**Distribution Agreement**

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

| | |
|---|---|
| Mengdi Wu | 2016.04.04 |

[Association between Clinic Characteristics and Positive Low-dose CT outcome among current and Former Smoker]

By

[Mengdi Wu]
[MSPH]

[Department of Biostatistics and Bioinformatics]

_____ [Chair's signature]
[Ying Guo, PhD]
Committee Chair

# Association between Clinic Characteristics and Positive Low-dose CT outcome among current and Former Smoker

By

[Mengdi Wu]

B.S.
University of Wisconsin Madison
2014

Thesis Committee Chair: Ying Guo, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Department of Biostatistics and Bioinformatics
2016

# Abstract

Association between Clinic Characteristics and Positive Low-dose CT outcome among current and Former Smokers

By Mengdi Wu

**Background**: Lung cancer is one of the leading cause of death now in the United States. According to Cancer statistics reported by Greenlee et al. in 2000, 28% of death were due to lung cancer among deaths due to cancer in United States. Lung Imaging Reporting and Data System (Lung-RADS) is a scoring system developed by American College of Radiology to help better interpret CT screening and manage lung cancer diagnosis based on the screening characteristics.

**Objective**: The aim of the study is to identify clinic factors that associated with change of screening classification during screening trials and further modify current screening procedures.

**Method**: Univariate and Multivariate analysis were applied to identify potential risk factors that would be associated with suspicious lung screening outcomes. Model selection were based on deviance analysis.

**Results**: The results indicated that heavy smokers with current or past smoking behaviors, increasing age would lead to potential risk of detection of malignant nodules at the time of screening. Male smokers would be more likely to develop suspicious screening results compared to female smokers. At the same time, patients with affected siblings are more likely to be later detected for suspicious nodules on low-dose CT screening. For patients with certain clinic characteristics, we would suggest for more frequent CT screening in order to detect for malignant disease at earlier stage.

# Association between Clinic Characteristics and Positive Low-dose CT outcome among current and Former Smokers

By

[Mengdi Wu]

B.S.
University of Wisconsin Madison
2014

Thesis Committee Chair: Ying Guo, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in
Department of Biostatistics and Bioinformatics
2016.

**Acknowledgement:**

I would first like to thank my thesis advisor Dr. Ying Guo of Department of Biostatistics and Bioinformatics at Emory University]. The door to Prof. Guo's office was always open whenever I ran into a trouble spot or had a question about my research or writing.

I would also like to thank the experts who were involved in the research project: Dr. Ashish Sharma, Dr. Andrew Post, Dr. Fred Prior, Dr. Adam Bernheim and Dr. Williams Shannon. Without their passionate participation and input, the study could not have been successfully conducted.

Finally, I must express my very profound gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

The thesis study was supported by grant from NCI U01 [1U01CA187013-01], Resources for development and validation of Radiomic Analyses & Adaptive Therapy, Fred Prior, Ashish Sharma (UAMS, Emory).

Table of Contents

# Chapter 1: Introduction

## 1.1 Lung Cancer as a Public Health concern

Lung cancer is one of the leading cause of death now in the United States. According to Cancer statistics reported by Greenlee et al. in 2000, 28% of death were due to lung cancer among deaths due to cancer in United States (Greenlee et al., 2000). According to reports from American Cancer Society, there will be 221,200 new lung cancer diagnosis in United States (American Cancer Society, 2015). Researchers estimated 158,040 deaths due to lung cancer in 2015. Since 1950, there is strong evidence to prove that smoking was primary cause for lung cancer. Study conducted by Wynder and Graham in 1950 had indicated that cigarettes smoking lead to carcinoma of the lung (Wynder and Graham, 1950). Evidence shown that risk of lung cancer would increase with quantity of cigarettes smoking as well as duration of smoking.

## 1.2 Screening among high risk individual

Researchers provided evidence to show that screening with chest x ray was not effective in order to reduce mortality rate for lung cancer (Flehinger et al., 1984). Henschke et al. conducted Early Lung Cancer Action Project to evaluate effect of low-dosed CT for people at high risk (Henschke et al., 1999). Their research reported that 23% non-calcified nodules were detected by low-dose CT at baseline and have better performance in detecting malignant disease comparing to performance of chest radiography. Several studies also indicate low-dose spiral CT help detect lung cancer at an early stage and make the disease more curable since it can detect smaller non-calcified nodules comparing to traditional chest x ray. However, health care provider would like to

screen patients based on their risk for lung cancer. High risk population would require more frequent screening schedule while patients with relatively low risk would receive screening in a longer period in order to reduce the cost and harm from low-dose CT.

### 1.3 Current Scoring System for Screening Interpretation

Lung Imaging Reporting and Data System (Lung-RADS) is a scoring system developed by American College of Radiology to help better interpret CT screening and manage lung cancer diagnosis based on the screening. Lung-RADS scores can change over the course of multiple screens, and have a classic temporal phenotyping pattern — the current score is a function of nodule size in the following screening, as well as the Lung-RADS scores at baseline. At the same time, Lung-RADS also categorize patients for different screening schedule. Patients ranked higher in Lung RADS category would receive a more frequent low-dose CT screening. Lung-RADS criteria primary categorize patients based on size of nodule as well as nodule attenuation. Study proved that application of Lung RADS would largely increase the precision of predictive value for lung cancer incidence in population receiving low-dosed CT lung screening (Mckee et al., 2014). Lung RADS depends solely on screening characteristics, but clinic characteristics were not included. According to Lung RADS, category 1 is defined as low-dose CT with no nodules or with benign nodules. Patients with nodules that would be unlikely to develop to clinically active cancer would be classified as category 2. Lung RADS defined patients categorized in categorize 3 and 4 as positive screening results. However, many patients initially categorized as Category 1 or 2 at the beginning of screening trial later shown positive screening results at the end of study due to potential risk of lung cancer. Such change of status draw attention of researchers since it might be

associated with active clinical lung cancer. We would like to include potential clinic risk factors for lung cancer into Lung RADS in order to better predict such change status in a shorter period of time instead of after three year screening.

