Manuscript Title: Estimation of the tumor size at cure threshold among aggressive non-small cell lung cancers (NSCLCs): Evidence from The Surveillance, Epidemiology, and End Results Program (SEER) and the National Lung Screening Trial (NLST).

Short Title: Tumor Size at NSCLC Cure Threshold: SEER and NLST.

Author: Deborah L. Goldwasser, PhD

Contact Information:

Address: Florida International University, Department of Mathematics and Statistics, 11200 SW 8th Street, Miami, FL 33199.

Phone: 305-725-9893

Email: dgoldwas@fiu.edu

Key Words: National Lung Screening Trial, Tumor Size, Simulation Model, Computed Tomography, Lung Cancer Screening

Abbreviations: National Lung Screening Trial (NLST), Non-Small Cell Lung Cancer (NSCLC), Surveillance, Epidemiology and End Results (SEER), Computed Tomography (CT), U.S. Preventive Services Task Force (USPSTF), Volume doubling time (VDT), Maximum Likelihood Estimation (MLE), Chest X-Ray (CXR), Extent of Disease (EOD), International Early Lung Cancer Action Program (I-ELCAP).

Novelty and Impact: The mortality benefit of CT screening for lung cancer is limited because the majority of aggressive NSCLCs are detected above their tumor size cure thresholds. We described novel statistical methodology that allowed us to estimate the

NSCLC cure threshold distribution for the most aggressive NSCLCs detected in the NLST and documented in SEER. We concluded that SEER and NLST data were most consistent with a median cure threshold for aggressive NSCLCs between 10 mm-15 mm.

Abstract:

The National Lung Screening Trial (NLST) demonstrated that non-small cell lung cancer (NSCLC) mortality can be reduced by a program of annual CT screening in high-risk individuals. However, CT screening regimens and adherence vary, potentially impacting the lung cancer mortality benefit. We defined the NSCLC cure threshold as the maximum tumor size at which a given NSCLC would be curable due to early detection. We obtained data from 518,234 NSCLCs documented in the U.S. SEER cancer registry between 1988 and 2012 and 1,769 NSCLCs detected in the NLST. We demonstrated mathematically that the distribution function governing the cure threshold for the most aggressive NSCLCs, $G(x|\Phi=1)$, was embedded in the probability function governing detection of SEER-documented NSCLCs. We determined the resulting probability functions governing detection over a range of $G(x|\Phi=1)$ scenarios and compared them with their expected functional forms. We constructed a simulation framework to determine the cure threshold models most consistent with tumor sizes and outcomes documented in SEER and the NLST.

Whereas the median tumor size for lethal NSCLCs documented in SEER is 43mm (males) and 40mm (females), a simulation model in which the median cure threshold for the most aggressive NSCLCs is 10mm (males) and 15mm (females) best fit the SEER

and NLST data. The majority of NSCLCs in the NLST were treated at sizes greater than our median cure threshold estimates. New technology is needed to better distinguish and treat the most aggressive NSCLCs when they are small (i.e. 5-15mm).

1. Introduction:

The National Lung Screening Trial (NLST) was the first randomized clinical trial to achieve a significant reduction in lung cancer mortality attributable to CT screening[1]. After three annual rounds of screening and a median of 6.5 years of follow-up, a 20.0% reduction in lung cancer mortality was observed in the CT arm relative to the chest x-ray arm. In contrast, there was no reduction in lung cancer mortality attributable to chest x-ray screening relative to usual care in the PLCO trial[2], consistent with earlier findings of the Mayo Lung Project[3]. The U.S. Preventive Services Task Force (USPSTF) has assigned a grade of B to lung screening by CT of high-risk individuals between the ages of 55-80[4]. Under the Affordable Care Act, Medicare is required to reimburse cancer screening services that receive a grade of B or higher.

In contrast with the United States, lung cancer screening by CT has not been universally endorsed across the European Union and the United Kingdom. Three small randomized CT studies in Europe did not produce a mortality benefit[5,6,7] and the larger randomized, Dutch-NELSON study mortality and survival results have not yet been reported[8]. Estimates of a mortality reduction attributable to CT screening have ranged from 0%[9] to as high as 64%[10]. Simulation models have closely approximated the NLST results[11]. Differences in screening regimens and adherence may influence mortality outcomes. A comprehensive review highlighted variation across CT lung screening studies in the U.S. and Europe[12]. For example, the smallest radiologic pulmonary nodule size warranting diagnostic testing, such as by fine-needle biopsy, varied by study site and ranged from 4 mm to 15 mm. Concerns have also been raised that a

lung cancer mortality reduction may be more difficult to achieve at a population level than in the clinical trial setting[13].

The median tumor diameter of aggressive NSCLCs detected in routine clinical practice exceeds 35 mm[14], whereas CT can detect nodules as small as 2mm. Nonetheless, despite over 90% adherence to annual screening in the CT arm of the NLST, the majority of aggressive lung cancers were not detected sufficiently early to be curable by surgical resection[1]. Populations of small, highly aggressive lung cancers have been reported across a range of CT screening cohorts[15,16]. Explicit knowledge of the cure threshold for the most aggressive non-small cell lung cancers (NSCLCs) can help to reconcile CT screening data with data from routine clinical practice and better define optimal screening parameters.

Stage for stage, lung cancers in women have higher survival[17]. As a result, the five-year NSCLC survival rate over 2004-2010 is higher in women compared to men (25.1% versus 18.7%)[18]. Some studies suggest that gender modifies lung cancer survival independently of other prognostic factors, such as age and smoking history[19,20]. Other studies attributed improved survival in women to a younger age at onset and shorter smoking history[21]. Furthermore, higher rates of over-diagnosis among women have been reported[22,23]. It is not known whether a more aggressive screening regimen is needed to reduce lung cancer mortality in men compared to women.

The use of statistical modeling and simulation to aid in the interpretation of lung cancer screening outcomes has increased in prominence, due in part to funding by the U.S. National Cancer Institute[24]. In order to simulate observed and hypothetical clinical

outcomes for a lung screening cohort, a natural history model of lung cancer must be specified. A classical type of natural history model incorporates estimates of the time that a cancer requires to pass through progressively advancing, asymptomatic stages, culminating in cancer death[25,26]. Alternatively, tumor size can be substituted for time as the primary dependent variable. A tumor size-driven natural history model incorporates estimates of tumor size at which lung cancers progress to a metastatic or advanced stage[27,28,29]. These two model types are related via a model of tumor volume doubling times (VDTs) over the natural course of lung cancer progression.

The estimation and fitting procedures for natural history models of lung cancer utilize maximum likelihood estimation (MLE). By definition, MLE produces a unique solution for a set of natural history model parameters, ostensibly reflecting the true underlying natural history model of lung cancer. As statistical modeling becomes more commonplace in the medical literature and a component of policy decisions[4,30], it is necessary to evaluate the credibility of the underlying models and the implicit assumptions used in model fitting procedures.

A change in the mode of clinical detection (i.e. chest x-ray (CXR), CT or symptomatic) influences the nature of the clinical data that is observed in practice[31]. Similarly, assumptions about the probability function governing detection (i.e. hazard function) affect the natural history model parameters obtained using MLE. Strict constraints, or model assumptions, are traditionally placed on the hazard function in order to ensure a unique solution from MLE is obtained. The size-dependent hazard function governing detection is distinct (and more complex) from the time-dependent hazard function governing detection. Previously, assumptions regarding these two distinct functions

have been erroneously conflated[27,28,29]. Furthermore, representation of the progression course of all lung cancers by a single unifying stochastic distribution precludes the precise estimation of over-diagnosis[30].

In this study, we directly estimated and determined the size-dependent hazard functions governing detection consistent with SEER data. Using a simulation framework, we determined the set of cure threshold models that were consistent with both SEER and NLST tumor size and outcome data. Our modeling framework also allowed for estimation of over-diagnosis in both SEER and the NLST.

2. Materials and Methods:

2.1 SEER data

The Surveillance, Epidemiology and End Results (SEER) cancer registry compiles clinico-pathologic data on the incidence of cancer from 15 U.S. states including the Cherokee Nation within the U.S. state of Oklahoma[32]. SEER-documented survival data was obtained from the National Center for Health Statistics.

We obtained data from SEER on non-small cell lung cancers (NSCLCs) diagnosed between 1988 and 2012 in individuals over the age of 45. For inclusion, the NSCLC must have been the only primary tumor in the patient's lifetime or the first among two or more primary tumors (sequence number: 00, 01). We identified NSCLCs by primary site (tumor site codes: C34.0-C34.3, C34.8-C34.9) and International Classification of Diseases for Oncology, Third Edition behavior and histology codes (behavior code: 2 or 3 and histology codes: 8010-8035, 8046-8078, 8140-8576). Carcinoid tumors (histology codes: 8240-8249) and in situ cancers (behavior code: 2) were included. Tumor size

was measured as the largest diameter of the primary tumor. At most a single tumor size measurement was reported, as given by the pathology report, operative report, endoscopic examination and radiographic report, in priority order. We compared the composition of demographic and histological data for NSCLCs with and without available tumor size information.

SEER documents the extent of disease (EOD) among NSCLCs as in situ, localized, regional or distant metastatic. We examined the relationship between tumor size and EOD for five tumor size categories: ≤ 15 mm, 16-25 mm, 26-35 mm, 36-45 mm, > 45 mm. We computed Kaplan-Meier five-year NSCLC survival curves for each tumor size and EOD category. We assumed that a cancer death occurring within five years following a primary NSCLC diagnosis resulted from dissemination from the primary tumor site. Any death by a cause other than cancer or loss to follow-up occurring within five years of diagnosis was treated as a censoring event. The relationship between five-year NSCLC survival and tumor size within each EOD category was examined. All analyses were carried out separately for each gender.

2.2 Kaplan-Meier estimation of the size-dependent probability function governing detection:

We defined a NSCLC as curable if the patient survived five years beyond a primary NSCLC diagnosis and incurable if the patient died of cancer within five years of a primary NSCLC diagnosis. For censored records, we imputed a curability status given the probability of NSCLC survival from the time of censoring to five years after NSCLC diagnosis conditional on tumor size and EOD category.

Kaplan-Meier estimation of the survival function governing detection, having tumor size as the dependent variable, was given by the following formula:

$$S_j(s) = \prod_{i:s_i \le s} \frac{n_{ij} - d_{ij}}{n_{ij}} \text{ (Equation 2.2.1)}$$

The binary index *j* indicates curability status, namely incurable (j=1) and curable (j=0). $S_j(s)$ equals the probability that a NSCLC remains undetected at size *s*, n_{ij} equals the number of NSCLCs at risk for detection at size s_i and d_{ij} equals the number of NSCLCs detected at size s_i. Whereas d_{ij} was observed for each *j* and size s_i, information concerning n_{ij} was incomplete. For example, a curable cancer detected at 30 mm was curable and at risk for detection at 15 mm. However, we did not know whether an incurable cancer detected at 30 mm would have been curable if it had been detected at 15 mm. We imposed the constraint that the sum of all n_{ij} equaled the total number of NSCLCs, *N,* ultimately detected and reported in SEER. Discrete hazard probabilities governing detection were derived from $S_j(s)$ by the following relationship.

$$h_j(s_i) = \frac{S_j(s_{i-1}) - S_j(s_i)}{S_j(s_{i-1})} \text{ (Equation 2.2.2)}$$

We smoothed the resulting hazard function using a rolling mean over a one centimeter interval.

