Critical Evaluation of Transbronchial Biopsy Practice in the Diagnosis of Lung Cancer at the Hershey Medical Center

A Dissertation in
Molecular Medicine
by
Anju Sinha

© 2008 Anju Sinha

Submitted in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

December 2008
The dissertation of Anju Sinha is reviewed and approved* by the following:

Rebecca Bascom  
Professor of Medicine  
Dissertation Advisor  
Committee Chair

John Richie  
Professor of Public Health Sciences and Pharmacology

Shantu Amin  
Professor of Pharmacology

Eugene J. Lengerich  
Associate Professor of Health Evaluation Sciences and Health Policy and Administration

Dani S. Zander  
Professor of Pathology

Craig Myers  
Molecular Medicine Graduate Program Chair

*Signatures are on file in the Graduate School.
Critical Evaluation of Transbronchial Biopsy Practice in the Diagnosis of Lung Cancer at the Hershey Medical Center

Abstract

The goal of this project is to determine the pre-test probability and diagnostic yield of transbronchial biopsy (TBBx) in detecting lung cancer and to identify gaps in the current process of care associated with use of TBBx for lung cancer diagnosis at the Hershey Medical Center.

The hypotheses of this project are that the current TBBx practice at the Hershey Medical Center has a low diagnostic yield, high false negative results, especially for small peripheral lesions, provides insufficient clinical data to the interpreting pathologist, and the current process of care for lung cancer patients has variable time gaps between CT scan, TBBx and lung resection.

The specific aims are:

1. To define the current clinical process of care and to identify clinical datasets used for key elements of lung cancer diagnosis.
2. To critically evaluate a retrospective cohort of bronchoscopically acquired TBBxs and estimate the pre-test probability of diagnosing lung cancer.
3. To identify and integrate matched pairs of TBBx and lung resections in order to validate the performance of TBBx against lung resection as the gold standard.
4. To benchmark TBBx pathology reports as a function of digitally acquired computerized tomography (CT) lesions using a novel classification scheme and identify characteristics of lesions commonly biopsied and lesions frequently reported to be positive for malignancy.
5. Assessment of pulmonary lesion characteristics accurately and inaccurately diagnosed to be malignant by TBBx.
6. To estimate the time intervals between CT scan, TBBx and lung resection in order to identify time gaps in the current process of care for lung cancer patients at the Hershey Medical Center.
**Methods:** Specific Aim 1: Process of care and quality measures are assessed by defining the audience of this project, the clinical area being evaluated in this study, the assessment team, the component of the process being measured, preliminary tests, scoring and analytic specifications. To understand and define the current process of care at the Hershey Medical Center for diagnosing, staging and managing patients with lung cancer, various physician resources utilized are conferences with pulmonologists, pathologists, oncologists, radiologists, and cardiothoracic surgeons. Electronic resources utilized to assess patient information are billing records, Power Chart, digital CT scan repository, “Natural Language Ila search”, a software program available in the pathology department, and AJCC guided lung cancer classification protocol adopted by the Department of Pathology.

Specific Aim 2: Patients who underwent TBBx are identified through billing records, their pathology reports are acquired from Power Chart. An epidemiological codebook is developed to document data. Outcome measures evaluated are: Pre-procedure suspicion of lung cancer, pathology report positive for malignancy, smoking history, specimen site biopsied, average aggregate volume of tissue, statistical analysis of average aggregate tissue volume differences between samples positive for malignancy and those negative for malignancy, histopathological type of malignancy, gender distribution by histopathological type of malignancy, estimate of the pre-test probability of a positive pathology report for malignancy among samples obtained from patients clinically suspected to have lung cancer, cellular atypia, non-malignant findings, and staining.

Specific Aim 3: “Natural Language Ila search”, software is used to identify and acquire pathology reports of lung resection surgeries. Patient name and MRN are used to identify patients common in the TBBx and lung resection datasets. An epidemiological codebook is developed to document lung resection data. TBBx reports are categorized as test positive and negative based on presence or absence of malignancy. Lung resection reports are used as gold standard to assess TBBx performance in diagnosing lung cancer. Data are
integrated to identify four groups of cases, true positive, true negative, false positive, and false negative.

Specific Aim 4: The digital CT scans IDX/GE electronic image repository is used to retrieve digital images of pulmonary lesions biopsied for patients included in the TBBx dataset. A single transverse slice of each CT scan image centered on the region of interest (ROI) is extracted using Snagit®. A novel image classification scheme is used to classify ROIs into focal and diffuse lesions. Each type of lesion if further divided into subcategories based on size. Outcome measures evaluated are: availability of digital CT scans with time, most commonly biopsied lesion characteristic, and characteristic of lesion most commonly reported to be malignant by TBBx.

Specific Aim 5: TBBx pathology report, lung resection pathology report and digital CT scan image data are integrated in order to identify lesion characteristics in the true and false positive and negative categories. The outcome measures established are: location of lesion biopsied and resected, congruency of site biopsied and site resected, histopathological type of malignancy reported by TBBx and lung resection, and digital CT scan lesion characteristics.

Specific Aim 6: Availability of lung resection and digital CT scan data with time are established. In order to assess time to diagnosis (TDx) and time to treatment (TRx) only pre-TBBx CT scans and post TBBx resections are included. TDx and TRx are established for the four groups of cases to identify gaps in process of care.

Results: Specific Aim 1: The audience of this project includes patients, physicians and researchers involved and interested in lung cancer. The clinical area being evaluated is TBBx and its efficiency in diagnosing lung cancer. The assessment team is primarily composed of the author and her thesis advisor. The component of the TBBx process being measured is primarily the pre-procedure differential diagnosis of cancer. Datasets utilized in this study include billing department records, Power Chart (Cerner), digital CT scan image
repository, “Natural Language Ila search” software available in the Department of Pathology.

Specific Aim 2: Pathology reports from 801 TBBx samples obtained from 653 procedures done over the 5 years 10 months period of this study are obtained and data organized and analyzed. Patients' age averaged 58yrs 2mths +/- 15yrs 2mths (mean +/- SD) and 360/653 (55.13%) are male. Lung Cancer was suspected pre-procedure in 122/653 (18.7%) cases. Pathology report is positive for malignancy in 111/801 (13.86%) TBBx samples. Pre-test probability of diagnosing lung cancer is highest (22.45%) among patients suspected to have lung cancer, intermediate (13.89%) in the group where differential diagnosis is not specified and minimum (0.7%) in the group where lung cancer is not suspected pre-procedure. Average aggregate volume of samples positive for malignancy is significantly higher (0.062cm$^3$) than the average aggregate volume of samples negative for malignancy (0.047cm$^3$). Adenocarcinoma is the most common histopathological type of cancer seen in our study. A lack of transfer of clinical information to the diagnosing pathologist is noticed.

Specific Aim 3: Fourteen of 72 cases common to TBBx and lung resection datasets are true positive, 28 are false negative, 29 are true negative and 1 is false positive. Sensitivity of TBBx in diagnosing lung cancer is estimated to be 0.33, specificity is 0.97, false negative rate (FNR) =0.67, false positive rate (FPR) =0.03, predictive value positive (PVP) =0.93, predictive value negative (PVN) =0.51.

Specific Aim 4: Digital CT scan data are available for 298/653 TBBx procedures. Focal ROIs are more commonly biopsied (180/298, 54.05%) compared to diffuse lesions (153/298, 45.95%). Focal ROIs 1-3cm in size are the most commonly biopsied lesions. Among diffuse lesions, ROIs involving <25% of the lobe are most commonly biopsied. However, malignancy is most commonly reported in focal ROIs >3cm in size. Among diffuse lesions, malignancy is most commonly reported in ROIs involving >50% of the lobe.

Specific Aim 5: Of the 29 true negative cases diffuse lesions involving <25% of the right lower lobe is most commonly seen. Among the 14 true positive
cases adenocarcinomas, focal lesions 1-3 cm in size, located in the right upper lobe are most commonly seen. One false positive sample is reported to be carcinoma in situ, squamous cell carcinoma in the left upper lobe. CT scan for this case is not available. Of the 28 false negative cases, adenocarcinomas, focal lesions 1-3cm in size and lesions in the right upper lobe are most commonly seen.

Specific Aim 6: Digital CT scan images are available for 298/653 (45.64%) and lung resection data are available for 74/653 (11.33%) of TBBx procedures. The median time to diagnosis (TDx) is 7 days pre-TBBx (range 0-57 days) and the median time to treatment (TRx) is 63.5 days post-TBBx (range 4-1493). No significant difference is seen in TDx and TRx between the true and false positive and negative groups. However the number of cases in these groups is small (4-14) therefore these results need to be confirmed in a study including larger number of patients.

Conclusions: Disparate electronic datasets can be linked to critically evaluate TBBx yield in diagnosing lung cancer. Essential clinical information is not provided to the diagnosing pathologist frequently. Pre-test probability of diagnosing lung cancer is highest among patients suspected to have lung cancer. Average aggregate volume is significantly higher in samples reported positive for malignancy compared to non-malignant samples. This is an important factor when assessing feasibility of using TBBx tissue for molecular markers. Validation of TBBx performance in diagnosing lung cancer against lung resection as gold standard provided a sensitivity= 0.33, and specificity= 0.97 with a high false negative rate of 0.67. Malignant focal lesions >3cm in size are more frequently detected accurately than adenocarcinomas, focal lesions 1-3cm in size that have a high false negative rate. The median TDx is 7 days pre-TBBx and the median TRx is 63.5 days post-TBBx. Variability is seen in TDx and TRx but no statistically significant difference is seen in TDx and TRx between the true and false positive and negative groups.
# Table of Contents

List of Figures ........................................................................................................... ix  
List of Tables ............................................................................................................ xii  
Acknowledgements .................................................................................................. xiii  

I  Goal, Hypotheses and Specific Aims ................................................................. 1  
II  Background and Significance:  
    Purpose ............................................................................................................. 4  
    Burden of Disease .............................................................................................. 4  
    Recent Trends in the Epidemiology of Lung Cancer ............................................ 6  
    Molecular Biology of Lung Cancer ................................................................. 13  
    Recent Advances in Technology and Treatment ............................................ 22  
    Gap Analysis .................................................................................................... 31  
    Rationale .......................................................................................................... 33  
    Goal, Hypotheses and Specific Aims ................................................................. 35  

III  Materials And Methods:  
    Specific Aim 1 .................................................................................................. 37  
    Specific Aim 2 .................................................................................................. 43  
    Specific Aim 3 .................................................................................................. 55  
    Specific Aim 4 .................................................................................................. 63  
    Specific Aim 5 .................................................................................................. 67  
    Specific Aim 6 .................................................................................................. 69  

IV  Results:  
    Specific Aim 1 .................................................................................................. 74  
    Specific Aim 2 .................................................................................................. 82  
    Specific Aim 3 .................................................................................................. 97  
    Specific Aim 4 .................................................................................................. 101  
    Specific Aim 5 .................................................................................................. 104  
    Specific Aim 6 .................................................................................................. 114  

V  Discussion .......................................................................................................... 127  

VI  Conclusions ...................................................................................................... 146  

VII  Applications, Implications and Future Directions: ....................................... 149  
    Appendix A: An example of TBBx pathology report ...................................... 151  
    Appendix B: Assessing the Potential of Sphingosine Kinase-1 as a Diagnostic and Prognostic Marker for Lung Cancer: A Pilot Study ........................................................................................................ 153  

VIII  Reference: ...................................................................................................... 169
List of Figures:

Figure 1: Incidence of commonly occurring cancers based on gender, demonstrating that lung cancer is currently the second most common cancer in both genders. Figure from (Jemal, Siegel et al. 2007).*

Figure 2: Figure from (Jemal, Siegel et al. 2007), demonstrating that lung cancer is currently the leading cause of death by cancer in both genders.*

Figure 3: Figure from (Wahbah M, Boroumand N et al. 2007). Change in incidence of histological type of lung cancer. As the incidence of squamous cell cancer decreased in the 1980’s, the incidence of adenocarcinoma increased.*

Figure 4: Algorithm showing data organization and variable definition.

Figure 5: Method used to assess the aggregate volume of TBBx tissue.

Figure 6: Algorithm showing the integration of TBBx and lung resection datasets. +/- denotes the group of patients with TBBx and resection reports positive for malignancy, +/ denotes the group of patients that are positive for malignancy on TBBx but negative on lung resection, -/+ denotes the group of patients negative for malignancy on TBBx but positive on lung resection, and +/- denotes the group of patients that are negative for malignancy on TBBx and lung resection.

Figure 7: Classification of CT scan digital images. A single slice of each CT scan centered at the region of interest (ROI) served as a representative sample. Lesions were classified as focal and diffuse, each category was further classified into 3 subcategories depending on size of ROI.

Figure 8: Algorithm for correlation between CT scan, TBBx and lung resection datasets.

Figure 9: Algorithm showing elements of process of care, in the diagnosis and management of lung cancer. Variety of tissue may be available from the same patient at different stages of the process. BAL= Bronchoalveolar Lavage, TBNA= transbronchial needle aspiration, TBBx= transbronchial biopsy.

Figure 10: Algorithm showing the various instruments used and tissues that may be obtained during bronchoscopy. The thick line highlights transbronchial biopsy tissue that has been utilized in this study.

Figure 11: Number of procedures done and samples acquired per year by bronchoscopic TBBx between August 1999 and June 2005 at the Hershey Medical Center. The first and last bars in figure show data for 6 months of clinical activity, while the middle 5 bars each represent one year data.

Figure 12: Average monthly trend in the TBBx practice at HMC. Average number of procedures done and samples acquired per month for each year by bronchoscopic TBBx between August 1999 and June 2005 at the Hershey Medical Center.

Figure 13: showing correlation between pre-test differential diagnosis and final pathology report with reference to presence or absence of malignancy. DDx= Differential diagnosis, pts= patients.

Figure 14: Pie chart for smoking history obtained from TBBx pathology reports. No smoking history is provided in pathology reports about 74% of times. Even in the 12% of procedures where there is mention of history of smoking present, no further details are provided. Passive smoke exposure history is rarely provided.

Figure 15: Distribution of parenchymal specimen site biopsied. RUL= right upper lobe, RML= right middle lobe, RLL= right lower lobe, LUL= left upper lobe, LL= left lingula, LLL= left lower lobe, post= posterior segment, ant= anterior, med= medial, lat= lateral, sup= superior, inf= inferior, apicopost= apico-posterior.

Figure 16: Distribution of tracheobronchial tree site biopsied. Rt= right, Lt= left, div= division, sec= secondary, all other abbreviations same as figure 15.
Figure 17: The aggregate volume of TBBx tissue obtained from 771/801 TBBxs (96%) performed between August 31st 1999 and June 30th 2005. Excluded are 30 samples for which volume is not provided. The average aggregate volume of 103 samples reported positive for malignancy is significantly more than the average aggregate volume of samples reported negative for malignancy.

Figure 18: Pie chart showing histopathological distribution of type of malignancies reported. One hundred and five samples are reported positive for malignancy. However, the total is 112 since samples reported to have mixed cellularity are counted more than once.

Figure 19: Distribution of various histopathological types of malignancies based on differentiation.

Figure 20: Gender distribution of histopathological types of malignancies diagnosed by TBBx performed between August 31st 1999 and June 30th 2005. Included are all conclusive/definite pathological diagnoses; excluded are specimens diagnosed as suspicious. Samples reported to be of mixed cellularity are assigned to each cell type. E.g.: adeno-squamous carcinoma is counted once as adeno carcinoma and once as squamous carcinoma.

Figure 21: The pre-test probability of pathology report positive for malignancy when it is suspected pre-procedure. It is highest for patients suspected to have malignancy and lowest in patients not suspected to have malignancy pre-bronchoscopy.

Figure 22: Algorithm demonstrating the cross link between TBBx and Lung Resection datasets. The two TBBx samples reported insufficient for evaluation could not be categorized and are reported separately.

Figure 23: Correlation between histopathological types of malignancy reported on TBBx and lung resections.

Figure 24: Algorithm showing CT scan data acquisition and distribution. ROI= region of interest.