### 1.4 Study Aim

The aim of the study is to identify clinic factors that associated with change of screening classification during screening trials and further modify current screening procedures. Risk factors including clinic characteristics at baseline of screening identified in the study can help healthcare practitioner to predict patient's chance of developing appropriate protocols for screening programs as well as help researchers start programs for lung cancer prevention.

## Chapter 2: Literature Review: Risk Factors for Lung Cancer:

A-priori knowledge of what characteristics would be important to impact patients' risk for lung cancer would need to be obtained. Researchers would be interested in demographic characteristics, patient's medical history, smoking history as well as working exposure history.

### 2.1 Demographic Characteristics

#### 2.1.1 Gender

There is a debate that whether lung cancer incidence rate was higher among men or women. Previous research indicated that incidence rate of lung cancer was lower in female population compared to male population (Doll and Peto, 1981) due to lower exposure to tobacco smoking. Recently, there are studies in the United States suggesting

that risk of female for lung cancer associated with smoking is higher than male (Risch et al., 1993) since exposure of use of tobacco among female in the United States is almost the same as among male population (Ries et al., 1994). At the same time, study indicated that female are more vulnerable since genetic polymorphisms in gene CYP1A1, the gene associated with lung cancer metabolism, lead to increasing risk for lung cancer comparing to male without considering smoking history (Dresler et al., 2000).

### 2.1.2 Race

Since smoking prevalence among African American population is higher than among white population, researchers conducted studies to examine the association of lung cancer and race. Results indicated that African American were at higher risk for developing lung cancer compared to white with evidence that relative risk for lung cancer for African American is 1.8 comparing to white population (Harris et al., 1993).

### 2.1.3 Socioeconomic Status

Scientists also examined whether socioeconomic status would impact risk for lung cancer. In neighboring country Canada, study confirmed that there is an increase in risk for lung cancer among low income population with adjusted OR of 1.7 (95% CI: 1.3-2.2) (Mao et al., 2000). The study also conclude that people with less than 8 years of education as well as those with greater than 14 years education were less likely to develop lung cancer. Therefore, we believe that education level would be another risk factors for lung cancer.

## 2.2 Medical History

### 2.2.1 Body Mass Index

Body mass index (BMI) is usually calculated to estimate level of obesity. Usually, researchers believed that increasing BMI is positively associated with cancer incidence (Renehan et al., 2008). In other word, fatter people are more likely to develop cancer. However, for lung cancer incidence, researchers noticed that there is an increasing trend for risk of lung cancer among current and former smokers with decreasing BMI (Kabat and Wynder, 1992).

### 2.2.2 Family History of Lung Cancer

From the systematic review of relationship between family history and lung cancer conducted by Matakidou et al. in 2005 (Matakidou et al., 2005), researchers confirmed that based on 31 case-control studies and 17 cohort studies, individual risk of lung cancer would be increased if he or she has an relative that was diagnosed with lung cancer. Study provided statistics that individual with five or more affected first degree relatives for lung cancer would have odds ratio 2.7 higher comparing with general population (Brownson et al., 1997). At the same time, with two or more siblings diagnosed as lung cancer would be 1.4 times more likely to develop lung cancer. It is believed that familial history of lung cancer can lead to increasing individual risk for lung cancer due to genetic factors as well as exposure to similar environment (Mayenm Buenconsejo and Janerich, 1999). However, most studies identifying association between lung cancer and family history of lung cancer is conducted among non-smokers due to confounding effect of smoking history.

### 2.2.3 Previous diagnosis of Lung Disease

Researchers were looking for diagnosis of lung disease and their association with lung cancer. Studies based on patients from Johns Hopkins Lung Project and Intermittent Positive Pressure Breathing Trial proved that population suffered from chronic obstructive pulmonary disease (COPD) would experience increase risk for developing lung cancer (Tockman et al., 1987). Statistical analysis shown that increase in degree of airways obstruction would direct to increase of risk developing lung cancer among smokers. Such impact would be even higher comparing to the effect of level of smoking. At the same time, researchers were also interested in identifying fibrosis of lung as risk factors for lung cancer. Relative risk for lung cancer among patients diagnosed with pulmonary fibrosis is 7 to 14 comparing to the general population (Turner-Warwick et al., 1980).