We defined the cumulative density function (cdf), G(x) = Pr(X≤x), as the probability distribution governing the largest tumor size that a SEER-documented NSCLC would have been curable by early detection (cure threshold). The probability distribution, G(x|Φ=1), is conditional on a NSCLC's eventual incurable detection and therefore excludes from the underlying sample space any SEER-documented NSCLCs which

may have been subject to over-diagnosis. The function G(x|Φ=1) determines E (n_{i1}) for all *i* by the following equation:

$$\widehat{E}(n_{i1}) = \sum_{m=i}^{k} d_{m1} Pr(X \leq s_i \,|X \leq s_m) = \sum_{m=i}^{k} d_{m1} G(s_i \,|\Phi = 1, Y = s_m)$$ (Equation 2.2.3)

X is a random variable representing the tumor size at the cure threshold, Y is a random variable representing the tumor size at detection; Φ is an indicator random variable reflecting whether the detected NSCLC was incurable and *k* equals the largest tumor size at NSCLC detection documented in SEER. E(n_{i0}) was estimated by the relationship:

$$\hat{E}(n_{i0}) = -\hat{E}(n_{i1}) + \sum_{m=i}^{k}(d_{m0} + d_{m1})$$ (Equation 2.2.4)

We estimated E (n_{i0}) and E (n_{i1}) for a set of conditional cure threshold distributions G(x|Φ=1) having a gamma distribution with a median cure threshold of either 5mm, 10mm, 15mm, 20mm, 25mm or 30mm with a standard deviation ranging from 5mm to 20 mm. For each scenario, we computed the resulting survival and discrete hazard functions governing detection by Kaplan-Meier estimation.

2.3 Simulation of SEER-documented outcomes and tumor size distributions:

We estimated G(x) for all SEER-documented NSCLCs using maximum likelihood estimation. We assumed that G(x) was a mixture distribution with two components and the following form:

$$G(x) = pG_1(x) + (1-p)G_2(x) = pG(x|\Phi = 1) + (1-p)N(x|\mu, \sigma)$$ (Equation 2.3.1)

Likelihood equations, excluding constant terms, were as follows:

$L(Y = s_d; \Phi = 1) = pG_1(s_d) \sum_{m=0}^{d-1} \frac{S_0(s_m)}{S_1(s_m)} Pr_1(X = s_m) + (1-p)G_2(s_d) \sum_{m=0}^{d-1} \frac{S_0(s_m)}{S_1(s_m)} Pr_2(X = s_m)$ (Equation 2.3.2)

and

$$L(Y = s_d; \Phi = 0) = [1 - pG_1(s_d) - (1-p)G_2(s_d)] \text{ (Equation 2.3.3)}$$

The standard deviation of the normal mixture component was set equal to σ=10mm. Therefore, we used maximum likelihood estimation to estimate the parameters *p* and *μ*, only. We simulated the expected tumor size distributions and cure fraction under G(x) and the associated discrete hazard functions h_j for each of the assumed G(x|Φ=1) scenarios.

2.4 Hazard Functions Governing the Detection of SEER-documented NSCLCs

For a fixed tumor diameter *s*, we set the probability that a curable NSCLC was detected over a time interval Δt equal to $\mu_0\Delta t$. We defined t_1 as the length of time spent in size *s*, so that $t_1 = n\Delta t$, where *n* is an integer. We expressed the hazard $h_0(s)$ as:

$$h_0(s) = 1 - \lim_{\substack{n\to\infty \\ \Delta t\to 0}} (1 - \frac{\mu_0 n\Delta t}{n})^n = 1 - e^{-\mu_0 t_1} \text{ (Equation 2.4.1)}$$

We assumed that the probability that an incurable NSCLC was detected over a time interval Δt was equal to $(\mu_0+\mu_1)\Delta t$, and we similarly expressed the discrete hazard $h_1(s)$:

$$h_1(s) = 1 - \lim_{\substack{n\to\infty \\ \Delta t\to 0}} (1 - \frac{(\mu_0+\mu_1) n\Delta t}{n})^n = 1 - e^{-(\mu_0+\mu_1) t_1} \text{ (Equation 2.4.2)}$$

As a function of t_1, the discrete hazard ratio of h_0/h_1 is bounded above by one since $\mu_0 > 0$ and $\mu_1 \geq 0$. However, t_1 is a random variable defined as the length of time spent in

tumor size *s* and follows a probability distribution which differs for curable and incurable NSCLCs. We expressed t_1=*k*v, where *k* was a constant equal to the number of volume doublings that occurred over size *s* and *v* was the volume doubling time (VDT). We assumed that the distribution of VDTs followed a gamma distribution with probability density function (pdf), $f_s(v)$, and expressed $h_0(s)$ in terms of the parameters of the gamma distribution:

$$h_0(s) = \int_0^\infty (1 - e^{-\mu_0 kv}) f_s(v) dv \quad \text{(Equation 2.4.3)}$$

The gamma pdf is: $f_s(v) = \frac{1}{\Gamma(\alpha_0)\beta_0{}^{\alpha_0}} v^{\alpha_0 - 1} e^{-v/\beta_0}$ and gives a closed-form solution for $h_0(s)$:

$$h_0(s) = 1 - (\frac{1}{1+\mu_0 k\beta_0})^{\alpha_0} \quad \text{(Equation 2.4.4)}$$

The function $h_1(s)$ was similarly obtained. First-order Taylor expansion of $f(x) = (1+x)^{\alpha_0}$ around x=0 where $1 + x = \frac{1}{1+\mu_0 k\beta_0}$ yields the result that $(1+x)^{\alpha_0} \approx 1 + \alpha_0 x$. It followed that the discrete hazard ratio (curable/incurable) simplified to:

$$\frac{h_0(s)}{h_1(s)} \approx \frac{\mu_0 \beta_0 \alpha_0}{(\mu_0+\mu_1)\beta_1 \alpha_1} \quad \text{(Equation 2.4.5)}$$

The discrete hazard ratio exceeds one when $\beta_0 \alpha_0 > \beta_1 \alpha_1$ and $(\mu_0+\mu_1) \approx \mu_0$. This condition is satisfied when the average VDT for curable NSCLCs is greater than the average VDT for incurable NSCLCs and the rates of symptomatic detection are approximately equal. As μ_1 increases with increasing tumor size, the discrete hazard ratio, h_0/h_1, will decline.

2.5 NLST Data

We obtained data on NSCLCs detected during the first seven years of the NLST study for both the CT and chest x-ray (CXR) study arms. We excluded small cell lung cancers (ICD-0-3 codes 8041-8045) and carcino-sarcomas (ICD-0-3 code 8980) as well as 44 lung cancers (20 CT arm, 24 CXR arm) for which a lung cancer diagnosis was first recorded at the time of death. For these cancers, histology, staging and tumor size data were missing. We reported tumor size as given in the pathology record. When pathology tumor size was missing, we reported tumor size as the radiologic tumor size, when a radiologic record existed (screening years 0,1,2 only) and was consistent with the location of the tumor and timing of detection. For radiologic data, we reported tumor size as the average of the longest diameter and the perpendicular diameter.

For NSCLCs detected in the NLST, we defined curability based on five-year NSCLC survival, as for SEER-documented NSCLCs. Censored records occurred due to loss-of-follow-up or other-cause death within five-years of a primary NSCLC diagnosis. We imputed missing tumor size data differently for censored and non-censored records. For NSCLCs with non-censored survival data, we sampled a tumor size from the population of NSCLCs having the same curability status, gender and randomization group. For NSCLCs with censored survival data, we sampled a tumor size from the population of NSCLCs having the same tumor stage, gender and randomization group. For censored records, we imputed a curability status based on the probability of NSCLC survival from the time of censoring to five years after NSCLC diagnosis conditional on tumor size, tumor stage, gender and randomization group.

2.6 Evaluation of Consistency of Cure Threshold Models with NLST-reported NSCLCs.

We expected a hypothetical “no-screening” arm to produce approximately the same number of incurable NSCLCs as those that arose in the CXR arm of the NLST. A low or no mortality benefit attributable to screening by CXR had been demonstrated in prior studies and was a key assumption in the NLST study design[1,2,3]. For each given fitted G(x) model depicted in Table 2, we determined G, the expected size of the underlying G(x|Φ=1) population, constant across the hypothetical “no-screening” arm, the CXR arm and the CT arm (Supplement Methods 2.S1). Although more G(x|Φ=1) cancers are expected to be identified during active screening in the CT arm, these NSCLCs will appear later in the CXR arm, due to their progressive nature.

We simulated the number of NSCLC deaths in the CT and CXR arms over the range of fitted G(x) models, for males and females separately. In our simulations, a NSCLC death was recorded when, for a NSCLC in the G(x|Φ=1) population, the tumor size at the time of treatment exceeded the simulated cure threshold value for that NSCLC. The simulation procedure was described in detail in Supplement Methods 2.S2.

2.7 Estimation of Over-diagnosis in the NLST and SEER Populations.

Over-diagnosis accounted for the persistent differences in total NSCLCs over the seven-year period between the CT and CXR arms. We estimated the percentage of over-diagnosis in the CT and CXR arms relative to the expected number of total NSCLCs in the hypothetical “no-screening” arm and relative to G, the expected size of the underlying G(x|Φ=1) population. For the latter calculation, we estimated the extent of over-diagnosis in the SEER population.

Results:

3a. SEER Data

We identified 354,779 cases of NSCLC diagnosed between 1988 and 2012 with available tumor size information and 163,455 cases of NSCLC with missing tumor size information (Table 1). NSCLCs with missing tumor size information were more likely to have a cell type other than adeno-carcinoma, squamous cell or large cell (35.0% versus 26.6%). NSCLCs with missing tumor size information were also more likely to have distant metastases or to not have been staged (79.1% versus 45.2%).

The proportion of NSCLC with locally confined disease decreased with increasing tumor size (Supplement Table S1) whereas the proportion of NSCLCs classified as distant metastatic increased with increasing tumor size for both males and females. Among tumors having a diameter ≤ 15 mm, the proportion of NSCLCs having localized disease was 46.1% for males and 55.6% for females. Among tumors having a diameter > 45 mm, this proportion decreased to 10.7% for males and 11.1% for females. Among tumors having a diameter ≤ 15 mm, the proportion of NSCLCs having distant metastatic disease was 32.1% for males and 24.2% for females. Among tumors having a diameter > 45 mm, this proportion increased to 53.4% for males and 54.6% for females.

Within EOD status, five-year NSCLC survival varied with increasing tumor size most widely for localized disease (Supplement Figure S1). In males, the five-year NSCLC survival for a localized NSCLC having a tumor diameter ≤ 15 mm was 70.0%, whereas for a localized NSCLC having a tumor diameter > 45mm, it was 35.2%. In females, the five-year NSCLC survival for a localized NSCLC having a tumor diameter ≤ 15 mm was 79.2%, whereas for a localized NSCLC having a tumor diameter > 45mm, it was 37.6%.

For distant metastatic disease, the five-year NSCLC survival varied with primary tumor size from 7.1% to 3.4% for males and from 11.0% to 3.9% for females.

We classified SEER-documented NSCLCs with available tumor size information as either curable or incurable based on 5-year NSCLC survival after detection. In males, 76.8% of these NSCLCs were incurable whereas in females, 69.8% of these NSCLCs were incurable. For males, the median size at curable detection was 30 mm and the median size at incurable detection was 43 mm. For females, the median size at curable detection was 25 mm and the median size at incurable detection was 40 mm.

3b. Simulation of SEER Data and Fitted Cure Threshold Models

We reproduced SEER outcomes over a range of fitted G(x) models in both males and females. Simulation results were summarized and compared to SEER outcomes in Table 2. The fitted cure threshold models, G(x), having a median cure threshold for the G(x|Φ=1) population ranging between 5-20 mm produced results nearly identical to SEER for males. The fitted cure threshold models, G(x), having a median cure threshold for the G(x|Φ=1) population ranging between 5-25 mm produced results nearly identical to SEER for females.