Figure 25: Histogram demonstrating correlation between CT scan ROI classification and TBBx pathology report with regard to malignancy.

Figure 26: Algorithm for correlation between digital CT scan ROI, TBBx and lung resection datasets.

Figure 27: Lesion characteristics as seen in digital CT scan for 29 true negative TBBx and lung resection matched pairs.

Figure 28: Lesion location for 29 true negative TBBx and lung resection matched pairs.

Figure 29: Congruency of specimen site for 29 true negative TBBx and lung resection matched pairs.

Figure 30: CT scan lesion characteristics for 14 true positive TBBx and lung resection matched pairs.

Figure 31: Lesion histopathology data for 14 true positive TBBx and lung resection matched pairs.

Figure 32: Lesion location for 14 true positive TBBx and lung resection matched pairs.

Figure 33: Overall congruency of sampling location for the 14 true positive TBBx and lung resection matched pairs.

Figure 34: Lesion location and congruency of specimen site for 14 true positive TBBx and lung resection matched pairs.

Figure 35: CT scan lesion characteristics for 28 false negative TBBx and lung resection matched pairs.

Figure 36: Lesion histopathology data for 28 false negative TBBx and lung resection matched pairs. Three of these are reported as indeterminate on TBBx, categorized as negative. Total is 29 since one specimen is mixed type carcinoma.
Figure 37: Lesion location for 28 false negative TBBx and lung resection matched pairs. Total number of TBBx is 30 because 2 patients had 2 sites biopsied.

Figure 38: Overall congruency of sampling location for the 28 false negative TBBx and lung resection matched pairs.

Figure 39: Lesion location and congruency of specimen site for 28 false negative TBBx and lung resection matched pairs.

Figure 40: Time trend demonstrating availability of the 74 matched lung resections.

Figure 41: Time trend demonstrating availability of 298 digital CT scan images for matched TBBx procedures.

Figure 42: Availability of 74 matched lung resections with time.

Figure 43: Availability of the 298 digital CT scans with time.

Figure 44: Time distribution of 298 digital CT scan data with respect to date of TBBx.

Figure 45: Time interval between TBBx and Lung Resection in the 74 matched pairs

Figure 46: Time intervals between CT scan date, TBBx date and Lung Resection data among the 31 patients common in three datasets. An estimate of process of care in diagnosis of lung cancer.

Figure 47: Graphical representation of data for time to diagnosis and time to treatment for the groups of patients common to all three datasets.

Figure 48: Time to diagnosis and time to treatment for four true positive TBBx samples.

Figure 49: Time to diagnosis and time to treatment for fourteen true negative TBBx samples.

Figure 50: Time to diagnosis and time to treatment for thirteen false negative TBBx samples.

Figure 51: Median and SEM of time to diagnosis for the three groups of patients showing that the difference between the groups is not statistically significant (p>0.05).

Figure 52: Median and SEM of time to treatment for the three groups of patients showing that the difference between the groups is not statistically significant (p>0.05).

Figure 53: A schematic of population characteristics in the diagnosis of lung cancer, indicating that TBBx may be a specific tool in the diagnosis process but is not sensitive. Representation is not to scale.

Figure 54: A Schematic showing the possible sources of error in acquiring tissue during bronchoscopic transbronchial biopsies. T= tumor, E1= Tissue acquired from wrong lobe, E2= Tissue acquired from wrong segment, E3= Tissue acquired form the edge of the tumor, may not be representative sample, E4= Tumor heterogeneity may result in incorrect interpretation of cell type. Other errors that not shown in schematic but listed in figure E5= Sample obtained may be too small therefore insufficient for evaluation, E6= tissue obtained may be damaged due to manipulations during procedure, E7= TBBx is done to acquire parenchymal tissue, however, sometimes mucosal tissue is obtained and the parenchymal lesion is not sampled. RUL= right upper lobe, RML= right middle lobe, RLL= right lower lobe, LUL= left upper lobe, LL= left lingula, LLL= left lower lobe.

* Figures are excerpted from cited references believing that they are “fair use” based on the purpose, nature, amount, and market effect of the use.
List of Tables:

Table 1: Tabular representation of frequency of diagnosing lung cancer at various stages and 5-year survival rates, demonstrating that most patients are diagnosed at late stage of disease and survival is poor. Data adopted from Fry WA, et al. Cancer. 1999;86:1867-1876 (Fry, Phillips et al. 1999).

Table 2: Represents the percent of patients diagnosed with various histological types of lung cancer and their corresponding 5-year and 10-year survival rates. Data for percent diagnosed are adopted from Ginsberg 1997(Ginsberg 1997). Data for 5-year and 10-year survival are adopted from Fry, et al. 1999 (Fry, Phillips et al. 1999). * Numbers do not sum to 100% because of differences in diagnostic criteria. NA=not available.

Table 3: Data adopted from (Brooks, Austin et al. 2005). Indirect evidence of shift in location of lung cancer, adenocarcinoma is the most common histological type of lung cancer and it occurs most frequently in the periphery, suggesting that peripheral lesions are the most commonly encountered lesions in lung cancer.

Table 4: Epidemiological codebook developed to document data obtained from pathology reports for all TBBxs performed during the period of the study.

Table 5: The codebook developed for documentation and analyzes of data obtained from lung resection pathology reports.

Table 6: Table to show algorithm used to assess sensitivity and specificity of diagnosing lung cancer by TBBx procedure.

Table 7: An overview of the pre-procedure differential diagnosis mentioned in the TBBx pathology reports.

Table 8a (top) shows categorization of smoking history data acquired from TBBx reports. Table 8b (bottom) shows detail information acquired for the 16 reports that are categorized as others in table 8a.

Table 9: Showing that the probability of volume measurements being unavailable is not statistically significantly different for samples reported positive for cancer versus samples reported negative for cancer.

Table 10: Overview of types of cellular atypia reported in pathology reports of all 801 samples.

Table 11: Overview of the non-malignant findings noted in pathology reports of all 801 samples.

Table 12: Overview of staining data acquired from pathology reports of all 801 samples.

Table 13: Correlation between TBBx and lung resection pathology reports with regard to presence or absence of malignancy. Two patients' data is excluded, suboptimal for evaluation.

Table 14: Digitally acquired CT scan pulmonary lesion characteristics of the 31 matched pairs categorized as a function of TBBx and lung resection matched pathology report with regard to presence or absence of malignancy. One patient in false positive category is reported separately. Digital CT scan data for this patient is not available.

Table 15: Availability of digital CT scan and lung resection data for all 653 TBBx done; categorized by date of service for TBBx.

Table 16: Evaluation of process of care for lung cancer diagnosis. Time interval between CT scan and TBBx, and time interval between TBBx and lung resection.
Acknowledgments

The author acknowledges the valuable contributions made by the members of the Bascom Laboratory for the success of this study. Special thanks to Dr. Henry Crist, and Dr. Saeid Safaee for their critical contribution during development and designing stages of this study. Also to Dr. Muhammad Khan and Dr. Syed Gilani who helped with some of the data collection and analysis. This study is supported by Endographics Imaging Systems, Inc. NIH NCI grants CA074325 and CA091534.
DEDICATION

I dedicate my dissertation to my son Sushruth for being my inspiration and strength through the toughest period of my life, professionally and personally, and to Kevin for having faith in me and making my work possible.
Critical Evaluation of Transbronchial Biopsy Practice in the Diagnosis of Lung Cancer at the Hershey Medical Center

I. Goal, Hypotheses and Specific Aims

The goal of this project is to determine the pre-test probability and diagnostic yield of transbronchial biopsy (TBBx) in detecting lung cancer and to identify gaps in the current process of care associated with use of TBBx for lung cancer diagnosis at the Hershey Medical Center. This project focuses on TBBx because it is part of the routine process of care and provides non-invasive histological samples acquired at the early stage of diagnosis and staging of lung cancer.

This project is developed with the long-term vision to incorporate molecular markers into routine process of care for diagnosing, staging and developing individualized treatment for patients with lung cancer. For maximum clinical impact, it is essential to incorporate appropriate markers at an early stage and in a procedure with a high accuracy of diagnosing lung cancer. To identify appropriate markers, an in-depth understanding of the molecular biology of lung cancer is essential. This dissertation provides a relatively up-to-date and comprehensive background on the current knowledge on molecular markers in lung cancer.

Earliest histological confirmation of lung cancer is often made during bronchoscopy when tissue is acquired for histological and or cytological evaluation. Transbronchial biopsy is one of the procedures during bronchoscopy when pulmonary parenchymal tissue is obtained for histological evaluation. TBBx is the main focus of this project since it is a critical step in the process of diagnosis and staging of lung cancer. Efficiency and accuracy at this stage is crucial in determining course of clinical management and survival of patients with
lung cancer. However many facets of TBBx practice are poorly understood and of concern. Therefore this project focuses on TBBx and aims to determine the pre-test probability and diagnostic yield of TBBx in diagnosing lung cancer. This project also aims to identify gaps in the current process of care associated with use of TBBx for diagnosing lung cancer that must be remedied to prepare for translation of basic science knowledge into clinical practice.

To critically evaluate yield of TBBx, parameters used in this project include pre-test probability of diagnosing lung cancer, frequency of false negative results by comparing TBBx results with a gold standard, surgical lung resection, computerized tomography (CT) scan acquired characteristics of lesions accurately and inaccurately identified, and finally the time intervals between CT scan (suspicion of lung cancer), TBBx (confirmation of diagnosis) and lung resection (surgical management).

The hypotheses of this project are that the current TBBx practice at the Hershey Medical Center has a low diagnostic yield, high false negative results, especially for small peripheral lesions, provides insufficient clinical data to the interpreting pathologist, and the current process of care for lung cancer patients has variable time gaps between CT scan, TBBx and lung resection.

The specific aims are:
1. To define the current clinical process of care and to identify clinical datasets used for key elements of lung cancer diagnosis.
2. To critically evaluate a retrospective cohort of bronchoscopically acquired TBBxs and estimate the pre-test probability of diagnosing lung cancer.
3. To identify and integrate matched pairs of TBBx and lung resections in order to validate the performance of TBBx against lung resection as the gold standard.
4. To benchmark TBBx pathology reports as a function of digitally acquired computerized tomography (CT) lesions using a novel classification scheme and
identify characteristics of lesions commonly biopsied and lesions frequently reported to be positive for malignancy.

5. Assessment of pulmonary lesion characteristics accurately and inaccurately diagnosed to be malignant by TBBx.

6. To estimate the time intervals between CT scan, TBBx and lung resection in order to identify time gaps in the current process of care for lung cancer patients at the Hershey Medical Center.
II: Background and Significance

Purpose:

Dr. Elias Zerhouni, director of National Institutes of Health articulates the gap in translation effort where vast amount of basic science research conducted over the last few decades has had little impact on clinical care of patients. For example, tremendous research and efforts have been invested in the development of preventive, screening, radiologic, genetic, molecular biomarkers, diagnostic, surgical and therapeutic regimens for lung cancer. Yet the 5-year survival of lung cancer patients has remained a dismal 13-16% for over 4 decades (ACS 2007). Similar efforts in other cancers like prostate, colon and breast cancer have resulted in improvement of survivorship. The current project attempts to address some aspects of the challenges with lung cancer care and address this need-of-the-hour by defining the process of care for lung cancer patients, critically evaluating the transbronchial biopsy practice in diagnosing lung cancer and identify gaps in the process of care.

Burden of Disease:

Lung cancer continues to be the leading cause of death by cancer in the United States of America. Lung cancer is estimated to be the cause for 29% of all cancer deaths, with about 213,380 estimated new cases and 160,390 estimated deaths in 2007 (Jemal, Siegel et al. 2007). Recent endeavors in community education, screening, early diagnosis, and treatment have improved the long term survival of patients with prostate, colon and breast cancer. However, lung cancer continues to have poor prognosis in spite of similar efforts (Fry, Phillips et al. 1999). Current 5-year survival rate of patients with lung cancer is only about 16% (Jemal, Siegel et al. 2007). One of the challenges in lung cancer care is the rarity of diagnosing lung cancer at an earlier stage. As shown in table 1, about
28% of patients are diagnosed in stages I and II which have a 5-year survival of 42% and 23% respectively. Five-year survival decreases to 11% in stage IIIa and 5% in Stage IIIb. About 32% of patients are diagnosed in stage III. Forty percent patients are diagnosed in stage IV when 5-year survival is only 1% (Fry, Phillips et al. 1999).

Table 1: Tabular representation of frequency of diagnosing lung cancer at various stages and 5-year survival rates, demonstrating that most patients are diagnosed at late stage of disease and survival is poor. Data adopted from Fry WA, et al. Cancer. 1999;86:1867-1876 (Fry, Phillips et al. 1999).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 0</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>Stage I</td>
<td>21%</td>
<td>42%</td>
</tr>
<tr>
<td>Stage II</td>
<td>7%</td>
<td>23%</td>
</tr>
<tr>
<td>Stage III</td>
<td>32%</td>
<td>11% (Stage IIIA)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5% (Stage IIIB)</td>
</tr>
<tr>
<td>Stage IV</td>
<td>40%</td>
<td>1%</td>
</tr>
</tbody>
</table>

For early stages of lung cancer including stages I, II and IIIa, surgical resection and treatment can result in longer survival of the patients (Fry, Menck et al. 1996). For later stage disease (III b and IV) chemotherapy is the treatment of choice but has a poor initial response rate of only 30%. Furthermore, the initial responses last for only 4-6 months, after which patients typically develop resistance (Rosell, Cecere et al. 2006). Thus, there is a pressing need to improve current screening and diagnostic techniques, identify therapeutic targets and develop more efficient therapeutic regimens in order to enhance response, prognosis and survival of patients with lung cancer. This projects aims to assess the diagnostic yield of TBBx in detecting lung cancer and identify gaps in the process of care.
Lung cancer diagnosis and treatment require a multi-disciplinary approach with repeated hand-offs of patients during the process of diagnosis, staging, treatment and follow-up. There are well established, international staging criteria that guide clinical decision-making (Sihoeh and Yim 2004; Detterbeck, Jantz et al. 2007; Silvestri, Gould et al. 2007). However, adherence to published guidelines is modest (Olsson, Schultz et al. 2008; Schultz, Powell et al. 2008; Schultz, Silvestri et al. 2008). Recent studies also show wide variation in time to treatment among VA facilities. Formal review of patient flow was associated with increased timeliness. Teaching hospital, curative (vs. palliative) radiotherapy, the presence of comorbidities, and initial referral to a non-respiratory physician were identified as factors associated with less timely care. Nurse-led care coordination, multidisciplinary meetings via teleconference, and an outpatient two-step diagnostic process were identified as interventions that improved timeliness of care. Timely care was associated with better survival.

**Recent Trends in the Epidemiology of Lung Cancer:**

In order to develop long-term effective strategies for improving screening and management of the lung cancer patients it is important to stay abreast with the recent changes in trend of the epidemiology of lung cancer. **Lung cancer incidence** increased dramatically in the mid-twentieth century and is currently the second most common cancer occurring in men and women in the United States of America (Jemal, Siegel et al. 2007) (figure 1). There appears to be a relative decrease in the incidence of lung cancer from 17.8% in 1980 to 13.5% in 2003, although the total number of cases has increased 1.5 folds (Wahbah M et al. 2007). Factors contributing to this apparent decrease in lung cancer incidence are not well defined. Pennsylvania and the economically disadvantaged parts of the Northeastern United States, the Appalachian region, has a significantly high incidence rate of lung cancer along with some other cancers compared to the national average (Lengerich, Tucker et al. 2005).
Figure 1: Incidence of commonly occurring cancers based on gender, demonstrating that lung cancer is currently the second most common cancer in both genders. Figure from (Jemal, Siegel et al. 2007)*.

