers combined. There are two major types of lung cancer which have been identified: about 15% of lung cancers are small cell lung cancer, and the most common type is non-small cell lung cancer. The age-specific lung cancer incidence rate rises with advancing age and reaches its peak between 65 and 74 [1].

Smoking is the major risk factor for development of lung cancer. The general prognosis of lung cancer is poor because symptoms tend not to show up until it is at an advanced stage. Five-year survival is 54.8% for stage I lung cancer, but only 4.2% in advanced, inoperable non-small cell lung cancer. The age-specific lung cancer incidence rate rises with advancing age and reaches its peak between 65 and 74 [1].

The general prognosis of lung cancer is poor because symptoms tend not to show up until it is at an advanced stage. Five-year survival is 54.8% for stage I lung cancer, but only 4.2% in advanced, inoperable non-small cell lung cancer. The age-specific lung cancer incidence rate rises with advancing age and reaches its peak between 65 and 74 [1].

We assume the commonly followed disease progression model and the disease develops by progressing through three states [3], denoted by $S_r \rightarrow S_p \rightarrow S_c$. The state $S_r$ refers to the disease-free state, where either a 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 state $S_p$ is the preclinical disease state, in which an asymptomatic individual unknowingly has the disease that a screening exam can detect. The state $S_c$ is a state at which the disease manifests itself with clinical symptoms. This was illustrated in Figure 1. The three key parameters in the probability model are: the sensitivity, the sojourn time and the transition probability. The sensitivity is the probability that the screening exam is positive given that the individual is in $S_p$. The sojourn time refers to the time interval between the beginning of the preclinical state, and the manifestation of clinical symptoms, i.e., $(S_p, S_c)$. The transition probability density from the disease free into the preclinical stage is the probability density function of making a transition from $S_r$ to $S_p$, it is in fact a sub-pdf.

We will focus on estimating the three key parameters in CT screening using the NLST data. There is a reason 1). CT screening is the

Keywords: Sensitivity; Transition probability density; Sojourn time; NLST

Abstract

In this study cancer screening likelihood method was used to analyze the CT scan group in the National Lung Screening Trial (NLST) data. Three key parameters: screening sensitivity, transition probability density from disease free to preclinical state, and sojourn time in the preclinical state, were estimated using Bayesian approach and Markov Chain Monte Carlo simulations. The sensitivity for lung cancer screening using CT scan is high; it does not depend on a patient’s age, and is slightly higher in females than in males. The transition probability from the disease-free to the preclinical state has a peak around age 70 for both genders, which agrees with the fact that the highest lung cancer incidence rate appears between age 65 and 74. The posterior mean sojourn time is around 1.5 years for all groups, and that explains why screening only have a short time interval to catch lung cancer. Accurate estimation of the three key parameters is critical for other estimations such as lead time and over-diagnosis, because these quantities are functions of the three key parameters.

Acknowledgments

This work was supported by University of Louisville Research Grant 20150225. The authors are also very grateful to Dr. Geng Li of the School of Public Health and Information Sciences, University of Louisville for providing the data used in this study.

Key Parameters in CT for the National Lung Screening Trial Data. J Biom Biostat 2015, 6:5
Rui Qi Liu, Beth Levitt, Tom Riley and Dongfeng Wu**

1Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, Louisville, KY, USA
2Information Management Services Inc, USA

Reference


Copyright: © 2015 Liu R, 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.
is the number of cases diagnosed in the clinical state $\tau^1$, where $t_i = t_0 + i$ for annual screenings. Let $q(x)$ be the survival function of the sojourn time in the preclinical state $S_p$. The density can help us determine which age group of people is at higher risk for the disease, so people can take preventive steps before the symptoms show up [4].

