

Projection of long-term outcomes using X-rays and pooled cytology in lung cancer screening

Dongfeng Wu¹

Diane Erwin²

Seongho Kim¹

¹Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, Louisville, KY,

²Information Management Services, Inc, Rockville, MD, USA

Purpose: To investigate the long-term outcome of using X-ray and the 3-day “pooled” sputum cytology in lung cancer screening using the Mayo Lung Project data.

Methods: In this study, the statistical method developed by the authors using the Mayo Lung Project data was applied. Each participant eventually falls into one of the four categories depending on whether they would be diagnosed with lung cancer and whether symptoms would have appeared before death. The derived probability of each outcome was presented. Bayesian inference was made for the percentages of true-early-detection, no-early-detection, over-diagnosis, and symptom-free-life. The authors present estimates of these proportions for male heavy smokers by using Markov Chain Monte Carlo samples in the simulation. Human lifetime was treated as a random variable.

Results: The probability that a male heavy smoker will live a lung cancer-free life is about 80%; the probability for developing lung cancer for a male heavy smoker is about 20%, much higher than in the general population (6.95%). The probability of over-diagnosis is low, about 6%–9% among the early detected cases. The probability of true-early-detection is greater than 90%. Finally, the probability of no-early-detection increases as the screening interval increases; among the diagnosed cases, it is about 15% if the screens are 6 months apart, and it is close to 50% if the screens are 2 years apart.

Conclusion: This research provides a framework and a systematic approach for policy makers regarding how to evaluate long-term outcomes of chest X-ray and pooled sputum cytology in lung cancer screening.

Keywords: true early detection, no early detection, over diagnosis, symptom-free life, pooled sputum cytology

Introduction

Lung cancer is the most common form of cancer in the world; in the USA, the incidence of lung cancer peaked in the 1970s and then leveled off, probably because the smoking rate was higher during World War II.^{1,2} It is estimated that in 2010, in the USA, 116,750 men and 105,770 women will be diagnosed with lung cancer, and 157,300 people will die of it.² There are two major types of lung cancer, small cell and non-small cell, and each type is treated differently.³ The age-specific lung cancer incidence rate rises with advancing age and reaches its peak between 65 and 74.²

Cigarette smoking is the major cause of lung cancer; other environmental health hazards include long-term exposure to radon or other carcinogens in polluted air. It may take years for lung cancer cells to develop, and usually people do not have any symptoms until the disease has progressed to a late stage.³ As a result, less

Correspondence: Dongfeng Wu
Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, Louisville, KY 40202, USA
Tel +1 502 852 1888
Fax +1 502 852 3294
Email dongfeng.wu@louisville.edu

than 15% of lung cancers are discovered in early stages when the possibility of curative treatment is the greatest.¹ However, there is no evidence of significant reduction in lung cancer mortality due to early detection.³ A recent large-scale screening study for lung cancer, the National Lung Screening Trial (NLST), is comparing spiral computed tomography (CT) scans with chest X-rays among high-risk people. Early results from the NLST study, announced in November 2010, found that people who received spiral CT had 20% lower chance of dying from lung cancer than those who received chest X-rays.^{3,4} However, there is no recommendation for regular screening for lung cancer so far, even for people at high risk. After all, survival depends not only on early detection but also on effective treatment following detection. That is, early detection may not directly lead to longer survival. In this paper, the long-term outcome of lung cancer screening was evaluated from another perspective. All initially superficially healthy males who planned to be screened for lung cancer were categorized into four mutually exclusive groups: (1) true-early-detection, (2) no-early-detection, (3) over-diagnosis, and (4) symptom-free-life. Which category a participant would be in depended on whether he would be diagnosed with lung cancer, and whether he would die from this cause.⁵

The Mayo Lung Project was started in August 1971. Its purpose was to determine whether the death rate from bronchogenic carcinoma can be reduced significantly by vigorous application of modern detecting techniques and aggressive treatment.⁶ Between 1971 and 1983, 9309 candidates for the Mayo Lung Project had been enrolled in the study. They were all male heavy smokers. As a group, 85% were smoking between 1 and 2.5 packs of cigarettes per day. More than 97% had smoked for at least 20 years. More than 90% of the participants were still smoking regardless of the warnings that they received. Each participant took a screening test every 4 months, amounting to 19 tests altogether. Each screening test included a chest X-ray and a 3-day pooled sputum cytology sampling. If any of the tests was positive, then the screen was considered positive and a definitive work-up exam, such as biopsy, was done. The data used included the total number of participants in each screening exam, the number of detected and confirmed cancer cases in each screening exam, and the number of interval cases. These data were stratified by age at entry. The age at entry ranged from 44 to 76 years old in the Mayo Lung Project. However, only the data from age 45 to age 69 was used because the

other age groups had too few participants, which may have caused large bias in the estimation.