2.3 Smoking History

Smoking is confirmed as the leading cause of lung cancer. 85%-90% lung cancer incidences in United States were estimated to be associated with smoking behavior. According to cohort study by Darby et al. in 2005, people smoke 15-24 cigarettes have 26 higher risk to develop lung cancer comparing with non-smokers (Darby et al., 2005). Another study conducted by Pope et al. shows that for people smoke more than 24 cigarettes would have at least 24 times higher risk to develop lung cancer (Pope et al., 2011). There is sufficient evidence to prove that increasing amount of cigarettes intake would lead to increasing risk of lung cancer. In the study, we decided to use pack-years, duration of smoking and age start smoking to measure patients' smoking pattern for both former and current smokers. At the same time, for former smokers, age that they quit smoking is regarded as an important factors impacting lung cancer risk. We also

considered the smoking history of cigar and pipes in the study since there is evidence that there is higher incidence of small-cell as well as squamous cell carcinomas among people smoking pipes and cigars.

## Chapter 3: Method

### 3.1 Data Source: National Lung Screening Trial (NLST)

NLST is a randomized screening trial with patients enrolled in either chest X-Ray and low-dose computed tomography (CT) (National Lung Screening Trial Team, 2011). 53454 patients with relative high risk for lung cancer were enrolled in 33 medical center in the United States. Eligible patients were aged between 55 and 74 years old at the beginning of randomization. All patients have smoking history of at least 30 pack-years. The study includes formers smokers that quit within 15 years. The screening trial excluded patients who have previously diagnosed with lung cancer as well as those who received chest X-ray within 18 months at the beginning of study. Each patient was randomly assigned for either chest X-Ray or low-dosed CT. From August 2002 to April 2004, patients received three annual screening based on their assigned screening group. 26722 patients received low-dose CT and 26732 patients received single-view poster-anterior chest radiography. According to NLST Team, 90% patients were adherent to the screening program. After randomization, patients participated in the screening trial filled in questionnaires asking for demographic, medical history, and smoking history. Our study would focus on low-dose CT since there is evidence shown that low-dose CT decrease mortality rate for lung cancer by 20% and has better performance in identifying smaller size of nodule (National Lung Screening Trial Team, 2011). At the same time, we

only focused on characteristics information collected at baseline to avoid potential correlation across serial screening observations for each individual.

### 3.2 Lung RADS and NLST

We applied Lung RADS to NLST population both at the beginning and at the end of screening trial. We matched patients for receiving low-dose CT screening at the two time points and compare their Lung RADS score. The study would focus on increase of Lung RADS ranking to positive screening results since such change represents potential risk for lung cancer that need more frequent screening schedule. We would like to identify risk factors for lung cancer and their association with such change in order to further modify the current scoring system.

### 3.3 Statistical Methods

The study is interested in learning about univariate and multivariate effect of potential risk factors for lung cancer discussed in the previous literature review section and their association with change of screening status. In order to identify association between potential risk factors and change of Lung RADS score, we performed Chi-Square test of independence to evaluate the univariate association. Continuous variable such as BMI, age were categorized into different levels. BMI were categorized into levels as underweight, normal, overweight and obese. Age were catergorized according to the quartile. We assessed normal assumption for Chi Square test by looking at histograms of the data. With relatively large sample size we assume that due to central limit theorem, the normality assumption for Chi Square test was satisfied. At the same time, contingency tables were created for Chi Square test with predictors and outcome

variables. With cell count greater than 5 in all contingency table, we believe that it would not be necessary to perform fisher's exact test to deal with sparse data. Crude odds ratio would be calculated for each risk factors, including 95% confidence interval. Such measurement shows association between risk factors and the outcome variables.

For covariates that were tested as significant in univariate analysis, we would want to include those variables into logistic regression model to learn about multivariate effect. Statisticians didn't recommend only include predictors that were considered as statistically significant with p-value less than 0.05 due to the possibility of residual confounding (Vittinghoff et al., 2005). Maldonado and Greenland suggested that potential confounders should be only eliminated with p-value greater than 0.2 (Maldonado and Greenland, 1993). Logistic regression model allows to estimate odds of whether a patient would be diagnosed with lung cancer or not. Final logistic regression model was selected based on method of deviance. First, researchers found variables that can reduce deviance significantly on its own, and fit a saturated model with those variables. Secondly, remove each individual variable from saturated model to see if any variable would become insignificant with the existence of other variables and drop such variable. Add back the omitted variables to the model and compared to the previous model to check if there is a significant increase in deviance to conclude the final model. After identify the final model, goodness of fit test was conducted to assess overall performance of the model. Pearson chi-squared goodness of fit test was tested to identify whether the final model has issue with lack-of-fit. Odds ratio were calculated to represent the association between predictors and outcome of lung cancer diagnosis since logistic regression were performed. Statistical analyses were performed in open-sourced

statistical programming software R and SAS software (version 9.4; SAS Institute Inc., Cary, NC).

# Chapter3: Results

## 3.1 Change in Lung RADS Classification

26449 patients enrolled in low-dose CT group in NLST study received initial screening at time of base line, while 25090 were screened at the end of screening trial. 25088 patients were matched for both baseline and end-of-study screening. At the beginning of trial, 18587 patients were classified as Category 1, which indicates patients without nodules or with benign nodules. 3004 patients were classified as Category 2. After calculation of Lung RADS score for at baseline and at end of NLST trial, 16141 patients were classified as Category 1 and 1643 patient were classified as Category 2 for both time points. According to Lung RADS categorization, both category were defined as negative screening results. At baseline, all patients were classified as Category 1, 2 or pending. At the end of study, we have observed shifts to Category 3, which is defined as positive screening results. Results of Lung RADS classification among NLST population enrolled in low-dose CT group were summarized in Table 2. 1051 patients shifted from Category 1 to Category 3 at the end of study while 492 patients shifted from Category 2 to Category 3. Such change would require a more frequent screening schedule changed from 12-month interval to 6-month interval.

