Development of a lung cancer prediction model for surgeons and factors affecting its national application

By

Stephen A. Deppen

Dissertation

Submitted to the Faculty of the Graduate School of Vanderbilt University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In

Epidemiology

August, 2013

Nashville, Tennessee

Approved:

Professor Eric L. Grogan

Professor Melinda Aldrich

Professor Jeffrey Blume

Professor Melissa McPheeters

Professor Pierre P. Massion

SHORT ABSTRACT

Lung cancer is deadly, killing more people than breast, colon and prostate cancer combined. Surgeons evaluating patients for lung cancer face a dilemma: to operate and subject the individual to operation associated mortality and morbidity or not operate and possibly miss early diagnosis and treatment. No models designed for surgeons evaluating lung lesions. We successfully estimated the TREAT model. A model designed for surgeons with an internally validated AUC of 0.87 and Brier score of 13.

If the TREAT model is applied to a national population, its accuracy may decrease due to local conditions. To determine the possible extent of such variation, benign disease prevalence after lung surgery was estimated using 2009 Medicare hospital discharge data. Significant variation in benign disease prevalence between states was observed with a low of 1.3% in Vermont and a high of 25% in Hawaii. The causes for this observed variation are unknown. Residence in a county with high fungal lung disease prevalence was not associated with increased likelihood of benign disease.

FDG-PET scan variance was observed in the national ACOGOS Z4031 trial. FDG-PET sensitivity (82%) and specificity (31%) were significantly lower than in previous published studies. Granuloma occurred in 68% of the false positive FDG-PET scans and sensitivity varied significantly between sites. Scan accuracy increased with increasing lung lesion size. Whether the observed variation is caused by practice variation, referral patterns, fungal lung disease, or other factors is unknown.

A meta-analysis examined FDG-PET accuracy to diagnose lung lesions sought to determine if other researchers had observed variance in FDG-PET accuracy. Seven studies reported false

positive scans arising from granulomas caused by infectious lung disease. Specificity of those studies was 59%, significantly lower than the specificity (77%) observed in the remaining 53 studies. Studies whose mean lesion size was less than or equal to 20 mm had significantly lower sensitivity than studies conducted in larger lesions.

The TREAT model shows clinical promise and should be externally validated. The causes of observed variation in benign disease prevalence and FDG-PET accuracy should be investigated with particular attention made to measuring infectious disease exposures that cause granulomas.

ABSTRACT

More people die in the United States (US) from lung cancer than from breast, colon, and prostate cancer combined with over 160,000 deaths occurring annually. Early detection, diagnosis and treatment improve cancer survival. Screening for early detection of disease with low dose computed tomography (LDCT) reduced lung cancer related mortality by 20% in a recently concluded national trial; however, screening resulted in many false positive results requiring further evaluation. Surgeons evaluating patients for lung cancer face a dilemma: to operate and potentially subject the individual to significant operation associated mortality and morbidity or not operate and possibly miss early diagnosis and opportunities for treatment. F^{18}-fluorodeoxyglucose positron emission tomography (FDG-PET) is one of the most accurate, non-invasive tests available for diagnosing lung nodules. However, lung granulomas with an infectious disease etiology appear physiologically and metabolically similar to cancer and generate false positive CT and FDG-PET scans. If an association exists between imaging detected granulomas and the infectious lung diseases that created them, then populations may exist that will not benefit from FDG-PET scans for diagnosis of lung cancer. If the prevalence of benign diagnosis after lung surgery varies across the US, then the benefit from a national screening program will likely be less in areas with higher benign disease prevalence. Exploration of possible causes of regional variation in benign disease prevalence and creation of a predictive model for surgeons to diagnose lung cancer has the potential to reduce the morbidity and mortality arising from unnecessary surgery for patients with benign disease.

Current models estimating lung cancer risk were not developed in populations being evaluated by surgeons. Using the diagnostic information available to surgeons at the point of decision to operate that includes the diagnostic work-up of all previous specialists combined, the TREAT

Lung Cancer model was estimated. Clinical and radiographic imaging information was retrospectively collected from a population being evaluated for known or suspected lung cancer by surgeons at Vanderbilt University Medical Center. The predictors for lung cancer included: age, sex, smoking history, pre-operative symptoms, previous history of cancer, body mass index, predicted forced expiratory volume in one second, lesion characteristics, and FDG-PET avidity. The TREAT model better discriminated lung cancer (AUC=0.89, 95%CI: 0.86-0.92) in the surgical evaluation population than the previously validated Mayo model (AUC=0.80, 95%CI: 0.76-0.85). The TREAT model was better calibrated using the Brier score metric (0.12) when compared to the Mayo model (0.17). When the variables from the Mayo model were re-estimated using the Vanderbilt population (AUC 0.83, 95% CI: 0.79-0.87), the more comprehensive TREAT model continued to better discriminate cancer from benign disease ($p<0.001$). Lesion size, age, avidity on a FDG-PET scan, and observed growth on serial CT scans were the strongest predictors for lung cancer. The model had little reduction in AUC (0.87) or Brier score (0.12) on internal validation with bootstrap methodology.

The TREAT model to predict lung cancer was more accurate than previously published models. Its 89% AUC is the highest among all published models. However, to be clinically useful in a surgical evaluation population with its high prevalence of lung cancer, all variables that might indicate benign disease or influence the primary predictive variables of growth or FDG-PET avidity must be explored. Granulomas arising from infectious lung diseases including histoplasmosis, coccidiomycosis, blastomycosis and tuberculosis have been reported in the literature and generate false positive imaging scans for lung cancer. Granulomas were the most common benign diagnosis after surgery for suspected lung cancer in this research, and they were 60% to 70% of the etiologies observed with false positive FDG-PET scans. Also, individuals with clinical early stage disease, a population similar to that likely found in the successful national CT screening program, have smaller lung nodules. Published literature and clinical data from a

national trial showed that FDG-PET scans failed to differentiate cancer from benign disease in populations with smaller lesions or with endemic infectious lung disease. This research calls into question the use of FDG-PET scans for diagnosis in populations with lesions under 2cm or having a high prevalence of infectious lung disease.

To investigate the possible range of variability in benign disease, benign disease point prevalence after surgical operation was investigated by state in the 2009 Medicare population. Each individual was matched to a probability of fungal lung disease prevalence as measured in a national surveillance program from the 1960s. Benign disease prevalence variation was observed between states, from a low of 1.3% in Vermont to a high of 25% in Hawaii. The median point prevalence by state was 8.9% (IQR: 7.8 – 10.9; $chi^2<0.001$). Historic fungal lung disease prevalence by county of residence at the time of surgery was not associated with benign disease at the individual level (p=0.9).

In a secondary analysis examining the variation of FDG-PET scan diagnostic accuracy in 682 individuals from a completed national trial, lower sensitivity and specificity was observed (82% and 31% respectively) compared to previously published studies. Wide variation in sensitivity across enrolling sites was observed (68% to 91%; p=0.03). Variation in the specificity of FDG-PET scans was also observed (15% to 44%; p=0.72); however, the small number of benign disease cases at each individual site resulted in little power to draw conclusions regarding the specificity of FDG-PET scan variation by enrolling site. Scan accuracy increased with increasing lung lesion size. Of the 80 false positive scans, 69% were granulomas. All positive FDG-PET scans were examined separately and false positive FDG-PET scans were not found to be associated with historic fungal lung disease exposure after adjusting for age and size of the lung lesion (p=0.12). The causes of the observed variation in FDG-PET scan results are unknown. Possible causes of observed variation include verification bias which results in higher sensitivity

and lower specificity, practice variation, infectious lung diseases which generate granulomas that mimic cancer, or other unknown causes.

A systematic review of the literature was conducted to examine whether other researchers had found similar poor performance of FDG-PET scans in the diagnosis of lung cancer and to update a meta-analysis conducted in 2001. A systematic review of the literature found 60 studies reporting FDG-PET scan accuracy to diagnose lung cancer. The studies showed that FDG-PET scan sensitivity to diagnose lung cancer declines as lesion size decreases and the specificity increases slightly. The advent of fusion PET and CT scanners slightly increased sensitivity (88.4, 95% CI: 84.2-91.7 to 90.0, 95% CI: 86.5-92.6) and significantly increased the specificity (69.4, 95% CI: 63.0-75.2 to 77.9, 95% CI: 70.7-83.8) for diagnosis in 29 studies when compared to 24 studies that reported using PET only scanners. In seven studies reporting endemic infectious lung disease, the sensitivity of FDG-PET was higher (91%, 95% CI: 90%-94%) to the 53 studies (90%, 95% CI: 82%-93%) that did not report infectious lung disease in the underlying study population. However, the specificity was lower in areas of endemic infectious lung diseases that cause granulomas (59%, 95% CI: 46%-70%) when compared to those studies that did not report underlying infectious disease (77%, 95%CI: 69%-86%)

No direct association was observed between fungal lung disease exposures and higher benign disease prevalence or false positive FDG-PET scans. No direct measurement of fungal lung disease exposure was possible in this analysis. The prevalence of benign disease after lung surgery varied by state and the cause of this observation, whether practice variation, verification bias, infectious lung disease or other unknown causes, is not known at this time. Investigating possible causes of the observed variation in FDG-PET scan accuracy and in benign disease prevalence after lung surgery is needed in future research. Such information will inform policy makers and health researchers as they examine the efficacy of a national lung cancer screening

program. Furthermore, the influence of local practice and locally endemic benign lung disease on the diagnosis of lung cancer must be better understood and incorporated into any lung cancer predictive models for surgeons before such a model can be clinically implemented nationally.

Table of Contents

Abstract

Table of Contents i

List of Figures iii

List of Tables v

Acknowledgements vii

1.0 Evaluation and diagnosis of lung cancer and its epidemiology with a focus on variables for predictive modeling 1

1.1 Epidemiology of lung cancer 1

1.2 The National Lung Screening Trial and the possibility of screening for lung cancer 3

1.3 Conceptual framework for the surgical evaluation of lung nodules and the diagnosis of lung cancer 4

1.4 Epidemiologic predictors of lung cancer 12

1.5 Summary of epidemiologic risk factors for lung cancer 22

1.6 Radiographic imaging and predicting lung cancer risk 23

1.7 Infectious lung diseases as a spectrum bias of FDG-PET for diagnosis of lung cancer 26

1.8 Geographic variation in the diagnosis of lung cancer 29

1.9 Conclusion 31

1.10 References 33

2.0 Development and validation of a clinical prediction model to estimate lung cancer risk for those being evaluated for surgery 59

2.1 Introduction 59

2.2 Methods 61

	2.3	Results	65
	2.4	Discussion	67
	2.5	Conclusion	74
	2.6	References	75
3.0	Geographic variation in the diagnosis of lung cancer with a focus on FDG-PET scans in the ACOSOG Z 4031 trial		87
	3.1	Background	87
	3.2	Introduction: Geographic variation in benign disease prevalence after surgical lung resection in the US.	89
	3.3	Methods	91
	3.4	Results	94
	3.5	Discussion	95
	3.6	Geographic variation in benign disease prevalence after surgical lung resection and implications for lung cancer screening	100
	3.7	Methods	103
	3.8	Results	106
	3.9	Discussion	107
	3.10	References	113
4.0	Variation in FDG-PET Accuracy to Diagnose Lung Cancer: a meta-analysis		125
	4.1	Introduction	125
	4.2	Methods	127
	4.3	Results	131
	4.4	Discussion	136
	4.5	Conclusion	143

4.6 References 144

5.0 Conclusions and future directions of research 173

5.1 References 181

Appendix 1 182

Appendix 2 193

Appendix 3 195

List of Figures

Chapter One

Figure 1	Annual Incidence and Mortality for the Most Common Cancers: 2012	52
Figure 2	Age Adjusted U.S. Lung and Bronchus Cancer Incidence and Mortality Rates by Sex, 1975-2009	52
Figure 3	Results of the National Lung Screening Trial. Detection of lung cancers by study arm and cumulative lung cancer related death by study arm	53
Figure 4	Examples of Lung Nodules, Masses and Opacities on Radiographic Imaging	54
Figure 5	Conceptual Framework for the Diagnosis of Lung Lesion	55
Figure 6	Diagnostic Guidelines for Lung Nodules Suspicious for Lung Cancer.	56
Figure 7	Mechanistic Framework for Understanding How Cigarette Smoking Causes Lung Cancer	56
Figure 8	FDG-PET Scan of a Granuloma	57
Figure 9A	Fungal Lung Disease Prevalence, Histoplasmin Skin Test Cross Reactive with Coccidioidomycosis and Blastomycosis	58
Figure 9B	Geographic Distribution of Histoplasmosis in Persons >65 Years of Age, United States, 1999-2008	58

Chapter Two

Figure 1	Hosmer-Lemeshow test showing poor ($p<0.001$) calibration of SPN and Mayo models in a surgical population	81
Figure 2	Comparison of three relationships between pack years and lung cancer.	81
Figure 3	Comparison of AUC for two Mayo models and TREAT model	82
Figure 4	Comparison of AUC for a model composed of only imaging data (ImageModel) and epidemiological data (EpiModel)	82

Figure 5	Bland – Altman plot of log(odds) predicted risk for the TREAT and Mayo model.	83
Figure 6A	Bland – Altman plot of the difference in the log odds of the TREAT and Mayo models across participants age.	83
Figure 6B	Scatter plot of the difference in the log odds of the TREAT and Mayo models across by smoking history.	84
Figure 6C	Box plot of the difference in the log odds of the TREAT and Mayo models across pack-years.	84
Figure 6D	Box plot of the difference in log odds of the TREAT and Mayo models across the three growth categories.	85
Figure 6E	Box plot of the difference in the log odds of the TREAT and Mayo models across based upon FDG Avidity.	85
Figure 7	Plot of relative contribution to explaining variance in dependent variable	86

Chapter Three

Figure 1	Fungal Lung Disease Prevalence, Histoplasmin Skin Test Cross Reactive with Coccidioidomycosis and Blastomycosis	122
Figure 2	CT scan of a spiculated granuloma in a patient presenting with hemoptysis	122
Figure 3	Consort diagram of benign disease after surgery for known or suspected lung cancer, 2009 MEDPAR	123
Figure 4	Benign disease point prevalence from 2009 MEDPAR data by state	123
Figure 5	ACOSOG Z-4031 Trial: Enrolling site location with size of circle corresponding to participation volume – 51 sites in 39 cities	124
Figure 6	Accuracy of FDG-PET in diagnosing lung cancer by lesion size in millimeters	124

Chapter Four

Figure 1	Consort diagram of systematic review of eligible studies	159

Figure 2	Forest plot of individual study estimates using simple pooled sensitivity and specificity and study heterogeneity, (I^2**)**	160
Figure 3	Summary ROC curve from the unadjusted random effects model and estimated 95% prediction contour	161
Figure 4	Deeks' Funnel Plot and Asymmetry Test for publication bias	162
Figure 5	Sub-group analysis of 7 studies reporting endemic infectious lung disease, Forest plot	163
Figure 6	SROC with random effects model for 5 studies reporting endemic infectious lung disease	164
Figure 7	Sub-group analysis of studies with mean or median lesion size less than or equal to 20mm in diameter, Forest plot	165
Figure 8	Sub-group analysis of studies reporting use of PET only scanners, Forest plot	166
Figure 9	Sub-group analysis of studies reporting blinding of readers to patient history, Forest plot	167
Figure 10	QUADAS quality metrics reported for each study by 2 reviewers	168
Figure 11	Likelihood ratio graph for all studies	169
Figure 12	Likelihood ratio graph for seven studies reporting endemic infectious lung disease	170
Figure 13	Likelihood ratio graph for seven studies reporting mean or median lesion size less than or equal to 20 mm	171

List of Tables

Chapter One

Table 1	National Comprehensive Cancer Network Risk Criteria for Lung Cancer Screening	50
Table 2	Comparison of existing predictive lung cancer models	51

Chapter Two

Table 1	Variables used in published, validated clinical lung cancer prediction models	78
Table 2	Univariate analysis of demographics and radiological data	79
Table 3	Multivariate logistic regression of lung cancer prediction model	80

Chapter Three

Table 1	Characteristics of MEDPAR inpatient discharges with lung surgery, United States 2009	119
Table 2	Descriptive Characteristics of ACOSOG Z4031 patients with FDG-PET Scans	119
Table 3	Accuracy of FDG-PET in diagnosing cancer among patients with clinical stage 1 NSCLC	120
Table 4	Pathology of false negative and false positive lesions	121
Table 5	FDG-PET sensitivity and specificity by enrolling city with at least 25 participants	121

Chapter Four

Table 1	Participant and Study Characteristics for Diagnosis of Pulmonary Nodules	152
Table 2	Meta-analysis bivariate random effects model	158

ACKNOWLEDGEMENTS

I want to thank Dr. Eric Grogan, a great mentor and friend, who introduced me to how surgeons think and made this PhD a possibility when he hired me to join his team six years ago. I also want to thank Dr. Melinda Aldrich for introducing me to the day-to-day discipline of epidemiology and the policy aspects of our work that have always been of great interest to me, and who, through her example, modeled how epidemiologists think and tackle problems in ways no class could communicate. I want to thank the other members of my committee: Dr. Pierre Massion, for his passion for research, smile, keen wit and scientific eye; Dr. Jeffery Blume, who unfailingly found my statistical gaffes and required learning along with clear analysis; and Dr. Melissa McPheeters, whose patience for my grammatical assaults is legendary and who helped me keep my eyes on the prize throughout the process.

I want to thank the TREAT lung cancer group for their hard work in making some of this research possible. They helped by holding my hand as I learned R, and with their insightful questions and critiques they make my research better in every way. I also want to thank the Department of Thoracic Surgery who supported me financially, and the surgeons Drs. Nesbitt, Lambright, and Bill Putnam. Without their clinical expertise and willingness to discuss the problems they face, not to mention the patients for whom they cared, none of this work would have occurred. Bill Putnam, who is as good of a boss as I've ever had and a great man to work for and mentor to learn from, I especially want to thank.

I want to thank my classmates Lan, Natalia and Webster, who shared the growing pains with me of being the first cohort to get an epidemiology PhD at Vanderbilt and turned a critical eye to my logic or lack thereof. I want to thank Spencer who covered my many administrative lapses and missed university deadlines. I want to thank my professors Drs. Sandy Deming, Aaron Kipp, Katherine Hartmann, Alicia Beeghly-Fadiel, Maureen Sanderson, Liana Castel, Chris Slaughter, Robert Greevy, and Chris Fonnesbeck for sharing their knowledge.

My deepest thanks go to Vonda Hunter, my wife and partner. She was a single parent for stretches of time as I pursued my dream and who knows me better than anyone else in this world. Thanks to my three kiddoes Hop, Ben and Jin Lan, whose unconditional love and beauty renew my spirit.

Finally, I want to thank my parents, John and Lorraine Deppen. In 1980, they both died from lung cancer. Their memory helped me work through the hard times and revel in the good. I hope that they can be proud of what I have become. This work is dedicated to them and their memory.

Chapter 1

1.0 An introduction to lung cancer, its epidemiology, evaluation and diagnosis with a focus on variables for predictive modeling

1.1 Epidemiology of lung cancer

Lung cancer is the most common cause of cancer-related death in the United States (US) and kills more people than breast, prostate, and colorectal cancers combined (Figure 1). An estimated 226,160 new cases and 160,340 deaths from lung cancer occurred in the US in 2012, accounting for approximately 17% of annual incident cancers and 28% of all cancer deaths.[1,2] One in fourteen individuals born today will be diagnosed with lung cancer at some time during their lifetime based upon incidence rates from 2007-2009. Lung cancer diagnosis and treatment exerts a significant economic burden as well. The US spends an estimated $10.3 billion on direct lung cancer care, which comprises 10% of all cancer related healthcare expenditures.[3] These direct costs are dwarfed by the estimated $83 billion lost in productivity from caregiving to those with the disease and lost wages from early mortality.[4]

Lung cancer, like most cancers, is a complex and diverse disease. It is caused primarily by environmental factors. Known risk factors for lung cancer include smoking tobacco, radon exposure, air pollutants, second-hand smoke, occupational exposures and individual genetic susceptibility. The primary risk factor for both small cell (SCLC), which comprises 14% of all lung cancers, and non-small cell lung cancer (NSCLC), which comprises the remaining 86%, is tobacco smoking. Approximately 85 to 90% of the population attributable risk for lung cancer is

due to tobacco smoking.[5] Among non-smokers lung cancer is a relatively rare disease, less than 15 per 100,000 person-years.[6]

Lung cancer incidence has declined as smoking prevalence has declined in the US. The age-adjusted incidence rates for lung cancer in the US, according to Surveillance Epidemiology and End Results (SEER) data, peaked in 1992 at 69.4 cases per 100,000 person-years. The trend in incidence has fallen with 58.8 per 100,000 person-years in 2009, the most recently available data.[7] Among women, age-adjusted incidence has decreased to 51.2 per 100,000 since reaching its peak of 53.5 in 2005 (Figure 2). The incidence rate of lung cancer in men has steadily declined from its peak of 102.1 per 100,000 person-years in 1984 to 69.2 in 2009. Age-adjusted mortality rates in the US have seen slight declines in recent years for women (38.6 per 100,000) from their peak in 2002 of 41.6 per 100,000 person-years. A more dramatic decline in mortality among men has been observed from an age-adjusted high in 1990 of 90.6 to 62 per 100,000 person-years in 2009.[8]

The overall 5-year survival rate for lung cancer is 15.2%. This poor prognosis is largely due to the cancer's typically advanced stage at the time of diagnosis.[9] In comparison, using the more granular American Joint Committee on Cancer staging schema, populations with localized, pathological stage 1A lung cancer have a 73% 5-year survival rate, while metastatic stage IV disease has a 1.8% survival rate.[10,11] To date, only surgical removal of cancerous tissue can cure lung cancer. Surgery for lung cancer is only indicated in those who can withstand the physical rigors of major surgery and have localized stage 1 or stage 2 disease.[12] To reduce the mortality burden from lung cancer, detection of disease in its earlier stages when surgical treatment is a possibility is necessary. Early diagnosis and treatment that reduces lung cancer related mortality has been the subject of much research in the lung cancer field; however, none of the methods for

early detection had reduced mortality from lung cancer prior to the advent of low dose computed tomography (LDCT) scans.[13-17]

1.2 The National Lung Screening Trial and the possibility of screening for lung cancer

The recently concluded National Lung Screening Trial (NLST) was a two-arm randomized controlled trial that screened 53,454 asymptomatic individuals at high risk for lung cancer with either LDCT or chest x-ray. A 20% decrease in lung cancer related mortality in the LDCT arm was observed compared to those randomized to the chest x-ray arm (Figure 3). All-cause mortality was 7% less in the LDCT arm as well. However, the false positive rate for LDCT was 96.4% and 39% of all participants had at least one positive scan between the three annual scans conducted during the trial. The success of LDCT in reducing mortality compared to chest radiograph in the NLST has caused physician and patient advocate groups to propose screening guidelines for lung cancer. The American College of Chest Physicians (ACCP), American Society of Clinical Oncology, National Comprehensive Cancer Network (NCCN), American Association of Thoracic Surgery, American Cancer Society, American Lung Association and American Thoracic Society all endorse using the NLST screening regimen as the basis for lung cancer screening.[18] In addition, the NCCN expanded the pool of those to potentially be screened from ages 55 through 74 to 50 through 74, with more than a 20 pack-year history of smoking or who also have any secondary risk factors including: family history of lung cancer among first degree relatives, previous cancer, asbestos or radon exposure, chronic obstructive pulmonary disease, or occupational exposure to a known lung carcinogen (Table 1).[19]

Cost effectiveness of LDCT screening has been estimated to be as low as $19,000[20] to over $110,000 per quality adjusted life year using the NLST based risk criteria[21] for screening inclusion and accepted screening guidelines described above. More accurate estimates of cost-

effectiveness within the NLST are expected to be published shortly.[22] The publication of results from the NLST, the promulgation of clinical guidelines, and endorsement of LDCT lung cancer screening by clinical and patient advocacy groups led the US Preventive Services Task Force to initiate a review of their 2007 recommendation against screening with LDCT. Public announcement of the Task Force's draft review of their recommendation is expected in of 2013.[23] Should the Task Force find strong evidence of benefit from lung cancer screening with LDCT, then Medicare and other insurers would be required to cover the expense of screening with LDCT. Thus, screening for lung cancer would join that of breast, colon, cervical and prostate cancer.

Using the proposed screening guidelines based upon the NLST definition of a high risk individual, an estimated 7.4 million current and former smokers would be eligible for lung screening.[24,25] In the National Lung Screening Trial (NLST), a lung abnormality was identified in 39% of patients during the screening protocol requiring additional diagnostic testing, but less than 4% of those were malignant.[26] If the false positive rate for national screening is similar to that found during the trial, an estimated 2.9 million abnormalities will be found on LDCT over the first three years of the screening program, each requiring additional surveillance and diagnostic testing.

1.3 Conceptual framework for the surgical evaluation of lung nodules and the diagnosis of lung cancer

An individual can present to a clinician for evaluation and diagnosis of a lung nodule suspicious for lung cancer three ways: symptomatically, incidental discovery of the nodule after imaging for another clinical indication, and from periodic screening. Irrespective of source, the diagnosis of lung cancer begins with radiographic imaging of the chest and detailed history and physical of the

individual. Previously, most lung cancer was diagnosed symptomatically (estimated 75%), but incidental discovery has increased recently with the proliferation of imaging modalities like CT scans.[27] If a screening program is initiated in the US, then an explosion in asymptomatic lung nodules requiring diagnosis will occur. Since the lung is difficult to access outside the immediate proximity to an airway, clinicians rely heavily on available non-invasive diagnostic tools, like radiographic imaging, to diagnose lung cancer.

Lung anomalies on radiograph are generally characterized as either lung nodules, opacities, lesions or masses. Abnormalities or lesions of the lung are generally classified according to size at maximum average diameter and morphological characteristics that are observable on radiograph. Lung nodules are larger than 3mm in maximum diameter and less than 30mm (Figure 4A). Lung masses are 30mm and larger in size. Nodules and masses appear as generally solid collections of tissue on CT scan (Figure 4B). Ground glass opacities and ground glass nodules can vary widely in size and are characterized by their non-solid appearance or shadow consistency on imaging (Figure 4C).[28] For the purpose of this dissertation, "nodule" and "lesion" are synonymous.

Figure 5 illustrates the three-step progression of a lung lesion in an individual from the time at which the lesion is "Undetectable" (step 1) to "Detectable" (step 2) and finally to the surgical evaluation of a "Suspicious" lesion (step 3). This conceptual framework begins with the "at risk" patient prior to developing a detectable suspicious lung lesion and moves through time until a definitive diagnosis of cancer or benign disease occurs at the time of an operation or through prolonged radiographic surveillance. At each step of their evaluation, patients accumulate additional clinical and imaging diagnostic information until presenting to the surgeon who must decide to either operate or continue following the lesion radiographically over time.

In the first step, the patient has an "undetectable lung lesion". Every person has genetic, demographic, environmental and behavioral risk factors for the development of lung cancer. While smoking and advanced age are the primary risk factors for developing lung cancer, clinical diagnoses of other lung diseases like chronic obstructive pulmonary disease (COPD), asbestosis, etc. may also increase or decrease the risk of lung cancer. Most individuals are asymptomatic at this point in the growth and detectability of a lung lesion. Models developed to estimate the likelihood of lung cancer in this population of undetectable lung lesions focus on common epidemiologic risk factors including age, sex, race, years smoked, number of cigarettes smoked per day, family history of cancer, occupational exposures to carcinogens, age at smoking initiation, or years since smoking cessation.[29-33] This first group of epidemiologic models uses population characteristics to estimate the likelihood of developing cancer over a one-year, five-year, or ten-year time frame. These population risk models categorize individuals into cancer risk strata, and those strata at highest risk are then likely candidates for LDCT screening or greater surveillance. The most widely used in the US is the Bach model.[29,34] Other models including those by Spitz, Tammemagi, and Cassidy have been validated.[30,35-38] One model by Hoggart and colleagues, though promising, has not been validated but performed similarly to the Bach model.[39] Model accuracy as measured by C-statistic or area under the curve (AUC) is 65% to 80% among the four models and 60% to 75% in external validation populations. The prevalence of disease in the populations from which these models were developed range from 0.01% to 1% (Table 2).[33,40] In high risk cohorts like that included in the NLST, lung cancer incidence was 6.45 per 1000 person years.[22]

Once a lung lesion is present, the individual can be clinically evaluated, the nodule discovered and categorized, and likelihood of cancer assessed. Discovery and evaluation of a lung lesion is encapsulated in Step 2 of the conceptual model. In this step, the lesion is radiographically detectable by CT scan and clinical symptoms may or may not be present. LDCT scans are fairly

sensitive (90-95%) in the detection of lung lesions down to 3 to 5mm, and high dose, thin slice CT scans can detect lung anomalies smaller than 3mm.[41] Common symptoms of possible lung cancer include unexplained weight loss, shortness of breath and hemoptysis. Discernible imaging characteristics of the lesion are now observable on CT scan and indicate the relative likelihood the lesion is cancerous. The prevalence of lung cancer in this population is much higher than that prior to lesion discovery. Four percent of lung anomalies discovered on LDCT scan were cancerous in the NLST and a higher prevalence of disease (25-55%) was observed in the training datasets for the Mayo, Solitary Pulmonary Nodule (SPN) and Veterans Affairs models developed to diagnose lung cancer in indeterminate lung nodules.[35,42,43] Clinical guidelines suggest an individual be evaluated by a primary care physician or pulmonologist who makes a clinical assessment of cancer risk.

The correct diagnosis of a nodule suspicious for lung cancer is as much art as science. The clinician must individualize care by weighing the risks of increasingly invasive procedures and the patient's physiological status with the need to treat in a timely fashion. The diagnosis of lung cancer, as described hereafter, relies upon NCCN and ACCP as well as Fleischner Society guidelines, which both the NCCN and ACCP incorporate into their clinical guidelines.[19,27,41,44-46] The assessment of a lung lesions cancer risk should be based upon either clinical judgment or a validated prediction model at the time of evaluation (Figure 6).[27,45]

The pre-operative diagnosis of lung cancer combines two classifications of clinical information. The first is primarily epidemiologic in nature and includes risk factors such as: age, smoking history, family history of cancer, occupational exposures to carcinogenic agents, environmental exposures and previous diagnostic history of other, non-cancerous lung diseases. These risk factors describe known population risks for developing lung cancer and are commonly obtained during initial clinical evaluation. Lesion or nodule specific information on lung cancer risk is

derived from imaging and various invasive procedures to obtain a tissue diagnosis. Data used to diagnose an individual can be placed on a continuum of increasingly invasive diagnostic tests. An examination starts with a history and physical, then progresses to an LDCT scan or chest x-ray, sputum cytology, a high dose CT scan with injected contrast or Positron Emission Tomography scan with injected radiotracer F^{18}-fluorodeoxyglucose (FDG-PET). Each of these tests generates non-invasive diagnostic information. CT scans with contrast and FDG-PET scans are currently the most sensitive and specific non-invasive tests available. According to a meta-analysis, combination PET/CT scans have the highest overall accuracy with sensitivity of 92-96% and specificity of 78-84%.[47,48] This combination of imaging modalities offers anatomic, morphologic and metabolic information on the lung lesion.[45] Sputum cytology is a non-invasive test that can provide a pathological diagnosis, but has poor sensitivity and specificity.[45] Increasingly invasive tests to obtain a tissue diagnosis include bronchoscopy, endobronchial ultrasound with biopsy, fine needle aspiration, thoracoscopic biopsy and finally thoracotomy.

According to national guidelines, if the likelihood of cancer is less than 5% then continued surveillance with follow-up LDCT scans is recommended. When the likelihood of cancer is greater than 60% then surgical biopsy including bronchoscopy with biopsy, CT guided biopsy, mediastinoscopy or surgical resection is recommended. When the likelihood of malignancy is between 5% and 60%, then diagnostic CT scan with imaging contrast, FDG-PET scan or other non-invasive test is recommended (figure 6). Nodules greater than 8mm in maximum diameter are candidates for evaluation by injected radiotracer F^{18}-fluorodeoxyglucose (FDG) followed by combined positron emission tomographic and computed tomographic scan (FDG-PET). A negative FDG-PET scan indicates a low likelihood of cancer. A scan is indicative of possible cancer when hypermetabolism, more commonly called avidity is observed. If the lesion is determined to be avid by FDG-PET scan, a biopsy or surgical excision should then be pursued to obtain pathological tissue diagnosis. If not avid, then continued surveillance by annual LDCT is

suggested. For any nodule that is stable or decreases in size on subsequent LDCT, continued surveillance with decreasingly frequent LDCT is warranted for an additional two years. Continued annual LDCT should be considered for those patients who are eligible for lung cancer screening until the individual is no longer eligible for definitive treatment of a lung cancer that could occur.[19,41]

Three validated models are available to estimate the likelihood of lung cancer after detection of a lung lesion on radiographic imaging.[35,36,42,43] Each model includes age, history of previous cancer, smoking history of the individual, the size of the lesion and one or more imaging characteristics from CT scans. Imaging characteristics associated with higher risk of cancer include a growth on serial CT scans; having a spiculated, diffuse coronal, spiky or pointed edge on the surface of the lesion compared to a lobulated or smooth lesion edge; and a lesion location in the upper lung. The SPN model developed by Gurney and colleagues also includes hemoptysis, lesion cavitation which indicates likely benign disease when cavitation is smaller than 16mm, calcification which also indicates benign disease, and FDG-PET avidity.[49] The prevalence of lung cancer observed in the populations from which these models were developed ranged from 25% to 55%. The AUCs for these models were between 70% and 80%. Table 2 lists the models from Steps one and two of the conceptual model described in Figure 5 as well as the variables used in each model for comparison purposes.

After a lesion has been detected, the clinician must estimate the likelihood of cancer using a predictive model or their clinical judgment. The evaluating clinician may also consider whether the likelihood for cancer is great enough to warrant referral to a surgeon for additional assessment and possible treatment. If so, the surgeon compiles all relevant diagnostic data acquired to date and may order additional testing to better estimate both the probability of cancer as well as whether the patient is a low enough operative risk to benefit from surgery. Additional testing may

include more invasive tests like endobronchial ultrasound, fine needle aspiration or navigational bronchoscopy. Assessment of operative risk includes pulmonary function tests and cardiac stress or diffusing capacity for carbon monoxide tests. At this point in the evaluation process, surgeons have access to all available data including patient risk factors, clinical and imaging characteristics, and the results of any additional procedural or operative evaluation tests. A final assessment of cancer risk occurs and a decision to operate or to continue surveillance through periodic radiographs is made. The surgeon weighs the likelihood of cancer with the possible benefits to the patient and possible harms from the procedure and plans a course of care which is personalized to the individual after conversation with the patient.[46]

Epidemiologic risk models, such as those developed by Cassidy, Spitz, Tammemagi and Bach, assess a patient's risk for developing cancer prior to imaging at step one in our conceptual model for evaluating lung lesions. These models assist in determining who benefits most from lung cancer screening by LDCT scan.[29,31,33,40] Swensen's Mayo model, Gould's VA model and Gurney's SPN calculator estimate the probability of cancer in patients with a known radiographic abnormality on CT scan or chest x-ray.[42,43] These diagnostic models estimate the probability of malignancy to help the clinician decide in step two who needs watchful waiting, additional testing, or surgical referral. No published models are available for the surgeon evaluating individuals for possible surgical biopsy. Isbell and colleagues showed that Swensen's Mayo model and Gurney's SPN calculator model had similar accuracy to predict lung cancer as reported in the original studies (AUC of 78 and 80 respectively) but were poorly calibrated in the validation dataset from a single institution's thoracic surgery practice. The prevalence of cancer in most surgical populations is between 50% and 80%. The three models available to evaluate lung nodules were developed or validated in populations with 25% to 55% cancer prevalence. Thus, all lung nodule risk models were developed in non-surgical populations with a lower prevalence of cancer compared to surgical populations. Currently available models to assess the

risk a lung nodule is cancerous tend to underestimate the likelihood of cancer in individuals with cancer and overestimate the probability of cancer in those with benign disease when applied to surgical populations.[50]

Unlike the colon, breast or prostate, the lung is a difficult organ to access for some biopsies, and many individuals seeking diagnosis have poor lung function from their prolonged tobacco smoke exposure. Less invasive biopsies, like fine needle aspiration and transthoracic needle lung biopsy, are frequently complicated by pneumothorax or atrial fibrillation (15-25%) or by less frequent and more severe complications including internal bleeding or lung failure requiring mechanical ventilation.[51-54] These biopsies are often not recommended for lesions occurring more centrally within the lung or near major vessels. Minimally invasive lung resection for biopsy of a lung nodule has a similar profile of complications but also has significant risk for mortality associated with it. In the NLST, which included primarily large academic cancer centers and procedures performed by thoracic specialists, surgeons experienced a 1.2% mortality rate within 60 days of their lung operation. Other researchers have found the mortality rate for thoracoscopic surgery to be even higher, between 2 and 5% within 30 days of the operation, depending upon the population and specialization of the surgeon.[55,56] In comparison, breast lumpectomy with lymph node sampling has a 0% 30-day postoperative mortality rate.[57] Thus, surgical lung biopsy has significant risk associated with the operation. The surgeon evaluating a lung nodule faces a significant dilemma: should he/she subject the patient to an operation with the possible range of complications associated with this major surgery, or miss catching the lung cancer and the only opportunity for a cure available if the cancer is early stage local disease.

Prior to surgical evaluation, all prediction models, radiographic studies, and clinical evaluations are designed to rule in cancer and maximize sensitivity. Consequently, 20% to 40% of diagnostic lung operations result in a benign diagnosis.[58-62] Thoracoscopic surgery in the NLST had a 24%

benign disease result prevalence at final pathology.[22] Clearly, a model developed in a surgical population, with its larger available information set and higher prevalence of disease, is warranted to aid surgeons in the evaluation of lung nodules and avoid unnecessary surgeries. Possible variables representing known risk factors or predictive variables will be discussed further below.

1.4 Epidemiologic predictors of lung cancer

Predictors of lung cancer in populations being evaluated for possible surgery can be categorized into two broad groups, epidemiologic and radiographic. Epidemiologic predictors, as discussed earlier, include age; sex; race; number of years of smoking; number of cigarettes smoked per day; family history of cancer; occupational exposures to carcinogens; age at smoking initiation; years of smoking cessation; presence of other lung diseases like emphysema or COPD; and presence of symptoms like hemoptysis, shortness of breath, unplanned weight loss, or fatigue. Factors unique to surgical populations that indicate relative tobacco exposure and the individual's response to that exposure include body mass and relative lung function. Radiographic imaging variables associated with lung cancer include the location of the lesion in the lung, rate of growth, physiological shape, and metabolic activity relative to surrounding lung tissue. Each predictor and its association with lung cancer will be examined and use in other lung cancer prediction models discussed below.

Host Factors and Lung Cancer Risk from Smoking

Smoking has been a known cause of lung cancer for nearly six decades[63,64]. Doll and Hill's seminal findings from their prospective cohort study confirmed previous retrospective studies that found tobacco smoking causes lung cancer.[63] Approximately 80% of women and 90% of men diagnosed with lung cancer were smokers. In the US smoking prevalence peaked among males in the 1940s and 50s at around 67%. Smoking prevalence peaked for US females in the late 1960s at

approximately 44%. Active tobacco smoking remains the greatest risk factor for lung cancer, although passive tobacco smoke exposure is a contributing risk factor especially among non-smokers. Tobacco smoke and the carcinogenic chemicals in the smoke act on the lung to generate cancer through a number of biological processes. Polycyclic aromatic hydrocarbons (PAH), carcinogenic metals, and N-nitrosamines are all present in tobacco smoke and exert their effects on the lung and carcinogenesis through both gene mutations and the formation of DNA adducts (Figure 7). The subsequent cellular changes and mutation proliferation result in lung cancer.[65,66]

There is an observed dose-response between smoking and lung cancer risk.[32,43,67] The relative risk per year of contracting lung cancer increases between 8% and 17% per 10 cigarettes per day additionally smoked.[68,69] The change in risk due to smoking is generally monotonic[70,71] and Tammemagi found a non-linear relationship between the number of cigarettes smoked per day and lung cancer risk.[33] The age at which smoking began, the number of cigarettes smoked per day, and the duration of smoking all influence the likelihood of developing lung cancer.[29,72] Current smokers have a 20-fold increase in lung cancer risk compared to never-smokers[5]. In one study, the cumulative risk of dying from lung cancer before age 85 was 22.1% for a male smoker and 11.9% for a female smoker of European descent, in the absence of competing causes of death. This compares to a 1.1% probability for a man and 0.8% probability for a woman non-smoker dying from lung cancer before age 85.[73] Smoking is often measured in pack-years[19,29,30,38,43], cigarettes per day[33], years of smoking[30,32] or as ever having smoked more than 100 cigarettes[30,35,42] when used in predictive modeling.

Smoking cessation decreases all cause and lung cancer related mortality irrespective of the age at which one stops smoking.[74,75] The benefits of quitting smoking for reducing lung cancer risk are also well documented.[76,77] The benefits of smoking cessation to lung cancer risk reduction stop at around two times the risk for former smokers when compared to never smokers.[78,79] Decreased

risk of lung cancer changes little after 15 years post-cessation, but risk remains elevated among smokers compared to non-smokers for at least 30 years after cessation.[69] Years of smoking cessation or having quit smoking as a dichotomous outcome are included in screening models.[19,29,30,33] Any model estimating lung cancer risk must include some measure of smoking exposure.

Sex Differences in Lung Cancer Incidence

Sex differences in lung cancer risk from smoking in the US have been observed in the past, and women have been considered more susceptible to developing lung cancer.[80-82] More men are diagnosed with lung cancer than women but smoking prevalence among men has historically been higher compared to women. The gap in smoking rates between men and women has greatly narrowed and in 2009 23.5% of men and 17.9% of women were smokers.[83] As the rates of smoking among both men and women have become more similar since the 1960s, their lung cancer incidences and age adjusted lung cancer death rates have converged (Figure 1).[84,85] After controlling for smoking duration, age at initiation, and intensity, no sex specific differences in lung cancer risk have been found.[29,70,85-87] However, death from lung cancer is lower in women compared to men, and this difference in cancer related death has been consistent over time.[88,89]

Separating sex specific risk for lung cancer from that associated with smoking habits has been difficult. Some researchers investigating incident lung cancers found women smokers had higher risks for lung cancer when compared to male smokers, especially after age 70.[81] Women have been found to have different distributions of lung cancer histology. Specifically, non-smoking women of all races are two to four times more likely to have bronchioloalveolar cancer compared to men.[90-92] Among non-smokers, age and temporal biases within the cohorts studied may have influenced the relationship given that women live longer than men.[93] A recent study examined sex specific risk for lung cancer by pooling non-smokers from 13 cohorts and 22 cancer registries

from around the world.[73] Thun and colleagues found no differences in lung cancer age-standardized incidence rates between men and women age 40 and higher.[73] Sex is not included as an independent variable in any lung cancer prediction models of lung nodules; however, the Bach model includes sex as a covariate and the models developed by Spitz and Cassidy matched cases and controls on sex (Table 2).[38,40]

Race

A number of population studies have found that racial differences in response to smoking and lung cancer risk may occur due to intensity, types of cigarettes smoked or other environmental factors such as access to care or socio-economic status. Incidence rates of lung cancer among smokers are higher among African American men compared to Caucasian men[67] Smoking prevalence among African American men is slightly higher than that observed among white men. However, the literature remains mixed as to whether race is an independent host factor for smoking related risk for lung cancer incidence. African Americans and Native Hawaiians appear to have higher risk of lung cancer at lower smoking exposure (<31 cigarettes a day). Lung cancer risks were similar among heavy smokers across races and sexes in one large multi-racial cohort[94], and only African American heavy smokers had higher lung cancer risks in a case-control study.[95]

One conjectured cause of the observed higher lung cancer risks among African Americans is their higher prevalence of mentholated cigarette smoking, 62% among African Americans compared to 23% among Caucasian smokers.[96] Menthol cigarettes are thought to be more difficult to quit compared to non-flavored cigarettes, thus contributing to the lung cancer burden from smoking.[96,97] A recent study examined menthol cigarette use, lung cancer and smoking cessation in the Southern Community Cohort. The authors compared incidence and rates of smoking cessation between African American smokers and Caucasian smokers. They found lower incidence of lung cancer among menthol smokers irrespective of race and similar likelihood of

quitting smoking irrespective of the type of cigarette smoked.[98] Other prospective studies also found no increased risk from mentholated cigarettes.[99] Race is included only in the Tammemagi PLCO model for screening (Table 2).[33] Years of education completed as a predictive variable was included in the Tammemagi PLCO model. However, race in combination with education may be more indicative of socio-economic status or as an indirect measurement of occupational or other environmental exposures that are not fully captured by smoking. As such, race may be an important variable in epidemiologic models of lung cancer risk, but not in models assessing the risk from a discovered lung nodule.

Second-hand Tobacco Smoke

Exposure to second-hand or environmental tobacco smoke as a cause of cancer was first reported in 1981 in a study of non-smoking Asian women married to smokers[100,101] More recent studies and meta-analyses found a dose-response relationship between lung cancer and environmental tobacco smoke exposure which strengthened the causal relationship found in earlier studies and reviews.[102-104] In nonsmokers exposed to second hand smoke had seven percent of the cotinine and other tobacco specific metabolites compared to current smokers.[105] Cotinine is a metabolite of nicotine and directly varies with the amount of nicotine inhaled. It is commonly used to estimate smoking exposure. Population exposure to second-hand smoke has been estimated to be around 40%, and an estimated one-fourth of lung cancer cases among non-smokers or 3,000 deaths per year are attributable to environmental tobacco smoke in the US.[103] Reduction of this exposure through recent workplace and public area restrictions of tobacco smoking has resulted in cotinine serum levels decreasing 2 to 3 fold compared to levels observed in the late 1980s.[106,107] Thus second hand smoke exposure, while a significant historical risk factor, will decrease in importance as a risk factor in the future with continued success of public health initiatives to reduce exposure to this carcinogen.

Second-hand smoke exposure has been included in some lung cancer risk models that estimated the likelihood of cancer among nonsmokers. The models were less accurate at estimating risk among nonsmokers when compared to former or current smokers as measured by AUC. Only Tammemagi and colleagues report evaluating second hand smoke exposure at a possible covariate.[32] The NCCN guidelines for defining populations likely to benefit from screening specifically excluded environmental tobacco smoke exposure from the high risk group who should seek screening with LDCT. Second-hand smoking exposure is not a predictive variable in any models estimating lung cancer risk given a lung lesion found on radiograph (Table 2). No lung cancer prediction models that combine smoking and nonsmoking populations include second-hand smoking exposure.

Non-smoking related exposures and risk factors for lung cancer

Other environmental exposures associated with increased lung cancer risk include: naturally occurring radon (10-15% population attributable risk), asbestos, metal dusts and other occupational exposures (9-15%), and outdoor air pollutants (1-2%).[5,107] Collectively, these environmental and occupational exposures represent much of the attributable risk in the 10% to 15% of all lung cancers that occur among nonsmokers.[107] Residential and occupational radon exposure is the second leading cause of lung cancer behind cigarettes.[108] Asbestos, arsenic, chromium, nickel, beryllium and silica have been shown to increase risk for developing lung cancer among exposed, non-smoking workers. Other chemicals, including organic solvents, pesticides and non-ferrous metal dust or fumes have not consistently been shown to cause cancer in non-smokers.[109-111] A number of studies have shown increased risk from radon and occupational carcinogen exposure concomitant with tobacco smoking.[112] Wide variances in estimated occupational exposures are not uncommon in the literature, making risk estimation for individual chemicals problematic. Occupational exposure remains a significant burden for lung cancer in the US and across much of the developed world.[111,113-115]

Radon, asbestos, and occupational exposures are included in NCCN guidelines for defining possible high risk individuals eligible for screening. Outdoor air pollution has been reported as increasing lung cancer risk; however, the biases from residual confounding from occupational and smoking exposures does not allow for accurate estimates of risk.[107,115] Radon and occupational exposure are significant risk factors for lung cancer at the population level, especially among nonsmokers. Asbestos exposure is a risk factor in the Bach, Spitz and Liverpool models. Occupationally occurring dust exposure is a risk factor in the Spitz model. No environmental risk factors are included in a lung cancer model that estimates risk after a lung nodule has been discovered (Table 2).

Age and Lung Cancer

Lung cancer is generally a disease of the elderly. Fewer than 1 in 100,000 individuals under age 30 develop lung cancer.[9] Lung cancer risk increases with age for both smokers and non-smokers. Lung cancer risk in non-smokers is 23 per 100,000 person years in individuals between the ages of 60 and 80. The increased risk appears to be similar among never-smokers irrespective of race, except for Hispanics and African-American men about which there is too little data.[73] Nearly two-thirds of incident lung cancers occur in those over age 65. The median age at initial diagnosis is 71 years.[116] Age is included in all published lung cancer risk models. The NCCN lung cancer screening recommendations consider those over age 55 years or over 50 years with additional risk factors as being in the high risk category (Table 2).

Body Mass Index as a Risk Factor for Lung Cancer

The evidence for body mass index (BMI) as a risk factor for lung cancer independent from smoking is generally weak and conflicting. A small case control study in non-smokers found those with a BMI over 30 had twice the odds of developing cancer when compared to those with

a BMI below 21.[117] In a more recent study of the Women's Health Initiative, BMI was found to be inversely related to lung cancer risk in smokers. In populations of non-smokers no association was found between BMI and lung cancer after adjusting for other lung cancer risk factors. BMI has been found to be inversely related to cancer incidence in populations with a history of tobacco use in a number of studies.[98,118,119] BMI, independent of smoking, has not been found to be associated with lung cancer in a prospective cohort to date. Studies examining the relationship between BMI and lung cancer may suffer from residual confounding from smoking or from some other unknown biological mechanism occurring or associated with higher BMI as protective for lung cancer. Additional studies are needed to determine whether BMI is a lung cancer risk factor independent of smoking. As a possible predictive variable for lung cancer modeling, BMI is included only in Tammemagi's PLCO model for screening (Table 2).[33] BMI, like chronic obstructive pulmonary disease, may indicate the individual's response to prolonged smoking exposure. It should be investigated as a predictor of lung cancer when evaluating lung nodules suspicious for lung cancer.

Non-cancer lung disease

Recent studies have shown a strong association between lung disease and lung cancer. An estimated 40 to 70% of those with lung cancer have co-morbid COPD.[120] COPD has been found to be an independent risk factor for lung cancer in a number of studies.[120-122] Individuals with COPD have a twofold higher risk for lung cancer and are more likely to have squamous cell histology if diagnosed with lung cancer. Other lung diseases not directly caused by smoking have mixed results in their association with lung cancer risk. Asthma, for example, was associated with decreased risk of lung cancer in males in one study.[122] On the other hand, pulmonary fibrosis and interstitial lung disease have been shown to exhibit an inflammatory response in airway epithelium leading to dysplasia. Subsequently, these two diseases have been found to be possible risk factors for lung cancer.[123,124] Non-cancer lung diseases like COPD or pulmonary fibrosis are

an inclusion criteria for proposed NCCN lung cancer screening guidelines and may be an important risk factor for early detection.[125] Tammemagi's PLCO model, Spitz's model, Cassidy's Liverpool lung cancer screening model and the NCCN guidelines all include either COPD, emphysema, or COPD with other non-cancerous lung diseases in their risk estimation.

Genetic Factors in Lung Cancer Risk

An estimated 11% of smokers eventually develop lung cancer.[126] The lack of lung cancer in the majority of those who smoke leads to the hypothesis that significant heritable risk factors exist and influence the tumorigenesis from tobacco exposure to lung cancer[105] Studies showing an association between family history of cancer and increased lung cancer risk support this hypothesis.[5,127] Aggregation of lung cancer within families has been shown in both case-control and cohort studies. One study examining the genetic evidence among twins and a second study in non-smokers and their first-degree relatives found higher risks for lung cancer within families. Furthermore, associations were stronger in lung cancers occurring at a younger ages.[128,129] Genetic mutations are grouped into two distinct categories: germline and somatic. Germline mutations are heritable and occur in a body's reproductive cells. These mutations are incorporated into the DNA of every cell in the body of the offspring. For example, a germline mutation to the p53 region may make one more susceptible to the carcinogens in tobacco smoke or increase cancer risk independent of tobacco smoke exposure. Somatic mutations arise after conception. These mutations are not heritable and can arise from natural copy errors that are maintained through the DNA replication process or from environmental exposures that alter DNA. Biological pathways that regulate cells and mechanistically detoxify tobacco smoke can each be affected by either type of mutation (Figure 7).

Cytochrome-p450 metabolizes the non-reactive compounds in tobacco smoke, primarily polycyclic aromatic hydrocarbons for example, into highly reactive intermediates. These reactive

compounds may bind to DNA and cause genetic injury (Figure 7). Glutathione s-transferase then transforms the intermediates created primarily by cytochrome-p450 into complexes that are more easily excreted. Germline mutations to genes which regulate either pathway may influence lung cancer risk.[130,131] Similarly, mutations that influence enzymes from these pathways have been shown to increase lung cancer risk.[132,133]

A second area of genetic interest is the tumor suppressor gene, p53. This gene influences multiple cellular response pathways relevant to carcinogenesis of various cancers and is not unique to lung cancer. The types of mutations in this gene in populations with lung cancer appear to be different between smokers and non-smokers. As such, the mutations in the p53 gene are likely somatic rather than germline. A third pathway for lung cancer susceptibility involves DNA repair and DNA repair capacity. Both germline and somatic mutations occur in the capacity of DNA repair enzymes to overcome miscoding or promote apoptosis. For example, DNA methylation, both hyper-methylation and hypo-methylation depending upon gene regions, increases risk.[134] Comprehensive review of genetic variations in the cause of lung cancer is beyond the scope of this chapter and other sources better serve this purpose.[135-138]

Currently, no serum biomarker or other genetic test is recommended to assess risk for screening or diagnostic purposes.[44] However, family history of lung cancer is a common risk factor among the risk models for screening populations. The Cassidy Liverpool model breaks family history of lung cancer into three categories: no history, early onset (<60 years), and late onset (≥60 years). Spitz defined family history as two or more relatives with history of any cancer. Tammemagi's PLCO model defined family history of lung cancer as a dichotomous (yes or no) variable of lung cancer risk.

1.5 Summary of epidemiological risk factors for lung cancer

When assessing the variety of risk factors used to predict lung cancer, only age and some measurement of smoking exposure are common across all lung cancer risk models. These two variables are included in both screening models and in models assessing risk after the discovery of a lung lesion. Age is modeled linearly or as a categorical variable in the instance of NCCN guidelines or in Gurney's Bayesian SPN model. Smoking exposure definition varies by model. The screening models, which rely solely on epidemiologic information, attempt to measure both the dose and duration of exposure by including cigarettes per day as well as years of smoking or the combination of the two as measured by pack-years. Length of smoking cessation is also considered in the Bach, Spitz and PLCO screening models as well as in the VA model for assessing lung lesions.