It was assumed that clinical lung cancer develops through three states: $S_0 \rightarrow S_p \rightarrow S_c$, where S_0 represents the disease-free state, S_p represents the preclinical disease state in which an asymptomatic individual unknowingly has disease that a screening exam can detect, and S_c represents the clinical state when the disease manifests itself in clinical symptoms. If a person enters the preclinical state, S_p , at age t_1 , and their clinical symptoms present later at age t_2 , then $(t_2 - t_1)$ is the sojourn time in the preclinical state. If they are offered a screening exam at time t within the interval (t_1, t_2) , and cancer is diagnosed, then the length of the time $(t_2 - t)$ is the lead time, the length of time the diagnosis is advanced by screening. The goal of screening is to detect lung cancer in the preclinical state; that is, when a person has a tumor but there are no symptoms. There are two other key parameters involved in the screening program: “sensitivity” of the screening technique and “transition probability” from disease-free to preclinical state. The sensitivity is the probability that the screening exam is positive, given that the individual is in the preclinical stage. The transition probability into the preclinical stage is the probability density function (PDF) of making a transition from the disease-free state to the preclinical state.

The three elements, sensitivity, sojourn time distribution, and transition probability density, are the three key parameters in cancer screening. Many other features can be expressed as a function of them; for example, the lead time distribution is a function of the three parameters, and the probability of true-early-detection and the probability of over-diagnosis are also functions of the three parameters. Hence accurate estimations of these three parameters are very important.⁷

Methods

All initially superficially healthy people who entered the screening program were, at time of death, categorized in the following way:⁵

- Outcome 1 (symptom-free-life [SympF]) – a male heavy smoker who took part in screening exams that never found lung cancer and ultimately died of other causes.
- Outcome 2 (no-early-detection [NoED]) – a male heavy smoker who took part in screening exams but whose disease manifested itself clinically and was not detected by scheduled screening.
- Outcome 3 (true-early-detection [TrueED]): a male heavy smoker whose lung cancer was diagnosed at a scheduled

screening exam and whose clinical symptoms would have appeared before his death.

- Outcome 4 (over-diagnosis [OverD]) – a male heavy smoker who was diagnosed with lung cancer at a scheduled screening exam but whose clinical symptoms would not have appeared before his death.

Here, the method developed by Wu and Rosner⁴ for the evaluation of the long-term outcomes is briefly reviewed. Define $\beta(t)$ to be the screening sensitivity at age t ; that is, the probability of a positive screening result if the individual is in the preclinical state, where t is the individual's age at the exam. Define $w(t)$ as the PDF of a transition from S_0 to S_p at age t . Let $q(x)$ be the 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. Throughout this paper, the time variable t represents an individual's age. The capital letter T represents a person's lifetime, which is a continuous random variable with a PDF of $f_T(t)$.

The probability for each of the four $Q(z) = \int_z^\infty q(x)dx$ categories was derived when human lifetime is a random variable, and hence subject to competing risks.⁴ Suppose an individual has taken K screening exams at his age $t_0 < t_1 < \dots < t_{K-1}$, given that his lifetime $T = t_K (> t_{K-1})$, then:

$$P(\text{Outcome1: SympF} | T = t_K) = 1 - \int_0^{t_K} w(x)dx + \int_{t_{K-1}}^{t_K} w(x)Q(t_K - x)dx + \sum_{j=0}^{K-1} (1 - \beta_j) \dots (1 - \beta_{K-1}) \times \int_{t_{j-1}}^{t_j} w(x)Q(t_K - x)dx \quad (1)$$