Underlying the nearness of our simulation results to SEER outcomes were disparate fitted parameters of G(x), G(x|Φ=1) distributions, and associated discrete hazard functions governing detection. We depicted the estimated discrete hazard ratios (curable/incurable) and absolute discrete hazard functions governing detection in Figure 1 and Figure 2, respectively. For both males and females, we found the largest peak discrete hazard ratios were associated with a median cure threshold for the G(x|Φ=1)

population equal to 5 mm (Figure 1a). For females, the peak discrete hazard ratio was 1.8 and for males, the peak discrete hazard ratio was 1.6 (Figure 1a). Based on Equation 2.4.5, the peak discrete hazard ratio can be interpreted as the approximate ratio of average doubling times between curable and incurable NSCLCs when $\mu_1 = 0$.

For both males and females, we found the smallest absolute discrete hazard function governing detection for incurable cancers was associated with a median cure threshold for the G(x|Φ=1) population equal to 5 mm (Figure 2a). For a NSCLC that was incurable at size 2 mm, the probability that it was detected before it reached 20 mm was 14.7% for males and 17.8% for females (Figure 2a). In contrast, when the median cure threshold for the G(x|Φ=1) population was 25 mm, the probability that the same NSCLC was detected before it reached 20 mm was 42.4% for males and 46.9% for females (Figure 2e). Models having the smallest median cure threshold for the G(x|Φ=1) population were associated with the longest latency periods in which incurable NSCLCs were largely undetectable in the absence of screening.

3c. NLST Data

We described the demographic information for 953 NSCLCs (570 male, 383 female) detected in the CT arm and 816 NSCLCs (491 male, 325 female) detected in the CXR arm of the NLST study in Table 3. We summarized the curability status of these NSCLCs and the timing of detection over the seven-year period after the initial randomization in Supplement Table S2. The median tumor size at detection (including imputed tumor sizes) in the CT arm was 23 mm for males and 18 mm for females. The

median tumor size at detection in the CXR arm was 30 mm for males and 25 mm for females.

3d. NLST Simulation Results

We summarized the simulated numbers of NSCLC deaths that arose in the CT and CXR arms of the NLST over the range of fitted cure threshold models G(x) in Table 4. For males, the best fit model had a median cure threshold for the G(x|Φ=1) population equal to 10 mm. Models having a median cure threshold for the G(x|Φ=1) population ranging from 5 mm to 15 mm provided a fit that bordered the 95% confidence range of the simulation. For females, the best fit model had a median cure threshold for the G(x|Φ=1) population equal to 15 mm. Models that had a median cure threshold for the G(x|Φ=1) population ranging from 10 mm to 20 mm provided a fit that bordered the 95% confidence range of the simulation.

3e. Estimates of Over-Diagnosis in the NLST and SEER

Given 5-year NSCLC survival in the overall SEER population (25.1% in males, 18.7% in females), we expected the total number of NSCLCs to arise in the no-screening arm to range between 368 and 392 for males and between 227 and 247 for females. For the CT arm, we estimated the over-diagnosis relative to a non-screened population was between 31% - 35% in males and between 36% - 41% in females. For the CXR arm, we estimated the over-diagnosis relative to a non-screened population was between 20% - 25% in males and between 24% - 30% in females.

We also estimated over-diagnosis with respect to the G(x|Φ=1) population for the models having the best fit to the NLST data, namely models 2a-c for males and models

3a-c for females. For males, the estimated size of the G(x|Φ=1) population ranged between 318 and 368 for models 2a-c. Our over-diagnosis estimates ranged between 35% - 44% (CT arm), 25% - 35% (CXR arm) and 7% - 16% (no-screening arm). For females, the estimated size of the G(x|Φ=1) population ranged between 191 and 227 for models 3a-c. Our over-diagnosis estimates ranged between 41% - 50% (CT arm), 30% - 41% (CXR arm) and 8% - 16% (no-screening arm).

3. Discussion:

The SEER database is the largest cancer registry in the U.S. and a key indicator of population-wide trends concerning lung cancer detection, survival and mortality. The vast majority of SEER-documented NSCLCs reported between 1988 and 2012 were detected in the absence of routine screening. Because detection in the SEER population occurs non-systematically, either due to symptoms or an incidental CXR or CT scan, SEER data reflects the natural history of the NSCLCs we sought to characterize. We modeled SEER data with respect to the distribution of tumor sizes at cure threshold among the most aggressive NSCLCs, namely the G(x|Φ=1) sub-population and examined the consistency of these cure threshold models with NLST data.

Applied to SEER data, our statistical methodology allowed for the direct estimation of the hazard function governing detection given SEER-documented NSCLC tumor sizes and a designated cure threshold distribution for the G(x|Φ=1) sub-population. Our methodology differed from existing methodology used to estimate size-driven natural history models in several key ways. We distinguished the G(x|Φ=1) population from the

overall G(x) population, which included NSCLCs that could potentially be classified as over-diagnosis. We further showed that direct estimation of the hazard function governing detection required specification of the cure threshold distribution for the G(x|Φ=1) sub-population. Without this result, we would have needed to incorporate *a priori* assumptions about the shape and form of the hazard function governing detection into the estimation framework.

Furthermore, we contradicted the assumption that the hazard function governing detection of incurable NSCLCs is uniformly greater than the hazard function governing detection of curable NSCLCs over the detectable tumor size range[21,28] or equivalently, that the time-dependent and size-dependent hazard functions are inter-changeable[29]. Based on Equation 2.4.5, we expected this assumption to be false since we expected average VDTs for curable SEER-documented NSCLCs to be larger than VDTs for incurable SEER-documented NSCLCs. Over the range of models in which the median tumor size at the cure threshold for the G(x|Φ=1) sub-population was smaller than 20 mm, the peak hazard ratio exceeded one for both males and females, consistent with our Equation 2.4.5. Therefore, the model solutions we found most likely based on SEER data (i.e. median G(x|Φ=1) cure threshold < 20 mm) would not have been obtained under existing statistical methodology since they violated *a priori* assumptions. Prior published estimates of the tumor size at onset of metastases have consistently been reported as greater than 40 mm[27,28,29].

Another assumption was that advanced-stage cancers have high detectability, frequently detected at or near the time of metastatic transition[27]. In a modeling framework, this assumption has the effect of either lengthening the estimate of an early-

stage pre-clinical phase relative to an advanced-stage pre-clinical phase or increasing estimates of the cure threshold. We found that for models having a median cure threshold for the G(x|Φ=1) population less than 20 mm, the initial detection probability for incurable NSCLCs was small and increased with tumor size. In the Lung Screening Study (pilot study to the NLST), the rate of detection for advanced-stage cancer by CT was twice as high as the rate of detection for advanced-stage NSCLC by chest x-ray[33] at baseline. Data from CT screening studies has routinely demonstrated small tumor sizes for aggressive NSCLCs that would typically be detected later in routine clinical practice[15,16]. Overall, these data support a latency period for aggressive NSCLCs in the absence of sensitive detection strategies, such as CT.

Modeling the G(x|Φ=1) sub-population has advantages other than pure mathematical convenience. The cure threshold distribution G(x|Φ=1) has greater clinical relevance than the cure threshold distribution G(x) since potentially over-diagnosed NSCLCs are excluded. Each aggressive NSCLC detected and treated prior to the cure threshold in the G(x|Φ=1) sub-population theoretically contributes to a population-wide reduction in NSCLC mortality. In modeling the NLST data, the sizes of the G(x|Φ=1) sub-populations were equivalent in expectation across the CT arm, the CXR arm and the hypothetical “no-screening” arm. In contrast, the sizes of the G(x) populations differed across these arms because the rates of over-diagnosis varied by screening modality. Furthermore, we were able to pool baseline and annual-repeat cancers over the seven-year period following randomization in the NLST. Because the G(x|Φ=1) sub-population excluded potentially over-diagnosed lung cancers, the effect of length-bias did not impact our model predictions.

The NLST study reported a 20% lung cancer mortality reduction attributable to CT screening, a watershed moment after a long period of contentious debate on whether to screen for lung cancer at all. However, the mortality benefit was limited in the NLST because the vast majority of aggressive NSCLCs were detected above their respective cure thresholds. Our simulation framework allowed us to test a given cure threshold model against the observed tumor sizes at treatment and detection among NSCLCs in the NLST. We concluded that SEER and NLST data were most consistent with a model in which the median cure threshold for aggressive NSCLCs was between 10 mm (males) to 15 mm (females).

A current challenge is to distinguish among pulmonary nodules in the size range < 15 mm, where the majority of pulmonary nodules are benign[34] and to minimize unnecessary surgeries[35]. The International Early Lung Cancer Action Program (I-ELCAP) screening regimen incorporates sensitive VDT measurements for annual-repeat lung cancers and diagnoses lung cancers at smaller tumor sizes than the NLST[36,37]. I-ELCAP has consistently reported the largest estimates of a mortality benefit attributable to CT screening[10] although reported five-year lung cancer survival rates of over 80% are for a single study arm (i.e. the complete G(x) population which may include over-diagnosed lung cancers)[36]. The screening regimen in the randomized Dutch-NELSON study also places emphasis on the detection of small pulmonary nodules demonstrating growth[8,38], although survival and mortality outcomes have not yet been reported.

In our view, this study has two key clinical implications. For one, it cautions against proposals to lengthen the interval between screens and to raise the tumor size

threshold for lung cancer diagnosis in order to reduce rates of over-diagnosis[39] and favors the development of novel classification technologies to distinguish among the most and least aggressive lung cancers detected at small tumor sizes. Furthermore, our results suggest that CT screening programs which diagnose lung nodules at the smallest possible size will have the highest lung cancer mortality reductions. Given the variation among studies in the timing of diagnosis and treatment, we anticipate that the mortality benefits associated with different CT screening studies may vary substantially.

References:

1. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD for the National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. N Engl J Med 2011; 365(5): 395-409.
2. Oken MM, Hocking WG, Kvale PA, Andriole GL, Buys SS, Church TR, Crawford ED, Fouad MN, Isaacs C, Reding DJ, Weissfeld JL, Yokochi LA, O'Brien B, Ragard LR, Rathmell JM, Riley TL, Wright P, Caparaso N, Hu P, Izmirlian G, Pinsky PF, Prorok PC, Kramer BS, Miller AB, Gohagan JK, Berg CD for the PLCO Project Team Screening by Chest Radiograph and Lung Cancer Mortality: The Prostate, Lung, Colorectal, and Ovarian (PLCO) Randomized Trial. JAMA 2011; 306(17): 1865-1873.
3. Marcus PM, Bergstralh EJ, Fagerstrom RM, Williams DE, Fontana R, Taylor WF, Prorok PC. Lung Cancer Mortality in the Mayo Lung Project: Impact of Extended Follow-up. J Natl Cancer Inst 2000; 92(16): 1308-1316.

4. Moyer VA for the U.S. Preventive Services Task Force. Screening for Lung Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med 2014; 160(5): 330-338.
5. Infante M, Cavuto S, Lutman FR, Brambilla G, Chiesa G, Ceresoli G, Passera E, Angeli E, Chiarenza M, Aranzulla G, Cariboni U, Errico V, Inzirillo F, Bottoni E, Voulaz E, Alloisio M, Destro A, Roncalli M, Santoro A, Ravasi G for the DANTE Study Group. A Randomized Study of Lung Cancer Screening with Spiral Computed Tomography: Three-Year Results from the DANTE Trial. Am J Respir Crit Care Med 2009; 180: 445-453.
6. Saghir Z, Dirksen A, Ashraf H, Bach KS, Brodersen J, Clementsen PF, Dossing M, Hansen H, Kofoed KF, Larsen KR, Mortensen J, Rasmussen JF, Seersholm N, Skov BG, Thorsen H, Tonnesen P, Pedersen JH. CT Screening for Lung Cancer Brings Forward Early Disease. The Randomised Danish Lung Cancer Screening Trial: Status after Five Annual Screening Rounds with Low-Dose CT. Thorax 2012; 67(4): 296-301.
7. Pastorino U, Rossi M, Rosato V, Marchiano A, Sverzellati N, Morosi C, Fabbri A, Galeone C, Negri E, Sozzi G, Pelosi G, la Vecchia C. Annual or Biennial CT Screening Versus Observation in Heavy Smokers: 5-year Results of the MILD Trial. Eur J Cancer Prev 2012; 21(3): 308-15.
8. Van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K, Vernhout R, van Iersel CA, van den Bergh K, Westeinde S, van der Aalst C, Thunnissen E, Xu DM, Wang Y, Zhao Y, Gietema HA, de Hoop BJ, Groen H, de Bock GH, van Ooijen P, Weenink C, Verschakelen J, Lammers JW, Timens W, Willebrand D,

Vink A, Mali W, de Koning HJ. Management of Lung Nodules Detected by Volume CT Scanning. N Engl J Med 2009; 361(23): 2221-2229.