Sex, COPD or emphysema, occupational exposures to dust or asbestos, and family history of cancer are each common predictive variables across three of the four screening models. COPD, occupational exposures and family history of lung cancer are additional risk factors that made individuals between ages 50 and 55 with between 20 and 30 pack years of smoking history eligible for LDCT screening in the NCCN screening guidelines (Table 1). Previous history of cancer is a risk factor in the PLCO and Liverpool screening models and in the Mayo and SPN lung nodule models. Race, education, and BMI are predictive variables only in the PLCO model.

When examining the totality of the four validated lung cancer screening models (Bach, Spitz, Tammemagi PLCO 2012, and Cassidy Liverpool), the Tammemagi PLCO 2012 model, by Tammemagi et al., included the same broad factors of risk found in the other three models except for occupational exposure to asbestos. The Tammemagi PLCO model used modern statistical prediction techniques and best model development practices. It also had the highest apparent

AUC (80) and experienced only slight reduction in AUC in external validation sets. Estimated AUC in validation sets varied between 70 and 78 depending on the cancer risk cut-off used to include patients in the validation sets. The intent of these screening models, which rely exclusively on epidemiologic information, is to determine the characteristics of a population that would most benefit from screening. This screened population is a subset of those who will develop lung cancer. Pinsky and Berg estimate that implementing a screening program based upon the NCCN guidelines or the more inclusive estimates of risk from the Tammemagi PLCO model will discover fewer than 10,000 of the 160,000 annual lung cancers.[139] Many of the remaining 150,000 cancers will present to clinicians as either incidentally discovered lung lesions or symptomatically. Irrespective of the presentation to a surgeon, imaging will be available and the host of information that imaging provides can be used to estimate the likelihood of cancer. One study found that information from a CT scan was better at discriminating individuals with lung cancer than all the epidemiologic information combined.[140]

1.6 Radiographic imaging and predicting lung cancer risk

Once a lung lesion has been discovered, the breadth of information that imaging provides can be used to further characterize the lesion and it's propensity to be cancerous. Imaging characteristics can be categorized broadly as either physical or metabolic. Physical characteristics include lesion shape, size and location in the lung. Lesions located in the upper lung are more likely to be of a smoking etiology and thus lung cancer, while lesions in the lower lung lobes are more commonly associated with infectious or granulomatous etiologies.[35,42] Spiculation, coronal, or spikey edge characteristics (Figure 4A and 4B) indicate undifferentiated cells and thus cancer. Smooth lesion edges are common to calcified and benign tumors.[27,141] Lung nodules smaller than 1 centimeter (cm), as measured by maximum diameter on a radiograph, are more often benign.[27] Nodules over 3 cm in diameter are 5 fold more likely to be cancerous than nodules about 1.5 cm in diameter.[142]

Lesion size as a predictive variable is common to all three models estimating lung cancer risk among patients with lung lesions.

Metabolic characteristics include growth, doubling rate or volume change and radiotracer estimated avidity. Clinicians often consider lesion growth as the strongest indicator of cancer risk.[143] Growth on serial radiographs is defined as an increase in mean diameter of 2 mm for nodules initially less than 15 mm in size and an increase of at least 15% compared to baseline scan for nodules more than 15 mm in size at baseline.[19] However, a rapid increase in size should raise the suspicion of an inflammatory process or the possibility of small cell lung cancer. Volumetric measurement of lesions is a fairly recent development that shows some promise in screening populations with repeated scans to estimate the likelihood a lesion is cancerous.[144] Lesion volume is generated with 3 dimensional radiographic measurement and estimates the doubling time of a lesion by measuring the change in volume between subsequent volumetric CT scans. Proponents of volumetric measurement state this method more accurately measures growth of a lesion.[145,146] This measurement has not been incorporated in any predictive models of lung lesions to date. Avidity measured by a PET scanner requires injection of a radiotracer and is not recommended until after a CT scan has been conducted. A combined PET-CT scan is considered among the most accurate, non-invasive diagnostic tests for lung cancer and is often conducted immediately prior to surgery.[19,46,48] FDG-PET avidity is included in the SPN model in addition to growth, lesion size and edge characteristics.[142]

FDG-PETs role in diagnosing lung cancer

Since the late 1990s, FDG-PET has become widely accepted for non-invasive clinical diagnosis and staging of lung cancer.[46,147] FDG-PET scans use a fluorine radiotracer (F^{18}) attached to a glucose receptor analogue (deoxyglucose). Hypermetabolism of lung tissue is measured by comparing tissue with higher concentrations of the radiotracer tagged glucose analogue (FDG)

which is above the background of surrounding lung tissue (Figure 8). When hypermetabolism is observed, the lesion is classified as being FDG avid or simply avid. Avidity is sometimes measured on a continuous scale called standard uptake value (SUV). SUV is estimated by measuring the average radionuclide activity concentration across the region of interest observed on the FDG-PET image at one time point and comparing that value with the background level measured in nearby lung tissue or by the injected radioactivity divided by body weight. An SUV above 2.5 is considered avid and suspicious for cancer by common convention among radiologists and thoracic clinicians.[148-150]

Metabolically active tissues preferentially consume the FDG glucose analogue. Neoplastic cells, inflammatory lesions, wounds and active benign tumors have higher glucose metabolism. Slow growing cancers, especially neuroendrocrine tumors or lesions with little metabolic activity, will often not be distinguished on a FDG-PET scan and such cancerous nodules are known to generate false negative results. Inflammatory lesions, wounds and growing benign tumors are often metabolically active and known to cause false positive scan results.[151,152]

A meta-analysis has shown FDG-PET to have high sensitivity (97%) and moderate specificity (78%) to identify lung cancer when the size of the lung nodule is greater than 1 cm.[47 47] The estimated AUC of FDG-PET to diagnose lesions greater than 1cm across all the studies in the meta-analysis was 91% and the likelihood ratio for a positive test was 7.1.[47 47] Based on the results of this study, FDG-PET was determined to be cost effective in the evaluation of patients with a solitary pulmonary nodule.[153 153] Subsequently, FDG-PET was approved by Centers for Medicare and Medicaid Services for diagnosis of lung nodules suspicious of lung cancer in 2001, and current clinical guidelines recommend its use.[46,154] In 2001, fusion FDG-PET/CT scanners became available. This scanning technology uses a computer program to adjust for breathing and other physiological factors to create a dynamic image of both the physical morphology of the

lung from the CT scan as well as the metabolic activity of the FDG-PET scan. Some studies have found the combination scanners to be more sensitive and specific in identifying lung cancer than PET scans alone.[155-157] A recent review of published fusion FDG-PET and CT scan studies found slightly higher accuracy but not significantly different than the earlier meta-analysis by Gould and colleagues.[48] The National Comprehensive Cancer Network guidelines advocate the use of combination FDG-PET/CT scans to diagnose solid or partially solid lung nodules whose maximum diameter is at least 8mm.

However, the accuracy of FDG-PET is inconsistent, and false positive results are associated with infectious fungal lung disease and other inflammatory or infectious disease processes. The accuracy of FDG-PET drops significantly in areas with endemic granulomatous disease and up to 30% of thoracic operations in these areas result in a benign diagnosis.[158,159] Unfortunately, most of the studies included in Gould's seminal 2001 meta-analysis were from medical centers in Europe, Japan and New England, areas that had low endemic granulomatous disease prevalence. The observed differences in the accuracy of FDG-PET to diagnose lung cancer may indicate an unintended selection bias in many of the earlier studies of FDG-PET accuracy used by the two meta-analyses. In some regions of the US where the prevalence of fungal lung disease is high or in Asia and Africa where tuberculosis is endemic [160-162], FDG-PET may not be effective in the diagnosis of lung cancer compared to CT alone.

1.7 Infectious lung diseases as a spectrum bias of FDG-PET for diagnosis of lung cancer

Infectious lung diseases that generate granulomas include fungal lung diseases from mycotic pathogens and mycobacteria, of which tuberculosis is the most common. Both classes of infectious pathogens have been shown to cause granulomas leading to false positive FDG-PET scans.[158,163-168] The three major mycotic infections in order of North American prevalence are

histoplasmosis, coccidioidomycosis and blastomycosis.[169,170] [169,170] Chronic exposure to each individual pathogen is not known, as cross-reactivity between pathogens to the diagnostic skin test occurs.[170] Prevalence of the three pathogens was 17.6% among naval recruits living in a single county screened between 1958 and 1969.[170] Acute episodes of disease requiring hospitalization or outpatient therapy have been reported in elderly populations. The highest incidence rate was for histoplasmosis at 3.3 per 100,000 person years, followed by coccidioidomycosis (3.2 per 100,000 person years) and blastomycosis (0.7 cases per 100,000 person years).[169] Other less common fungal diseases include cryptococcosis, aspergillosis, actinomyocis, nocardiosis, candidiasis, and echinococcosis. Histoplasmosis is endemic across the Mississippi, Ohio, Missouri and Tennessee River valleys and an estimated 50 million individuals have been infected with *histoplasmosis capsulatum.*[171] Coccidioidomycosis is endemic across the Southwest and San Joaquin valley in California.[172] The highest incidence of coccidioidomycosis occurs in Arizona.[173] Blastomycosis is the most broadly dispersed fungal disease of the three. Endemic areas occur across the southeastern US and up through the Great Lakes and as far west as Texas. Highest incidences are observed in the Ohio and Mississippi river valleys and lower Great Lakes area[169] (Figure 9A). Mycotic fungal spores reside in soil as well as in bat and avian feces. Farmers, construction workers and those who engage in outdoor activity in rural areas have the highest risk of inhaling the spores. Acute and deadly single site outbreaks have been documented and were associated with nearby soils having a high concentration of avian fecal material and subsequently being disturbed by construction equipment.[171,174]

Fungal lung disease is generally asymptomatic or causes flu-like symptoms. Immunocompetent individuals exposed to a fungal load that does not generate an acute response will create antibodies and typically eliminate the infection in less than 14 days. Reinfection is common in endemic areas. Current diagnosis of acute or disseminated pulmonary histoplasmosis is determined by serological or urine test.[175,176] Only the skin antigen test is sensitive to exposures

that occurred more than two to three years prior. Similar to the tuberculosis skin test, the skin test for histoplasmosis is cross-reactive to coccidioidomycosis and blastomycosis. A skin reaction occurred if the individual was previously exposed to one of the fungi, but the test is no longer commercially produced and reagents for the skin test are no longer available.[171]

An estimated 1 in 2,000 people infected by histoplasmosis will develop chronic pulmonary disease. Among individuals with symptomatic coccidioidomycosis disease who require antifungal treatment, 5% develop irreversible bronchiectasis, pulmonary nodules, or residual lung cavities.[177] Granuloma formation in the lung is one possible result for any fungal infection. Granulomatous disease of the lung can look identical to cancerous tissue on a CT or PET scan (Figure 8) and can have all the hallmarks of a cancerous lesion including symptomatic hemoptysis, growth on repeated CT scans, spiculated edge characteristics and lack of calcification.[178] While fungal lung diseases are the primary etiology of granulomas in the US and North America, tuberculosis is the more common cause of granulomas in Asia, Africa and developing countries.[161,162,164,177,179-184]

If infectious lung diseases like histoplasmosis in North America and tuberculosis in Southeast Asia and Africa generate granulomas and these granulomas reduce the specificity of FDG-PET scans to diagnose lung cancer, then the lack of published studies reflecting the impact of these infectious diseases on FDG-PET specificity may indicate a spectrum bias in the published literature and meta-analyses. Spectrum bias is a form of selection bias wherein the breadth or complexity of a disease is not well reflected in the population undergoing a diagnostic test[185]. Spectrum bias more commonly occurs when outliers in risk stratification are not included in the test population. For example, a cholesterol test is very accurate at classifying heart disease when the population with disease was recruited from a heart clinic and had moderate to high levels of cholesterol. An unbiased assessment of the test requires high and low levels of cholesterol in both

those with and without disease in order to reflect the possible population in which the test would be applied. Similarly, a diagnostic test may be highly accurate in moderately high or low risk strata but perform poorly in the tails of the risk distribution. With respect to FDG-PET scans, the prevalence of infectious lung disease and the granulomatous disease created by those infections may be causing the bias in published FDG-PET scan results.

For infectious lung diseases the geographic variation in the prevalence of these diseases may have inadvertently created just such a bias for FDG-PET scans. This bias occurs in those with granulomas that, in turn, only influences or confounds diagnostic imaging and not the probability of lung cancer. This bias only becomes important in the context of diagnosing lung cancer with diagnostic imaging. No published reports have found granulomatous disease increasing or decreasing the likelihood someone develops lung cancer. Thus infectious lung diseases are not a confounder of lung cancer. Rather, granulomas from the disease generate false positive FDG-PET results which then influence the clinician's estimate of the probability of lung cancer.

1.8 Geographic variation in the diagnosis of lung cancer

Evidence of geographic variation in the diagnosis of lung cancer after surgery or in the diagnostic accuracy of FDG-PET has implications for any models attempting to predict lung cancer in populations being evaluated by surgeons. Evidence of either implies that local variation in benign disease prevalence, patient work-up, fungal lung disease or other unknown factors occur and may influence a clinical prediction model.

There is a fourfold geographic variation in lung cancer incidence across the US. Kentucky has the highest rate among both men (125.1 per 100,000) and women (80.3 per 100,000), and Utah has the lowest rates (29.2 for men and 19.0 per 100,000 for women).[8] These incidence rates mirror

the historical smoking prevalence in those states and have remained fairly consistent over time. The variation in lung cancer rates follows smoking rates across the country. Fungal lung disease varies greatly by region of the country in the US. For example, the histoplasmosis prevalence in the central and western portions of Tennessee and Kentucky is between 80% and 97%, while in the Boston area and Pacific Northwest, it is a rare disease (Figure 9A and 9B).[170] Variation in benign lung disease rates may occur across the US as well and reflect underlying infectious disease prevalence. The interaction between local smoking prevalence and infectious lung disease prevalence is unknown. The influence of local practice or referral patterns, clinical expertise and availability of specialists on the prevalence of benign disease after a lung operation for suspected lung cancer is also unknown.

Whether geographic variation in the diagnosis of lung cancer after thoracoscopic surgery for suspected lung cancer occurs is unknown in the US. Answering this basic question is the first step in determining the scope of the variation in the surgical diagnosis of lung cancer. If such variation occurs and is clinically significant then examination of possible causes of that variation becomes important. One possible cause of variation in the US could be fungal lung diseases as a direct cause of benign lung diseases. Another indirect cause would be fungal lung diseases' influence on FDG-PET scan accuracy to diagnose lung cancer.

We can exploit the observed variation in the geographic distribution of fungal lung disease in North America to test the hypothesis that fungal lung disease confounds FDG-PET accuracy to diagnose lung cancer. If FDG-PET specificity varies inversely with endemic fungal lung disease, then we have both a possible measurement of the amount of bias in specificity caused by fungal lung disease and a strong argument supporting current clinical observation that fungal lung disease reduces FDG-PET accuracy to diagnose lung cancer. Secondly, if fungal lung disease is associated with benign disease after surgery for suspected lung cancer, then this is further

evidence of another factor that influences the surgical diagnosis of lung cancer. Some measurement of fungal lung disease should be considered in predicting lung cancer. Finally, reports of granulomas of the lung and their underlying infectious diseases should appear in the literature as other researchers seek to improve the diagnosis of lung cancer. A systematic review examining FDG-PET diagnostic accuracy could describe current knowledge as to the tests characteristics and factors that affect the diagnosis of lung cancer.

1.9 Conclusion

No clinical prediction model exists designed for surgeons evaluating a lung nodule prior to surgical biopsy. Existing models are either poorly calibrated to the higher prevalence of lung cancer in surgical populations or do not incorporate the breadth of information available to the surgeon at the point of decision to operate. Because lung cancer is so deadly, patients and providers must aggressively pursue a diagnosis to rule out cancer. The high rates of benign disease after surgery observed in many of the lung cancer studies will be unchanged without an intervention for surgeons to aid their decision.[50,56,59,62,186] The need for such a model will only increase as chest imaging technology progresses or a lung cancer screening regimen is implemented.

A prediction model estimated with data from a single institution has limited clinical usefulness if local variation in cancer or benign disease after surgery occurs across the US. Determining whether such variation is present and then exploring the possible causes of that variation are necessary prior to implementing a predictive model for lung cancer nationally. Some evidence indicates radiographic imaging accuracy to diagnose

lung cancer varies across the US. Since a predictive model for surgeons will rely heavily on such imaging, defining the extent of variation and possible causes of that variation will strengthen any prediction model for lung cancer applied nationwide.

This dissertation begins the process of creating a clinical prediction model, the TREAT model, for surgeons evaluating individuals for lung cancer. It explores whether variation in the diagnosis of lung cancer currently exists. This research also begins to examine possible causes of variation that influence what is currently considered the most accurate radiographic imaging, FDG-PET scans. These questions reflect current knowledge regarding predicting lung cancer and explore possible new information of import and for surgeons and patients struggling to answer the question, “Is it cancer?”

1.10 References

1. Kohler BA, Ward E, McCarthy BJ, et al. Annual report to the nation on the status of cancer, 1975-2007, featuring tumors of the brain and other nervous system. *J Natl Cancer Inst.* May 4 2011;103(9):714-736.

2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin.* 2012;62(1):10-29.

3. National CI. Cancer Trends Progress Report - 2009-2010 Update. *Cancer Trends Progress Report* 2010; http://www.cancer.gov/newscenter/newsfromnci/2010/ProgressReport2010. Accessed 01/20/2012.

4. Bradley CJ, Yabroff KR, Dahman B, Feuer EJ, Mariotto A, Brown ML. Productivity Costs of Cancer Mortality in the United States: 2000–2020. *J Natl Cancer Inst.* December 17, 2008 2008;100(24):1763-1770.

5. Alberg A, Ford J, Samet J. Epidemiology of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest.* 2007;132:29S - 55S.

6. Wakelee HA, Chang ET, Gomez SL, et al. Lung Cancer Incidence in Never Smokers. *J Clin Oncol.* February 10, 2007 2007;25(5):472-478.

7. National CI. Fast Stats. *Fast Stats* 2012. Accessed 01/23/2012, 2012.

8. National Cancer Institute. State cancer profiles. 2011; http://statecancerprofiles.cancer.gov/. Accessed 12/01/2011, 2011.

9. Howlader N, Noone A, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2009. *SEER Cancer Statistics Review* 2012; http://seer.cancer.gov/csr/1975_2009_pops09/index.html. Accessed 2/22/2013, 2013.

10. Ries L, Young J, Keel G, Eisner M, Lin Y, Horner M-J. SEER Survival Monograph: Cancer Survival Among Adults: U.S. SEER Program, 1988-2001, Patient and Tumor Characteristics. In: Institute NC, ed. NIH Pub. No. 07-6215 ed: National Institutes of Health; 2007.

11. Vallieres EMDF, Shepherd FAMDF, Crowley JP, et al. The IASLC Lung Cancer Staging Project: Proposals Regarding the Relevance of TNM in the Pathologic Staging of Small

Cell Lung Cancer in the Forthcoming (Seventh) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol.* 2009;4(9):1049-1059.

12. Scott WJ, Howington J, Feigenberg S, Movsas B, Pisters K. Treatment of non-small cell lung cancer stage I and stage II: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest.* Sep 2007;132(3 Suppl):234S-242S.

13. Oken Mm HWGKPA, et al. Screening by chest radiograph and lung cancer mortality: The prostate, lung, colorectal, and ovarian (plco) randomized trial. *JAMA.* 2011;306(17):1865-1873.

14. *Guide to Clinical Preventive Services: An Assessment of the Effectiveness of 169 Interventions. Report of the US Preventive Services Task Force.* Baltimore, MD: Williams & Wilkins;1989.

15. Gennari R, Veronesi U, Andreoli C, et al. Early detection of cancer: ideas for a debate. *Crit Rev Oncol Hematol.* Feb 2007;61(2):97-103.

16. Bach PB, Silvestri GA, Hanger M, Jett JR. Screening for Lung Cancer ACCP guidelines *Chest.* 2007;132:69S-77S.

17. Sox HC. Better Evidence about Screening for Lung Cancer. *N Engl J Med.* 2011;365(5):455-457.

18. Bach PB, Mirkin JN, Oliver TK, et al. Benefits and harms of ct screening for lung cancer: A systematic review. *JAMA.* 2012;307(22):2418-2429.

19. National Comprehensive Cancer Network. Lung Cancer Screening v1.2012. *NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines).* 2011. http://www.nccn.org/professionals/physician_gls/recently_updated.asp. Accessed 10-27-2011.

20. Pyenson BS, Sander MS, Jiang Y, Kahn H, Mulshine JL. An Actuarial Analysis Shows That Offering Lung Cancer Screening As An Insurance Benefit Would Save Lives At Relatively Low Cost. *Health Aff (Millwood).* April 1, 2012 2012;31(4):770-779.

21. McMahon PM, Kong CY, Bouzan C, et al. Cost-Effectiveness of Computed Tomography Screening for Lung Cancer in the United States. *J Thorac Oncol.* 2011;6(11):1841-1848 1810.1097/JTO.1840b1013e31822e31859b31823.

22. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *N Engl J Med.* 2011;365(5):395-409.

23. US Preventive Services Task Force. Topics in Progress. 2012; http://www.uspreventiveservicestaskforce.org/uspstf/topicsprog.htm. Accessed 11-27-2012.

24. CDC. Cigarette Smoking Among Adults and Trends in Smoking Cessation - US , 2008. In: Health and Human Services Centers for Disease Control and Prevention, ed. Vol MMWR CDC Surveill Summ. 2009:1227-1232.

25. Institute NC. Tobacco Use Supplement from Current Population Survey. *TUS-CPS* 2010; http://riskfactor.cancer.gov/studies/tus-cps/info.html. Accessed 01/09/2013.

26. NCI. Lung cancer screening trial results show mortality benefit with low-does CT. In: Institute NC, editor: NCI office of Media Relations, 2010.

27. Gould MK, Fletcher J, Iannettoni MD, et al. Evaluation of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest.* Sep 2007;132(3 Suppl):108S-130S.

28. Austin J, Muller N, Friedman P, et al. Glossary of terms for CT of the lungs: recommendations of the Nomenclature Committee of the Fleischner Society. *Radiology.* 1996;200:327 - 331.

29. Bach PB, Kattan MW, Thornquist MD, et al. Variations in lung cancer risk among smokers. *J Natl Cancer Inst.* Mar 19 2003;95(6):470-478.

30. Spitz MR, Etzel CJ, Dong Q, et al. An Expanded Risk Prediction Model for Lung Cancer. *Cancer Prev Res.* September 1, 2008 2008;1(4):250-254.

31. Cassidy A, Duffy SW, Myles JP, Liloglou T, Field JK. Lung cancer risk prediction: A tool for early detection. *Int J Cancer.* 2007;120(1):1-6.

32. Tammemagi MCP, Freedman MTMD, Pinsky PFP, et al. Prediction of True Positive Lung Cancers in Individuals with Abnormal Suspicious Chest Radiographs-A Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial Study. *J Thorac Oncol.* 2009;4(6):710-721.

33. Tammemägi MC, Katki HA, Hocking WG, et al. Selection Criteria for Lung-Cancer Screening. *N Engl J Med.* 2013;368(8):728-736.

34. Cronin KA, Gail MH, Zou Z, Bach PB, Virtamo J, Albanes D. Validation of a model of lung cancer risk prediction among smokers. *J Natl Cancer Inst.* May 3 2006;98(9):637-640.

35. Gould MK, Ananth L, Barnett PG. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest.* Feb 2007;131(2):383-388.

36. Schultz EM, Sanders GD, Trotter PR, et al. Validation of two models to estimate the probability of malignancy in patients with solitary pulmonary nodules. *Thorax.* Apr 2008;63(4):335-341.

37. Raji OY, Duffy SW, Agbaje OF, et al. Predictive Accuracy of the Liverpool Lung Project Risk Model for Stratifying Patients for Computed Tomography Screening for Lung CancerA Case–Control and Cohort Validation Study. *Ann Intern Med.* 2012;157(4):242-250.

38. Cassidy A, Myles JP, van Tongeren M, et al. The LLP risk model: an individual risk prediction model for lung cancer. *Br J Cancer.* 2007;98(2):270-276.

39. Hoggart C, Brennan P, Tjonneland A, et al. A Risk Model for Lung Cancer Incidence. *Cancer Prev Res.* April 11, 2012 2012.

40. Spitz MR, Hong WK, Amos CI, et al. A risk model for prediction of lung cancer. *J Natl Cancer Inst.* May 2 2007;99(9):715-726.

41. MacMahon H, Austin J, Gamsu G, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology.* 2005;237:395 - 400.

42. Swensen S, Silverstein M, Ilstrup D, Schleck C, Edell E. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med.* 1997;157:849 - 855.

43. Gurney JW, Lyddon DM, McKay JA. Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part II. Application. *Radiology.* February 1993 1993;186(2):415-422.

44. Alberts WM. Epidemiology of Lung Cancer ACCP Evidence-Based Clinical Practice Guidelines (2nd Edition). *Chest.* 2007;132:29S-55S.

45. Rivera MP, Mehta AC. Initial diagnosis of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest.* Sep 2007;132(3 Suppl):131S-148S.

46. Wahidi MM, Govert JA, Goudar RK, Gould MK, McCrory DC. Evidence for the treatment of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest.* Sep 2007;132(3 Suppl):94S-107S.

47. Gould MK, Maclean CC, Kuschner WG, Rydzak CE, Owens DK. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis. *JAMA.* Feb 21 2001;285(7):914-924.

48. Cronin P, Dwamena B, Kelly A, Bernstein S, Carlos R. Solitary pulmonary nodules and masses: a meta-analysis of the diagnostic utility of alternative imaging tests. *Eur Radiol.* 2008;18(9):1840-1856.

49. Probability of Malignancy in SPN: Bayesian Analysis. 2000; http://www.chestx-ray.com/SPN/SPNProb.html. Accessed January 1, 2010.

50. Isbell JM, Deppen S, Putnam JB, Jr., et al. Existing general population models inaccurately predict lung cancer risk in patients referred for surgical evaluation. *Ann Thorac Surg.* Jan 2011;91(1):227-233; discussion 233.

51. Wiener RS, Schwartz LM, Woloshin S, Welch HG. Population-Based Risk for Complications After Transthoracic Needle Lung Biopsy of a Pulmonary Nodule: An Analysis of Discharge Records. *Ann Intern Med.* August 2, 2011 2011;155(3):137-144.

52. Anderson JM, Murchison J, Patel D. CT-guided lung biopsy: factors influencing diagnostic yield and complication rate. *Clin Radiol.* Oct 2003;58(10):791-797.

53. Ernst AMD, Eberhardt RMD, Krasnik MMD, Herth FJFMD. Efficacy of Endobronchial Ultrasound-Guided Transbronchial Needle Aspiration of Hilar Lymph Nodes for Diagnosing and Staging Cancer. *J Thorac Oncol.* 2009;4(8):947-950.

54. Li H, Boiselle PM, Shepard JO, Trotman-Dickenson B, McLoud TC. Diagnostic accuracy and safety of CT-guided percutaneous needle aspiration biopsy of the lung: comparison of small and large pulmonary nodules. *AJR Am J Roentgenol.* Jul 1996;167(1):105-109.

55. Grogan EL, Jones DR. VATS lobectomy is better than open thoracotomy: what is the evidence for short-term outcomes? *Thorac Surg Clin.* Aug 2008;18(3):249-258.

56. Silvestri GA. Screening for Lung Cancer: It Works, but Does It Really Work? *Ann Intern Med.* September 5 2011:E-364.

57. El-Tamer MB, Ward BM, Schifftner T, Neumayer L, Khuri S, Henderson W. Morbidity and Mortality Following Breast Cancer Surgery in Women: National Benchmarks for Standards of Care. *Ann Surg.* 2007;245(5):665-671

58. Deppen SA, Aldrich MC, Walker R, et al. A new model improves lung cancer prediction in the preoperative evaluation of patients with suspicious lung lesions. . Paper presented at: Annual Meeting of American Association of Cancer Research; April 6-10, 2013; Washington D.C.

59. Grogan E, Deppen SA, Ballman K, et al. Accuracy of FDG-PET to diagnose lung cancer in the ACOSOG Z4031 trial. Paper presented at: Alliance for Clinical Trials in Oncology; June 29-July1, 2012, 2012; Chicago, IL

60. Veronesi G, Bellomi M, Scanagatta P, et al. Difficulties encountered managing nodules detected during a computed tomography lung cancer screening program. *J Thorac Cardiovasc Surg.* Sep 2008;136(3):611-617.

61. Crestanello JA, Allen MS, Jett JR, et al. Thoracic surgical operations in patients enrolled in a computed tomographic screening trial. *J Thorac Cardiovasc Surg.* Aug 2004;128(2):254-259.

62. Kozower BD, Meyers BF, Reed CE, Jones DR, Decker PA, Putnam Jr JB. Does Positron Emission Tomography Prevent Nontherapeutic Pulmonary Resections for Clinical Stage IA Lung Cancer? *Ann Thorac Surg.* 2008;85(4):1166-1170.

63. Doll R, Hill A. The mortaltiy of doctors in relation to theri smoking habits; a preliminary report. *BMJ.* 1954;1(4877):1451-1455.

64. Emmert-Streib F, Dehmer M, SpringerLink (Online service). Information theory and statistical learning. New York ; [London]: Springer; 2009: http://dx.doi.org/10.1007/978-0-387-84816-7.

65. Paz-Elizur T, Krupsky M, Blumenstein S, Elinger D, Schechtman E, Livneh Z. DNA repair activity for oxidative damage and risk of lung cancer. *J Natl Cancer Inst.* Sep 3 2003;95(17):1312-1319.

66. Hecht SS. Lung carcinogenesis by tobacco smoke. *Int J Cancer*. 2012;131(12):2724-2732.

67. Siegel R, Ward E, Brawley O, Jemal A. Cancer statistics, 2011. *CA Cancer J Clin.* 2011;61(4):212-236.

68. Song Y-M, Sung J, Cho H-J. Reduction and Cessation of Cigarette Smoking and Risk of Cancer: A Cohort Study of Korean Men. *J Clin Oncol.* November 1, 2008 2008;26(31):5101-5106.

69. Ebbert JO, Yang P, Vachon CM, et al. Lung Cancer Risk Reduction After Smoking Cessation: Observations From a Prospective Cohort of Women. *J Clin Oncol.* March 1, 2003 2003;21(5):921-926.

70. Alberg AJ, Nonemaker J. Who Is at High Risk for Lung Cancer? Population-Level and Individual-Level Perspectives. *Semin Respir Crit Care Med.* 02.06.2008 2008;29(03):223,232.

71. Lubin JH, Caporaso N, Wichmann HE, Schaffrath-Rosario A, Alavanja MCR. Cigarette Smoking and Lung Cancer: Modeling Effect Modification of Total Exposure and Intensity. *Epidemiology.* 2007;18(5):639-648.

72. Burns D, Anderson C, Gray N. Has the lung cancer risk from smoking increased over the last fifty years? *Cancer Causes and Control.* 2011;22(3):389-397.

73. Thun MJ, Hannan LM, Adams-Campbell LL, et al. Lung Cancer Occurrence in Never-Smokers: An Analysis of 13 Cohorts and 22 Cancer Registry Studies. *PLoS Med.* 2008;5(9):e185.

74. Westmaas J, Brandon T. Reducing Risk in Smokers. *Curr Opin Pulm Med.* 2004;10(4):284-288.

75. Halpern MT, Gillespie BW, Warner KE. Patterns of Absolute Risk of Lung Cancer Mortality in Former Smokers. *J Natl Cancer Inst.* March 17, 1993 1993;85(6):457-464.

76. Peto R, Darby S, Deo H, Silcocks P, Whitley E, Doll R. Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. *BMJ.* 2000;321(7257):323-329.

77. Services USDoHaH. The Health Consequences of Smoking: A Report of the Surgeon General. In: U.S. Department of Health and Human Services PHS, ed. Rockville, MD: Centers for Disease Control 2004.

78. Ando M, Wakai K, Seki N, et al. Attributable and absolute risk of lung cancer death by smoking status: Findings from the Japan collaborative cohort study. *Int J Cancer.* 2003;105(2):249-254.

79. Burns DM. Primary prevention, smoking, and smoking cessation: implications for future trends in lung cancer prevention. *Cancer.* 2000;89(11 Suppl):2506-2509.

80. Henschke CI, Miettinen OS. Women's susceptibility to tobacco carcinogens. *Lung Cancer.* 2004;43(1):1-5.

81. Henschke CI, Yip R, Miettinen OS. Women's susceptibility to tobacco carcinogens and survival after diagnosis of lung cancer. *JAMA.* Jul 12 2006;296(2):180-184.

82. Risch HA, Howe GR, Jain M, Burch JD, Holowaty EJ, Miller AB. Are Female Smokers at Higher Risk for Lung Cancer than Male Smokers? *Am J Epidem.* September 1, 1993 1993;138(5):281-293.

83. American Lung Association. *Trends in tobacco use.* 2011.

84. Jemal A, Travis WD, Tarone RE, Travis L, Devesa SS. Lung cancer rates convergence in young men and women in the United States: Analysis by birth cohort and histologic type. *Int J Cancer.* 2003;105(1):101-107.

85. Bain C, Feskanich D, Speizer FE, et al. Lung Cancer Rates in Men and Women With Comparable Histories of Smoking. *J Natl Cancer Inst.* June 2, 2004 2004;96(11):826-834.

86. Freedman N, Leitzmann M, Hollenbeck A, Schatzkin A, Abnet C. Cigarette smoking and subsequent risk of lung cancer in men and women: analysis of a prospective cohort study. *Lancet Oncol.* 2008;9:649 - 656.

87. De Matteis S, Consonni D, Pesatori AC, et al. Are Women Who Smoke at Higher Risk for Lung Cancer Than Men Who Smoke? *Am J Epidem.* February 20, 2013 2013.

88. Thun MJ, Jemal A. How much of the decrease in cancer death rates in the United States is attributable to reductions in tobacco smoking? *Tob Control.* Oct 2006;15(5):345-347.

89. Jemal A, Ward E, Thun M. Declining death rates reflect progress against cancer. *PLoS One.* 2010;5(3):e9584.

90. Wong MP, Fung L-F, Wang E, et al. Chromosomal aberrations of primary lung adenocarcinomas in nonsmokers. *Cancer.* 2003;97(5):1263-1270.

91. Patel JD. Lung Cancer in Women. *J Clin Oncol.* May 10, 2005 2005;23(14):3212-3218.

92. Kligerman S, White C. Epidemiology of Lung Cancer in Women: Risk Factors, Survival, and Screening. *American Journal of Roentgenology.* February 1, 2011 2011;196(2):287-295.

93. Davis DL. Trends in Nonsmoking Lung Cancer. *Epidemiology.* 1993;4(6):489-491.

94. Haiman CA, Stram DO, Wilkens LR, et al. Ethnic and Racial Differences in the Smoking-Related Risk of Lung Cancer. *N Engl J Med.* 2006;354(4):333-342.

95. Stellman SD, Chen Y, Muscat JE, et al. Lung cancer risk in white and black Americans. *Ann Epidemiol.* Apr 2003;13(4):294-302.

96. Kabat GC, Shivappa N, Hébert JR. Mentholated cigarettes and smoking-related cancers revisited: An ecologic examination. *Regul Toxicol Pharmacol.* 2012;63(1):132-139.

97. Okuyemi KS, Faseru B, Sanderson Cox L, Bronars CA, Ahluwalia JS. Relationship between menthol cigarettes and smoking cessation among African American light smokers. *Addiction.* 2007;102(12):1979-1986.

98. Blot WJ, Cohen SS, Aldrich M, McLaughlin JK, Hargreaves MK, Signorello LB. Lung cancer risk among smokers of menthol cigarettes. *J Natl Cancer Inst.* May 18 2011;103(10):810-816.

99. Murray RP, Connett JE, Skeans MA, Tashkin DP. Menthol Cigarettes and Health Risks in Lung Health Study Data. *Nicotine Tob Res.* January 1, 2007 2007;9(1):101-107.

100. Hirayama T. Non-smoking wives of heavy smokers have a higher risk of lung cancer: a study from Japan. *BMJ.* 1981-01-17 00:00:00 1981;282(6259):183-185.

101. Trichopoulos D, Kalandidi A, Sparros L, MacMahon B. Lung cancer and passive smoking. *Int J Cancer.* 1981;27(1):1-4.

102. Fontham Eh CPRP, et al. Environmental tobacco smoke and lung cancer in nonsmoking women: A multicenter study. *JAMA.* 1994;271(22):1752-1759.

103. Agency UEP. Respiratory health effects of passive smoking: lung cancer and other disorders. In: EPA, ed. Vol EPA/600/600F. Washington, DC: US Government Printing Office; 1992.

104. Hackshaw AK, Law MR, Wald NJ. The accumulated evidence on lung cancer and environmental tobacco smoke. *BMJ.* October 18, 1997 1997;315(7114):980-988.

105. Wolfinger RD, Gibson G, ED. W. Assessing gene significance from cDNA microarray expression data via mixed models. Submitted.

106. Pirkle JL, Bernert JT, Caudill SP, Sosnoff CS, Pechacek TF. Trends in the Exposure of Nonsmokers in the U.S. Population to Secondhand Smoke: 1988–2002. *Environ Health Perspect.* 2006;114(6).

107. Samet JM, Avila-Tang E, Boffetta P, et al. Lung Cancer in Never Smokers: Clinical Epidemiology and Environmental Risk Factors. *Clin Cancer Res.* September 15, 2009 2009;15(18):5626-5645.

108. Committee on Health Risks of Exposure to Radon BV. *Health Effects of Exposure to Radon.* National Academies Press; 1999.

109. Zeka A, Mannetje At, Zaridze D, et al. Lung Cancer and Occupation in Nonsmokers: A Multicenter Case-Control Study in Europe. *Epidemiology.* 2006;17(6):615-623 610.1097/1001.ede.0000239582.0000292495.b0000239585.

110. Veglia F, Vineis P, Overvad K, et al. Occupational Exposures, Environmental Tobacco Smoke, and Lung Cancer. *Epidemiology.* 2007;18(6):769-775 710.1097/EDE.1090b1013e318142c318148a318141.

111. Boffetta P, Autier P, Boniol M, et al. An Estimate of Cancers Attributable to Occupational Exposures in France. *J Occup Environ Med.* 2010;52(4):399-406 310.1097/JOM.1090b1013e3181d1095e1355.

112. Darby S, Hill D, Auvinen A, et al. Residential radon and lung cancer—detailed results of a collaborative analysis of individual data on 7148 persons with lung cancer and 14 208 persons without lung cancer from 13 epidemiologic studies in Europe. *Scand J Work Environ Health.* 2006;32(Supplement 1).

113. Olsson AC, Gustavsson P, Zaridze D, et al. Lung Cancer Risk Attributable to Occupational Exposures in a Multicenter Case-Control Study in Central and Eastern Europe. *J Occup Environ Med.* 2011;53(11):1262-1267 1210.1097/JOM.1260b1013e318234e318232d318232.

114. Brown T, Darnton A, Fortunato L, Rushton L. Occupational cancer in Britain. *Br J Cancer*. 2012;107(S1):S56-S70.

115. Whitrow MJ, Smith BJ, Pilotto LS, Pisaniello D, Nitschke M. Environmental exposure to carcinogens causing lung cancer: Epidemiological evidence from the medical literature. *Respirology*. 2003;8(4):513-521.

116. National CI. Table I-11 Median Age of Patients at Diagnosis, 2004-2008. 2011; http://seer.cancer.gov/csr/1975_2008/results_single/sect_01_table.11_2pgs.pdf. Accessed 01/23/2012, 2012.

117. Rauscher GH, Mayne ST, Janerich DT. Relation between Body Mass Index and Lung Cancer Risk in Men and Women Never and Former Smokers. *Am J Epidem.* September 15, 2000 2000;152(6):506-513.

118. Smith L, Brinton LA, Spitz MR, et al. Body Mass Index and Risk of Lung Cancer Among Never, Former, and Current Smokers. *J Natl Cancer Inst.* March 27, 2012 2012.

119. Tarnaud C, Guida F, Papadopoulos A, et al. Body mass index and lung cancer risk: results from the ICARE study, a large, population-based case–control study. *Cancer Causes and Control.* 2012;23(7):1113-1126.

120. Young RP, Hopkins RJ, Christmas T, Black PN, Metcalf P, Gamble GD. COPD prevalence is increased in lung cancer, independent of age, sex and smoking history. *Euro Resp J.* August 1, 2009 2009;34(2):380-386.

121. de Torres JP, Marín JM, Casanova C, et al. Lung Cancer in Patients with Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med.* October 15, 2011 2011;184(8):913-919.

122. Koshiol J, Rotunno M, Consonni D, et al. Chronic Obstructive Pulmonary Disease and Altered Risk of Lung Cancer in a Population-Based Case-Control Study. *PLoS One.* 2009;4(10):e7380.

123. Usui K, Tanai C, Tanaka Y, Noda H, Ishihara T. The prevalence of pulmonary fibrosis combined with emphysema in patients with lung cancer. *Respirology.* 2011;16(2):326-331.

124. Mizuno S, Takiguchi Y, Fujikawa A, et al. Chronic obstructive pulmonary disease and interstitial lung disease in patients with lung cancer. *Respirology.* 2009;14(3):377-383.

125. Sekine Y, Katsura H, Koh E, Hiroshima K, Fujisawa T. Early detection of COPD is important for lung cancer surveillance. *Euro Resp J.* November 16, 2011 2011.

126. Amos CI, Xu W, Spitz MR. Is there a genetic basis for lung cancer susceptibility? *Recent Results Cancer Res.* 1999;151:3-12.

127. Matakidou A, Eisen T, Houlston RS. Systematic review of the relationship between family history and lung cancer risk. *Br J Cancer.* 2005;93(7):825-833.

128. Braun MM, Caporaso NE, Hoover RN, Page WF. Genetic component of lung cancer: cohort study of twins. *The Lancet.* 1994;344(8920):440-443.

129. Bailey-Wilson JE, Amos CI, Pinney SM, et al. A major lung cancer susceptibility locus maps to chromosome 6q23-25. *Am J Hum Genet.* Sep 2004;75(3):460-474.

130. Ji Y-n, Wang Q, Lin X-q, Suo L-j. CYP1A1 MspI polymorphisms and lung cancer risk: An updated meta-analysis involving 20,209 subjects. *Cytokine.* 2012;59(2):324-334.

131. Pavanello S, Fedeli U, Mastrangelo G, et al. Role of CYP1A2 polymorphisms on lung cancer risk in a prospective study. *Cancer Genet.* 2012;205(6):278-284.

132. McWilliams JE, Sanderson BJ, Harris EL, Richert-Boe KE, Henner WD. Glutathione S-transferase M1 (GSTM1) deficiency and lung cancer risk. *Cancer Epidemiol Biomarkers Prev.* 1995;4(6):589-594.

133. Ye Z, Song H, Higgins JP, Pharoah P, Danesh J. Five glutathione s-transferase gene variants in 23,452 cases of lung cancer and 30,397 controls: meta-analysis of 130 studies. *PLoS Med.* Apr 2006;3(4):e91.

134. Tsou JA, Hagen JA, Carpenter CL, Laird-Offringa IA. DNA methylation analysis: a powerful new tool for lung cancer diagnosis. *Oncogene.* Aug 12 2002;21(35):5450-5461.

135. Schwartz AG, Prysak GM, Bock CH, Cote ML. The molecular epidemiology of lung cancer. *Carcinogenesis.* Mar 2007;28(3):507-518.

136. Chen L-S, Saccone NL, Culverhouse RC, et al. Smoking and Genetic Risk Variation Across Populations of European, Asian, and African American Ancestry—A Meta-Analysis of Chromosome 15q25. *Genet Epidemiol.* 2012;36(4):340-351.

137. Truong T, Hung RJ, Amos CI, et al. Replication of Lung Cancer Susceptibility Loci at Chromosomes 15q25, 5p15, and 6p21: A Pooled Analysis From the International Lung Cancer Consortium. *J Natl Cancer Inst.* July 7, 2010 2010;102(13):959-971.

138. Hung RJ, Christiani DC, Risch A, et al. International Lung Cancer Consortium: Pooled Analysis of Sequence Variants in DNA Repair and Cell Cycle Pathways. *Cancer Epidemiol Biomarkers Prev.* November 1, 2008 2008;17(11):3081-3089.

139. Efron B, Tibshirani R. Improvements on Cross-Validation: The .632+ Bootstrap Method. *J Am Stat Assoc.* 1997;92(438):548-560.

140. Pecot CV, Li M, Zhang XJ, et al. Added value of a serum proteomic signature in the diagnostic evaluation of lung nodules. *Cancer Epidemiol Biomarkers Prev.* May 2012;21(5):786-792.

141. Gurney JW. Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part I. Theory. *Radiology.* February 1993 1993;186(2):405-413.

142. Gurney JW. Chest x-ray SPN risk estimator. 2009; http://www.chestx-ray.com/SPN/SPNProb.html.

143. Swensen S, Silverstein M, Edell E, et al. Solitary pulmonary nodules: clinical prediction model versus physicians. *Mayo Clin Proc.* 1999;74:319 - 329.

144. Xu DM, van der Zaag-Loonen HJ, Oudkerk M, et al. Smooth or Attached Solid Indeterminate Nodules Detected at Baseline CT Screening in the NELSON Study: Cancer Risk during 1 Year of Follow-up1. *Radiology.* January 2009 2009;250(1):264-272.

145. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of Lung Nodules Detected by Volume CT Scanning. *N Engl J Med.* 2009;361(23):2221-2229.

146. Zhao Y, Bock G, Vliegenthart R, et al. Performance of computer-aided detection of pulmonary nodules in low-dose CT: comparison with double reading by nodule volume. *Eur Radiol.* 2012/10/01 2012;22(10):2076-2084.

147. Rivera MP, Detterbeck F, Mehta AC. Diagnosis of lung cancer: the guidelines. *Chest.* Jan 2003;123(1 Suppl):129S-136S.

148. Bryant AS, Cerfolio RJ. The Maximum Standardized Uptake Values on Integrated FDG-PET/CT Is Useful in Differentiating Benign From Malignant Pulmonary Nodules. *Ann Thorac Surg.* 2006;82(3):1016-1020.

149. Keyes JW, Jr. SUV: Standard Uptake or Silly Useless Value? *J Nucl Med.* October 1, 1995 1995;36(10):1836-1839.

150. Menda Y, Bushnell DL, Madsen H, McLaughlin K, Kahn D, Kernstine KH. Evaluation of various corrections to the standardized uptake value for diagnosis of pulmonary malignancy. *Nucl Med Commun.* 2001;22(10):1077-1081.

151. Delbeke D, Coleman RE, Guiberteau MJ, et al. Procedure Guideline for Tumor Imaging with 18F-FDG PET/CT 1.0. *J Nucl Med.* May 2006 2006;47(5):885-895.

152. Erasmus JJ, Patz Jr EF. Positron emission tomography imaging in the thorax. *Clin Chest Med.* 1999;20(4):715-722.

153. Gould MK, Sanders GD, Barnett PG, et al. Cost-Effectiveness of Alternative Management Strategies for Patients with Solitary Pulmonary Nodules. *Annals of Internal Medicine.* May 6, 2003 2003;138(9):724-735.

154. Centers for Medicare & Medicaid Services. U.S. Department of Health and Human Services - Medicare National Coverage Determinations Manual. http://www.cms.gov/manuals/downloads/ncd103c1_Part4.pdf. Accessed 9/14/2011, 2011.

155. Chang C-Y, Tzao C, Lee S-C, et al. Incremental Value of Integrated FDG-PET/CT in Evaluating Indeterminate Solitary Pulmonary Nodule for Malignancy. *Mol Imaging Biol.* 2010;12(2):204-209.

156. Kagna O, Solomonov A, Keidar Z, et al. The value of FDG-PET/CT in assessing single pulmonary nodules in patients at high risk of lung cancer. *Eur J Nucl Med Mol Imaging.* Jun 2009;36(6):997-1004.

157. Bar-Shalom R, Kagna O, Israel O, Guralnik L. Noninvasive diagnosis of solitary pulmonary lesions in cancer patients based on 2-fluoro-2-deoxy-D-glucose avidity on positron emission tomography/computed tomography. *Cancer.* 2008;113(11):3213-3221.

158. Deppen S, Putnam JB, Jr., Andrade G, et al. Accuracy of FDG-PET to diagnose lung cancer in a region of endemic granulomatous disease. *Ann Thorac Surg.* Aug 2011;92(2):428-432; discussion 433.

159. Croft DR, Trapp J, Kernstine K, et al. FDG-PET imaging and the diagnosis of non-small cell lung cancer in a region of high histoplasmosis prevalence. *Lung Cancer.* 2002;36(3):297-301.

160. Hung GU, Shiau YC, Tsai SC, Ho YJ, Kao CH, Yen RF. Differentiation of radiographically indeterminate solitary pulmonary nodules with. *Jpn J Clin Oncol.* Feb 2001;31(2):51-54.

161. Li Y, Su M, Li F, Kuang A, Tian R. The value of 18-F-FDG-PET/CT in the differential diagnosis of solitary pulmonary nodules in areas with a high incidence of tuberculosis. *Ann Nucl Med.*1-8.

162. Sathekge M, Maes A, Pottel H, Stoltz A, van de Wiele C. Dual time-point FDG-PET-CT for differentiating benign form malignant solitary pulmonary nodules in a TB endemic area. *S Afr Med J.* 2010;100(9):598-601.

163. Adachi K, Watanabe F. [Lung cancer detected by fluorodeoxyglucose-positron emission tomography/computed tomography in the course of Mycobacterium avium infection; report of a case]. *Kyobu Geka.* 2009;62(11):1019-1021.

164. Chundru S, Wong CY, Wu D, et al. Granulomatous disease: is it a nuisance or an asset during PET/computed tomography evaluation of lung cancers? *Nucl Med Commun.* Jul 2008;29(7):623-627.

165. Gurney JW, Conces DJ. Pulmonary histoplasmosis. *Radiology.* May 1, 1996 1996;199(2):297-306.

166. Mukhopadhyay S, Gal AA. Granulomatous Lung Disease: An Approach to the Differential Diagnosis. *Arch Pathol Lab Med.* 2010/05/01 2010;134(5):667-690.

167. Salhab K, Baram D, Bilfinger T. Growing PET positive nodule in a patient with histoplasmosis: case report. *J Cardiothorac Surg.* 2006;1(1):23.

168. Smith MA, Battafarano RJ, Meyers BF, Zoole JB, Cooper JD, Patterson GA. Prevalence of Benign Disease in Patients Undergoing Resection for Suspected Lung Cancer. *Ann Thorac Surg.* 2006;81(5):1824-1829.

169. Baddley J, Winthrop K, Patkar N, et al. Geographic Distribution of Endemic Fungal Infections among Older Persons, United States. *Emerg Inf Dis.* 2011;17(9):1664-1669.

170. Edwards L, Acquaviva F, Livesay V, Cross F, Palmer C. An altas of sensitivity to tuberculin, PPD-B and histoplasmin in the United States. *Am Rev Resp Dis.* 1969;99(4):S1-132.

171. Lenhart SW, Schafer M, Singal M, Hajjeh RA. Histoplasmosis: Protecting workers at risk. In: Services DoHaH, ed. 2 ed. Cincinnati, OH: NIOSH; 2004:1-39.

172. Limper AH, Knox KS, Sarosi GA, et al. An Official American Thoracic Society Statement: Treatment of Fungal Infections in Adult Pulmonary and Critical Care Patients. *Am. J. Respir. Crit. Care Med.* January 1, 2011 2011;183(1):96-128.

173. CDC. MMWR. Vol 52. Atlanta, GA 2003:109-112.

174. Hage CA, Wheat LJ, Loyd J, Allen SD, Blue D, Knox KS. Pulmonary Histoplasmosis. *Semin Respir Crit Care Med.* 26.03.2008 2008;29(02):151,165.

175. Hage CA, Ribes JA, Wengenack NL, et al. A Multicenter Evaluation of Tests for Diagnosis of Histoplasmosis. *Clin Inf Dis.* September 1, 2011 2011;53(5):448-454.

176. Kauffman CA. Histoplasmosis: a Clinical and Laboratory Update. *Clinical Microbiology Reviews.* January 2007 2007;20(1):115-132.

177. Doty J. Pulmonary Fungus and Tuberculosis. 2002; http://www.ctsnet.org/doc/5753. Accessed 10/09/2011, 2011.

178. Nguyen BD. F-18 FDG PET/CT Imaging of Disseminated Coccidioidomycosis. *Clin Nucl Med.* 2006;31(9):568-571.

179. Mason R, Broaddus V, Martin T, et al., eds. *Murray and Nadel's Textbook of Respiratory Medicine.* 5th ed: Saunders; 2010. Murray J, Nadel J, eds; No. 2.

180. Bradsher RW. Histoplasmosis and Blastomycosis. *Clin Inf Dis.* 1996;22:S102-S111.

181. Hahm C, Park H, Jeon K, et al. Solitary Pulmonary Nodules Caused by Mycobacterium tuberculosis and Mycobacterium avium Complex. *Lung.*188(1):25-31.

182. Lam WK, White NW, Chan-Yeung MM. Lung cancer epidemiology and risk factors in Asia and Africa State of the Art. *Int J Tuberc Lung Dis.* 2004;8:1045-1057.

183. Low S-Y, Eng P, Keng GHW, Ng DCE. Positron emission tomography with CT in the evaluation of non-small cell lung cancer in populations with a high prevalence of tuberculosis. *Respirology.* 2006;11(1):84-89.

184. Weber A, Stammberger U, Inci I, Schmid RA, Dutly A, Weder W. Thoracoscopic lobectomy for benign disease--a single centre study on 64 cases. *Eur J Cardiothorac Surg.* Sep 2001;20(3):443-448.

185. Feinstein AR. *Clinical Epidemiology: The architecture of clinical research.* W. B Saurers Co; 1985.

186. Kuo E, Bharat A, Bontumasi N, et al. Impact of Video-Assisted Thoracoscopic Surgery on Benign Resections for Solitary Pulmonary Nodules. *Ann Thorac Surg.* 2012;93(1):266-273.

TABLES Chapter 1

Table 1. Screening Criteria Risk Status – National Comprehensive Cancer Network		
High Risk	**Moderate Risk**	**Low Risk**
Age 55 -74	Age ≥ 50	Age < 50
≥ 30 pack year history	≥ 20 pack year history OR second-hand smoke exposure	< 20 pack year history
Current smoker OR Smoking cessation ≤ 15 years	No additional risk factors	
OR[a] Age ≥ 50		
≥ 20 pack year history AND		
One additional risk factor (other than second-hand smoke)[b]		

a NCCN criteria category 2B level of evidence.
b Risks include - Radon exposure, asbestos exposure, occupational exposure, cancer history, family history, other lung disease (COPD or pulmonary fibrosis).