$P(\text{Outcome2: NoED} | T = t_K) = I_{K,1} + I_{K,2} + \dots + I_{K,K}$, where for $j = 1, \dots, K$,

$$I_{K,j} = \sum_{i=0}^{j-1} (1 - \beta_i) \dots (1 - \beta_{j-1}) \times \int_{t_{i-1}}^{t_i} w(x)[Q(t_{j-1} - x) - Q(t_j - x)]dx + \int_{t_{j-1}}^{t_j} w(x)[1 - Q(t_j - x)]dx, \quad (2)$$

$$P(\text{Outcome3: TrueED} | T = t_K) = \sum_{j=1}^{K-1} \beta_j \left\{ \sum_{i=0}^{j-1} (1 - \beta_i) \dots (1 - \beta_{j-1}) \times \int_{t_{i-1}}^{t_i} w(x)[Q(t_j - x) - Q(t_K - x)]dx + \int_{t_{j-1}}^{t_j} w(x)[Q(t_j - x) - Q(t_K - x)]dx \right\} + \beta_0 \int_0^{t_0} w(x)[Q(t_0 - x) - Q(t_K - x)]dx \quad (3)$$

$$P(\text{Outcome4: OverD} | T = t_K) = \sum_{j=1}^{K-1} \beta_j \left\{ \sum_{i=0}^{j-1} (1 - \beta_i) \dots (1 - \beta_{j-1}) \times \int_{t_{i-1}}^{t_i} w(x)Q(t_K - x)dx + \int_{t_{j-1}}^{t_j} w(x)Q(t_K - x)dx \right\} + \beta_0 \int_0^{t_0} w(x)Q(t_K - x)dx. \quad (4)$$

For an individual at age t_0 , the number of future screens is unknown; however, if they plan to follow a prefixed screening schedule, $t_0 < t_1 < \dots < t_K < \dots$, then the probability of each outcome can be obtained by doing the integration:

$$P(\text{Outcome } i | T \geq t_0) = \int_{t_0}^\infty P(\text{Outcome } i | K = K(t), T = t) f_T(t | T \geq t_0) dt, \quad i = 1, 2, 3, 4, \quad (5)$$

where the probability for "Outcome i " was derived in equations (1) to (4), and the probability for their lifetime is defined as $f_T(t | T \geq t_0) = f_T(t) / 1 - F_T(t_0)$ for $t \geq t_0$, and is 0 for $t < t_0$. Let event A be defined as being asymptomatic (or superficially healthy) of lung cancer before t_0 . Then:

$$P(A | T \geq t_0) = 1 - \int_0^{t_0} w(x)dx + \int_0^{t_0} w(x)Q(t_0 - x)dx \quad (6)$$

and

$$P(\text{Outcome } i | T \geq t_0, A) = \frac{P(\text{Outcome } i | T \geq t_0)}{P(A | T \geq t_0)} \quad (7)$$

It was proved that:

$$\sum_{i=1}^4 P(\text{Outcome } i | A, T \geq t_0) = 1. \quad (8)$$

That is, the probabilities for these four outcomes always add up to 1, given that an individual is asymptomatic before they take the first exam. The probability $P(\text{Outcome } i | A, T \geq t_0)$, for $i = 1, 2, 3, 4$, using simulation is reported in the following section.

To make predictive inference for lung cancer screening using chest X-ray, accurate estimation of the three key parameters, sensitivity, sojourn time distribution, and transition probability density, is first required because the derived probability is a function of these parameters.^{6,7} These parameters were estimated using Bayesian inference

and the Mayo Lung Project data; the results were published in the authors' earlier works.^{7,8} The parametric models are:

$$\beta(t) = \beta = \frac{1}{1 + \exp(-b_0)} \quad (9)$$

$$w(t) = \frac{0.3}{\sigma t \sqrt{2\pi}} \exp \left\{ -\frac{(\log t - \mu)^2}{2\sigma^2} \right\} \quad (10)$$

$$q(x) = \frac{\kappa x^{\kappa-1} \rho^{\kappa}}{[1 + (x\rho)^{\kappa}]^2}, \quad x > 0, \quad (11)$$

The transition density function is a sub-PDF, with 0.3 as its upper bound, because not everyone will develop lung cancer in their lifetime, and according to the National Cancer Institute (NCI) "SEER Fast Fact Stats" database,^{9,10} the lifetime risk of being diagnosed with lung and bronchus cancer is about 6.94% for males. According to Villeneuve and Mao,¹⁰ the lifetime risk for male smokers is 17.2%. Since the Mayo Lung Project participants were male heavy smokers, the risk should be much higher than that. Therefore, in this paper, 30% was chosen as a reasonable upper limit. The parameters that need to be estimated in the above model are $\theta = (b_0, \mu, \sigma^2, \kappa, \rho)$. For detailed justifications on how these age effect functions were chosen, see Wu et al.^{7,8}