9. Bach PB, Jett JR, Pastorino U, Tockman MS, Swensen SJ, Begg CB. Computed Tomography Screening and Lung Cancer Outcomes. JAMA 2007; 297(9): 953-961.
10. Henschke CI, Boffetta P, Gorlova O, Yip R, DeLancey JO, Foy M. Assessment of Lung Cancer Mortality Reduction from CT Screening. Lung Cancer 2011; 71: 328-332.
11. Meza R, Haaf KT, Kong CY, Erdogan A, Black W, Tammemagi M, Choi SE, Jeon J, Han S, Munshi V, van Rosmalen J, Pinsky P, McMahon PM, de Koning H, Feuer EJ, Hazelton WD, Plevritis SK. Comparative Analysis of Five Lung Cancer Natural History and Screening Models that Reproduce Outcomes of the NLST and PLCO Trials. Cancer 2014; 120(11): 1713-1724.
12. Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW, Byers T, Colditz GA, Gould MK, Jett JR, Sabichi AL, Smith-Bindham R, Wood DE, Qaseem A, Detterbeck FC. Benefits and Harms of CT Screening for Lung Cancer: A Systematic Review. JAMA 2012; 307(22): 2418-2429.
13. Sox HC Better Evidence about Screening for Lung Cancer. N Engl J Med 2011; 365(5): 455-457.
14. Wisnivesky JP, Yankelevitz DF, Henschke CI. Stage of Lung Cancer in Relation to its Size: Part 2. Evidence. Chest 2005; 127(4): 1136-1139.
15. Henschke CI, Yankelevitz DF, Miettinen O for the International Early Lung Cancer Action Program (I-ELCAP) Investigators. Computed Tomographic

Screening for Lung Cancer: The Relationship of Disease Stage to Tumor Size. Arch Intern Med 2006; 166: 321-325.

16. Goldwasser DL, Kimmel M. Small Median Tumor Diameter at Cure Threshold (<20 mm) among Aggressive Non-Small Cell Lung Cancers in Male Smokers Predicts Both Chest X-Ray and CT Screening Outcomes in a Novel Simulation Framework. Int J Cancer 2013; 132(1): 189-197.
17. Fu JB, Kau Y, Severson RK, Kalemkerian GP. Lung Cancer in Women: Analysis of the National Surveillance, Epidemiology, and End Results Database. Chest 2005; 127(3): 768-777.
18. American Lung Association. Trends in Lung Cancer Morbidity and Mortality 2014; Epidemiology and Statistics Unit, Research and Program Services Division.
19. Ebbert JO, Williams BA, Sun Z, Aubry MC, Wampfler JA, Garces YI, Meyer RL, Yang P. Duration of Smoking Abstinence as a Predictor for Non-Small Cell Lung Cancer Survival in Women. Lung Cancer 2005; 47: 165-172.
20. Wisnivesky JP, Halm EA. Sex Differences in Lung Cancer Survival: Do Tumors Behave Differently in Elderly Women? J Clin Oncol 2007; 25(13): 1705-1712.
21. Hsu LH, Chu NM, Liu CC, Tsai S, You DL, Ko JS, Lu MC, Feng AC Sex-Associated Differences in Non-Small Cell Lung Cancers in the New Era: Is Gender an Independent Prognostic Factor? Lung Cancer 2009; 66(2): 262-267.
22. Lindell RM, Hartman TE, Swensen SJ, Jett JR, Midthun DE, Tazelaar HD, Mandrekar JN. Five-Year Lung Cancer Screening Experience: CT Appearance, Growth Rate, Location, and Histologic Features of 61 Lung Cancers. Radiology 2007; 242(2): 555-562.

23. Hazelton WD, Goodman G, Rom WN, Tockman M, Thornquist M, Moolgavkar S, Weissfeld JL, Feng Z. Longitudinal Multistage Model for Lung Cancer Incidence, Mortality and CT Detected Indolent and Aggressive Cancers. Math Biosci 2012; 240: 20-34.
24. Cancer Intervention and Surveillance Modeling Network. National Cancer Institute, United States, accessed June 24, 2016, www.cisnet.cancer.gov/lung.
25. Walter SD, Day NE. Estimation of the Duration of a Pre-Clinical Disease State Using Screening Data. Am. J. Epidemiol 1983; 118(6): 865-886.
26. Flehinger BJ, Kimmel M. The Natural History of Lung Cancer in a Periodically Screened Population. Biometrics 1987; 43: 127-144.
27. Kimmel M, Flehinger BJ. Nonparametric Estimation of the Size-Metastasis Relationship in Solid Cancers. Biometrics 1991; 47: 987-1004.
28. Xu JL, Prorok PC. Estimating a Distribution Function of the Tumor Size at Metastasis. Biometrics 1998; 54: 859-864.
29. Pashkevich MA, Sigal BM, Plevritis SK. Modeling the Transition of Lung Cancer from Early to Advanced Stage. Cancer Causes Control 2009; 20: 1559-1569.
30. Patz EF, Pinsky P, Gatsonis C, Sicks JD, Kramer BS, Tammemagi MC, Chiles C, Black WC, Aberle DR for the NLST Overdiagnosis Manuscript Writing Team. Overdiagnosis in Low-Dose Computed Tomography Screening for Lung Cancer. JAMA Intern Med 2014; 174(2): 269-274.
31. Feinstein AR, Sosin DM, Wells CK. The Will Rogers Phenomenon: Stage Migration and New Diagnostic Techniques as a Source of Misleading Statistics for Survival in Cancer. N Engl J Med 1985; 312(25): 1604-1608.

32. Surveillance, Epidemiology and End Results Program. National Cancer Institute, United States, accessed June 24, 2016, www.seer.cancer.gov.
33. Gohagan JK, Marcus PM, Fagerstrom RM, Pinsky PF, Kramer BS, Prorok PC, Ascher S, Bailey W, Brewer B, Church T, Engelhard D, Ford M, Fouad M, Freedman M, Gelmann E, Gierada D, Hocking W, Inampudi S, Irons B, Johnson CC, Jones A, Kucera G, Kvale P, Lappe K, Manor W, Moore A, Nath H, Neff S, Oken M, Plunkett M, Price H, Reding D, Riley T, Schwartz M, Spizarny D, Yoffie R, Zylak C for the Lung Screening Study Research Group. Final Results of the Lung Screening Study, a Randomized Feasibility Study of Spiral CT versus Chest X-ray Screening for Lung Cancer. Lung Cancer 2005; 47: 9-15.
34. Horeweg N, van Rosmalen J, Heuvelmans MA, van der Aalst CM, Vliegenthart R, Scholten ET, ten Haaf K, Nackaerts K, Lammers JJ, Weenink C, Groen HJ, van Ooijen P, de Jong PA, de Bock GH, Mali W, de Koning HJ, Oudkerk M. Lung Cancer Probability in Patients with CT-Detected Pulmonary Nodules: A Pre-Specified Analysis of Data from the NELSON trial of Low-Dose CT Screening. Lancet Oncol 2014; 15: 1332-1341.
35. Flores R, Bauer T, Aye R, Andaz S, Kohman L, Sheppard B, Mayfield W, Thurer R, Smith M, Korst R, Straznicka M, Grannis F, Pass H, Connery C, Yip R, Smith JP, Yankelevitz D, Henschke C, Altorki N for the I-ELCAP Investigators. Balancing Curability and Unnecessary Surgery in the Context of Computed Tomography Screening for Lung Cancer. J Thorac Cardiovasc Surg 2014; 147(5): 1619-1625.

36. Yip R, Henschke CI, Yankelevitz DF, Boffetta P, Smith JP for the International Early Lung Cancer Investigators. The Impact of the Regimen of Screening on Lung Cancer Cure: A Comparison of I-ELCAP and NLST. Eur J Cancer Prev 2015; 24(3): 201-208.
37. Henschke CI, Li K, Yip R, Salvatore M, Yankelevitz DF. The Importance of the Regimen of Screening in Maximizing the Benefit and Minimizing the Harms. Ann Transl Med 2016; 4(8): 153-165.
38. Walter JE, Heuvelmans MA, de Jong PA, Vliegenthart R, van Ooijen PM, Peters RB, Ten Haaf K, Yousaf-Khan U, van der Aalst CM, de Bock GH, Mali W, Groen HJ, de Koning HJ, Oudkerk M. Occurrence and Lung Cancer Probability of New Solid Nodules at Incidence Screening with Low-Dose CT: Analysis of Data from the Randomised, Controlled NELSON trial. Lancet Oncol 2016; 17(7): 907-16.
39. Esserman L, Thompson I. Solving the Overdiagnosis Dilemma. J Natl Cancer Inst 2010; 102(9): 582-583.

Table 1: Demographic, cell type and extent of disease distribution among SEER non-small cell lung cancers with and without complete tumor size information.

	All Cases	With Complete Tumor Size Information	Without Complete Tumor Size Information
N	518,234	354,779	163,455
Age (+/- Standard deviation)	68.8 +/- 12.0	68.5 +/- 10.3	69.5 +/- 15.0
Percentage Male	55.9%	55.1%	57.6%
Ethnicity			
White	81.8%	81.9%	81.5%
African-American	11.5%	11.4%	11.9%
Other ethnicity	6.7%	6.7%	6.6%
Tumor Cell Type			
Adeno-carcinoma	41.6%	42.8%	39.1%
Squamous Cell	23.9%	25.7%	20.0%
Large Cell Carcinoma	5.3%	5.0%	5.8%
Other Cell Type	29.2%	26.6%	35.0%
Extent of Disease			
In Situ	0.1%	0.1%	0.2%
Local	19.1%	25.1%	6.1%
Regional	25.0%	29.7%	14.7%
Distant	49.3%	43.4%	62.3%
Unstaged	6.5%	1.8%	16.8%

Table 2: Comparison of simulated non-small cell lung cancer tumor sizes and cure fraction with observed SEER findings among males and females.

a) Males

				SEER Data (Simulated[1] and Observed)			
Scenario	G(x|Φ=1) (Median, mm)	G(x|Φ=1) (Stdev, mm)	Estimated (p,μ)	Median Size at Curable Detection (mm)	Median Size at Incurable Detection (mm)	% Incurable	Median Size at Cure Threshold among Incurable Cancers (mm)
Observed	**---**	**---**	**---**	**30 mm**	**43 mm**	**76.8%**	**unobservable**
1a	5	5.75	0.78619, 100	30 mm	42 mm	77.1%	5 mm
1b	5	9.0	0.80711, 100	30 mm	42 mm	77.0%	5 mm
1c	5	14.5	0.84014, 100	30 mm	42 mm	77.1%	5 mm
2a	10	7.61	0.81609, 100	30 mm	42 mm	77.2%	10 mm
2b	10	11.5	0.84419, 100	30 mm	42 mm	76.9%	9 mm
2c	10	17.9	0.88555, 100	30 mm	42 mm	76.9%	8 mm
3a	15	5.59	0.82465, 100	30 mm	43 mm	77.1%	15 mm
3b	15	9.15	0.85156, 100	29 mm	43 mm	76.8%	15 mm
3c	15	13.5	0.88151, 100	29 mm	43 mm	76.7%	14 mm
4a	20	7.23	0.86345, 100	29 mm	45 mm	76.5%	20 mm
4b	20	10.4	0.88600, 100	29 mm	44 mm	76.5%	19 mm
4c	20	15.2	0.91712, 100	29 mm	44 mm	76.0%	18 mm
5a	25	8.05	0.89428, 100	29 mm	47 mm	75.7%	25 mm
5b	25	11.55	0.91670, 100	29 mm	46 mm	75.4%	24 mm
5c	25	16.8	0.94869, 100	29 mm	46 mm	74.8%	22 mm
6a	30	8.77	0.92014, 100	29 mm	49 mm	74.3%	29 mm
6b	30	12.54	0.94200, 100	30 mm	49 mm	73.8%	28 mm
6c	30	18.27	0.97301, 100	30 mm	48 mm	72.8%	27 mm