Table 2: Comparison of existing predictive lung cancer models							
	Used for Screening				Used for surgical referral after discovery of a lung lesion		
	Bach*	Spitz	Liverpool	$PLCO_{2012}$	Mayo	VA	SPN
Age	√	m	m	√	√	√	√
Race		m		√			
Sex	√	m	m				
Smoking Status Ever/never					√	√	
Years since quitting smoking	√	√		√		√	
Pack-Years		√	√				√
Duration of smoking	√			√			
Cigarettes per day	√			√			
Family History of lung cancer		√	√	√			
Education				√			
BMI				√			
Emphysema/COPD		√	√	√			
Environmental exposures**	√	√	√				
Nodule Size (max diameter)					√	√	√
Nodule lung location					√		√
Nodule shape					√		√
Nodule growth							√
Previous cancer history (not lung)			√	√	√		√
Contrast CT scan positive							√
Hemoptysis							√
FDG-PET Avidity							√
Population Prevalence							√
ROC in validation and development	61-77	57-73	71	70-80	78-83	73-79	80

m = Variable matched in case-control study
*** = Predictive for smokers or ex-smokers only**
**** = Asbestos (Bach, Spitz, Liverpool), Dust (Spitz), and Hayfever (Spitz)**

FIGURES

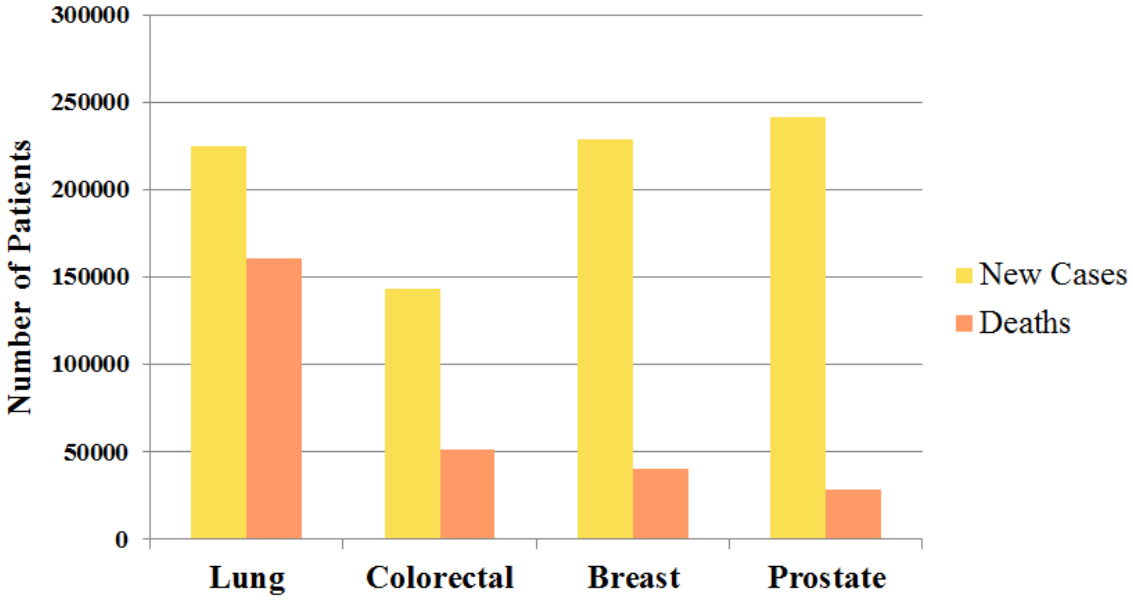


Figure 1. Annual Incidence and Mortality for the Most Common Cancers: 2012. Source: American Cancer Society, *Cancer Statistics 2012*

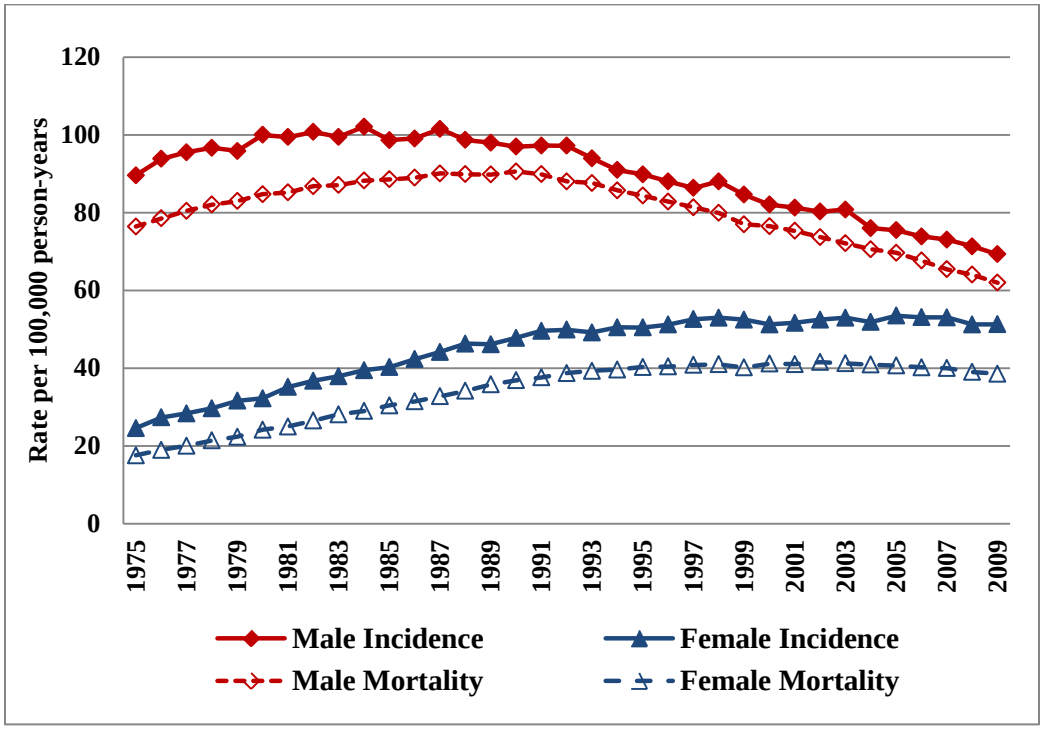


Figure 2.Age Adjusted U.S. Lung and Bronchus Cancer Incidence and Mortality Rates by Sex, 1975-2009. Source: SEER 2012

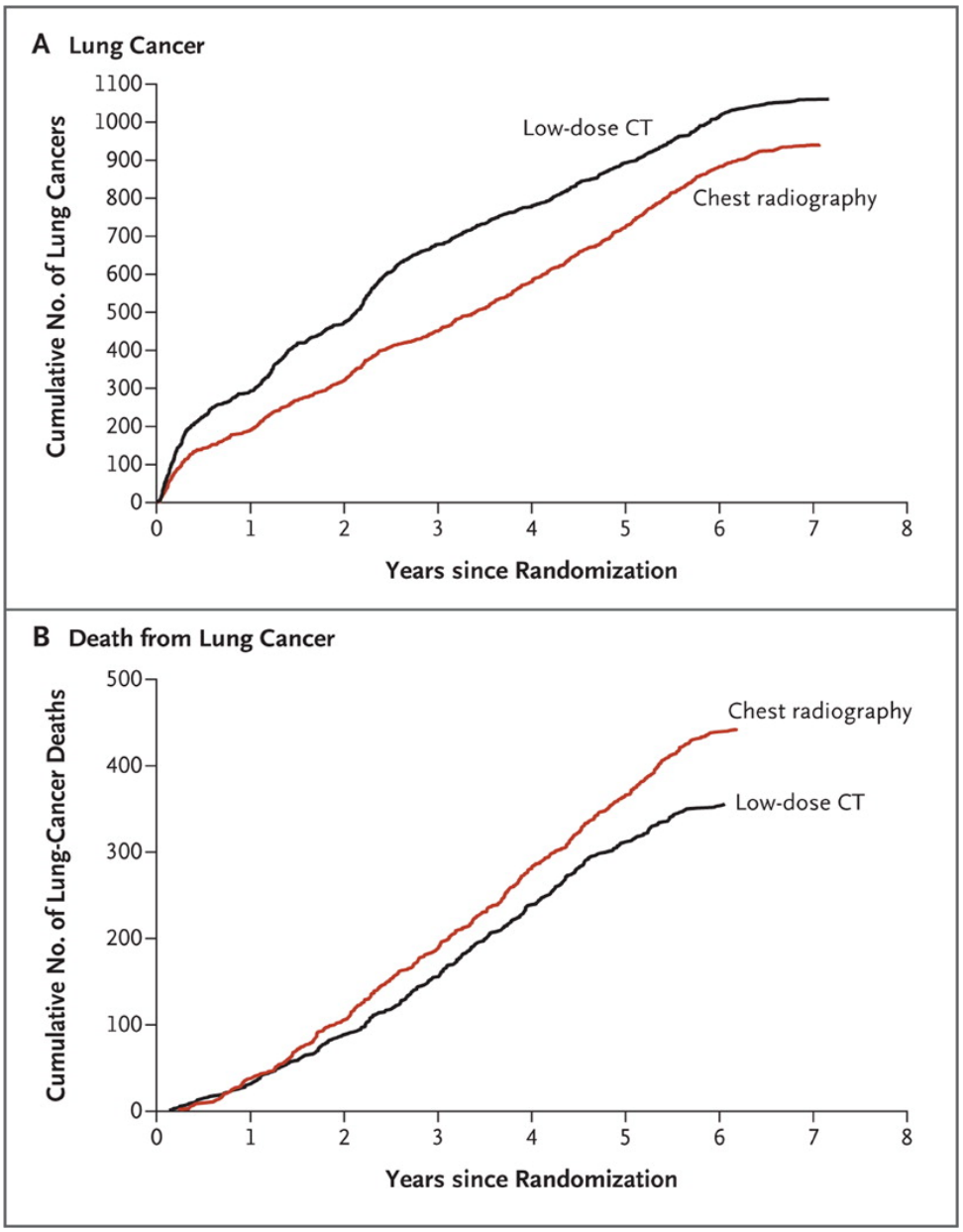
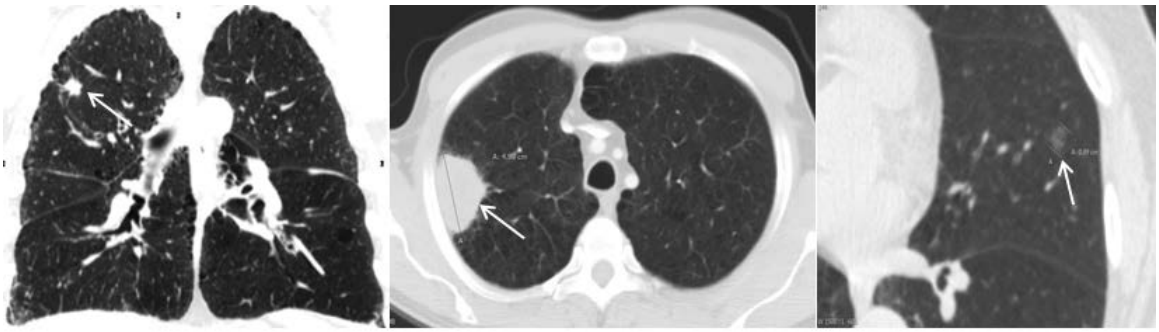


Figure 3. Results of the National Lung Screening Trial. Detection of lung Cancers by study arm (A) and Cumulative lung cancer related death by study arm (B). Source: NLST, 2012.



4A. Spiculated Lung nodule **4B. Spiculated Lung Mass** **4C. Ground Glass Opacity**

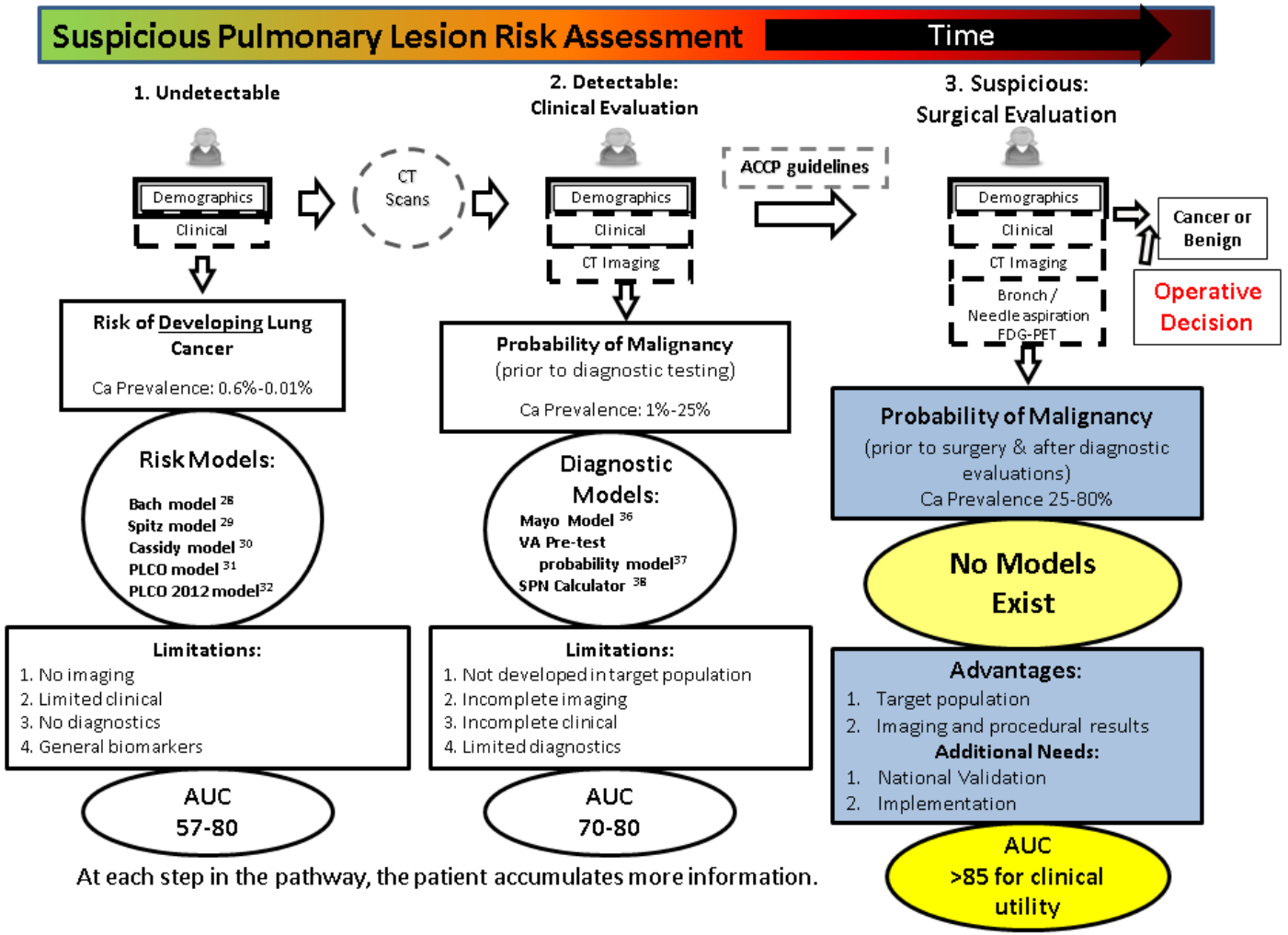


Figure 5. Conceptual Framework for the detection and diagnosis of lung cancer

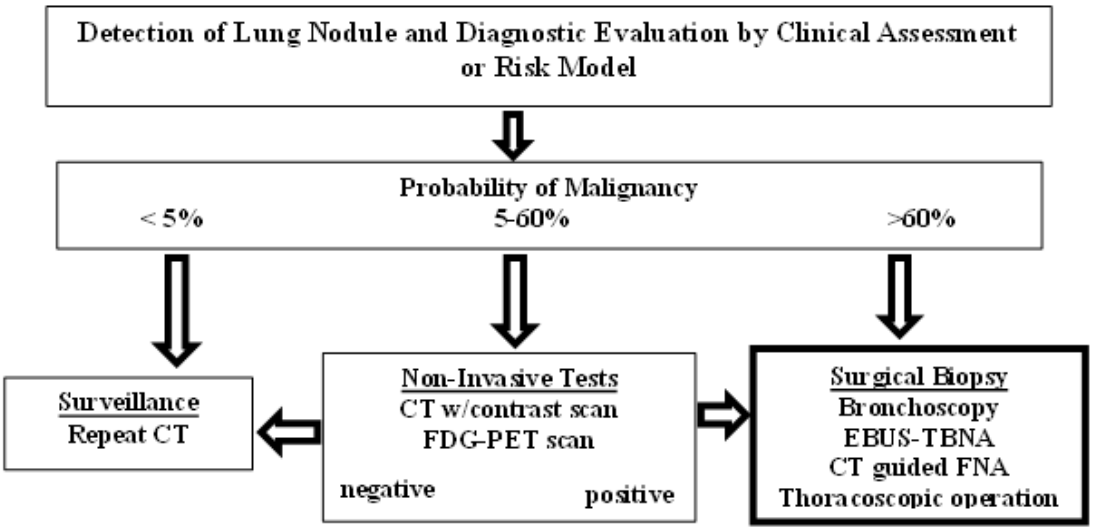
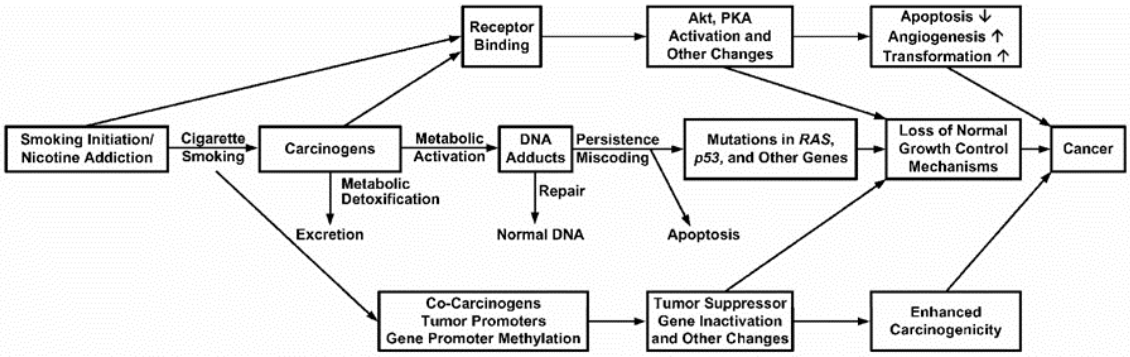


Figure 6. Diagnostic guidelines of lung nodules suspicious for lung cancer. Adapted from Wahidi, Chest 2007



Conceptual model for understanding mechanisms of tobacco carcinogenesis. The central track Conceptual model for understanding mechanisms of tobacco carcinogenesis. The central track involving carcinogen-DNA adduct formation and consequent mutations in critical genes is the major accepted pathway. The top and bottom tracks also contribute but their roles are less well defined.

Figure 7. Mechanistic framework for understanding how cigarette smoking causes lung cancer. Source: Hecht, Int J Cancer 2012.

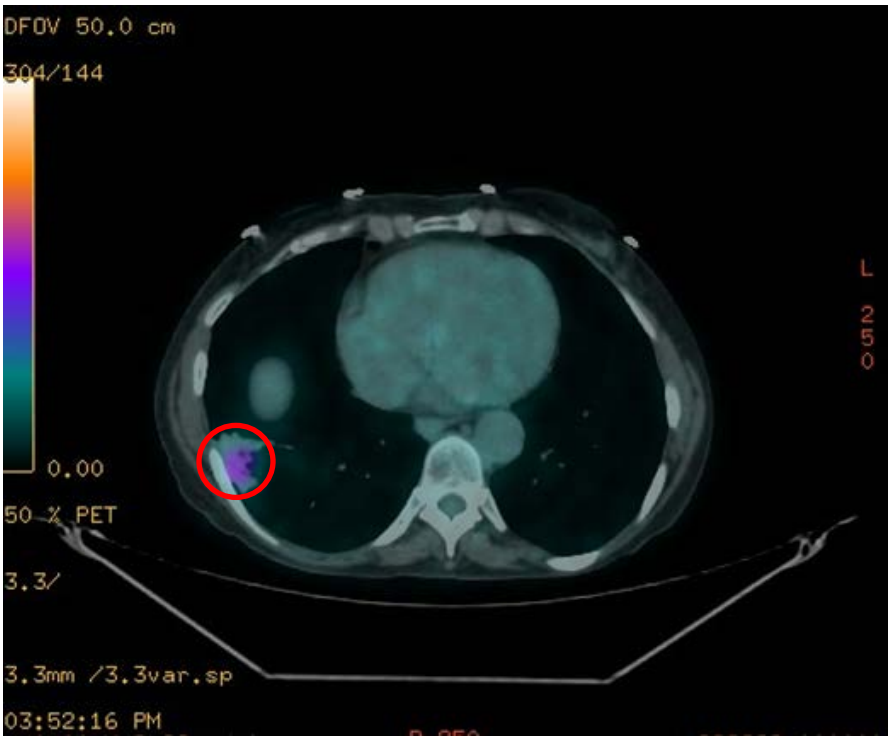


Figure 8. Avid FDG-PET/CT scan of a granuloma

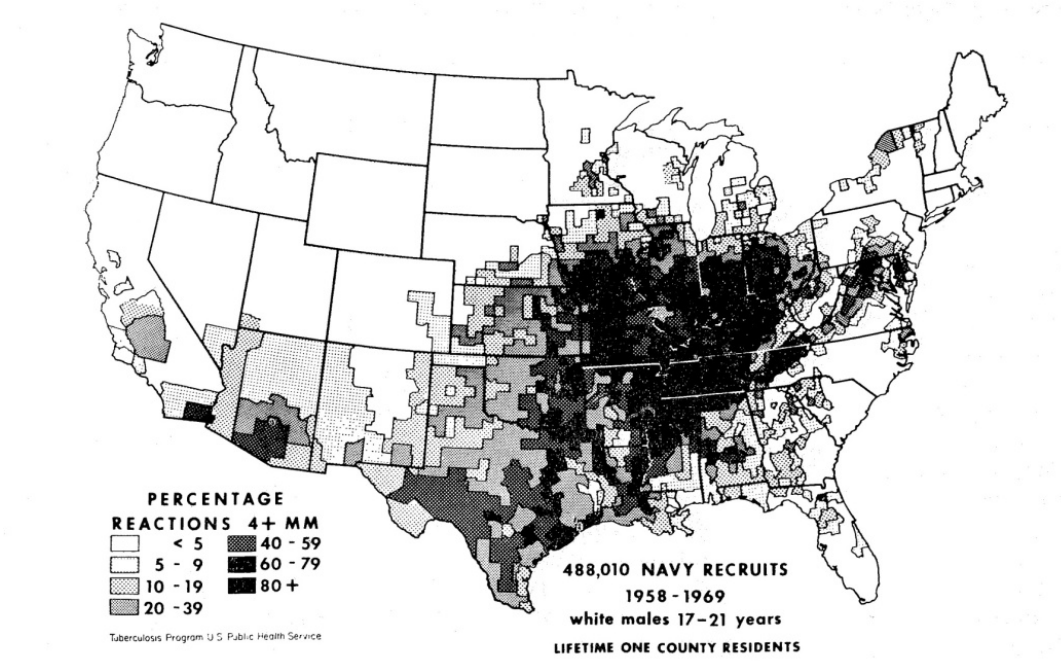


Figure 9A. Fungal lung disease prevalence, histoplasmin skin was cross reactive with coccidioidomycosis and blastomycosis (from Edwards et al, Am Rev Resp Dis 1969)

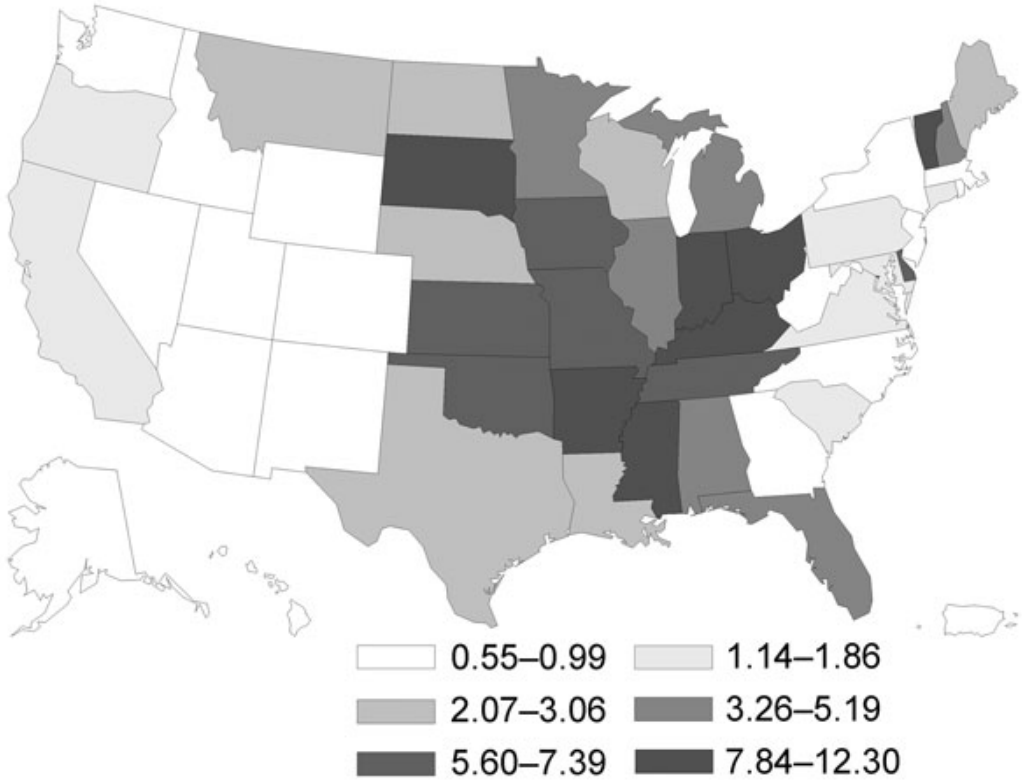


Figure 9B. Geographic distribution of histoplasmosis in persons >65 years of age, United States, 1999-2008. Values are number of cases/100,000 person years (from Baddley et al, Emerg Inf Dis 2011).

Chapter 2

II. Development and validation of a clinical prediction model to estimate lung cancer risk for those being evaluated for surgery

2.1 Introduction

Lung cancer is the leading cause of cancer-related mortality in the U.S. among men and women.[1,2] Poor prognosis for lung cancer remains common despite the steady decline in incidence of lung cancer likely due to , declining smoking rates, increased awareness in the general population, and the advent of new technologies to detect lung cancer in early stages of the disease. The average five-year survival rate for lung cancer is 16%, due to the late stage of the disease at the time of diagnosis.[3] To decrease mortality from lung cancer beyond reducing smoking prevalence in the population, early detection and treatment are necessary. The National Lung Screening Trial (NLST) recently found a 20% reduction in deaths from lung cancer by screening a high risk population with low dose computed tomography (LDCT). However, 39% of patients screened with LDCT scans had at least one positive screening test (identification of suspicious nodule) requiring additional diagnostic evaluation.[4] In addition, after nodule discovery and radiographic surveillance, 24% of procedures diagnosed benign disease. Other studies describing surgery for known or suspected lung cancer report benign disease between 20% and 40%.[5-8]

Unlike biopsy for other cancers with a screening regimen, the lung is difficult to access, and lung biopsy has significant risks associated with the procedure. Reviews of outcome after lung surgery have found 1% to 4% mortality rates within 30 days of surgery and rates as high as 7% at 90

days.[9-12] The surgeon's dilemma is balancing the likelihood that a lung nodule is cancerous with the possible harm caused by a surgical biopsy. A national screening program with LDCT would likely result in a reduction of lung cancer deaths but an increase in unnecessary deaths from invasive diagnostic procedures unless preoperative non-invasive evaluation for lung cancer becomes more accurate than it is in current practice.

Diagnostic Evaluation of Patients with Lung Nodules

The surgeon evaluating a patient with a lung nodule is in a unique clinical position. Surgeons have a relatively large body of diagnostic information often compiled from a variety of other specialists. Tests to determine whether the patient can survive an operation that may culminate in a lobectomy also influence the decision process. The American College of Chest Physicians and the National Cancer Comprehensive Network practice guidelines detail an evidence-based algorithm for the diagnosis of lung cancer.[13,14] The guidelines require clinicians to estimate the likelihood of lung cancer using their clinical expertise or a validated prediction model. If the probability of lung cancer is greater than 60%, then patients are referred to a surgeon for further evaluation and diagnostic resection. If the patient's probability of cancer is less than 5% then continued periodic surveillance with LDCT is suggested. If the patient has an intermediate probability of cancer, between 5% and 60%, then additional testing for lung cancer is recommended. Testing may include F^{18}-fluorodeoxyglucose positron emission tomography (FDG-PET) or minimally invasive biopsy like bronchoscopy, endobronchial ultrasound, or fine needle aspiration.[15,16]

The Mayo model[17], the VA model[18] and the Solitary Pulmonary Nodule (SPN) model[19,20] have been externally validated for the prediction of lung cancer. Each of the three models uses epidemiological risk factors, such as age and smoking history, or symptomatic indications like hemoptysis, as well as imaging data including lesion size, growth, location and edge

characteristics (Table 1). These three models were developed in populations with lower prevalence of lung cancer than is observed by surgeons evaluating lung nodules suspicious for lung cancer. A clinical prediction model in a population whose prevalence of disease is 50% to 80% and includes FDG-PET scans has not yet been published and current predictive models perform poorly in this population.[20] Surgeons need a clinical model to help avoid the significant mortality and morbidity that accompanies thoracoscopic surgery for diagnosis of lung cancer without missing the lung cancer itself.[21]

2.2 Methods

Study Population

The Vanderbilt Lung Cancer Cohort was composed of two sources of individuals: 1) using Vanderbilt University Medical Center's Thoracic Surgery Quality Improvement database, 453 patients were identified who underwent a surgical procedure for known or suspected non-small cell lung cancer (NSCLC) from January 2005 to October 2010. Demographic and clinical data for each surgery was abstracted using the Society of Thoracic Surgeons National Database for General Thoracic Surgery specifications and guidelines.[22] Each lesion was confirmed by pathologic examination after thoracotomy, thoracoscopy, mediastinoscopy, or bronchoscopy with biopsy; and 2) a separate group of 39 individuals who had been evaluated by a thoracic surgeon for possible cancer but who underwent radiographic surveillance rather than surgery was included as well. Individuals with known metastatic disease, individuals without a definitive clinical diagnosis after surgery, or those who underwent re-operation on a known malignancy were excluded. Non-surgical patients with less than 18 months of radiographic surveillance or clinical follow-up were excluded. Patients with no physician or radiological report of maximum pre-operative nodule size were also excluded. Vanderbilt University Institutional Review Board approved this study (IRB# 090781).

Radiographic data was abstracted by medical reviewers with experience from previous studies using radiologist reports of the most recent pre-operative CT scans for lesion growth, edge characteristics, and FDG-PET avidity.[5,6,20,23] Lesion edge characteristics were defined by the terms smooth, lobulated, lobular, lobed, GGO, GGN, coronal, corona, spiky, or spiculated in the radiologists' reports and designated by medical reviewers as either smooth, lobulated, ground glass opacity, spiculated or indeterminate. Growth on serial radiographs occurring at least 60 days apart is defined as an increase in mean diameter of 2mm for nodules initially less than 15mm in size and an increase of at least 15% compared to a baseline scan for nodules more than 15mm in size at baseline.[13] For cases with one preoperative radiograph or whose subsequent radiograph was fewer than 60 days and deemed too short a time span to record lesion growth, the case was designated as "insufficient data." FDG-PET scan results were categorized into four groups based on copies of the original physician reports or original scans. Categorization was determined by either physician report or by maximum standard uptake value (SUV). The categories of avidity and their corresponding SUVs are: Not avid/Not cancerous (SUV=0), Low avidity/Not likely cancerous (SUV 0.1 to 2.5), Avid/Likely cancerous (SUV 2.5 to 5) and Highly avid/Cancerous (SUV>5). Any radiological reports of insufficient quality to determine diagnosis, shape characteristics, or FDG-PET avidity by chart review were reviewed for determination by a thoracic surgeon. If no designation could be made, then original scans were reviewed by a thoracic radiologist blinded to clinical pretest data and pathological outcome. Pre-operative symptoms were defined as any documented evidence in the medical record of the following: shortness of breath, unplanned weight loss, pneumothorax, fatigue, pain, COPD, chronic bronchitis, pneumonia or bronchiolitis. Predicted forced exhaled volume in one second (FEV_1) was a continuous variable based on the most recent pulmonary function test prior to surgery.

Statistical analysis

A multivariable logistic regression model, hereafter referred to as the TREATmodel, was developed to predict lung cancer, following the methodology of Harrell and Steyerberg.[24,25] Missing data was imputed using multiple imputation with chained equations. Multinomial logistic regression was used to impute categorical variables, and predictive mean matching was the method of imputation for all other variables. Missing data was analyzed by variable to determine the percentage missing of each variable, correlation between variables with missing data, and whether patterns of missing data suggest violation of the assumption that data are missing at random (MAR).[26] A 10 imputed dataset burn-in was used and an additional 50 imputed datasets were then generated. These fifty datasets were used for all subsequent model estimation. After imputation, variable distributions were visually examined. Imputed data outside the range of observed data was limited to the maximum or minimum observed value. A sensitivity analysis was conducted comparing the estimated model using imputation results with chained equations to the same models using a multivariate normal imputation model that solely used predictive mean matching. A separate sensitivity analysis compared the multiply imputed data with a model using complete data.

Model development was guided by examination of published models predicting lung cancer in a population with lung nodules or lesions. Nonlinear associations between continuous variables and lung cancer were evaluated using restricted cubic splines of 3 and 5 knots, and linearity was tested using the Wald statistic. The model's ability to discriminate between cancer and benign disease was evaluated by area under the receiver-operating-characteristic curve (AUC). Model calibration, the comparison of a model's predicted probabilities to observed probabilities, was assessed with Brier score. Graphical interpretation of the model used partial effect plot and nomogram of the model (Appendix 1.3 and 1.4).

The developed model was internally validated using the .632 bootstrap method to estimate optimism corrected measures of AUC and Brier score.[27] The .632 bootstrap method creates a new population by randomly sampling with replacement from the Vanderbilt Lung Cancer Cohort. The complete dataset was created with missing data being replaced by the mean value of imputed data across all 50 imputations. The model was then re-estimated for each bootstrap iteration and an iteration-specific AUC and Brier score created. This bootstrap process was repeated 500 times. Optimism corrected AUC and Brier score were estimated using the weighted equation from Efron.[27,28] Large decreases in AUC or increases in Brier score indicate high optimism in model estimation and possible overfitting or lack of a stable, valid model. Due to the limited sample size and the limited power available to detect interactions between variables, no interactions were included in the model. A separate sensitivity analysis was conducted to estimate maximum possible AUC from the available variables by estimating a model with all clinically feasible interactions between variables included in a purposely overfitted model.

AUC and Brier score for the Mayo model were estimated using published variable coefficients from the original article (MayoOriginal).[17] The Mayo model probability of a malignant pulmonary nodule is of the form:

$$= e^x/(1+e^x);$$

where x =−6.8272 + (0.0391*age) + (0.7917*smoking history) + (1.3388*previous non-thoracic cancer) + (0.1274*lesion size) + (1.0407*spiculated lesion edge) + (0.7838*upper lobe location). Unlike the originally estimated and validated Mayo model, all cases with previous cancer were included in the comparative analysis. In their original study, Swensen and colleagues only included those with extra-thoracic cancer who were treated more than five years previous to the current evaluation.[17]

Using the same variables as the original Mayo model cancer risk was predicted in the same population as the fitted TREAT model and variable coefficients re-estimated (MayoVUMC). A model with only non-imaging variables was estimated (EpiModel) and compared to a model composed of only imaging (ImageModel). Differences between model AUCs were measured using the binomial exact test. Graphic comparisons of poor model fit used Bland-Altman plots for differences in predicted likelihood of cancer between the two models.

Analysis of demographic variables and pre-specified predictors of lung cancer according to lung cancer status were conducted utilizing only observed data. The T-test was used to examine differences in continuous variables, the Wilcoxon signed-rank test for differences in medians, and the Fisher's exact test for categorical variables. Analysis for imputation, model estimation and internal validation was performed in R v3.0.1 and Stata v12(College Station, TX). Code is found in Appendix 1.

2.3 Results

Lung cancer prevalence was 72%. Diagnosis was determined pathologically in 453 (92%) and by radiographic follow-up in 39 (8%) individuals. Complete data was available for 264 individuals. Those with a cancer diagnosis were more likely to have complete data (58%) than those with benign disease (43%). Missing data occurred with FDG-PET scan (22%), growth on serial CT scans (13%), predicted FEV_1 (10%) and pre-operative disease symptoms (7%). The remaining variables of interest had less than 5% missing data. In univariate analysis, self-reported race, pre-operative symptoms and previous history of cancer were not significantly different between those with or without lung cancer. Lung cancer was inversely related to body mass index (BMI) and predicted FEV_1. Lung cancer was associated with sex, age, smoking, increasing lesion size, lesion

location in either upper lobe of the lung, lesion growth on serial radiographs, spiculated edge characteristics and FDG-PET avidity (Table 2).

The imputation model consisted of the outcome (cancer or benign disease), height as a component of BMI, and any previous history of smoking as a dichotomous variable. Independent variables used as predictors of lung cancer risk in the fitted TREAT model were used in the imputation model and included: age, sex, pack years, lesion size, lesion location, lesion growth, spiculated lesion edge characteristic, predicted FEV_1, BMI, previous history of cancer, any pre-operative symptoms and FDG-PET avidity. FDG-PET avidity had the highest variance inflation due to imputation (1.37) on the estimated model followed by lesion growth for the insufficient data category (1.25) and then by any pre-operative symptoms (1.24).

Comparison of lung cancer risk models

In the TREAT model lung cancer risk significantly increased with age, pre-operative lesion size, lesion growth, history of previous cancer and FDG-PET avidity (Table 3). Smoking intensity measured by pack years had a non-linear association with lung cancer in univariate analysis but not in multivariate analysis (p=0.1). The non-linear association between pack years and lung cancer risk was maintained in the TREAT model (Figure 2).

The AUC for the TREAT model was 0.89 (95% CI: 86 - 92) and Brier score was 0.12. Internal validation with .632 bootstrap estimated an optimism adjusted AUC of 0.87 and Brier score of 0.13. The MayoOriginal model, using published coefficients to estimate lung cancer risk, had an AUC of 0.80 (95% CI: 76 - 85) which was significantly less (P<0.001) than the AUC observed for the TREAT model. The Mayo model generally overestimated risk and its Brier score was 0.17, showing poorer calibration than the new model. The MayoVUMC model used the original Mayo model variables, re-estimated coefficient values based on the Vanderbilt lung cancer

population, and had lower AUC 0.83 (95%CI: 79 - 87, p<0.001) than the TREAT model and similar Brier score (0.12) (Figure 3). The overfitted TREAT model had slightly higher AUC (0.90; 95%CI: 86 - 92) and slightly lower Brier score (0.11) than the TREAT model without interaction terms.

In a separate analysis of classes of models, the lung cancer predictive model that included only epidemiologic data (age, smoking history, any pre-operative symptoms, predicted FEV_1, previous history of cancer, BMI, race, and sex) had an AUC of 0.77 (95% CI: 72 – 82). The model relying on imaging data only (lesion location, size, FDG-PET avidity, lesion edge characteristics, growth on serial radiographs) had an AUC of 0.84 (95% CI: 80 – 88) and was significantly higher than the model with only epidemiologic variables (p=0.02) (Figure 4).

Bland-Altman plots showed the Mayo model more frequently overestimated risk of lung cancer when compared to the TREAT model (Figure 5). No systematic differences in model risk estimation were observed between the TREAT and Mayo models for the covariates of age, smoking status, and pack years (Figure 6A, 6B and 6C). FDG-PET avidity and growth showed a systematic difference in predictive values between the two models. These two variables indicate metabolic activity of the lesion and were included only in the TREAT model. The Mayo model on average predicted a slightly higher risk of lung cancer compared to the TREAT model in individuals with no growth (Figure 6D) and a significantly higher risk among individuals with non-avid scans (Figure 6E).

2.4 Discussion

To date, to our knowledge, no clinical prediction model for lung cancer in a thoracic surgery population has been published. The TREAT lung cancer model found a high and consistent

predictive discrimination for lung cancer. The TREAT model used pre-specified variable selection based on covariates of lung cancer risk published in validated screening and lung lesion diagnostic models. Given the TREAT model's high level of discrimination, it may be of value providing clinical guidance in estimating individual likelihood of lung cancer if externally validated.

Although the TREAT model is relatively simple and includes no interaction terms and only one non-linear variable, a more complex over-fitted model found negligible improvement in AUC (0.90) compared to the apparent TREAT model AUC of 0.89. This small increase in discriminatory power indicates the TREAT model explains well the relationship between the variables in the model and the likelihood of lung cancer as currently specified. Little additional information is likely to be gleaned from the variables used in the TREAT model.

Surgeons evaluating pulmonary nodules are faced with a basic question of equipoise. Is the risk for cancer such that the individual should undergo the possible harms arising from thoracoscopic surgery to determine diagnosis of the nodule? Diagnostic lung surgery has a 1-4% mortality rate associated with the procedure and 1.2% of those with benign disease in the NLST died within 60 days of their procedure.[9,29] Complications including prolonged air leak (5%) and atrial fibulation (11%) are also common from the procedure and recovery takes 4-6 weeks among non-complicated patients.[21,29,30] If the patient under evaluation has marginal lung function and other pre-operative comorbidities, then the likelihood of a poor outcome increases. This procedural risk is juxtaposed against the danger of missing a curable lung cancer. One suggested solution is for the clinician to delay biopsy and treatment until a more definitive non-invasive diagnosis is possible. The window of best prognosis from the time of lesion discovery to stage progression and metastasis is not well known; however, one study found that untreated clinical stage 1 lung

cancer had a median survival time of only 9 months.[31] Thus clinical belief is that even among small and localized cancer, treatment should not be delayed if possible.

The superior performance of the TREAT lung cancer model (AUC 0.89) compared to the Mayo model in its original published form (MayoOriginal AUC 0.80) or re-estimated within the Vanderbilt population (MayoVUMC AUC 0.83) arises from two factors. First, the population being evaluated by surgeons has a higher prevalence of lung cancer than other lung nodule evaluation populations. Current predictive models for lung nodules were established in populations being evaluated by pulmonologists (Mayo model) and radiologists (SPN model). A finding of risk from these models results in a referral to a proceduralist to establish tissue diagnosis and course of treatment. The performance of the Mayo model was similar to the AUC observed in the original study (0.83) and in the validation study in a population of veterans (0.80).[17,18] The Mayo model performed well in the Vanderbilt population although the prevalence of disease was higher (72%) than that found in the original population in which the model was developed (23%) or validated (44%).[17-19] However, the Mayo and SPN models underestimate risk for cancer in the lower quintiles of lung cancer risk.[20] The higher prevalence of lung cancer is not uncommon in surgical populations being evaluated for lung cancer,[6,32,33] which was one of the motivating factors for development of the TREAT model specifically for surgeons.

The second factor contributing to the TREAT model's superior discrimination is the addition of FDG-PET avidity and lesion growth as predictive variables. Each variable represents metabolic activity of the lesion which is information relatively independent from other imaging variables like lesion shape or size. The TREAT model takes advantage of the additional information available to clinicians at the time of decision to operate. The addition of FDG-PET, lesion growth, predicted FEV_1 and presentation with any symptoms improved discrimination between benign disease and lung cancer. Each variable contributed to the variance explained, and FDG-

PET avidity and lesion growth were the third and fourth strongest explanatory variables behind lesion size and age (Figure 7).

The Bland-Altman plots reinforced the importance of FDG-PET avidity and lesion growth among the factors that discriminate a higher likelihood of lung cancer. When a lesion was not FDG-PET avid or did not have growth on serial radiographs, then the Mayo model had a consistently higher predicted probability of lung cancer and was associated with the largest differences in predicted risk between the Mayo and TREAT models. The distribution of this systematic bias was concentrated in the lower spectrum of risk (Figures 6D and 6E). Variables common to both models like age and smoking status did not show a systematic bias in predicted risk for lung cancer between the two models (Figures 6A-6C). Measurement of metabolic activity through FDG-PET and lesion growth appear to be an important biological factor missing from the Mayo model.

As with most research this study has a number of weaknesses. The Vanderbilt Lung Cancer Cohort was a retrospective cohort which relied on chart review. Vanderbilt is a tertiary referral academic medical center and may not be representative of other surgical practices evaluating lung lesions. Other populations in other parts of the country may have differing prevalence of disease, differing referral patterns, radiologist expertise or other underlying factors, like fungal lung disease prevalence, that are relevant to a clinical prediction model but either not measured or not included in the TREAT model.

Another weakness with this study is the high amount of missing data for variables of interest in the Vanderbilt Lung Cancer Cohort. Missing FDG-PET scan results (22% missing) occurred most frequently overall. Among the 39 who were managed by active surveillance, four had a FDG-PET scan. Of the 66 missing an FDG-PET scan in the entire cohort, 35 (53%) were in the

active surveillance group and did not undergo surgery. Conversely, among those with cancer, individuals with a high likelihood of lung cancer also were more likely to have missing FDG-PET scans. The lack of FDG-PET scan among those at the highest and lowest risk for lung cancer reflects clinical practice in which tests are not ordered that will not materially change the clinical decision to operate. This pattern of work-up bias between individuals indicates that the data is not missing completely at random. Similarly, predicted FEV_1 is generally not performed unless surgery is likely and clinical guidelines suggest everyone receive a pulmonary function test, of which predicted FEV_1 is a component, prior to surgery. The observed pattern and cause of the missing data may be a violation of the MAR assumption required for multiple imputation to be robust in its representation of the underlying true population.

Patterns that suggest high correlation between seemingly unrelated variables like FEV_1 and FDG-PET scan may indicate an unknown underlying cause of missing data that is not captured in the observed data. Cluster analysis of the variables in the dataset showed moderate correlation (Appendix 1.1). The multiple imputation methodology depends on missing data being missing at random (MAR) or the more restrictive missing completely at random mechanism. The MAR mechanism of missing data assumes that "missingness" depends upon or is explained by the observed variables and observations within the dataset and not on unobserved variables or unobserved individuals outside the dataset. Inclusion of the 39 non-surgical patients increased the amount of missing data for FDG-PET and predicted FEV_1. However, their inclusion into the cohort captured a broader population of risk and better represented the entire population of lung nodules evaluated by surgeons. Logistic regression models were developed to predict the likelihood of a record's variable being missing. If the outcome (cancer) remained a significant variable after controlling for the other relevant variables, then the missing at random assumption is violated. For FDG-PET, lesion growth, and predicted FEV_1 the cancer outcome was not significantly associated with predicting the cancer a record would have missing data. Of the three

variables with missing data, cancer was most predictive of FEV_1 (p=0.11) of having missing data. Therefore, the mechanism of missing data was likely MAR. If the likelihood an individual will have missing data had been associated with the outcome of cancer then non-ignorable missing data occurs and the assumptions behind valid imputation were violated. Such was not the case here. Furthermore, this dataset is retrospective, includes all persons evaluated by a thoracic surgeon and reflects the spectrum of risk likely to be encountered by a thoracic surgeon which removed the problem of non-ignorable missing data due to selection bias.[34,35]

The FDG-PET scan and lesion growth results were primarily derived from the radiologist's report and most individual patient FDG-PET scans were not reviewed by a radiologist or surgeon for this analysis. Reliance on radiologist's report, while a limitation, also mirrors clinical practice where often the surgeon must rely on external expert interpretation if the original scans are not available or are of poor quality. Individuals missing information on lesion growth were reviewed by a thoracic radiologist. After review a third category was created representing an individual with a single study or someone with serial studies who did not have a large enough interval between scans to determine growth. Approximately 40 of the 65 with lung cancer and missing data from this variable arose from the lack of stored scans in the radiology system. These missing scans were more likely to occur in individuals diagnosed between 2005 and 2007.

Another limitation of this study was that family history of cancer was not collected in the Vanderbilt Lung Cancer Cohort. Screening models whose focus is on population-level lung cancer risk have included first degree family member history of cancer in their cancer risk predictions.[36,37] Tammemagi et al. found that only among individuals with significant tobacco exposure was family history of cancer a risk factor.[37] It is unknown whether family history adds new information to the existing cancer risk model, or to the TREAT model, since models that

include family history do not include radiographic predictive factors and the existing epidemiologic variables explain much of the observed variation with an AUC of 0.77.

The Vanderbilt Lung Cancer Cohort is composed of individuals with known or suspected lung cancer. Ideally, a population composed of only suspected lung cancer would be used. Those with known lung cancer tend to have larger, growing, FDG-PET avid lesions. These cancers are more easily biopsied or diagnosed through sputum cytology. Inclusion of these individuals may bias our point estimates and standard errors for lesion size, FDG-PET avidity and growth. Another weakness of this analysis is the lack of an external validation dataset in which to compare the robustness of our results. Although little optimism was observed in the bootstrap adjusted AUC or Brier scores, the model must be validated in separate, external populations prior to use in the clinical setting.

This study is unique in a number of ways which add strength and import to its results. This is the first study to predict lung cancer in a cohort with a lung nodule being evaluated by thoracic surgeons. The TREAT model includes a combination of strong predictors of lung cancer including FDG-PET avidity and lesion growth that have not been previously published. The high AUC and optimism adjusted AUC indicate a model that discriminates lung cancer better than other published models. If the TREAT model is externally validated then it can act as the base model of lung cancer risk when evaluating new biomarkers and diagnostic tests for lung cancer in surgical populations. Future work will validate this model in external datasets and prospectively evaluate the impact of the model in the clinical setting.

2.5 Conclusion

In a population with a radiographically confirmed lung lesion being evaluated for possible surgery, the TREAT lung cancer model predicted the risk for lung cancer with high accuracy. The model was internally validated and showed little optimism in its estimate accuracy to discriminate between lung cancer and benign disease. The TREAT model incorporates the full spectrum of epidemiological and radiographic evidence and better predicts lung cancer with a higher AUC than existing published models. This model should be validated in external datasets and if valid applied in a prospective study to reduce unnecessary lung surgeries.

2.6 References

1. Siegel R, Ward E, Brawley O, Jemal A. Cancer statistics, 2011: The impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA Cancer J Clin.* Jul-Aug 2011;61(4):212-236.

2. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin.* Sep-Oct 2010;60(5):277-300.

3. Howlander N, Noone A, Krapcho M. SEER Cancer Statistics Review, 1975-2008. In: Institute NC, ed. Bethesda, MD2012.

4. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *N Engl J Med.* 2011;365(5):395-409.

5. Deppen S, Putnam JB, Jr., Andrade G, et al. Accuracy of FDG-PET to diagnose lung cancer in a region of endemic granulomatous disease. *Ann Thorac Surg.* Aug 2011;92(2):428-432; discussion 433.

6. Grogan E, Deppen SA, Ballman K, et al. Accuracy of FDG-PET to diagnose lung cancer in the ACOSOG Z4031 trial. Paper presented at: Alliance for Clinical Trials in Oncology; June 29-July1, 2012, 2012; Chicago, IL

7. Kozower BD, Meyers BF, Reed CE, Jones DR, Decker PA, Putnam Jr JB. Does Positron Emission Tomography Prevent Nontherapeutic Pulmonary Resections for Clinical Stage IA Lung Cancer? *Ann Thorac Surg.* 2008;85(4):1166-1170.

8. Veronesi G, Bellomi M, Scanagatta P, et al. Difficulties encountered managing nodules detected during a computed tomography lung cancer screening program. *J Thorac Cardiovasc Surg.* Sep 2008;136(3):611-617.

9. Grogan EL, Stukenborg GJ, Nagji AS, et al. Radiotracer-guided thoracoscopic resection is a cost-effective technique for the evaluation of subcentimeter pulmonary nodules. *Ann Thorac Surg.* Sep 2008;86(3):934-940; discussion 934-940.

10. Berry MF, Hanna J, Tong BC, et al. Risk factors for morbidity after lobectomy for lung cancer in elderly patients. *Ann Thorac Surg.* Oct 2009;88(4):1093-1099.

11. von Meyenfeldt EM, Gooiker GA, van Gijn W, et al. The Relationship Between Volume or Surgeon Specialty and Outcome in the Surgical Treatment of Lung Cancer: A Systematic Review and Meta-Analysis. *J Thorac Oncol.* 2012;7(7):1170-1178

12. Cheung M, Hamilton K, Sherman R, et al. Impact of Teaching Facility Status and High-Volume Centers on Outcomes for Lung Cancer Resection: An Examination of 13,469 Surgical Patients. *Ann Surg Oncol.* 2009;16(1):3-13.

13. National Comprehensive Cancer Network. Lung Cancer Screening v1.2012. *NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines).* 2011.

http://www.nccn.org/professionals/physician_gls/recently_updated.asp. Accessed 10-27-2011.

14. Gould MK, Fletcher J, Iannettoni MD, et al. Evaluation of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest.* Sep 2007;132(3 Suppl):108S-130S.

15. Wahidi MM, Govert JA, Goudar RK, Gould MK, McCrory DC. Evidence for the treatment of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest.* Sep 2007;132(3 Suppl):94S-107S.

16. Rivera MP, Detterbeck F, Mehta AC. Diagnosis of lung cancer: the guidelines. *Chest.* Jan 2003;123(1 Suppl):129S-136S.

17. Swensen S, Silverstein M, Ilstrup D, Schleck C, Edell E. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med.* 1997;157:849 - 855.

18. Schultz EM, Sanders GD, Trotter PR, et al. Validation of two models to estimate the probability of malignancy in patients with solitary pulmonary nodules. *Thorax.* Apr 2008;63(4):335-341.

19. Gurney JW, Lyddon DM, McKay JA. Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part II. Application. *Radiology.* February 1993 1993;186(2):415-422.

20. Isbell JM, Deppen S, Putnam JB, Jr., et al. Existing general population models inaccurately predict lung cancer risk in patients referred for surgical evaluation. *Ann Thorac Surg.* Jan 2011;91(1):227-233; discussion 233.

21. Grogan EL, Jones DR. VATS lobectomy is better than open thoracotomy: what is the evidence for short-term outcomes? *Thorac Surg Clin.* Aug 2008;18(3):249-258.

22. Society of thoracic surgeons. General Thoracic Surgery Database. 2013; http://www.sts.org/sts-national-database/database-managers/general-thoracic-surgery-database. Accessed 07/21/2013.

23. Grogan EL, Weinstein JJ, Deppen SA, et al. Thoracic Operations for Pulmonary Nodules Are Frequently Not Futile in Patients with Benign Disease. *J Thorac Oncol.* Oct 2011;6(10):1720-1725.

24. Harrell FEJ. *Regression Modeling Strategies.* Charlottesville, VA: Springer; 2001.

25. Steyerberg EW. *Clinical Prediction Models.* 1st ed: Springer; 2009.

26. Little R, Rubin D. Statistical Analysis with Missing Data, Second edition. 2002.

27. Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol.* Aug 2001;54(8):774-781.

28. Efron B, Tibshirani R. Improvements on Cross-Validation: The .632+ Bootstrap Method. *J Am Stat Assoc.* 1997;92(438):548-560.

29. Kozower BD, Sheng S, O'Brien SM, et al. STS database risk models: predictors of mortality and major morbidity for lung cancer resection. *Ann Thorac Surg.* Sep 2010;90(3):875-881; discussion 881-873.

30. Bradley CJ, Yabroff KR, Dahman B, Feuer EJ, Mariotto A, Brown ML. Productivity Costs of Cancer Mortality in the United States: 2000–2020. *J Natl Cancer Inst.* December 17, 2008 2008;100(24):1763-1770.