Results

Markov Chain Monte Carlo (MCMC) was used to generate a random posterior sample from the joint posterior distribution of the parameters for a Bayesian inference.⁷ The posterior simulation was partitioned into three subchains. Gibbs sampling was then used to sample the posteriors for b_0 , (μ, σ^2) , (κ, ρ) separately. Each MCMC simulation was run for 30,000 steps, with a burn-in of 5000 steps. After the burn-in time, the posteriors were sampled every 100 steps, providing 250 posterior samples for the parameter vector θ . Four chains were simulated, each with different starting values that were overdispersed with respect to the target distribution. Bayesian output diagnosis showed that the chains had converged. The 250 posterior samples from each of the four chains were pooled for the analysis, giving a total of 1000 posterior samples θ_i^* , $i = 1, \dots, 1000$. The posterior estimates for parameters θ and the standard errors are listed in Table 1 in Wu et al.⁷

The posterior samples θ_i^* were then used to estimate the probabilities for each category. Given the Mayo Lung Project data, the posterior predictive probability can be estimated by:

Table 1 Projected probability of each category using Mayo Lung Project data

Δ , ^a months	Mean probability (SD), %			
	SympF	NoED	TrueED	OverD
Age at first screening $t_0 = 45$				
6	80.50 (0.90)	2.95 (0.55)	15.47 (0.79)	1.01 (0.20)
12	80.61 (0.91)	5.89 (0.85)	12.54 (0.72)	0.90 (0.19)
18	80.70 (0.91)	8.02 (0.98)	10.40 (0.68)	0.81 (0.17)
24	80.77 (0.92)	9.57 (1.03)	8.85 (0.64)	0.74 (0.16)
Age at first screening $t_0 = 50$				
6	80.55 (0.94)	2.93 (0.54)	15.47 (0.82)	1.04 (0.21)
12	80.66 (0.95)	5.84 (0.85)	12.55 (0.75)	0.92 (0.19)
18	80.76 (0.95)	7.96 (0.98)	10.43 (0.71)	0.83 (0.18)
24	80.83 (0.95)	9.50 (1.03)	8.89 (0.66)	0.76 (0.16)
Age at first screening $t_0 = 55$				
6	81.11 (0.96)	2.79 (0.51)	14.96 (0.84)	1.06 (0.21)
12	81.25 (0.96)	5.57 (0.80)	12.18 (0.78)	0.92 (0.19)
18	81.32 (0.97)	7.59 (0.93)	10.16 (0.73)	0.85 (0.18)
24	81.40 (0.97)	9.06 (0.98)	8.69 (0.68)	0.76 (0.17)
Age at first screening $t_0 = 60$				
6	82.53 (0.91)	2.52 (0.46)	13.80 (0.81)	1.13 (0.22)
12	82.66 (0.91)	5.03 (0.71)	11.30 (0.77)	1.00 (0.21)
18	82.76 (0.91)	6.86 (0.81)	9.47 (0.72)	0.90 (0.19)
24	82.83 (0.91)	8.19 (0.86)	8.13 (0.68)	0.83 (0.18)

Notes: Each row should add up to 100%, however, due to simulation accuracy, it is not exactly 100%; ^a $\Delta = t_i - t_{i-1}$ is the proposed time interval between screenings.

Abbreviations: Outcomes: NoED, no-early-detection; OverD, over-diagnosis; SympF, symptom-free-life; TrueED, true-early-detection; SD, standard deviation.

$$P(\text{Outcome } i | T \geq t_0, A, \text{Data}) = \frac{1}{n} \sum_{j=1}^n P(\text{Outcome } i | T \geq t_0, A, \theta_j^*) \quad (12)$$

where $P(\text{Outcome } i | T \geq t_0, A, \theta_j^*)$ represents the probability for each outcome as defined in equation (7), and θ_j^* is the 1000 posterior samples from the MCMC simulation.

The above method was applied to make predictive inference in the case of a screening program consisting of periodic lung screening tests for male heavy smokers. If it is assumed that there are four cohorts with ages 45, 50, 55, and 60 at the initial screening exam, then the probability for each category was calculated under different screening intervals: 6, 12, 18, and 24 months. Since lifetime is a random variable, the distribution of lifetime was obtained using the actuarial life table from the Social Security Administration, published online.¹² For details about how to transform the period life table into the PDF, please see the appendix of Wu and Rosner.⁵ The conditional PDF for the male lifetime is plotted in Figure 1. The simulation results are summarized in Table 1.