## 3.2 Univariate Analysis

Characteristics of demographic, medical history, smoking history and working exposure history were demonstrated in Table 1 for patients changed from Category 1 to

category 3 and Table 2 for patients changed from category 2 to category 3. For patients shifted status to Catefory 3 from Category 1, gender was statistically significantly associated with such change. Results of Chi-square test indicated that odds of such change would be 1.11 times (95% CI: 0.98-1.25, p-value =0.09) higher among male patients compared to female patients. Race was not considered as important factors for patients' shift of Lung RADS classification from Category 1 to Category 3 since there is no sufficient statistical evidence. Although test result shows that education has weak association with suspicious screening outcome at the end of study, we won't exclude the effect of education from multivariate analysis. With p-value less than 0.001, we believe that age is another important factors for explaining the change of screening classification at the end of study. Comparing with patients age from 54 to 57, with increasing age, we can observe increasing odds of suspicious screening outcome at the end of study. Odds of change of classification from Category 1 to Category 3 at the end of study is 1.09 (95% CI: 0.88-1.34) times higher for patients from 57 to 60 years comparing with patients aged from 54-57. Odds would be 1.30 (95% CI: 1.08-1.58) times higher for patients aged from 60 to 65. For patients greater than 65 years old, the risk for suspicious screening outcome would be 1.78 (95% CI: 1.47-2.15) times higher comparing with patients from 54-57 years old. Medical history of patients seem not significantly associated with suspicious screening outcome at the end of study. However, there is significant association between suspicious screening outcome and sibling's history of lung cancer. Results of statistical test indicated that patients with affected siblings would be more likely to experience suspicious screening outcome. At the same time, current smokers are more likely to be categorized for higher Lung RADS score compared with former smokers. The odds for

higher Lung RADS score at the end of study would be 1.11 (95% CI: 0.98-1.25) times higher among current smokers compared with the odds among former smokers. With one unit increase of age starting smoking, the odds for shift to category 3 would be 0.96 (95% CI: 0.93-0.00) times lower. The amount of cigarettes smoked as well as the duration of smoking is not significantly associated with suspicious screening outcome. Also, for current and former smokers, living or working with smokers won't affect the change of developing positive screening outcome. Occupational exposure to asbestos won't impact chance of developing to positive screening outcome among current and former smokers.

Among patients initially detected with benign nodules (Category 2) changed to positive screening outcome (Category 3), we identified different factors that would impact such change. Education were no longer identified as important factors that lead to suspicious screening outcomes for current and former smokers initially screened with benign nodules. Gender is again statistically significant associated with the change to positive screening results from patients initially detected with benign nodules. The odds of such change is 1.26 time higher among male smokers comparing to female smokers. At the same time, both affected parents and siblings, smoking behaviors, occupational exposure as well as previous history of benign lung disease were not statistically significant factors for suspicious screening outcomes for current and former smokers initially screened with benign nodules. Patients were less likely to obtain suspicious screening results with initial screening with benign nodules if they have habit smoking with pipe (OR: 0.81, 95% CI: 0.64-1.01).

For high risk populations with benign nodules (Category 2) at baseline screening, age is still an important factors that raise chance for malignant finding for later screening.

With increasing age, the odds of being categorized in 3rd category would be increasing. Odds of change of classification from Category 2 to Category 3 at the end of study is 1.34 (95% CI: 0.97-1.87) times higher for patients from 57 to 60 years comparing with patients aged from 54-57. Odds would be 1.42 (95% CI: 1.05-1.95) times higher for patients aged from 60 to 65. For patients greater than 65 years old, the risk for suspicious screening outcome would be 1.66 (95% CI: 1.22-2.27) times higher comparing with patients from 54-57 years old. We would observed similar effect of age among patients classified as Category 1 and Category 2 at the beginning of study.

With strong statistical evidence (p-value = 0.0211), occupational exposure to asbestos has impact on chance of developing suspicious screening outcome. For patients who were exposed to asbestos in working environment, the chance of developing malignant lung disease that would be detected in final screening would be 1.65 (95% CI: 1.06 -2.51) times higher compared to people with normal BMI.

3.3 Multivariate Analysis

Based on analysis of deviance and model selection, we finally concluded two final logistic regression model to model the change from category 1 to category 3 and change from category 2 to category 3 during NLST screening process. Multivariate effect of risk factors associated with certain change were summarized in Table 4 and Table 5. According to results of lack-of-fit test, with p-value extremely close to 1, we don't have concerns of lack-of-fit for those two models. In other words, there is almost no difference between change of screening classification between observed data and fitted value.

The final model that identify effect of clinic factors on change from Category 1 to Category 3 includes effect of gender, age as well as history of lung cancer of individual's siblings. Adjusting for other characteristics, with affected siblings would increase chance for suspicious screening outcomes. Among high risk populations, the risk of developing positive screening outcome among would be 1.23 (95% CI: 1.03-1.60) higher compared to patients without affected siblings. As similar results in univariate analysis, in the final fitted model, increasing age leads to increasing risk of developing positive screening results. The final fitted model excludes effect of habits related to smoking history and medical history.