31. Raz DJ, Zell JA, Ou S-HI, Gandara DR, Anton-Culver H, Jablons DM. Natural History of Stage I Non-Small Cell Lung Cancer*. *Chest.* July 2007 2007;132(1):193-199.

32. Allen MS, Darling GE, Pechet TT, et al. Morbidity and mortality of major pulmonary resections in patients with early-stage lung cancer: initial results of the randomized, prospective ACOSOG Z0030 trial. *Ann Thorac Surg.* Mar 2006;81(3):1013-1019; discussion 1019-1020.

33. Smith MA, Battafarano RJ, Meyers BF, Zoole JB, Cooper JD, Patterson GA. Prevalence of Benign Disease in Patients Undergoing Resection for Suspected Lung Cancer. *Ann Thorac Surg.* 2006;81(5):1824-1829.

34. Rubin D, Schenker N. Multiple imputation in health-care databases: an overview and some applications. *Stat Med.* 1991;10(4):585 - 598.

35. Janssen KJ, Vergouwe Y, Donders AR, et al. Dealing with missing predictor values when applying clinical prediction models. *Clin Chem.* May 2009;55(5):994-1001.

36. Spitz MR, Etzel CJ, Dong Q, et al. An Expanded Risk Prediction Model for Lung Cancer. *Cancer Prev Res.* September 1, 2008 2008;1(4):250-254.

37. Tammemagi MCP, Freedman MTMD, Pinsky PFP, et al. Prediction of True Positive Lung Cancers in Individuals with Abnormal Suspicious Chest Radiographs-A Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial Study. *J Thorac Oncol.* 2009;4(6):710-721.

Table 1. Variables used in published, validated clinical lung cancer prediction models			
Variable	**Mayo Model**	**VA Model**	**SPN Model**
Age	√	√	√
Smoking (Y/N)	√	√	
Smoking - Pack Years			√
Years quit smoking		√	
Hemoptysis			√
Previous Cancer	√		
Lesion Size	√	√	√
Lesion Growth			√
Spiculation	√		√
Lesion Location	√		√
FDG-PET Avidity			√

Table 2. Univariate analysis of demographics and radiological data

	Cancer N=352	Benign N=140	P-value
Male (%)	192 (55)	60 (43)	0.02
Caucasian (%) Missing N=2	324 (92) 0	128 (93) 2	0.79
Age (SD)	65 (10.7)	55 (13.9)	<0.001
Smoking Status – Ever (%)	297 (84)	81 (58)	<0.001
Median Pack – years among smokers (IQR) Missing N = 8	45 (30, 60) 6	35 (20, 50) 2	0.001
BMI, kg/m^2 (SD) Missing N=2	26.9 (5.5) 0	29.6 (7.1) 2	<0.001
Mean Predicted FEV_1 (SD) Missing N = 50	75.4 (18.8) 17	84.5 (20.8) 33	<0.001
Pre-operative symptom[a] (%) Missing N = 33	82 (25.5) 30	36 (26.3) 3	0.85
Previous history of cancer	136 (38.6)	45 (32.1)	0.18
Upper lobe location Missing N = 13	214 (62.2) 8	69 (51.1) 5	0.03
Lesion Size, Ave mm (SD)	31.6 (19.4)	18.8 (12.0)	<0.001
Growth Missing N = 65	160 (55.4) 63	41 (29.7) 2	<0.001
Spiculation Missing N = 19	176 (52.1) 14	38 (28.2) 5	<0.001
FDG-PET Avidity[b] Missing N = 109	283 (86.2) 43	45 (60.1) 66	<0.001

a – Symptoms include: hemoptysis, unplanned weight loss, shortness of breath, fatigue or chest pain

b - Includes avidity categories “avid/likely cancerous – SUV 2.5-5” and “highly avid/cancerous SUV >5”

Table 3. Multivariate logistic regression of lung cancer prediction model

	N[a]	Odds Ratio (95% CI)	p-value
Age (per year)	492	1.05 (1.03, 1.08)	<0.001
BMI	490	0.97 (0.93, 1.02)	0.24
Gender - Male	492	0.94 (0.54, 1.64)	0.84
Pack Years[b]	484	1.03 (1.01, 1.05)	0.02
Pack Years'[b]	484	0.98 (0.95, 1.00)	0.09
Lesion Size (per mm)	492	1.06 (1.04, 1.08)	<0.001
Spiculated lesion edge	473	1.42 (0.78, 2.59)	0.26
Lesion location – upper lobe	479	1.00 (0.57, 1.75)	0.99
Lesion Growth	427		
No lesion growth		(ref)	(ref)
Insufficient data on growth		1.25 (0.56, 2.79)	0.59
Growth observed		2.97 (1.45, 6.12)	0.003
History of previous cancer	492	1.95 (1.07, 3.55)	0.03
Predicted FEV_1	442	0.99 (0.97, 1.00)	0.07
Any pre-operative symptoms	459	0.63 (0.33, 1.19)	0.16
FDG-PET Avid	382	6.81 (3.04, 15.3)	<0.001

a reported N is number of individuals with complete data in the dataset

b Pack-years was modeled with a non-linear relationship with cancer. Odds ratios are not directly interpretable and are included here for reporting purposes only. See Figure 2 for graph of relationship.

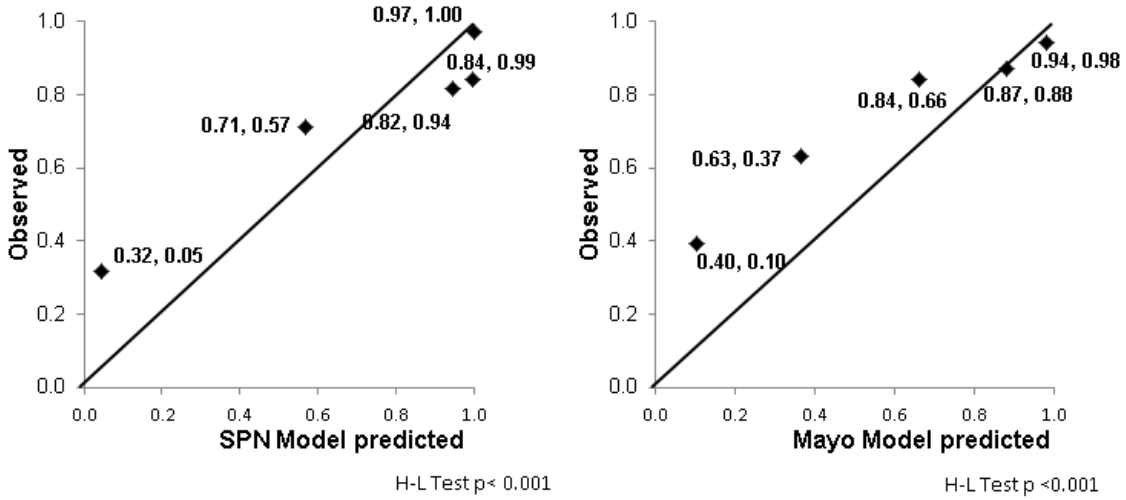


Figure 1. Hosmer-Lemeshow test showing poor (p<0.001) calibration of SPN and Mayo models in a surgical population (from Isbell et al, Annals of Thoracic Surgery, 2011)

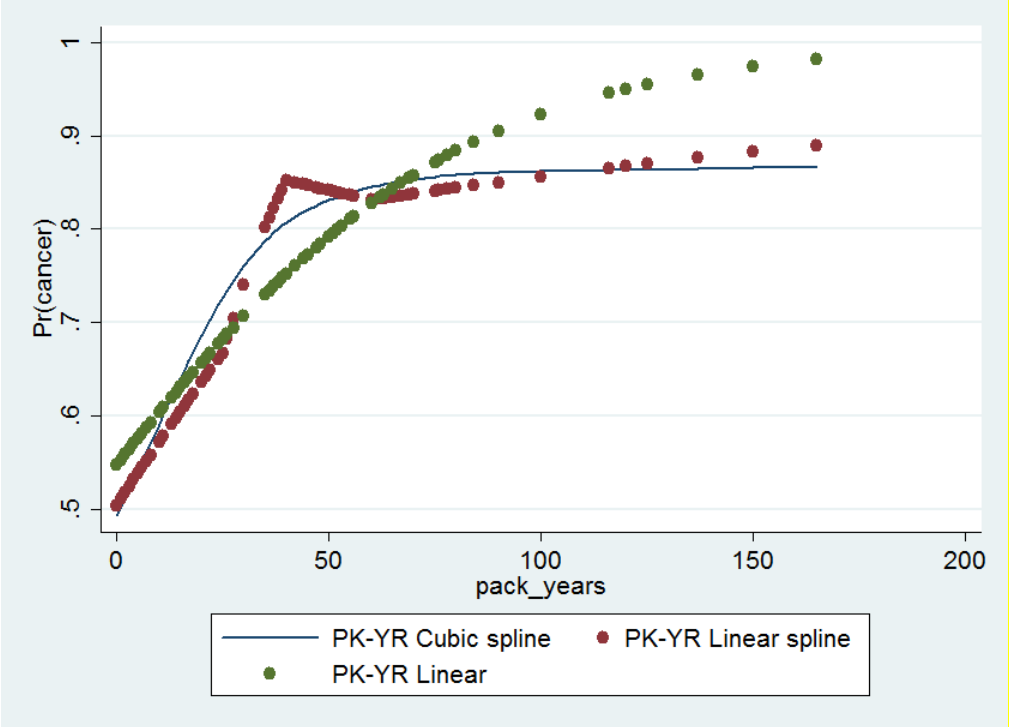


Figure 2. Comparison of three relationships between pack years and lung cancer.

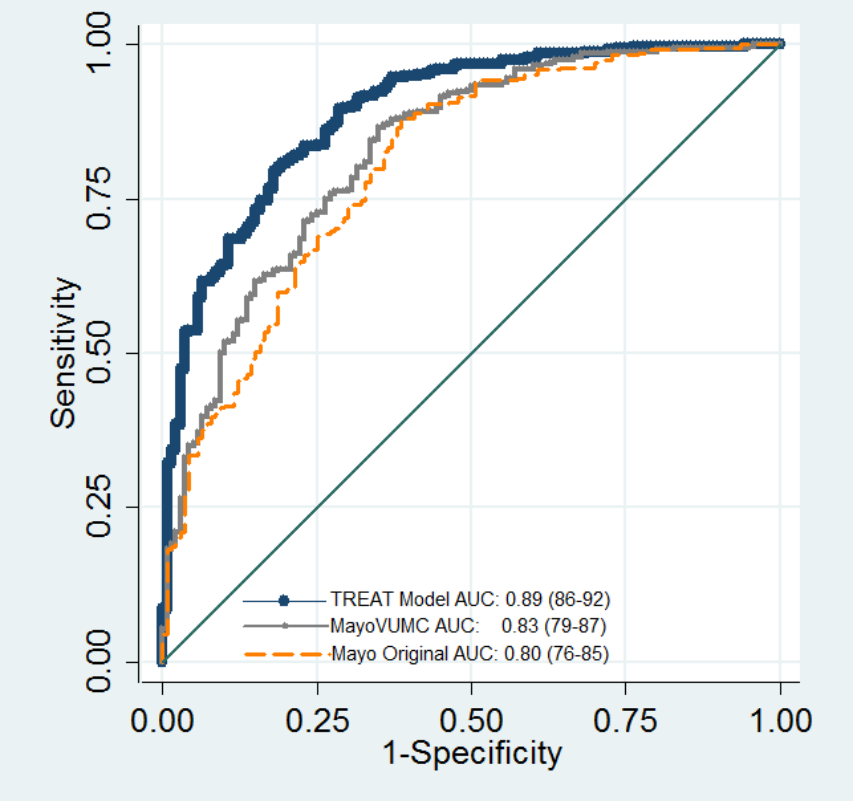


Figure 3. Comparison of AUC for two Mayo models and TREAT model

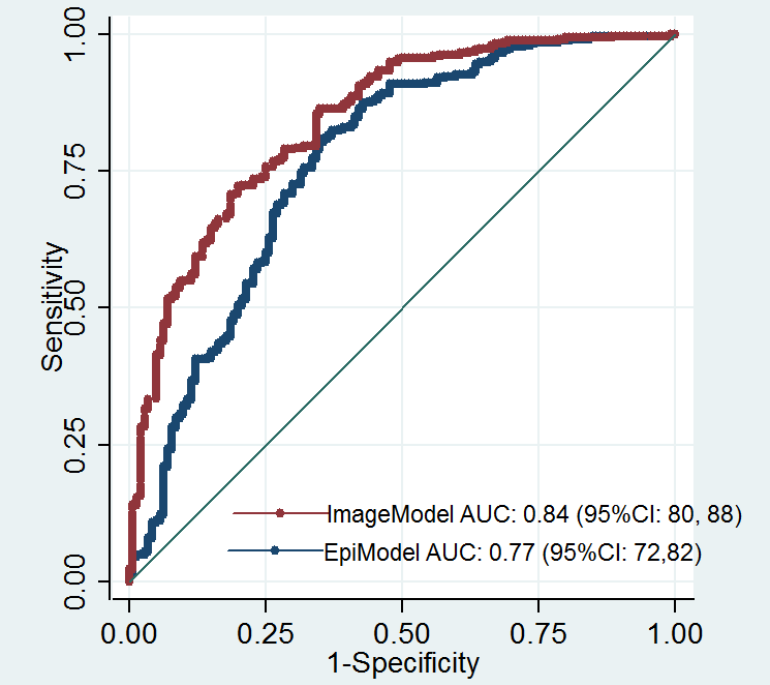


Figure 4. Comparison of AUC for a model composed of only imaging data (ImageModel) and epidemiological data (EpiModel) (p=0.02)

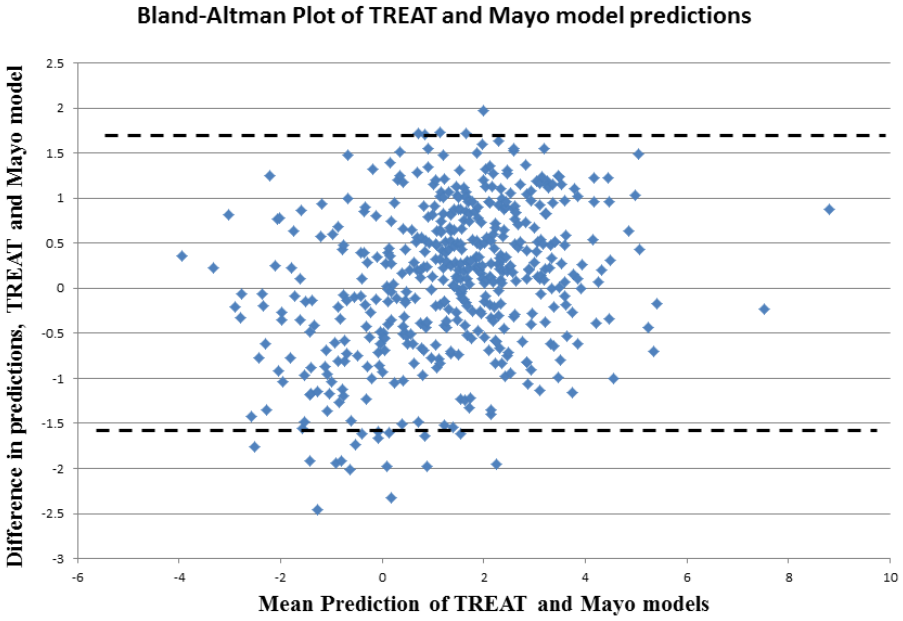


Figure 5. Bland – Altman plot of log(odds) predicted risk for the TREAT and Mayo model. X axis is the log odds of lung cancer for TREAT model plus the log odds of lung cancer in the mayo model divided by 2. Y axis is the log odds of lung cancer for TREAT model minus the log odds of lung cancer in the mayo model.

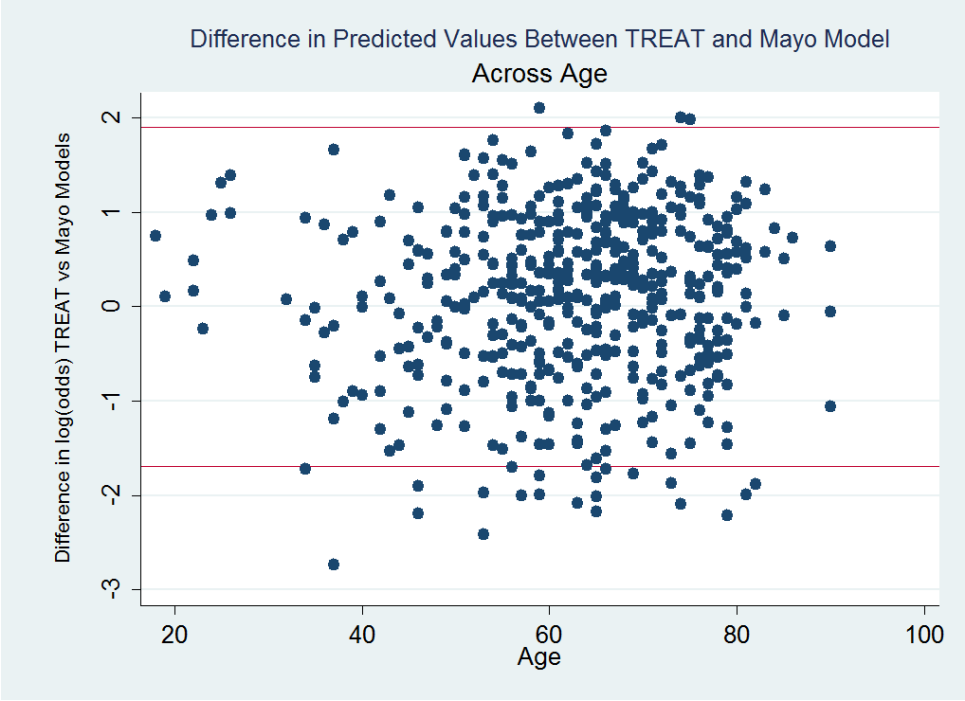


Figure 6A. Bland – Altman plot of the difference in the log odds of the TREAT and Mayo models across participants age. Red lines are the upper and lower confidence interval +/- 2 standard deviations.

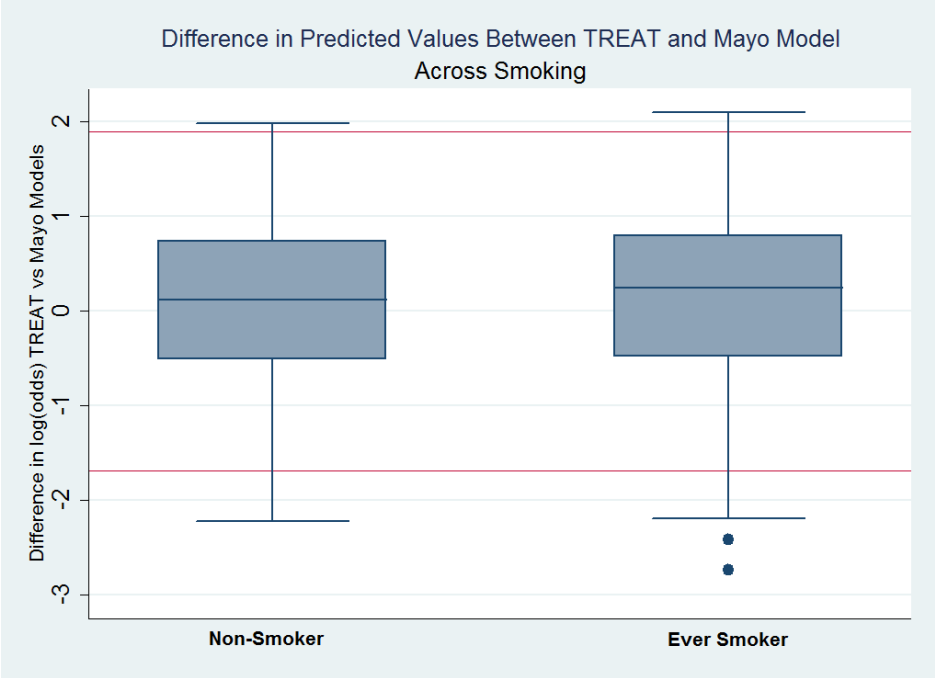


Figure 6B. Box plot of the difference in the log odds of the TREAT and Mayo models across by smoking history. Red lines are the upper and lower confidence interval +/- 2 standard deviations.

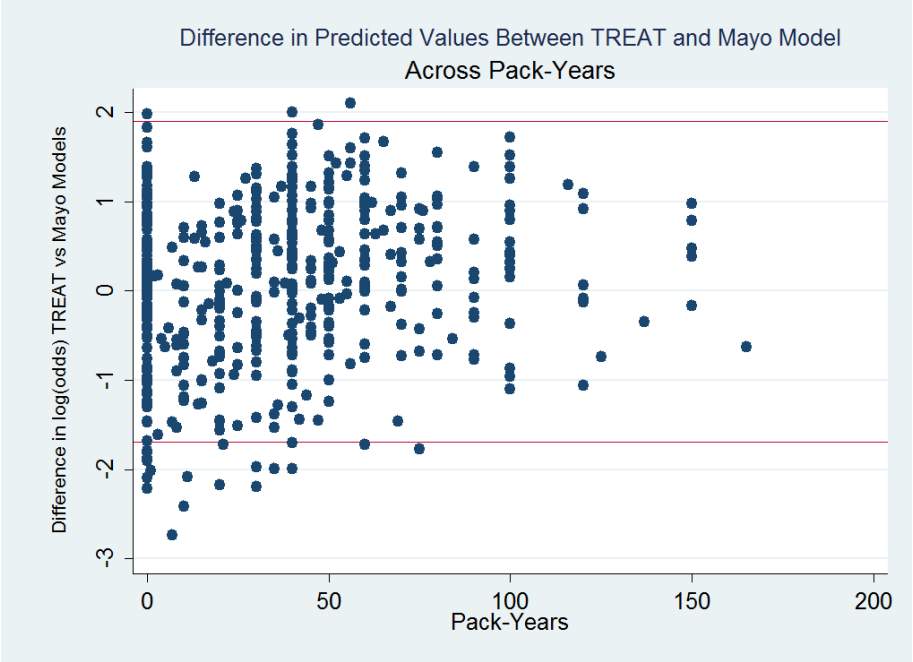


Figure 6C. Scatter plot of the difference in the log odds of the TREAT and Mayo models across pack-years. Red lines are the upper and lower confidence interval +/- 2 standard deviations.

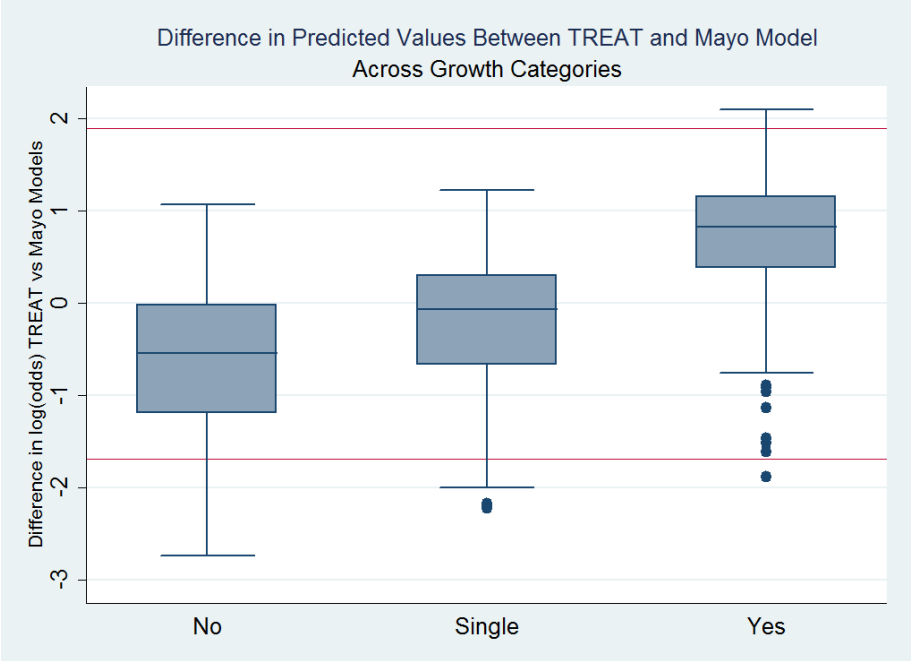


Figure 6D. Box plot of the difference in log odds of the TREAT and Mayo models across the three growth categories. "Single" growth refers to participants with only one LDCTG scan or too short a per-operative period for serial scans to report growth. Red lines are the upper and lower confidence interval +/- 2 standard deviations.

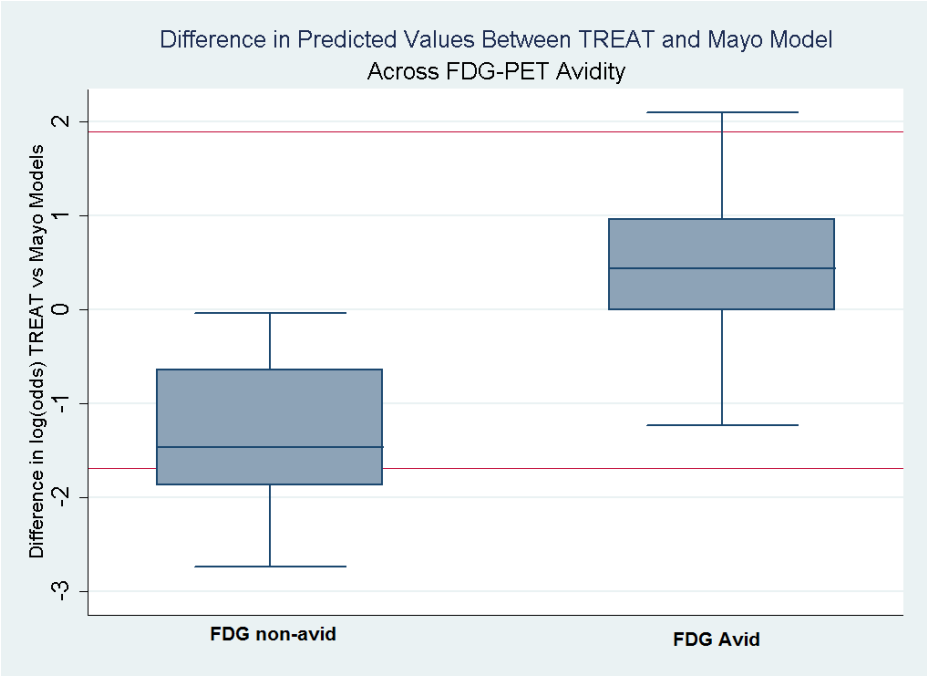


Figure 6E. Box plot of the difference in the log odds of the TREAT and Mayo models across based upon FDG Avidity. Red lines are the upper and lower confidence interval +/- 2 standard deviations.

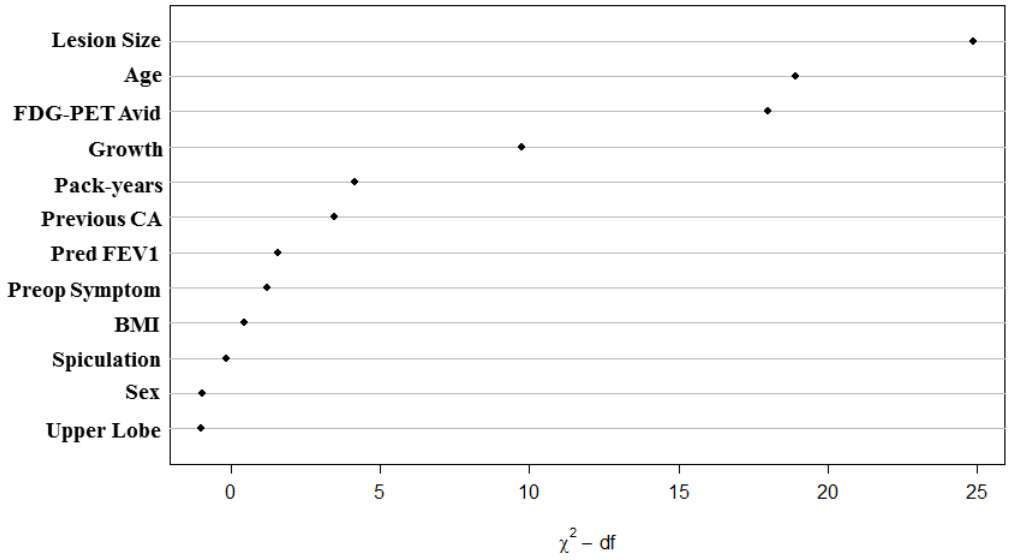


Figure 7. Chi-square plot of relative contribution to explaining variance by each dependent variable.

Chapter 3

III. Geographic variation in the diagnosis of lung cancer and implications for screening and diagnosis of lung cancer.

3.1 Background

Hundreds of thousands of lung nodules are evaluated annually throughout the US.[1] This number is likely to increase with the implementation of a national screening program for lung cancer. The National Lung Screening Trial (NLST) conducted a randomized control trial comparing annual screening with annual low dose computed tomography (LDCT) scans to chest x-rays in a population at high risk for lung cancer. LDCT reduced lung cancer related mortality by 20% and all-cause mortality by 7.6% compared to screening with chest x-ray.[2] A number of physician societies and patient advocate groups have since established guidelines recommending lung cancer screening using LDCT scans in populations similar to that enrolled in the NLST.[3-5] The Affordable Care Act tasked the US Preventive Services Task Force with reviewing disease screening programs. If after their review the Task Force finds that a screening program provides significant benefit to those screened, then their recommendation for screening must be covered by insurers. The Task Force is currently reviewing the new evidence provided by the NLST and will publish their recommendations regarding lung cancer screening in 2013.[6,7] Given the results of the NLST and the support for screening from both patient advocate groups and physician societies, a national lung cancer screening program for high risk populations using LDCT scans is likely in the near future.

Although there was a reduction in lung cancer related mortality in the NLST, 96% of lung abnormalities found by LDCT were false positive findings.[2] Other studies have found similarly high false positive rates from CT scans.[8] Furthermore, 24% of all lung procedures to diagnose lung cancer identified benign disease. Using the results of the NLST and the study's definition of the appropriate population to screen, an estimated 7.4 to 8.7 million individuals per year could be eligible for screening.[9,10] If 80% of the 8 million eligible were screened, then approximately 2.6 million anomalies would be discovered requiring further diagnostic testing over the first three years of a screening program. These findings will generate over 1 million repeat CT scans, an additional 170,000 FDG-PET scans and 80,000 diagnostic operations, 20,000 of which would have a pathological finding of benign disease.

Should a screening program be introduced in the US, then variation in the diagnosis and treatment of lung cancer may arise. Geographic variation in screening efficacy and outcomes due to healthcare access, provider bias and carcinogenic exposures were observed in screening programs for breast, colon and prostate.[11-14] Similar issues may occur with the implementation of screening for lung cancer. Understanding variation in existing practice as well as identifying factors that induce variation and are unique to lung cancer screening, can inform clinicians and policy makers prior to its implementation. Initiatives and policy can then be crafted, informed by epidemiology and health services research, to improve the administration of lung cancer screening.[15] This research addresses two factors of possible importance to the implementation of a lung cancer screening program.

First, the variation of benign disease prevalence in the US is unknown. If variation in benign disease occurs between states or regions of the country then this observed variation may be indicative of systemic underlying causes that should be determined, measured and where possible addressed as part of implementing a screening program. Should clinically significant variation in

benign disease prevalence exist, then a screening program may have geographically varying results. Measuring the amount of variation and possible patterns of variation in benign disease is the first step in determining whether a problem exists and the extent of the problem.

Second, clinicians rely heavily on radiographic imaging for the non-invasive diagnosis of lung cancer. Diagnostic guidelines recommend using F^{18}-fluorodeoxyglucose positron emission tomography (FDG-PET) for the diagnosis of lung cancer after lung lesion discovery.[3] FDG-PET scans are considered among the most accurate and cost effective non-invasive tests available for the diagnosis of lung cancer.[16-18] Some investigators have questioned the accuracy of FDG-PET in regions of the country were granulomatous disease is highly prevalent.[19-21] They hypothesize that some populations would not benefit from a diagnostic FDG-PET scan and should only receive CT scans. The following analysis examines each issue in turn and the possible impact on screening for lung cancer.

3.2 Geographic variation in benign disease prevalence after surgical lung resection in the US.

Introduction

The NLST reported screening with LDCT led to the discovery and treatment of more early stage lung cancer. This stage shift resulted in a 20% decrease in lung cancer related mortality compared to screening with chest x-ray. The participants in the NLST were recruited and received most of their care at National Cancer Institute designated cancer centers by teams with extensive experience and specialization in lung cancer diagnosis and treatment.[2,22] Yet even with this expertise, 24% of procedures performed to diagnosis lung cancer indicated benign disease and not lung cancer. A screening regimen for lung cancer with LDCT will likely lead to care being conducted by more clinical generalists and the efficacy of the screening intervention may

be reduced and the number of individuals needing to be screened to cure one lung cancer increase.[23]

Geographic variability of benign lung disease prevalence has significant implications for a national lung cancer screening program. Geographic variation may indicate a mismatch of specialist availability or practice variation. Population risks for lung cancer, like smoking prevalence, or underlying risk factors for benign disease, like granuloma caused by endemic fungal lung disease, may not be evenly distributed over the landscape. The first step in identifying whether geographic variation exists and is clinically significant is to estimate the amount of variation across the US. Then possible causes of the observed variation can be explored.

One possible cause of variation in the US could be fungal lung diseases which are the most common cause of benign granulomas. Granuloma formation is possible from any fungal lung disease. The three most common fungal lung diseases occurring in the US are histoplasmosis, coccidioidomycosis, and blastomycosis. Soils act as reservoirs for fungal spores that cause fungal lung disease. Historically, fungal lung disease prevalence varied greatly by region of the country (Figure 1). If benign granulomas are causing variation in benign disease prevalence then regions with higher prevalence of fungal lung disease should be associated with high benign disease prevalence after lung surgery compared to regions with low fungal lung disease.

We describe the benign disease point prevalence after surgical lung resection at the state level using the 2009 Medicare Provider Analysis and Review Hospital National Limited Data Set (MEDPAR). This administrative billing dataset includes all 14.7 million inpatient discharges for calendar year 2009 covered by Medicare. Medicare insurance covers hospital care for over 95% of all individuals over age 65 and is among the most comprehensive representation of hospital activity among the elderly.[24] The median age for diagnosis of lung cancer is 71 years of age.[25]

Thus using Medicare hospital data will include the majority of the US population who had surgery for known or suspected lung cancer. Fungal lung disease prevalence is examined as possibly being associated with geographic variation in benign disease point prevalence in the same surgical population.

3.3 Methods: geographic variation in benign disease prevalence after surgical resection

Study population

We conducted a retrospective cohort study using the 2009 MEDPAR dataset. The MEDPAR dataset contains claims data for services provided to beneficiaries admitted to Medicare certified inpatient hospitals and skilled nursing facilities during the 2009 calendar year. The accumulation of claims occurs from the individual's date of admission to date of discharge represents one hospital stay. Records represent final action on the claims data submitted by the hospital after all adjustments have been resolved. This dataset was made available by the Center for Medicare Services (CMS) and included ICD-9CM codes for as many as 10 diagnoses. The primary diagnosis was determined by CMS as the diagnosis responsible for the majority of care delivered during that hospital stay. Each record also included sex, county of residence at the time of admission and as many as 10 procedures that occurred during that stay. The primary procedure was determined by the procedure responsible for the majority of hospital days or healthcare services provided during that stay.[26]

All individuals with a primary procedure ICD-9 code of lung resection or thoracoscopic biopsy (32.2, 32.9, 32.3, 32.40, 32.41, 32.49, 32.6, 32.9, 33.1, 33.2, or 33.28) were included in the study (N=33,655). Those who had a second hospitalization with a primary procedure of lung resection or biopsy (N=538) were excluded from further analysis. These procedures were likely re-

operations from the initial hospitalization, treatment for later stage disease or palliative in nature. Outcome of the surgery was categorized by primary diagnosis. Diseases were classified into benign, cancerous or a disease not arising from a lung nodule. Benign diseases included: histoplasmosis, aspergillosis, blastomycosis, coccidiodomycosis, candida, other mycotic disease not otherwise specified, sarcoidosis, lung abscess, tuberculosis, other mycobacterial disease, pulmonary fibrosis, benign neoplasm, coin lesion, Wegeners granuloma, dermatophytosis, other non-neoplastic diseases of the bronchus, pulmonary diseases due to other mycobacteria and benign disease of the lung not otherwise specified. Cancer diagnoses included all malignant lung neoplasms of the lung or bronchus, secondary neoplasms of the lung, malignant neoplasm of the pleura, carcinoma *in situ* of the lung or bronchus, neuroendrocrine neoplasms, and neoplasms of unspecified nature in the respiratory system.

The diseases not arising from a lung nodule included: metastatic lung cancer, empyema, idiopathic pulmonary fibrosis, pneumonia, pneumonopathy, pneumothorax, emphysema, acute respiratory failure, septicemia and unspecified pleural effusion. Each individual with a disease not arising from a lung nodule was excluded from further analysis. Total surgeries were grouped by state and point prevalence of benign disease was estimated by dividing the number of individuals with benign disease after lung surgery by the sum of lung surgeries for cancerous and benign disease.

Fungal Lung Disease Exposure

Over 1.2 million men age 17 to 21 received a histoplasmin skin test during physical examination upon entry into the Navy between 1959 and 1968. The histoplasmin skin test was similar to the modern tuberculosis skin test in that a solution with *histoplasmosis capsulatum* was injected subcutaneously and if a skin reaction occurred, the host had been exposed to histoplasmosis or coccidioidomycosis or blastomycosis with which the test was cross reactive.[27] For purposes of the

national surveillance, a skin induration of at least 4mm was considered a positive reaction for fungal exposure. Among this population 488,010 white males reported residing in only one county during their lifetime prior to recruitment.[28] Other races were not reported due to lack of sufficient numbers to allow county level public reporting of prevalence. Positive histoplasmin skin tests occurred in 16.5% of the white males from a single county of residence. Prevalence among the 488,010 participants was collated across years and reported in tabular form by county or groups of contiguous counties within a state.

Fungal lung disease prevalence was assigned to 498 individual counties or state economic areas as reported in the 1969 National Surveillance Survey.[28] This historical prevalence was used to assign to MEDPAR lung surgery patients their likely prevalence of fungal lung disease. State economic area lung disease prevalence estimates based on the 1969 survey were entered into a geographical information system (GIS) (ArcGIS v10, Redlands, CA). The patient's geographic residence was defined as their self-reported county of residence at the time of hospitalization and was added to the GIS. New counties created since 1969 were assigned to state economic areas based upon the location of the county seat. All counties created since 1969 were within the 1969 state economic areas and no population based weighting of exposure was necessary. Outcome of cancer or benign disease after surgery was input into this GIS for each individual. Estimated fungal lung disease exposure based upon participant's county of residence at hospitalization, were examined for association with individual outcome of benign disease or lung cancer.

Statistical Analysis

Benign diagnosis point prevalence was estimated at the state level by dividing the total number of benign cases by the sum of the benign cases and malignant cases. Benign disease point prevalence was compared between states using Pearson chi-square test. Sensitivity analysis was conducted by removing the states with the highest and lowest benign disease prevalence and

benign disease point prevalence was estimated in this sample of 48 states. Fungal lung disease prevalence based on county of residence at time of hospitalization and derived from 1969 national survey was compared between individuals with benign disease and those with cancer using univariate analysis with the Wilcoxon rank-sum test. All statistical analysis was conducted with Stata v12 (College Station, TX)

3.4 Results

There were 25,378 Medicare beneficiaries that had a lung operation for known or suspected lung cancer in 2009. Among these, 2,328 (9.2%) had a benign diagnosis and 23,050 had a diagnosis of lung cancer. Benign diagnosis was more frequent among women (9.8%) than men (8.5%) after surgery. Crude in hospital mortality rate for all patients and among those with cancer was 2.3%. The mortality rate for patients with benign disease was 2.1%. A significant difference in crude mortality rate was not observed between those diagnosed with lung cancer and individuals diagnosed with benign disease ($p=0.86$). Prevalence of benign disease varied significantly (chi-square $p<0.001$) across states from a low of 1.3% in Vermont to a high of 25.0% in Hawaii (Figure 3). Median benign disease by state was 8.9% (IQR: 7.8 – 10.9). Other states with high prevalence included Wyoming (18.4%), New Mexico (16.7%) and Montana (14.5%). States with low prevalence of benign disease were Idaho (5.2%), Maine (5.7%) and Rhode Island (5.7%). After excluding Vermont and Hawaii as possible outliers, significant differences in benign disease point prevalence between states remained ($p=0.001$).

Median fungal lung disease prevalence mapped from the 1969 surveillance data was 5.5% (IQR: 2.2%, 23.8%) among those with a diagnosis of lung cancer. Median fungal lung disease prevalence was 5.2% (IQR: 2.2%, 23.9%) among those with a benign diagnosis. Fungal lung disease prevalence was highly skewed with mean prevalence of 17.5% (standard deviation

24.0%) for those diagnosed with cancer. Fungal lung disease prevalence was estimated to be17.6% (standard deviation 24.3%) among individuals with a benign diagnosis. significant differences in fungal lung disease prevalence between those with cancer or a benign diagnosis was not observed (p=0.90).

3.5 Discussion

The benign disease point prevalence between states varied significantly in the 2009 Medicare population in univariate analysis. The four Rocky Mountain States, Montana, Wyoming, Colorado, and New Mexico as well as Arizona, Nevada, Nebraska and Kansas had benign disease point prevalence between 10.8% (Kansas) and 18.4% (Wyoming). The three states of Maine (5.7%), New Hampshire (7.4%) and Vermont (1.3%) had low point prevalences. These were the main observable clustering of benign disease by state.

Screening for lung cancer in states with higher prevalence of benign disease will likely result in higher frequencies of benign diagnosis and exposing individuals to possible harm from surgery related mortality, given the variation in benign disease after lung surgery across the US. Both outcomes act to decrease the efficacy of a screening program within high benign prevalence states. Understanding the causes of the observed variation can inform clinical and policy interventions to decrease variation and improve the outcome of surgery for suspected lung cancer.

Locally endemic fungal lung diseases were explored as a possible cause for higher benign disease prevalence. In univariate analysis, no statistically significant relationship was evident between individually estimated fungal lung disease exposure and benign disease after lung surgery. Estimated median exposures among those with cancer or benign disease (5.5% and 5.2% respectively) were one third the mean exposure and median exposure among those with benign

disease was actually lower than exposure among those with lung cancer. We estimated individual exposure based upon a national survey conducted from 1959 to 1968 of young, healthy men who lived in a single county. Using this survey of lung disease prevalence is problematic in a number of ways.

First, the MEDPAR population is an elderly population compared to those surveyed in 1969 and may well have lived in multiple residences with a variety of occupations prior to surgery in 2009. Their exposure history is likely much more complex than the simple model used here to estimate individual exposure as being equal to the average likelihood of exposure based upon residence at time of diagnosis.

Second, exposure prevalence from fungal lung diseases may be difference today, 44 years after the 1969 survey was completed. Fungal spores reside in soils and are preferential to floodplains and farmland in relatively humid climes. Baddley and colleagues found higher prevalence of acute histoplasmosis in the elderly in the Nebraska and the Northern Great Lakes area than previously found.[29] This may indicate a migration over time of fungal spores up the Missouri River valley and from the lower to the upper Great Lakes and New England or, in the case of the Northeastern US, increases seasonal mobility of New England populations to regions of higher fungal lung disease prevalence in the south.

Finally, our individual exposure estimate based upon aggregate county level prevalence may suffer from the ecological fallacy. Ecological fallacy occurs when individual characteristics are assumed to be represented by an aggregated measure. We used the average prevalence in each county or economic area from the Edward study to represent individual exposure. This assumed average, based upon a geographic area, may misclassify individual exposure. Increased urbanization and impermeable cover of soils may have greatly reduced possible exposure. The

reduction in rural population and farming occupations over this time period would act to reduced individual exposure as well. However, exposure is increased among those in agricultural industries or who frequent the outdoors. The lack of individual data on occupation also hampers estimation from another important vector of fungal lung disease exposure.

Given the national demographic trends in the last 40 years of increased urbanization and increased mobility, one can conjecture that both trends cause misclassification of fungal lung disease exposure in such a way to attenuate any association toward the null. For example, increased urbanization concentrates populations in urban areas as well as spreads the extent of the urban geography into previously rural environments. The decrease in exposure to soils and transmission of fungal spores in urbanized areas would likely lead to overestimation prevalence of fungal lung disease in both those with and without cancer when using the 1969 study. On the other hand, greater population mobility would cause those currently residing in urban areas to be assigned a low prevalence when their historical exposure could be higher.

While fungal lung disease acting through the mechanism of granulomas has biological plausibility as a cause of geographic variation of benign disease prevalence, the available geographically based estimates of fungal lung disease exposure are methodologically questionable. Independent verification of exposure through either direct measurement or a better estimate of fungal lung disease reservoirs in soils through land cover maps is necessary. Ideally, this land cover mapping would also be coupled with data on an individual's occupational and residence history.

Another possible cause of the observed variation in benign disease prevalence is the varying smoking rates by state. The prevalence of smoking in adults over age 18 who smoked more than 100 cigarettes in 2009 ranged from a low of 9.8% in Utah to a high of 25.6% in West Virginia

and Kentucky.[30] However, current smoking prevalence would not accurately represent smoking exposure since the latency period for the development of lung cancer is approximately 20 years. Jemal and colleagues found that the smoking prevalence between 1992 and 2007 decreased in all states except Wyoming. Nationally, smoking among men decreased from 26.5% to 20.1% and among women decreased from 21.5% to 15.5%. The largest decreases occurred in western states and the smallest decrease occurred in the Midwest. Therefore, state level or ideally county level smoking prevalence from the early 1990's would be necessary to control for smoking prevalence as a confounder. Higher prevalence of smoking results in more lung cancer in that state's population.[31,32] In our analysis we were unable to discriminate between those with known lung cancer and those with suspected lung cancer at the time of surgery. This resulted in a lower national mean prevalence of benign disease (9.2%) in this study when compared to the NLST (24%)[2] or other surgical studies (20% to 40%).[19,33,34] Therefore, higher rates of smoking and subsequent higher lung cancer prevalence in the state's population will inflate the denominator in estimating prevalence and decrease the benign disease prevalence. States with higher smoking prevalence will tend to have lower benign prevalence compared to states with low smoking prevalence. Not accounting for smoking prevalence likely biases the variance between states upward. Smoking prevalence from the 1992 Tobacco Use Supplement to the Current Population Survey was not available publically by state at the time of analysis. However, a free source of this data has been located reporting smoking prevalence by sex by state. Additional analysis using this resource will be future work.

Surgical practice variation is well documented in outcomes literature related to diagnosis, hospital mortality, complications, and length of stay.[35-39] In a recent study examining changing benign surgery rates at a single institution, the authors found that implementation of minimally invasive surgical technique reduced morbidity and mortality associated with the surgery, but they also observed that benign disease diagnosis rates increased after implementation of video assisted

thoracoscopic surgery. The authors conjectured that due to a broader range of patients able to undergo surgery with the new and lower risk from the new surgical method, clinicians were willing to more aggressively pursue a diagnosis. The risk threshold missing a cancer had effectively been reduced because the risk for undergoing surgery had decreased[34]. Varying risk thresholds, training and specialist availability may all play a role in the observed benign disease prevalence variation. The impact of new technology in the diagnosis and treatment of lung cancer should be explored to determine whether differential adoption by geography or across time influenced benign disease prevalence.

Another clinical group pursuing best practices in lung cancer screening and treatment found that resection for benign disease could be minimized with adherence to a clinical diagnostic protocol even in an area with endemic histoplasmosis.[40] The success of this clinical group in diagnosing lung cancer suggests that best practices do exist and that promulgation of such practices could reduce some of the observed variation in benign disease prevalence. Much like the efforts to establish best practices in mammography screening,[41,42] creation and promulgation of evidence based best clinical practices for diagnosis and treatment of lung cancer is critical to the success of a national lung cancer screening program.[3,43]

One other weakness of this study is its analysis of one year of Medicare data to estimate benign disease variability. Benign disease prevalence may not remain stable over time and individual states may jump between quartiles of disease prevalence. Such variation over time by a state would weaken these results. Additional years of data should be pursued to estimate the relative stability of a state's prevalence and their association with changes in practice described above. A more complete model of benign disease prevalence variation that includes beneficiary demographics, smoking and fungal lung disease exposures, and practice changes that can vary over time and across the US would best be able to estimate benign disease prevalence.

In conclusion this exploratory study found significant variation by state in the point prevalence of benign disease after surgery for known or suspected lung cancer in the elderly. The robustness of these results should be explored further by exploring variation in benign disease prevalence by state over time and associations with smoking and fungal lung disease prevalence. The high prevalence of benign disease in some states would decrease the efficacy of a screening program for early detection and treatment of lung cancer in those states.

3.6 Geographic variation in the FDG-PET accuracy to diagnosis of early stage lung cancer in the ACOSOG Z4031 Trial

Background

Once a lung lesion has been discovered, surgeons rely heavily on radiographic imaging for the non-invasive diagnosis of lung cancer. According to national guidelines if a lung nodule is larger than 8mm in diameter and either the probability of cancer is between 5% and 60% or the nodule does not exhibit characteristics indicative of a benign etiology like calcification, then a F^{18}-fluorodeoxyglucose positron emission tomography (FDG-PET) combined with computed tomography scan is suggested for the non-invasive diagnosis of a lung nodule.[3,43,44] Meta-analyses of FDG-PET diagnostic accuracy reported a 96% sensitivity and 78% specificity.[45,46] FDG-PET scan is among the most accurate non-invasive tests available for the diagnosis of lung nodules.[46]

FDG-PET scans were used as part of the diagnostic process in 8.3% of the individuals with a positive CT scan during the NLST. If a national lung screening program is implemented using the NLST criteria and proposed guidelines, then an estimated 170,000 FDG-PET scans at a cost of $2,000 to $3,000 per scan will be performed to diagnose the screening discovered lung lesions.[2,47,48] However, some studies have questioned the efficacy of FDG-PET scans to diagnose

lung cancer where granuloma generating diseases of the lung are endemic.[19,21] Granulomatous disease of the lung can look identical to cancerous tissue on CT or FDG-PET scan. It can have all the hallmarks of a cancerous lesion including symptomatic hemoptysis, growth on repeated CT scans, spiculated edge characteristics and the lack of calcification (Figure 2)[49,50] Lung granulomas are often caused by fungal diseases in the US.[34,51-54] Researchers in countries where tuberculosis is the etiological cause of most granuloma also question the use of FDG-PET to diagnose lung cancer.[55,56] In these studies the specificity of FDG-PET was 25% to 50%, much lower than that reported in either meta-analysis.

Granuloma is a common benign diagnosis of lung nodules in the US and accounts for 45% to 65% of pathologically determined benign disease after FDG-PET imaging.[16,19,52] Approximately 15% to 40% of benign granulomas demonstrated an active fungal disease in these studies. As described earlier, fungal lung disease prevalence varies widely across the US (Figure 1). What is not well understood is the interaction of endemic fungal lung disease, subsequent lung granuloma formation and its impact on FDG-PET imaging. The studies by Croft and Deppen occurred in regions of the country were histoplasmosis prevalence was over 80% in the 1969 national survey. The relationship between endemic fungal lung disease and degradation of FDG-PET specificity may be localized to areas of extreme fungal disease prevalence or it may directly reflect local conditions and exposure patterns unique to a locale and the activities of the population or practice variation. If FDG-PET scan accuracy to diagnose lung cancer varies with endemic fungal lung disease, then populations may exist that would not benefit from this expensive diagnostic test.

Furthermore, FDG-PET scans may not be as sensitive to detecting lesions below 2cm in maximum diameter.[45] A screening population like that in the NLST has 50% pathologically determined stage 1 disease at time of diagnosis. This compares to 15% stage 1 disease at diagnosis under current symptomatically and incidentally discovered lung cancer. Clinical stage

1 disease is broken down further into stage 1A where the tumor is less than 3 cm in maximum diameter on radiograph (40% prevalence in the NLST) and stage 1B where the tumor is greater than 3 cm (10% prevalence in the NLST). In clinical stage 1 disease, no lymph nodes are enlarged indicating possible local metastasis and no other tumors are evident in the lungs, bone or other organs. The high prevalence of smaller lesions that typify the type of lesion discovered in LDCT screening may reduce the efficacy of FDG-PET to detect cancerous lesions and decrease the sensitivity of the test in a screening population. The NLST did not include FDG-PET outcomes in its study, but another national surgical trial found limited usefulness in FDG-PET for preventing unnecessary surgery of clinical stage 1 disease.[33] The combination of reduced sensitivity due to the larger numbers of smaller lesions coupled with possibly poorer specificity among populations with endemic fungal lung disease may make FDG-PET scans a particularly ineffective test in screening populations in the US. A national study of participants with clinical stage 1 lung cancer who had a FDG-PET scan would emulate lesions arising from a screening program and presenting to surgeons for diagnosis.

The recently concluded American College of Surgeons Oncology Group (ACOSOG) Z4031 trial evaluated participants with clinical stage 1 Non-small Cell Lung Cancer (NSCLC). The ACOSOG Z4031 trail obtained FDG-PET results in addition to bio-specimens associated with the studies primary research aim and provides the largest national sample to determine the accuracy of FDG-PET to diagnose early stage lung cancer to date. We conducted a secondary analysis of the ACOSOG trial to estimate the accuracy of FDG-PET to diagnose lung cancer in patients with known or suspected clinical stage 1 lung cancer. The possible exposure to fungal lung disease in this trial was likely to vary widely as this was a national trial with participants enrolling from a variety of cities and regions of the country. With this unique population we examined whether FDG-PET scan accuracy varied by the size of the lesion, by study site and whether false positive FDG-PET scans were associated with an estimated fungal lung disease exposure. We

hypothesized that a population with smaller lesions and residing in a region of endemic fungal lung disease would likely not benefit from a FDG-PET scan for diagnosis prior to surgery.

3.7 Methods

Study Population

The primary objective of the ACOSOG Z4031 study "Use of Proteomic Analysis of Serum Samples for Detection of Non-Small Cell Lung Cancer" (5U10CA076001-11) was to determine prospectively whether a serum proteomic profile can predict the presence of primary NSCLC in patients with suspicious lung lesions who are candidates for lung resection. The study design was a prospective study of 1000 patients undergoing lung resection for clinically known or suspicious clinical stage 1 lung lesions. The ACOSOG trial was a national study that occurred across 23 states and Ontario, Canada in 51 hospitals in 39 cities (Figure 3). Patients were enrolled in the ACOSOG Z4031 trial from February 2004 to May 2006. Inclusion criteria were: 1) 18 years or older, 2) clinically suspicious lung lesion that was possible stage 1 lung cancer, 3) CT scan < 60 days prior to the lung resection and no evidence of metastatic disease, 4) no untreated malignancies, 5) no malignancy within the past 5 years except effectively treated basal cell or squamous cell skin cancer, surgically treated carcinoma in situ of the cervix, or surgically treated lobular carcinoma in situ of the ipsilateral or contralateral breast with low risk for recurrence, and 6) able to provide informed consent. Exclusion criteria included patients who had: 1) undergone previous lung resection within the preceding 30 days, 2) received prior chemotherapy or radiotherapy and 3) received a blood product transfusion of any kind within the past 60 days of the operative procedure.[57] The ACOSOG Z4031 prospective clinical trial contains data on 969 patients who met all the eligibility criteria.