(1) With a 2mm bias correction in the probability function for curable detection.

b) Females

				SEER Data (Simulated[1] and Observed)			
Scenario	G(x|Φ=1) (Median, mm)	G(x|Φ=1) (Stdev, mm)	Estimated (p,μ)	Median Size at Curable Detection (mm)	Median Size at Incurable Detection (mm)	% Incurable	Median Size at Cure Threshold among Incurable Cancers (mm)
Observed	**---**	**---**	**---**	**25 mm**	**40 mm**	**69.8%**	**unobservable**
1a	5	5.75	0.72524, 100	26 mm	38 mm	70.1%	5 mm
1b	5	9.0	0.75139, 100	26 mm	38 mm	69.9%	5 mm
1c	5	14.5	0.78980, 100	26 mm	38 mm	70.1%	4 mm
2a	10	7.61	0.76773, 100	26 mm	38 mm	70.0%	10 mm
2b	10	11.5	0.80235, 100	26 mm	38 mm	70.3%	9 mm
2c	10	17.9	0.84976, 100	26 mm	38 mm	70.1%	8 mm
3a	15	5.59	0.78577, 100	26 mm	39 mm	69.8%	15 mm
3b	15	9.15	0.82017, 100	26 mm	38 mm	69.9%	14 mm
3c	15	13.5	0.85597, 100	25 mm	38 mm	70.0%	13 mm
4a	20	7.23	0.84349, 100	25 mm	40 mm	69.3%	19 mm
4b	20	10.4	0.87129, 100	25 mm	40 mm	69.2%	19 mm
4c	20	15.2	0.90741, 100	25 mm	40 mm	69.0%	17 mm
5a	25	8.05	0.88841, 100	25 mm	42 mm	68.0%	24 mm
5b	25	11.55	0.91487, 100	25 mm	42 mm	67.6%	23 mm
5c	25	16.8	0.95068, 100	26 mm	41 mm	67.1%	21 mm
6a	30	8.77	0.92309, 100	26 mm	46 mm	66.3%	29 mm
6b	30	12.54	0.94757, 100	26 mm	45 mm	65.6%	28 mm
6c	30	18.27	0.98010, 100	26 mm	44 mm	64.8%	25 mm

(1) With a 2 mm bias correction in the probability function for curable detection.

Table 3: Demographic, cell type, tumor size and tumor stage distribution among non-small cell lung cancers detected in the NLST.

	CT Screened Arm			Chest X-Ray Screened Arm		
	Total NSCLCs	NSCLC Deaths	NSCLC Alive or Censored	Total NSCLCs	NSCLC Deaths	NSCLC Alive or Censored
N (Total NSCLCs)	953	341	612	816	410	406
Median Age at Randomization (+/- Standard deviation)	63 +/- 5.3	64 +/- 5.5	63 +/- 5.3	63 +/- 5.2	64 +/- 5.3	63 +/- 5.2
Percentage Male	59.8%	67.2%	55.7%	60.2%	62.7%	57.6%
Median post-diagnosis FUP time (years)	---	0.80	4.47	---	0.62	2.70
Ethnicity						
White	91.0%	90.9%	91.0%	90.0%	88.8%	91.1%
African-American	4.8%	4.1%	5.2%	5.8%	6.1%	5.4%
Other ethnicity or Unknown	4.2%	5.0%	3.8%	4.3%	5.1%	3.4%
Histology						
Adeno-carcinoma	52.9%	42.5%	58.7%	46.1%	46.3%	45.8%
Squamous Cell	26.1%	29.0%	24.5%	26.2%	19.3%	33.3%
Large Cell Carcinoma	4.2%	5.0%	3.8%	5.4%	5.9%	4.9%
Other Cell Type	15.3%	21.1%	12.1%	21.1%	26.8%	15.3%
Missing Data	1.5%	2.3%	1.0%	1.2%	1.7%	0.7%
Tumor Size						
≤ 10 mm	15.6%	10.9%	18.3%	5.5%	3.2%	7.9%
11 – 20 mm	33.8%	17.0%	43.1%	25.4%	18.5%	32.3%
21 – 30 mm	19.5%	20.2%	19.1%	21.4%	18.0%	24.9%
31 – 40 mm	8.8%	11.1%	7.5%	14.8%	16.6%	13.1%
41 – 50 mm	6.5%	11.4%	3.8%	7.4%	6.8%	7.9%

> 50 mm	9.5%	16.4%	5.7%	15.9%	21.4%	10.3%
Missing Data	6.2%	12.9%	2.5%	9.6%	15.4%	3.7%
Tumor Stage						
Stage IA	43.7%	13.2%	60.6%	23.8%	7.3%	40.4%
Stage IB	11.5%	6.7%	14.2%	11.4%	4.9%	18.0%
Stage IIA	3.3%	3.5%	3.1%	3.6%	2.7%	4.4%
Stage IIB	3.7%	3.2%	3.9%	4.4%	2.0%	6.9%
Stage IIIA	8.6%	16.4%	4.2%	12.1%	13.7%	10.6%
Stage IIIB	10.1%	15.2%	7.2%	11.6%	14.4%	8.9%
Stage IV	17.8%	39.9%	5.6%	32.4%	54.6%	9.9%
Missing Data	1.4%	1.8%	1.1%	0.7%	0.5%	1.0%

Table 4: Comparison of simulated non-small cell lung cancer deaths with observed NLST findings among males and females (1).

a) Males

Scenario	G(x|φ=1) (Median, mm)	G(x|φ=1) (Stdev, mm)	Total (G(x|φ=1) LCs	Total LC Deaths, CT Arm	Total LC Deaths , CXR Arm	Difference in Total Deaths (CXR-CT)	Difference in the Percentage of Total NSCLC Deaths (CXR-CT)/CXR
Obs.	**---**	**---**	**unobs.**	269 [260, 279]	309 [299, 319]	40 [26, 53]	12.9% [8.6%,16.9%]
1a	5	5.75	316 [306, 327]	294 [270, 318]	308 [293, 331]	15 [-10, 42]	4.8% [-3.3%, 13.0%]
1b	5	9.0	324 [314, 335]	290 [265, 312]	304 [286, 327]	16 [-12,45]	5.2% [-4.0%, 14.3%]
1c	5	14.5	338 [327, 349]	287 [264, 309]	302 [280, 322]	14 [-13, 43]	4.6% [-4.7%, 13.8%]
2a	10	7.61	328 [318, 339]	275 [253,296]	300 [280, 321]	25 [-4, 51]	8.3% [-1.3%, 16.3%]
2b	10	11.5	340 [329, 351]	273 [251, 296]	296 [275, 317]	23 [-6, 52]	7.8% [-2.2%, 16.6%]
2c	10	17.9	356 [345, 368]	271 [248, 294]	290 [268, 311]	18 [-10, 48]	6.3% [-3.7%, 16.0%]
3a	15	5.59	332 [321, 343]	247 [228, 265]	285 [263, 306]	37 [12, 65]	13.1% [4.4%, 21.5%]
3b	15	9.15	343 [332, 354]	249 [230, 268]	282 [263, 303]	33 [7, 59]	11.5% [2.6%, 19.7%]
3c	15	13.5	355 [343, 366]	251 [229, 272]	279 [257, 298]	28 [0, 56]	10.1% [0.0%, 18.9%]
4a	20	7.23	347 [336, 359]	219 [201, 236]	257 [239, 277]	38 [13, 61]	15.0% [5.3%, 22.7%]
4b	20	10.4	356 [345, 368]	221 [203, 240]	258 [237, 278]	37 [13, 62]	14.3% [5.1%, 23.1%]
4c	20	15.2	369 [357, 381]	224 [205, 244]	255 [236, 275]	30 [3, 58]	12.0% [1.3%, 21.6%]
5a	25	8.05	360 [348, 371]	189 [173, 205]	224 [205, 242]	36 [10, 59]	16.0% [4.9%, 25.2%]
5b	25	11.55	369 [357, 381]	191 [176, 209]	226 [209,245]	36 [13, 58]	15.7% [6.0%, 24.6%]
5c	25	16.8	382 [369, 394]	195 [178, 213]	228 [209, 250]	32 [7, 59]	14.2% [3.1%, 25.1%]
6a	30	8.77	370 [358, 382]	160 [145, 174]	191 [174, 207]	30 [10, 52]	15.9% [5.6%, 25.6%]
6b	30	12.54	379 [367, 391]	164 [149, 178]	195 [177, 214]	32 [6, 55]	16.4% [3.4%, 26.3%]
6c	30	18.27	391 [379, 404]	170 [153, 186]	200 [180, 220]	30 [6, 56]	15.0% [3.2%, 26.5%]

(1) Results are reported as the median and the 95% confidence interval.

b) Females

Scenario	G(x|φ=1) (Median, mm)	G(x|φ=1) (Stdev, mm)	Total (G(x|φ=1) LCs	Total LC Deaths, CT Arm	Total LC Deaths , CXR Arm	Difference in Total Deaths (CXR-CT)	Difference in the Percentage of Total NSCLC Deaths (CXR-CT)/CXR
Obs.	**---**	**---**	**unobs.**	134 [127, 141]	178 [170, 185]	44 [33, 54]	25.0% [18.9%,29.7%]
1a	5	5.75	185 [177, 192]	173 [155, 193]	179 [168, 196]	7 [-14, 28]	3.7% [-8.0%, 14.9%]
1b	5	9.0	192 [183, 199]	170 [152, 189]	178 [163, 196]	8 [-13, 32]	4.3% [-7.8%, 16.9%]
1c	5	14.5	201 [192, 209]	168 [149, 187]	176 [159, 193]	8 [-16, 31]	4.5% [-9.3%, 16.8%]
2a	10	7.61	196 [187, 203]	161 [145, 178]	174 [159, 192]	13 [-7, 35]	7.7% [-4.3%, 18.8%]
2b	10	11.5	205 [195, 213]	159 [143, 177]	174 [156, 190]	15 [-7, 36]	7.5% [-4.4%, 19.2%]
2c	10	17.9	217 [207, 225]	158 [140, 177]	172 [156, 190]	14 [-9, 37]	7.9% [-5.7%, 20.5%]
3a	15	5.59	200 [191, 208]	143 [126, 160]	164 [149, 180]	22 [0,41]	13.0% [0.0%, 23.7%]
3b	15	9.15	209 [200, 217]	144 [127, 160]	165 [148, 182]	22 [2,42]	13.2% [1.2%, 24.2%]
3c	15	13.5	218 [208, 227]	144 [127, 161]	164 [149, 181]	21 [-2, 43]	12.7% [-1.3%, 24.4%]
4a	20	7.23	215 [205, 224]	120 [106, 135]	147 [132, 163]	27 [8, 47]	18.4% [5.4%, 29.9%]
4b	20	10.4	222 [212, 231]	122 [107, 135]	150 [135, 166]	29 [6, 47]	19.3% [4.8%, 30.0%]
4c	20	15.2	231 [221, 241]	126 [108, 142]	151 [135, 169]	25 [4, 48]	16.4% [2.7%, 30.2%]
5a	25	8.05	227 [216, 235]	96 [85, 109]	128 [115, 142]	32 [16, 48]	24.8% [13.1%, 35.4%]
5b	25	11.55	233 [223, 242]	101 [88, 113]	133 [118, 145]	32 [13,48]	23.9% [10.7%, 34.5%]
5c	25	16.8	242 [232, 252]	106 [92, 122]	135 [121,150]	29 [9,49]	21.7% [7.0%, 34.2%]
6a	30	8.77	235 [225, 245]	77 [67, 89]	110 [98, 122]	32 [18,49]	29.6% [17.3%, 41.5%]
6b	30	12.54	242 [231, 251]	82 [70, 93]	114 [100, 126]	32 [14,49]	28.0% [13.6%, 40.0%]
6c	30	18.27	250 [239, 260]	89 [76, 103]	118 [103, 133]	30 [10,47]	25.0% [8.8%, 37.2%]

(1) Results are reported as the median and the 95% confidence interval.