The probability of SympF is quite high, above 80% for all age groups, and is almost constant as the screening interval changes. This could be due to the fact that about 80% of male heavy smokers will not develop clinical lung cancer before

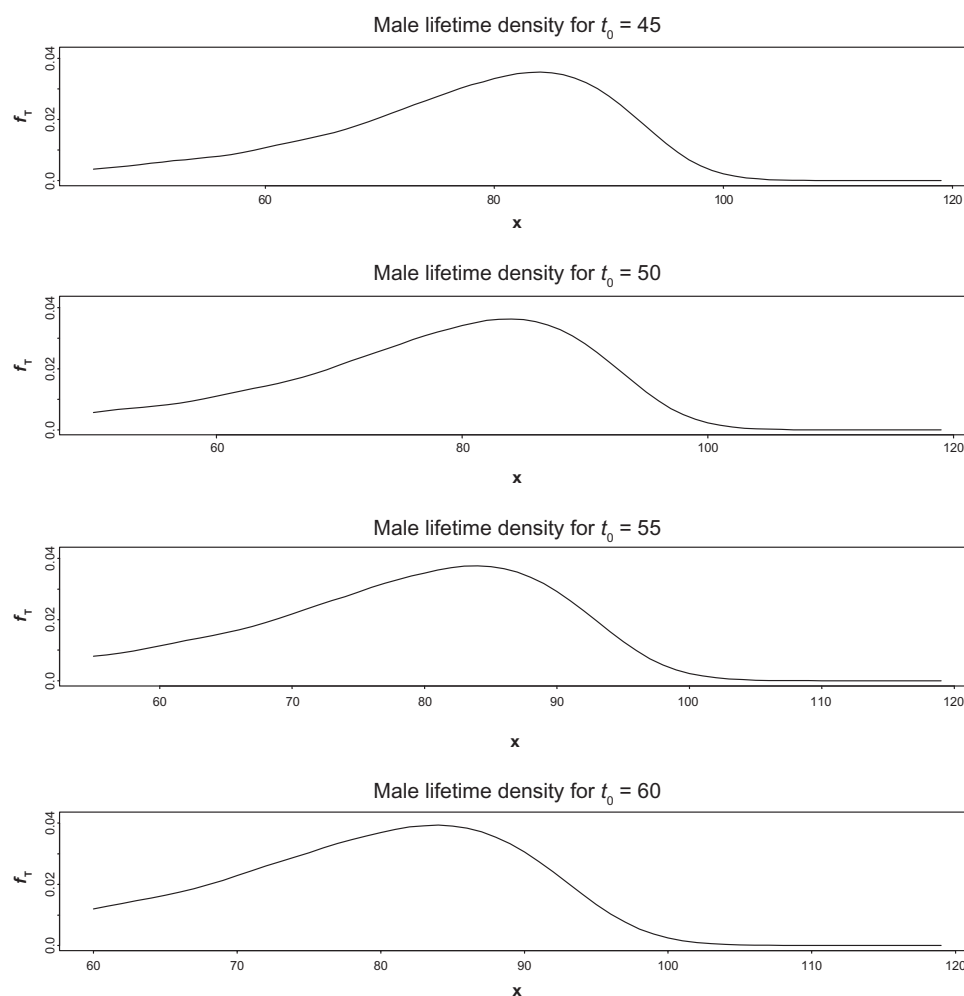


Figure 1 The conditional probability density function for male lifetime $f_T(t|T \geq t_0)$.

death, or about 80% of male heavy smokers die from causes other than lung cancer. The probability of NoED increases as the screening interval increases, ranging from 2.5% to 9.5% for all cohorts. The probability of TrueED decreases as the screening interval increases, from 15.5% to about 8%. The probability of OverD is low; it is barely over 1% for all hypothetical cohorts. The standard deviations are reported in parentheses in Table 1.

The trend of the probability for different screening intervals is clearly shown in Figure 2. The probability of SympF and OverD does not vary much over different screening intervals, while the probability of TrueED and NoED show the opposite directions as the screening interval increases.

Table 2 summarizes the predictive conditional probability of NoED, TrueED, and OverD for all diagnosed cancer cases. The probability for each column shows the same pattern as in Table 1 but in larger magnitude. For example, for the cohort who is age 50 years at initial screening, the probability of

TrueED is about 80% if screened every 6 months, given an eventual lung cancer diagnosis; whereas the probability will drop to 65% if screened annually. The corresponding standard errors are reported in parentheses.