Lung cancer mortality saw an almost exponential increase through latter half of the twentieth century from its onset in the mid-twentieth century (ACS 2007). As seen in figure 2, lung cancer is currently the leading cause of death by cancer among men and women in the United States of America.
There seems to be no change in survival of lung cancer patients according to a ten-year survey done by Fry et al. (Fry, Phillips et al. 1999). However, according to the American Cancer Society, lung cancer mortality has declined significantly in men by about 1.9% per year between 1991 and 2003. This decline has been attributed mainly to the decrease in tobacco consumption in men (Hermens, Van Engelenburg et al. 2003; ACS 2007; Kabir, Connolly et al. 2007). Improvements in screening, diagnosis and treatment have also contributed to a slight but significant increase in survival (Breathnach, Freidlin et al. 2001; Henschke, Yankelevitz et al. 2006). In Pennsylvania, mortality due to lung cancer has decreased by 1 per 100,000 per year over a 5 year period (2000-2004) (http://cancercontrolplanet.cancer.gov/ 2008). Although this decrease is statistically significant, it is minimal and of insignificant solace to the large number of people suffering from lung cancer. On the other hand, all-cause mortality in some counties in the United States has increased in recent years (Ezzati, Friedman et al. 2008). Lung Cancer is one of the few factors contributing to this increased death rate in spite of significant declines in cardiovascular mortality.

Last few decades have also seen a changing trend in the gender distribution of lung cancer. As seen in figures 1 and 2, since its onset in the mid-
twentieth century, lung cancer epidemic was predominantly a male disease, but the later part of the twentieth century saw an increase in the incidence of lung cancer among women. Lung cancer mortality also increased among women with a 20 year lag period. This increase is attributed to the increase in tobacco consumption among women which started in the 1960s. Lung cancer is now the leading cause of death by cancer among both men and women (Janssen-Heijnen and Coebergh 2001; ACS 2007). Besides smoking, various other factors have been identified that influence tumor development and proliferation in women that are different from the factors influencing tumor biology in men (Belani, Marts et al. 2007). There is some evidence suggesting that women who smoke are more susceptible to developing lung cancer, especially adenocarcinoma, and that life-long non-smoking women with lung cancer have a poorer prognosis. However, Thun et al. challenge this perception based on their temporal study of lung cancer among life-long never smokers, in which women with lung cancer did not have a higher death rate than men (Thun, Henley et al. 2006).

The trend in the histopathology of lung cancer has also seen a change in the last few decades. Squamous cell carcinoma was the most commonly occurring histological type of lung cancer in the 1980’s. However, its incidence decreased in the 1990’s and the incidence of adenocarcinoma of the lung increased during this period. As seen in figure 3, adenocarcinoma is currently the most common type of lung cancer (Janssen-Heijnen and Coebergh 2001; Wahbah M, Boroumand N et al. 2007) (Brooks, Austin et al. 2005). This change in trend has been attributed to the use of low-tar cigarettes, addition of filters, and changes in smoking behavior (Wahbah M, Boroumand N et al. 2007). Adenocarcinoma is the most frequent form of lung cancer seen among women, which also contributes to this change in trend (Belani, Marts et al. 2007). Whether gender is an independent factor from low-tar cigarette and filter use remains to be determined (Wahbah M, Boroumand N et al. 2007).
Figure 3: Figure from (Wahbah M, Boroumand N et al. 2007). Change in incidence of histological type of lung cancer. As the incidence of squamous cell cancer decreased in the 1980’s, the incidence of adenocarcinoma increased*.

Table 2 demonstrates the percent of patients diagnosed with various histological types of lung cancer and their corresponding 5-year and 10-year survival rates. As all stages of each histological type of cancer are grouped together in this table the survival rates are not significantly different, suggesting that histological type of lung cancer does not influence survival whereas stage at diagnosis does as seen in table 1.
Table 2: Represents the percent of patients diagnosed with various histological types of lung cancer and their corresponding 5-year and 10-year survival rates. Data for percent diagnosed are adopted from Ginsberg 1997 (Ginsberg 1997). Data for 5-year and 10-year survival are adopted from Fry, et al. 1999 (Fry, Phillips et al. 1999). ** Numbers do not sum to 100% because of differences in diagnostic criteria. NA=not available

<table>
<thead>
<tr>
<th>Histological Type</th>
<th>Percent Diagnosed**</th>
<th>5-Year Survival</th>
<th>10-Year Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Cell Carcinoma</td>
<td>18%</td>
<td>5%</td>
<td>2%</td>
</tr>
<tr>
<td>Squamous Cell Carcinoma</td>
<td>30%</td>
<td>14%</td>
<td>8%</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>40%</td>
<td>17%</td>
<td>10%</td>
</tr>
<tr>
<td>Large Cell Carcinoma</td>
<td>15%</td>
<td>10%</td>
<td>6%</td>
</tr>
<tr>
<td>Sarcoma</td>
<td>NA</td>
<td>23%</td>
<td>18%</td>
</tr>
<tr>
<td>All Other/Unknown</td>
<td>NA</td>
<td>12%</td>
<td>9%</td>
</tr>
</tbody>
</table>

Along with the shift in the histology of lung cancer, there is evidence of shift in the location of lung cancer. Squamous cell carcinoma was the most common type of cancer and along with small cell carcinoma, they occurred most commonly in the central tracheo-bronchial tree, but over the last few decades the incidence of adenocarcinoma has increased and it commonly presents in the peripheral lung (Brooks, Austin et al. 2005). This change in location of lung cancer is also believed to be a consequence of low-tar cigarettes and use of filters which may have changed particle size deposited in the airway, therefore the relative dose of various carcinogens delivered. The smoking behavior too is believed to have changed leading to more cigarettes being smoked, longer puffs, and deeper inhalation with deposition of carcinogens further into the periphery of lungs. As the incidence of adenocarcinoma is increasing the incidence of peripheral tumors is also increasing.
Table 3: Data adopted from (Brooks, Austin et al. 2005). Indirect evidence of shift in location of lung cancer, adenocarcinoma is the most common histological type of lung cancer and it occurs most frequently in the periphery, suggesting that peripheral lesions are the most commonly encountered lesions in lung cancer.

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Tumor Location</th>
<th>Peripheral (%)</th>
<th>Central (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenocarcinoma</td>
<td></td>
<td>80</td>
<td>20</td>
</tr>
<tr>
<td>Squamous cell ca</td>
<td></td>
<td>55</td>
<td>45</td>
</tr>
<tr>
<td>Small cell ca</td>
<td></td>
<td>23</td>
<td>77</td>
</tr>
<tr>
<td>Large cell ca</td>
<td></td>
<td>78</td>
<td>22</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>71</td>
<td>29</td>
</tr>
</tbody>
</table>

Although lung cancer is primarily a smoker's disease, about 10-15% (17,000 to 26,000) of lung cancer deaths per year occur among lifetime non-smokers (Thun, Henley et al. 2006). About 300 deaths from lung cancer per year in the United States are believed to occur secondary to environmental smoke among non-smokers (Belani, Marts et al. 2007). Other risk-factors that are considered contributory to the development of lung cancer among non-smokers are: a positive family history of lung cancer especially adenocarcinoma in young women, a variety of genetic, hormonal, environmental factors, and interactions among them, contribute to the development of lung cancer among non-smokers. (Blair and Freeman 2006; Zeka, Mannetje et al. 2006; Belani, Marts et al. 2007).

Besides smoking and genetics other risk factors have also been identified to play an essential role in the development of lung cancer, especially in smokers. These risk factors include occupational and environmental exposure to secondhand smoke, indoor and outdoor air pollution, radon, asbestos, metals like arsenic, cadmium and chromium, organic chemicals, radiation, diet, physical inactivity and occupation (Blair and Freeman 2006) (Parkes 1983; Steenland, Loomis et al. 1996; ACS 2007). Zeka et al. investigated the role of high-risk occupations in development of lung cancer among non-smokers. Women painters and rubber workers appeared to be at a
higher risk of developing lung cancer, especially if employed for more than 12 years. They also report a high odds ratio for developing lung cancer, in both genders, among workers exposed to nonferrous metal dust and fumes, silica and organic solvents (Zeka, Mannetje et al. 2006). Cassidy A. et al. discussed the need to develop a lung cancer risk prediction model for assessing risk for lung cancer which includes all factors like age, gender, smoking history, occupation, genetic polymorphism, dietary factors and inflammation (Cassidy, Duffy et al. 2007).

Last but not the least, with all that has been learned about cancer risk factors, tumor biology, carcinogenesis, techniques that are rapidly developing and becoming available for clinical care, another often overlooked factor is socioeconomic status as described by Ward et al. and Lengerich et al. (Lengerich, Wyatt et al. 2004; Ward, Jemal et al. 2004; Lengerich, Tucker et al. 2005). Many parts of the country and around the world are unable to have access to quality healthcare, prevention and education programs, screening facilities, and advanced technologies, therefore continue to have increased morbidity and mortality from various cancers including lung cancer.

The occurrence of lung cancer among non-smokers and the growing body of knowledge about other factors confounding tumor development emphasize the need for obtaining and providing accurate smoking and exposure histories to pathologists at time of tissue acquisition. Pathogenesis, susceptibility factors, gene expression and molecular markers may differ between smokers and non-smokers and may be further confounded by exposure misclassification.

**Molecular Biology of Lung Cancer:**

Carcinogens in cigarette smoke have been continuously identified since the onset of the lung cancer epidemic decades ago, but only over recent years are the cellular and molecular mechanisms underlying the pathogenesis of lung
cancer being understood and are rapidly developing. It is currently believed that an accumulation of multiple genetic, epigenetic alterations are required to develop lung cancer, including chromosomal abnormalities, aberrant DNA methylation, proto-oncogene activation, and tumor suppressor gene inactivation. These abnormalities cause inefficient detoxification and metabolism of carcinogens, deregulation of cell signaling pathways, malfunctioning of cell cycle checkpoints, dysfunctional cell repair mechanism and imbalance in cell proliferation and apoptosis signals resulting in carcinogenesis.

Activation of proto-oncogenes by DNA rearrangement, point mutation, gene amplification, chromosomal translocation, or overexpression results in oncogenes. More than 30% of NSCLCs especially adenocarcinomas have activated mutations of the Ras gene family (major K-ras2). SCLCs are not known to have ras mutations (Xiao 2006). myc family gene activation is more frequently seen in SCLC than in NSCLC. Some studies suggest that L-myc polymorphism may play a role in deferential susceptibility to lung cancer in smokers. Bcl-2 overexpression is seen in 75% of SCLC, but lesser in NSCLC.

Inactivation of tumor suppressor gene p53, due to mutation in its DNA binding motif, is known to occur in many cancers including lung cancer. Smokers have a higher frequency of p53 gene mutations than non-smokers. More than 90% of SCLCs and 50% of NSCLC have p53 mutations. Excessive p53 degradation, resulting in lesser activity can also result from deregulation of p14/ARF, an upstream regulator of p53. Mutations in p14/ARF were seen in 19-34% of NSCLC and absence of p14/ARF protein was seen in 65% SCLCs. CDK4/cyclin D1 complex kinase activity is inhibited by p16 protein, thereby preventing cells from proceeding into G1/S phase transition. Inactivation of p16 can thus lead to unregulated proliferation of cells and is seen in 30-70% NSCLCs. Retinoblastoma (Rb) protein, another tumor suppressor protein, binds to numerous proteins, including E2F transcription factor and inhibits them thus
preventing cells from proceeding into S phase. Inactivation of Rb protein is seen in 90% SCLCs and upto 30% of NSCLCs (Xiao 2006).

Many distinct deletions on chromosome 3p have been seen in lung cancer, suggesting the presence of possible tumor suppressor genes. The FHIT gene, with pro-apoptotic activity, was discovered as a result of homozygous deletions at 3p14.2 in lung cancer. About 40% of NSCLCs and 80% of SCLCs have abnormal transcripts of FHIT. Loss of homozygosity (LOH) of the FHIT gene is seen in 80% smokers and only about 22% in non-smokers, suggesting this site to be a target for carcinogenesis in response to tobacco exposure. Epigenetically inactivated RASSF1A is found in 90% SCLCs and 40% NSCLCs. RASSF1A promoter methylation is significantly associated with poor survival of lung cancer patients and RASSF1A silencing was associated with poor prognosis, poor differentiation, aggressive tumor, and invasion. It is thus a potential prognostic marker for lung cancer. Numerous other chromosomal abnormalities have been identified in various types of lung cancers (Xiao 2006).

Similarly, Chen et al found a 5 gene signature that significantly associated with relapse-free and survival of patients with NSCLC (Chen, Yu et al. 2007). The five genes are v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 3 (ERBB3), a member of the epidermal growth factor receptor family of tyrosine kinases associated with shortened cell survival, lymphocyte-specific protein tyrosine kinases (LCK), a member of the Src family of protein tyrosine kinases associated with cell differentiation, mobility and induction of apoptosis. The third genes is dual-specificity phosphatase 6 (DUSP6) known to inactivate ERK2 resulting in tumor suppression and apoptosis. Monocyte-to-macrophage differentiation associated protein (MMD) is the fourth gene identified by them. It is known to be associated with cancer metastasis. Finally, signal transducer and activator of transcription 1 (STAT1) gene that causes growth arrest and apoptosis. Bianchi et al. conducted a meta-analysis of all published data on microarray in lung cancer patients and tested their 80 gene model on a cohort of
patients to report a 10-gene predictive marker for stage I adenocarcinoma patients (Bianchi, Nuciforo et al. 2007). Ein-Dor et al predict that it will require thousands of samples to robustly identify predictive genes for cancers including lung cancer (Ein-Dor, Zuk et al. 2006).

There is rapidly growing information about the role of micro-RNA in numerous malignancies including lung cancer. micro-RNAs are non-protein-coding and can regulate gene expression by hybridizing with complementary sequences in the 3’ untranslated region (3’ UTR) of target messenger RNA (mRNA), thereby repressing mRNA translation and causing them to become unstable (Eder and Scherr 2005). Johnson et al showed that micro-RNA in the let-7 family downregulate let-60/RAS, which in turn suppresses let-7. The 3’ UTRs in human RAS gene contain numerous let-7 complementary sites, thus RAS expression may be regulated by let-7. Human lung cancer tissue demonstrated a decreased expression of let-7 compared to adjacent normal lung tissue, while RAS protein expression was increased, suggesting a potential role of this micro-RNA in lung cancer pathogenesis (Johnson, Grosshans et al. 2005). Takamizawa et al. report an association between low let-7 micro-RNA expression in lung cancer patients and shortened post-operative survival (Takamizawa, Konishi et al. 2004).

Epigenetic aberrations including promoter methylation and gene silencing are considered to be critical aspects of cancer development. Generalized DNA hypomethylation and regional promoter hypermethylation is commonly seen in cancers. The methylation profile of genes in neuroendocrine tumors of the lung including carcinoid and SCLC is very different from that of NSCLC and from adenocarcinoma and squamous cell carcinoma. Promoter hypermethylation of tumor suppressor genes is another potential biomarker for risk assessment and early diagnosis of lung cancer (Xiao 2006). There is plethora of research on the potential role of DNA methyltransferase (DNMT1 and DNMT3b) inhibitors and histone deacetylase inhibitors in cancer management.
RASSF1A methylation is another potential biomarker for diagnosis and prognosis of numerous cancers including lung (Fruhwald 2003; Lodygin and Hermeking 2006; Hesson, Cooper et al. 2007).

**Telomerase** activity is associated with immortal phenotype and is detected in almost all SCLC and about 80% of NSCLC.

Abnormal expression and activity of numerous **growth factors and their receptors** also result in signal pathways that may activate cell proliferation and tumor growth. Gastrin-releasing peptide (GRP) and bombesin-like peptide are two such factors believed to be associated with signal transduction pathway activation in lung cancer. Mutation in the epidermal growth factor receptor (EGFR) also called erBB1 are association with response to tyrosine kinase inhibitors and antibodies like sunitinib (Socinski, Novello et al. 2008), gefitinib, cetuximab (Vassilakopoulos, Troupis et al. 2001; Ramalingam, Forster et al. 2008). Increased expression of EGFR is present in 60-80% of NSCLCs but seldom in SCLCs and is associated with shorter survival (Xiao 2006). Heregulin receptor (**HER2/neu**) also belongs to the EGFR type-1 family and regulates epithelial cell proliferation and differentiation, the origin of lung cancer. HER2/neu expression is increased in 30% NSCLC, especially adenocarcinoma and is associated with drug-resistance and poor survival. Co-expression of EGFR and HER2/neu has a synergistic effect on poor survival. c-kit RTK and its ligand stem cell factor are co-expressed in 70% SCLC. Inhibition of this pathway causes growth inhibition and cell death in SCLC cell lines.