Figure Legends:

Figure 1: Estimated discrete hazard ratios governing detection between curable and incurable NSCLCs over a range of G(x|ϕ=1) conditional distributions among males and females.

a) Median of G(x|ϕ=1) = 5 mm, Standard Deviation of G(x|ϕ=1)=8.9 mm
b) Median of G(x|ϕ=1) =10 mm, Standard Deviation of G(x|ϕ=1)=11.5 mm
c) Median of G(x|ϕ=1) = 15 mm, Standard Deviation of G(x|ϕ=1)=13.5 mm
d) Median of G(x|ϕ=1) = 20 mm, Standard Deviation of G(x|ϕ=1)=15.2 mm
e) Median of G(x|ϕ=1) = 25 mm, Standard Deviation of G(x|ϕ=1)=16.8 mm
f) Median of G(x|ϕ=1) = 30 mm, Standard Deviation of G(x|ϕ=1)=18.27 mm

Figure 2: Estimated discrete hazard functions governing detection for curable and incurable NSCLCs over a range of G(x|ϕ=1) conditional distributions among males and females.

a) Median of G(x|ϕ=1) = 5 mm, Standard Deviation of G(x|ϕ=1)=8.9 mm
b) Median of G(x|ϕ=1) =10 mm, Standard Deviation of G(x|ϕ=1)=11.5 mm
c) Median of G(x|ϕ=1) = 15 mm, Standard Deviation of G(x|ϕ=1)=13.5 mm
d) Median of G(x|ϕ=1) = 20 mm, Standard Deviation of G(x|ϕ=1)=15.2 mm
e) Median of G(x|ϕ=1) = 25 mm, Standard Deviation of G(x|ϕ=1)=16.8 mm
f) Median of G(x|ϕ=1) = 30 mm, Standard Deviation of G(x|ϕ=1)=18.27 mm

Supplement Figure S1: Five-year NSCLC survival as a function of tumor size within each extent of disease category among males and females.

a) Males
b) Females

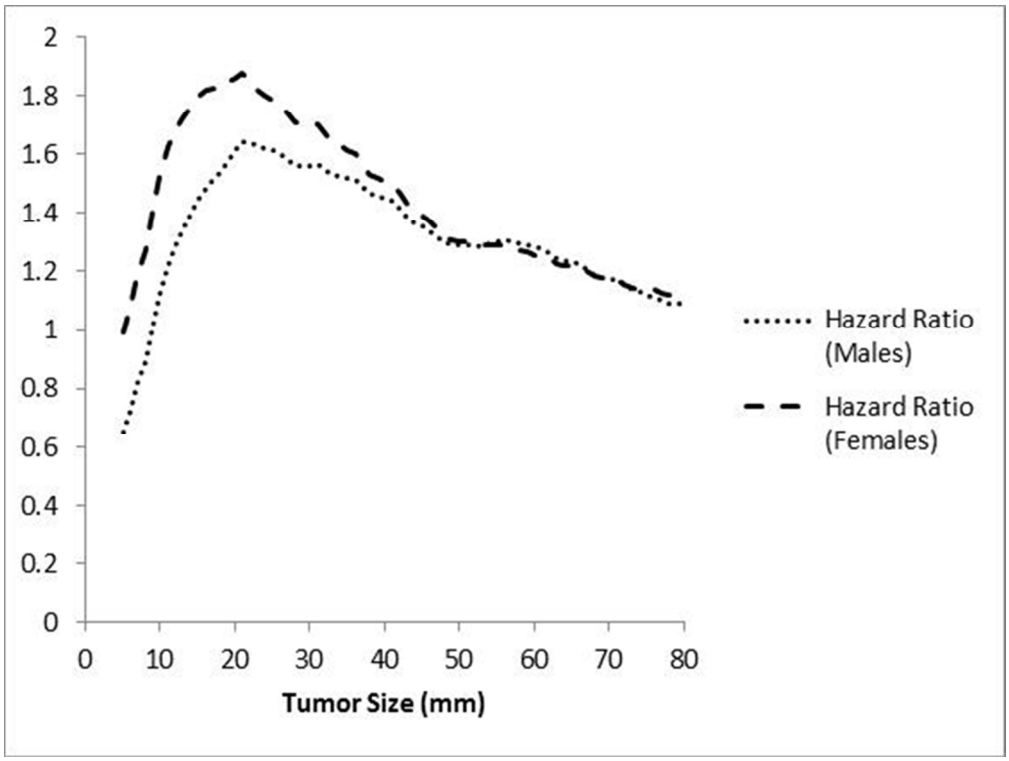

Figure 1a) Median of G(x|φ=1) = 5 mm, Standard Deviation of G(x|φ=1)=8.9 mm

254x190mm (96 x 96 DPI)

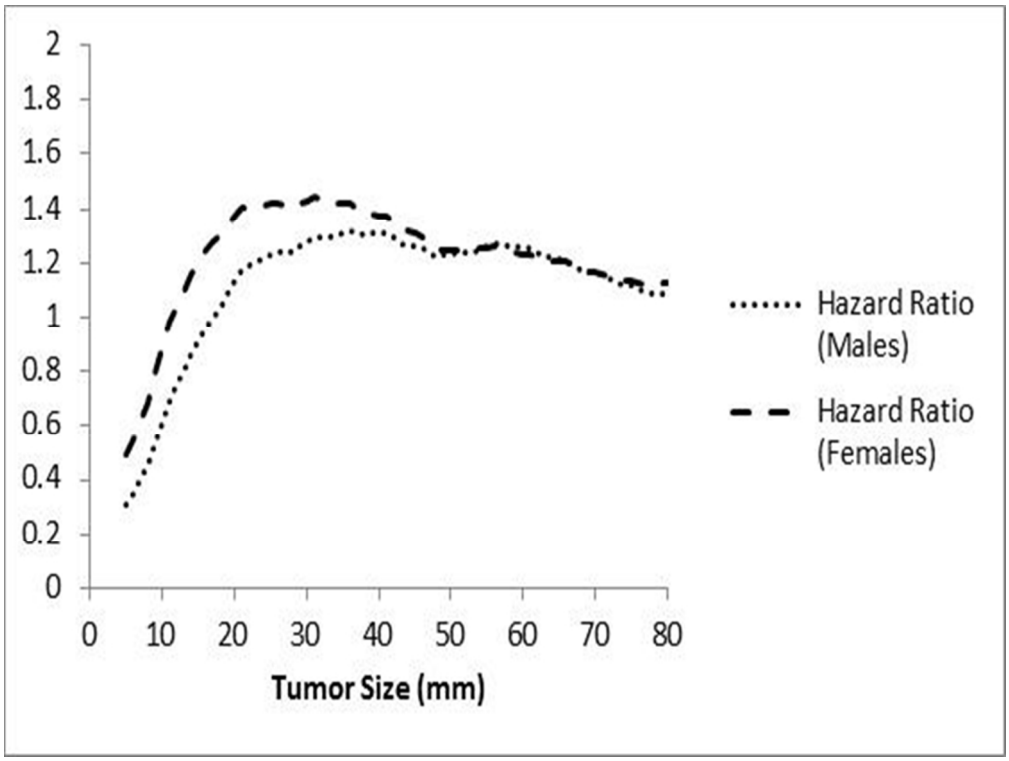

Figure 1b) Median of G(x|φ=1) =10 mm, Standard Deviation of G(x|φ=1)=11.5 mm

254x190mm (96 x 96 DPI)

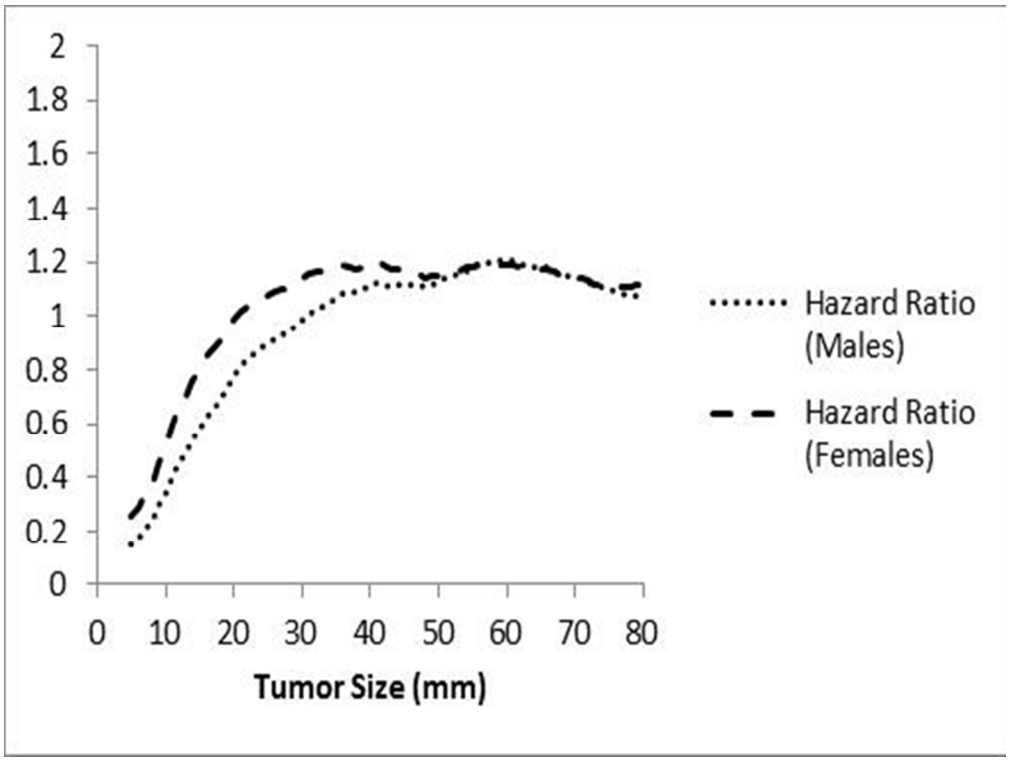

Figure 1c) Median of G(x|φ=1) = 15 mm, Standard Deviation of G(x|φ=1)=13.5 mm

254x190mm (96 x 96 DPI)