Table 3 reports the predictive probability of TrueED and OverD among screen-detected cases. These are the probabilities that most researchers are eager to explore and discuss. It shows that among the screen-detected cases, only about 6%–9% are OverD, which means, if left untreated, 6%–9% of patients may die of other causes before clinical symptoms show up. More than 90% of the screen-detected cases are TrueED that need treatment. The 95% credible intervals (CIs) are reported in Table 3 as well, ranging from 83% to 95% for the probability of TrueED and 4% to 12% for OverD.

Discussion

In this study, the probability calculation method in Wu and Rosner⁴ was applied to the Mayo Lung Project study and

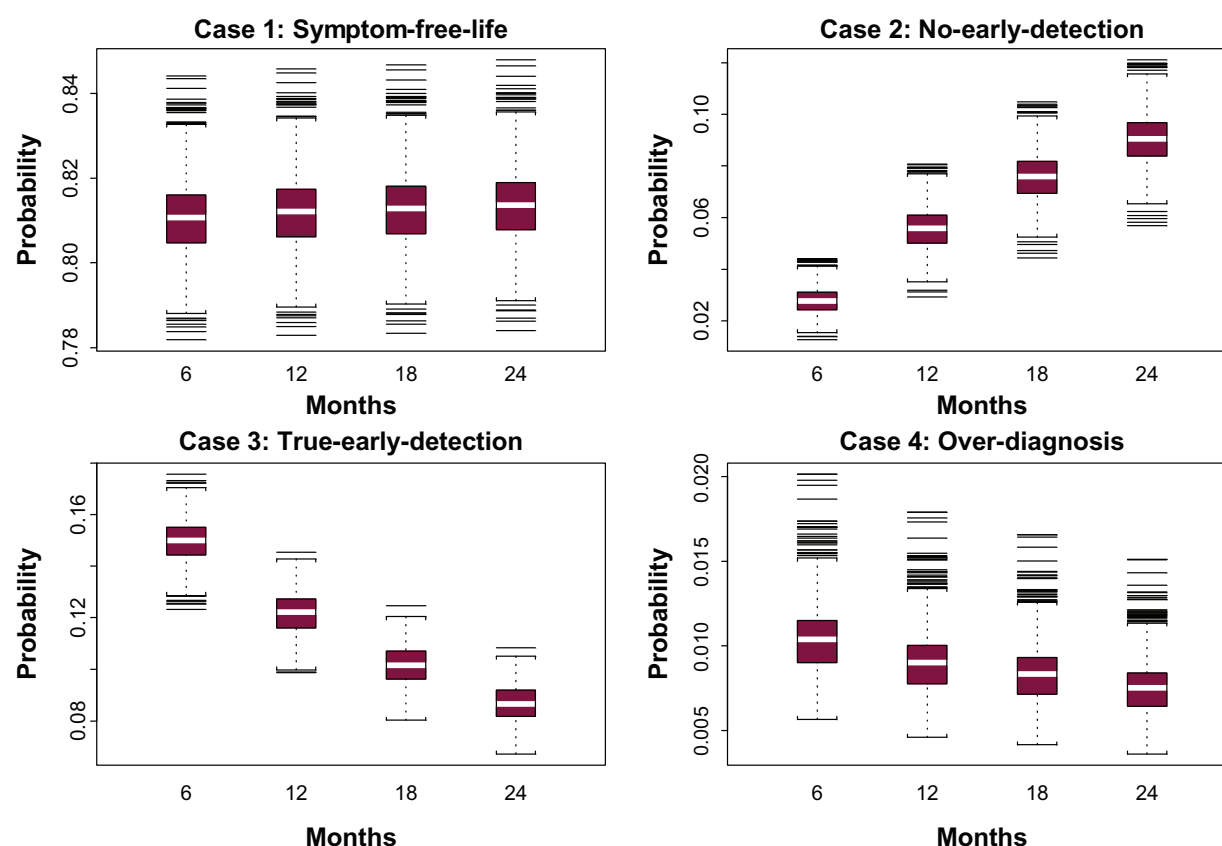


Figure 2 The trend of the probability for each category for the cohort with initial age $t_0 = 55$.