Different clinic characteristics is associated with change from Category 2 to Category 3 among NLST population. Similar to the previous model, increasing age is strongly associated with increasing risk of developing suspicious screening outcome adjusting other clinic characteristics. Male smokers were more likely to develop positive screening outcomes at the end of study. At the same time, comparing with white population, Asian is less likely to develop positive lung screening (OR: 0.086, 95% CI: 0.005-0.43). Smoking pipe is likely to reduce chance of increase of Lung RADS score (OR: 0.77, 95% CI: 0.61-0.97). Smokers that were exposed to asbestos in working environment have higher chance to be detected with suspicious nodules in low-dose CT screening (OR: 1.83, 95% CI: 1.15-2.83).

## Chapter 4: Discussion

Although smoking habit is known as an important risk factor for lung cancer, it is not significantly associated with developing positive screening result during the process of low-dose CT screening for current and former smokers. Such controversial finding

may be related to continuous smoking behavior. For heavy smokers included in NLST population, their smoking habits won't be changed during the three-year screening projects. Therefore, smoking habits including amount of cigarettes smoked per day as well as smoking duration would not be considered to differentiate patients for risk of developing positive screening results.

According to the two final fitted model in the study, we should differentiate patient's risk of developing suspicious lung screening results according to their age and gender. With increase in age, patients would be more likely to be detected with nodules that are associated with clinically active lung cancer, especially for patients aged greater than 65 years old. Current NLST screening procedure provides same screening schedule for all patients without considering their risk for lung cancer as well as their medical needs. Although Lung RADS successfully developed protocols to provide different screening frequency according to characteristics of screening image, our study indicated that it is necessary to add clinical characteristics such as gender, race as well as age in order to further modify the system and better detect suspicious nodule with reduced time.

According results of analysis, our study indicated that for heavy smokers with current or past smoking behaviors, increasing age would lead to potential risk of detection of malignant nodules at the time of screening. Male smokers would be more likely to develop suspicious screening results compared to female smokers. At the same time, patients with affected siblings are more likely to be later detected for suspicious nodules on low-dose CT screening. For patients with certain clinic characteristics, we would suggest for more frequent CT screening in order to detect for malignant disease at earlier stage.

Reference

Brady J. McKee, Shawn M. Regis, Andrea B. McKee, Sebastian Flacke, Christoph Wald, Performance of ACR Lung-RADS in a Clinical CT Lung Screening Program, Journal of the American College of Radiology, Volume 12, Issue 3, March 2015, Pages 273-276, ISSN 1546-1440, http://dx.doi.org/10.1016/j.jacr.2014.08.004.

Brownson, R. C., Alavanja, M. C., Caporaso, N., Berger, E., & Chang, J. C. (1997). Family history of cancer and risk of lung cancer in lifetime non-smokers and long-term ex-smokers. *International journal of epidemiology*, *26*(2), 256-263.

Darby S, Hill D, Auvinen A, et al. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies(link is external). BMJ 2005;330(7485):223.

Doll R and Peto R (1981) The causes of cancer: quantitative estimates of avoidable risksin the United States today. J Natl Cancer Inst 66: 1191–1308

Dresler, C. M., Fratelli, C., Babb, J., Everley, L., Evans, A. A., & Clapper, M. L. (2000). Gender differences in genetic susceptibility for lung cancer. *Lung Cancer*, *30*(3), 153-160.

Greenlee RT, Murray T, Bolden S, Wingo PA. Cancer statistics, 2000. CA Cancer J Clin 2000;50:7–33

Harris, R. E., Zang, E. A., Anderson, J. I., & Wynder, E. L. (1993). Race and sex differences in lung cancer risk associated with cigarette smoking. *International Journal of Epidemiology*, *22*(4), 592-599.

Henschke CI, McCauley DI, Yankelevitz DF, Naidich DP, McGuinness G, Miettinen OS, Libby DM, Pasmantier MW, Koizumi J, Altorki NK, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet. 1999 Jul 10;354(9173):99–105. [PubMed]

Kabat, G. C., & Wynder, E. L. (1992). Body mass index and lung cancer risk. *American journal of epidemiology*, *135*(7), 769-774.

Mao, Y., Hu, J., Ugnat, A. M., Semenciw, R., Fincham, S., & Canadian Cancer Registries Epidemiology Research Group. (2001). Socioeconomic status and lung cancer risk in Canada. *International Journal of Epidemiology*, *30*(4), 809-817.

Mayne, S. T., Buenconsejo, J., & Janerich, D. T. (1999). Familial cancer history and lung cancer risk in United States nonsmoking men and women. *Cancer Epidemiology Biomarkers & Prevention*, *8*(12), 1065-1069.

Matakidou, A., Eisen, T., & Houlston, R. S. (2005). Systematic review of the relationship between family history and lung cancer risk. *British journal of cancer*, *93*(7), 825-833.

MidthunDE, Swensen SJ, Jett JR, Hartman TE. Evaluation of nodules detected by screening for lung cancer with low dose spiral computed tomography. *Lung Cancer*2003;41(suppl 2):S40.

National Lung Screening Trial Research Team. (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening. *The New England journal of medicine*, *365*(5), 395.

Pope CA, Burnett RT, Turner MC, et al. Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: shape of the exposure-response relationships(link is external). Environ Health Persp 2011;119(11):1616-21.