Disruption in the **apoptosis** signaling pathways are believed to contribute in tumorigenesis, metastasis and drug-resistance. Caspases, a family of aspartate-specific cystein proteases, regulate apoptosis by two cascade mechanisms; one via cell surface death receptors, and the other via the mitochondria. Caspase-8 known to be involved in the death receptor mediated apoptosis is also found to be activated in drug induced apoptosis, which is
primarily believed to be a mitochondrial phenomenon. Caspase-8 mRNA expression is absent in 79% SCLCs but retained in NSCLCs, suggesting its role as a potential biomarker for drug-resistance (Xiao 2006).

Akt/PKB, Serine/threonine kinase, is activated in response to growth factor stimulation and is involved in cell proliferation, apoptosis inhibition and cancer progression. Constitutive Akt activity was observed in pre-malignant and malignant human bronchial epithelial cells but not in non-malignant cells. Nicotine also activates Akt in vitro, in vivo and in smokers (Xiao 2006).

Protein Kinase C, a family of serine/threonine protein kinases, is believed to be critical in regulating two survival pathways, PI3K/Akt and MEK/ERK pathways, and is implicated in induction of drug-resistance in cancer cells. SCLC cells acquire chemoresistance in response to induced expression of PKC-ε via apoptosis inhibition mechanisms. PKC-δ inhibition renders NSCLC cells dramatically sensitive to chemotherapeutic drugs by increasing apoptosis.

Death-associated protein kinase (DAPK), a family of pro-apoptotic calmodulin-regulated serine/threonine protein kinases, although ubiquitously present in numerous tissues is usually inhibited by an auto-inhibitory mechanism. The gene DAPK is located on chromosome 9q34.1. This region is frequently found to have allelic loss in 50-64% of lung cancer. Aberrant promoter methylation causing inactivation is seen in a large number of lung cancers and is associated with poor prognosis and advanced pathological stage in NSCLC. However, promoter methylation is not seen with exposure to tobacco or asbestos (Xiao 2006).

**Histopathology in Lung Cancer:** Immunohistochemical staining for the enzyme cylooxegenase 2 (COX2) expression done by Laga et al. demonstrated a significant correlation between increased COX2 expression and decreased survival in patients with stage I and II NSCLC (Laga, Zander et al. 2005). Patel et
al. provide a scholarly review of literature providing evidence for the role of COX2 in NSCLC (Patel and Chiplunkar 2007).

The anti-apoptotic protein survivin is reported to be increased in numerous tumors including NSCLC. It is positively associated with advanced disease and recurrence after treatment, therefore is a potential prognostic marker (Sela 2002).

Mucin-type glycoprotein abnormalities are associated with numerous cancers. Normal type II pneumocytes and atypical lesions of type II pneumocytes demonstrate intense immunoreactivity to membrane bound MUC1 mucin, which is also found in more than 60% primary and metastatic NSCLC tissues. MUC1 mucin expression is also correlated with adenocarcinoma histology. MUC1 mRNA increases prior to histological appearance of tumors in a mouse model. NSCLC cell lines demonstrate a significant down-regulation of MUC1 expression as they differentiate away from type II pneumocyte lineage, thus identifying it to be a potential marker for histological lineage association during lung cancer progression (Jarrard, Linnoila et al. 1998).

The enzyme, Sphingosine kinase 1, (SK1) is also identified as an oncogene, since its up-regulation has been associated with carcinogenesis (Xia, Wang et al. 2002). French et al. and Johnson et al. showed that various human tumors, including lung cancer, demonstrated about a two fold higher expression of SK1 mRNA than adjacent normal tissue from same patients (French, Schrecengost et al. 2003; Johnson, Johnson et al. 2005). Johnson et al. also demonstrated by immunohistochemistry that SK expression was higher in non-small-cell human lung cancer tissue than in adjacent normal lung tissue from same subjects (Johnson, Johnson et al. 2005). Research in this field suggests that SK1 may be a potential marker for many cancers including lung cancer. Other alterations in the sphingolipid metabolism have been shown to influence cancer cell survival and response to radiation and chemotherapeutic drugs,

Vassilakopoulos et al. very elegantly designed and evaluated the diagnostic and prognostic value of squamous cell carcinoma antigen (SCC-Ag), a glycoprotein secreted by NSCLC tumors, in patients with NSCLC (Vassilakopoulos, Troupis et al. 2001).

Aldehyde dehydrogenase (ALDH) isoenzymes 1A1 and 3A1 are increased significantly in squamous cell cancer and adenocarcinoma compared to SCLC. These isoenzymes are also significantly increased in atypical pneumocyte and in normal pneumocyte of smokers, suggesting their role as markers for field carcinogenic exposure and tumorigenic transformation (Patel, Lu et al. 2008).

Glutathione S-transferase (GST) and glutathione synthase (GSH2) are involved with oxidative stress and carcinogenesis. GST-pi nuclear staining of more than 10% in NSCLC tissue is associated with decreased survival of stage I and II squamous cell carcinoma patients. GSH2 cytoplasmic staining of greater than 80% in NSCLC tissue is associated with improved survival of stage I adenocarcinoma, suggesting the potential of these enzymes as prognostic markers and therapeutic targets for treatment in early stage lung cancers (Allen, Granville et al. 2007).

**Cytology:** Bronchial brushing, washing and lavage during bronchoscopy are common sources of cytological material in the diagnosis of lung cancer. Halling et al. reported that fluorescence *in situ* hybridization (FISH) was more sensitive than cytology for brushing and washing but more so for brushing specimens (Halling, Rickman et al. 2006). In this study FISH was more sensitive
in detecting NSCLCs especially peripheral tumors. Cytology however, is more specific and they recommend that the two techniques can be utilized together to increase diagnostic sensitivity. Zander makes a very valid point in reference to Halling et al. study, that the time has come that lung cancer diagnosis should be parallel to pap-smear for cervical cancer where recent molecular techniques have improved the diagnostic yield for detecting varying degrees of cellular atypia in the process of screening for cervical cancer (Zander 2006). Similar molecular markers need to be developed, tested, validated and integrated into standard protocol for diagnosing lung cancer.

Gallagher et al. are studying the role of UDP-glucuronosyltransferases (UGTs) in cancer risk and development. 4-(Methylnitrosamo)-1-(3-pyridyl)-1-butanone (NNK) is a potent procarcinogen abundantly present in cigarette smoke. Its metabolite 4-(methylnitrosamo)-1-(3-pyridyl)-1-butanol (NNAL) is detoxified by glucuronidation by the UGT enzymes including UGT2B17. In a recent study Gallagher et al. report that both copies of the UGT2B17 gene deletion, seen in 10% of whites, is associated with a decreased urinary NNAL-Gluc:NNAL ratio in healthy smoking women, but not in healthy smoking men (Gallagher, Muscat et al. 2007). The decreased ratio is associated with decreased enzyme activity. In a case-control study including lung cancer patients and healthy controls, they report that women with the deletion of UGT2B17 gene have a higher risk of developing adenocarcinoma than women with heterogygous or wildtype gene. No such association is found in men or in any other histological type of lung malignancy. This study provides some insight into the various factors that influence susceptibility to cancer.

In summary, tumorigenesis is a complex phenomenon involving intricate interactions between underlying genetic predispositions, environmental, dietary, and behavioral factors. The presence of any one carcinogenic factor is usually not sufficient to result in cancer due to the presence of repair mechanisms and apoptosis process. DNA damage or mutations are recognized and repaired by
the repair proteins or the cell undergoes apoptosis. Therefore, it usually requires more than one mechanism to be dysfunctional for a malignant tumor to develop. For example, DNA may get damaged due to exposure to carcinogenic elements, or carcinogenic products may build up due to dysfunctional detoxification enzymes. This damaged DNA is not recognized and repaired, either due to lack of recognition of damage or due to inactive repair proteins. The damaged DNA may under synthesis in the S-phase of cell cycle, but this may cause genetic instability in the double helical structure of the DNA. This instability is usually recognized and apoptosis is triggered. However, if the cells apoptosis mechanism is also dysfunctional, then the damaged DNA may undergo proliferation. Thus the presence of multiple factors leads to carcinogenic transformation and proliferation into a malignant tumor (Kastan and Bartek 2004).

The intricate nature and interactions between type of carcinogenic exposure, particular enzyme dysfunction, and specific protein abnormalities with underlying genetic predisposition may play a role in development of specific types of cancers.

**Recent Advances in Technology and Treatment:**

**Radiology:** CT Screening programs are being developed for early detection of lung cancer among high-risk population (Henschke, Yankelevitz et al. 2006). New York Early Lung Cancer Action Plan (NY-ELCAP) recently published the results of their CT screening study (2007). They had a diagnostic yield of only 1.6% compared to 2.7% in the ELCAP study. However, all these patients were detected at an early stage where resection and treatment results in excellent long-term survival of up to 92% 10-year survival. Some of the concerns that question the feasibility of CT screening are the high cost, low detection, risk of over diagnosis, and increased risk due to radiation exposure. With time these questions may be answered and an effective screening protocol is likely to develop. Other developments in the detection techniques are underway,
including algorithms for automated detection of small pulmonary nodules (Enquobahrie, Reeves et al. 2007).

**Bronchoscopy:** Bronchoscopy has been a part of pulmonary clinical care and a tool for diagnosing lung cancer for over 50 years (Poppe 1960; Sabot 1960; Sabot 1960; Umiker 1960). However, it was flexible bronchoscopy, introduced in 1970, that revolutionized pulmonary care (Sackner 1975). Flexible bronchoscopy made it possible for bronchoscopists to go beyond the first bifurcation of the trachea into more tertiary divisions of the tracheobronchial tree to acquire mucosal and alveolar tissue. With recent increase in peripheral lesions in lung cancer, it is especially challenging but essential for bronchoscopists to be able to reach the desired site for tissue biopsy with accuracy. Bronchoscopy continues to evolve with decreasing diameters of bronoscopes to venture further into smaller bronchioles, numerous instruments that can be introduced through the bronchoscope, procedures that can be done and the tissues that can be acquired. However, since the lung is a complex organ that has over 22 generations of divisions in the tracheobronchial tree to reach the alveolar tissue, the efficiency of acquiring tissue accurately still remains a challenge. Lesions that are too far out in the periphery to be reached by bronchoscopy are biopsied using CT guided trans-thoracic needle aspiration technique.

Roth et al. reported that first bronchoscopy was the final diagnostic method to detect lung cancer in 163/363 (44.9%) of patients with lung cancer. In this study bronchoscopy results included biopsy, small volume lavage, bronchial washing, brushing, TBNA, and aspiration from the bronchoscope. Biopsy alone diagnosed 122/201 (60.7%) patients with lung cancer. In this study, predictors of diagnostic yield were endobronchial visibility, tumor size and distance from carina (Roth, Hardie et al. 2008). It is important to note here that their population included all patients diagnosed with lung cancer and not all patients suspected to have lung cancer. Also, in this study they do not distinguish between forceps biopsy and transbronchial biopsy. Forceps biopsy is usually intended for mucosal
lesions, tends to be central and probably endobronchially visible, while transbronchial biopsy is usually intended for parenchymal lesions, tends to be peripheral, and frequently endobronchially not visible.

Ahmed et al. conducted a study where they compared bronchoalveolar lavage (BAL) cytology results with transbronchial biopsy results that served as gold standard, to assess the sensitivity and specificity of diagnosing lung cancer by BAL. Such studies need to be reviewed with caution since they do not provide rationale as to why transbronchial biopsy was chosen as gold standard. This procedure itself may have low sensitivity and is a poor choice to be used as gold standard unless a clear rationale is provided (Ahmed and Ahmed 2004).

Koh et al. conducted an audit to assess the yield of diagnosing lung cancer at their centre using endobronchial ultrasound guided transbronchial lung biopsy (EBUS-TBLB) and transbronchial needle aspiration (EBUS-TBNA). They report a diagnostic yield of 62% with EBUS-TBLB and 88% with EBUS-TBNA in a population of 38 patients (Koh, Tee et al. 2008). They however, do not compare the yield of these processes without EBUS guidance. Also, their definition of yield includes non-malignant findings. This project focuses on malignancies. These differences will be further described in the discussion section of this dissertation. Wallace et al. recently reported sensitivity of diagnosing lung cancer by TBNA alone and compared it with fine needle aspiration using endobronchial bronchoscopic ultrasound (EBUS-FNA) and endoscopic ultrasound (EUS-FNA). They report a sensitivity of 36% by TBNA, 69% each by EBUS-FNA and EUS-FNA and 93% with EBUS-FNA plus EUS-FNA combined (Wallace, Pascual et al. 2008). This study did not assess diagnostic yield by transbronchial biopsy. Simultaneously Toumoy et al. conducted a similar study but focused on patients with central lesions and initial negative bronchoscopy and report sensitivity of diagnosing lung cancer with EBUS-TBNA to be 83% with a negative predictive value of 25% (Toumoy, Rintoul et al. 2008). They too did not study TBBx yield.
Halling et al. report sensitivity of diagnosing lung cancer by bronchial brush using cytology only to be 51%, and FISH only to be 71%. However the sensitivity increased to 75% when cytology and FISH results were combined. Similarly for bronchial washing specimens sensitivity of detecting lung cancer was 44% using cytology only and 49% by FISH only, but increased to 61% using combined results. Specificity of diagnosing lung cancer, in patients whose initial bronchoscopy results were negative, using bronchial brush specimen by cytology was 100% and by FISH was 83%. Specificity for bronchial wash specimen using cytology was 100% and using FISH was 95% (Halling, Rickman et al. 2006).

It is important to note that FISH, EBUS-FNA and EUS-FNA are not routinely used procedures to diagnose lung cancer. Conventional methods to diagnose lung cancer are cytology, forceps biopsy, transbronchial biopsy and TBNA. Although these techniques have a specificity of 100%, their sensitivity is in the range of 36 to 60%. Multiple techniques are used in combination to improve the sensitivity of diagnosing lung cancer.

The British Thoracic Society recommendations for suspected lung cancer include biopsy, brushing, and washing for visible lesions. TBNA is not included in this guideline and there is no consensus on guidelines to sample peripheral lesions (Roth, Hardie et al. 2008).

Detterbeck et al. and Silvestri et al. provide a very good comprehensive coverage of the various techniques that are available for accurate staging, each techniques strengths and weakness and how they compliment each other in different groups of patients (Detterbeck, Jantz et al. 2007; Silvestri, Gould et al. 2007; Schultz, Sanders et al. 2008). Sihoe et al. emphasis on the revolutionary changes in non-invasive techniques in diagnosing and staging lung cancer and include updates on the role of molecular biology in staging (Sihoe and Yim 2004).
Advances are being made by numerous physicians, scientists and engineers to improve the efficiency of acquiring the desired tissue accurately. Right here at the Pennsylvania State University, Hershey Medical Centre there is an ongoing collaborative project to develop a virtual navigation system (VNS) from fusion of 3-dimensional CT image data with the live bronchoscopic video that enables the bronchoscopist to arrive at the lesion with more confidence and accuracy (Higgins, Ramaswamy et al. 1998; Kiraly, Helferty et al. 2004; Dolina, Cornish et al. 2008; Higgins, Helferty et al. 2008).

Besides VNS there are numerous other techniques being developed to assist in obtaining accurate tissue including electromagnetic guided bronchoscopy (Gildea, Mazzone et al. 2006), fluorescent bronchoscopy (Gilbert, Luketich et al. 2004; Stanzel 2004), and lasers (El-Bayoumi and Silvestri 2008; Short, Lam et al. 2008). Laser capture microdissection technology is being utilized on bronchoscopically acquired specimens to identify and correlate molecular markers with conventional histological findings (Flake, Rivera et al. 2007). This line of research is perfectly aligned with the long-term vision of this project. Lasers in bronchoscopy are also being researched as therapeutic tools to achieve relatively non-invasive techniques for treating certain pulmonary lesions including low grade tracheobronchial mucoepidermoid carcinoma (Li, Huang et al. 2004). Sutedja et al. highlight the therapeutic potential of numerous techniques like lasers, electrocautery, cryotherapy, photodynamic therapy, and brachytherapy that may be utilized via bronchoscopy to treat early stage NSCLC. However, the important point to note is that these patients need to be selected carefully, they must be detected very early and must be staged accurately to have limited disease (Sutedja, van Boxem et al. 2001). To harness the maximum benefit of these technologies, and to improve patient outcome, it is critical to first improve the diagnostic and staging techniques.