Data Collection

At time of enrollment age, sex, race, ethnicity, body mass index, date of operation, pre-operative clinical stage, CT radiologist reports, FDG-PET radiologist reports, enrolling site and zip or postal code of patient residence were collected. Follow up of the study participants was 5 years since time of enrollment. Operative notes and pathological reports were collected along with 30 day mortality status, status at last follow-up, date of last follow-up, and cause of death if dead at last follow-up. All data were stored and maintained by the ACOSOG data center. Additional data was extracted from ACOSOG case report forms and study analysis was conducted under a separate ACOSOG approved protocol (Z4095). Data abstracted from the study case report forms by two trained medical reviewers included: clinical maximum lesion diameter according to CT or PET/CT immediately prior to surgery, smoking status, smoking pack years, pre-operative FDG-PET scan result and standard uptake value (SUV), pathological result for benign disease, and cancer histology. FDG-PET scan results were categorized based on the radiologist descriptor of avidity or maximum SUV. Four categories of FDG-PET scan results were not avid/not cancerous (SUV=0), low avidity/not likely cancerous (SUV 0.1 to 2.5), avid/likely cancerous (SUV≥2.5 to 5) and highly avid/cancerous (SUV≥5). Categories were based on radiological guidelines and clinical convention.[52,58] No original scans were reviewed. Pathological reports and operative notes were reviewed to determine etiology of benign disease and the specific cancer diagnosis. This secondary analysis of the ACOSOG Z4031 study was approved by the Vanderbilt IRB.

Association with Fungal Lung Disease Exposure

Fungal lung disease prevalence used the 1969 National Surveillance Survey and the details of this survey were described earlier in this chapter.[28] Unlike the MEDPAR data, individually reported zip codes were available in the Z4031 trial. The patient's residence was defined as the geographic center of their self-reported residential zip code at the time of study enrollment and was entered into the GIS. When a zip code crossed multiple state economic areas or counties and

possible differing fungal lung disease prevalence exposure, then fungal lung disease prevalence was weighted according to the proportion of the population in the zip code that was within a state economic area. Fungal lung disease exposure was the population weighted average disease prevalence for each county the zip code occurred. Population data were derived from 2000 census track data and mapped to each zip code occurring in 2000.[59,60] Census data from 2010 was not available at the time of analysis and since the majority ACOSOG Z4031 study participant of enrollment occurred prior to 2006, the 2000 census was assumed to be more representative of the national distribution of adults than 2010 census data. FDG-PET results, pathological diagnosis and estimated fungal lung disease exposure for each participant were added into the GIS. Estimated fungal lung disease exposure based upon participant zip code at study enrollment, were examined for association with individual pre-operative FDG-PET results.

Statistical Analysis

Differences in the demographics between the benign and malignant participants were compared using a t-test for continuous variables (age and lesion size) and binomial proportions test for differences in proportions (sex, race and FDG-PET avidity). Enrolling clinics in the same city or clinics in the same practice group were combined for geographic analysis of FDG-PET accuracy and outcome reported by city. Sensitivity, specificity, positive and negative predictive values were calculated for the ACOSOG study population who had FDG-PET scans available and among cities having at least 25 enrollees (N=8). The FDG-PET accuracy to diagnose lung cancer equals true positives plus true negatives divided by the total population tested and was calculated using the pathological diagnosis as the gold standard. Differences in the sensitivity and specificity between institutions were estimated using the chi-square statistic. The accuracy of FDG-PET by CT size group was compared with an analysis of variance. A logistic regression model was used to estimate the association of historical fungal lung disease prevalence of the patient's residence

at time of enrollment with false positive FDG-PET scan results among all positive FDG-PET scans, adjusting for age at diagnosis and maximum lesion size reported on pre-operative CT.

3.8 Results

Our current analysis had 682 participants which met the ACOSOG Z4031 trial eligibility criteria and had a preoperative FDG-PET scan. Benign disease was found in 116 patients (17%) and lung cancer in 566 patients (83%). Microbiology results were abstracted from pathological or operative reports and only 11% of benign cases have an unknown etiology. Of the 116 benign cases, 75 (65%) were documented in the pathology report as being granulomatous and 30 (26%) of the granulomas had documented histoplasmosis etiology in the pathology report. Smoking data was not collected at time of study enrollment but was abstracted for this analysis, resulting in 280 (29%) patients with smoking exposure information. Patients with cancer were more likely to be older, non-Caucasian, and have larger lesions that were FDG-PET avid than patients with benign disease (Table 1).

The overall accuracy of FDG-PET to diagnose lung cancer was 73% when compared to the pathologic diagnosis. The sensitivity and specificity were 82% and 31% and the positive and negative predictive values were 85% and 26%, respectively (Table 2). Table 3 shows the pathology of the enrollees with false positive and false negative FDG-PET scans. The majority of patients with false positive FDG-PET scans had granulomas and the majority of false negative FDG-PET scan results had adenocarcinoma. Of the 80 false positives, 69% of these were granulomas. Twenty-one of the 36 true positive lesions (58%) wee granulomas. Among the 101 lung nodules with a false negative FDG-PET scan, 11 were less than or equal to 1 cm in maximum diameter measured by CT pre-operatively, and 62% of these lesions were

adenocarcinoma, 11% were squamous, 10% were broncho-alveolar cell and 9% had a neuroendocrine pathology.

There were 8 cities with more than 25 participants (Table 4). The observed sensitivity by city varied from 68% to 91% (p=0.03) and the specificity ranged from 15% to 44% (p=0.72). FDG-PET accuracy improved with lesion size (Figure 2) from 67% in lesions that were one to two cm (sensitivity 76% and specificity 35%) to 84% in lesions that were three to five cm (sensitivity 90% and specificity 18%) ($p<0.001$).

The median estimated prevalence of fungal lung disease in patients with FDG-PET avid lesions was 5.6% (IQR 2.7, 28.5). We evaluated the association of fungal lung disease prevalence and false positive FDG-PET scans among those with a positive FDG-PET scan using logistic regression, controlling for age and pre-operative CT size. Of 545 positive FDG-PET scans, 480 had a valid residential zip code which could be assigned a 1969 fungal lung disease prevalence based on their residence at enrollment. Age (OR 1.04, 95% CI 1.01-1.07) and pre-operative CT size (OR 1.04, 95% CI 1.02-1.07) were strongly associated with a positive FDG-PET scan ($p<0.01$). The prevalence of fungal lung disease increased the likelihood of a false positive scan (OR 0.99, 95% CI 0.98-1.002), but the association was not statistically significant ($p = 0.12$).

3.9 Discussion

The poor results of FDG-PET to diagnose lung cancer in the ACOSOG Z4031 trial give insight into possible issues for this imaging modality when applied to lung nodules arising from a screening population. The ACOSOG Z4031 study had participants from 17 cities with enrolling centers that also participated in the NLST. Pathological stage 1A disease was observed in 45% of the ACOSOG trial and in 40% of the NLST. The NLST had more stage 3 and 4 lung cancer, and

the ACOSOG trial had more operable stage 2 disease. Individuals with CT evidence of possible metastatic disease would receive a FDG-PET scan for non-invasive staging and not for diagnosis.[61,62] Thus, some number of participants in the NLST who received a FDG-PET scan did so for staging purposes and not for diagnosis. Individuals with clinically determined stage 3 and stage 4 lung cancer are not candidates for surgery. This analysis does not address the accuracy of FDG-PET for staging. The NLST did not collect data on FDG-PET scans, and direct comparison of FDG-PET accuracy between the Z4031 trial and the NLST is not possible. Even with these differences between the two populations, a number of conclusions can be drawn regarding the use of FDG-PET scans and extrapolated to the types of lesions likely to be found in a low dose CT screening environment.

Currently available PET scanners have limited ability to detect metabolically active lesions smaller than 8 mm, and FDG-PET isn't recommended for lesions smaller than 8 mm due to high false negative rates.[3,63] For those lesions between one and three cm in diameter, a recent meta-analysis found FDG-PET to be accurate (sensitivity 94% and specificity 83%).[64] Stage 1A and stage 2A disease are characterized by T1 tumors that are smaller than three cm in diameter.[65] The much lower sensitivity observed in the ACOSOG trial may be caused by this population having more lesions less than three cm compared to the studies used in the meta-analysis. The accuracy in the 448 lesions smaller than three cm was 68%. If smaller lesions can be expected to arise from a screening population than a clinical population, then the lower sensitivity of FDG-PET may be more similar to that observed in our study than previously reported in the literature.

The specificity of FDG-PET to diagnose lung cancer in this trial is similar to two previously reported surgical series from Iowa City, Iowa and Nashville, Tennessee.[19,21] Both of these regions have a high prevalence of fungal lung disease. Croft and colleagues reported a sensitivity and specificity of 93% and 40% respectively in their smaller (N=74), Midwest cohort, with all

imaging performed at a tertiary referral center.[21] Deppen et al reported a sensitivity of 92% and specificity of 40% in a patient population (N=279) predominantly from the south central U.S., with imaging performed at a variety of regional imaging centers. Both studies noted elevated numbers of false positive scans and lowered specificity. The underlying populations from each study had a high prevalence of endemic fungal lung disease and granulomas were the most common benign results observed. The current study also had a preponderance of granulomas.

The results from this study differ from prior meta-analyses by Gould et al which reported a sensitivity and specificity of 96% and 78%, respectively and by Cronin et al who report similar results for combined FDG-PET/CT scans (94% and 83%, respectively). [45,64] The differences in our current analysis compared to the two meta-analyses may arise from a number of factors. First, verification bias may be present in our current study due to the entire ACOSOG study population having a pathologic diagnosis after surgical resection. Not all results reported in either meta-analysis relied solely on participants having a pathological diagnosis. Verification bias may explain the observed low specificity but does not explain the lower sensitivity in this study when compared to previous reports.[66]

The population of patients in the Z 4031 trail may also be different from those collected from publications used by the meta-analyses. The inclusion criteria of the Z4031 trial required clinical stage 1 disease prior to surgery. Other single institution series include those with known or suspected lung cancer, as well as, all pre-diagnosis clinical stages. A large number of sites contributing patients in the ACOSOG study are in regions of the United States with a high prevalence of fungal lung disease; and consequently, a large number of granulomas were observed in this series. This was not true for the sites used in the meta-analyses. Granulomatous disease and the fungal lung disease and tuberculosis that cause them are rare in Europe and Japan. In Gould's meta-analysis 727 of the 1474 (49%) lesions included were from either Japan or

European populations. At least another 15% of the lesions in the meta-analysis were from regions of the US where granulomatous disease is rare. Finally, the FDG-PET scans in the ACOSOG study were performed at many different institutions as well as from community imaging centers and were read by a variety of radiologists. This lack of uniformity in test administration and reporting likely introduced variability in scanning quality and interpretation by the radiologists. The studies included in both meta-analyses where conducted primarily at academic medical centers by thoracic radiologist specialists and not in the community setting. Overall, our current study reports a reduced sensitivity and specificity and reduced accuracy of FDG-PET to diagnose lung cancer when compared to published meta-analysis.

No association was observed between false positive FDG-PET scans and historical fungal lung disease after controlling for an individual's age and lesion size. The fungal lung disease increased the likelihood of a false positive FDG-PET scan but not significantly ($p=0.12$). The issues surrounding use of the 1969 survey were described in detail above. In summary estimating fungal lung disease with the 1969 national survey may lead to exposure misclassification due to: 1) ecological fallacy of assigning individual exposure from a population average, 2) historical data are being applied to an older population that has more varied exposure profile than that surveyed, 3) there is some evidence of migration of fungal spores over time, and 4) the distribution of both population to more urban settings and urbanization of the landscape over the last 44 years likely reduced exposure. Thus, it is not possible to know the direction of bias within the ACOSOG study population due to the counteracting effect of the various possible mechanisms of misclassification. Available estimates of exposure are methodologically questionable and require independent verification.

Except for cytopathology or microbiology testing of the lung lesion after surgery, there is no method of directly measuring fungal exposure. The skin test used in the 1969 survey is no longer

in production. There are three possibilities to solve the problem of determining whether an individual was exposed to fungal lung disease: reproduce the old skin test, better estimate fungal spore reservoirs in soils with land cover modeling or find another radiographic or biological biomarker that indicates the lesion is a granuloma. Land cover mapping of environments preferential to fungus spores coupled with an individual's movement over time in the landscape could estimate individual fungal lung disease exposure. Such an estimate should be a significant improvement over assigning a prevalence based upon representation of the population average within the arbitrary political boundary of a county.

Our study is one of the largest series evaluating the accuracy of FDG-PET to diagnose lung cancer in clinical stage 1 disease and represents a national sample with over 650 patients from 39 cities in the United States. Cancer or benign disease was determined pathologically as all patients had a surgical resection. Because it is a clinical study in a large national sample from multiple institutions with multiple surgeons and interpreting radiologists, the results from this analysis are generalizable to clinical practice for early stage patients being evaluated for surgical resection. However, as our study is a secondary analysis of a clinical trial designed and powered for other purposes, biases associated with retrospective reviews of the FDG-PET results are possible. To reduce these biases, reviewers were used who had experience with these types of chart reviews, were blinded to the final pathology and staging and did not conduct the statistical analyses. Because FDG-PET scans were performed at multiple academic and community centers, there were no standard FDG-PET scan administration or interpretation protocols. We believe this is both a strength and weakness of the study as it increases the generalizability of the study nationally but the results may not be applicable to high volume centers with expertise in FDG-PET scans. In addition, this study does not address the role of FDG-PET scan for clinical staging of lung cancer.

In conclusion, the accuracy of FDG-PET scan to diagnose lung cancer in a national sample of patients with known or suspected clinical stage 1 NSCLC is less than previously published meta-analyses. Clinicians must assess the pre-test probability of disease and consider whether a positive or negative FDG-PET scan will material alter their treatment decisions. Smaller, stage 1 lesions less than three cm in diameter may reduce the diagnostic yield of FDG-PET to such an extent that FDG-PET should not be used for diagnostic purposes. A positive scan is more likely to be cancer, because the prevalence of disease in a population being evaluated by surgeons is generally above 60%. False positive scan are common. Therefore, results of FDG-PET should be interpreted cautiously when diagnostic or treatment decisions are being made for patients with suspicious pulmonary lesions. Further research is needed to determine the impact of fungal lung disease on false positive FDG-PET results. In addition, additional diagnostic tests should be developed and used to minimize false negative results when adenocarcinoma, carcinoid or bronchoalveolar cell tumors are suspected.

3.10 References

1. Ost D, Fein AM, Feinsilver SH. Clinical practice. The solitary pulmonary nodule. *N Engl J Med.* Jun 19 2003;348(25):2535-2542.

2. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine.* 2011;365(5):395-409.

3. National Comprehensive Cancer Network. Lung Cancer Screening v1.2012. *NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines).* 2011. http://www.nccn.org/professionals/physician_gls/recently_updated.asp. Accessed 10-27-2011.

4. Bach PB, Mirkin JN, Oliver TK, et al. Benefits and harms of ct screening for lung cancer: A systematic review. *JAMA: The Journal of the American Medical Association.* 2012;307(22):2418-2429.

5. Lung Cancer Research Group. *American Cancer Society Interim Guidenace on Lung Cancer Screening.* 2012.

6. US Government Printing Office. Public Law 111-148 - Patient Protection and Affordable Care Act. In: GPO, ed. *H.R. 3590*: GPO; 2010.

7. US Preventive Services Task Force. Topics in Progress. 2012; http://www.uspreventiveservicestaskforce.org/uspstf/topicsprog.htm. Accessed 11-27-2012.

8. Swensen SJ, Jett JR, Hartman TE, et al. CT screening for lung cancer: five-year prospective experience. *Radiology.* Apr 2005;235(1):259-265.

9. Doria-Rose VP, White MC, Klabunde CN, et al. Use of Lung Cancer Screening Tests in the United States: Results from the 2010 National Health Interview Survey. *Cancer Epidemiology Biomarkers & Prevention.* July 1, 2012 2012;21(7):1049-1059.

10. Pinsky PF, Berg CD. Applying the National Lung Screening Trial eligibility criteria to the US population: what percent of the population and of incident lung cancers would be covered? *Journal of Medical Screening.* September 1, 2012 2012;19(3):154-156.

11. Logan RFA, Patnick J, Nickerson C, Coleman L, Rutter MD, von Wagner C. Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests. *Gut.* December 7, 2011 2011.

12. Fletcher SW. Breast Cancer Screening: A 35-Year Perspective. *Epidemiologic Reviews.* July 1, 2011 2011;33(1):165-175.

13. Ward E, Halpern M, Schrag N, et al. Association of insurance with cancer care utilization and outcomes. *CA Cancer J Clin.* Jan-Feb 2008;58(1):9-31.

14. Jemal A, Ward E, Wu X, Martin HJ, McLaughlin CC, Thun MJ. Geographic Patterns of Prostate Cancer Mortality and Variations in Access to Medical Care in the United States. *Cancer Epidemiology Biomarkers & Prevention.* March 1, 2005 2005;14(3):590-595.

15. Deppen SA, Aldrich MC, Hartge P, et al. Cancer Screening: The Journey from Epidemiology to Policy. *Annals of Epidemiology.* 2012;22(6):439-445.

16. Gould MK, Fletcher J, Iannettoni MD, et al. Evaluation of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest.* Sep 2007;132(3 Suppl):108S-130S.

17. Gould M. Cost Effectiveness of Positron Emission Tomography for Characterizing Pulmonary Nodules. *PET Clinics.* 2006;1(4):339 - 346.

18. Gambhir SS, Czernin J, Schwimmer J, Silverman DH, Coleman RE, Phelps ME. A tabulated summary of the FDG PET literature. *J Nucl Med.* May 2001;42(5 Suppl):1S-93S.

19. Deppen S, Putnam JB, Jr., Andrade G, et al. Accuracy of FDG-PET to diagnose lung cancer in a region of endemic granulomatous disease. *The Annals of Thoracic Surgery.* Aug 2011;92(2):428-432; discussion 433.

20. Grogan E, Deppen SA, Ballman K, et al. Accuracy of FDG-PET to diagnose lung cancer in the ACOSOG Z4031 trial. Paper presented at: Alliance for Clinical Trials in Oncology; June 29-July1, 2012, 2012; Chicago, IL

21. Croft DR, Trapp J, Kernstine K, et al. FDG-PET imaging and the diagnosis of non-small cell lung cancer in a region of high histoplasmosis prevalence. *Lung Cancer.* 2002;36(3):297-301.

22. Aberle DR, Berg CD, Black WC, et al. The National Lung Screening Trial: overview and study design. *Radiology.* Jan 2011;258(1):243-253.

23. Silvestri GA. Screening for Lung Cancer: It Works, but Does It Really Work? *Annals of Internal Medicine.* September 5, 2011 2011:E-364.

24. Trustees Bo, Tavenner MB. 2012 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplemental Medical Insurance Trust

Funds. In: Services HaH, ed. Washington DC: Center for Medicare Services; 2012.

25. Howlader N, Noone A, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2009. *SEER Cancer Statistics Review* 2012; http://seer.cancer.gov/csr/1975_2009_pops09/index.html. Accessed 2/22/2013, 2013.

26. Research data distribution center. Medicare provider analysis and review (MEDPAR) record--dictionary for SAS and CVS datasets. In: Services CfM, ed: CMS; 2009.

27. Edwards LB, Acquaviva FA, Livesay VT. Further Observations on Histoplasmin Sensitivity in the United States. *American Journal of Epidemiology.* 1973;98(5):315-325.

28. Edwards L, Acquaviva F, Livesay V, Cross F, Palmer C. An altas of sensitivity to tuberculin, PPD-B and histoplasmin in the United States. *American Review of Respiratory Disease.* 1969;99(4):S1-132.

29. Baddley J, Winthrop K, Patkar N, et al. Geographic Distribution of Endemic Fungal Infections among Older Persons, United States. *Emerging Infectious Diseases.* 2011;17(9):1664-1669.

30. Institute NC. Tobacco Use Supplement from Current Population Survey. *TUS-CPS* 2010; http://riskfactor.cancer.gov/studies/tus-cps/info.html. Accessed 01/09/2013.

31. Jemal A, Thun M, Yu XQ, et al. Changes in Smoking Prevalence among U.S. Adults by State and Region: Estimates from the Tobacco Use Supplement to the Current Population Survey, 1992-2007. *BMC Public Health.* Jun 29 2011;11(1):512.

32. Thun MJ, Jemal A. How much of the decrease in cancer death rates in the United States is attributable to reductions in tobacco smoking? *Tob Control.* Oct 2006;15(5):345-347.

33. Kozower BD, Meyers BF, Reed CE, Jones DR, Decker PA, Putnam Jr JB. Does Positron Emission Tomography Prevent Nontherapeutic Pulmonary Resections for Clinical Stage IA Lung Cancer? *The Annals of Thoracic Surgery.* 2008;85(4):1166-1170.

34. Kuo E, Bharat A, Bontumasi N, et al. Impact of Video-Assisted Thoracoscopic Surgery on Benign Resections for Solitary Pulmonary Nodules. *The Annals of Thoracic Surgery.* 2012;93(1):266-273.

35. Dilts D. Practice variation: the Achilles' heel in quality cancer care. *J Clin Oncol.* 2005;23:5881 - 5882.

36. Kozower BD, Sheng S, O'Brien SM, et al. STS database risk models: predictors of mortality and major morbidity for lung cancer resection. *Ann Thorac Surg.* Sep 2010;90(3):875-881; discussion 881-873.

37. Shapiro M, Swanson SJ, Wright CD, et al. Predictors of major morbidity and mortality after pneumonectomy utilizing the Society for Thoracic Surgeons General Thoracic Surgery Database. *Ann Thorac Surg.* Sep 2010;90(3):927-934; discussion 934-925.

38. Wright CD, Gaissert HA, Grab JD, O'Brien SM, Peterson ED, Allen MS. Predictors of Prolonged Length of Stay after Lobectomy for Lung Cancer: A Society of Thoracic Surgeons General Thoracic Surgery Database Risk-Adjustment Model. *The Annals of Thoracic Surgery.* 2008;85(6):1857-1865.

39. Goodney PP, Lucas FL, Stukel TA, Birkmeyer JD. Surgeon Specialty and Operative Mortality With Lung Resection. *Annals of Surgery.* 2005;241(1):179-184.

40. Starnes SL, Reed MF, Meyer CA, et al. Can lung cancer screening by computed tomography be effective in areas with endemic histoplasmosis? *The Journal of Thoracic and Cardiovascular Surgery.* 2011;141(3):688-693.

41. Swan J, Breen N, Coates RJ, Rimer BK, Lee NC. Progress in cancer screening practices in the United States. *Cancer.* 2003;97(6):1528-1540.

42. Aspy CB, Enright M, Halstead L, Mold JW. Improving Mammography Screening Using Best Practices and Practice Enhancement Assistants: An Oklahoma Physicians Resource/Research Network (OKPRN) Study. *The Journal of the American Board of Family Medicine.* July 1, 2008 2008;21(4):326-333.

43. Wahidi MM, Govert JA, Goudar RK, Gould MK, McCrory DC. Evidence for the treatment of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest.* Sep 2007;132(3 Suppl):94S-107S.

44. Bach P. NSCLC - Screening Guidelines. *NCCN.* 2008.

45. Gould MK, Maclean CC, Kuschner WG, Rydzak CE, Owens DK. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis. *Jama.* Feb 21 2001;285(7):914-924.

46. Cronin P, Dwamena B, Kelly A, Bernstein S, Carlos R. Solitary pulmonary nodules and masses: a meta-analysis of the diagnostic utility of alternative imaging tests. *European Radiology.* 2008;18(9):1840-1856.

47. Centers for Medicare & Medicaid Services. U.S. Department of Health and Human Services - Medicare National Coverage Determinations Manual. http://www.cms.gov/manuals/downloads/ncd103c1_Part4.pdf. Accessed 9/14/2011, 2011.

48. Berger M, Gould MK, Barnett PG. The Cost of Positron Emission Tomography in Six United States Veterans Affairs Hospitals and Two Academic Medical Centers. *American Journal of Roentgenology.* 2003/08/01 2003;181(2):359-365.

49. Nguyen BD. F-18 FDG PET/CT Imaging of Disseminated Coccidioidomycosis. *Clinical Nuclear Medicine.* 2006;31(9):568-571.

50. Detterbeck FCMD, Khandani AHMD. The Role of PET Imaging in Solitary Pulmonary Nodules. *Clinical Pulmonary Medicine.* 2009;16(2):81-88.

51. Grogan EL, Weinstein JJ, Deppen SA, et al. Thoracic Operations for Pulmonary Nodules Are Frequently Not Futile in Patients with Benign Disease. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer.* Oct 2011;6(10):1720-1725.

52. Bryant AS, Cerfolio RJ. The Maximum Standardized Uptake Values on Integrated FDG-PET/CT Is Useful in Differentiating Benign From Malignant Pulmonary Nodules. *The Annals of Thoracic Surgery.* 2006;82(3):1016-1020.

53. Kauffman CA. Histoplasmosis: a Clinical and Laboratory Update. *Clinical Microbiology Reviews.* January 2007 2007;20(1):115-132.

54. Coleman RE. PET in Lung Cancer. *J Nucl Med.* May 1, 1999 1999;40(5):814-820.

55. Sathekge M, Maes A, Pottel H, Stoltz A, van de Wiele C. Dual time-point FDG-PET-CT for differentiating benign form malignant solitary pulmonary nodules in a TB endemic area. *South African Medical Journal.* 2010;100(9):598-601.

56. Higashi K, Ueda Y, Sakuma T, et al. Comparison of [(18)F]FDG PET and (201)Tl SPECT in evaluation of pulmonary nodules. *J Nucl Med.* Oct 2001;42(10):1489-1496.

57. Harpole D, Ballman, K. V., Oberg, A.L., Whiteley, G, Cerfolio, R., Keenan, R., Jones, D.R., D'Amico, T.A., Schrager, J., Putnam, B. Proteomic analysis for detection of NSCLC: Results of ACOSOG Z4031. *J Clin Oncol* 2011 2011;29:Suppl: abstr 7003.

58. Bunyaviroch T, Coleman RE. PET Evaluation of Lung Cancer. *J Nucl Med.* March 1, 2006 2006;47(3):451-469.

59. U.S. Census Bureau GD. ZIP Code Tabulation Areas (ZCTAs). 2011; http://www.census.gov/geo/ZCTA/zcta.html. Accessed 09/30/2011.

60. U.S. Census Bureau GD. Guide to Census Tract Resources. *2000 Census Data for the United States* 2012; http://www.census.gov/geo/www/tractez.html. Accessed 10/1/2012.

61. Rivera MP, Mehta AC. Initial diagnosis of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest.* Sep 2007;132(3 Suppl):131S-148S.

62. Silvestri GA, Gould MK, Margolis ML, et al. Noninvasive staging of non-small cell lung cancer: ACCP evidenced-based clinical practice guidelines (2nd edition). *Chest.* Sep 2007;132(3 Suppl):178S-201S.

63. Delbeke D, Coleman RE, Guiberteau MJ, et al. Procedure Guideline for Tumor Imaging with 18F-FDG PET/CT 1.0. *Journal of Nuclear Medicine.* May 2006 2006;47(5):885-895.

64. Cronin P, Dwamena B, Kelly A, Carlos R. Solitary pulmonary nodules: meta-analytic comparison of cross-sectional imaging modalities for diagnosis of malignancy. *Radiology.* 2008;246:772 - 782.

65. Goldstraw PF, Crowley JP, Chansky KMS, et al. The IASLC Lung Cancer Staging Project: Proposals for the Revision of the TNM Stage Groupings in the Forthcoming (Seventh) Edition of the TNM Classification of Malignant Tumours. *Journal of Thoracic Oncology.* 2007;2(8):706-714.

66. Lauer MS, Murthy SC, Blackstone EH, Okereke IC, Rice TW. [18F]Fluorodeoxyglucose Uptake by Positron Emission Tomography for Diagnosis of Suspected Lung Cancer: Impact of Verification Bias. *Arch Intern Med.* January 22, 2007 2007;167(2):161-165.

Table 1. Characteristics of MEDPAR inpatient discharges with lung surgery, United States 2009

Characteristic	Cancer N=23050	Benign Disease N=2,328
Male sex (% of diagnosis)	**11,619 (49.6)**	**1,266 (45.6)**
White race (%)	**20,734 (90.0)**	**2,063 (88.6)**
Age category		
<65	**2,129 (9.4)**	**423 (10.7)**
65-69	**6,088 (26.4)**	**687 (26.7)**
70-74	**5,989 (26.0)**	**591 (25.9)**
75-79	**5,011 (21.7)**	**389 (21.3)**
80-84	**2,950 (12.8)**	**177 (12.3)**
>84	**883 (3.8)**	**61 (3.7)**

Table 2. Descriptive Characteristics of ACOSOG Z4031 patients with FDG-PET Scans

Characteristic	Cancer N=566	Benign N=116	p-value[a]
Male (%)	253 (45)	54 (47)	0.71
Caucasian (%)	517 (91)	113 (97)	0.03
Mean Age (SD)	67 (10)	61 (11)	<0.001
Lesion Size mm (SD)	26 (0.61)	20 (0.95)	<0.001
FDG-PET Avid[b] (%)	465 (82)	80 (69)	0.002

[a] Continuous variable statistics use t-test (Age and Lesion Size) and binomial proportions test for differences in proportions (Gender, Race and FDG-PET Avidity).

[b] The categories of avidity and their corresponding SUV are: Not avid/Not cancerous (SUV=0), Low avidity/Not likely cancerous (SUV 0.1 to 2.5), Avid/Likely cancerous (SUV 2.5 to 5) and Highly avid/Cancerous (SUV>5). PET avid was the sum of Avid/ likely cancerous and Highly avid/Cancerous (SUV≥2.5).

Table 3. Pathology of false negative and false positive lesions

Malignant	**FDG-PET Non-Avid (False Negatives) ***
Adenocarcinoma	62
Squamous Cell	11
Bronchoalveolar Cell	11
Carcinoid/Neuroendocrine	9
Other NSCLC	4
Other Cancer	1
Small Cell	1
Unknown	2
Benign	**FDG-PET Avid (False Positives)**
Granuloma**	55
Benign Tumor	8
Active Infectious disease***	9
Fibrosis	4
Other	4

***11 of the false negatives were <1cm

**Granuloma includes histoplasmosis, atypical mycobacteria, blastomycosis, cryptococcus, coccidiodomycosis, aspergillosis and nonspecific granulomas.

***Infectious disease includes active Mycobacterium tuberculosis and active pneumonia

Table 4. Accuracy of FDG-PET to diagnose cancer among patients with clinical stage1 NSCLC

FDG-PET	Cancer[a]	Benign	
Avid[b]	**465**	**80**	**PPV[c] 85%** 95%CI: (82, 88)
Not-Avid	**101**	**36**	**NPV[d] 26%** 95%CI: (19, 35)
			Prevalence 83% 95%CI: (80, 86)

Sensitivity 82% 95%CI: (79, 85)

Specificity 31% 95%CI: (23, 40)

[a]Diagnosis was based upon pathological result of the surgically resected specimen.
[b]FDG-PET avidity was defined by an SUV > 2.5 or moderate or intense uptake.
[c]PPV = positive predictive value
[d]NPV = negative predictive value

Table 5. FDG-PET sensitivity and specificity by enrolling city with at least 25 participants

	N	Sensitivity (%)	Cancer	Specificity (%)	Benign
	462		**N=378**		**N=84**
Durham, NC	41	91	33	25	8
Birmingham, AL	111	89	98	15	13
Philadelphia, PA	78	85	66	46	12
Pittsburg, PA	68	78	60	25	8
Charlottesville, VA	52	76	34	33	18
Cincinnati, OH	31	73	22	33	9
St. Louis, MO	54	68	47	29	7
Los Angeles, CA	27	67	18	44	9
Chi-square test		p = 0.03		p = 0.72	

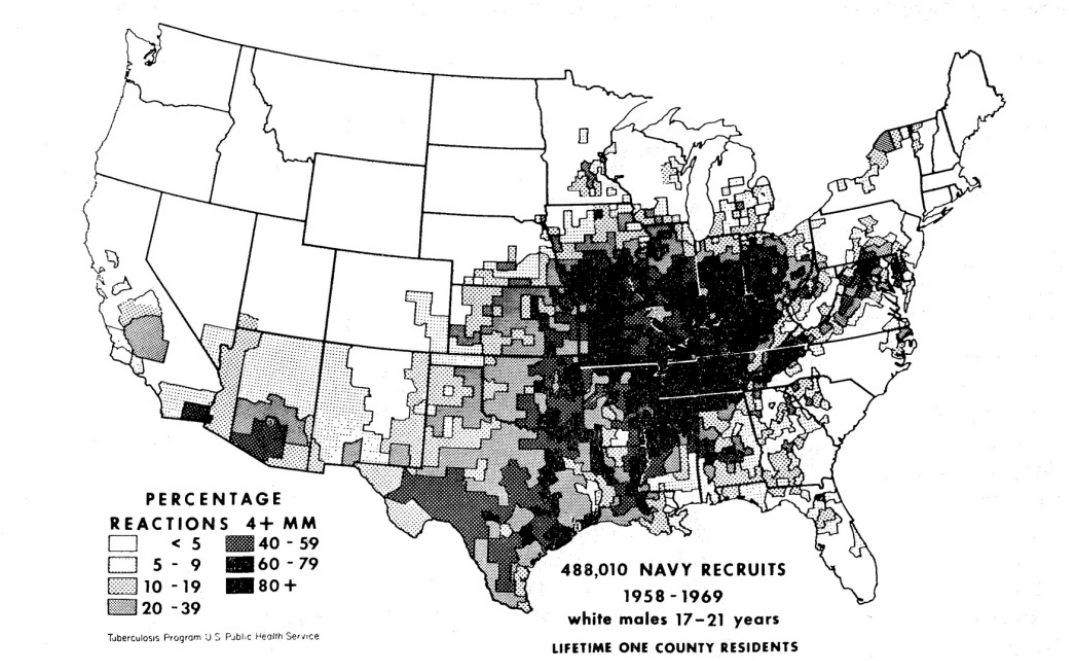


Figure 1. Fungal lung disease prevalence, histoplasmin skin was cross reactive with coccidioidomycosis and blastomycosis (from Edwards et al, 1969)

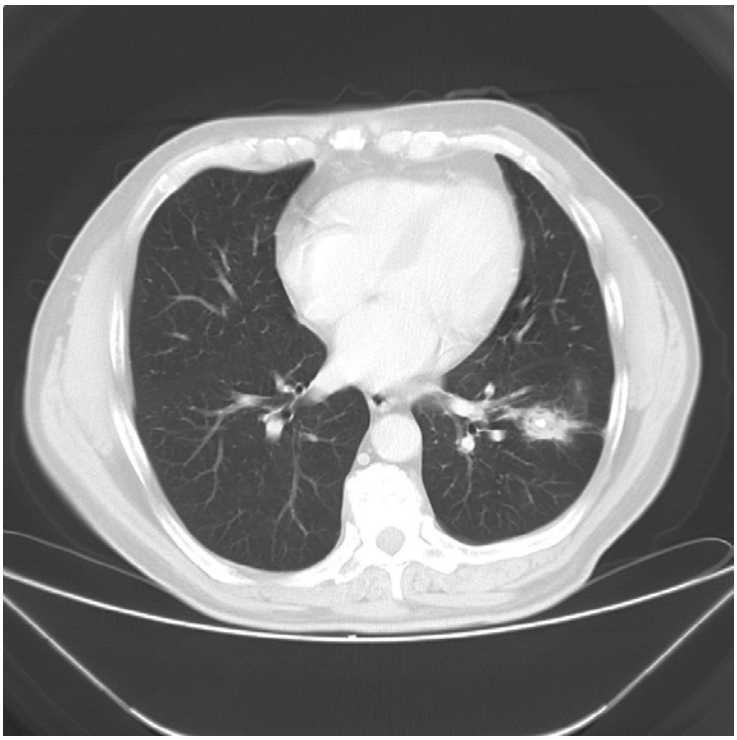


Figure 2. CT scan of a spiculated granuloma in a patient presenting with hemoptysis

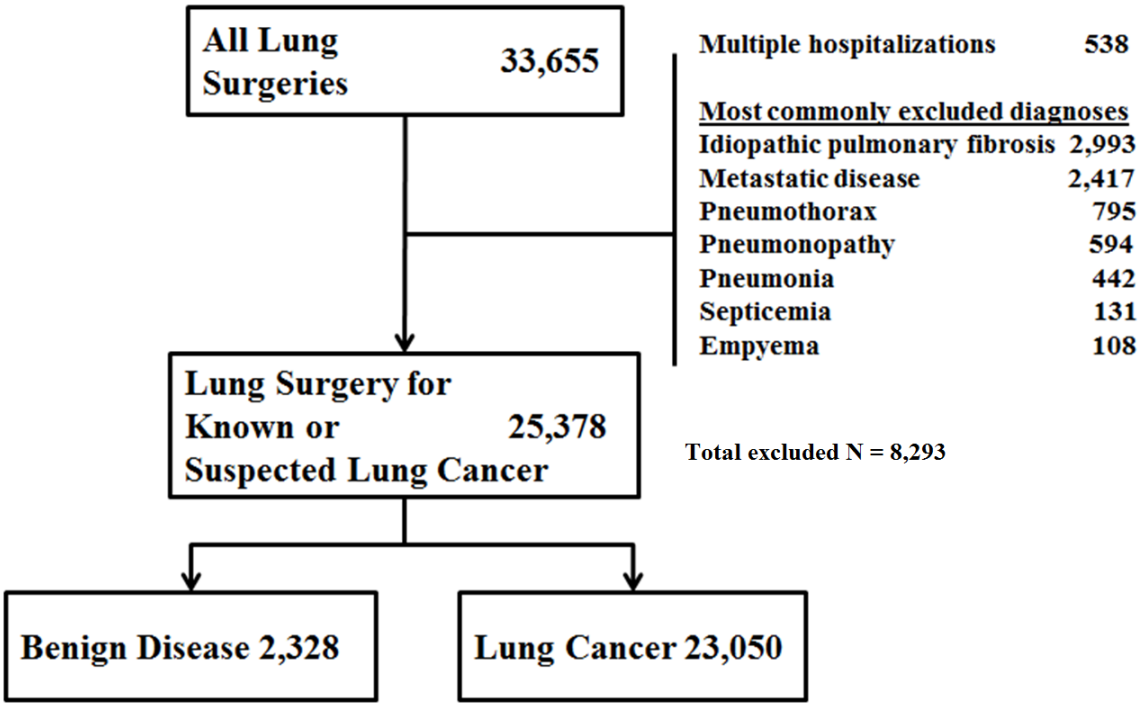


Figure 3. Consort diagram of benign disease after surgery for known or suspected lung cancer, 2009 MEDPAR

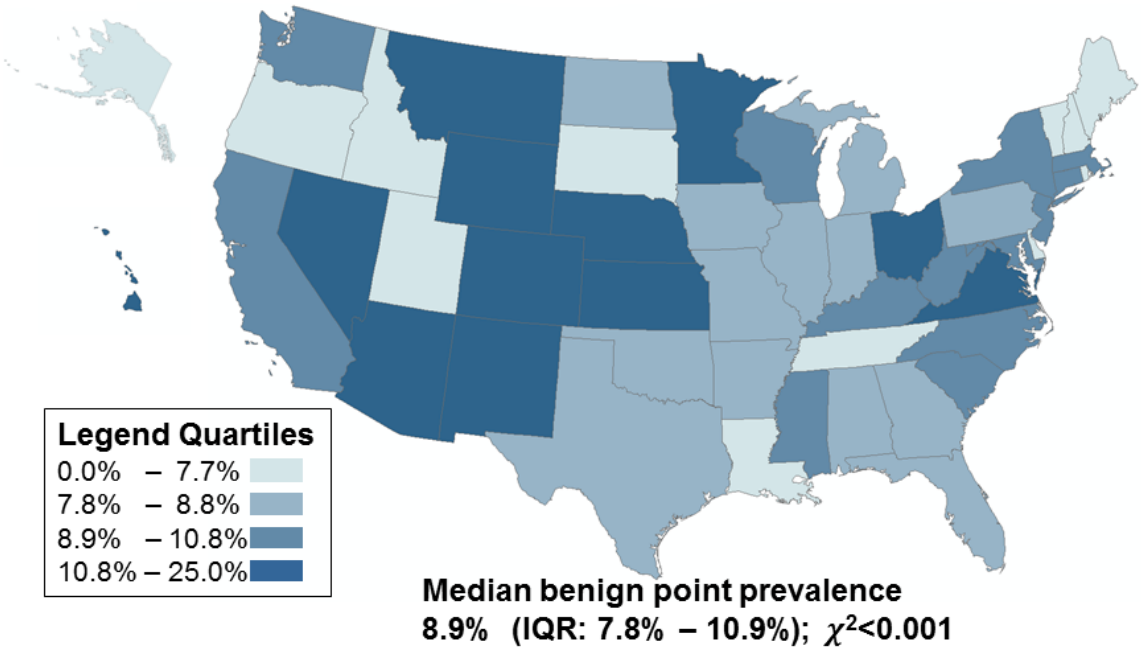


Figure 4. Point prevalence of benign disease after lung surgery by state in 2009 MEDPAR

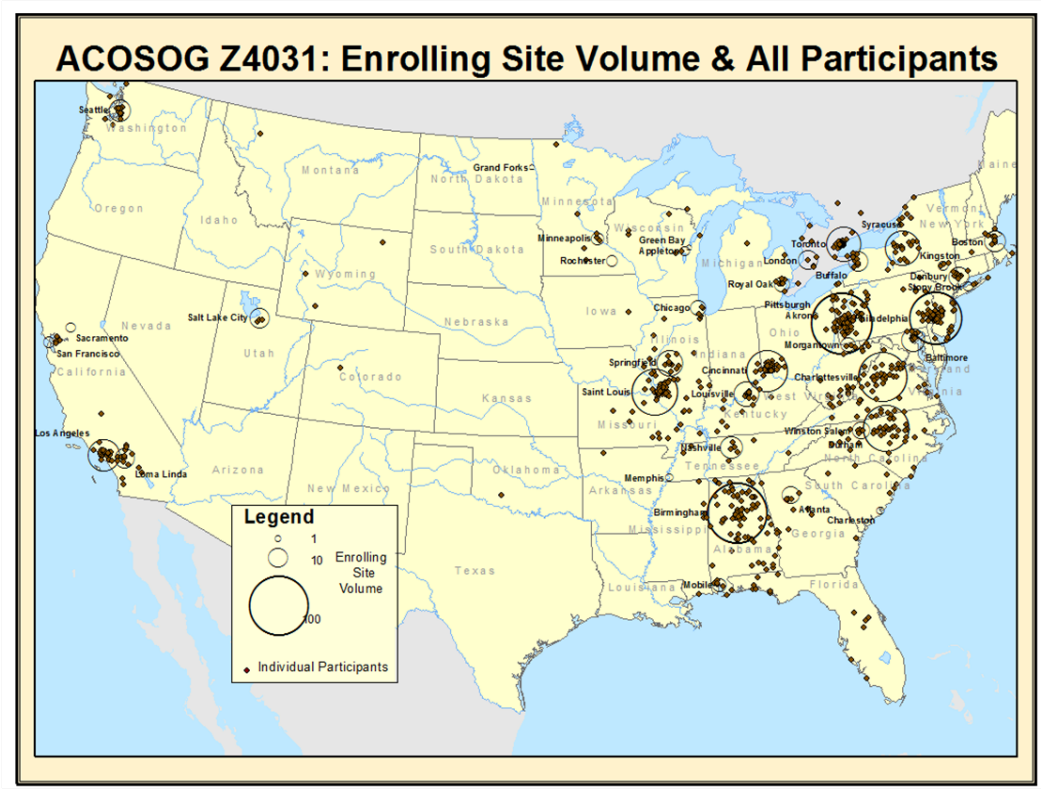


Figure 5. Enrolling site location with size of circle corresponding to participation volume – 51 sites in 39 cities. Individual dots are study participants residence by zip code at time of enrollment. Dots are overlapping for those with identical zip codes.

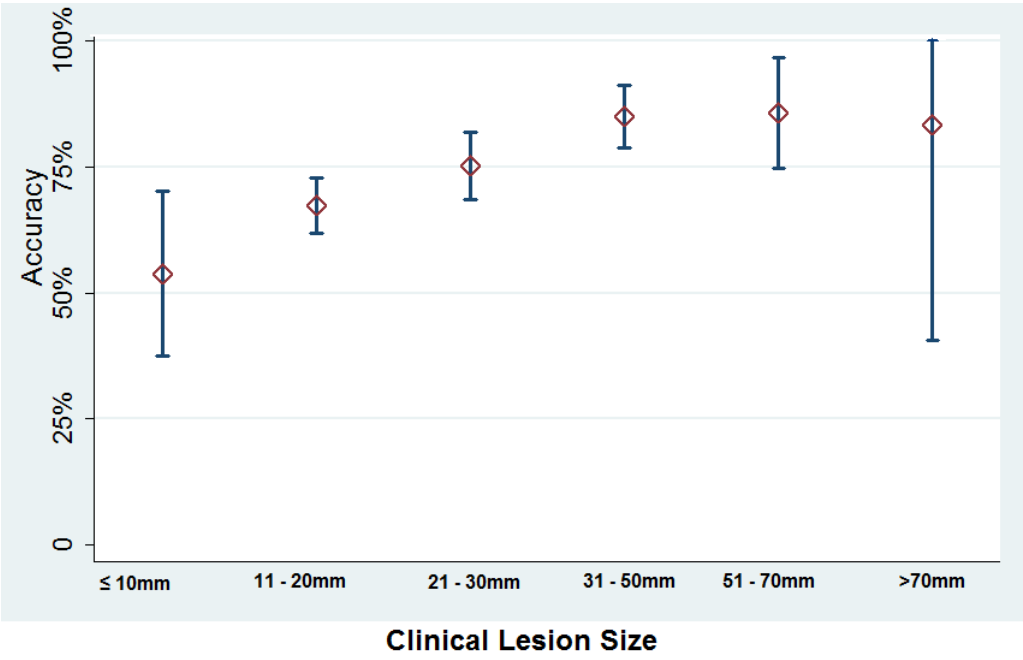


Figure 6. Accuracy of FDG-PET to diagnose lung cancer by lesion size in millimeters. Accuracy = (True Positives + True Negatives) / Total Population in size group

Chapter 4

IV. Variation in FDG-PET Accuracy to Diagnose Lung Cancer: a meta-analysis

4.1 Introduction

Clinicians rely heavily on radiographic imaging to discover and non-invasively diagnose lung nodules between 3 and 30mm in diameter. Lung nodules smaller than 30 mm are frequently asymptomatic and discovered incidentally from non-invasive imaging for other medical indications. One-third of these incidentally discovered nodules are malignant in persons over age 35.[1] Guidelines for the management of lung nodules suggest relying on imaging to assess the risk that a lung nodule is cancerous and using increasingly intensive and invasive tests as the likelihood of cancer increases.[2,3] As described in previous chapters, if the lesion is larger than 8 mm and the individual's assessed risk for lung cancer is between 5% and 60%, then F^{18}-Fluorodeoxyglucose positron emission tomography combined with helical CT (FDG-PET/CT) is suggested for non-invasive imaging to characterize the lesion's metabolic activity.

FDG is a glucose analog (deoxyglucose) attached to an 18-Fluorine radionuclide. When injected into an individual, the glucose is consumed by metabolically active tissue and the radionuclide preferentially accumulates in the active tissue. Metabolically active organs like the heart, brain and gonads concentrate the radionuclide.[4] Positron emissions from the radionuclide are detected by a positron emission tomographic scanner ring. In a modern combination PET/CT scanner, the PET image is combined with a CT image that adds anatomical structures to the PET image. Normal lung tissue is not metabolically active and thus does not accumulate the radionuclide. Neoplastic cells, pneumonia and other inflammatory diseases, wounds, and active benign tumors like hamartomas and granulomas have higher

glucose metabolism than the surrounding normal tissue and thus accumulate the 18F radionuclide attached to the glucose analog within the lung.[5,6] This differential in normal, compared to abnormal, metabolism makes FDG-PET scans useful in the diagnosis of lung cancer.[4] Less active tissue including slow-growing adenocarcinomas or carcinoid tumors do not concentrate the radionuclide. Lesions with low metabolic activity and lesions smaller than 8 mm commonly generate false negative scans.

FDG-PET is 90% to 94% accurate in the characterization of cancer or benign disease in lung nodules, with a sensitivity of 96% to 99% and specificity of 78% to 82% according to previous meta-analyses.[7,8] FDG-PET has been demonstrated to reduce non-therapeutic resections (e.g. resection for benign lesions or metastatic disease) by 17% to 20%[9-12]. For these reasons, FDG-PET is widely accepted for the clinical diagnosis and staging of lung cancer in patients with suspicious lung nodules.[13,14]

Since the publication of the meta-analysis in 2001, FDG-PET has been adopted worldwide for the clinical diagnosis and staging of lung cancer. When a test becomes widely adopted and is applied outside the controlled environment of clinical trials, the diagnostic accuracy of the test is usually diminished.[15] The observed decrease in accuracy may be due to practice variation and poor quality control. Poorer accuracy may also arise in sub-populations within which some test-influencing factor is prevalent. If the confounding factor was not prevalent in the original studies measuring the diagnostic accuracy of the test, then a form of selection bias may have occurred in the original research.

Recent studies observed reduced FDG-PET accuracy in diagnosing lung cancer in patients with lung lesions where histoplasmosis and other fungal lung diseases are endemic.[16,17] Histoplasmosis, coccidioidomycosis and blastomycosis are the most prevalent fungal lung diseases in the US and are common etiologies of lung granulomas. Histoplasmosis and blastomycosis are endemic across much of the Mississippi, Ohio and Missouri River valleys through southern Ontario while coccidioidomycosis is

prevalent in the southwest US.[18] Some international studies also found reduced FDG-PET accuracy in diagnosing lung cancer.[19,20] These studies occurred in areas of endemic tuberculosis which is the primary cause of lung granulomas outside of North America.[18,21] We undertook a systematic review and meta-analysis of publications describing FDG-PET accuracy to diagnose lung cancer among patients being evaluated with lung nodules or masses and published since the earlier meta-analysis by Gould and colleagues.[8] This meta-analysis investigated whether FDG-PET accuracy varies between studies according to characteristics associated with the tests, settings, participants or methodology. Our meta-analysis also specifically investigated possible variation in diagnostic test accuracy caused by locally endemic infectious lung diseases in study participants.

4.2 Methods

Study Selection

Studies evaluating individuals for possible lung cancer using FDG-PET or combined FDG PET/CT scans were reviewed, including published and unpublished studies. We searched Medline using the Pubmed interface, Embase and the New York Academy of Medicine Grey Literature Report. The literature search in each database included any of the terms lung cancer, pulmonary nodule, lesion, non-small cell lung cancer, or NSCLC. From this set of publications additional descriptors were required that included any of the terms diagnostic, positron emission tomography, PET, fluorodeoxyglucose, FDG, or combinations of those terms (see Appendix 2). Unpublished abstracts were reviewed and extracted by a research librarian. Publication could be in any language, but the abstract had to appear in English in one of the above databases to be included. Bibliographies from meta-analyses and literature reviews were examined individually and papers of interest were included in the final list of abstracts for review. A search was conducted for published studies between October 2000 and March 2011. Complete citations including authors, Pubmed identification number, abstract and year of publication were imported electronically into a dedicated REDCap database for abstract review and data extraction.

Exclusion Criteria

All published abstracts were examined independently by two reviewers for inclusion. If either investigator deemed a study worthy of consideration after reviewing the abstract, then the publication received subsequent review and data extraction. Studies with 100% cancer prevalence or 100% benign disease prevalence were excluded. Studies reporting staging using FDG-PET and not reporting the results of lung nodule scans were excluded. Case reports were excluded. Studies with 10 or fewer participants were also excluded. Those studies not reporting enough information to determine the number of true positive, true negative, false positive and false negative results were excluded. Studies that preselected specific histologies, tumor characteristics in imaging including minimum standard uptake value, ground glass opacity observed in CT prior to FDG-PET scan, or lesion location in the lung were excluded. Studies using only gamma camera PET scanners were excluded. One study could be excluded for multiple reasons. Unpublished abstracts not excluded based on the above criteria but could be matched to a subsequent publication were excluded. Publications whose population, in part or in whole, was included in multiple studies were excluded so that the population under study contributed only once to the meta-analysis. Authors were contacted to determine the uniqueness of a study population when multiple studies that possibly included the same population were observed and eligibility was uncertain.

Study quality

Two reviewers independently assessed the quality of each study according to prospective criteria using a modified QUADAS set of 12 questions.[22,23] The questions addressed the technical quality of the index test, the technical quality of the reference test, the independence and accuracy of the test interpretation, and the sample size and population representation. To evaluate agreement between the raters for assessments of study eligibility, we calculated the observed percentage of agreement and the kappa coefficient for inter-rater reliability. Study quality was graphically reviewed and subanalysis was conducted on prospective versus retrospective studies. Studies that used only pathological diagnosis were

compared to studies that used both pathological and radiographic determination of diagnosis, and studies that blinded reviewers to patient demographic and history were compared to those that did not.

Data Synthesis and Analysis

After abstract review, articles designated for full text review and data extraction were independently coded into a database by two reviewers. All articles were reviewed and coded by one investigator (SD). Two other investigators (CK and AM) divided the articles into two groups and independently reviewed the full text and extracted relevant data. Each publication was reviewed independently twice. Discrepancies in coding were reviewed by the investigators and an independent clinician. Changes were agreed upon by consensus. Additional data extraction beyond citation, authors and year of publication included: type of population studied, method of determining final diagnosis, method of determining benign disease, scanner type, patient demographics, reported presence of infectious lung diseases, infectious lung disease prevalence, and country or region where the participant population was recruited (see Appendix 2 for database metadata). The methodology of designating the PET scan results included the metric of measuring FDG-PET avidity, which was either standard uptake value (SUV), modified SUV, radiologist's assessment, or other. The number of levels of risk or avidity and the SUV threshold used to differentiate benign and cancerous diagnosis of disease were recorded. For each study a 2x2 contingency table of test and disease result was created. Sensitivity and specificity, diagnostic odds ratio, and positive and negative likelihood ratios for each article were estimated, and 95% confidence intervals were calculated based upon normal approximations to the binomial distribution.

FDG-PET test performance was estimated in a pooled fashion using forest plots, and hierarchical summary receiver operator curves (SROC) were generated. Study heterogeneity was quantitatively measured by Cochrane Q and I^2 statistics and assessed graphically by forest plot and SROC curve. Publication bias was graphically charted by funnel plot and quantitatively measured by Deek's Asymmetry Test. Sub-group analysis was conducted by forest plots.