**Methodology**

Let the time variable $t$ represents the participants’ age. Then let $\beta(t)$ represents the sensitivity of the screening. Define $w(t)dt$ as the probability of a transition from $S_0$ to $S_p$ during $(t, t+dt)$. Define $w(t)$ as the time interval between the $i$-th and the $(i+1)$-th screening exams. Let $n_{i,k}$ be the total number of individuals in this cohort examined at the $i$-th screening exam, and $R_{i,k}$ is the number of cases diagnosed at the $i$-th screening exam, and $C_{i,k}$ is the number of cases diagnosed in the clinical state $S_p$, within the interval $(t_{i-1}, t_i)$, which is the interval cases. For the NLST study entry, the likelihood function for all groups is:

$$L = \prod_{k=1}^{3} \prod_{i=55}^{74} D_{i,k}^{n_{i,k}} f_{i,k}^{R_{i,k}} \left(1 - D_{i,k}^{n_{i,k}} - I_{i,k}^{R_{i,k}}\right)^{n_{i,k} - R_{i,k} - I_{i,k}}$$

where $D_{i,k}$ is the probability that an individual will be diagnosed at the $k$-th scheduled exam given that he or she is in $S_p$, and $I_{i,k}$ is the probability of being incident in the $k$-th screening interval. Two probabilities of these intervals were originally derived in [5]:

$$D_{i,k} = \beta(t_i) \left\{ 1 - \beta(t_i) \right\} \left\{ 1 - \beta(t_i) \right\} \int_{c_i} f(x) q(t_i,s) \, dx + \int_{c_i} f(x) q(t_i,s) \, dx$$

$$I_{i,k} = \beta(t_i) \left\{ 1 - \beta(t_i) \right\} \int_{c_i} f(x) q(t_i,s) \, dx$$

The three key parameters were estimated from the NLST data using the following parametric models:

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

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

and

$$Q(x) = \exp(-\lambda x^\alpha), \alpha > 0, \lambda > 0.$$
Figure 2: The MCMC trace plots of the parameters $\theta=(b_0, b_1, \mu, \sigma^2, \lambda, \alpha)$ using CT arm overall group in NLST data.

Figure 3: The posterior density plots of the parameters $\theta=(b_0, b_1, \mu, \sigma^2, \lambda, \alpha)$ using CT arm overall group in NLST data.
Figure 4: The posterior density plots of the parameters $\theta=(b_0, b_1, \mu, \sigma^2, \lambda, \alpha)$ using CT arm male group in NLST data.

Figure 5: The posterior density plots of the parameters $\theta=(b_0, b_1, \mu, \sigma^2, \lambda, \alpha)$ using CT arm female group in NLST data.
The age-dependent Bayesian estimates of the sensitivity $\beta$ and the transition density $w(t)$ for each group are listed in Table 2. Figures 6-8 show posterior quantiles of sensitivity and transition probability for each group.

From equation (4), we can see $\beta(t)$ will be monotonic increasing with age $t$ if $b_1>0$. In our cases, $b_1$ is greater than but is also closed to 0 in all cases. We did a Bayes hypothesis test for $H_0: b_1 \leq 0$ versus $H_1: b_1>0$. For the overall group which includes both genders, the posterior probability of a positive slope is $P(b_1>0|Data) = 0.532$; For males, this posterior probability is $P(b_1>0|Data) = 0.513$; for females, this posterior probability is $P(b_1>0|Data) = 0.594$. Hence, the evidence of age effect is not significant in all groups.

The age-dependent transition probability is itself a sub-pdf from our model construction. The posterior density curve of the transition probability could be seen from Figures 6-8. The transition probability is not a monotone function of age, having a single maximum around age 70 for both males and females.

The posterior mean sojourn time is 1.48 years for CT overall, 1.44 years for CT male and 1.62 years for CT female, with a posterior median of 1.47 years for CT overall, 1.41 years for CT male and 1.58 years for CT female, respectively. The 95% highest posterior density (HPD) interval is (1.22, 1.77) for overall, (1.11, 1.78) for males and (1.21, 2.04) for females. The standard error for the sojourn time is 0.144 for CT overall, 0.185 for CT male and 0.221 for CT female.

### Discussion

In this paper, the three key parameters, screening sensitivity, the transition probability density and the sojourn time distribution, were estimated using Bayesian approach. The NLST CT arm data have been used for the estimation.