In summary, improvements in CT scan screening will help detect lung cancer earlier and improvements in bronchoscopy techniques including
transbronchial biopsy and TBNA will improve accuracy and staging of disease. These advances are absolutely essential to improve detection of patients with lung cancer in earlier stages when lesions are smaller, especially with the current trend towards more peripheral lesions. Improvements in these fields will pave the way to the next steps of planning appropriate and effective treatments and utilizing numerous other recent advances in technology for maximum effective impact on patient outcome.

**Surgery:** Surgery is the treatment of choice for early stage cancer and limited stage disease. The advantages of surgery are that it is localized and limited to target area therefore avoids systemic side effects. *En bloc* resection is the usual standard and involves removal of the tumor along with surrounding normal tissue, and the regional lymph nodes at risk. It is most suitable form of treatment for solid tumors. Various surgical treatment options include single and multiple lobectomy, pneumonectomy, and wedge resection. Lymph node dissection is an independent surgical procedure usually performed with all lobectomies for lung cancer and pneumonectomies.

However, since recurrence is a common event, recently combined treatment has received some attention with the goal to achieve better long term survival of patients by combining thoracoscopic lobectomy and chemotherapy. Peterson et al. report that thoracoscopic lobectomy was associated with shorter length of stay in hospital and better compliance with post-operative chemotherapy (Petersen, Pham et al. 2007). Ginsberg et al. investigated if limited resection would have better outcome for peripheral T1N0 stage NSCLC when compared to lobectomy. Unfortunately their study results showed that limited resection was associated with increased recurrence and mortality from cancer (Ginsberg and Rubinstein 1995). Numerous other studies have investigated the outcomes of various surgical techniques in different stages and types of lung cancers (Lederle 1996; Fibla Alfara, Gomez Sebastian et al. 2003; Nakamura, Aute et al. 2005; Birdas, Koehler et al. 2006; El-Sherif, Gooding et al. 2007).
Some recent advances in surgical techniques include non-invasive surgery for early lung cancer using lasers, photocoagulation, cryotherapy, electrocautery, photodynamic therapy, and brachytherapy (Sutedja, van Boxem et al. 2001).

It is believed that recurrence at site of original tumor may be a consequence of “incomplete” resection or “inaccurate” staging. The so called “normal” surrounding tissue, reported by conventional histological examination that is removed during resection is found to have the genetic and molecular characteristics of malignant cells and it is a matter of time this tissue will eventually be malignant on histological examination. This information highlights the need to redefine normal and malignant cells at a genetic and molecular level. This new guideline can then be used to redefine staging criteria and surgical margins.

**Radiation therapy:** External beam therapy and brachytherapy are two commonly used modalities of treatment for lung cancer. An energized particle, usually a photon, causes release of free radicals and DNA damage that triggers apoptosis of the tumor cells. Side effects can be acute and involve irritation of the skin, esophagus and lung, or chronic and involve fibrosis and radiation damage of the lung, radiation damage to the heart. Yet another complication of radiation therapy is development of a second primary.

Resistance and recurrence are commonly seen in response to all treatment modalities for lung cancer. Resistance can occur if the malignant cells have dysfunctional apoptosis mechanism. Resistance can also develop if the cells can enhance repair mechanisms for damaged DNA. Yet another molecular mechanism known to play a role in development of resistance is the upregulation of cell survival proteins like sphingosine kinase (Nava, Cuvillier et al. 2000).
Chemotherapy: Conventionally chemotherapeutic agents used for lung cancer treatment are classified by their targets. Like radiation, platinum based agents and antimetabolites target DNA. Platinum agents include cisplatin and carboplatin. Their mechanism of action includes attaching large platinum adducts on the DNA. This triggers apoptosis of tumor cells. Antimetabolite agents include gemcitabine and pemetrexed. These agents act by competing with known metabolites. Gemcitabine is a pyrimidine neoleoside analogue. The cytotoxic effects of gemcitabine are exerted by its active metabolites diphosphate (dFdCDP) and triphosphate (dFdCTP) neoleosides. dFdCDP-assisted incorporation of dFdCTP into DNA inhibits DNA synthesis and triggers apoptosis. Pemetrexed is a folate antimetabolite, and inhibits the enzymes thymidylate synthase, dihydrofolate reductase and glycinamide ribonucleotide transformylase which are essential for DNA and RNA formation and cell replication.

Topoisomerases (Topo) are a family of enzymes required for unwinding DNA and making it accessible for transcription during replication. Topoisomerase inhibitors are another group of chemotherapeutic agents used to treat lung cancer. Topo-I inhibitors include drugs like topotecan, irinotecan, and Camptothecin. They act by blocking the rejoining step of the cleavage/religation reaction to topo-I. Topo-II inhibitors include etoposide and doxorubicin. Doxorubicin inhibits DNA replication in the S-phase of cell cycle by stabilizing topo-II complex after it has broken the DNA chain for replication preventing the DNA double helix from being released.

Vinca alkaloids and taxanes are groups of chemotherapeutic drugs that target microtubules. Drugs included in this group are vinorelbine, thiotepa, docetaxel, paclitaxel, vincristine, and chlorambucil. Vinca alkaloids act by depolimerizing microtubules by binding to tubulin monomers during mitosis. Newly synthesized chromosomes are thus unable to separate, aborting cell replication and triggering cell death. Taxanes stabilize GDP-bound tubulin and cause “frozen mitosis”. Taxanes are also believed to be radiosensitizers.
The most recent target in chemotherapy is the epidermal growth factor receptor tyrosine kinase (EGFR-TK). Drugs in this group include gefitinib, erlotinib, sunitinib, sorafenib, and cetuximab. These drugs inhibit the receptor tyrosine kinase and prevent all down stream signaling effects of cell survival and proliferation. Monoclonal human antibodies against vascular endothelial growth factor (VEGF) like bevacizumab have also been recently introduced for their anti-angiogenic effect (Belani, Marts et al. 2007; Heymach, Johnson et al. 2007; Owonikoko, Ragin et al. 2007; Ramalingam and Belani 2007; Ramalingam and Belani 2007; Ramalingam, Parise et al. 2007; Ramalingam and Belani 2008; Ramalingam, Forster et al. 2008; Ramalingam, Dahlberg et al. 2008; Socinski, Novello et al. 2008).

Since chemotherapy is systemic treatment it is associated with side effects that can be so severe as to warrant discontinuation of treatment. Side effects can include myelosupression causing low platelet count resulting in easy or spontaneous bleeding, low immune status causing recurrent infection, anemia resulting in easy fatigability, mucositis, nausea, vomiting, diarrhea, loss of hair, and loss of appetite. Some cause peripheral neuropathy and other side effects resulting from specific groups of drugs.

Some groups are working towards enhancing current treatment modalities to achieve better long term survival of patients by combining thoracoscopic lobectomy and chemotherapy (Petersen, Pham et al. 2007). Although early detection and management are associated with better survival, the majority of cases are currently diagnosed at a latter stage and about 70% of these cases are resistant to standard treatment regimens. The remaining 30% cases develop resistance within 4-6 months of treatment (Rosell, Cecere et al. 2006). Therefore, the obvious pressing need is to develop better treatment regimens in order to improve lung cancer patient survival in the long-term.
Research done on the sphingolipid metabolic pathway and prostate cancer promises that SK1 is a potential diagnostic and prognostic marker (Pchejetski, Golzio et al. 2005). Inhibitors of SK1 have proved to be promising therapeutic agents in mouse models of breast cancer (French, Upson et al. 2006).

Selenium rich diet and compounds like 1,4-phenebis (methylene) selenocyanate (p-XSC) are being actively studied on in vitro and mouse lung cancer models to evaluate their potential as chemopreventive agents. There is evidence suggesting that they may prevent tumor development by decreasing oxidative stress damage (Richie, Kleinman et al. 2006).

**Gap Analysis:**

**General:** As seen from the literature reviewed thus far, extensive amounts of research have been done and are continuing to develop with the goal of improving lung cancer detection and survival. However, as Dr. Elias Zerhouni, director of National Institutes of Health points out that all these efforts have had little impact on routine clinic care of patients. Five-year survival rate of lung cancer patients has remained a dismal 13-16% over 4 decades now.

In this section the author will attempt to highlight some of the gaps which may be responsible for this lag. The approach taken is to compare lung cancer with other cancers and advancements made in those areas.

Starting with screening, cancers like breast, prostate, cervical and colon cancer have routine screening procedures recommended and increasingly implemented to detect them at earlier stages. Mammography and manual breast examination are regularly being done to detect breast cancer early. Breast cancer incidence and death rates have declined significantly recently. Incidence has decreased by 4.1 annual percent change (APC) for the time period 2001-
2003 and death from breast cancer has decreased by 1.4 APC for the time period 1999-2003 (Jemal, Siegel et al. 2007).

Prostate specific antigen (PSA) and digital examination are routinely used procedures to screen and detect prostate cancer at an early stage. Due to the increase in screening technique the incidence of prostate cancer spiked, probably due to lead-time bias, increasing by 16.4 APC between 1988 and 1992 and decreased by 10.8 APC in the following time period between 1992-1995. However, the death rate from prostate cancer decreased by 0.6 APC in 1991-1994 and by 4.0 APC in 1994-2003 (Jemal, Siegel et al. 2007).

Routine colonoscopy is now recommended to screen for early colon and rectum cancer. The incidence of colon and rectum cancer has decreased by 2.1 APC in 1998-2003 and the death rate from these cancers has decreased by 2.8 APC in 2001-2003 (Jemal, Siegel et al. 2007).

Unlike the breast, prostate and colon, the lung, especially peripheral bronchioles and alveoli, is not an easily accessible organ amenable to physical examination. Lack of approved routine screening procedure for lung cancer also contributes to delayed diagnosis. Limited sensitivity of diagnostic procedures like bronchoscopy (BAL, brushing, TBBX, TBNA etc) also contributes to delay in diagnosis, and probably inaccurate staging of disease, which directly impact treatment and prognosis. In spite of tremendous progress and efforts in the development of diagnostic and prognostic biomarkers for lung cancer there is currently no approved, routinely used biomarker to detect lung cancer. Recent advances in diagnostic and therapeutic regimens have also contributed little to the long-term survival of patients with lung cancer. Improvements in each of these stages of lung cancer care are essential in detecting patients at an earlier stage of disease, accurately staging and efficiently treating them in order to decrease the mortality rate.
Rationale:

The Institute of Medicine’s publication “Crossing the Quality Chasm” states that the aim for 21st century health care system ought to be safe, effective, patient-centered, timely, efficient and equitable (Corrigan, Donaldson et al. 2004). In accordance with these guidelines, lung cancer diagnosis and staging ideally ought to be rapid, accurate, minimally invasive, pertinent to clinical treatment, and prognosis. A detailed understanding of process of lung cancer care will lead to identification and development of strategies to improve the standard of diagnosis.

Molecular markers are the new great tool and hold the promise of individualized therapy. Identifying various tissue acquisition processes and establishing associations between spatial and temporal tissue biomarker profiles, diagnosis, management, response to treatment and prognosis has the potential to efficiently integrate molecular medicine with clinical care of patients. However, use of molecular markers will be most useful if bronchoscopy yield is accurate. Therefore it is first important to assess the bronchoscopy yield at the Hershey Medical Center if the long term goal is to integrate molecular markers as a part of routine diagnostic and clinical care for lung cancer patients. This project is thus the first step towards the long term goal.

Effective utilization of molecular markers in the diagnosis and management of lung cancer needs to fulfill the following criteria:

- Obtain primary tumor tissue, distinguishable from surrounding tissue that may have field carcinogenic changes or genetic polymorphism. Primary tumor biology may also be distinguishable from markers present in nodal tissue, distant metastases and blood.
- Molecular markers would be more useful on histological tissue samples rather than cytological samples in order to evaluate tissue architecture.
This would exclude bronchial brushing, washing lavage and fine needle aspiration.

- It would be desirable to obtain tissue by a non-invasive technique during initial stage of diagnosing and staging, in advance of therapy. This excludes surgical resections.

- It would also be desirable to obtain tissue by a technique that can be repeated during follow-up for assessment of response to therapy, development of resistance, recurrence etc.

TBBx fulfills these criteria and promises to be an ideal step for incorporation of molecular markers into routine process of care, diagnosis and management of patients with lung cancer. TBBx is chosen as the focus of this project because of its potential in contributing to the overall process of care. This is a procedure routinely done, usually at an early stage of diagnosis and may be repeated for staging, follow-up, recurrence etc. However, the efficiency and accuracy of TBBx in obtaining primary tumor tissue has not been validated at the Hershey Medical Center. This project is undertaken with the short-term goal of assessing the accuracy of transbronchial biopsies in diagnosing lung cancer at the Hershey Medical Center. Although TBBx is identified for its potential, it has its limitations that need to be recognized. Some of the concerns associated with TBBx are:

- Frequency of performance at HMC
- Accuracy of obtaining tissue from desired site
- Quantity of tissue
- Diagnostic yield
- Tumor heterogeneity, hence not representative
- Accuracy of returning to same site for follow-up
Goal, Hypotheses and Specific Aims:

The long-term vision is to incorporate molecular markers as part of routine process of care for diagnosing and managing patients with lung cancer. To achieve this vision it is essential to identify appropriate markers and to utilize them on accurately acquired tissue at time of diagnosis. TBBx provides tissue for histological diagnosis and is a procedure done at the early stage of diagnoses and staging. Molecular markers promise to provide maximum clinical benefit if utilized at this stage. However, it is first essential to establish the diagnostic yield of TBBx and identify gaps in the current process of care for lung cancer patients in order to prepare for incorporation of molecular markers.

Thus the goal of this project is to determine the pre-test probability and diagnostic yield of TBBx in detecting lung cancer and to identify gaps in the current process of care associated with use of TBBx for lung cancer diagnosis at the Hershey Medical Center.

The hypotheses of this project are that the current TBBx practice at the Hershey Medical Center has a low diagnostic yield, high false negative results, especially for small peripheral lesions, provides insufficient clinical data to the interpreting pathologist, and the current process of care for lung cancer patients has variable time gaps between CT scan, TBBx and lung resection. These hypotheses are addressed with the following specific aims:

1. To define the current clinical process of care and to identify clinical datasets used for key elements of lung cancer diagnosis.
2. To critically evaluate a retrospective cohort of bronchoscopically acquired TBBxs and estimate the pre-test probability of diagnosing lung cancer.
3. To identify and integrate matched pairs of TBBx and lung resections in order to validate the performance of TBBx against lung resection as the gold standard.
4. To benchmark TBBx pathology reports as a function of digitally acquired CT lesions using a novel classification scheme and identify characteristics of lesions accurately and inaccurately identified by TBBx.

5. Assessment of pulmonary lesion characteristics accurately and inaccurately diagnosed by TBBx.

6. To estimate the time intervals between CT scan, TBBx and lung resection in order to identify time gaps in the current process of care for lung cancer patients at the Hershey Medical Center.
III: Materials and Methods

**Specific Aim 1:** To define the current clinical process of care and to identify clinical datasets used for key elements of lung cancer diagnosis.

As per the guidelines suggested by Rubin *et al.* (Rubin, Pronovost *et al.* 2001; Rubin, Pronovost *et al.* 2001) process of care and quality measures are assessed by defining the following:

i) the audience of this project,

ii) The clinical area being evaluated in this study,

iii) The assessment team,

iv) The component of the process being measured,

v) Preliminary tests,

vi) Scoring and analytic specifications.