Table 2 The projected probability given that it is a diagnosed cancer case^a

Δ , ^b months	Mean probability ^c (SE), %		
	NoED	TrueED	OverD
Age at first screening $t_0 = 45$			
6	15.14 (2.51)	79.61 (2.00)	5.24 (1.14)
12	30.41 (3.72)	64.93 (3.01)	4.67 (1.05)
18	41.62 (4.04)	54.14 (3.30)	4.24 (0.98)
24	49.86 (4.01)	46.24 (3.28)	3.90 (0.92)
Age at first screening $t_0 = 50$			
6	15.03 (2.50)	79.59 (1.99)	5.38 (1.15)
12	30.18 (3.71)	65.02 (2.99)	4.80 (1.07)
18	41.33 (4.04)	54.34 (3.29)	4.33 (1.00)
24	49.51 (4.01)	46.48 (3.27)	4.01 (0.94)
Age at first screening $t_0 = 55$			
6	14.82 (2.48)	79.55 (1.98)	5.63 (1.20)
12	29.79 (3.69)	65.28 (2.96)	4.93 (1.10)
18	40.76 (4.03)	54.66 (3.25)	4.58 (1.04)
24	48.89 (4.01)	46.96 (3.25)	4.15 (0.97)
Age at first screening $t_0 = 60$			
6	14.45 (2.45)	79.09 (1.94)	6.46 (1.30)
12	29.01 (3.66)	65.20 (2.87)	5.79 (1.21)
18	39.77 (4.02)	54.97 (3.18)	5.26 (1.13)
24	47.70 (4.02)	47.44 (3.19)	4.86 (1.06)

Notes: ^aDiagnosed cancer case = interval-incident cases and screen-detected cases; ^b $\Delta = t_i - t_{i-1}$ is the proposed time interval between screenings; ^cthe estimated probability was calculated by $p_i^*/(p_i^* + p_j^* + p_k^*)$, $i = 2, 3, 4$ for each of the 1000 posterior Markov Chain Monte Carlo samples, then taking the average.

Abbreviations: Outcomes: NoED, no-early-detection; OverD, over-diagnosis; TrueED, true-early-detection; SE, standard error.

Table 3 The probability of over-diagnosis and true-early-detection for the screen-detected cases

Δ ^a	Mean probability ^b (95% CI), %	
	TrueED	OverD
Age at first screening $t_0 = 45$		
6	93.84 (90.79–95.76)	6.16 (4.24–9.21)
12	93.34 (90.24–95.31)	6.66 (4.69–9.76)
18	92.80 (89.66–94.81)	7.20 (5.19–10.34)
24	92.29 (89.12–94.39)	7.71 (5.61–10.88)
Age at first screening $t_0 = 50$		
6	93.68 (90.63–95.64)	6.31 (4.36–9.37)
12	93.17 (90.02–95.17)	6.83 (4.83–9.98)
18	92.68 (89.45–94.73)	7.32 (5.27–10.55)
24	92.13 (88.92–94.25)	7.87 (5.75–11.08)
Age at first screening $t_0 = 55$		
6	93.41 (90.23–95.41)	6.59 (4.59–9.77)
12	93.02 (89.77–95.08)	6.98 (4.92–10.23)
18	92.33 (89.01–94.46)	7.67 (5.54–10.99)
24	91.96 (88.64–94.13)	8.04 (5.87–11.36)
Age at first screening $t_0 = 60$		
6	92.47 (89.08–94.65)	7.53 (5.35–10.92)
12	91.89 (88.46–94.12)	8.11 (5.88–11.54)
18	91.32 (87.84–93.58)	8.68 (6.42–12.16)
24	90.78 (87.27–93.09)	9.22 (6.91–12.73)

Notes: ^a $\Delta = t_i - t_{i-1}$ is the proposed time interval between screenings; ^bthe estimated probability was calculated by $p_i^*/(p_i^* + p_j^* + p_k^*)$, $i = 3, 4$ for each of the 1000 posterior Markov Chain Monte Carlo samples, then taking the average.

Abbreviations: Outcomes: OverD, over-diagnosis; TrueED, true-early-detection; CI, credible interval.

some useful information regarding screening among male heavy smokers was obtained. This research can provide policy makers with important estimates of the probability of TrueED, NoED, OverD, and other outcome categories that result from a periodic screening program. Bayesian analysis was used because it can incorporate uncertainty, and it is easy to calculate the variations and the CIs of the percentages.

According to the NCI SEER database,¹⁰ the lifetime risk for lung and bronchus cancer for both genders during their lifetime is 6.95% for all races, with a 95% CI (6.91%–6.99%). In other words, the accepted lifetime risk of lung cancer is 1/14 during one's lifetime for the general population. The authors' estimated probability of SympF is about 80% for male heavy smokers; that is, male heavy smokers have a lifetime risk of 20% for lung cancer, which is about three times larger than the accepted lifetime risk for the general population.