For lung cancer, the estimated sensitivity was 56.8% for JHLP control group data, where only X-ray screenings were administered, from the study of Jang et al. [1]. Kim et al. [8] estimated the sensitivity as 79.9% using the JHLP study group data, in which both X-ray and sputum cytology were used. By using Mayo Lung Project male heavy smokers data, Wu et al. [6] estimated combined X-rays and sputum cytology sensitivity is 89.4%. Chen et al. [9] estimated the screening sensitivity of sputum cytology as a supplement to the chest X-ray using MSKC-LCSP data was 86.64%. Compared with these previous results, the sensitivity estimated in this study was around 95% for all the groups, which is much larger. This confirms that CT scan improves the lung cancer screening sensitivity compared to X-rays. In addition, it seems that the sensitivity of lung cancer screening using CT scan does not depend on the age of patients. For the NLST data CT arm, Pinsky et al. [10] and Aberle et al. [11] estimated the sensitivity was 93.5% and 94.4%, respectively, which is also closed to our sensitivity estimation.

<table>
<thead>
<tr>
<th>Age</th>
<th>Sensitivity $\beta$</th>
<th>Transition probability for $w(t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>Mean</td>
</tr>
<tr>
<td>55</td>
<td>0.9642</td>
<td>0.9551</td>
</tr>
<tr>
<td>60</td>
<td>0.9657</td>
<td>0.9581</td>
</tr>
<tr>
<td>65</td>
<td>0.9642</td>
<td>0.9587</td>
</tr>
<tr>
<td>70</td>
<td>0.9616</td>
<td>0.9570</td>
</tr>
<tr>
<td>75</td>
<td>0.9613</td>
<td>0.9529</td>
</tr>
<tr>
<td>55</td>
<td>0.9484</td>
<td>0.9360</td>
</tr>
<tr>
<td>60</td>
<td>0.9496</td>
<td>0.9398</td>
</tr>
<tr>
<td>65</td>
<td>0.9497</td>
<td>0.9396</td>
</tr>
<tr>
<td>70</td>
<td>0.9495</td>
<td>0.9355</td>
</tr>
<tr>
<td>75</td>
<td>0.9506</td>
<td>0.9274</td>
</tr>
<tr>
<td>55</td>
<td>0.9601</td>
<td>0.9499</td>
</tr>
<tr>
<td>60</td>
<td>0.9641</td>
<td>0.9563</td>
</tr>
<tr>
<td>65</td>
<td>0.9665</td>
<td>0.9599</td>
</tr>
<tr>
<td>70</td>
<td>0.9666</td>
<td>0.9610</td>
</tr>
<tr>
<td>75</td>
<td>0.9710</td>
<td>0.9596</td>
</tr>
</tbody>
</table>
Figure 8: Posterior quantiles (5%, 50% and 95%) of sensitivity and transition probability for CT female group.

The transition probability from disease-free to preclinical state has a peak around age 70 for both males and females. The transition probability also has a peak around age 70 from Chen’s study [9]. The “SEER Fast Fact Stats” [2] show that the highest percent of new lung cancer cases is in 65-74 age group. Our results are consistent with that fact.

In the Mayo Lung Project study, the mean sojourn time was 2.2 years [6], the mean sojourn time for male heavy smokers in MSKC-LCSP data is about 3.35 years. The posterior mean sojourn in this study is around 1.5 years for both gender groups in this study. Since these two studies were carried out about one or two decades ago, it maybe that today’s heavy smokers have a shorter sojourn time. That is, the tumor grows faster than before to present clinical symptoms, and makes it harder to catch the disease during the preclinical state.

Conclusion

In summary, this project focuses on the estimation of the three key parameters: sensitivity, sojourn time distribution and transition probability density from the disease-free to the preclinical state, to lay a foundation for the estimation of other interesting terms, such as lead time, over diagnosis, long term outcomes in the future, because all these interesting terms can be expressed as a function of the three key parameters.

Conflict of Interest Declaration

We authors declare here that we have no conflict of interest with other researchers.

Acknowledgements

We authors thank the National Cancer Institute’s Cancer Data Access System for allowing us to use the NLST data. This work was conducted in part using the resources of the University of Louisville’s research computing group and the Cardinal Research Cluster.

References

2. SEER Fast Stats Results, NIH.