To understand and define the current process of care at the Hershey Medical Center for diagnosing, staging and managing patients with lung cancer, various resources are utilized. Attending and discussions at pulmonary and lung cancer clinics, observation of bronchoscopy procedure conducted by pulmonologists including R. Bascom M.D., J. Toth M.D. and K. Gleeson M.D, clinical and research conferences involving interaction between pulmonologists, pathologist (D. Zander M.D., H. Crist M.D., S. Saffee M.D.), oncologists (C. Belani M.D.), radiologists (R. Maharaj M.D.) and cardiothoracic surgeons (M. Lazar M.D. and D. Campbell M.D) are among the various resources accessed.

Access to patients’ files for comprehensive information from the medical records department is feasible under the IRB regulations for this project (IRB # 21529EP). However, clinical files consist of free text and have no consistent structure. The focus of this project is to look for clinical information transmitted among members of the diagnostic team. Electronic record datasets, on the other hand, are defined elements using consistent language and coding. Therefore
numerous such disparate electronic datasets are identified and utilized in this project to retrieve data. These datasets include:

i) billing records to identify TBBx (information source Ruth Kreider, Department of Medicine, HMC),

ii) Power Chart to access TBBx pathology reports (Cerner, electronic database available on all computers connected to the HMC network),

iii) CT scan digital image repository (retrieved by M. Khan M.D. research assistant, Pulmonary Allergy Critical Care Division, Department of Medicine, HMC),

iv) “Natural Language IIA search”, a software program available in the pathology department used to access lung resections repository (retrieved by S. Saffee M.D. fellow in Pulmonary Allergy Critical Care Division, Department of Medicine, HMC),

v) The American Joint Committee on Cancer (AJCC) guided lung cancer classification protocol adopted by the Department of Pathology at the Hershey Medical Center (from H. Crist M.D. Assistant Professor, Department of Pathology).

Authorization to access these datasets is obtained as per institution policies and IRB regulations, after CITI training and inclusion of author’s name into the IRB investigators list.

These datasets are found to be disparate with each source containing essential, but partial information. However, all datasets contain patient name, date of birth, and medical record number. These are utilized to develop links between the datasets in order to assemble the comprehensive information required to achieve the aims of this project.

Billing records are used to identify all TBBx procedures conducted at the Hershey Medical Center between August 31st 1999 and June 30th 2005. Access to power chart (Cerner), an electronic dataset with individual patient information
is utilized to acquire TBBx pathology reports. American Joint Committee on Cancer (AJCC) guided lung cancer classification protocol adopted by the Department of Pathology at the Hershey Medical Center is utilized to develop the epidemiological codebook to document and organize data. These sources served as the key elements utilized to fulfill goals of aim 2 i.e. to critically evaluate a retrospective cohort of bronchoscopically acquired TBBxs and estimate the pre-test probability of diagnosing lung cancer.

Lung resection pathology report repository is the dataset that serves as the gold standard for lung cancer diagnosis in this project and serves as the key element for achieving aim 3. "Natural Language IIa search", a software program available in the Department of Pathology, is used to identify all patients that underwent lung resection between 2\textsuperscript{nd} June 2000 and 8\textsuperscript{th} December 2006. AJCC lung cancer classification protocol is again utilized to develop the epidemiological codebook to document and organize lung resection pathology report data. This dataset is linked with the TBBx dataset using patient name and medical record number. Patients common to both datasets are identified and provide subsequent data to achieve goals of aim 3 i.e. to identify and integrate matched pairs of TBBx and lung resections in order to validate the performance of TBBx against lung resection as the gold standard.

Digital CT scan image repository (IDX/GE) is yet another dataset utilized to collect and organize data about pulmonary lesion characteristics that are biopsied. A novel classification scheme is developed based on the lesion characteristics. Digital CT scan image repository is the key dataset utilized to fulfill goals of aim 4 i.e. to benchmark TBBx pathology reports as a function of digitally acquired computerized tomography (CT) lesions using a novel classification scheme and identify characteristics of lesions commonly biopsied and lesions frequently reported to be positive for malignancy.
Data from all three datasets, TBBx pathology report, digital CT scan, and lung resection repository are integrated and analyzed to validate the outcome of TBBx pathology reports. These data identify lesions accurately and inaccurately identified by TBBx and fulfill goals of aim 5: to assess CT scan acquired pulmonary lesion characteristics accurately and inaccurately diagnosed by TBBx.

Finally, the date of service for CT scan, TBBx and lung resection are utilized to achieve goals of aim 6 i.e. to estimate the time intervals between CT scan, TBBx and lung resection in order to identify time gaps in the current process of care for lung cancer patients at the Hershey Medical Center.

After acquiring a working knowledge of the various medical specialists involved in the care of lung cancer patients and identifying the numerous datasets where patient information is stored, the current clinical process of care is defined. A simplistic overview of how a particular patient suspected of lung cancer is diagnosed and managed is outlined. The numerous investigations conducted, procedures performed, tissues acquired, treatment options, rehabilitation, recovery and follow up is outlined. The understanding of process of care guided the development of all the other specific aims of this project.

The potential short comings of this project are:

i) Inability to access patient files may limit acquisition of comprehensive individual patient information including detailed smoking history, history of passive smoke exposure, family history, past history of lung cancer or any other cancer, number of TBBx procedures done, stage of cancer, number of lung resection surgeries done etc. An attempt to overcome this limitation led to development of establishing structured links between the disparate datasets to get relatively more comprehensive information specifically with regard to diagnostic yield of TBBx as a diagnostic procedure for lung cancer. This potential limitation led to identifying the strengths of the various established electronic
datasets and the need to provide relevant information to the physician team in these datasets.

ii) Billing records may have some misclassification and may not accurately distinguish between various procedures conducted during bronchoscopy including TBBx and forceps biopsy. This limitation is mainly due to the retrospective design of the study and may be overcome in a prospective study. Misclassifications is documented and excluded from statistical analysis of data. However, it is judged that misclassification is a rare error and does not contribute significantly to the outcome of the study.

iii) Due to the project being a retrospective study, it is possible that detailed information may not be available including exact site of TBBx beyond the lobar bronchus, pre-procedure differential diagnosis etc. For TBBx, there may be instances where multiple sites are biopsied from a patient during a procedure and processed in a single jar. The retrospective design of this study may limit the accuracy of site of lesions biopsied. However, it is presumed that it occurs rarely and does not skew the outcome of the study significantly.

iv) The TBBx data collection, organization and analysis do not incorporate methods to identify patients that undergo multiple TBBxs. Therefore, if a patient underwent more than one TBBx procedures on different dates, they are recorded as if they were two or more different patients. However, when integrating TBBx data with lung resection data (specific aim 3), if a patient had more than one TBBx, the most recent pre-op TBBx report is included. This limitation prevents accurate assessment and quantification of these elements’ contributions to the delay in diagnosis of lung cancer thus compromising the outcome of aim 5.

v) The software used to identify lung resections, may erroneously include a few TBBx as lung resections. These will be identified and excluded. It is
presumed that all errors are recognized and excluded therefore do not affect the outcome of the study.

vi) Surgical lung resection is treatment of choice for early stages (I, II, IIIa) of lung cancer. Using lung resection as gold standard thus limits the population of lung cancer patients included in aim 3 to those with early stage disease. Inability to have another histopathological method to serve as gold standard for advanced stage disease limited patients with advanced disease from being included in aim 3 of the study. Again inability to access patient files did not provide information on stage of disease, whether TBBx is performed on patients known to have lung cancer pre-procedure, pre- or post-operative radio/chemotherapy, and survival of patients. This also limited an accurate assessment of quality and process of care for lung cancer patients. However, maximum possible analyses are conducted from data available.

vii) Digital image repository for CT scan data (IDX/GE) has only recently been established at the Hershey Medical Center, since 2002. However, this project includes TBBxs done between August 31st 1999 and June 30th 2005. Therefore many patients' CT scan data may not be available in the digital image repository, limiting the number of patients included in aim 4 of the project. However, it is considered to be a critical resource for future research purposes and this project aims to provide an initial assessment of the strengths of this resource.

The outcomes of this study, in spite of its limitations, provide critical information about current process of care, diagnostic yield of TBBx as a tool for diagnosing lung cancer. The results from this study establish the feasibility, significance and guidelines for organizing and conducting a prospective study with more comprehensive information and accurate data.
Specific Aim 2: To critically evaluate a retrospective cohort of bronchoscopically acquired TBBxs and to estimate the pre-test probability of diagnosing lung cancer.

The protocol for this part of the study is approved by the IRB (# 21529EP). Financial services provide pulmonary division bronchoscopy billing records (47 series) to identify consecutive TBBxs (TM_ID 4731628) performed between August 31st 1999 and June 30th 2005 (5 years 10 months) at the Hershey Medical Center. Pathology reports of TBBxs done during this period are acquired from the Power Chart (Cerner) database. An example of the pathology report is shown in appendix A.

An assessment of pathology reports available is performed to develop an epidemiological code book. An Excel worksheet format from Microsoft Office is used to document data according to the codebook developed (table 4). Data elements include demographics, performing bronchoscopist, clinical history, smoking history, pre-procedure differential diagnosis as mentioned in pathology report, site from which tissue is obtained, formalin fixation of tissue or freezing or any other tissue processing, aggregate volume of tissue obtained, number of tissue fragments, adequacy of tissue, presence of malignancy, type of malignancy, cellular atypia (metaplasia, hyperplasia), non-malignant findings including inflammation, granulation, fibrosis, vasculitis, hemorrhage, and staining done etc.

| Table 4: Epidemiological codebook developed to document data obtained from pathology reports for all TBBxs performed during the period of the study. |
|---|---|---|---|---|
| TRANSBRONCHIAL BIOPSY PATHOLOGY REPORTS CODEBOOK |  |  |  |
| PI: Rebecca Bascom, M.D., M.P.H. |  |  |  |
| IRB: 21529 EP (approved through August 31st 2008) |  |  |  |
| Cl# | Field # | Variable | Field Name | Field Type | Field Width |
| A | 1 | Subject's research case # (21529 EP) | RES_CASENUM | N | 10 |
| B | 2 | Accession # (pathology report #) | CAT_ACC_NUM | N | 10 |
| C | 3 | Date of Birth | DOB | D | mm/dd/yyyy |

43
<table>
<thead>
<tr>
<th>Column</th>
<th>Code</th>
<th>Description</th>
<th>Value</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>D 4</td>
<td>Age (in years) at time of procedure</td>
<td>AGE</td>
<td>N 3</td>
<td></td>
</tr>
<tr>
<td>E 5</td>
<td>Sex</td>
<td>SEX</td>
<td>N 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 = Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 = Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F 6</td>
<td>Bronchoscopists (Responsible physician's) code</td>
<td>MD_CODE</td>
<td>N 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>47080 = Wojner M M</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47085 = Bascom R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47300 = Gleeson K</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47535 = Robinson R W</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47180 = Tucakovic M</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47590 =</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47075 = Imadojemu V A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47175 = Vendor Roberts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47025 =</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47045 =</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G 7</td>
<td>Clinical History (Symptoms)</td>
<td>CLIN_HISTORY</td>
<td>N 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>position 1 = cough</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>position 2 = hemoptysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>position 3 = fever</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>position 4 = weight loss</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>position 5 = past h/o cancer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>position 6 = past h/o pneumonia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>position 7 = other</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>position coding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = absent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 = present</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9 = not provided</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H 8</td>
<td>Smoking History</td>
<td>SMOK_HISTORY</td>
<td>N 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>position 1 = current smoking status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>position 2 = pack years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>position 3 = time since quit smoking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>position 4 = never smoker (&lt;20 packs in lifetime)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>position 5 = passive smoke exposure (PSE)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>position coding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>for position 1 (current smoking status)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = not provided</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 = currently smoking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 = currently not smoking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 = h/o heavy smoking + but details not provided</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>for position 2 (pack years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = not provided</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 = &lt;20 pack years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 = &gt;20 pack years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 = not applicable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>for position 3 (time since quit)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position</td>
<td>Code</td>
<td>Description</td>
<td>Diagnosis</td>
<td>Suspected</td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>-------------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>not provided</td>
<td>Lung cancer</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>&lt;5 years ago</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>&gt;5 years ago</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>no exposure as child or adult</td>
<td>Metastasis</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>exposure as child only</td>
<td>Lymphoma</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>exposure as adult only</td>
<td>Sarcoïdosis</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>exposure + as child and adult</td>
<td>Diffuse Lung Disease</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>not provided</td>
<td>Hemorrhage</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>exposure as child only</td>
<td>Infection</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>exposure as adult only</td>
<td>Other</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>exposure + as child and adult</td>
<td>Other</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>not provided</td>
<td>Infection</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>exposure as child only</td>
<td>Other</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>exposure as adult only</td>
<td>Other</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>exposure + as child and adult</td>
<td>Other</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>not provided</td>
<td>Other</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>exposure as child only</td>
<td>Other</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>exposure as adult only</td>
<td>Other</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>exposure + as child and adult</td>
<td>Other</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>not provided</td>
<td>Other</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>exposure as child only</td>
<td>Other</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>exposure as adult only</td>
<td>Other</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>exposure + as child and adult</td>
<td>Other</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Column</td>
<td>Value</td>
<td>Description</td>
<td>Code</td>
<td>Count</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>--------------------------------------------------</td>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td>T</td>
<td>20</td>
<td>Specimen site (segmental airway and Lung)</td>
<td>SPEC_SITE_AIR</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>1st position</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>not provided</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Right</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Left</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Trachea and primary carina</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>other</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>not specified</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2nd position</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>not provided</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>upper lobe</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>middle lobe</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>lower lobe</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>lingula</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3rd/4th position</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>00</td>
<td>not provided</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>01</td>
<td>apical</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>02</td>
<td>posterior</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>03</td>
<td>anterior</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>04</td>
<td>superior</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>05</td>
<td>inferior</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>06</td>
<td>lateral</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>07</td>
<td>medial</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>08</td>
<td>anterior basal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>09</td>
<td>lateral basal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>posterior basal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>anteromedial basal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>apicoposterior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>21</td>
<td>Fixation</td>
<td>FIXATION</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Formalin fixed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Unfixed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>not specified</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>22</td>
<td>Tissue Volume</td>
<td>TISS_VOL</td>
<td>N</td>
</tr>
<tr>
<td>W</td>
<td>23</td>
<td>Number of Tissue Fragments</td>
<td>TISS_NUM</td>
<td>N</td>
</tr>
<tr>
<td>X</td>
<td>Adequacy of TBBx specimen for alveolar tissue</td>
<td>ALV_SPEC_ADEQ</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>----</td>
<td>---------------------------------------------</td>
<td>----------------</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>0= not specified</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1= No alveolar tissue present</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2= 1-5 alveoli seen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3= blood or debris obscuring tissue</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4= &gt; 5 alveoli present (adequate for diagnosis)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5= Muhamad alveoler tissue present</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6= Muhamad inadequate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Y</th>
<th>Adequacy of TBBx specimen for airway mucosa tissue</th>
<th>MUC_SPEC_ADEQ</th>
<th>N</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0= not specified</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1= No airway mucosal tissue present</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2= Blood or debris obscuring tissue</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3= Airway mucosal tissue adequate for diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4= airway tissue present according to path report</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5= Airway tissue absent according to path report</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Z</th>
<th>Neoplasia present or not</th>
<th>NEO_PR_AB</th>
<th>N</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0= absent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1= present</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8= inadequate for evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9= not specified</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AA</th>
<th>Type of Neoplasia</th>
<th>TYPE_NEO</th>
<th>N</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0= not specified</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1= benign tumor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2= carcinoma-in-situ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3= Malignant tumor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4= Indeterminate (includes suggestive of Ca)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Type of malignancy present:**