Renehan, A. G., Tyson, M., Egger, M., Heller, R. F., & Zwahlen, M. (2008). Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *The Lancet*, *371*(9612), 569-578.

Risch HA, Howe GR, Jain M, Burch JD, Holowaty EJ and Miller AB (1993) Are female smokers at higher risk for lung cancer than male smokers? A case–control analysis by histologic type. Am J Epidemiol 138:281–293

Ries LA, Miller BA and Hankey BF (1994) SEER Cancer Statistics Review,1973–1991: Tables and Graphs. Bethesda: National Cancer Institute. NIH Publications, 264

TOCKMAN, M. S., ANTHONISEN, N. R., WRIGHT, E. C., & DONITHAN, M. G. (1987). Airways obstruction and the risk for lung cancer. *Annals of internal medicine*, *106*(4), 512-518.

Turner-Warwick, M., Lebowitz, M., Burrows, B., & Johnson, A. (1980). Cryptogenic fibrosing alveolitis and lung cancer. *Thorax*, *35*(7), 496-499.

US Department of Health and Human Services. (2004). The health consequences of smoking: a report of the Surgeon General. *Atlanta, GA: US Department of Health and*

*Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health*, *62*.

Vittinghoff, E., Glidden, D. V., Shiboski, S. C., & McCulloch, C. E. (2011). *Regression methods in biostatistics: linear, logistic, survival, and repeated measures models*. Springer Science & Business Media.

Wynder EL, Graham EA. Tobacco smoking as a possible etiologic factor in bronchogenic carcinoma. J Am Med Assoc 1950;143:329–36

**TABLE1: LUNG RADS CLASSIFICATION OF NLST AT BEGINNING AND END OF TRIAL**

| | | At the end of trial | | | | |
|---|---|---|---|---|---|---|
| | | Cat 1 | Cat 2 | Cat 3 | Pending | Total |
| **AT THE BEGINNING OF TIRAL** | Cat 1 | 16141 | 1299 | 1051 | 96 | 18587 |
| | Cat 2 | 793 | 1643 | 492 | 76 | 3004 |
| | Pending | 828 | 572 | 1975 | 122 | 3497 |
| | Total | 17762 | 3514 | 3518 | 294 | 25088 |

**TABLE 2: DISTRIBUTION OF STUDY SPECIFIC CHARACTERISTICS AMONG PATIENTS CHANGED FROM CAT 1 TO CAT 3**

| | Change From Cat 1 to Cat 3 n=18587 | | | |
|---|---|---|---|---|
| **VARIABLE** | Yes | No | p-value | Crude Odds Ratio (95% CI) |
| **N** | 1051 | 24037 | | |
| ***DEMOGRAPHIC*** | | | | |
| **GENDER** | | | 0.09 | |
| **MALE** | 670 | 10353 | | 1.11(0.98-1.25) |
| **FEMALE** | 381 | 7183 | | Reference |
| **RACE** | | | 0.4357 | |
| **WHITE** | 959 | 15901 | | Reference |
| **AMERICAN INDIAN OR ALASKAN NATIVE** | 2 | 64 | | 0.61(0.09-1.93) |
| **ASIAN** | 20 | 389 | | 0.92(0.57-1.41) |
| **BLACK OR AFRICAN-AMERICAN** | 55 | 821 | | 1.25(0.93-1.63) |
| **NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER** | 9 | 230 | | 0.68(0.33-1.25) |
| **MORE THAN ONE RACE** | 4 | 56 | | 1.40(0.42-3.39) |
| **EDUCATION** | | | 0.1589 | |
| **8TH GRADE OR LESS** | 13 | 206 | | Reference |
| **9TH-11TH GRADE** | 59 | 787 | | 1.27(0.71-2.45) |
| **ASSOCIATE DEGREE/ SOME COLLEGE** | 270 | 4141 | | 1.13(0.67-2.11) |
| **BACHELORS DEGREE** | 175 | 3056 | | 1.00(0.58-1.87) |
| **GRADUATE SCHOOL** | 133 | 2588 | | 0.89(0.52-1.69) |
| **HIGH SCHOOL GRADUATE/GED** | 130 | 3943 | | 1.05(0.62-1.95) |

| | | | | |
|---|---|---|---|---|
| **POST HIGH SCHOOL TRAINING, EXCLUDING COLLEGE** | 130 | 2468 | | 0.89(0.52-1.68) |
| **OTHER** | 19 | 323 | | 1.04(0.51-2.20) |
| **AGE** | | | <0.001 | |
| **54-57** | 169 | 3681 | | Reference |
| **57-60** | 210 | 4189 | | 1.09(0.88-1.34) |
| **60-65** | 322 | 5385 | | 1.30(1.08-1.58) |
| **>65** | 350 | 4281 | | 1.78(1.47-2.15) |
| ***MEDICAL HISTORY*** | | | | |
| **BMI** | | | 0.1576 | |
| **NORMAL** | 300 | 4802 | | Reference |
| **OBESE** | 289 | 5085 | | 0.91(0.77-1.07) |
| **OVERWEIGHT** | 439 | 7463 | | 0.94(0.81-1.09) |
| **UNDERWEIGHT** | 14 | 136 | | 1.66(0.91-2.82) |
| **PARENTAL HISTORY OF LUNG CANCER** | | | 0.5676 | |
| **YES** | 171 | 2740 | | 1.05(0.88-1.24) |
| **NO** | 846 | 17036 | | Reference |
| **SIBLING HISTORY OF LUNG CANCER** | | | 0.0174 | |
| **YES** | 96 | 1262 | | 1.30(1.04-1.61) |
| **NO** | 921 | 15757 | | Reference |
| **DIGNOSED WITH COPD** | | | 0.2762 | |
| **YES** | 60 | 872 | | 1.16(0.88-1.51) |
| **NO** | 984 | 16607 | | Reference |
| **DIAGNOSED WITH FIBROSIS OF THE LUNG** | | | 0.2313 | |
| **YES** | 2 | 34 | | 1.05(0.15-3.48) |
| **NO** | 1044 | 17445 | | Reference |
| **DIAGNOSED WITH PNEUMONIA** | | | 0.512 | |