The proportion of OverD among the screen-detected cases is very small, about 6%–9% among all age groups, showing that more than 90% of the screen-detected cases are TrueED cases and immediate treatment is needed. The ultimate goal of a screening program in cancer is to reduce cancer mortality. Reduction in cancer mortality is discussed in Fontana et al.⁶ However, no reduction in lung cancer mortality has been confirmed from screening, except some encouraging news from the NLST recently. That will be the next project for the authors of this paper: to explore whether early detection of lung cancer may or may not contribute to patient survival.

The model used in this study contributes to the study of a screening program by providing a framework for the evaluation of long-term effects. The model can be used to evaluate and compare the characteristics of different cohorts under different screening frequencies. For example, the model can be applied to data obtained from female heavy smokers, male nonsmokers, or female nonsmokers and provide answers to questions such as: What is the percentage of SympF in the cohort? What is the percentage of TrueED versus OverD among the screen-detected cases? What is the percentage of NoED for different screening schedules? The authors of this paper hope policy makers will be able

to use the information in their decision making regarding public health issues.

Acknowledgment

The authors thank Dr Philip Prorok of the NCI for allowing access to the Mayo Lung Project data.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Cancer Research UK. Lung cancer and smoking statistics – key facts [updated 2011 Feb 14]. <http://info.cancerresearchuk.org/cancerstats/keyfacts/lung-cancer/>. Accessed July 1, 2011.
2. Altekruse SF, Kosary CL, Krapcho M, et al, editors. *SEER Cancer Statistics Review, 1975–2007*. Bethesda, MD: National Cancer Institute. Available from: http://seer.cancer.gov/csr/1975_2007/, based on November 2009 SEER data submission, posted to the SEER web site, 2010.
3. American Cancer Society. *Cancer Facts and Figures 2011*. New York, NY: American Cancer Society; 2011. Available from: <http://www.cancer.org/acs/groups/content/@epidemiologysurveillance/documents/document/acspc-029771.pdf>. Accessed July 15, 2011.
4. The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011. Epub 2011 Jun 29. Available from: <http://www.nejm.org/doi/full/10.1056/NEJMoa1102873#t=article>. Accessed July 19, 2011.
5. Wu D, Rosner GL. A projection of true-early-detection, no-early-detection, over-diagnosis and not-so-necessary probabilities in tumor screening. Proceedings of the American Statistical Association, Biopharmaceutical Section. Alexandria, VA: American Statistical Association; 2010:1144–1157.
6. Fontana RS, Sanderson DR, Woolner LB, et al. The Mayo Lung Project for early detection and localization of bronchogenic carcinoma: a status report. *Chest*. 1975;67:511–522.
7. Wu D, Erwin D, Rosner GL. Sojourn time and lead time projection in lung cancer screening. *Lung Cancer*. 2011;72:322–326.
8. Wu D, Rosner GL, Broemeling LD. MLE and Bayesian inference of age-dependent sensitivity and transition probability in periodic screening. *Biometrics*. 2005;61(4):1056–1063.
9. National Cancer Institute. SEER fast stats results: lung and bronchus [webpage on the Internet]. Bethesda, MD: National Cancer Institute; nd. Available from: <http://seer.cancer.gov/statfacts/html/lungb.html>. Accessed July 15, 2011.
10. National Cancer Institute. SEER Cancer Statistics Review 1975–2007 [2010]. Available from: http://seer.cancer.gov/csr/1975_2007/results_merged/topic_lifetime_risk.pdf. Accessed July 1, 2011.
11. Villeneuve PJ, Mao Y. Lifetime probability of developing lung cancer, by smoking status, Canada. *Can J Public Health*. 1994;85(6):385–388.
12. Social Security Online. Actuarial Publications. Period Life Table [updated 2011 Apr 5]. Woodlawn, MD: US Social Security Administration; 2011. Available from: <http://www.ssa.gov/OACT/STATS/table4c6.html>. Accessed July 1, 2011.

Open Access Medical Statistics

Publish your work in this journal

Open Access Medical Statistics is an international, peer-reviewed, open access journal publishing original research, reports, reviews and commentaries on all areas of medical statistics. The manuscript management system is completely online and includes a very quick and fair

Submit your manuscript here: <http://www.dovepress.com/open-access-medical-statistics-journal>

Dovepress

peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.