**Cytology**

<table>
<thead>
<tr>
<th>AB</th>
<th>Carcinoid</th>
<th>CARCINOID</th>
<th>N</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>Atypical carcinoid</td>
<td>ATY_CAR</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>AD</td>
<td>Small Cell Lung Cancer</td>
<td>SCLC</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>AE</td>
<td>Non Small Cell Lung Cancer</td>
<td>NSCLC</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>AF</td>
<td>Adenocarcinoma</td>
<td>ADENOCA</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>AG</td>
<td>Bronchoalveolar Ca</td>
<td>BRONCHO</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>AH</td>
<td>Large Cell Ca</td>
<td>LCC</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>AI</td>
<td>Squamous Cell Ca</td>
<td>SCC</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>AJ</td>
<td>Other type of Ca (includes immature myeloid cells, malignant lymphoproliferative disease, metastasis)</td>
<td>OTHER CA</td>
<td>N</td>
<td>1</td>
</tr>
</tbody>
</table>

**position coding**
<table>
<thead>
<tr>
<th>Position</th>
<th>Code</th>
<th>Description</th>
<th>Staining</th>
<th>0</th>
<th>1</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>AV</td>
<td>H&amp;E</td>
<td>H&amp;E</td>
<td>N</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>AW</td>
<td>Mucicarmine (for mucinous origin - adenocarcinoma)</td>
<td>MUCIN</td>
<td>N</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>AX</td>
<td>LCA</td>
<td>LCA</td>
<td>N</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>AY</td>
<td>CD20 ( for B cells)</td>
<td>CD20</td>
<td>N</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>AZ</td>
<td>CD43</td>
<td>CD43</td>
<td>N</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>BA</td>
<td>Synaptophysin for neuro endocrine)</td>
<td>SYNAP</td>
<td>N</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>BB</td>
<td>GMS (for fungi)</td>
<td>GMS</td>
<td>N</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>BC</td>
<td>S100 (stains mesothelial tissue eg: neuro endocrine or melanoma)</td>
<td>S100</td>
<td>N</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>BD</td>
<td>HMB 45 (for smooth muscle or melanoma)</td>
<td>HMB 45</td>
<td>N</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>BE</td>
<td>Cytokeratin 7 (stains epithelial cells (non-mucinous origins) eg: lung, ovary, breast, endometrium, may stain mesothelium)</td>
<td>CYTO 7</td>
<td>N</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>BF</td>
<td>Cytokeratin 20 (stains epithelial cell (non-mucinous origins) eg: Gl tract)</td>
<td>CYTO 20</td>
<td>N</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>BG</td>
<td>AFB</td>
<td>AFB</td>
<td>N</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
To evaluate the strength and feasibility of conducting this study, first of all an assessment is made of how many TBBx procedures are done at the Hershey Medical Center every year and how many samples are obtained. This assessment also helped identify if there were significant changes in practice during the time period of this study.

Demographics including age, gender distribution and performing bronchoscopist for the patient population included are established. Frequency and quality information available is evaluated to identify parameters for assessment of TBBx diagnostic yield.

Pre-procedure differential diagnosis is established from columns I-P in the codebook. Patients suspected to have lung cancer would have the number 1 in column I. The number 0 in column I indicates that lung cancer is not suspected and 9 indicates that the pre-procedure diagnosis of lung cancer is not specified. Data in column I are used to categorize patients by pre-procedure differential diagnosis mentioned in pathology report into three categories based on information provided in pathology reports:

i. Lung cancer suspected (I-1),

ii. Lung cancer not suspected (I-0), and

iii. Indeterminate (I-9)

If the pathology report mentions any differential diagnosis other than lung cancer (as indicated in columns J-P) then, that patient is categorized as “lung cancer not suspected” (1 in column I). If the pathology report mentioned lung
cancer as one of the various differential diagnoses, then the patient is categorized as “lung cancer suspected” (0 in column I). However, if no differential diagnosis is mentioned and/or a clinical finding is mentioned such as mass, cavity or infiltrate, then the patient is categorized as “indeterminate” (9 in columns I-P). An overview of population characterization is shown below.

POPULATION CHARACTERIZATION:

- Suspicion of having lung cancer in advance of bronchoscopy (Pre-test probability)
  - Not Suspected (0 in column I and 1 in appropriate column J-P) e.g. Sarcoidosis, ILD, Infection, T.B., metastasis (not lung cancer)
  - Indeterminate (9 in column I-P) e.g. Mass, cavity, infiltrate, GGO, or no differential diagnosis mentioned.
  - Suspected (1 in column I and appropriate number in columns J-P) e.g. Lung Cancer (with or without other differential diagnosis)
Data are organized as shown in figure 4, with the focus on presence or absence of malignancy and histopathological type of malignancy. All pathology reports billed as TBBx are included in the study. If the pathology report specifically reported inadequate sample resulting in no definitive histopathological diagnosis, it is documented as inadequate sample. However, if the pathology report mentions presence of mucosal tissue and any
histopathological findings in the tissue, they are documented as adequate samples and coded accordingly in columns Z to BI.

All data available are documented, analyzed and reported in the results section. Numerous shortcomings are identified in this aim of the study due its retrospective design:

i) Detailed clinical history may not be mentioned in the pathology reports and thus may not be feasible to use as an outcome measure.

ii) Similarly, smoking history may be infrequently mentioned thus not feasible to be used as an outcome measure. Correlation between histopathological type of malignancy and smoking may not be possible to evaluate.

iii) Detailed site location for tissue acquisition too may not be specified beyond the level of lobar bronchus and therefore may not serve as a feasible outcome measure. Therefore no correlation may be established between histopathological type of malignancy and central versus peripheral site of lesion.

iv) Due to the retrospective design of the study and billing record as the source of identifying TBBx, it is not possible to discern with certainty that parenchymal tissue is the intended tissue of biopsy. Therefore in cases where only mucosal tissue or airway tissue is seen and reported it is not possible to discern if the sample is sufficient or not. Samples are documented as insufficient only when specifically stated to be insufficient in the pathology report.

v) It is possible that tissue may be acquired from two different sites but processed and submitted to pathology in one jar. This may make it impractical to distinguish the pathological differences between the two sites biopsied. In such case, the first site mentioned in the pathology report is documented as the site biopsied.

vi) Special staining of tissue to identify receptors and markers may not be routinely done therefore this data too may be inadequate for final outcome evaluation.
Finally, patients whose TBBx samples are reported to be insufficient for evaluation are included in the final analysis. The rationale for this is that from a patients’ perspective it is a missed opportunity to diagnose the disease which leads to delay in diagnosis and inconvenience to the patient requiring another procedure for re-evaluation. A delayed diagnosis of cancer could result in progression of disease and shorter survival.

In spite of these shortcomings critical data are obtained to make it feasible to measure outcomes including average aggregate volume and its correlation with presence or absence of malignancy. Pre-procedure differential diagnosis, although not mentioned in all cases, is mentioned sufficiently to allow for pre-test probability assessment. Similarly, histopathological type of malignancy could be correlated with gender. However, correlations between smoking history and histopathological type of malignancy, correlation between location of lesion and histopathological type of malignancy etc could not be made. However, raw data is presented in the results.

Aggregate tissue volume (column V) is measured by the pathology technician at the time of sample processing. Multiple fragments of tissue floating in formalin are aggregated on filter paper using an eyedropper or fine toothed forceps. The tissues are aligned and then the aggregate is measured in 3 dimensions (XYZ), using a centimeter ruler with millimeter gradations (figure 5). The pathology report provides these three dimensions consistently. Aggregate tissue volume is calculated as the product of the XYZ dimensions and is used as an outcome variable (column V).
Since the focus of this project is on malignancy, statistical analysis is conducted on the average aggregate volume (column V) of tissue obtained from all samples reported positive for malignancy in the pathology report (1 in column Z), compared with the average aggregate volume of samples reported negative for malignancy (0 in column Z). The p-value is calculated using the Student’s t-test, unpaired, two-way variance with a p-value of 0.05 specified as significant.

The gender (column E) distribution by histopathological type of malignancy (columns AB-AJ) is also established. A pre-test probability of malignancy is assessed by comparing the pre-procedure differential diagnosis (column I) with TBBx pathology report (column Z). Final outcome measures established are:

- Pre-procedure suspicion of lung cancer (column I)
- Pathology report positive for malignancy (column Z)
- Smoking History (column H)
- Specimen Site Biopsied (column S and T)
- Average aggregate volume of tissue (column V)
• Statistical analysis of average aggregate tissue volume differences between samples positive for malignancy and those negative for malignancy.

- Histopathological type of malignancy (column AB-AJ)
- Gender distribution by histopathological type of malignancy (sorting, correlating and analyzing columns E with columns AB - AJ)
- Estimate the pre-test probability of a positive pathology report for malignancy among samples obtained from patients clinically suspected to have lung cancer. (sorting, correlating and analyzing data in columns I with columns Z)
- Cellular atypia (column AK)
- Non-malignant findings (columns AL-AU)
- Staining (columns AV-BI)

**Specific Aim 3:** To identify and integrate matched pairs of TBBx and lung resections in order to validate the performance of TBBx against lung resection as the gold standard.

"Natural Language Ilia search", a software program available in the HMC Department of Pathology is used to identify and acquire pathology reports of all lung resection surgeries done at the Hershey Medical Center between 2\textsuperscript{nd} June 2000 and 8\textsuperscript{th} December 2006 (6.5 years). Results of this search were acquired from Saied Saffee M.D, fellow in the Division of Pulmonary Allergy Critical Care Medicine at the Hershey Medical Center.

A dataset is created on Excel spreadsheet from Microsoft office program that includes patient medical record number (MRN) and patient name of all patients who underwent lung resection surgery during this period of time. These data elements are used to identify patients common to both datasets; lung resection repository and TBBx billing records. These patients are selected for this validation study of TBBx reports. A codebook is developed (table 5) similar to the
one used for TBBx to document information. Variables A – W in table 5 are developed based on information acquired from the lung resection pathology reports.

AJCC lung cancer classification derived protocol used by the pathology department to report lung cancer findings is obtained by personal correspondence with Henry S. Crist, M.D. Assistant Professor, Division of Anatomic Pathology, Medical Director, Surgical Pathology, HMC and utilized to develop variables M-W in table 5. Personal discussion with cardiothoracic surgeon Michael J. Lazar, M.D., Assistant Professor of Surgery, HMC, provided the classification of different types of surgeries conducted and guided the development of variable L in the codebook (table 5).

Table 5: The codebook developed for documentation and analyzes of data obtained from lung resection pathology reports.

<table>
<thead>
<tr>
<th>Sr.No</th>
<th>Variable</th>
<th>Field Name</th>
<th>Field Type</th>
<th>Field Width</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Medical Record Number</td>
<td>MRN</td>
<td>N</td>
<td>11</td>
</tr>
<tr>
<td>B</td>
<td>Name</td>
<td>NAME</td>
<td>Text</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>C</td>
<td>Research Case # (IRB#.SR.NO)</td>
<td>RES_CASENUM</td>
<td>N</td>
<td>10</td>
</tr>
<tr>
<td>D</td>
<td>Accesion Number</td>
<td>ACC_NUM</td>
<td>N</td>
<td>9</td>
</tr>
<tr>
<td>E</td>
<td>Date of Service</td>
<td>ACC_DATE</td>
<td>D</td>
<td>mm/dd/yyyy y</td>
</tr>
<tr>
<td>F</td>
<td>Gender</td>
<td>SEX</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1= Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2= Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Blank= Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>Age</td>
<td>AGE</td>
<td>N</td>
<td>3</td>
</tr>
<tr>
<td>H</td>
<td>Pathologist</td>
<td>PATH_IST</td>
<td>L</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>AAE= Ashraf Abou-Elella</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ABA= Arthur B Abt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CP= Cunfeng Pu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CSA= Catherine S. Abendroth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DSZ= Dani S. Zander</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EEF= Elizabeth E. Frauenhoffer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FMR= Francesca M. Ruggiero</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HSC= Henry S. Crist</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LAF= Llewellyn A. Foulke</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MGB= Michael G. Bayeri</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Field</td>
<td>Code</td>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>------</td>
<td>------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lung Specimen Site</td>
<td>I</td>
<td>SIDE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>LOBE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>SEGMENT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>Operation Performed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>Neoplasia present or not</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SIDE**
- 0 = not provided
- 1 = Right
- 2 = Left
- 3 = Trachea and primary carina
- 4 = other
- 5 = not specified

**LOBE**
- 0 = not provided
- 1 = upper lobe
- 2 = middle lobe
- 3 = lower lobe
- 4 = lingula

**SEGMENT**
- 00 = not provided
- 01 = apical
- 02 = posterior
- 03 = anterior
- 04 = superior
- 05 = inferior
- 06 = lateral
- 07 = medial
- 08 = anterior basal
- 09 = lateral basal
- 10 = posterior basal
- 11 = anteromedial basal
- 12 = apicoposterior

**Operation Performed**
- 0 = Not specified
- 1 = Pneumonectomy
- 2 = Single Lobectomy
- 3 = Multiple Lobectomy
- 4 = Lobectomy (detail not provided)
- 5 = Wedge Resection
- 6 = Biopsy
- 7 = others

**Neoplasia present or not**
- 0 = absent
- 1 = present
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8= inadequate for evaluation</td>
<td></td>
</tr>
<tr>
<td>9= not specified</td>
<td></td>
</tr>
<tr>
<td>N 14</td>
<td><strong>Type of Neoplasia</strong></td>
</tr>
<tr>
<td>0= not specified</td>
<td></td>
</tr>
<tr>
<td>1= benign tumor</td>
<td></td>
</tr>
<tr>
<td>2= carcinoma-in-situ</td>
<td></td>
</tr>
<tr>
<td>3= Malignant tumor</td>
<td></td>
</tr>
<tr>
<td>4= Indeterminate (includes suggestive of Ca)</td>
<td></td>
</tr>
</tbody>
</table>

**Type of malignancy present:**

**Cytology**

| O 15 | Carcinoid | CARCINOID | N 1 |
| P 16 | Atypical carcinoid | ATY_CAR | N 1 |
| Q 17 | Small Cell Lung Cancer | SCLC | N 1 |
| R 18 | Non Small Cell Lung Cancer | NSCLC | N 1 |
| S 19 | Adenocarcinoma | ADENOCA | N 1 |
| T 20 | Bronchoalveolar Ca | BRONCHO | N 1 |
| U 21 | Large Cell Ca | LCC | N 1 |
| V 22 | Squamous Cell Ca | SCC | N 1 |
| W 23 | Other type of Ca (includes immature myeloid cells, malignant lymphoproliferative disease, metastasis) | OTHER CA | N 1 |

**position coding**

| 0= absent |   |
| 1= present, differentiation not specified |   |
| 2= present, well differentiated |   |
| 3= present, poorly differentiated |   |
| 9= not specified |   |

Inclusion criterion is availability of matched data pairs for each patient common to both datasets with a pre-operative TBBx and lung resection pathology reports. If a patient had two or more TBBxs, each followed by a lung resection surgery, each pair is included as a separate pair and analyzed. An algorithm of data organization is shown in figure 6.

Exclusion criteria are:

- Patients whose TBBx is misclassified as lung resection, identified by reading the lung resection pathology report and comparing with TBBx pathology report.
- Patients whose TBBx or lung resection pathology reports are not available.
- Patients whose TBBx is done post-resection.
Patients whose TBBx samples are reported to be insufficient for evaluation. These patients are excluded only from the final categorization of TBBx yield. However, their TBBx lung resection pathology reports are reviewed and results presented, since insufficient tissue at TBBx contributes to delay in diagnosis of lung cancer, thus is part of overall evaluation of process of care.

Patients who have had more than one pre-resection TBBx, the most recent TBBx report is included, the others are excluded.

Figure 6: Algorithm showing the integration of TBBx and lung resection datasets. +/+ denotes the group of patients with TBBx and resection reports positive for malignancy, +/- denotes the group of patients that are positive for malignancy on TBBx but negative on lung resection, -/+ denotes the group of patients negative for malignancy on TBBx but positive on lung resection, and -/- denotes the group of patients that are negative for malignancy on TBBx and lung resection.
In order to determine the presence or absence of cancer in the pathology report, all TBBx reports are stratified into two groups, one TBBx report positive for cancer (3 in column AA of table 4), and second TBBx report negative for cancer (0 in column Z of table 4). All lung resection reports are similarly categorized into two groups, one lung resection pathology report positive for cancer (3 in column N of table 5), and second lung resection pathology report negative for cancer (0 in column M of table 5). Each patient report is then identified and categorized into one of 4 groups as shown in figure 6 and table 6:

i) First, patients whose TBBx and lung resection report is positive for malignancy (True Positive, aka +/+).

ii) Second, patients whose TBBx report and lung resection report is negative for malignancy (True Negative, aka -/-).

iii) Third, patients whose TBBx report is negative for malignancy, but lung resection report is positive for malignancy (False Negative, aka -/+).

iv) Finally, patients whose TBBx report is positive for malignancy but resection report is negative for malignancy (False Positive, aka +/-).