| | | | | |
|---|---|---|---|---|
| **YES** | 220 | 3842 | | 0.94(0.82-1.10) |
| **NO** | 824 | 13637 | | Reference |
| ***SMOKING HISTORY*** | | | | |
| ***CURRENT SMOKE STATUS*** | | | 0.0994 | |
| **FOMER SMOKER** | 481 | 8484 | | Reference |
| **CURRENT SMOKER** | 570 | 9052 | | 1.11(0.98-1.25) |
| **PACK YEAR** | 56.67 | 55.77 | 0.229 | 1.01(1.00-1.01) |
| **AGE START SMOKING** | 16.11 | 16.53 | 0.0097 | 0.96(0.93-0.99) |
| **PIPE** | | | 0.2559 | |
| **YES** | 258 | 4061 | | 1.08(0.94-1.25) |
| **NO** | 783 | 13305 | | Reference |
| **CIGAR** | | | 0.1385 | |
| **YES** | 245 | 3792 | | 0.91(0.72-1.15) |
| **NO** | 18724 | 13635 | | |
| **LIVED WITH SMOKERS** | | | 0.9058 | |
| **YES** | 911 | 15229 | | 0.99(0.75-1.32) |
| **NO** | 133 | 2189 | | Reference |
| **WORK WITH SMOKERS** | | | 0.6581 | |
| **YES** | 359 | 14957 | | 1.04(0.87-1.25) |
| **NO** | 57 | 2429 | | Reference |
| ***WORK EXPOSURE*** | | | | |
| **HAS EVER WORKED FOR 1 YEAR OR MORE WITH ASBESTOS** | | | 0.5999 | |
| **YES** | 46 | 859 | | 0.92(0.67-1.24) |
| **NO** | 1004 | 16632 | | Reference |

**TABLE 3: DISTRIBUTION OF STUDY SPECIFIC CHARACTERISTICS AMONG PATIENTS CHANGED FROM CAT 2 TO CAT 3**

| | Change from Cat 2 to Cat3 n=3004 | | | | |
|---|---|---|---|---|---|
| **VARIABLE** | Yes | No | p-value | Crude Odds | 95% CI |
| **N** | 492 | 24596 | | | |
| ***DEMOGRAPHIC*** | | | | | |
| **GENDER** | | | 0.0163 | | |
| **MALE** | 287 | 1317 | | 1.26(1.04-1.54) | |
| **FEMALE** | 205 | 1195 | | Reference | |
| **RACE** | | | 0.00133 | | |
| **WHITE** | 465 | 2324 | | Reference | |
| **AMERICAN INDIAN OR ALASKAN NATIVE** | 0 | 10 | | 0.98(0.95-1.01) | |
| **ASIAN** | 1 | 58 | | 0.98(0.97-0.99) | |
| **BLACK OR AFRICAN-AMERICAN** | 20 | 69 | | 1.00(0.99-1.01) | |
| **NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER** | 1 | 7 | | 0.99(0.96-1.03) | |
| **MORE THAN ONE RACE** | 3 | 39 | | 0.99(0.97-1.00) | |
| **EDUCATION** | | | 0.8364 | | |
| **8TH GRADE OR LESS** | 8 | 26 | | Reference | |
| **9TH-11TH GRADE** | 25 | 105 | | 1.00(0.98-1.01) | |
| **ASSOCIATE DEGREE/ SOME COLLEGE** | 105 | 587 | | 0.99(0.98-1.00) | |

| | | | | |
|---|---|---|---|---|
| **BACHELORS DEGREE** | 83 | 423 | | 0.99(0.98-1.01) |
| **GRADUATE SCHOOL** | 65 | 368 | | 0.98(0.97-1.01) |
| **HIGH SCHOOL GRADUATE/GED** | 124 | 592 | | 1.00(0.98-1.01) |
| **POST HIGH SCHOOL TRAINING, EXCLUDING COLLEGE** | 70 | 371 | | 0.99(0.98-1.01) |
| **OTHER** | 11 | 25 | | 1.00(0.98-1.02) |
| **AGE** | | | 0.0132 | |
| **54-57** | 67 | 476 | | Reference |
| **57-60** | 112 | 590 | | 1.34(0.97-1.87) |
| **60-65** | 153 | 762 | | 1.42(1.05-1.95) |
| **>65** | 160 | 684 | | 1.66(1.22-2.27) |
| ***MEDICAL HISTORY*** | | | | |
| **BMI** | | | 0.4915 | |
| **NORMAL** | 150 | 741 | | Reference |
| **OBESE** | 122 | 696 | | 0.87(0.67-1.12) |
| **OVERWEIGHT** | 211 | 1047 | | 1.00(0.79-1.25) |
| **UNDERWEIGHT** | 6 | 20 | | 1.51(0.53-3.64) |
| **PARENTAL HISTORY OF LUNG CANCER** | | | 0.2184 | |
| **YES** | 86 | 381 | | 1.06(0.73-1.49) |
| **NO** | 396 | 2062 | | Reference |
| **SIBLING HISTORY OF LUNG CANCER** | | | 0.73 | |
| **YES** | 42 | 202 | | 1.115(0.82-1.57) |
| **NO** | 440 | 2246 | | Reference |
| **DIGNOSED WITH COPD** | | | 0.3830 | |
| **YES** | 21 | 131 | | 0.81(0.49-1.28) |
| **NO** | 469 | 2373 | | |
| **DIAGNOSED WITH FIBROSIS OF THE LUNG** | | | 1 | |