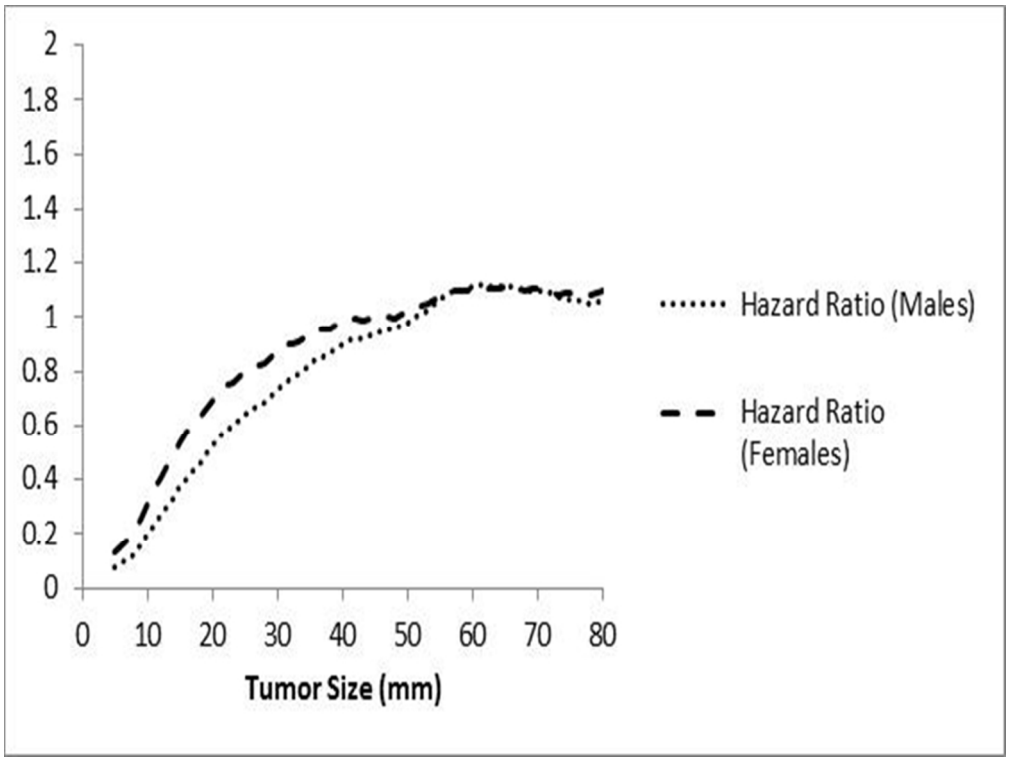

Figure 1d) Median of G(x|φ=1) = 20 mm, Standard Deviation of G(x|φ=1)=15.2 mm

254x190mm (96 x 96 DPI)

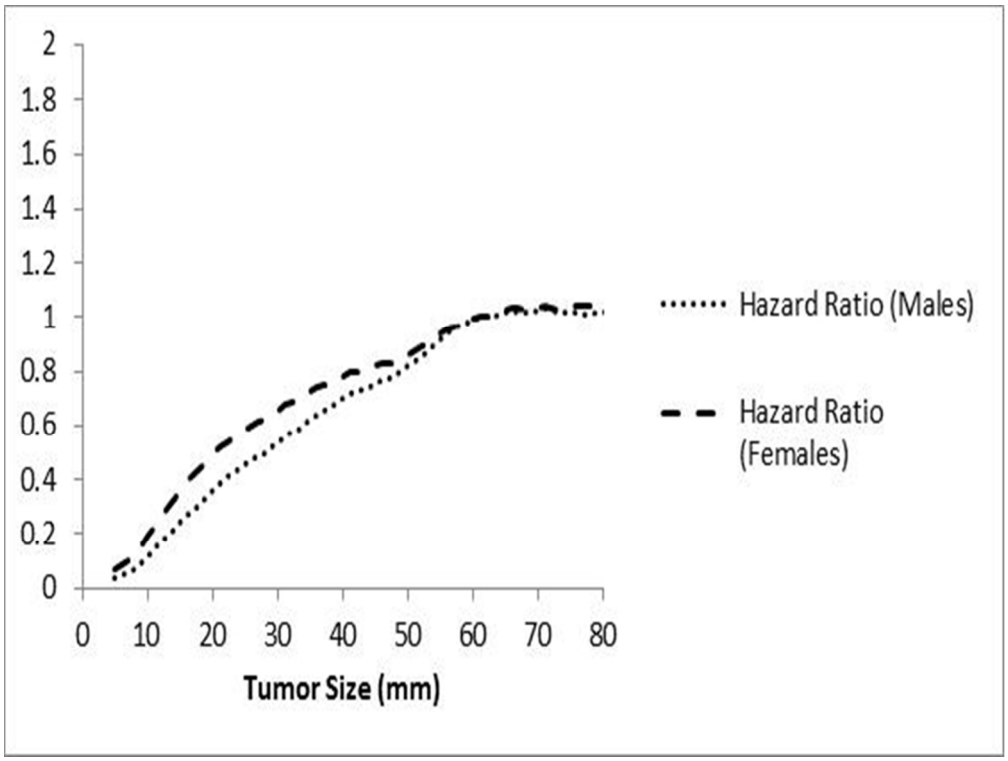

Figure 1e) Median of G(x|φ=1) = 25 mm, Standard Deviation of G(x|φ=1)=16.8 mm

254x190mm (96 x 96 DPI)

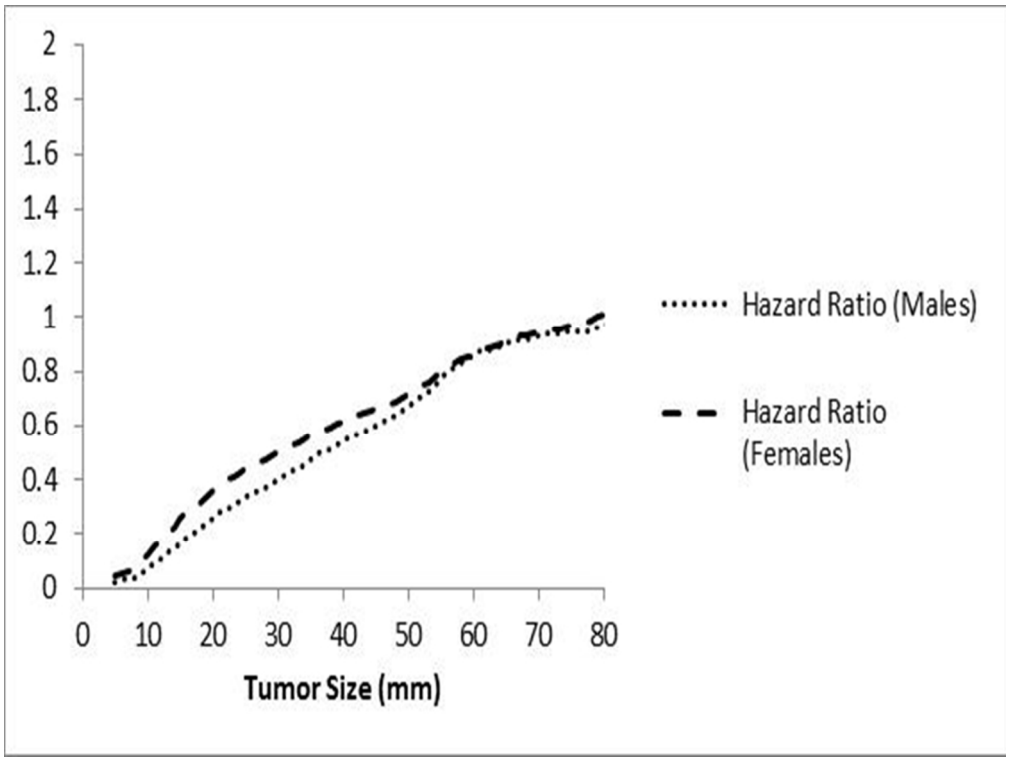

Figure 1f) Median of G(x|φ=1) = 30 mm, Standard Deviation of G(x|φ=1)=18.27 mm

254x190mm (96 x 96 DPI)

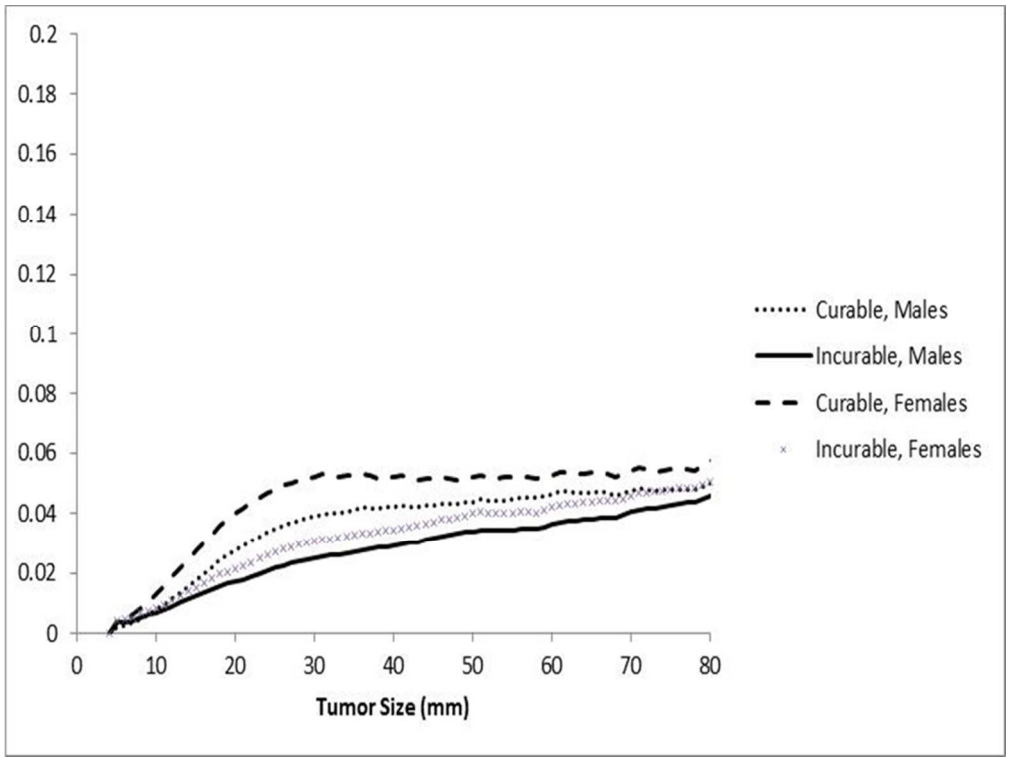

Figure 2a) Median of G(x|φ=1) = 5 mm, Standard Deviation of G(x|φ=1)=8.9 mm

254x190mm (96 x 96 DPI)

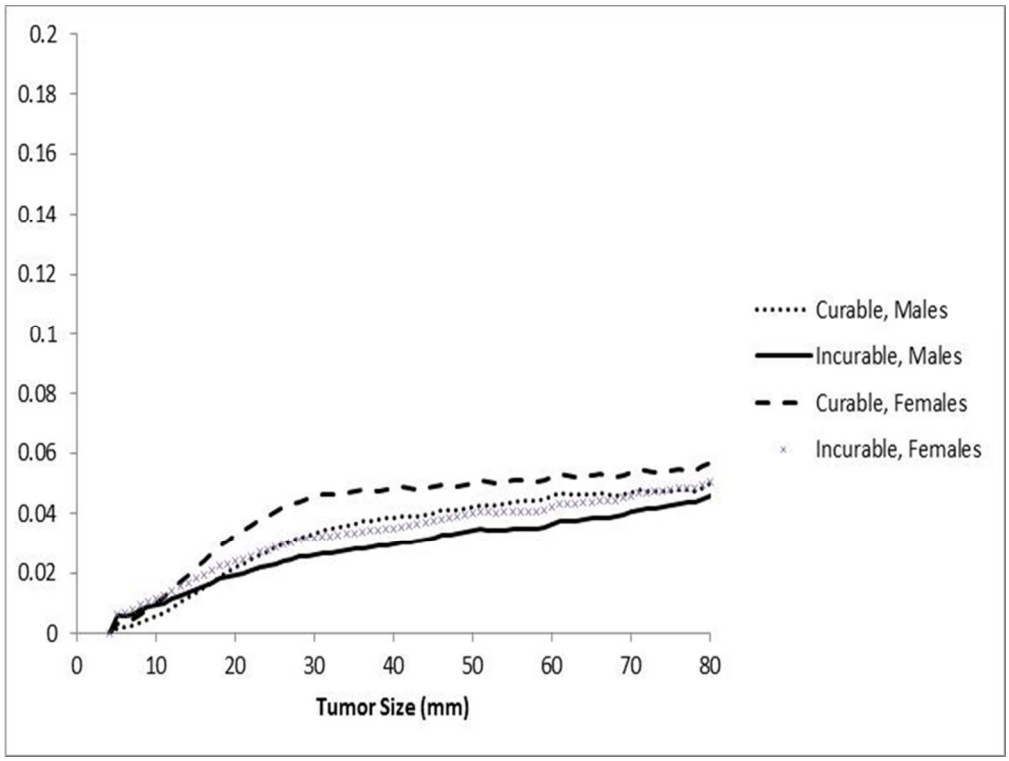

Figure 2b) Median of G(x|φ=1) =10 mm, Standard Deviation of G(x|φ=1)=11.5 mm

254x190mm (96 x 96 DPI)