Considerable heterogeneity is expected in diagnostic studies and a meta-analysis model was created to summarize test performance by using a bivariate random-effects binomial regression model. This model was formulated so that the test results were conditioned on the probability of disease. This approach allowed fixed and random effects modeling of clinically relevant variables. For variables with missing data, multiple imputation with chained equations was performed. Predictive mean matching models estimated all missing data. An imputed dataset burn-in was used and an additional 10 imputed datasets were then generated. These 10 datasets were used for all subsequent model estimation.

Study characteristics likely to generate heterogeneity were chosen for sub-group analysis and included in the meta-analysis model as linear effects within the random and fixed effect model. Those characteristics included: endemic infectious lung disease in the study population, mean or median lesion diameter less than 2 cm, the method of diagnosis, method of blinding scan readers, and scanner type as a categorical variable. Other study characteristics measuring relative quality were examined in sensitivity analysis to determine if they materially changed the results of the meta-analysis model.

Variables measuring domains of study quality examined in sensitivity analysis included whether the study was prospective or retrospective, the method of final diagnosis and year of publication. Publication year was examined in both a linear and non-linear fashion. To determine whether the diagnostic accuracy of FDG-PET has changed over time, we ordered the studies chronologically, estimated each study's diagnostic log odds ratio and performed a cumulative meta-analysis.[24] The diagnostic odds ratio is defined as true positives divided by the false negative results; this fraction is then divided by the ratio of false positive to true negative results (TP/FN) / (FP/TN). Analysis of clinical relevance was conducted by estimating positive and negative likelihood ratios for the combined studies and plotting the estimated likelihoods with 95% confidence intervals graphically. All analysis was performed with STATA (v12, College Station, TX).

4.3 Results

Study identification

1,218 articles and 33 meta-analyses or reviews were found. An additional 13 articles were added from the meta-analyses or reviews for a total of 1,231 articles that met the search criteria. Upon initial abstract review, 1013 articles were excluded. An article could be excluded for multiple reasons, but the most common reason for exclusion during either portion of the review was 100% cancer prevalence (637) (Figure 1).

Five abstracts were included from foreign language journals and four of these were translated into English.[25-28] The remaining abstract that was not translated had enough information in the English abstract to estimate sensitivity and specificity and test accuracy.[29] One unpublished conference abstract was included in the initial review and its later publication was used in the secondary review.[16] Two hundred and eighteen studies received full review and 158 were excluded upon this secondary review. The remaining 60 studies met all inclusion criteria and were used for final analysis. The total number of participants among the 60 studies was 6,347 and median study participant size was 75 (IRQ: 47, 127). Cancer prevalence among all 6,347 participants was 63.8%. Individual study cancer prevalence varied from 21% to 90% with a median prevalence of 62.6% across studies. Twenty-five of the 60 studies (42%) were conducted prospectively. Seven of 60 studies documented endemic infectious lung disease in the population scanned.[16,17,20,30-33]

In a pooled analysis of all 60 studies using an unadjusted random effects model, sensitivity was 89% (95% CI: 87%, 91%) and specificity was 74% (95% CI: 70%, 78%) (Figure 2). Pooled diagnostic odds ratio was 24 (95%CI: 18, 33). The unadjusted area under the hierarchical summary receiver operator curve was 0.90 (95%CI: 0.87, 0.93) (Figure 3). Significant heterogeneity beyond that explained by

differences in thresholds chosen for diagnosis within studies was observed across studies for both sensitivity, with I^2 of 81% (95%CI: 77%, 86%), and specificity, with an I^2 of 76% (95%CI: 70%, 82%). Therefore, estimates of pooled sensitivity and specificity and the SROC should be interpreted with caution as they did not reflect differences in population and the assumptions for pooling data were violated. Heterogeneity of diagnostic odds ratio measured by I^2 was 98 (95%CI; 95, 99) and an estimated 5% of observed variability in sensitivity and specificity likely arose from variation in cutpoints across the 60 studies. I^2 values greater than 75% are indicative of strong heterogeneity between studies.[34] Published studies of FDG-PET scan accuracy included small studies reporting low to moderate accuracy, and Deeks' asymmetry test was not statistically significant for publication bias (p = 0.18) (Figure 4).

Differences in FDG-PET accuracy over time were examined with a cumulative analysis of the diagnostic odds ratio accumulated over the years of publication (Appendix 2, Figure 1). Studies were grouped into four three-year populations of studies and their diagnostic odds ratios pooled. No statistically significant differences in accuracy measured by diagnostic odds ratio were observed between any two groups of studies. A bivariate boxplot was constructed to examine the interrelationship between sensitivity and specificity in the unadjusted random effects model (Appendix 2, Figure 2). The slight oblong distribution of the covariance region indicates a slight preference for sensitivity of FDG-PET scan over specificity and an asymmetric SROC.

Meta-analysis model

A random-effects logistic regression model was created. Heterogeneity between studies remained (rho =0.08 95%CI: 0.04, 0.12) after accounting for variability between studies caused by whether the study reported granulomas arising from endemic infectious lung disease or the scanner type, whether or not readers were blinded to patient information, and the average size of the lung lesion in the cohort being less than or equal to 20 mm (Table 2). Missing data occurred in 25 (42%) studies. Multiple imputation was performed with chained equations. Missing data occurred in a monotone pattern between the two (3%

of the 60 studies) reported neither blinding method nor a mean or median lesion size.[35,36] The multiple imputation model was estimated using pathological diagnosis of cancer or benign disease, the year of the study, the method of diagnosis, scanner type, study specific cancer prevalence, whether the study was retrospective or prospective, and whether the study reported locally endemic infectious lung disease. The point estimates of the meta-analysis model variables using complete case data were generally higher than those from the imputed data, but the overall results were similar (Table 2).

Odds ratios for the interaction terms with cancer that are greater than one increase sensitivity; odds ratios less than one decrease sensitivity. Odds ratios for variables without interaction with cancer influence specificity inversely. Therefore, odds ratios increased specificity when their values were less than one, and odds ratios decreased specificity when their odds ratios were greater than one. The variable's overall effect on sensitivity was determined by the sum of coefficients for direct effect and for the interaction between the variable and a cancer diagnosis. For example, studies using combined PET/CT scanners had higher specificity (OR 0.63; 95%CI: 0.43, 0.94) and sensitivity (OR 1.47; 95%CI: 1.21, 2.31) when compared to studies that used PET only variables, the method of blinding scan readers (17%), and mean or median lesion size (23%). Two studies scanners. The pooled sensitivity from the meta-analysis model was 89% (95%CI: 82%, 94%). Estimated pooled specificity was 75% (95%CI: 46%, 86%).

Sub-group analysis

Seven studies reporting infectious lung disease endemic to the local population had significantly lower specificity, 59% (95%CI: 46%, 70%; $p<0.001$)[16,17,20,30-33] compared to a specificity of 77% (95%CI: 70%, 86%) in the remaining 53 studies (Figure 5). A sensitivity of 91% (95%CI: 90%, 93%) for the endemic disease studies were slightly higher compared to the other 53 studies, specificity 90% (95%CI 82%, 93%). One study by Chundru and colleagues was an outlier in terms of cutpoint choice.[30] -The Bryant paper exerted significant influence due to its large size (N=585) compared to the remaining 6 studies that had a combined population of 606[32]. Excluding these two studies resulted in a sensitivity of 90% (95%CI

87%, 93%) that was the same as the 53 studies that did not report endemic infectious lung disease. The remaining 5 studies had a much lower specificity (39%; 95%CI: 29%, 51%) and excluding these two studies from the pool of studies that reported endemic infectious lung disease removed all heterogeneity for sensitivity and reduced heterogeneity for specificity measured by I^2 by half to 43%. The SROC for these five studies was 0.88 (95%CI: 0.85, 0.91) (Figure 6).

Among the 24 studies reporting average or median lesion diameter less than or equal to 20mm, the estimated sensitivity was 88% (95%CI: 82%, 91%).[20,26,28,32,37-56] The 19 studies with average or median diameter greater than 20 mm had higher sensitivity, 91% (95%CI: 87%, 94%)[16,17,27,31,57-71] compared to the studies with smaller mean lesion size (Figure 7). Specificity of FDG-PET to diagnose lung cancer was not significantly different between studies based upon mean lesion size. Specificity was 71% (95%CI: 42%, 83%) among studies with mean lesion size above 20 mm and 75% (95%CI: 54%, 85%) specificity for studies reporting mean lesion size less than or equal to 20 mm.

Combined PET/CT scanners were significantly more specific ($p=0.02$) and sensitive ($p<0.001$) when compared to PET only scanners (Figure 8). Twenty-five studies reported using PET only scanners in their diagnostic processes.[17,27,29,31,33,36,39-41,46-48,56-59,61,72-79] Twelve of the 25 studies were published between 2000 and 2002; combined PET/CT scanners became clinically available in 2001. The sensitivity for PET only scanners was 87.3% (95%CI: 82%, 91%). The sensitivity for the 28 studies using combined PET/CT scanners was 90.5% (95%CI: 87%, 94%).[16,20,25,26,28,30,32,35,37,38,49-55,65,66,69-71,80-85] The sensitivity for the seven studies that used a PET or PET/CT scanner in combination with another type of scan was 91.2% (95%CI: 88, 93) which was significantly ($p=0.03$) higher than PET only scanners and similar to PET/CT scanner sensitivity.[42,45,63,64,68,86,87] The specificity for PET only scanners was 70% (95%CI: 56%, 79%) which was significantly lower ($p=0.02$) than the observed specificity for combined PET/CT scanners, 77.5% (95%CI: 57%, 86%). The specificity reported for scanners that used PET and some other radiographic modality was 76% (95%CI: 71%, 80%) and was not significantly different from PET only

scanners (p=0.65). Among the other imaging modalities reported, three studies used single-photon emission computerized tomography (SPECT) as the alternative secondary scanning modality.[45,63,64] Two reported using F^{18}-fluorothyminidine (FLT) in conjunction with FDG.[68,87] One used a sodium iodide detector[86] and one created an algorithm of staggered PET scans and a model in conjunction with standard uptake values.[42]

Study Quality

Thirty-five (58%) of the included studies were retrospective and 25 (42%) were prospective studies (Table1). There was no statistically significant difference in sensitivity (90% and 89%) or specificity (76% and 74%) based upon study enrollment method. Method of diagnosis was either entirely based upon pathological diagnosis or a combination of pathological diagnosis and radiographic surveillance. There was no statistically significant difference in sensitivity between studies reporting using pathology only (90%; 95%CI: 82%, 94%) and those using a combination of methods for diagnosis (89%; 95%CI: 82%, 93%). Studies using only pathological diagnosis had lower specificity (70%: 95%CI: 48%, 86%) than those reporting using a combination of methods (77%; 95%CI: 48%, 88%), but the differences were not statistically significant.

Twelve studies blinded the scan reader to both patient history and outcomes of previous tests (Figure 9).[25,28-30,33,41,53,56,60,84,86] Scan readers were not blinded to patient history and outcome in 37 studies (60%). Another 12 studies did not report enough information to determine the method of reader blinding.[31,35,36,40,42,46,48,51,55,57,58,70] The twelve studies that reported blinding of readers had slightly lower sensitivity (88%; 95%CI: 82%, 92%) and specificity (70%; 95%CI: 57%, 86%) when compared to the 37 studies that did not (sensitivity 90%; 95%CI: 84%, 94% and specificity 74%; 95%CI: 46%, 80%). This was the only study quality variable used in the meta-analytic model and multiple imputation was used to estimate missing data for blinding in those 12 studies not reporting method of blinding.

The quality metric that most studies failed to meet was patients receiving the same reference standard regardless of index test result (58% No). Many studies examined the test accuracy in the clinical setting. Since FDG-PET scans are part of the diagnostic work up and considered standard of care for diagnosing lung cancer, this result is to be expected. Twenty-seven studies had fewer than 25 cancer or 25 benign cases. Most often studies lacked sufficient benign cases, and this was reflected in the higher variability of specificity across all studies. Most studies (83%) emulated the use of the test as it would be used in practice.

Agreement between reviewers for the initial examination for eligibility of the 1,231 abstracts was examined quantitatively. The observed agreement for study eligibility between the three reviewers was 94.1%, and Cohen's Kappa was 0.85 showing strong agreement between reviewers. Consensus was used when reviewers disagreed for the remaining data abstracted and agreement was not reviewed quantitatively.

4.4 Discussion

For the last decade molecular imaging with FDG-PET has become part of the diagnostic process for lung nodules suspicious for non-small cell lung cancer. The limitation of FDG in smaller lesions or slower growing cancers has been well documented.[3,4,88] Recently, significant research efforts have been undertaken by radiologists to find a complement or a replacement for FDG-PET scans.[87,89-93] To date, no replacement for FDG has been suggested for the diagnosis of lung cancer.[2,88] Previous meta-analyses found FDG-PET to be highly sensitive and fairly specific in the diagnosis of lung cancer.[7,8] In our study, the sensitivity (89%; 95%CI: 82%, 94%) was less than the median sensitivity (94.2%; 95%CI: 89.1%, 97.0%) reported by Gould.[8] Specificity (75%; 95%CI: 46%, 86%) was also less than the specificity observed (83.3%) in the earlier meta-analysis.[8] The unadjusted model estimated summary receiver

operating curve (SROC) was 0.90 (95%CI: 0.87, 0.93) and this estimated SROC was similar to that reported by Gould and colleagues (SROC 0.91; 95%CI: 0.89, 0.93). However, under conditions of a correctly pooled meta-analysis, differences between studies should be illustrated by a choice of different combinations of sensitivity and specificity and movement along the curve. Thus, studies should cluster along the SROC curve and within the 95% prediction contour. The large dispersion of studies and large prediction area illustrates the wide variance or heterogeneity between studies (Figure 3). This dispersion of study points suggests multiple SROC curves for the different populations observed across the studies being evaluated. Thus, the plotted SROC curve does not comply with the assumptions for a single pooled SROC. Both the SROC curve in this study and in the Gould meta-analysis showed significant heterogeneity across studies.

This meta-analysis had slightly greater variance in both sensitivity and specificity when compared to either earlier meta-analysis. Yet our meta-analysis had a pooled population that was four times larger than that reported in Gould and five times that reported in Cronin's meta-analysis. The more sensitive nature of the systematic review that included studies comparing FDG-PET to other imaging modalities and seeking the causes of the observed heterogeneity results in recently published articles are likely the causes of the greater variance in this study's FDG-PET accuracy. Significant heterogeneity arose from three sources not related to the quality of study in this meta-analysis. They were the type of scanner used, the mean or median size of the lesion examined, and endemic infectious lung disease in the study population.

The newer technology of combination PET/CT scanners has generally replaced the stand-alone PET scanner since their introduction into clinical practice in 2001.[94,95] This analysis found significant improvement in sensitivity and specificity among the studies of combination PET/CT scanners when compared to studies using stand-alone PET scanners. Others have found the newer combined scanners to also improve non-invasive staging of lung cancer.[96,97]

Among the 24 studies that reported average or median lesion diameters of 20mm or less, pooled sensitivity was 88% which was significantly less (p=0.001) than the sensitivity (91%) in those studies whose mean or median diameter was greater than 20 mm. The populations undergoing FDG-PET scans in the studies reviewed had high prevalences of lung cancer; the median prevalence was 63% across all studies. The high prevalence of disease coupled with the low survival rate for lung cancer resulted in clinicians preferring sensitivity over specificity. The observed 3% difference in sensitivity appears clinically insignificant on its face; however, this seemingly small difference arising from studies with lesions less than 20mm in diameter and with 63% lung cancer prevalence translated into missing two cancers per 100 individuals scanned. High dose CT scans with less than 5 mm thin slices have a high sensitivity (95%), but the specificity of such scans is only 50%.[98-100] The high specificity of FDG-PET scans is one of the primary reasons they have been recommended for the diagnosis of lung cancer.

Seven studies reported confounding of FDG-PET specificity due to false positive granulomas generated by endemic infectious lung disease.[16,17,20,30-33] The pooled specificity among the seven studies was 59%. Granuloma prevalence in the population with benign diagnoses within these studies ranged from 45%[32] to over 75%.[20] Studies reporting endemic infectious lung disease included the largest two individual studies (N=585 and 211) in the systematic review.[16,32] Mean lung cancer prevalence in these seven studies was 76.7% compared to the 60.7% prevalence observed in the remaining 53 studies. Two of the seven studies were retrospective[16,31] and five studies determined diagnosis with pathology only.[16,17,20,32,33]

Of the seven studies, Bryant et al. and Chundru et al. were identified as possible outliers. The article by Bryant and colleagues is of interest in that it was both the largest study in the analysis (9% of the total); it reported benign granulomas as a common diagnosis from false positive FDG-PET scans; and the reported sensitivity (93%) and specificity (75%) were the most accurate among the seven studies. The study by

Chundru and colleagues may not have been from consecutive patients and may have selected an enriched population with benign disease, which would cause it to be excluded from all analysis. The novel method of assessing the likelihood of cancer in the Chundru was unique among all studies.[30] For the five homogeneous studies SROC was 0.88 with a sensitivity of 90% and a specificity of 39%, which is a striking difference from the specificity of 59% when the Bryant and Chundru studies are included in the pooled estimates. One possible explanation for the difference in specificity results from the Bryant et al. paper can be found by comparing it to the study by Deppen et al. Bryant and colleagues derived their population from central and southern Alabama, and Deppen and colleagues reported on FDG-PET scans of the population immediately to the north. Both studies are from large tertiary academic medical centers and report results from a thoracic surgery population. But the prevalence of fungal lung disease in middle Tennessee, northern Alabama and southern Kentucky is twice that found in central and southern Alabama. Therefore, the prevalence of benign granulomas among those being evaluated was less in Bryant's study. The lower specificity of FDG-PET scans in populations with endemic infectious lung disease puts into question the cost effectiveness of FDG-PET in the diagnosis of lung cancer when compared to thin slice, high dose CT scans with contrast in such populations. FDG-PET scans are over five times more expensive than high dose CT scans. High dose CT scans cost between $350 and $500 compared to the $2000-$3000 for a FDG-PET scan.[96,101] The strikingly low specificity in the seven studies and in the five more homogeneous studies raises the question of how such a sub-population could have been missed in prior meta-analyses.

In the meta-analysis by Gould and colleagues, 727 of the 1,474 (49%) lesions included were from studies that reported on either Japanese or European populations. Fungal lung disease and tuberculosis that causes granulomas are rare in Europe and Japan. At least another 15% of the lesions in the meta-analysis were from regions of the US where granulomatous disease is rare. Similarly, in the meta-analysis by Cronin, 860 of the 1,190 lesions (72%) reported in the 22 studies reviewed were from populations where infectious lung disease was rare. One of the studies from Germany included in both meta-analyses

concluded that the prevalence of histoplasmosis and coccidioidomycosis was far lower in Europe and that their results may not translate to North America.[102] These imaging centers that dominated the early reporting of FDG-PET results may have introduced a spectrum bias into the evaluation of FDG-PET scans to diagnose lung nodules. The lack of infectious lung disease in the populations being evaluated for lung cancer resulted in optimistic estimates of the test's specificity. In regions where infectious lung disease is highly prevalent, the specificity of FDG-PET scans to diagnose lung nodules suspicious for lung cancer appears to be near 40%.

These two conclusions have significant implications for clinicians. In individuals who are being evaluated for lung cancer and reside in a region of the US or in other countries with significant endemic infectious lung disease with lesions less than 20 mm, a FDG-PET/CT scan will not likely differentiate the lesion as cancer or benign disease. The false positive rate is high, which reduces the positive likelihood ratio and the false negative rate is also too high to elicit a differential diagnosis. The lower positive and negative likelihood ratios observed in this meta-analysis in populations with small nodules from areas with endemic infectious lung disease greatly limit the usefulness of FDG-PET. This limitation is illustrated in likelihood quadrant plots.

A likelihood quadrant graph plots the positive and negative likelihood ratio of a test. The plot is divided into four quadrants based upon a designation of a "discriminating" test. A highly discriminating diagnostic test will have a positive likelihood ratio above 10 and a negative likelihood ratio less than 0.1 (Upper Left Quadrant). Such a test, if negative, will definitively rule out disease due to its low negative likelihood ratio, and if positive will rule in disease due to its high positive likelihood ratio. The 2001 meta-analysis reported a positive likelihood ratio of 7.1 and a negative likelihood ratio of 0.06. Tests with high positive likelihood ratios (>10) and high negative likelihood ratios (>0.1) are best at confirming presence of disease (Upper Right Quadrant) while tests with low positive likelihood ratios (<10) and low negative likelihood ratios (<0.1) best discriminate those without disease (Lower Left Quadrant). Studies

that have low positive likelihood ratios (<10) and high negative likelihood ratios (>0.1) are relatively poor tests for discriminating either those with or without disease (Lower Right Quadrant).

For each of the plots (Figures 11 to 13) the pre-test probability of disease was the pooled lung cancer prevalence (63.8%) and each study is plotted according to its individually estimated positive and negative likelihood ratio. Figure 11 displays all the studies in this meta-analysis. As we can see, a majority of studies lie in the Lower Right Quadrant. The pooled likelihood ratios and their respective 95% confidence intervals (positive LR 3.56, 95%CI: 1.51, 6.71 and negative LR 0.15, 95%CI: 0.07-0.39) are the brown diamond and crosshairs bisecting the diamond. The high sensitivity of most studies placed them in the lower left quadrant. Only one study is in the preferred upper left quadrant.[86] Figure 12 displays only the 7 studies that reported endemic infectious lung disease. Except for the Bryant study (point 3) with its high sensitivity (93%)[30] and low negative likelihood ratio (0.09) and the study by Chundru et al. (point 5) with its high specificity and high positive likelihood ratio (7.5)[30], none of the other six studies were near either discriminating quadrant. The pooled positive likelihood ratio was 2.21 (95%CI: 1.67-3.10) and the negative likelihood ratio was 0.15 (95%CI: 0.10-0.22). For the clinician deciding whether to order a FDG-PET scan for a patient from an area with high prevalence of endemic infectious lung disease, there is little likelihood that the test will discriminate whether the individual has or does not have the disease.

Lesions under 15 mm are known to reduce the sensitivity of FDG-PET in lung cancer diagnosis.[8] The test is not recommended in nodules smaller than 8 mm^3. In Figure 13, the 24 studies reporting a mean or median lesion diameter less than or equal to 20 mm had a pooled positive likelihood ratio of 3.75 (95%CI: 2.07, 6.35) and a negative likelihood ratio of 0.16 (95% CI: 0.11, 0.30). Sixteen of the twenty-four studies lie within the non-discriminating lower right quadrant and the combined likelihood ratios and their confidence intervals lie entirely within this quadrant as well. These two aspects of poor discrimination for FDG-PET scan, endemic infectious lung disease, and smaller lesion size should cause the clinician to question the need to order a FDG-PET scan when they occur in patients. In these

instances, when no indication of metastatic disease is evident in prior imaging, the additional expense of a FDG-PET scan may not be warranted as the results will not likely change the diagnostic plan to obtain a biopsy or watch for further indications of cancer or benignity.

The limitations of this analysis are those common to any meta-analysis and include publication bias, selection bias, and heterogeneity of those studies included. While no significant publication bias was observed according to Deek's funnel plot and test, this is not proof that no bias exists. Since FDG-PET has been well established for the diagnosis of lung cancer, there may be a bias for reporting poor results of FDG-PET accuracy in more recent years. Among those studies testing a novel radiopharmaceutical there may have been reporting or other study biases not well captured in this review which was more general in its nature. An argument can be made that studies reporting results from PET only scanners no longer reflect clinical practice and should not be included in this analysis. However, we attempted to control for the shortcomings of PET only scanners in the random effects model so that the impact on FDG-PET accuracy in studies reporting results from smaller lesions and from regions with endemic infectious lung disease could be explored. To avoid selection bias, this meta-analysis reviewed unpublished studies and attempted to broadly review studies reporting use of FDG-PET to characterize lung nodules and evaluate them for lung cancer.

Heterogeneity across studies reporting diagnostic test accuracy is to be expected. We attempted to control for such heterogeneity through a bivariate random-effects binomial regression model with a number of clinically important covariates. Significant differences between studies remained after including those covariates and as such, the pooled analysis is only an approximation. Another weakness of this meta-analysis was the use of multiple imputations for two variables, the blinding of readers and lesion size. The missing data pattern was monotonic which allows most imputation methods to be statistically robust[103], and there did not appear to be systematic bias in the mechanism of why either variable was missing.

4.5 Conclusion

This meta-analysis found a reduction in FDG-PET specificity in certain localized populations that reported endemic infectious lung disease. FDG-PET may not offer differential diagnosis for lung cancer in individuals with lesions smaller than 20 mm or who have had significant exposure to granuloma causing lung infections. Therefore, it is likely that some populations exist that would not benefit from a FDG-PET scan for the diagnosis of a lung nodule suspicious for lung cancer. The effectiveness of FDG-PET/CT scans compared to high dose CT scans in these populations may be much lower than previously reported.[101]

4.6 References

1. Bunyaviroch T, Coleman RE. PET Evaluation of Lung Cancer. *J Nucl Med.* March 1, 2006 2006;47(3):451-469.

2. Gould MK, Fletcher J, Iannettoni MD, et al. Evaluation of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest.* Sep 2007;132(3 Suppl):108S-130S.

3. National Comprehensive Cancer Network. Lung Cancer Screening v1.2012. *NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines).* 2011. http://www.nccn.org/professionals/physician_gls/recently_updated.asp. Accessed 10-27-2011.

4. Delbeke D, Coleman RE, Guiberteau MJ, et al. Procedure Guideline for Tumor Imaging with 18F-FDG PET/CT 1.0. *J Nucl Med.* May 2006 2006;47(5):885-895.

5. Marom EM, Erasmus JJ, Patz Jr EF. Lung cancer and positron emission tomography with fluorodeoxyglucose. *Lung Cancer.* 2000;28(3):187-202.

6. Oyen WG, Mansi L. FDG-PET in infectious and inflammatory disease. *Euro J Nucl Med Mol Imaging.* 2003;30(11):1568-1570.

7. Cronin P, Dwamena B, Kelly A, Carlos R. Solitary pulmonary nodules: meta-analytic comparison of cross-sectional imaging modalities for diagnosis of malignancy. *Radiology.* 2008;246:772 - 782.

8. Gould MK, Maclean CC, Kuschner WG, Rydzak CE, Owens DK. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis. *JAMA.* Feb 21 2001;285(7):914-924.

9. Herder GJ, Golding RP, Hoekstra OS, et al. The performance of 18F-fluorodeoxyglucose positron emission tomography in small solitary pulmonary nodules. *Euro J Nucl Med Mol Imaging.* 2004;31(9):1231-1236.

10. Silvestri GA, Gould MK, Margolis ML, et al. Noninvasive staging of non-small cell lung cancer: ACCP evidenced-based clinical practice guidelines (2nd edition). *Chest.* Sep 2007;132(3 Suppl):178S-201S.

11. Cerfolio RJ, Ojha B, Bryant AS, Raghuveer V, Mountz JM, Bartolucci AA. The accuracy of integrated PET-CT compared with dedicated pet alone for the staging of patients with nonsmall cell lung cancer. *Ann Thorac Surg.* 2004;78(3):1017-1023.

12. Reed CE, Harpole DH, Posther KE, et al. Results of the American College of Surgeons Oncology Group Z0050 trial: the utility of positron emission tomography in staging potentially operable non-small cell lung cancer. *J Thorac Cardiovasc Surg.* Dec 2003;126(6):1943-1951.

13. Wahidi MM, Govert JA, Goudar RK, Gould MK, McCrory DC. Evidence for the treatment of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest.* Sep 2007;132(3 Suppl):94S-107S.

14. Rivera MP, Mehta AC. Initial diagnosis of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest.* Sep 2007;132(3 Suppl):131S-148S.

15. Feinstein AR. *Clinical Epidemiology: The architecture of clinical research.* W. B Saurers Co; 1985.

16. Deppen S, Putnam JB, Jr., Andrade G, et al. Accuracy of FDG-PET to diagnose lung cancer in a region of endemic granulomatous disease. *Ann Thorac Surg.* Aug 2011;92(2):428-432; discussion 433.

17. Croft DR, Trapp J, Kernstine K, et al. FDG-PET imaging and the diagnosis of non-small cell lung cancer in a region of high histoplasmosis prevalence. *Lung Cancer.* 2002;36(3):297-301.

18. Mason R, Broaddus V, Martin T, et al., eds. *Murray and Nadel's Textbook of Respiratory Medicine.* 5th ed: Saunders; 2010. Murray J, Nadel J, eds; No. 2.

19. Li Y, Su M, Li F, Kuang A, Tian R. The value of 18-F-FDG-PET/CT in the differential diagnosis of solitary pulmonary nodules in areas with a high incidence of tuberculosis. *Ann Nucl Med.*1-8.

20. Sathekge M, Maes A, Pottel H, Stoltz A, van de Wiele C. Dual time-point FDG-PET-CT for differentiating benign form malignant solitary pulmonary nodules in a TB endemic area. *S Afr Med J.* 2010;100(9):598-601.

21. Mukhopadhyay S, Gal AA. Granulomatous Lung Disease: An Approach to the Differential Diagnosis. *Arch Pathol Lab Med.* 2010/05/01 2010;134(5):667-690.

22. Fontela PS, Pant Pai N, Schiller I, Dendukuri N, Ramsay A, Pai M. Quality and Reporting of Diagnostic Accuracy Studies in TB, HIV and Malaria: Evaluation Using QUADAS and STARD Standards. *PLoS One.* 2009;4(11):e7753.

23. Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol.* 2003;3(1):25.

24. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA.* Sep 15 1999;282(11):1061-1066.

25. Schillaci O, Travascio L, Bolacchi F, et al. Accuracy of early and delayed FDG PET-CT and of contrast-enhanced CT in the evaluation of lung nodules: a preliminary study on 30 patients. *La Radiologia Medica.* 2009/09/01 2009;114(6):890-906.

26. Orlacchio A, Schillaci O, Antonelli L, D'Urso S, Sergiacomi G, Nicoli P. Solitary pulmonary nodules: morphological and metabolic characterisation by FDG-PET-MDCT. *Radiol Med (Torino).* 2007;112:157 - 173.

27. Núñez R, Kalapparambath A, Varela J. Improvement in sensitivity with delayed imaging of pulmonary lesions with FDG-PET. *Revista Española de Medicina Nuclear (English Edition).* 2007;26(4):196-207.

28. Ding QY, Hua YQ, Zhang GZ, et al. A controlled study of positron-emission-tomography and positron-emission-tomography/computed tomography in differential diagnosis of solitary pulmonary nodules--report of 60 cases. *Chin Med J (Engl).* Sep 20 2005;118(18):1572-1576.

29. Ning XH, Meng QL, Wang YZ, Bai CM. Application of 18F-fluorodeoxyglucose positron emission tomography in diagnosis of malignant diseases. *Chin Med Sci J.* Jun 2009;24(2):117-121.

30. Chundru S, Wong CY, Wu D, et al. Granulomatous disease: is it a nuisance or an asset during PET/computed tomography evaluation of lung cancers? *Nucl Med Commun.* Jul 2008;29(7):623-627.

31. Kim S, Machac J, Krynyckyi B, et al. Fluoro-deoxy-glucose positron emission tomography for evaluation of indeterminate lung nodules: assigning a probability of malignancy may be preferable to binary readings. *Ann Nucl Med.* 2008;22(3):165-170.

32. Bryant AS, Cerfolio RJ. The Maximum Standardized Uptake Values on Integrated FDG-PET/CT Is Useful in Differentiating Benign From Malignant Pulmonary Nodules. *Ann Thorac Surg.* 2006;82(3):1016-1020.

33. Mamede M, Higashi T, Kitaichi M, et al. 18F-Fdg Uptake and PCNA, Glut-1, and Hexokinase-II expressions in cancer and inflammatory lesions of the lung. *Neoplasia.* 2005;7(4):369-379.

34. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. *Introduction to Meta-Analysis.* 1 ed: Wiley; 2009.

35. Lan XL, Zhang YX, Wu ZJ, Jia Q, Wei H, Gao ZR. The value of dual time point 18F-FDG PET imaging for the differentiation between malignant and benign lesions. *Clin Radiol.* 2008;63(7):756-764.

36. Sasaki M, Kuwabara Y, Yoshida T, et al. Comparison of MET-PET and FDG-PET for differentiation between benign lesions and malignant tumors of the lung. *Ann Nucl Med.* 2001;15(5):425-431.

37. Barnett PG, Ananth L, Gould MK, Group ftVAPETIitMoPwSPNCS. Cost and Outcomes of Patients With Solitary Pulmonary Nodules Managed With PET Scans. *Chest.* January 1, 2010 2010;137(1):53-59.

38. Bastarrika G, Garcia-Velloso MJ, Lozano MD, et al. Early Lung Cancer Detection using Spiral Computed Tomography and Positron Emission Tomography. *Am J Respir Crit Care Med.* Mar 24 2005.

39. Christensen JA, Nathan MA, Mullan BP, Hartman TE, Swensen SJ, Lowe VJ. Characterization of the Solitary Pulmonary Nodule: 18F-FDG PET Versus Nodule-Enhancement CT. *Am. J. Roentgenol.* November 1, 2006 2006;187(5):1361-1367.

40. Higashi K, Ueda Y, Sakuma T, et al. Comparison of [(18)F]FDG PET and (201)Tl SPECT in evaluation of pulmonary nodules. *J Nucl Med.* Oct 2001;42(10):1489-1496.

41. Lee J, Aronchick JM, Alavi A. Accuracy of F-18 fluorodeoxyglucose positron emission tomography for the evaluation of malignancy in patients presenting with new lung abnormalities: a retrospective review. *Chest.* Dec 2001;120(6):1791-1797.

42. Alkhawaldeh K, Bural G, Kumar R, Alavi A. Impact of dual-time-point 18F-FDG PET imaging and partial volume correction in the assessment of solitary pulmonary nodules. *Euro J Nucl Med Mol Imaging.* 2008/02/01 2008;35(2):246-252.

43. Matthies A, Hickeson M, Cuchiara A, Alavi A. Dual Time Point 18F-FDG PET for the Evaluation of Pulmonary Nodules. *J Nucl Med.* July 1, 2002 2002;43(7):871-875.

44. Nomori H, Watanabe K, Ohtsuka T, Naruke T, Suemasu K, Uno K. Visual and Semiquantitative Analyses for F-18 Fluorodeoxyglucose PET Scanning in Pulmonary Nodules 1 cm to 3 cm in Size. *Ann Thorac Surg.* 2005;79(3):984-988.

45. Halley A, Hugentobler A, Icard P, et al. Efficiency of 18F-FDG and 99mTc-depreotide SPECT in the diagnosis of malignancy of solitary pulmonary nodules. *Eur J Nucl Med Mol Imaging.* Sep 2005;32(9):1026-1032.

46. Yi CA, Lee KS, Kim B-T, et al. Tissue Characterization of Solitary Pulmonary Nodule: Comparative Study Between Helical Dynamic CT and Integrated PET/CT. *J Nucl Med.* March 2006 2006;47(3):443-450.

47. Naalsund A, Maublant J. The Solitary Pulmonary Nodule -- Is It Malignant or Benign? Diagnostic Performance of 99m Tc-Depreotide SPECT. *Respiration.* 2006;73(5):634-641.

48. Tsunezuka Y, Shimizu Y, Tanaka N, Takayanagi T, Kawano M. Positron Emission Tomography in Relation to Noguchi's Classification for Diagnosis of Peripheral Non-Small-Cell Lung Cancer 2 cm or Less in Size. *World J Surg.* 2007/02/01 2007;31(2):314-317.

49. Kim SK, Allen-Auerbach M, Goldin J, et al. Accuracy of PET/CT in Characterization of Solitary Pulmonary Lesions. *J Nucl Med.* February 1, 2007 2007;48(2):214-220.

50. Jeong SY, Lee KS, Shin KM, et al. Efficacy of PET/CT in the characterization of solid or partly solid solitary pulmonary nodules. *Lung Cancer.* 2008;61(2):186-194.

51. Ohno Y, Koyama H, Takenaka D, et al. Dynamic MRI, dynamic multidetector-row computed tomography (MDCT), and coregistered 2-[fluorine-18]-fluoro-2-deoxy-D-glucose–positron emission tomography (FDG-PET)/CT: Comparative study of capability for management of pulmonary nodules. *Journal of Magnetic Resonance Imaging.* 2008;27(6):1284-1295.

52. Degirmenci B, Wilson D, Laymon CM, et al. Standardized uptake value-based evaluations of solitary pulmonary nodules using F-18 fluorodeoxyglucose-PET/computed tomography. *Nucl Med Commun.* 2008;29(7):614-622.

53. Kagna O, Solomonov A, Keidar Z, et al. The value of FDG-PET/CT in assessing single pulmonary nodules in patients at high risk of lung cancer. *Eur J Nucl Med Mol Imaging.* Jun 2009;36(6):997-1004.

54. Ohba Y, Nomori H, Shibata H, et al. Evaluation of Semiquantitative Assessments of Fluorodeoxyglucose Uptake on Positron Emission Tomography Scans for the Diagnosis of Pulmonary Malignancies 1 to 3 cm in Size. *Ann Thorac Surg.* 2009;87(3):886-891.

55. Ohno Y, Koyama H, Matsumoto K, et al. Differentiation of Malignant and Benign Pulmonary Nodules with Quantitative First-Pass 320–Detector Row Perfusion CT versus FDG PET/CT. *Radiology.* February 1, 2011 2011;258(2):599-609.

56. MacDonald K, Searle J, Lyburn I. The role of dual time point FDG PET imaging in the evaluation of solitary pulmonary nodules with an initial standard uptake value less than 2.5. *Clin Radiol.* 2011;66(3):244-250.

57. Halter G, Storck M, Guhlmann A, Frank J, Grosse S, Liewald F. FDG positron emission tomography in the diagnosis of peripheral pulmonary focal lesions. *Thorac Cardiovasc Surg.* Apr 2000;48(2):97-101.

58. Hung GU, Shiau YC, Tsai SC, Ho YJ, Kao CH, Yen RF. Differentiation of radiographically indeterminate solitary pulmonary nodules with. *Jpn J Clin Oncol.* Feb 2001;31(2):51-54.

59. Imdahl A, Jenkner S, Brink I, et al. Validation of FDG positron emission tomography for differentiation of unknown pulmonary lesions. *Eur J Cardiothorac Surg.* August 1, 2001 2001;20(2):324-329.

60. Menda Y, Bushnell DL, Madsen H, McLaughlin K, Kahn D, Kernstine KH. Evaluation of various corrections to the standardized uptake value for diagnosis of pulmonary malignancy. *Nucl Med Commun.* 2001;22(10):1077-1081.

61. Yang SN, Liang JA, Lin FJ, Kwan AS, Kao CH, Shen YY. Differentiating benign and malignant pulmonary lesions with FDG-PET. *Anticancer Res.* Nov-Dec 2001;21(6A):4153-4157.

62. Hickeson M, Yun M, Matthies A, et al. Use of a corrected standardized uptake value based on the lesion size on CT permits accurate characterization of lung nodules on FDG-PET. *Euro J Nucl Med Mol Imaging.* 2002/12/01 2002;29(12):1639-1647.

63. Demura Y, Tsuchida T, Ishizaki T, et al. 18F-FDG accumulation with PET for differentiation between benign and malignant lesions in the thorax. *J Nucl Med.* Apr 2003;44(4):540-548.

64. Kahn D, Menda Y, Kernstine K, et al. The Utility of 99mTc Depreotide Compared With F-18 Fluorodeoxyglucose Positron Emission Tomography and Surgical Staging in Patients With Suspected Non-small Cell Lung Cancer*. *Chest.* February 2004 2004;125(2):494-501.

65. Ferran Na, Ricart Ya, Lopez Mb, et al. Characterization of radiologically indeterminate lung lesions: 99mTc-depreotide SPECT versus 18F-FDG PET. *Nucl Med Commun.* 2006;27(6):507-514.

66. Wang F, Wang Z, Yao W, Xie H, Xu J, Tian L. Role of 99mTc-Octreotide Acetate Scintigraphy in Suspected Lung Cancer Compared with 18F-FDG Dual-Head Coincidence Imaging. *J Nucl Med.* September 2007 2007;48(9):1442-1448.

67. Veronesi G, Bellomi M, Veronesi U, et al. Role of Positron Emission Tomography Scanning in the Management of Lung Nodules Detected at Baseline Computed Tomography Screening. *Ann Thorac Surg.* 2007;84(3):959-966.

68. Tian J, Yang X, Yu L, et al. A Multicenter Clinical Trial on the Diagnostic Value of Dual-Tracer PET/CT in Pulmonary Lesions Using 3′-Deoxy-3′-18F-Fluorothymidine and 18F-FDG. *J Nucl Med.* February 2008 2008;49(2):186-194.

69. Pauls S, Buck A, Halter G, et al. Performance of Integrated FDG-PET/CT for Differentiating Benign and Malignant Lung Lesions -Results from a Large Prospective Clinical Trial. *Mol Imaging Biol.* 2008/03/01 2008;10(2):121-128.

70. Baram D, Bilfinger T. Interaction of clinical suspicion and PET in the diagnosis of suspected thoracic malignancy. *Med Sci Monit.* 2008;14(7):CR379-383.

71. Huang Y-E, Pu Y-L, Huang Y-J, et al. The utility of the nonattenuation corrected 18F-FDG PET images in the characterization of solitary pulmonary lesions. *Nucl Med Commun.* 2010;31(11):945-951 910.1097/MNM.1090b1013e32833ed32857d.

72. Chhajed P, Bernasconi M, Gambazzi F, Bubendorf L, Rasch H, Kneifel S. Combining bronchoscopy and positron emission tomography for the diagnosis of the small pulmonary nodule < or = 3 cm. *Chest.* 2005;128:3558 - 3564.

73. Herder GJ, van Tinteren H, Golding RP, et al. Clinical Prediction Model To Characterize Pulmonary Nodules*. *Chest.* October 2005 2005;128(4):2490-2496.

74. Menda Y, Kahn D. Somatostatin receptor imaging of non-small cell lung cancer with 99mTc depreotide. *Semin Nucl Med.* 2002;32(2):92-96.

75. Nomori H, Watanabe K, Ohtsuka T, Naruke T, Suemasu K, Uno K. Evaluation of F-18 fluorodeoxyglucose (FDG) PET scanning for pulmonary nodules less than 3 cm in diameter, with special reference to the CT images. *Lung Cancer.* Jul 2004;45(1):19-27.

76. Skehan S, Coates G, Otero C, O'Donovan N, Pelling M, Nahmias C. Visual and semiquantitative analysis of 18F-fluorodeoxyglucose positron emission tomography using a partial-ring tomograph without attenuation correction to differentiate benign and malignant pulmonary nodules. *Can Assoc Radiol J.* 2001;52(4):259-265.

77. Uppot R, Conde K, Sagar V, Manzone T. Positron emission tomography (PET) imaging for solitary pulmonary nodules--review of the Delaware experience. *Del Med J.* 2001;73(10):381-385.

78. Pastorino U, Bellomi M, Landoni C, et al. Early lung-cancer detection with spiral CT and positron emission tomography in heavy smokers: 2-year results. *Lancet.* Aug 23 2003;362(9384):593-597.

79. Roman M, Rossleigh M, Angelides S, Walker B, Dixon J. Staging and Managing Lung Tumors Using F-18 FDG Coincidence Detection. *Clin Nucl Med.* 2001;26(5):383-388.

80. Aukema TS, Valdés Olmos RA, Klomp HM, et al. Evaluation of 18F-FDG PET-CT for Differentiation of Pulmonary Pathology in an Approach of Outpatient Fast Track Assessment. *J Thorac Oncol.* 2009;4(10):1226-1230 1210.1097/JTO.1220b1013e3181b1222b1782.

81. Chang C-Y, Tzao C, Lee S-C, et al. Incremental Value of Integrated FDG-PET/CT in Evaluating Indeterminate Solitary Pulmonary Nodule for Malignancy. *Mol Imaging Biol.* 2010;12(2):204-209.

82. Grgic A, Yüksel Y, Gröschel A, et al. Risk stratification of solitary pulmonary nodules by means of PET using 18F-fluorodeoxyglucose and SUV quantification. *Euro J Nucl Med Mol Imaging.* 2010/06/01 2010;37(6):1087-1094.

83. Sachs S, Bilfinger TV. The impact of positron emission tomography on clinical decision making in a university-based multidisciplinary lung cancer practice*. *Chewst.* 2005;128(2):698-703.

84. Veronesi G, Bellomi M, Veronesi U, et al. Role of positron emission tomography scanning in the management of lung nodules detected at baseline computed tomography screening. *Ann Thorac Surg.* Sep 2007;84(3):959-965; discussion 965-956.

85. Yamamoto Y, Nishiyama Y, Ishikawa S, et al. 3'-Deoxy-3'-18F-Fluorothymidine as a Proliferation Imaging Tracer for Diagnosis of Lung Tumors: Comparison With 2-Deoxy-2-18F-Fluoro-D-Glucose. *J Cardiovasc Comput Tomogr.* 2008;32(3):432-437 410.1097/RCT.1090b1013e3180980db3180989.

86. Keith C, Miles K, Griffiths M, Wong D, Pitman A, Hicks R. Solitary pulmonary nodules: accuracy and cost-effectiveness of sodium iodide FDG-PET using Australian data. *Eur J Nucl Med Mol Imaging.* 2002;29(8):1016-1023.

87. Buck A, Hetzel M, Schirrmeister H, et al. Clinical relevance of imaging proliferative activity in lung nodules. *Euro J Nucl Med Mol Imaging.* 2005/04/01 2005;32(5):525-533.

88. Acker MR, Burrell SC. Utility of 18F-FDG PET in Evaluating Cancers of Lung. *J Nucl Med Tech.* June 1, 2005 2005;33(2):69-74.

89. Yang W, Zhang Y, Fu Z, Sun X, Mu D, Yu J. Imaging proliferation of (1)(8)F-FLT PET/CT correlated with the expression of microvessel density of tumour tissue in non-small-cell lung cancer. *Euro J Nucl Med Mol Imaging.* Aug 2012;39(8):1289-1296.

90. Dittman H, Dohmen B, Paulsen F. 18-FLT PET for diagnosis and staging of thoracic tumors. *Eur J Nucl Med Mol Imag.* 2003;30:1407 - 1412.

91. Baylin SB, Jackson RD, Goodwin G, Gazdar AF. Neuroendocrine-related biochemistry in the spectrum of human lung cancers. *Exp Lung Res.* Nov 1982;3(3-4):209-223.

92. Zhuang H, Pourdehnad M, Lambright ES, et al. Dual time point 18F-FDG PET imaging for differentiating malignant from inflammatory processes. *J Nucl Med.* Sep 2001;42(9):1412-1417.

93. Bar-Shalom R, Kagna O, Israel O, Guralnik L. Noninvasive diagnosis of solitary pulmonary lesions in cancer patients based on 2-fluoro-2-deoxy-D-glucose avidity on positron emission tomography/computed tomography. *Cancer.* 2008;113(11):3213-3221.

94. Beyer T, Antoch G, Müller S, et al. Acquisition Protocol Considerations for Combined PET/CT Imaging. *J Nucl Med.* January 1, 2004 2004;45(1 suppl):25S-35S.

95. Beyer T, Townsend DW, Brun T, et al. A Combined PET/CT Scanner for Clinical Oncology. *J Nucl Med.* August 1, 2000 2000;41(8):1369-1379.

96. Lejeune C, Al Zahouri K, Woronoff-Lemsi M-C, et al. Use of a decision analysis model to assess the medicoeconomic implications of FDG PET imaging in diagnosing a solitary pulmonary nodule. *The European Journal of Health Economics.* 2005;6(3):203-214.

97. Liao CY, Chen JH, Liang JA, Yeh JJ, Kao CH. Meta-analysis study of lymph node staging by 18 F-FDG PET/CT scan in non-small cell lung cancer: Comparison of TB and non-TB endemic regions. *Eur J Radiol.* Nov 2012;81(11):3518-3523.

98. Henschke C, Shaham D, Yankelevitz D, Altorki N. CT Screening for Lung Cancer: Past and Ongoing Studies. *Semin Thorac Cardiovasc Surg.* 2005;17:99 - 106.

99. Libby DM, Smith JP, Altorki NK, Pasmantier MW, Yankelevitz D, Henschke CI. Managing the small pulmonary nodule discovered by CT. *Chest.* Apr 2004;125(4):1522-1529.

100. Cronin P, Dwamena B, Kelly A, Bernstein S, Carlos R. Solitary pulmonary nodules and masses: a meta-analysis of the diagnostic utility of alternative imaging tests. *Eur Radiol.* 2008;18(9):1840-1856.

101. Gould M. Cost Effectiveness of Positron Emission Tomography for Characterizing Pulmonary Nodules. *PET Clinics.* 2006;1(4):339 - 346.

102. Präuer HW, Weber WA, Römer W, Treumann T, Ziegler SI, Schwaiger M. Controlled prospective study of positron emission tomography using the glucose analogue [18F]fluorodeoxyglucose in the evaluation of pulmonary nodules. *Br J Surg.* 1998;85(11):1506-1511.

103. Little R, Rubin D. Statistical Analysis with Missing Data, Second edition. 2002.

Table 1: Participant and Study Characteristics for Diagnosis of Pulmonary Nodules.