| | | | | | |
|---|---|---|---|---|---|
| | **YES** | 0 | 4 | | - |
| | **NO** | 490 | 2500 | | Reference |
| **DIAGNOSED WITH PNEUMONIA** | | | | 0.3028 | |
| | **YES** | 117 | 545 | | 1.12(0.90-1.41) |
| | **NO** | 373 | 1959 | | Reference |
| ***SMOKING HISTORY*** | | | | | |
| ***CURRENT SMOKE STATUS*** | | | | 0.3078 | |
| **FOMER SMOKER** | | 245 | 1314 | | Reference |
| **CURRENT SMOKER** | | 247 | 1198 | | 0.92(0.74-1.10) |
| **PACK YEAR** | | 56.18 | 56.30 | 0.9254 | 1.00(0.997-1.004) |
| **AGE START SMOKING** | | 16.65 | 16.52 | 0.8937 | 0.99(0.97-1.01) |
| **PIPE** | | | | 0.0632 | |
| | **YES** | 96 | 582 | | 0.81(0.64-1.01) |
| | **NO** | 389 | 1906 | | Reference |
| **CIGAR** | | | | 0.3453 | |
| | **YES** | 97 | 535 | | 0.90(0.71-1.12) |
| | **NO** | 390 | 1957 | | Reference |
| **LIVED WITH SMOKERS** | | | | 0.9547 | |
| | **YES** | 426 | 2177 | | 0.99(0.76-1.31) |
| | **NO** | 62 | 319 | | Reference |
| **WORK WITH SMOKERS** | | | | 0.5473 | |
| | **YES** | 422 | 2149 | | 1.08(0.84-1.43) |
| | **NO** | 63 | 338 | | Reference |
| ***WORK EXPOSURE*** | | | | | |
| **HAS EVER WORKED FOR 1 YEAR OR MORE WITH ASBESTOS** | | | | 0.0211 | |
| **YES** | | 29 | 92 | | 1.65(1.06-2.51) |

| | | | |
|---|---|---|---|
| **NO** | 461 | 2412 | Reference |

**TABLE 4: SUMMARY OF MULTIVARIATE EFFECT OF FACTORS FOR CHANGE FROM CAT 1 TO CAT 3**

| **VARIABLES** | Adjusted Odds Ratio | 95% CI |
|---|---|---|
| **AMOUNT OF CIGARETTES SMOKED PER DAY** | 1.004 | 0.99-1.01 |
| **SIBLING HISTORY OF LUNG CANER** | | |
| **NO** | Reference | |
| **YES** | 1.23 | 1.03-1.60 |
| **GENDER** | | |
| **FEMALE** | Reference | |
| **MALE** | 1.23 | 1.08-1.41 |
| **AGE** | | |
| **54-57** | Reference | |
| **57-60** | 1.06 | 0.85-1.31 |
| **60-65** | 1.22 | 1.006-1.49 |
| **>65** | 1.56 | 1.28-1.89 |

**TABLE 5: SUMMARY OF MULTIVARIATE EFFECT OF FACTORS FOR CHANGE FROM CAT 2 TO CAT 3**

| VARIABLES | Adjusted Odds Ratio | 95% CI |
|---|---|---|
| **AGE** | | |
| **54-57** | Reference | |
| **57-60** | 1.29 | 1.02-1.93 |
| **60-65** | 1.41 | 1.09-1.98 |
| **>65** | 1.62 | 1.33-2.40 |
| **RACE** | | |
| **WHITE** | Reference | |
| **AMERICAN INDIAN OR ALASKAN NATIVE** | <0.001 | 0-0.00383 |
| **ASIAN** | 0.086 | 0.005-0.43 |
| **BLACK OR AFRICAN-AMERICAN** | 1.52 | 0.87-2.51 |
| **NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER** | 0.42 | 0.03-4.01 |
| **MORE THAN ONE RACE** | 0.71 | 0.10-1.81 |
| **PIPE** | | |

| | | | |
|---|---|---|---|
| **NO** | Reference | | |
| **YES** | | 0.77 | 0.59-1.72 |
| **GENDER** | | | |
| **FEMALE** | Reference | | |
| **MALE** | | 1.32 | 0.92-1.82 |
| **OCCUPATIONAL EXPOSURE TO ASBESTOS** | | | |
| **NO** | Reference | | |
| **YES** | | 1.83 | 1.15-2.83 |