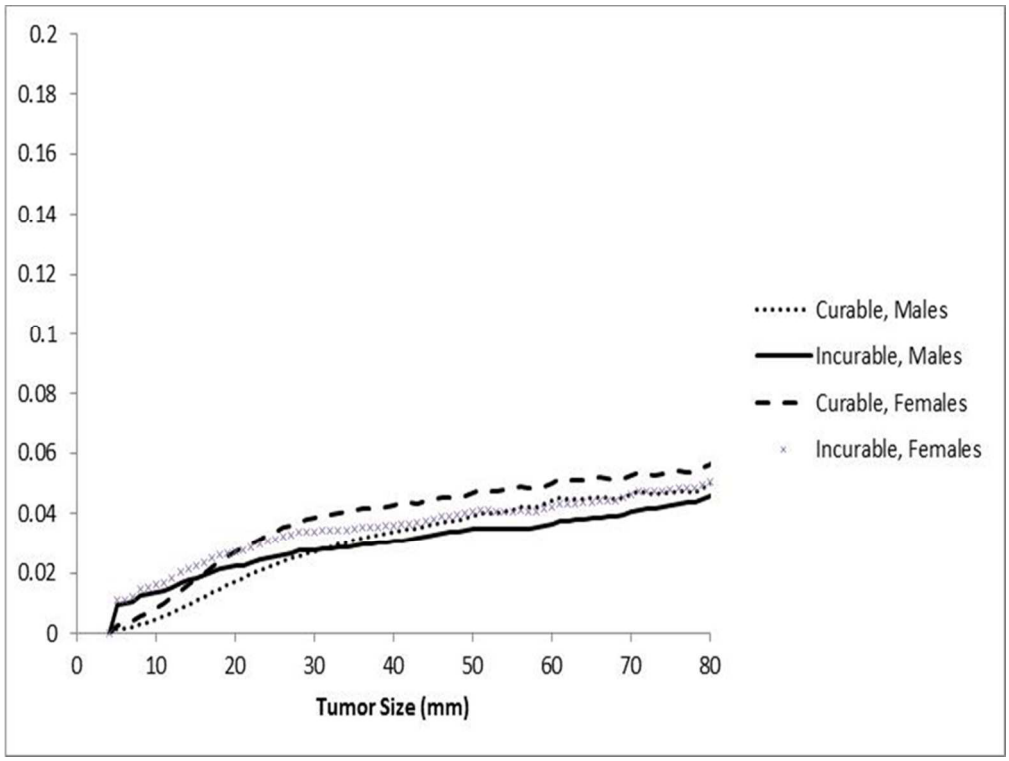

Figure 2c) Median of G(x|φ=1) = 15 mm, Standard Deviation of G(x|φ=1)=13.5 mm

254x190mm (96 x 96 DPI)

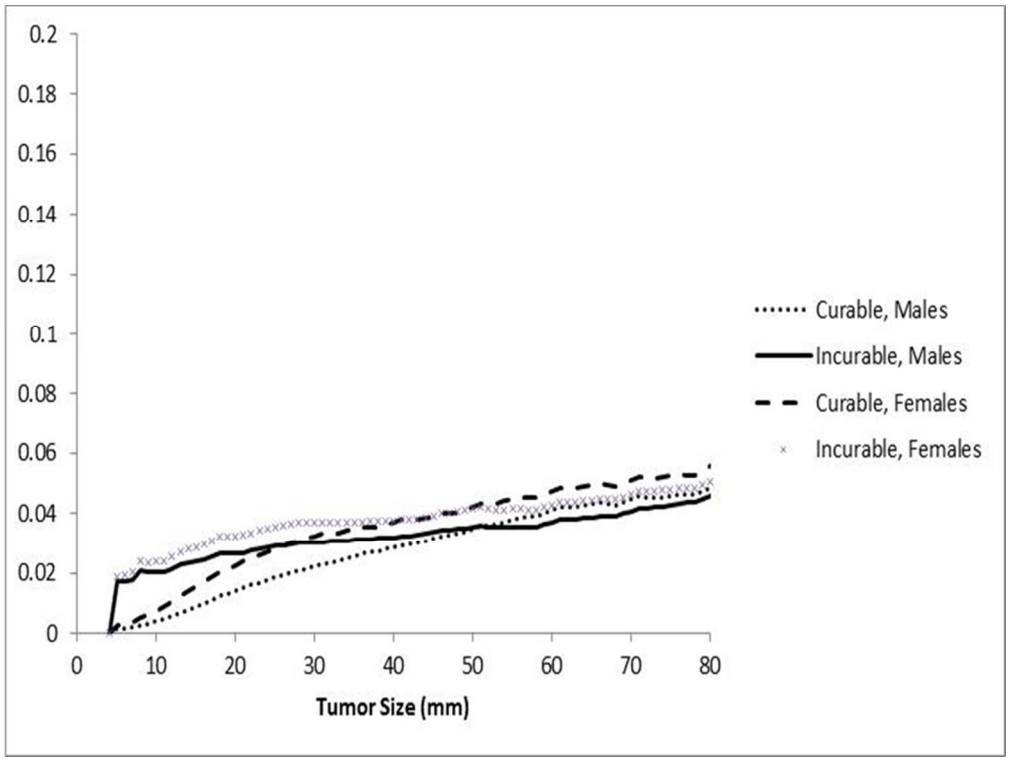

Figure 2d) Median of G(x|φ=1) = 20 mm, Standard Deviation of G(x|φ=1)=15.2 mm

254x190mm (96 x 96 DPI)

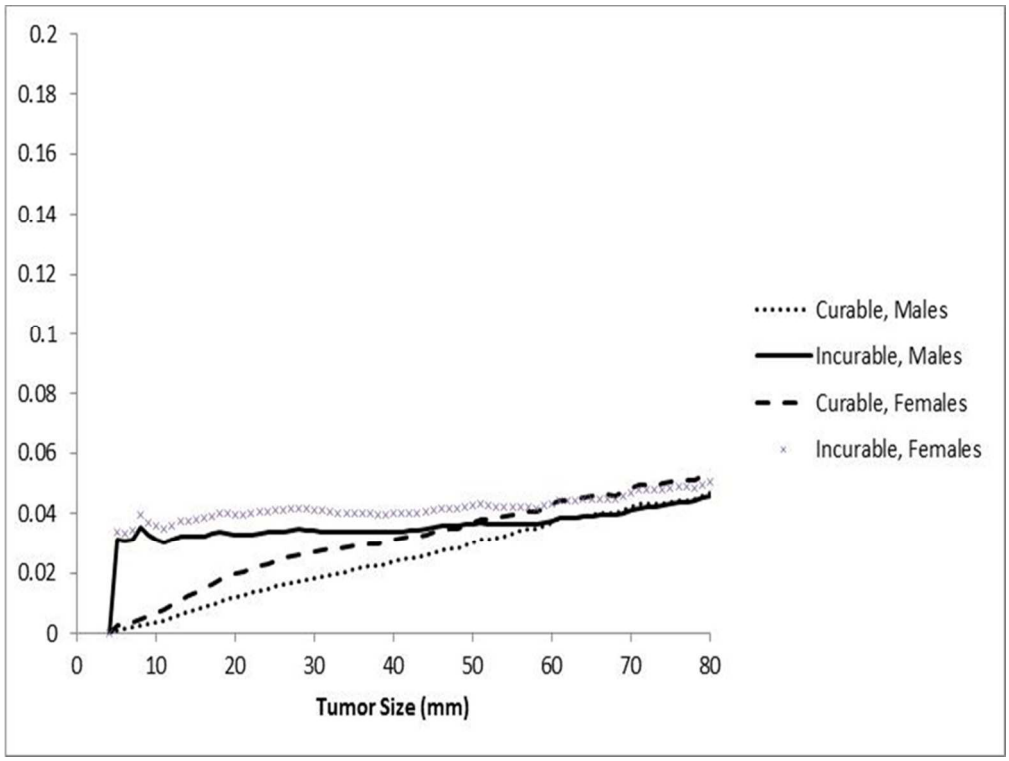

Figure 2e) Median of G(x|φ=1) = 25 mm, Standard Deviation of G(x|φ=1)=16.8 mm

254x190mm (96 x 96 DPI)

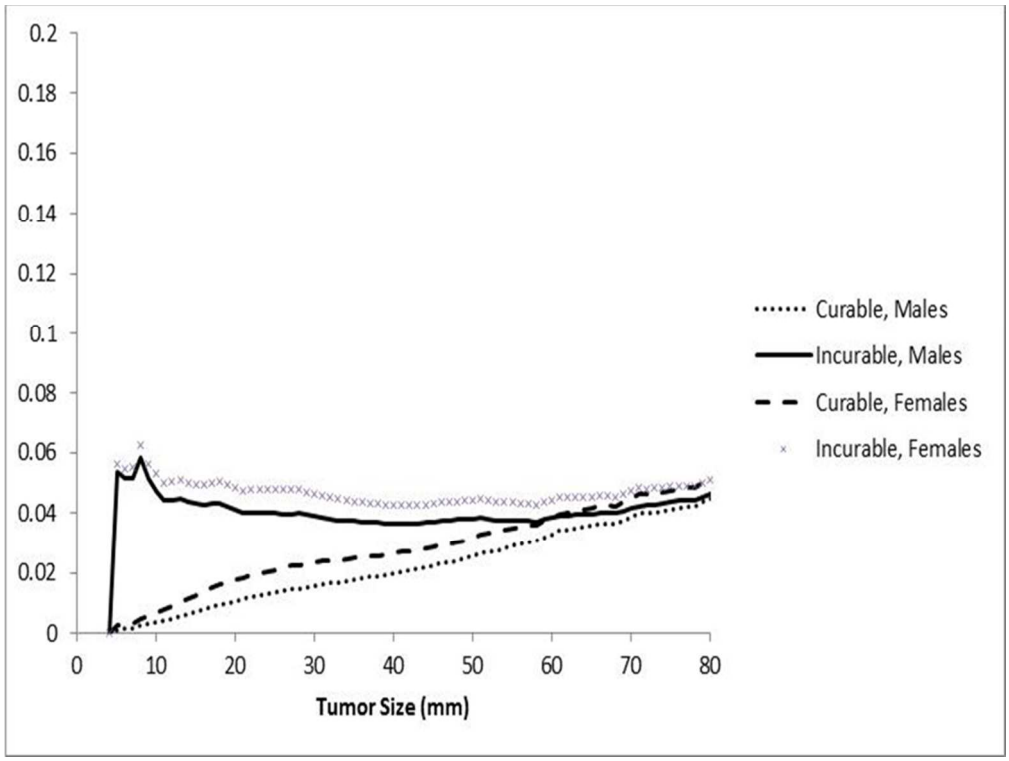

Figure 2f) Median of G(x|φ=1) = 30 mm, Standard Deviation of G(x|φ=1)=18.27 mm

254x190mm (96 x 96 DPI)