Citation	Study Method	Scanner Type	Number of scans	Mean or Median Age	% CA Prevalence	Study Population	Diagnosis Method	Blinding
Halter G, 2000 [57]	Prospective	PET	67	44	63%	Surgical	Pathology Only	NR
Higashi K, 2001[40]	Prospective	PET	66	65	82%	Radiological	Pathology Only	NR
Hung G, 2001[58]	Prospective	PET	26	53	77%	Undetermined	Pathology Only	NR
Imdahl A, 2001[59]	Prospective	PET	87	61	79%	Surgical	Pathology Only	Yes
Menda Y, 2001[60]	Retrospective	PET	127	64	68%	Radiological	Pathology Only	No
Roman M, 2001[79]	Prospective	PET	61	67	90%	Oncological	Pathology & Observation	No
Sasaki M, 2001[36]	Retrospective	PET	94	NR	80%	Radiological	Pathology & Observation	NR
Skehan S, 2001[76]	Retrospective	PET	77	64	74%	Radiological	Pathology & Observation	Yes
Uppot R, 2001[77]	Prospective	PET	25	63.5	44%	Radiological	Pathology & Observation	Yes
Yang S, 2001[61]	Not reported	PET	56	65	63%	Undetermined	Pathology & Observation	Yes
Croft, D 2002[17]	Retrospective	PET	85	63.2	82%	Pulmonary	Pathology Only	Yes
Keith C, 2002[86]	Retrospective		89	66.7	54%	Pulmonary	Pathology & Observation	No
Lee J, 2001 [41]	Prospective	PET	71	NR	61%	Radiological	Pathology & Observation	No
Demura Y, 2003[63]	Prospective	PET/CT-SPECT	80	65	63%	Radiological	Pathology Only	Yes
Pastorino U, 2003[78]	Prospective	PET	42	58	48%	Radiological	Observation*	Yes
Buck A, 2005[87]	Prospective	PET	43	61.9	40%	Radiological	Pathology Only	Yes
Kahn D, 2004[64]	Prospective	PET/CT-Other	157	68	78%	Radiological	Pathology & Observation	Yes
Bastarrika G, 2005[38]	Prospective	PET/CT	25	54.7	52%	Oncological	Pathology & Observation	Yes
Chhajed P, 2005[72]	Retrospective	PET	74	64	69%	Pulmonary	Pathology & Observation	Yes
Ding Q, 2005 [28]	Prospective	PET/CT	60	56	50%	Radiological	Pathology Only	No

Citation	Study Method	Scanner Type	Number of scans	Mean or Median Age	% CA Prevalence	Study Population	Diagnosis Method	Blinding
Halley A, 2005 [45]	Multi-arm Trial	PET/CT-SPECT	28	59	64%	Radiological	Pathology & Observation	Yes
Herder GJ, 2005[73]	Prospective	PET	106	64	58%	Oncological	Pathology & Observation	Yes
Mamede, M 2005[33]	Prospective	PET	60	65	77%	Surgical	Pathology Only	No
Nomori H, 2005[44]	Prospective	PET	139	59	66%	Surgical	Pathology & Observation	Yes
Sachs S, 2005[83]	Retrospective	PET/CT	161	62.4	44%	Pulmonary	Pathology & Observation	Yes
Bryant A ,2006[32]	Prospective	PET/CT	585	66	85%	Surgical	Pathology Only	Yes
Christensen J, 2006[39]	Retrospective	PET	42	66	60%	Pulmonary	Pathology & Observation	Yes
Ferran N, 2006[65]	Prospective	PET/CT	29	52	69%	Radiological	Pathology Only	Yes
Naalsund A, 2006[47]	Prospective	PET	29	62	69%	Radiological	Pathology & Observation	Yes
Yi C, 2006[46]	Prospective	PET/CT	119	55	66%	Radiological	Pathology & Observation	NR
Kim SK, 2007[49]	Retrospective	PET/CT	42	67	69%	Pulmonary	Pathology Only	Yes
Núñez R, 2007[27]	Prospective	PET	83	69	86%	Radiological	Pathology Only	Yes
Orlacchio A, 2007[26]	Prospective	PET/CT	56	63	46%	Undetermined	Pathology Only	Yes
Tsunezuka Y, 2007[48]	Prospective	PET	150	65	55%	Surgical	Pathology Only	NR
Veronesi G, 2007[84]	Retrospective - screening	PET/CT	157	57	37%	Radiological	Pathology & Observation	No
Wang F, 2007[66]	Prospective	PET/CT	44	62	71%	Radiological	Pathology & Observation	Yes
Alkhawaldeh K, 2008[42]	Retrospective	PET/CT Dual	265	67	27%	Radiological	Pathology & Observation	NR
Baram D, 2008[70]	Retrospective	PET/CT	313	62	69%	Surgical	Pathology & Observation	NR
Chundru S, 2008 [30]	Prospective	PET/CT	62	68	21%	Radiological	Pathology & Observation	Yes

Citation	Study Method	Scanner Type	Number of scans	Mean or Median Age	% CA Prevalence	Study Population	Diagnosis Method	Blinding
Degirmenci B, 2008[52]	Retrospective	PET/CT	49	69	53%	Radiological	Pathology & Observation	Yes
Jeong S, 2008[50]	Retrospective	PET/CT	100	58	40%	Radiological	Pathology & Observation	Yes
Kim SC, 2008[31]	Retrospective	PET	158	70.3	65%	Radiological	Pathology & Observation	NR
Lan X, 2008[35]	Prospective	PET/CT	45	53	62%	Radiological	Pathology & Observation	NR
Ohno Y, 2008[51]	Prospective	PET/CT	202	72	75%	Radiological	Pathology & Observation	NR
Pauls S, 2008[69]	Prospective	PET/CT	261	64.3	83%	Radiological	Pathology & Observation	Yes
Tian J, 2008[68]	Prospective	PET/CT Other	55	55	29%	Radiological	Pathology & Observation	Yes
Yamamoto Y, 2008[85]	Retrospective	PET/CT	54	70	67%	Radiological	Pathology & Observation	Yes
Aukema T, 2009[80]	Prospective	PET/CT	114	63	84%	Pulmonary	Pathology Only	Yes
Kagna O, 2009 [53]	Retrospective	PET/CT	93	67	40%	Surgical	Pathology & Observation	No
Ning X, 2009[29]	Prospective	PET	101	53.6	65%	Undetermined	Pathology & Observation	No
Ohba Y, 2009[54]	Prospective	PET/CT	130	NR	78%	Pulmonary	Pathology Only	Yes
Schillaci O, 2009[25]	Prospective	PET/CT	30	59.3	60%	Radiological	Pathology & Observation	No
Barnett P, 2010[37]	Prospective	PET/CT	375	66	54%	Pulmonary	Pathology & Observation	Yes
Chang C, 2010[81]	Retrospective	PET/CT	117	62	37%	Radiological	Pathology & Observation	Yes
Grgic A, 2010[82]	Retrospective	PET/CT	140	62	57%	Radiological	Pathology & Observation	Yes
Huang Y, 2010[71]	Retrospective	PET/CT	56	59	61%	Undetermined	Pathology & Observation	Yes
Sathekge M, 2010 [20]	Prospective	PET/CT	30	60	47%	Radiological	Pathology Only	Yes
Deppen S, 2011[16]	Retrospective	PET/CT	211	64	80%	Surgical	Pathology Only	Yes

Citation	Study Method	Scanner Type	Number of scans	Mean or Median Age	% CA Prevalence	Study Population	Diagnosis Method	Blinding
Macdonald K, 2011[56]	Retrospective	PET	54	65.2	48%	Radiological	Pathology & Observation	No
Ohno Y, 2011[55]	Prospective	PET/CT	76	73	57%	Radiological	Pathology & Observation	NR

NR= Not reported

Table 1. Participant and Study Characteristics for Diagnosis of Pulmonary Nodules. (Continued)

Citation	Mean or Median Size	Lesion Size Range	Sensitivity (95%CI)	Specificity (95%CI)	Endemic Infectious Disease (Type, Country or Region)
Halter G, 2000 [57]	NR	14-73	84	81	No
Higashi K, 2001[40]	20	8 to 63	81 (69-91)	42	No
Hung G, 2001[58]	26	12 45	95	50	No
Imdahl A, 2001[59]	30	NR	90	72	No
Menda Y, 2001[60]	33	5-100	94	76	No
Roman M, 2001[79]	NR	NR	96	83	No
Sasaki M, 2001[36]	NR	NR	81	79	No
Skehan S, 2001[76]	NR	NR	95	85	No
Uppot R, 2001[77]	NR	NR	91	71	No
Yang S, 2001[61]	NR	18-72	94	71	No
Croft, D 2002[17]	44	7-170	93	40	Yes (Histoplasmosis, Iowa, USA)
Keith C, 2002[86]	NR	NR	92	95	No
Lee J, 2001 [41]	18.5	7-30	88	75	No
Demura Y, 2003[63]	NR	11 to 60	76	56	No
Pastorino U, 2003[78]	NR	NR	90	82	No
Buck A, 2005[87]	NR	NR	94	73	No
Kahn D, 2004[64]	22	5 - 105	96	71	No
Bastarrika G, 2005[38]	13	8-20	69	91	No

Citation	Mean or Median Size	Lesion Size Range	Sensitivity (95%CI)	Specificity (95%CI)	Endemic Infectious Disease (Type, Country or Region)
Chhajed P, 2005[72]	NR	≤ 30	94	70	No
Ding Q, 2005 [28]	18	5-30	90	93	No
Halley A, 2005 [45]	20	5-30	94	70	No
Herder GJ, 2005[73]	NR	NR	97	71	No
Mamede, M 2005[33]	NR	NR	87	21	Yes (Tuberculosis, Japan)
Nomori H, 2005[44]	20	10 to 30	58	77	No
Sachs S, 2005[83]	NR	NR	90	83	No
Bryant A ,2006[32]	NR	<25	93	75	Yes (Granuloma, Alabama, USA)
Christensen J, 2006[39]	15	7-25	88	76	No
Ferran N, 2006[65]	26	10-60	100	88	No
Naalsund A, 2006[47]	17	5-30	90	67	No
Yi C, 2006[46]	20	6.2-30	96	88	No
Kim SK, 2007[49]	15	7-30	97	85	No
Núñez R, 2007[27]	25	10-120	85	41	No
Orlacchio A, 2007[26]	18	≤ 30	77	100	No
Tsunezuka Y, 2007[48]	NR	≤ 20	76	64	No
Veronesi G, 2007[84]	20.5	NR	88	93	No
Wang F, 2007[66]	48	12-110	100	46	No
Alkhawaldeh K, 2008[42]	17.5	5-30	90	80	No
Baram D, 2008[70]	28.8	11-180	82	78	No
Chundru S, 2008 [30]	20	5-30	65	92	Yes (Granuloma, Michigan, USA)
Degirmenci B, 2008[52]	16.5	NR	62	80	No
Jeong S, 2008[50]	21	9 to 30	88	77	No
Kim SC, 2008[31]	29	2-110	87	53	Yes (Inflammation and Granuloma, New York, USA)
Lan X, 2008[35]	NR	NR	86	65	No
Ohno Y, 2008[51]	15.7	5-30	93	54	No
Pauls S, 2008[69]	41	7-140	96	87	No
Tian J, 2008[68]	28.2	6-110	88	59	No

Citation	Mean or Median Size	Lesion Size Range	Sensitivity (95%CI)	Specificity (95%CI)	Endemic Infectious Disease (Type, Country or Region)
Yamamoto Y, 2008[85]	NR	NR	97	50	No
Aukema T, 2009[80]	NR	NR	97	56	No
Kagna O, 2009 [53]	16	3-30	77	83	No
Ning X, 2009[29]	NR	NR	79	77	No
Ohba Y, 2009[54]	20	10-30	74	79	No
Schillaci O, 2009[25]	19	3.5-30	83	91	No
Barnett P, 2010[37]	17	NR	95	87	No
Chang C, 2010[81]	NR	NR	88	89	No
Grgic A, 2010[82]	NR	NR	94	63	No
Huang Y, 2010[71]	24	5-100	79	77	No
Sathekge M, 2010 [20]	19	NR	86	25	Yes (Tuberculosis, South Africa)
Deppen S, 2011[16]	29	5-140	92	40 (25-56)	Yes (Histoplasmosis, Tennessee, USA)
Macdonald K, 2011[56]	16.5	4-30	58	89 (72-98)	No
Ohno Y, 2011[55]	15.8	8-30	91(78-97)	52 (34-69)	No

NR= Not reported

Table 2. Meta-analysis bivariate random effects model[a]

	Complete Case Analysis N=35			Multiple Imputation N=60		
	Odds Ratio	**95% CI**	**p-value**	**Odds Ratio**	**95% CI**	**p-value**
Cancer	24.30	14.4 - 40.9	<0.001	23.6	16.1 - 34.5	<0.001
Endemic Disease	3.51	1.96 - 6.27	<0.001	2.57	1.53 - 4.31	<0.001
Endemic Disease w/CA [b]	0.41	0.25 - 0.67	<0.001	0.50	0.34 - 0.75	0.001
PET only scanner	Ref	Ref	Ref	Ref	Ref	Ref
PET/CT scanner	0.49	0.29 - 0.84	0.01	.63	0.43 - 0.94	0.02
PET/CT scanner w/CA [b]	3.04	1.87 - 4.93	<0.001	2.31	1.64 - 3.25	<0.001
PET+Other scanner	1.28	0.59 - 2.78	0.53	.88	0.50 - 1.53	0.64
PET+Other w/CA	0.82	0.39 - 1.77	0.63	1.77	1.05 - 3.00	0.03
Lesion size ≤ 20mm	0.66	0.40 - 1.07	0.09	0.89	0.60 - 1.30	0.53
Lesion size ≤ 20mm w/CA[b]	0.63	0.41 - 0.99	0.04	0.73	0.53 - 1.00	0.05
Blinding of readers	1.73	1.01 - 2.96	0.05	1.53	0.98 - 2.38	0.06
Blinding w/CA	0.86	0.51 - 1.43	0.55	0.76	0.51 - 1.12	0.16
Rho for random effects	0.07	0.03 - 0.13	<0.001	0.08	0.04 - 0.12	<0.001

a Odds ratios for interaction terms with cancer which are greater than one increase sensitivity and odds ratios less than one decrease sensitivity. Odds ratios for variables without interaction with cancer influence specificity inversely. Odds ratios with values less than one increased specificity. Odds ratios with values greater than one decreased specificity. The variable's overall effect on sensitivity is determined by the sum of coefficients for direct and interaction term.
b Interaction term for variable interacting with those diagnosed with cancer.

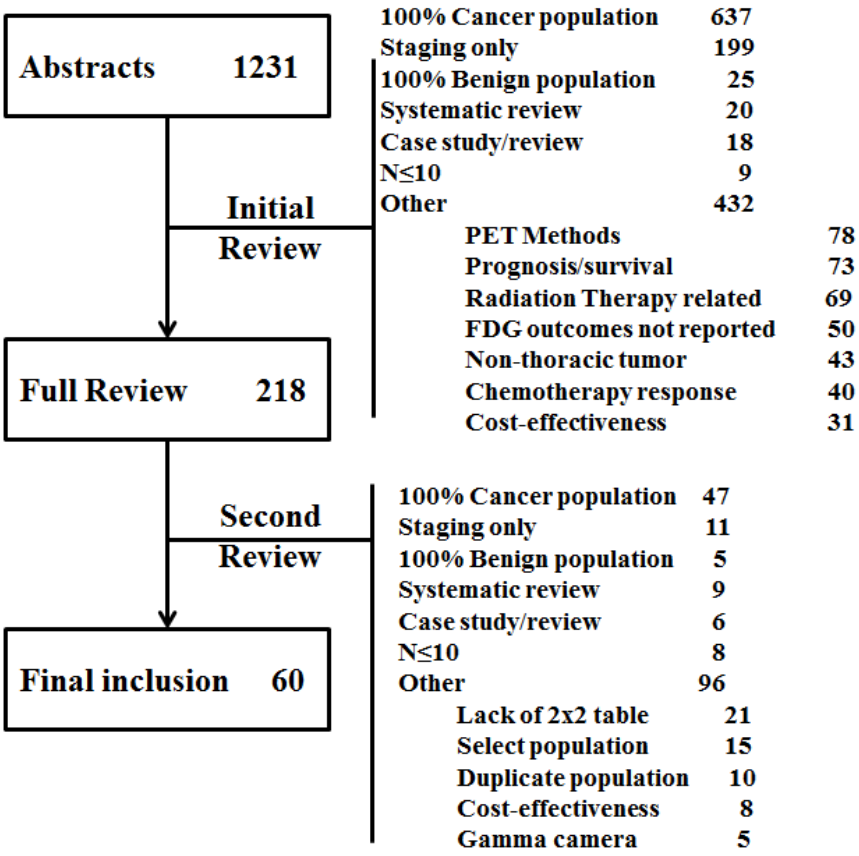


Figure 1. Consort diagram of systematic review of eligible studies.

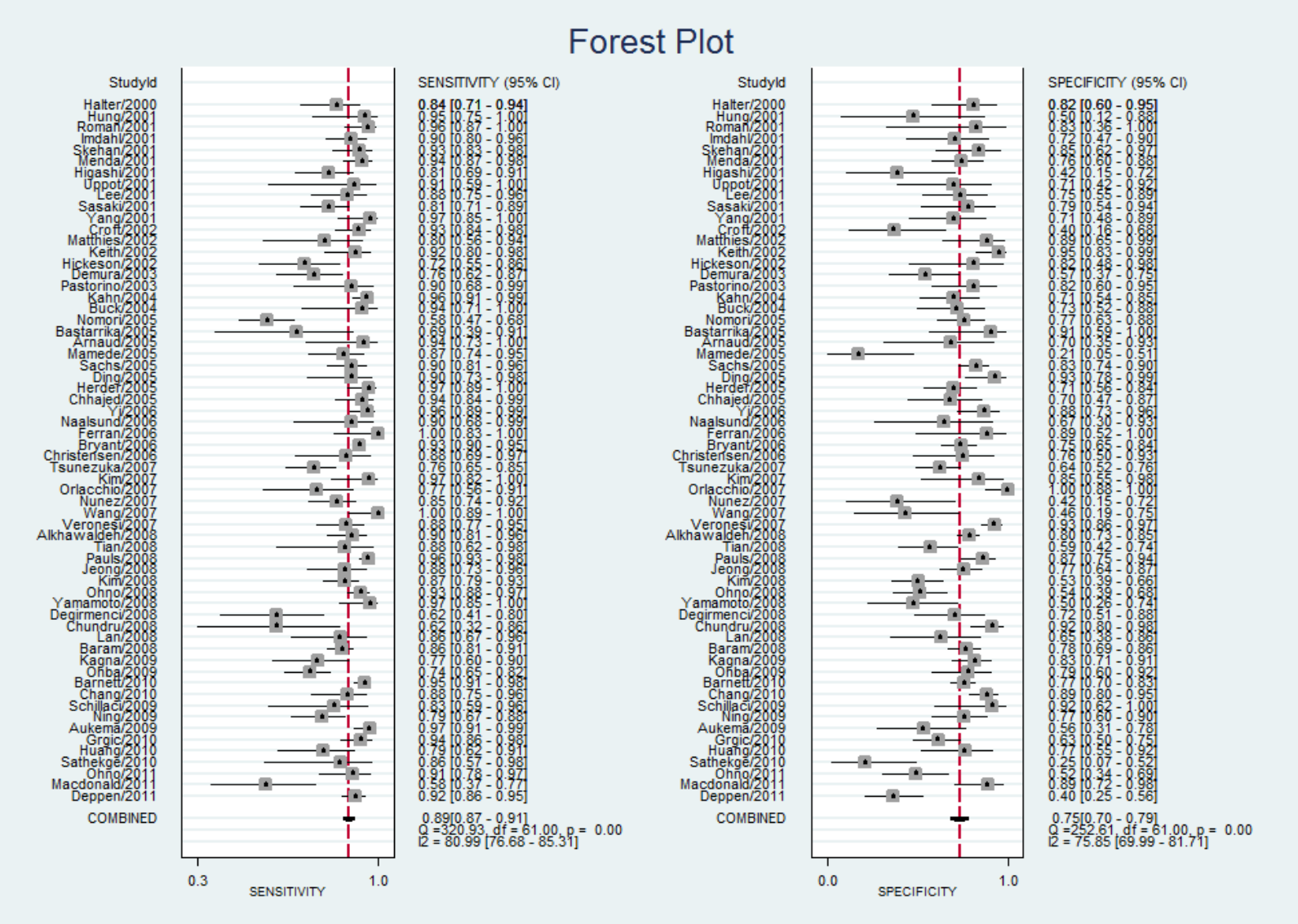


Figure 2: Forest plot of individual study estimates using simple pooled sensitivity and specificity and study heterogeneity, (I^2**)**

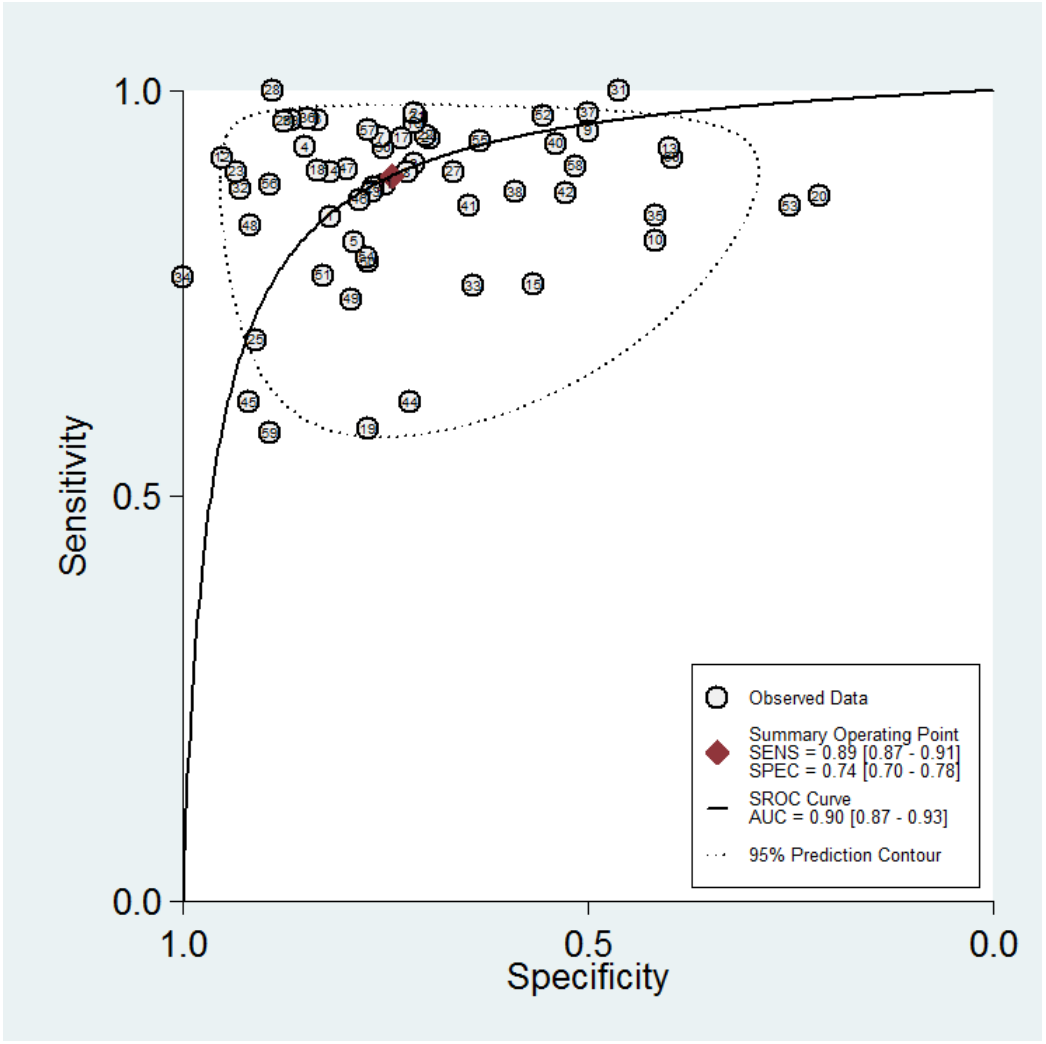


Figure 3. Summary ROC curve from the unadjusted random effects model and estimated 95% prediction contour within which the curve and summary operating point may be located.

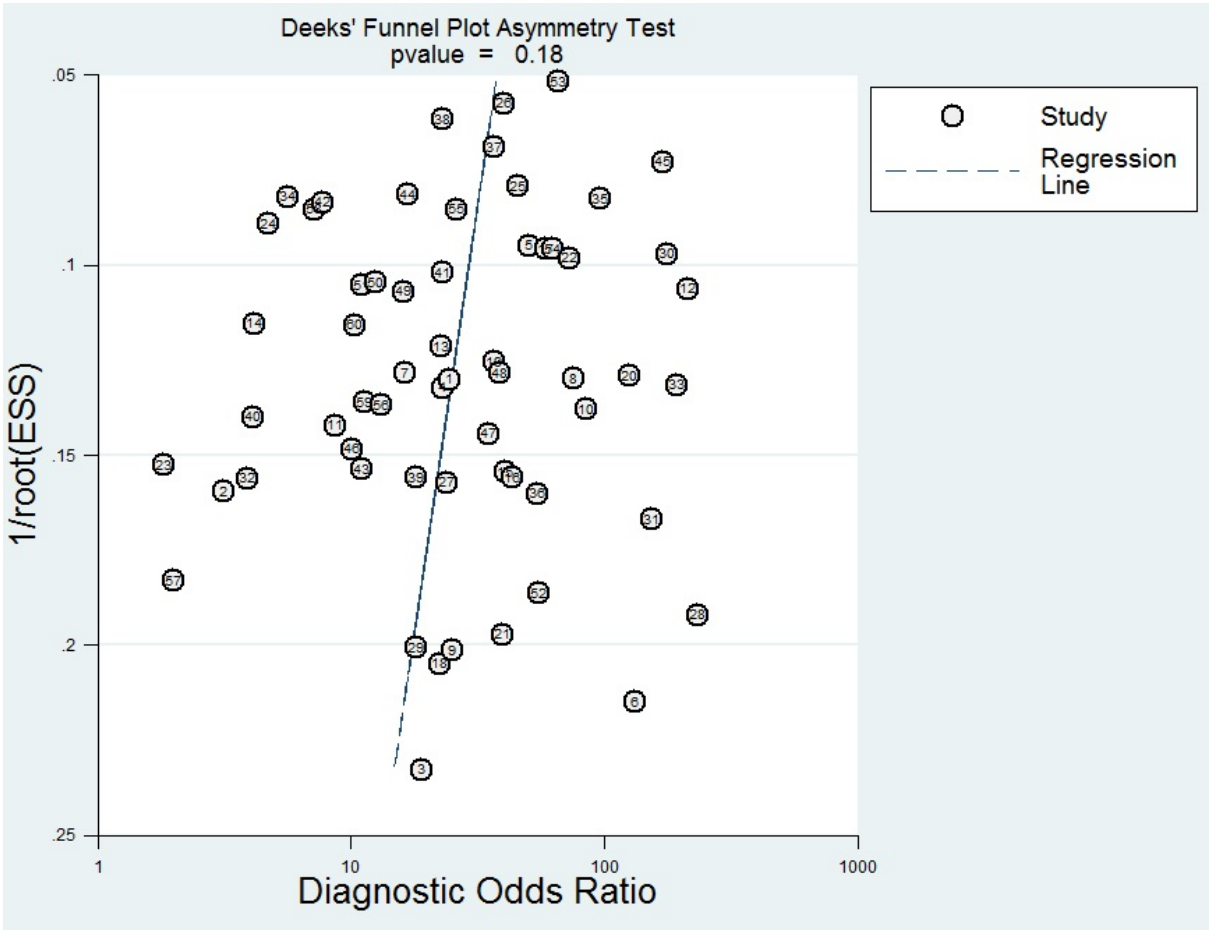


Figure 4. Deeks' Funnel Plot and Asymmetry Test for publication bias. No significant slope coefficient (p=0.18) was observed and generally random study diagnostic odds ration distribution is illustrative of likely lack of publication bias. When the regression line has a significant slope (p<0.05 or <0.10), then a relationship is observed between diagnostic odds ratio and the size of the study and indicated likely publication bias.

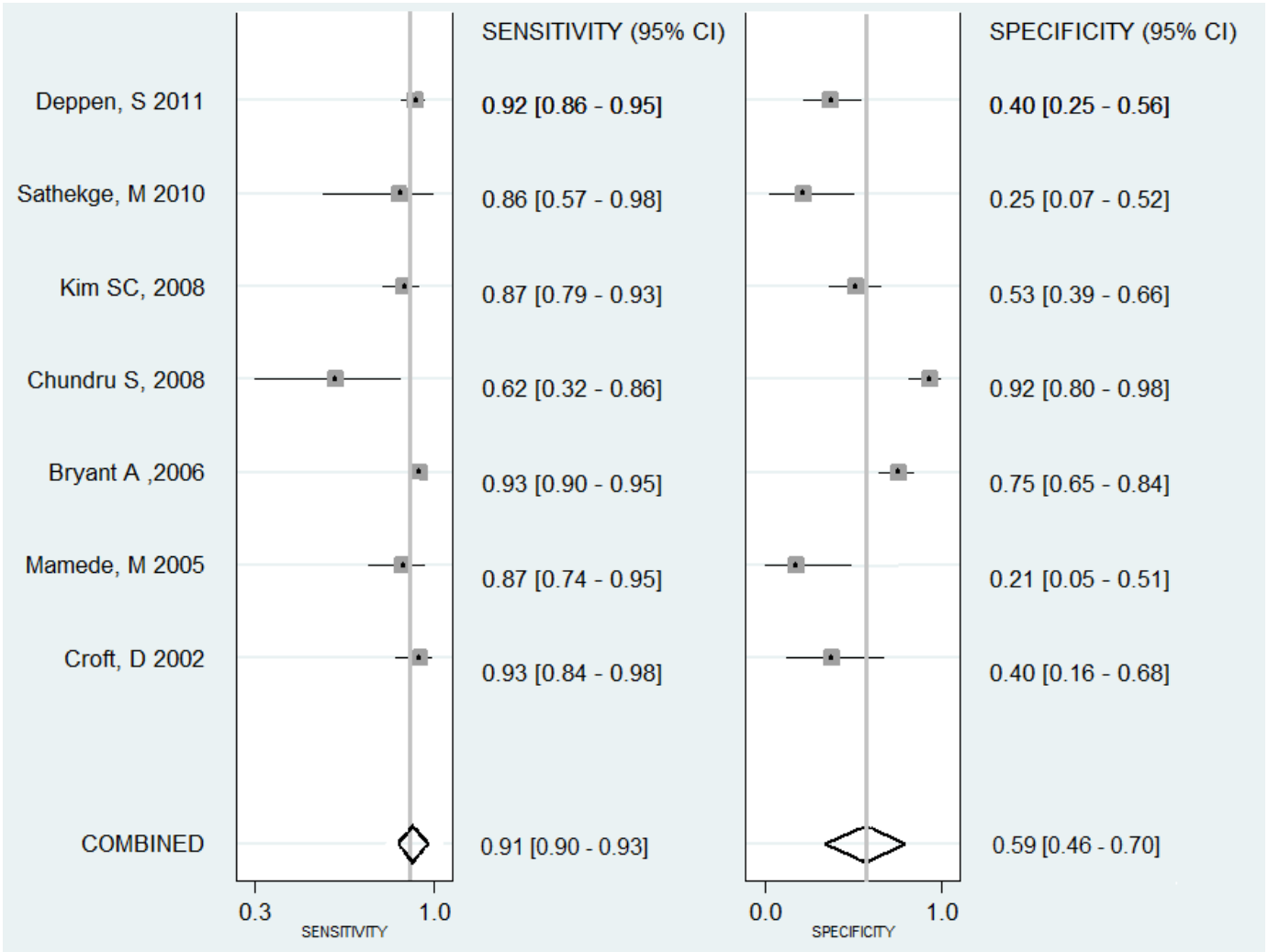


Figure 5. Sub-group analysis of 7 studies reporting endemic infectious lung disease, Forest plot. Sensitivity, specificity and their confidence intervals are estimated using meta-analysis model with multiple imputation for covariates of reader blinding and mean lesion size.

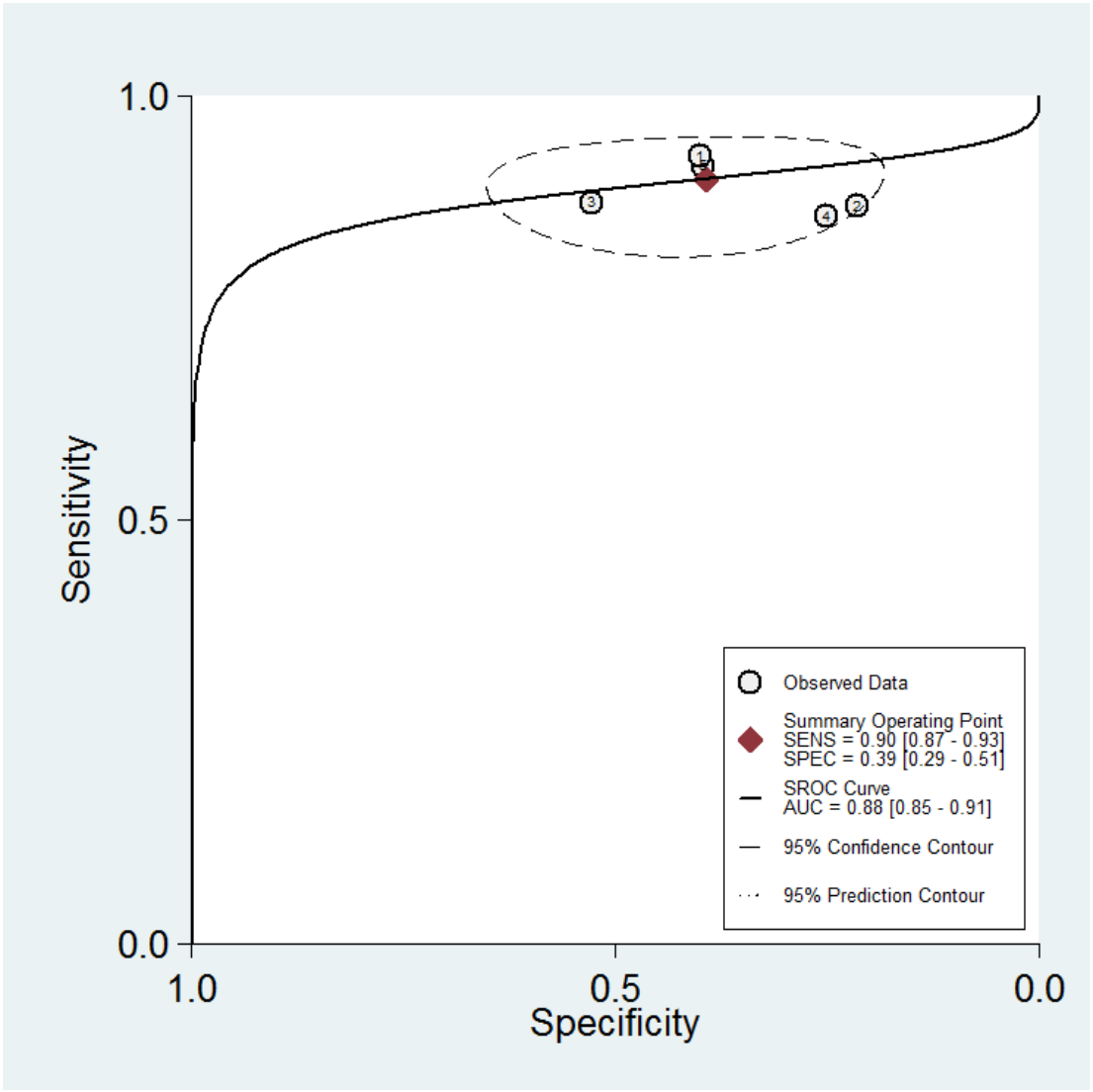


Figure 6. SROC with random effects model for 5 studies reporting endemic infectious lung disease.

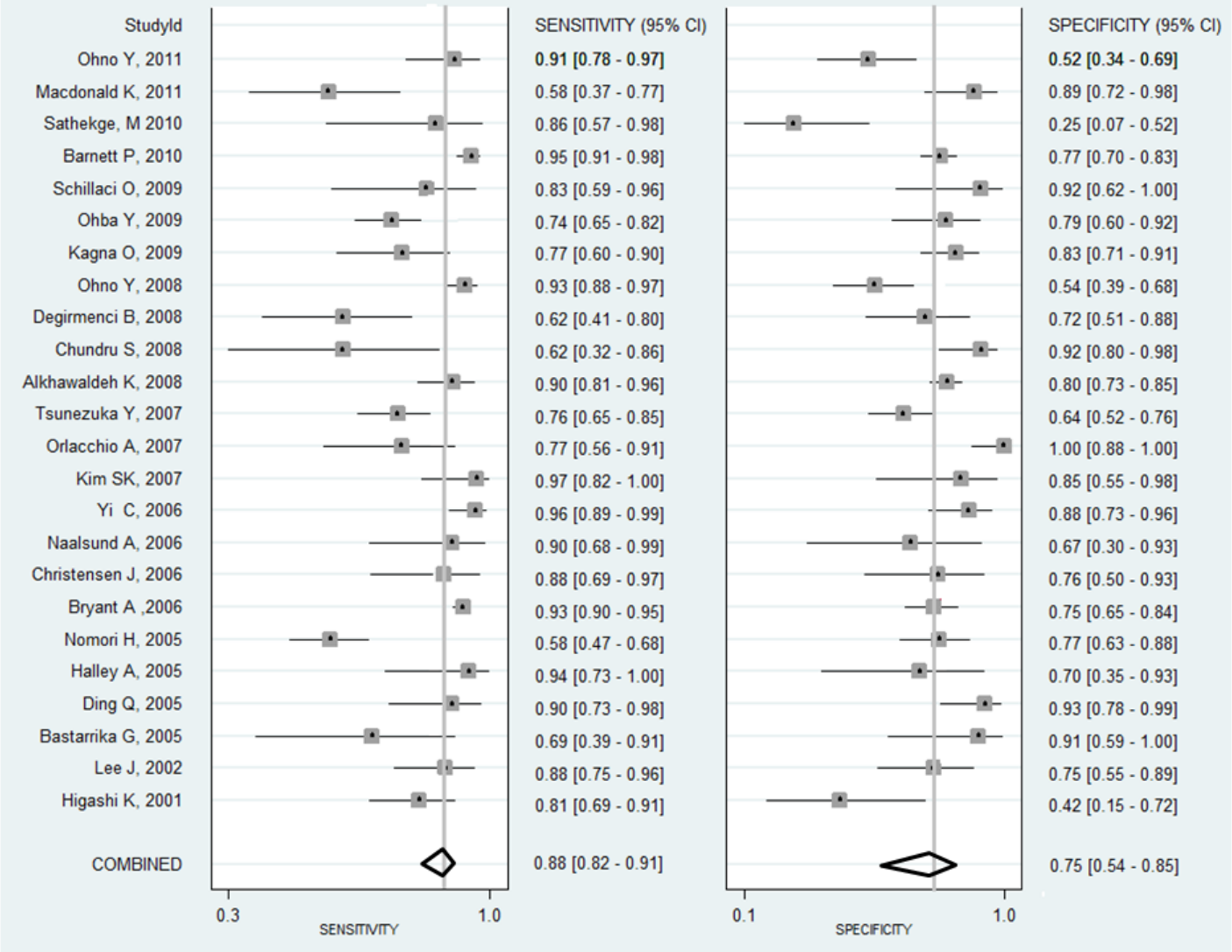


Figure 7. Sub-group analysis of studies with mean or median lesion size less than or equal to 20mm in diameter, Forest plot. Combined sensitivity, specificity and their confidence intervals are estimated using meta-analysis model with multiple imputation for covariates of reader blinding and mean lesion size.

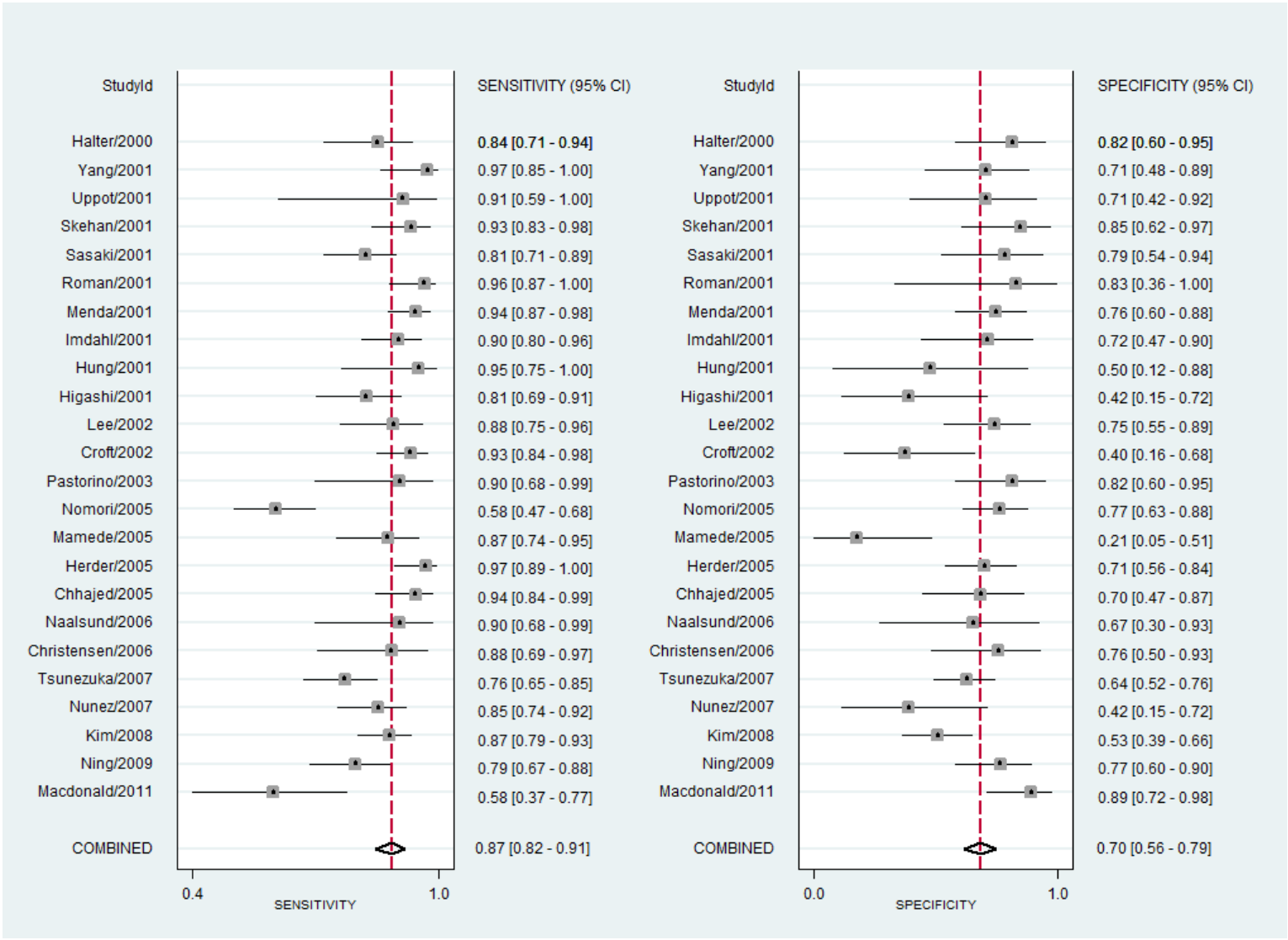


Figure 8. Sub-group analysis of studies reporting use of PET only scanners, Forest plot. Combined sensitivity, specificity and their confidence intervals are estimated using meta-analysis model with multiple imputation for covariates of reader blinding and mean lesion size. Reported studies are only those without missing data.

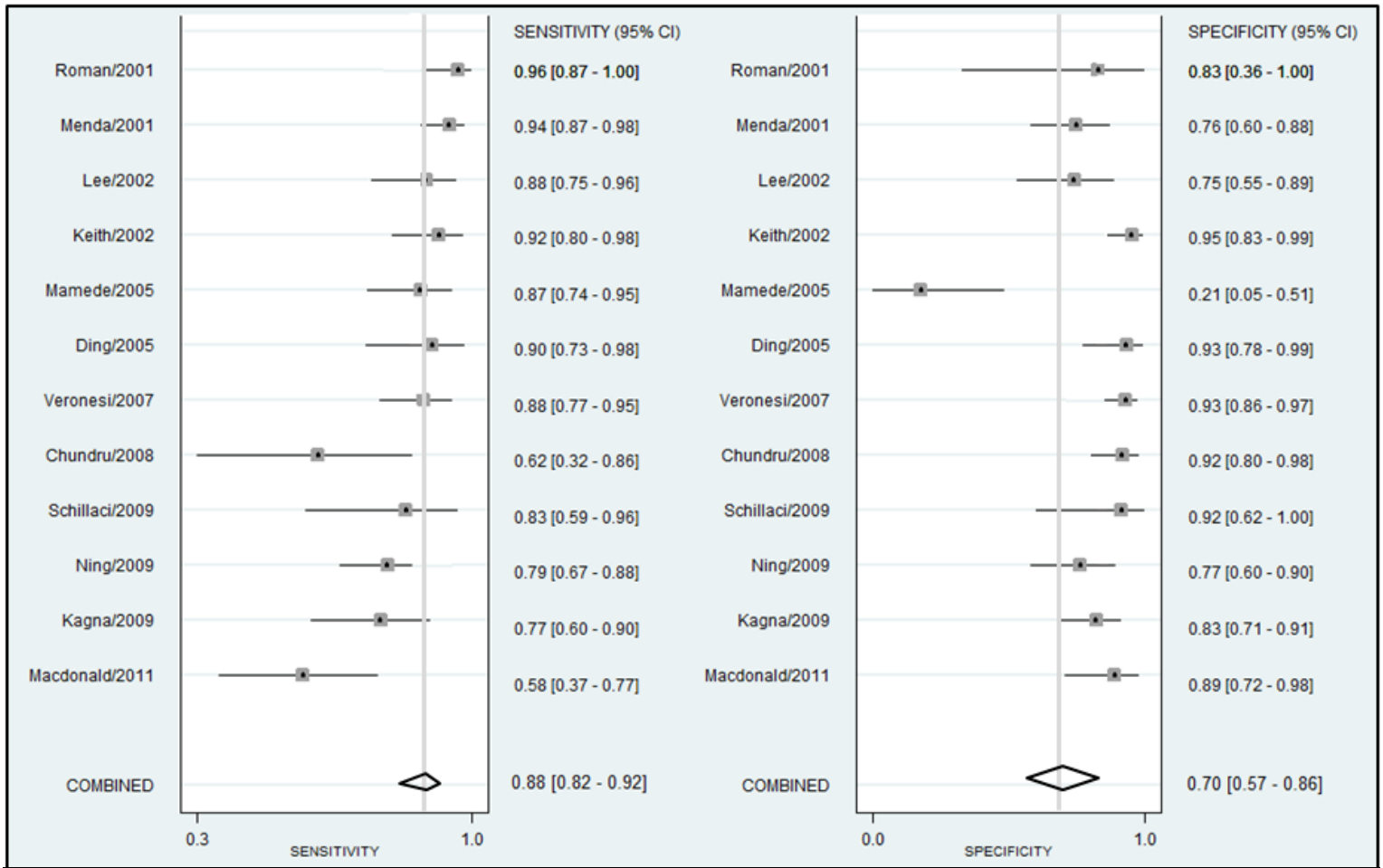


Figure 9. Sub-group analysis of studies reporting blinding of readers to patient history, Forest plot. Combined sensitivity, specificity and their confidence intervals are estimated using meta-analysis model with multiple imputation for covariates of reader blinding and mean lesion size. Reported studies are only those without missing data.

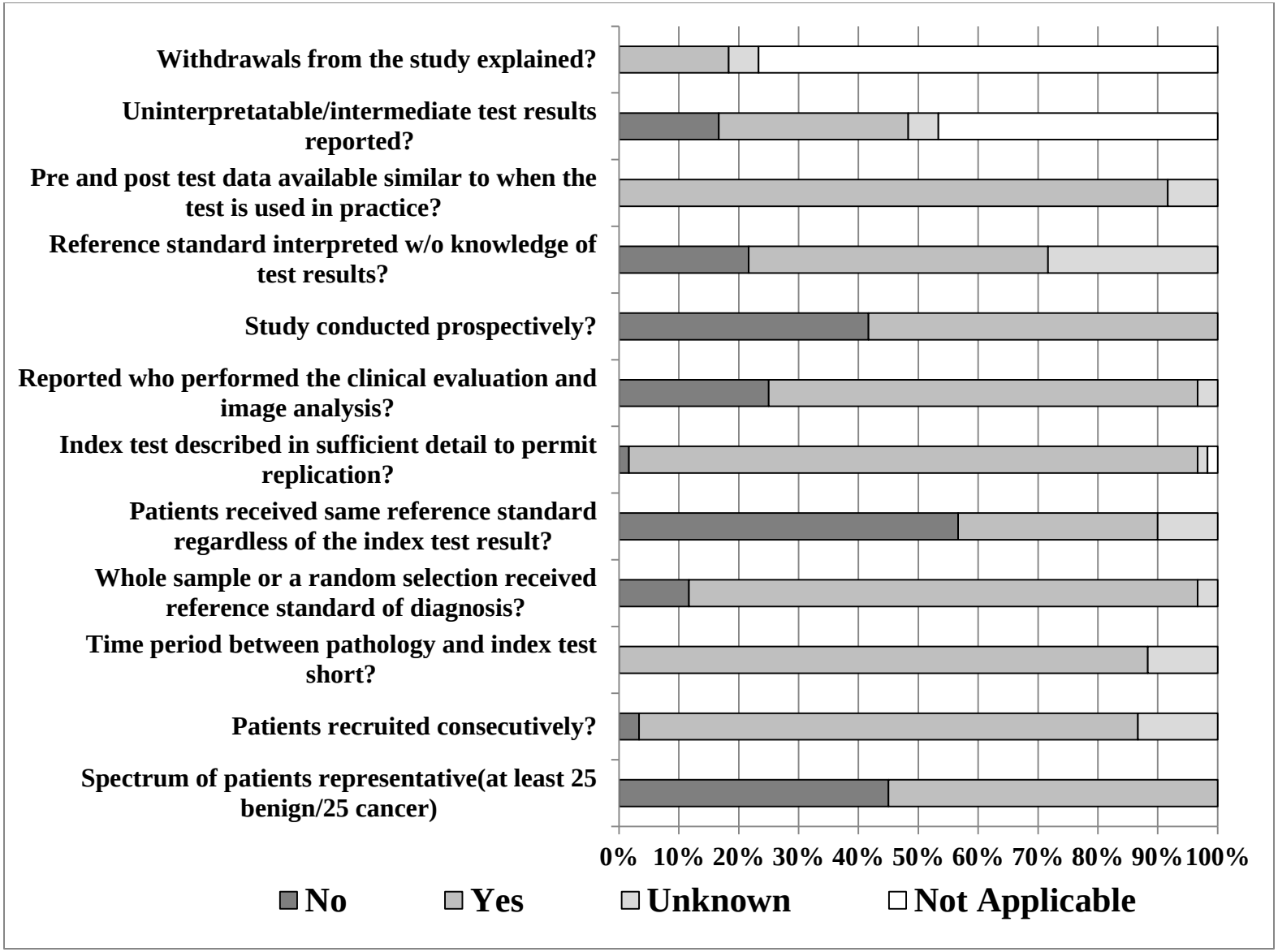


Figure 10. QUADAS quality metrics reported for each study by 2 reviewers.

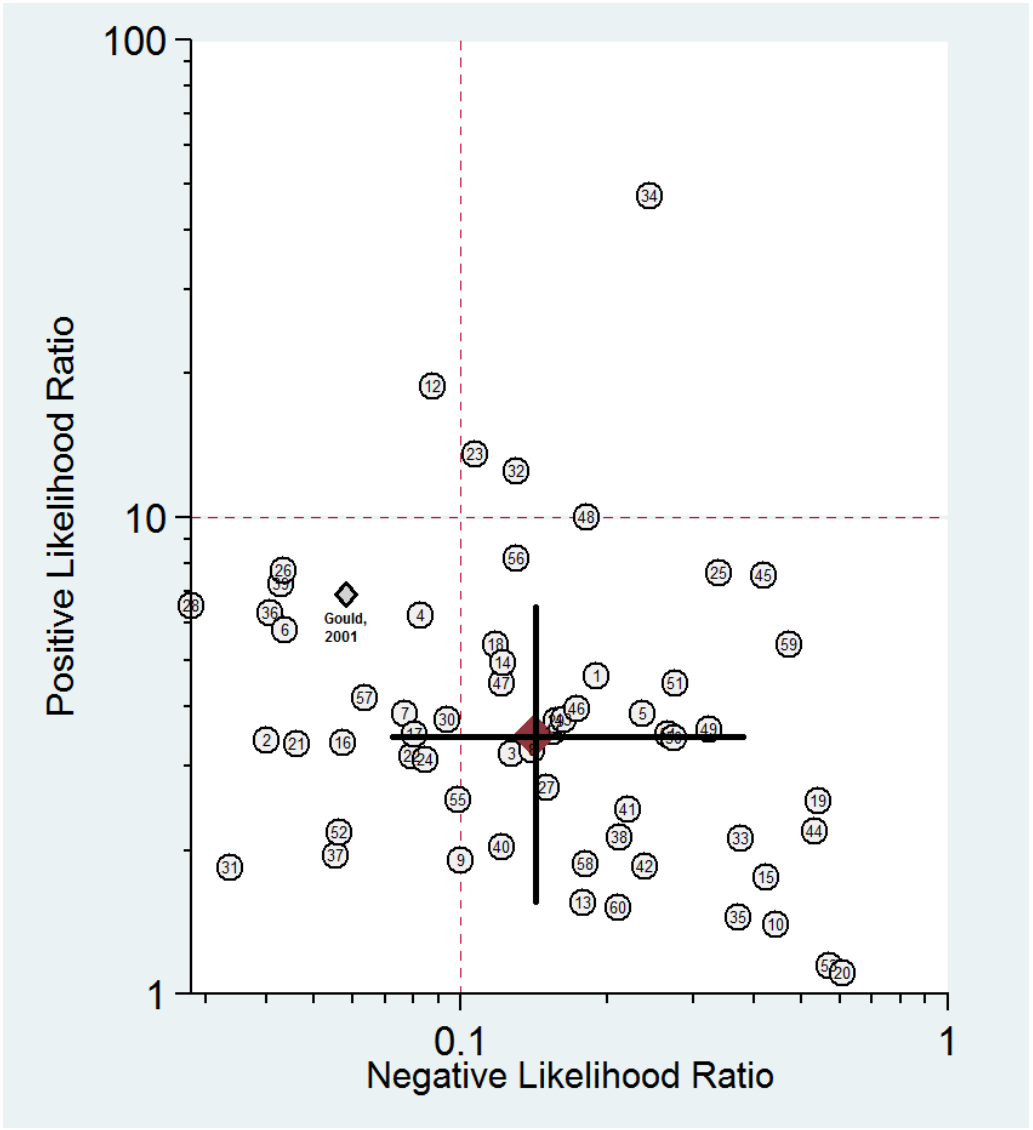


Figure 11. Likelihood ratio graph for all studies. Diamond and crosshatch is the estimated positive (3.56) and negative (0.147) likelihood ratio from the random effects meta-analysis model with covariates and the 95% confidence interval for positive (1.52, 6.71) and negative (0.07, 0.39) likelihood ratio. The gray diamond represents the results from the Gould 2001 meta-analysis.

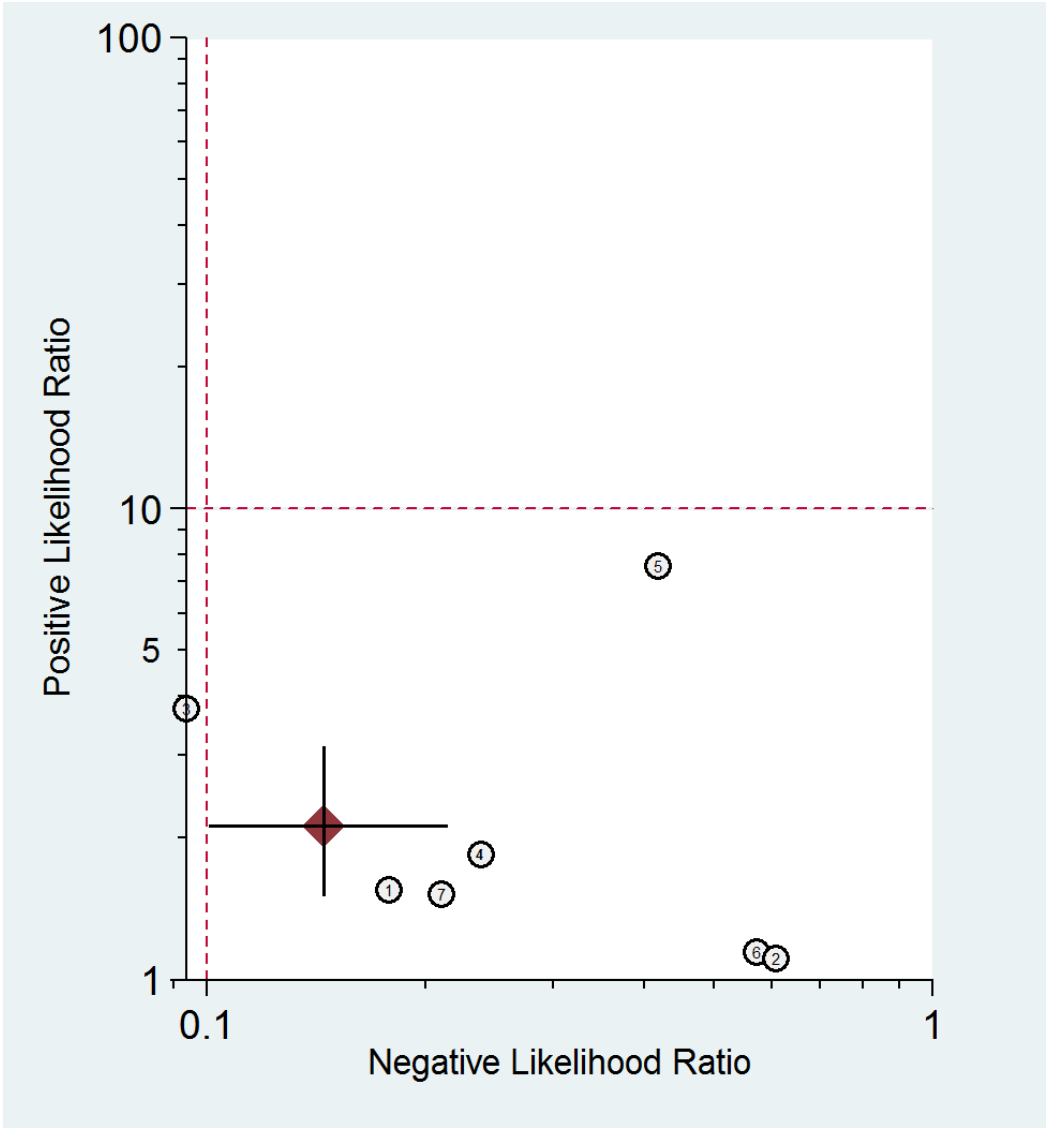


Figure 12. Likelihood ratio graph for seven studies reporting endemic infectious lung disease. Diamond and crosshatch is the estimated positive (2.21) and negative (0.15) likelihood ratio from the random effects meta-analysis model with covariates and the 95% confidence interval for positive (1.67, 3.1) and negative (0.10, 0.22) likelihood ratio.

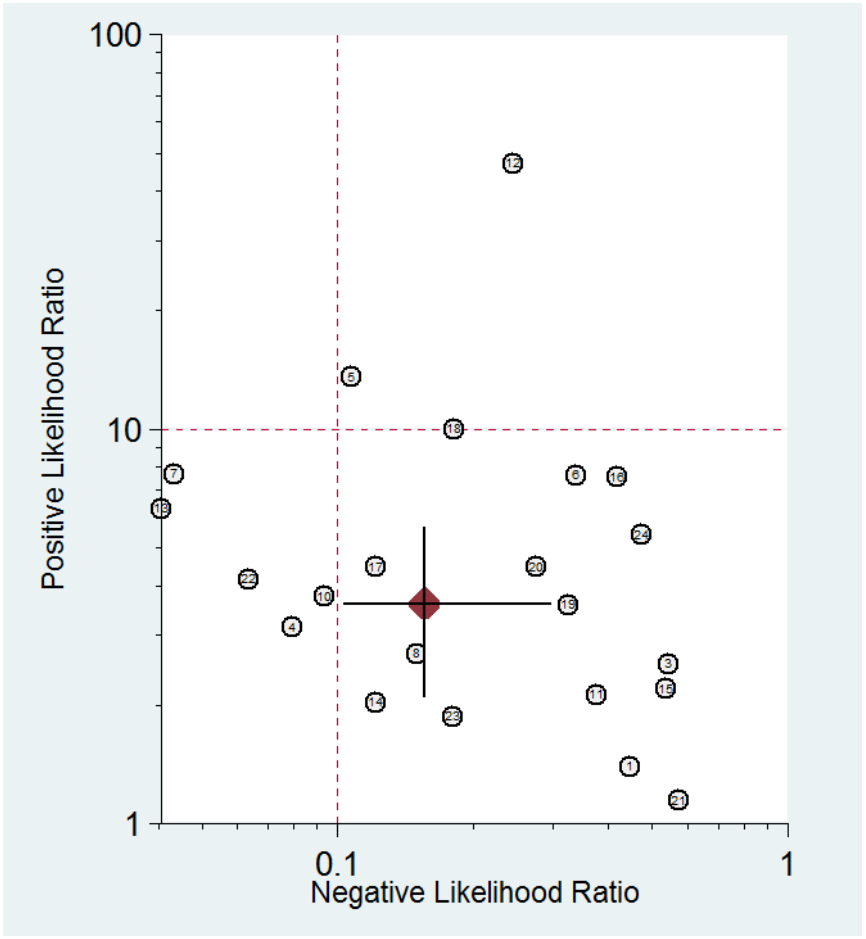


Figure 13. Likelihood ratio graph for seven studies reporting mean or median lesion size less than or equal to 20 mm. Diamond and crosshatch is the estimated positive (3.75) and negative (0.16) likelihood ratio from the random effects meta-analysis model with covariates and the 95% confidence interval for positive (2.01, 6.35) and negative (0.11, 0.30) likelihood ratio.

Chapter 5

V. Summary and Future Directions

This dissertation sought to help surgeons diagnose individuals with lung lesions suspicious for lung cancer. Increasingly, clinicians rely on radiographic imaging for the identification and diagnosis of possible lung cancer, and they use the same imaging to determine the lesions' characteristics non-invasively. Discriminating lung cancer from other benign etiologies is complicated by benign granulomatous disease of the lung that can physically and metabolically appear similar to common lung neoplasms upon imaging. Using a retrospectively collected cohort of individuals being evaluated by thoracic surgeons for possible lung cancer, I estimated and internally validated a clinical predictive model using data available to surgeons at the time of operative evaluation. This highly enriched dataset contains multiple diagnostic tests, some of which were not available prior to this point in the diagnostic work-up of the patient.

Variation in benign disease prevalence after lung surgery by state was also estimated in a separate analysis of Medicare administrative data. If a screening program using low dose CT scans to detect lung cancer is implemented nationally, then variation in the prevalence of benign disease after lung surgery may cloud the efficacy of screening and indicate possible local or systemic factors that gave rise to the observed variation. Given the importance of FDG-PET scans in the diagnostic process, factors that influence the accuracy of FDG-PET scans were explored through a systematic review and meta-analysis of the published literature and through secondary analysis of the nationally conducted ACOSOG Z4031 trial in clinical stage 1 lung cancer. The association between observed benign disease and fungal lung disease exposure in the Medicare population and in a separate analysis of false positive FDG-PET scans

in the ACOSOG trial was conducted using historic surveillance information with fungal lung disease prevalence by county. A summary of the results is outlined below:

1. Currently available models for estimating lung cancer risk after the discovery of a lung nodule or lesion were calibrated to populations with far lower prevalence of disease than what was found among individuals being evaluated for surgical biopsy. The more extensive diagnostic data available to surgeons allowed for the estimation of a predictive model, the TREAT model, that better discriminated cancer from benign disease (AUC 0.89 95%CI: 0.86, 0.92) when compared to the Mayo model, using the Mayo model's published point estimates (AUC 0.80: 95%CI: 0.76, 0.85) as well as the Mayo model's variables and re-estimating the coefficients of those variables within the Vanderbilt Lung Cancer Cohort (AUC 0.83; 95%CI: 0.79, 0.87). The TREAT model was better calibrated within the Vanderbilt Lung Cancer Cohort (Brier score 0.12) when compared to the Mayo model which estimated cancer likelihood using published point estimates (Brier score 0.17) and was the same for the Mayo model re-estimated within the Vanderbilt cohort (Brier score 0.12). Little optimism was observed after internal validation of the TREAT model using a bootstrap method with an AUC of 0.87 and Brier score of 0.13 (Chapter 2).

2. Benign disease prevalence after lung surgery varied widely by state in the 2009 Medicare population. The mean prevalence across the US was 9.1% and the median state point prevalence was 8.9% (IRQ 7.8%, 10.9%). The lowest prevalence was observed in Vermont (1.2%) and the highest in Hawaii (25%). After excluding these two states as possible outliers, significant differences in benign disease prevalence after surgery remained (p=0.001) in the 48 states examined. An in-hospital mortality rate of 2.1% was observed among those patients with benign disease. No differences in estimated fungal disease exposure between patients with cancer and patients with benign disease were observed (p=0.90), although the measurement of infectious disease exposure using historical exposure in young white men was likely flawed. The cause of

observed differences in benign disease prevalence between states is not known but may be due to practice variation, work-up bias or locally endemic lung diseases. (Chapter 3).

3. The accuracy of FDG-PET scans in diagnosing lung cancer was examined in the completed ACOSOG Z4031 trial. The trial included only those individuals with clinical stage 1 lung cancer. All participants underwent surgery and pathological diagnosis of their lung lesion. The clinical characteristics of the lesions evaluated in the ACOSOG trial were similar to those found in the recently completed National Lung Screening Trial, thus giving possible insight into the accuracy of FDG-PET scans in diagnosing lung cancer in screening populations. Among the 692 individuals with usable FDG-PET scans, lung cancer prevalence was 83%. The sensitivity (82%; 95%CI: 79%, 85%) and specificity (31%; 95%CI: 23%, 40%) of FDG-PET to diagnose lung cancer were much lower than that published in previous meta-analyses. Significant variation in sensitivity ($p=0.03$) was found between eight enrolling cities that recruited at least 25 participants. The sensitivity varied from a low of 67% to a high of 91%. FDG-PET scan accuracy increased with increasing lung lesion size. A separate analysis found no association between false positive scan results and exposure to fungal lung diseases (histoplasmosis, blastomycosis and coccidiomycosis) using the historical prevalence of fungal disease during the 1960s to estimate fungal lung disease exposure. No association was observed between false positive scans and fungal lung disease after adjusting for age and lung lesion size on pre-operative CT scans. (Chapter 3).

4. A systematic review of published and unpublished literature found 1,231 abstracts that investigated FDG-PET scan accuracy to diagnose lung cancer. After initial review, data abstraction was performed on 218 published articles. Articles that reported staging and not diagnosis, or included only lung cancer cases and not benign disease cases, were the most commonly excluded from the review. Sixty of the 218 abstracted articles were included for meta-

analysis. Overall pooled sensitivity among the 60 reports was 89% (95%CI: 87%, 91%) and specificity was 74% (95%CI: 70%,78%); however, significant heterogeneity was observed across studies beyond what was expected from differences in threshold sensitivity and specificity choice between studies. A binomial random effects model was estimated that included covariates of clinical significance. Those covariates were: lesion size, scanner type, method of blinding radiograph readers and presence of endemic infectious lung disease. Heterogeneity between studies remained after inclusion of these covariates. The estimated sensitivity was 89% (95%CI: 82%, 92%) and specificity was 75% (95%CI: 46%, 86%). Sub-group analysis found differences in FDG-PET accuracy by scanner type. PET only scanners (n=25 studies) had significantly lower sensitivity (88%) ($p<0.001$) and specificity (69%) ($p=0.02$) when compared to combined PET/CT scans (sensitivity 90% and specificity 78%, n=29 studies). In 24 studies with lung lesions less than or equal to 20 mm in diameter, sensitivity was 88% compared to 91% in the 19 studies whose mean or median lesion size was greater than 20 mm ($p=0.05$). The specificity was similar between the two groups. Seven studies from populations with high endemic infectious lung disease reported lower specificity (59%) and similar sensitivity (90%) to the 53 studies that did not report infectious lung disease in the studied population. The diagnostic accuracy of FDG-PET varied little across the 11 years reviewed, and no evidence of publication bias was observed across the 60 studies (Chapter 4).

A clinically viable predictive model designed for surgeons evaluating lung nodules suspicious for lung cancer was successfully estimated and internally validated. Collectively this analysis also found that FDG-PET's performance in diagnosing lung cancer was much poorer than previously published. These results call into question the use of this expensive, non-invasive test for diagnosis of smaller lung nodules or in populations where endemic infectious disease is high. This sensitive test remains an important predictor for lung cancer, and the use of FDG-PET for staging of lung cancer was not addressed in any of

the research conducted in this dissertation. The prevalence of benign disease after surgery was found to vary widely across the US, but the reasons for the observed variation are not well known. Practice variation and work-up bias, infectious lung disease induced granuloma, and other unknown factors are all possible causes. No association between fungal lung disease and false positive FDG-PET scans was observed. Understanding the factors that influence benign disease prevalence and patient selection prior to surgery are necessary steps in exploring relevant predictors of local variation to be added to a nationally applicable model to predict lung cancer for surgeons. This research raised a number of questions to be addressed in future work. They include:

1. The proposed TREAT model for lung cancer appears promising and should be externally validated. External validation would ideally be conducted in a variety of datasets that vary across the spectrum of fungal lung disease prevalence. This is an indirect method of determining whether a variable is missing from the model. One would expect that as the prevalence of fungal lung disease decreases, the accuracy of radiographic imaging to diagnose lung cancer increases. Thus the amount of variance explained by those variables collectively should increase. More complex time–to-event models or a linear model with interactions between variables and non-linear relationships should be explored as more data becomes available. Finally, local factors like cancer prevalence at the practice level, availability of specialists, and diagnostic practice variation should be explored to make the model more robust in a national setting.

2. The estimates for fungal lung disease exposure used in these analyses were indirect and likely flawed. The estimates used county and regional level prevalence measured in naval recruits from the 1960s. This young, healthy population who lived in one location for their entire lives was not representative of the much older and more mobile population being evaluated for lung cancer. Misclassification in the current analysis was likely. Current demographics and land use have changed greatly over the past 50 years and some evidence shows migration of fungal lung disease

to the upper Midwest. Since histoplasmosis and other fungal lung diseases reside primarily in soils, a land cover analysis using geographical information systems and available satellite imaging of land use could be conducted to estimate exposure based upon soil conditions, land cover impermeability and areas of possible exposure.

3. Measurement of fungal lung disease exposure using land maps and historical prevalences are both indirect measurements of exposure and are subject to misclassification and ecological bias. Using serum and lung tissue samples on hand in the Vanderbilt Lung Spore repository, researchers can directly measure the presence of fungal lung disease in these biological samples and link those results to available imaging results. Such data would allow the estimation of the prevalence of fungal lung disease in false positive imaging. Subsequently, investigation into radiographic or biomarker evidence indicating presence of fungal lung disease or granuloma should be pursued, as currently available biomarkers for lung cancer diagnosis have achieved mixed results.[1]

4. If fungal lung disease complicates the diagnosis of lung cancer and leads to higher rates of benign disease, then a natural experiment should be possible where clinicians trained in areas with low prevalence of infectious lung disease and who subsequently moved to areas with high prevalence of disease would exhibit a change in their patients' benign disease prevalence over time. This change in practice would occur as clinicians learn how to incorporate fungal lung disease into their diagnostic process. Essentially, one would expect clinicians to learn how to recognize fungal lung disease over time, and their prevalence of benign disease after procedures should decrease. Using contacts through clinical societies and training programs, we would contact individual surgeons and pulmonologists who were trained in areas where fungal lung disease is rare and currently practice in areas of high prevalence and measure their prevalence of benign disease over time. A similar experiment could be conducted for radiologists.

5. The causes of the observed variation in benign disease prevalence by state in the Medicare population are unknown at this time. The implementation of technological innovations like video assisted thoracotomy and robotic assisted surgery has increased the population who can receive surgical biopsy for the diagnosis of lung cancer. One center found their benign disease prevalence after surgery doubled with the implementation of the newer video assisted thoracotomy when compared to the older surgical technique.[2] These and other factors that represent practice variation or work-up bias may influence the benign disease prevalence across states and over time. Prevalence variation between states and across years should be explored in national datasets. Understanding how broad changes to clinical practice and local factors influence benign disease prevalence in populations being evaluated for lung cancer can inform policy makers evaluating the efficacy of a national screening program for lung cancer.

6. The systematic review and meta-analysis for FDG-PET diagnostic accuracy was conducted through 2001 and should be updated to include research published since 2011 and the results should be disseminated. Using the results of the meta-analysis, we will conduct a cost-effectiveness study to determine whether populations exist that would not likely benefit from a FDG-PET scan to diagnose lung cancer.

Given the results of the National Lung Screening Trial (NLST) and the support for screening healthy, high risk individuals with low dose computed tomography by clinical and patient advocacy groups, we will likely see screening for lung cancer in the near future. The implementation of a national screening program for lung cancer will increase the volume of CT-discovered lung anomalies requiring diagnosis by approximately 2.6 million, extrapolating from the results of the NLST.[3] An estimated 80,000 diagnostic operations will be conducted as part of the diagnostic process. Surgeons evaluating lung nodules and

weighing the decision to undertake a diagnostic operation need a clinically useful model to help in that decision process. This research is the first step toward developing such a model. The initial results of the TREAT model are promising and should be externally validated. The impact of fungal lung disease on diagnostic imaging remains unclear. Whether the observed variances in FDG-PET scans arose from local practice variation or from endemic infectious lung disease was not established in this research. However, a growing body of evidence has shown that FDG-PET scans perform poorly in the diagnosis of smaller lesions or in regions where the patient population has been exposed to infectious lung disease. As the US moves toward screening for lung cancer with low dose CT scans, researchers should examine why the prevalence of benign disease varies in order to inform health policy and clinical guidelines as we seek to better diagnose and treat this deadly disease.

5.1 References

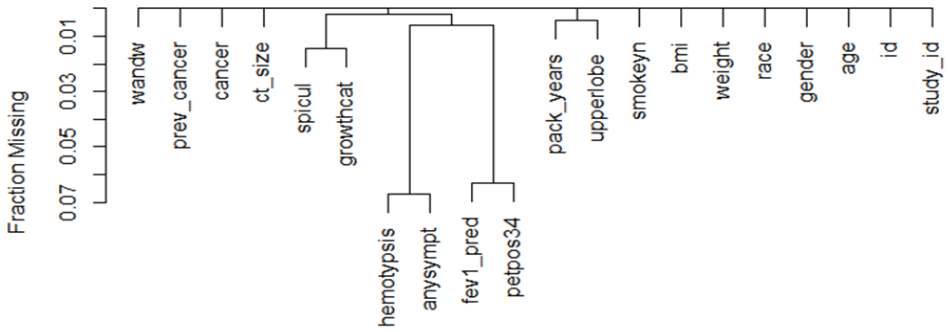
1. Hassanein M, Callison JC, Callaway-Lane C, Aldrich MC, Grogan EL, Massion PP. The state of molecular biomarkers for the early detection of lung cancer. *Cancer Prev Res.* Aug 2012;5(8):992-1006.

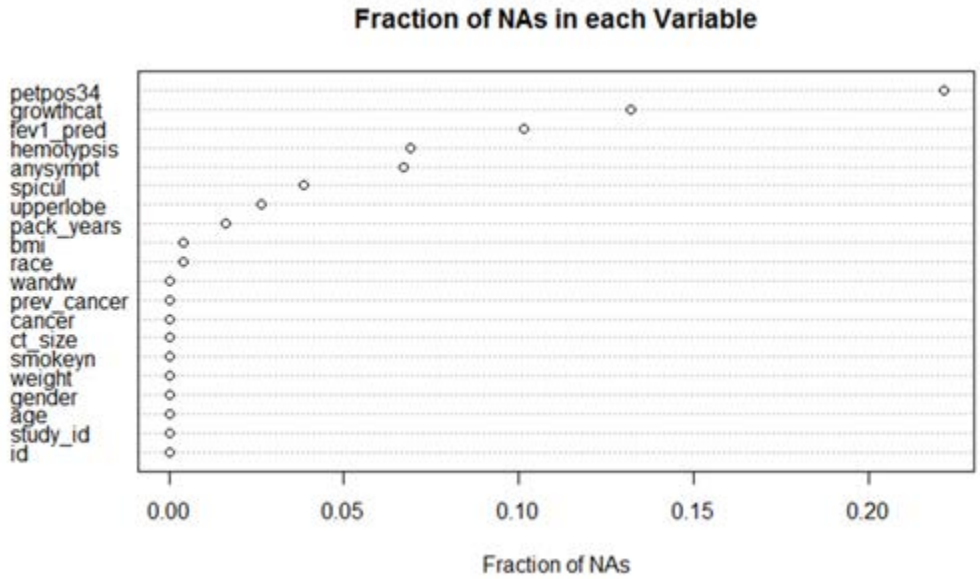
2. Kuo E, Bharat A, Bontumasi N, et al. Impact of Video-Assisted Thoracoscopic Surgery on Benign Resections for Solitary Pulmonary Nodules. *Ann Thorac Surg.* 2012;93(1):266-273.

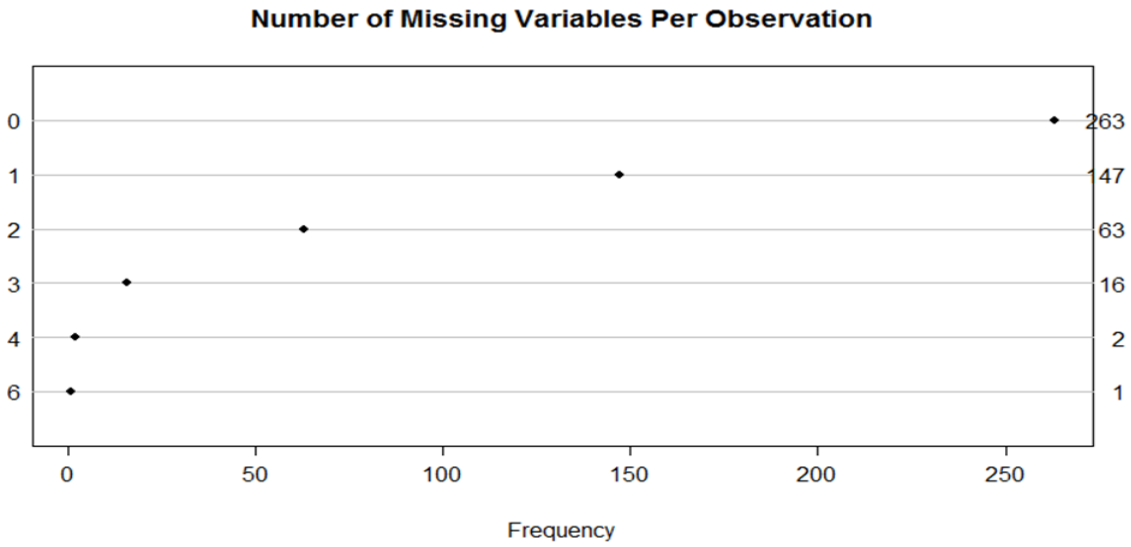
3. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *N Engl J Med.* 2011;365(5):395-409.

Appendix 1

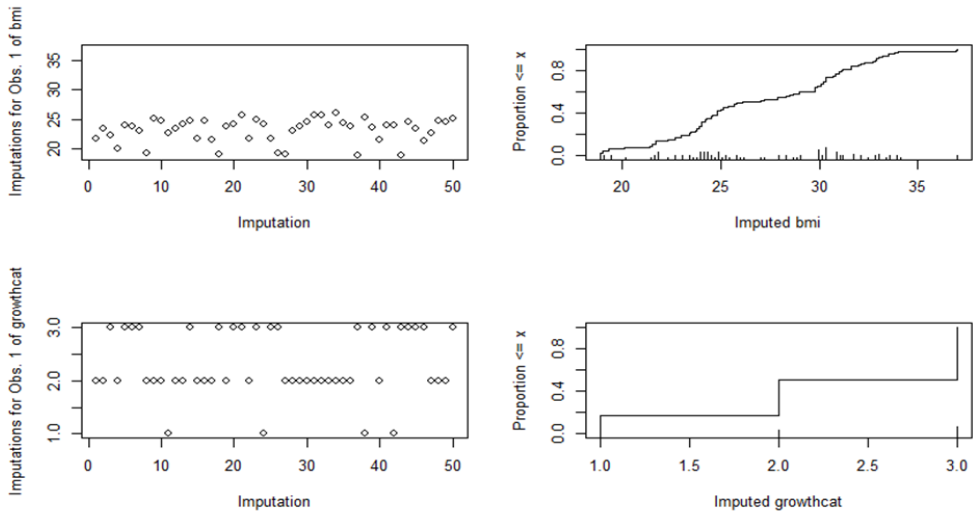
Appendix 1.1 Cluster analysis – missing data and missing data patterns



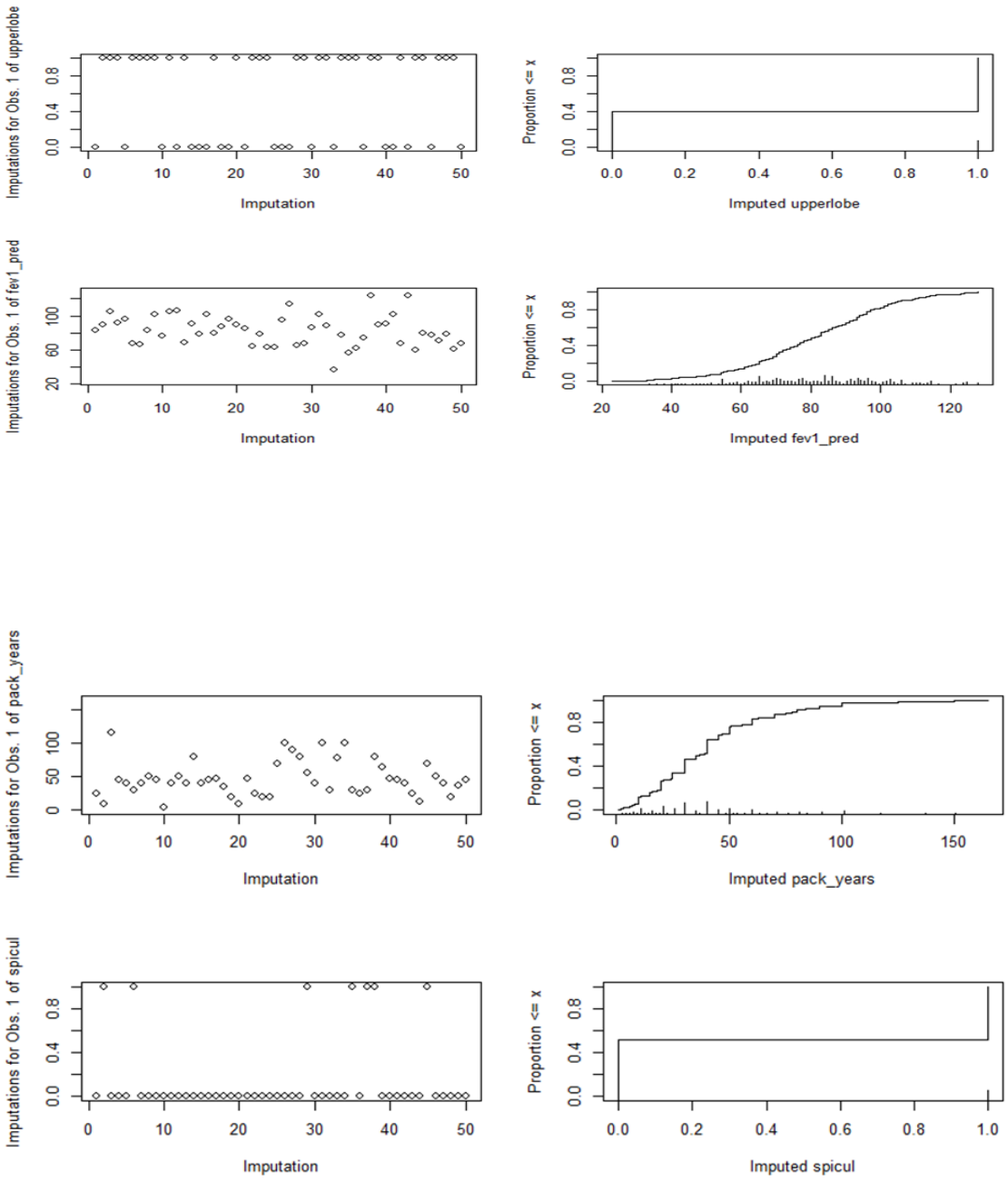




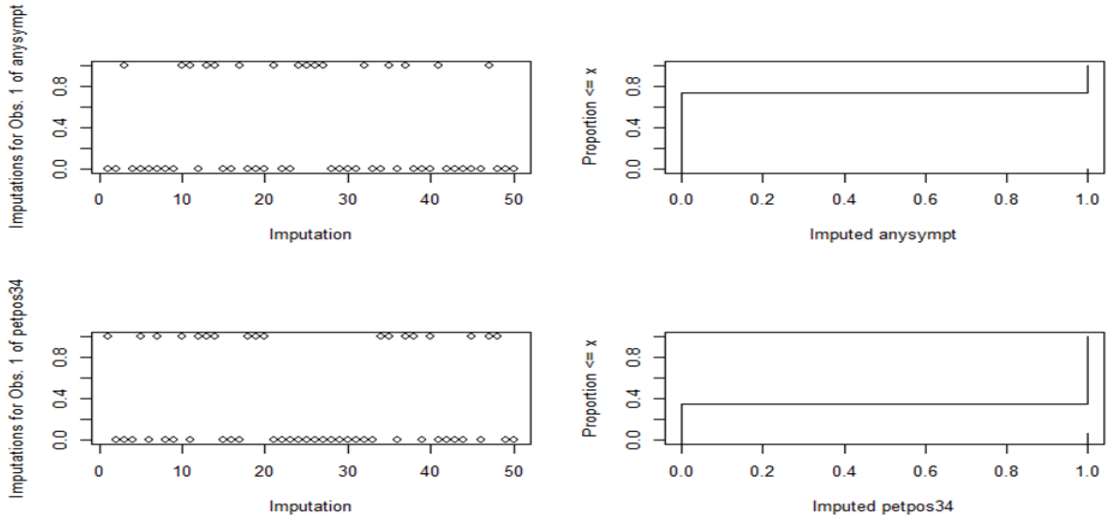
Appendix 1.2 Results of imputation for the 1st observation across 50 imputed datasets.



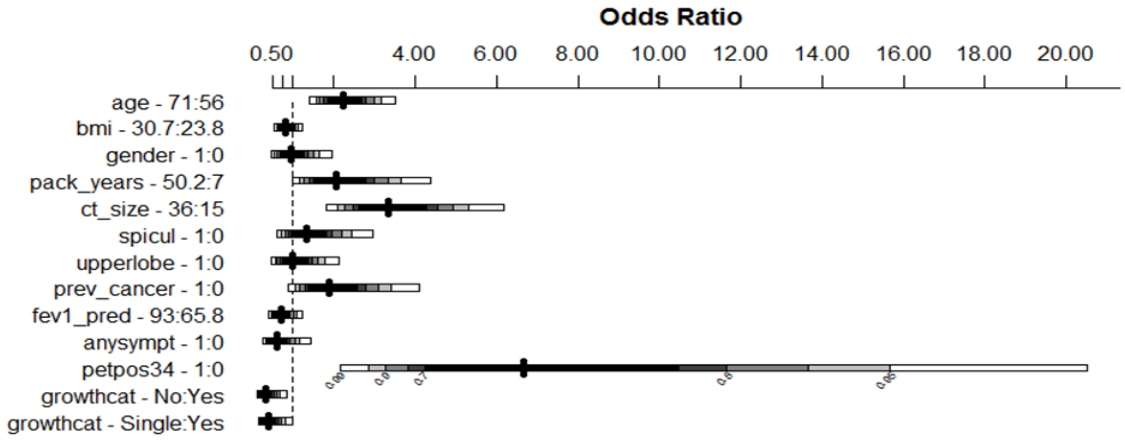
Appendix 1.2 Results of imputation for the 1st observation across 50 imputed datasets. (Continued)



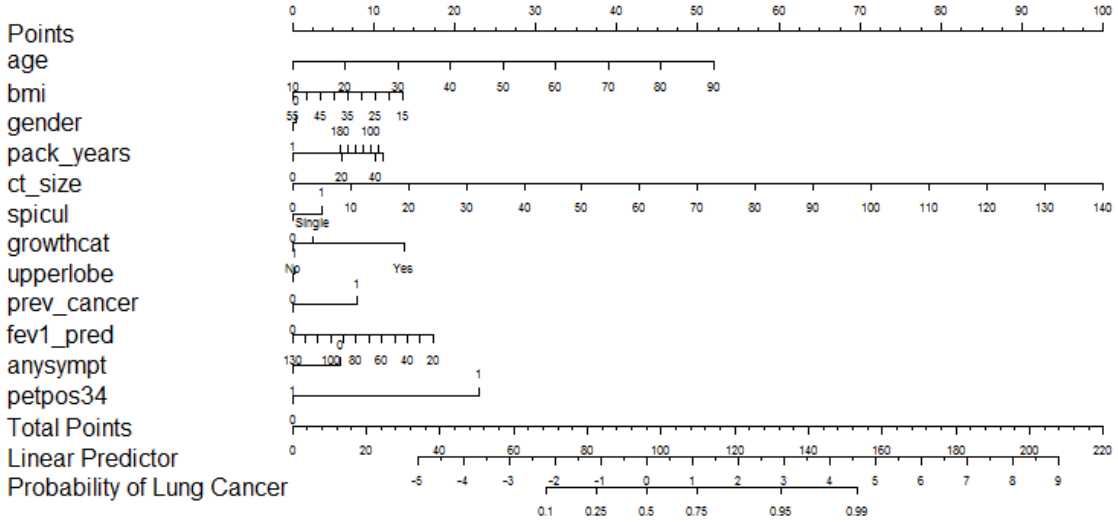
Appendix 1.2 Results of imputation for the 1st observation across 50 imputed datasets. (Continued)



Appendix 1.3 Odds Ratios and confidence bars, using dichotomous differences or quartiles of continuous variables for assessing the effects on the odds of lung cancer.



Appendix 1.4 Nomogram for TREAT lung cancer model



R code

```
library("Hmisc")
library("rms")
library("boot")
library("rpart")

## get data from shared directory 496 observations and 30 variables
predict.out.09062012 <- read.delim("F:/Thoracic Surgery Research/Common folder/Lung Cancer Cohort/Nodule study/predictive model/predict out 10162012.txt")

describe (na.detail.response=TRUE, predict.out.09062012)
s<-summary(cancer~ age + gender + bmi + smokeyn + pack_years + ct_size + spicul + upperlobe + prev_cancer +  fev1_pred + anysympt + growthcat + petpos34, data=predict.out.09062012)
s
w<-latex(s)
plot(s)
par(mfrow=c(1,1))
nimp <- 50   ### number of imputation
set.seed (5690)
nx<-naclus(predict.out.09062012)
plot(nx); naplot(nx) #Show patterns of NAs - missing data

f  <- aregImpute(~cancer + age + wandw+ bmi + gender + growthcat+ smokeyn + pack_years + ct_size + spicul + upperlobe + prev_cancer +  fev1_pred + anysympt  + weight + petpos34, n.impute=nimp, x=TRUE, nk=3, tlinear=F, data=predict.out.09062012)
par(mfrow=c(2,2))
print (f, digits=3)
plot (f, nclass=NULL, type=c('ecdf','hist'),datadensity=c("hist", "none", "rug", "density"), diagnostics=T, maxn=1)
```

```
m<-lrm(is.na(petpos34) ~ cancer + age + wandw+ bmi + gender + growthcat+ smokeyn +
pack_years + ct_size + spicul + upperlobe + prev_cancer +  fev1_pred + anysympt,
data=predict.out.09062012 )
m

#fit the logistic regression model assuming additive variables and non-linear pack years
lungCA <- fit.mult.impute(cancer~ age + bmi + gender+rcs(pack_years,3) + ct_size + spicul +
growthcat + upperlobe +  prev_cancer +  fev1_pred + anysympt + petpos34, lrm, f,
data=predict.out.09062012)

varcov0<-vcov(lungCA)
serror0<-sqrt(diag(vcov(lungCA)))

options(digits=3)
post.anova<-anova(lungCA)
par(mfrow=c(1,1))
plot(post.anova)
print(post.anova)
print(lungCA)
exp(cbind(OR=coef(lungCA), confint.default(lungCA)))##OR and confidence interval using
reported std errors


lungCA.test<-lungCA[[3]]
Ctrain<-lungCA.test [6]## Apparent C index
Btrain<-lungCA.test [11]## Apparent Brier Score
Btrain
Ctrain

##Heidi's OR and 95% CI code
lrm_OR <- function(lrmin) {

  temp <- data.frame("coef"=lrmin$coefficients)
  temp$se <- NA
  for ( i in 1:nrow(temp)) {
    temp$se[i] <- sqrt(lrmin$var[i,i])
  }

  temp$OR <- round(exp(temp$coef),3)
  temp$LCI <- round(exp(temp$coef-1.96*temp$se),3)
  temp$UCI <- round(exp(temp$coef+1.96*temp$se),3)
  temp$CI <- paste(temp$LCI,temp$UCI,sep=" to ")
  temp$p <- round(1-pchisq((temp$coef/temp$se)^2,1) ,4)
  colnames(temp) <- c("Coef", "se", "Odds Ratio","LCI", "UCI", "CI","p-value")
  print(temp[,c("Coef", "se", "Odds Ratio","CI","p-value" )])

}
lrm_OR(lungCA)

#complete case analysis
```

```
lungCA.complete <- lrm(cancer~ age + bmi + gender+rcs(pack_years,3) + ct_size + spicul +
growthcat + upperlobe +  prev_cancer +  fev1_pred + anysympt + petpos34,
data=predict.out.09062012)
varcov0<-vcov(lungCA.complete)
serror0<-sqrt(diag(vcov(lungCA.complete)))

options(digits=3)
post.anova.complete<-anova(lungCA.complete)
par(mfrow=c(1,1))
plot(post.anova.complete)
print(post.anova.complete)
lungCA.test<-lungCA.complete[[3]]
Ctrain.c<-lungCA.test [6]## Apparent C index
Btrain.c<-lungCA.test [11]## Apparent Brier Score
Btrain.c
Ctrain.c

#overfitted model
overfit.lungCA <- fit.mult.impute(cancer~ age + bmi + gender+rcs(pack_years,4) + rcs(ct_size,4)
+ spicul + growthcat + upperlobe +  prev_cancer +  rcs(fev1_pred,4) + anysympt + petpos34 +
petpos34%ia%rcs(ct_size,4) + anysympt%ia%rcs(fev1_pred,4) + petpos34*growthcat +
growthcat%ia%rcs(ct_size,4), lrm, f, data=predict.out.09062012)
##Heidi's OR and 95% CI code
lrm_OR <- function(lrmin) {

  temp <- data.frame("coef"=lrmin$coefficients)
  temp$se <- NA
  for ( i in 1:nrow(temp)) {
    temp$se[i] <- sqrt(lrmin$var[i,i])
  }

  temp$OR <- round(exp(temp$coef),3)
  temp$LCI <- round(exp(temp$coef-1.96*temp$se),3)
  temp$UCI <- round(exp(temp$coef+1.96*temp$se),3)
  temp$CI <- paste(temp$LCI,temp$UCI,sep=" to ")
  temp$p <- round(1-pchisq((temp$coef/temp$se)^2,1) ,4)
  colnames(temp) <- c("Coef", "se", "Odds Ratio","LCI", "UCI", "CI","p-value")
  print(temp[,c("Coef", "se", "Odds Ratio","CI","p-value" )])

}
lrm_OR(overfit.lungCA)
overfit.lungCA.test<-overfit.lungCA[[3]]
over.Ctrain<-overfit.lungCA.test [6]## Apparent C index
over.Btrain<-overfit.lungCA.test [11]## Apparent Brier Score
over.Btrain
over.Ctrain

#fit Mayo model in VUMC population for comparison of AUC and brier
mayoCA<-fit.mult.impute (cancer~age + ct_size + smokeyn + spicul +  prev_cancer + upperlobe,
lrm, f, data=predict.out.09062012)
varcov.m<-vcov(mayoCA)
```

```
serror.m<-sqrt(diag(vcov(mayoCA)))

options(digits=3)
postm.anova<-anova(mayoCA)
plot(postm.anova)
print(postm.anova)
exp(cbind(OR=coef(mayoCA), confint.default(mayoCA)))##OR and confidence interval using
reported std errors
print (mayoCA)

mayoCA.test<-mayoCA[[3]]
Cmayotrain<-mayoCA.test [6]## Apparent C index
Bmayotrain<-mayoCA.test [11]## Apparent Brier Score
Bmayotrain
Cmayotrain

#fit SPN model in VUMC population for comparison
SPN.CA<-fit.mult.impute (cancer~age + ct_size + spicul + pack_years + prev_cancer +
upperlobe + growthcat +petpos34 + hemotypsis, lrm, f, data=predict.out.09062012)
varcov.m<-vcov(SPN.CA)
serror.m<-sqrt(diag(vcov(SPN.CA)))

options(digits=3)
postSPN.anova<-anova(SPN.CA)
plot(postSPN.anova)
print(postSPN.anova)
exp(cbind(OR=coef(SPN.CA), confint.default(SPN.CA)))##OR and confidence interval using
reported std errors
print (SPN.CA)


SPN.CA.test<-SPN.CA[[3]]
CSPN.train<-SPN.CA.test [6]## Apparent C index
BSPN.train<-SPN.CA.test [11]## Apparent Brier Score
BSPN.train
CSPN.train

###getting predicted probabilities for ROC analysis and bootstrap optimism

phat.data<-data.frame(predict.out.09062012$study_id)

for (j in 1:nimp) {
compl<-predict.out.09062012
  train.impute<-impute.transcan(f, imputation=j, data=predict.out.09062012, list.out=TRUE,
pr=F, check=F) ##get imputed values using f model above and imputed dataset j
  compl[names(train.impute)]<-train.impute  ###put imputed values and names into new dataset

  imagefit<-predict(overfit.lungCA, compl, type="fitted") ##get prediction coefficients from
lungCA model and estimated predicted cancer using withheld data in test that has become a
complete cases dataset using imputation
 phat.data<-cbind(phat.data,imagefit)
```

```
  }
xhead<-c("study_id","phat1", "phat2","phat3", "phat4","phat5", "phat6","phat7", "phat8","phat9",
"phat10","phat11", "phat12","phat13", "phat14","phat15", "phat16","phat17", "phat18","phat19",
"phat20", "phat21", "phat22","phat23", "phat24","phat25", "phat26","phat27", "phat28","phat29",
"phat30","phat31", "phat32","phat33", "phat34","phat35", "phat36","phat37", "phat38","phat39",
"phat40", "phat41", "phat42","phat43", "phat44","phat45", "phat46","phat47", "phat48","phat49",
"phat50")
names(phat.data)<-xhead
phat.data<-cbind(phat.data, phatbar=rowMeans(phat.data[,-1]))

d<-datadist(compl)
options(datadist="d")
nomogram(overfit.lungCA)
fsum<-summary(overfit.lungCA)
fsum
plot(fsum)
plot (overfit.lungCA, xlim=c(0,1), ylim=c(0,1))

###getting predicted probabilities MAYO model for ROC analysis
phatmayo.data<-data.frame(predict.out.09062012$study_id)

for (j in 1:nimp) {
  compl.mayo<-predict.out.09062012
  trainmayo.impute<-impute.transcan(f, imputation=j, data=predict.out.09062012, list.out=TRUE,
pr=F, check=F) ##get imputed values using f model above and imputed dataset j
  compl.mayo[names(trainmayo.impute)]<-trainmayo.impute  ###put imputed values and names
into new dataset

  phatfit.mayo<-predict(overfit.lungCA, compl.mayo, type="fitted") ##get prediction coefficients
from lungCA model and estimated predicted cancer

  phatmayo.data<-cbind(phatmayo.data,phatfit.mayo)

}

names(phatmayo.data)<-xhead
phatmayo.data<-cbind(phatmayo.data, pmayobar=rowMeans(phatmayo.data[,-1]))
phatmayo.data

pbar.out<-merge (phat.data, phatmayo.data,by="study_id")
write.table(pbar.out, file="F:/Thoracic Surgery Research/Common folder/Lung Cancer
Cohort/Nodule study/predictive model/imputed mayo treatover.ROC.csv", sep=",")

plot (phatbar, pmayohat)

##epidemiology only predictive model
epi.ca<-fit.mult.impute (cancer~age +gender + rcs(pack_years, 4) + prev_cancer  + anysympt +
hemotypsis + bmi, lrm, f, data=predict.out.09062012)
varcov.m<-vcov(epi.ca)
serror.m<-sqrt(diag(vcov(epi.ca)))
```

```
options(digits=3)
postSPN.anova<-anova(epi.ca)
plot(postSPN.anova)
print(postSPN.anova)
exp(cbind(OR=coef(epi.ca), confint.default(epi.ca)))##OR and confidence interval using reported
std errors
print (epi.ca)
##imaging only model
image.ca <- fit.mult.impute(cancer~  rcs(ct_size, 4) + spicul + growthcat + upperlobe +
petpos34, lrm, f, data=predict.out.09062012)
varcov.m<-vcov(image.ca)
serror.m<-sqrt(diag(vcov(image.ca)))

options(digits=3)
postimage.anova<-anova(image.ca)
plot(postimage.anova)
print(postimage.anova)
exp(cbind(OR=coef(image.ca), confint.default(image.ca)))##OR and confidence interval using
reported std errors
print (image.ca)

#END TRAINING MODEL ESTIMATION
nboot <- 500  ### number bootstrap resampling
sum_c<-0
sum_b<-0
n<-nrow(predict.out.09062012)

for (i in 1:nboot){
  x1<-sample(n,n,replace=T) #bootstrap with replacement from the original
  x_train<-compl[x1,]  #training dataset
  x_test<-compl[-x1,]   #test dataset from original dataset and not in training

  ##estimated model with bootstrap training set
  lungCA.train <- fit.mult.impute(cancer~ age +  gender+ bmi + rcs(pack_years,3) + ct_size +
spicul + growthcat + upperlobe +  prev_cancer +  fev1_pred + anysympt + petpos34, lrm, f,
data=x_train, pr=F)

  ########### calculated C-index and Brier from bootstrap test set by each imputed dataset
iteration
  Ctest<-0
  Briertest<-0
     for (j in 1:nimp){


  pfit<-predict(lungCA.train, x_test, type="fitted") ##get prediction coefficients from
lungCA.train model and estimated predicted cancer using withheld data in test that has become a
complete cases dataset using imputation

  ##get C-index from likelihood ratio test and get brier score within this imputation iteration
  lrm_test <- lrm(x_test$cancer ~ pfit)  #c-index using current test
  cindex.test<-lrm_test[[3]]
```

```
  if (lrm_test$coefficient[2]>0) Ctest_temp<-cindex.test[6]else
  Ctest_temp<-1-cindex.test[6]  ##occasionally if predictive relationship is inverted, then c-index
<0.5, adjusting for this possibility
  Btest_temp<-cindex.test[11]
  Ctest<-Ctest+Ctest_temp
  Briertest<-Briertest+Btest_temp

  }
  sum_b<-sum_b+(Briertest/nimp)
  sum_c<-sum_c+(Ctest/nimp)

}
 boot632.c<-(0.368*Ctrain)+(0.632*(sum_c/nboot))
 boot632.brier<-(0.368*Btrain)+(0.632*(sum_b/nboot))
boot632.c
boot632.brier
```

Appendix 2

```
/*****************************************************************/
/* Diagnosis for lung cancer *************************************/
/*****************************************************************/

data nis.biopsy nis.no_biopsy;
      set 'F:\PATH NAME HERE';

array out1 PR1;

 do over out1;

      if out1 in: ('3220','3229') then lung_excision=1;
      if out1 in: ('3230') then segmental_resection=1;
      if out1 in: ('3240','3241','3249') then lobectomy=1;
      if out1 in: ('3260') then disection_lung=1;
      if out1 in: ('3290') then other_lung_excision=1;
      if out1 in: ('3310') then lung_incision=1;
      if out1 in: ('3320') then thora_lung_biopsy=1;
      if out1 in: ('3328') then open_lung_biopsy=1;


end;

      if lung_excision=1 or segmental_resection=1 or lobectomy=1 or
disection_lung=1 or other_lung_excision=1 or
         lung_incision=1 or thora_lung_biopsy=1 then output nis.biopsy;
         else output nis.no_biopsy;
run;

proc freq data=nis.biopsy;
tables lung_excision segmental_resection lobectomy disection_lung
other_lung_excision
         lung_incision thora_lung_biopsy;
run;

data nis.lung;
      set nis.biopsy;

array out3 DX1;

 do over out3;

      if out3 in: ('1150','1151','1159') then histoplasmosis=1;
      if out3 in: ('1160','1161','1162') then blastomycotic_inf=1;
      if out3 in:
('0110','0111','0112','0113','0114','0115','0116','0117','0118','0119')
then tuberculosis=1;
      if out3 in: ('1140','1143','1144','1145','1149') then coccidio=1;
```

```
      if out3 in:
('1170','1171','1172','1173','1174','1175','1176','1177','1178','1179')
then other_mycoses=1;
      if out3 in: ('5130','5131') then lung_abscess=1;
      if out3 in: ('135') then sarcoidosis=1;
      if out3 in: ('51889') then other_lung_dis=1;
      if out3 in: ('2123') then benign_neoplasm=1;
      if out3 in: ('5198','51919') then other_dis_bronchus=1;
      if out3 in: ('1124') then candida=1;
      if out3 in: ('1620','1622','1623','1624','1625','1628','1629')
then malignant_lung=1;
      if out3 in: ('1630','1631','1638','1639') then
malignant_pleura=1;
      if out3 in: ('1970') then sec_malig_resp=1;
      if out3 in: ('2312') then brunchus_insitu=1;
      if out3 in: ('2357') then bronchus_uncertain=1;
      if out3 in: ('2391') then bronchus_unsp=1;
      if out3 in: ('7931') then coin_lesion_nodule=1;
      if out3 in: ('4464') then wegeners=1;
      if out3 in: ('0310','20961','7866') then other_benign=1;
      if out3 in: ('20921') then other_malig=1;

      end;

array out4 DX1;

do over out4;

      if out4 in: ('5120','5121','5128') then pneumothorax=1;
      if out4 in: ('5100','5109') then empyema=1;
      if out4 in:
('1980','1981','1982','1983','1984','1985','1986','1987','1988') then
sec_malig_nos=1;
      if out4 in: ('5163') then ipf=1;
      if out4 in: ('515') then chronic_ipf=1;
      if out4 in: ('1961') then intrathoracic_lymph_node=1;
      if out4 in:
('1971','1972','1973','1974','1975','1976','1977','1978') then
sec_malig_resp_ex=1;
      end;


run;

data nis.lung;
      set nis.lung;

/*********************************/
/*dichotomize age for calculations*/
/*********************************/

if age<18 and age ne "." then agec=0;
if age>=18 and age ne "." then agec=1;
if age=". " then agec=".";

run;
```

Appendix 3

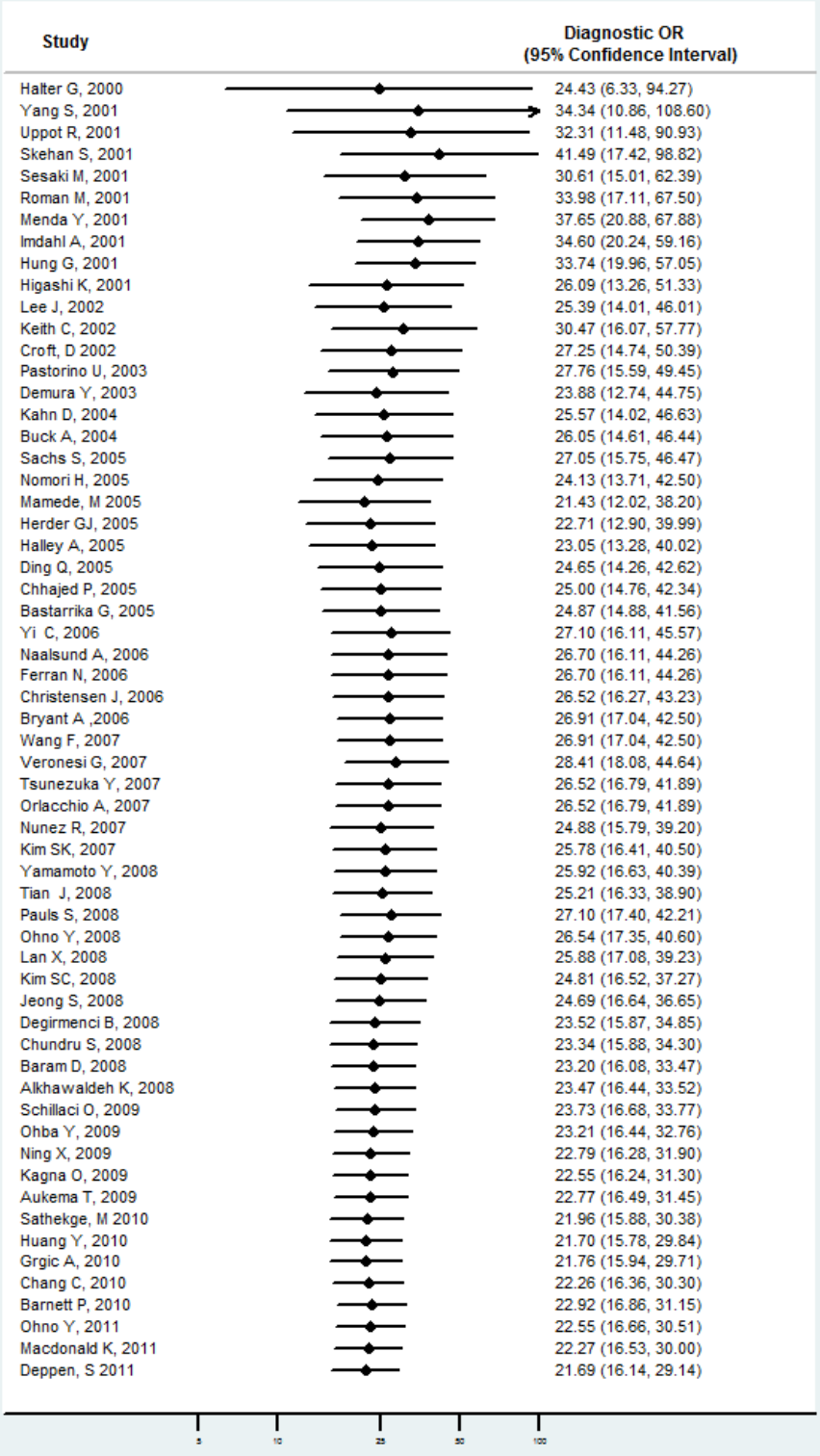
Appendix 3.1 FDG-PET imaging in the diagnosis of lung cancer

Last updated March 30, 2011

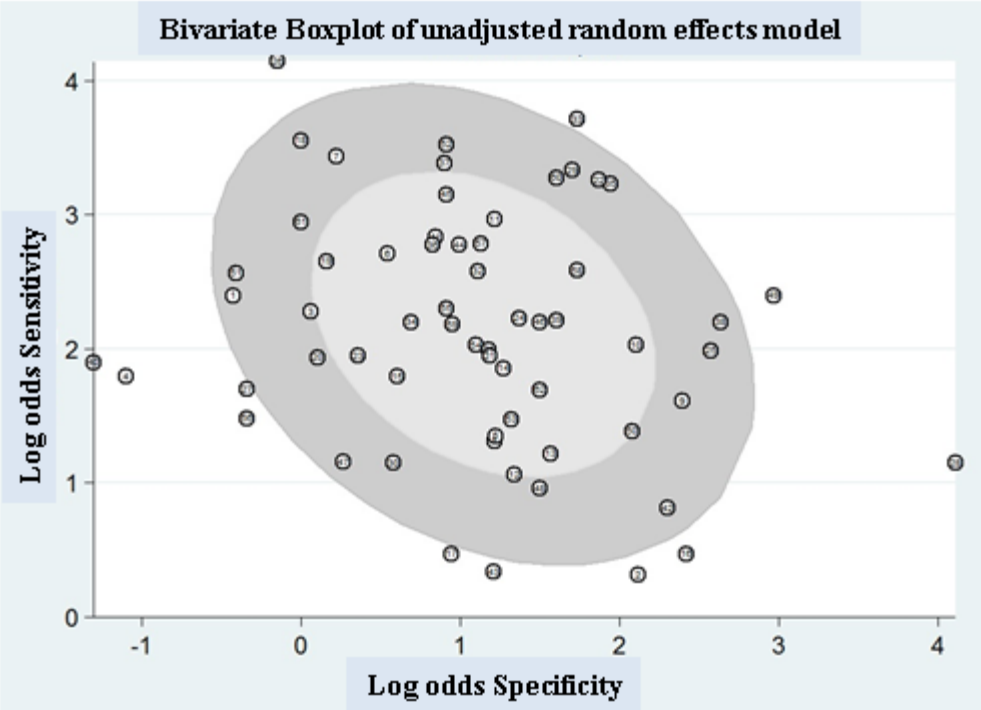
Search terms		Search resultS
#1	lung neoplasms[mh] OR lung cancer[tiab] OR lung nodule[tiab] OR lung nodules[tiab] OR pulmonary nodule[tiab] OR pulmonary nodules[tiab] OR lung lesion[tiab] OR lung lesions[tiab] OR pulmonary lesion[tiab] OR pulmonary lesions[tiab]	166,206
#2	Positron-Emission Tomography[mh] OR fluorodeoxyglucose F18[mh] OR fluorodeoxyglucose F18[nm] OR FDG-PET[tiab] OR FDG-PET/CT[tiab] OR positron emission tomography[tiab]	39401
#3	#1 AND #2 AND eng[la] AND humans[mh] AND 2000:2011[dp]	2244
#4	#3 AND case reports[pt]	542
#5	#3 AND letter[pt]	77
#6	#3 AND review[pt]	370
#7	#3 AND editorial[pt]	32
#9	#3 AND comment[pt]	97
#9	#3 AND practice guideline[pt]	9
#10	#3 AND historical article[pt]	4
#11	#3 AND news[pt]	4
#12	#3 AND meta-analysis[pt]	11
#13	#4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15	1026
#14	#3 NOT #13	1218

Key: [la] language; [mh] medical subject heading; [nm] substance name; [pt] publication type; [dp] publication date; [tiab] keyword in title or abstract Science Citation Index Search, Web of Science Interface
TS=(fdg) AND TS=(pet) AND TS=(lung) AND TS=((screen OR screening OR diagnosis OR diagnose OR diagnostic))

Refined by: Document Type=(PROCEEDINGS PAPER OR MEETING ABSTRACT)



Appendix 3.1. Cumulative meta-analysis of diagnostic odds ratio.



Appendix 3. 2. Bivariate boxplot of log odds sensitivity and log odds specificity for all studies. Oblong shape of boxplot indicates threshold preference for higher sensitivity and an asymmetric SROC curve.