Table 6: Table to show algorithm used to assess sensitivity and specificity of diagnosing lung cancer by TBBx procedure.

<table>
<thead>
<tr>
<th>Surgery</th>
<th>Neoplasia Positive</th>
<th>Neoplasia Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBBx</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neoplasia Positive</td>
<td>True Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td>Neoplasia Negative</td>
<td>False Negative</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

These data provide an assessment of the bronchoscopy acquired TBBx yield at the Hershey Medical Center, and are used to estimate frequency of false positive, false negative, sensitivity and specificity of diagnosing lung cancer by TBBx procedure.
Test Performance Characteristics:
Sensitivity = True positive / (True positive + False negative)
Specificity = True negative / (False positive + True negative)
False Negative Rate (FNR) = False Negative / (True positive + False negative)
False Positive Rate (FPR) = False positive / (False positive + True negative)
Predictive Value Positive (PVP) = True positive / (True positive + False positive)
Predictive Value Negative (PVN) = True negative / (False negative + True negative)

Next, a correlation is established between the histopathological type of malignancy reported by TBBx and lung resection pathology reports. This is done as an attempt to evaluate the quality of TBBx pathology reports since it is based on the small volume of sample acquired, while the lung resection procedure provides sufficient tissue to report occurrence of mixed tumor, rules out edge effects, crush effects and other artifacts that may compromise the quality of TBBx reports. An assessment correlating histopathological type of malignancy in each of the four groups defined above is done as part of aim 5 which aims to identify lesions characteristics accurately and inaccurately diagnosed by TBBX.

The limitations of this part of the study are:

i) Since only the most recent pre-operative TBBx is included, it is possible that a patient had multiple TBBxs and only after the most recent TBBx was reported to be positive for malignancy was is followed by surgery. Therefore not taking into account the delay in diagnosis and the previous other false negative biopsies. Thus data may be skewed towards a higher “true positive” outcome than in reality. This limitation may be overcome by reanalyzing the TBBx data for individual patients, or having access to patient chart which would indicate the number of TBBx conducted before a diagnosis of lung cancer was made.
ii) During a bronchoscopy procedure, TBBx is only one of the many methods to acquire tissue/samples for evaluation. Broncho-alveolar lavage, lymph node biopsy, cytology from bronchial brushing and washing are other methods to diagnose lung cancer during a bronchoscopy procedure. True sensitivity and specificity of diagnosing lung cancer by bronchoscopy can be assessed by taking into account all these factors. However the current project is only assessing the contribution of the TBBx procedure in diagnosing lung cancer for reasons stated in the introduction.

iii) In addition, bronchoscopically acquired TBBx is only one of the many methods used clinically to diagnose lung cancer. A high index of suspicion based on clinical presentation, rapidly growing pulmonary lesion as seen on X-ray or CT scan, biopsy from metastatic site positive for malignancy, may all lead to a clinical diagnosis of lung cancer even if the TBBx report is negative for cancer. To assess the efficiency of diagnosing cancer clinically would require all these factors to be included. However, that is not the goal of this study. This study focuses on the efficiency of TBBx procedure in diagnosing lung cancer.

iv) If a TBBx is reported positive for malignancy and the patient is diagnosed to have advanced stage disease (stage IIIb, IV), this patient is not a candidate for surgery, and therefore that patients data will not be available on the lung resection repository. Therefore the population included in this part of the study is skewed towards including earlier stage patients (stages I, II and IIIa). The design of this study gives an assessment of the performance of TBBx in diagnosing pulmonary malignancies, it is not an assessment of how efficiently pulmonary malignancies are diagnosed at the Hershey Medical Center.

v) Patients whose TBBx samples are reported to be insufficient for evaluation cannot be included for lung resection report analysis. However, the lung resection histopathology reports of these patients are mentioned separately. The rationale for this is that from a patients' perspective the TBBx insufficient
sample report is a missed opportunity to diagnose the disease which leads to delay in diagnosis and inconvenience to the patient requiring another procedure for re-evaluation. A diagnosis of cancer later could imply progression of disease and shorter survival. Thus a gap in the quality and process of care.

vi) Finally, patients whose TBBx samples are reported to be indeterminate, usually implying that some amount of cellular atypia is seen and is suggestive of malignancy but cannot be definitively diagnosed, are usually clinically considered negative unless other clinical criteria lead to a diagnosis of malignancy or require a repeat bronchoscopy and TBBx for definitive diagnosis. Therefore, in this study indeterminate TBBx pathology reports are categorized as negative for statistical analysis. Each of these patients’ lung resection pathology reports is then individually mentioned.

Specific Aim 4: To benchmark TBBx pathology reports as a function of digitally acquired computerized tomography (CT) lesions using a novel classification scheme and identify characteristics of lesions commonly biopsied and lesions frequently reported to be positive for malignancy.

A retrospective cohort study is conducted using the digital CT scans IDX/GE electronic image repository. This data source is used to retrieve digital images of pulmonary lesions biopsied for patients included in the TBBx dataset. Patients for whom a digital image of their CT scan is available in the repository, a single transverse slice of each CT scan image centered on the region of interest (ROI) is extracted using Snagit® (software available in the Bascom laboratory) and saved in a Microsoft Word document. A few examples of images acquired are seen in figure 7.

ROIs are coded by a physician using a simple yet novel image classification scheme. Lesions are first classified as focal or diffuse depending on the margin characteristics. Lesions with well defined margins are classified as
focal and those with ill defined or blurry margins are classified as diffuse. Focal ROIs are further classified into four categories depending on size (figure 7).

i. The length of the short axis (SA) is <1cm.
ii. SA 1-3 cms (clinically T1 tumor).
iii. SA >3cms (clinically T2 tumor).
iv. Multifocal ROI (clinically T4 tumor).

Diffuse ROIs are classified into three categories (figure 7) by extent of lobe involved into:

i. Less than 25% lobe involved.
ii. 25-50% lobe involved
iii. Greater than 50% lobe involved.
Figure 7: Classification of CT scan digital images. A single slice of each CT scan centered at the region of interest (ROI) served as a representative sample. Lesions were classified as focal and diffuse, each category was further classified into 3 subcategories depending on size of ROI.

If a patient had multiple sites biopsied and each site was submitted in a separate jar for pathological evaluation, then each ROI image is saved separately and coded separately. If multiple sites were biopsied and submitted in one jar, then the site included in the TBBx reported is documented.

Exclusion criteria are:
- Digital CT scan image not available in electronic repository.
- TBBx pathology report not available or sample insufficient for evaluation.
If multiple digital CT scans were done and available in electronic repository, the most recent pre-TBBx CT scan image is included, others are excluded.

If no pre-TBBx digital CT scan is available, but post-TBBx digital CT scan images are available, the image obtained from the first post-TBBx CT scan is included.

TBBx pathology reports were stratified by CT lesion category. Lesion type and size is correlated with TBBx pathology report. Pathology results are benchmarked as a function of chest computerized tomography (CT) lesions. A correlation is established between CT scan lesion category and TBBx pathology report with respect to malignancy.

Outcome measures evaluated are:

- Availability of digital CT scans in the HMC electronic repository with time.
- Establish most commonly biopsied lesion characteristic within this dataset.
- Establish characteristic of lesion most commonly reported to be malignant by TBBx within this dataset.

An assessment of lesion characteristics of true positive, true negative, false positive and false negative categories is done in aim 5.

The shortcomings of this part of the study are:

i) IDX/GE digital CT scan repository is a data source recently established at the Hershey Medical Center, beginning December 2002, therefore many patient images may not be available from this source.

ii) Often patient CT scans are done elsewhere and patients are referred to the Hershey Medical Center for further evaluation. Digital images of these patients too may not be available in the IDX/GE data source.
iii) If a recent pre or post-TBBx CT scan image is not available in IDX/GE, but a later image is available, it was included for analysis. This may cause some inaccuracy in correlation of lesion type and pathology report. However, based on clinician derived information, biopsies are rarely conducted for acute, rapidly changing pulmonary conditions. Usually biopsy is done for lesions that are relatively slow changing. Therefore, it is presumed that the CT scan image is a good representation of the lesion biopsied.

**Specific Aim 5:** Assessment of pulmonary lesion characteristics accurately and inaccurately diagnosed to be malignant by TBBx.

In order to assess pulmonary lesion characteristics accurately and inaccurately identified by TBBx, a detailed analysis of data acquired for the patients common to the TBBx and lung resection datasets is conducted. The TBBx samples reported to be inadequate for evaluation, but lung resection data of these patients if available is reported separately. Data of the remaining patients for who matched pairs of TBBx and lung resection pathology reports is available and are categorized by the four groups created in aim 3 are analyzed for each group:

i. True Positive
ii. True Negative
iii. False Negative
iv. False Positive

Data are analyzed to assess characteristics of pulmonary lesions that are accurately diagnosed to be malignant (true positive) by TBBx and conversely, characteristics of pulmonary lesions usually missed or inaccurately diagnosed to be malignant (false negative) by TBBx. The outcome measures established for these patients and compared for each of the four groups separately are:

- Location of lesion biopsied and resected
- Congruency of site biopsied and site resected
- Histopathological type of malignancy reported by TBBx and lung resection.

Finally, in order to assess CT scan lesion characteristics among the four groups of patients, patient name and MRN are used as linking criteria to identify patients common to all three datasets TBBx, lung resection and digital CT scan. Inclusion criterion is availability of 3 data elements (N=B in figure 8):
- TBBx pathology report (as included in aim 2, N=X in fig8)
- Lung resection pathology reports (as included in aim 3, N=Y in fig8) and
- Digital CT scan image (as included in aim 4, N=Z in fig8).

Figure 8: Algorithm for correlation between CT scan, TBBx and lung resection datasets.

Data acquired thus far from TBBx pathology reports, lung resection pathology reports and digital CT scans are integrated. Data from patients common to all three datasets are identified (N=B in figure 8). The final outcome measure assessed is:
- CT scan lesion characteristic for each of the four groups of patients.

The limitations of this part of the study include all limitations mentioned earlier for each dataset included in the study thus far. In addition, congruency of
ROI seen on CT scan, site biopsied during TBBx and site resected is determined with TBBx data as centre of reference. Therefore, if a patient has a multifocal lesion seen on CT scan it is theoretically possible that the patient had one site biopsied and another resected. However, it is highly unlikely to occur if the patient has lung cancer, because a patient with lung cancer and multi-focal lesions on TBBx is in an advanced stage of the disease and thus not a candidate for surgery, therefore unlikely to be included in this study. Such an occurrence is more likely in non-malignant cases involving diffuse pulmonary pathology. Since that is not the focus of this study, incongruency due to this limitation is unlikely to occur.

**Specific Aim 6:** To estimate the time intervals between CT scan, TBBx and lung resection in order to identify time gaps in the current process of care for lung cancer patients at the Hershey Medical Center.

**Availability:** All patients included in aim 2 (TBBX dataset N=X in figure 8) serve as the baseline. Time intervals for TBBx are stratified by date into 7 categories:

1. All TBBx done between 08/31/1999 and 12/31/1999 are categorized as 1999. Digital CT scans and lung resections done on these patients are included in this category even if the date when they are done is outside this time period.
2. TBBx done between 01/01/2000 and 12/31/2000 are categorized as 2000. Digital CT scan and lung resection data for these patients is included in this category.
3. TBBx done between 01/01/2001 and 12/31/2001 are categorized as 2001.
4. TBBx done between 01/01/2002 and 12/31/2002 are categorized as 2002.
5. TBBx done between 01/01/2003 and 12/31/2003 are categorized as 2003.
6. TBBx done between 01/01/2004 and 12/31/2004 are categorized as 2004.
7. TBBx done between 01/01/2005 and 06/30/2005 are categorized as 2005.
Of note here is that the first and last time intervals are only 4 and 6 months in duration respectively, while all others are one calendar year.

For these patients, availability of CT scan (N=Z in fig 8) and lung resection data (N=A in figure 8) are quantified for each time interval. For example for all TBBxs done in 2000, how many matched CT scans and lung resection pathology reports are available, irrespective of the year in which the CT scan and resection is done.

Dates of service (DOS) for TBBx, lung resection and digital CT scan included in the above parts of the study are documented. Availability of CT scan and lung resection reports is re-analyzed depending on the DOS. This analysis illustrates the change in availability of data with time.

To assess time intervals in the process of care, two parameters are established:

- **Time to Diagnosis (TDx):** Time interval between DOS for CT scan and TBBx.
- **Time to Treatment (TRx):** Time interval between DOS for TBBx and lung resection surgery.

**Time to diagnosis (TDx):** The DOS for TBBx as the central point of reference, CT scans are classified into three categories: pre-TBBx, same day as TBBx and post-TBBx. Pre and post-TBBx CT scans are further classified into 7 categories each:

i. One to 2 days,
ii. 3-7 days (1 week),
iii. 8-14 days (2 weeks),
iv. 15-30 days (1 month),
v. 31-90 days (3 months),
vi. 91-180 days (6 months), and
vii. Greater than 180 days (> 6 months).

To assess time to diagnosis the post TBBx CT scans are excluded. The average TDx is established from these data.

**Time to treatment (TRx):** Lung resections included in this study are all post-TBBx and are classified by time interval between TBBx and surgery into 8 categories:

i. Same day,

ii. 1-2 days,

iii. 3-7 days (1 week),

iv. 8-14 days (2 weeks),

v. 15-30 days (1 month),

vi. 31-90 days (3 months),

vii. 91-180 days (6 months), and

viii. Greater than 180 days (> 6 months).

The average TRx is established from these data.

An independent time interval analysis (TDx and TRx) is done for patients common to all 3 datasets (N=B in figure 8). These patients are identified by the categories of true and false positive and negative and an analysis of time intervals (TDx and TRX) is done for each category. Statistical analysis is done to identify any significant difference in TDx and TRx between the groups. (Student’s t-test, unpaired with two way variance, p=0.05 defined as significant)

Outcome measures:

- Availability of TBBx pathology reports from Power Chart.
- Availability of CT scan images with time.
- Availability of lung resection pathology reports with time.
- Time interval between CT scan and TBBx (Time to diagnosis, TDx).
Time interval between TBBx and lung resection (Time to treatment, TRx).

Establish TDx and TRx for each of the four groups of patients previously defined: (true positive, true negative, false positive, false negative)

The limitations of this part of the study include all limitations mentioned earlier in each specific aim of the study thus far, but some require special mention here:

i) The TBBx dataset is acquired for a time period from 08/31/1999 to 06/30/2005, while the lung resection dataset is acquired for a time period from 06/02/2000 to 12/08/2006. Therefore the overlap in time periods in these two datasets is only from 06/02/2000 to 06/30/2005, missing the first six months of TBBx and last year and half of lung resection patients. Due to the differences in the time intervals of each dataset included in this study, patients at the ends of the time intervals are unlikely to be common in all three datasets. This may reduce the total number of patients in this part of the study, thus data available for analysis and the power of conducting statistical analyses.

ii) Since digital CT scan repository is a recently developed resource at the Hershey Medical Center, since December 2002, many patients especially in the earlier part of the study may not have a digital CT scan image in this dataset. This factor may contribute significantly to the final population included in this part of the study and thus reduce the power of statistical analyses.

iii) It is possible that a patient may have had a pre-TBBx CT scan done outside the Hershey Medical Center and then had a TBBx done here followed by a post-TBBx CT scan. This may increase the post-TBBx CT scan category of patients. However, since TBBX is usually done for stable or slowly enlarging lesions, it is assumed that the findings of the post-TBBx CT scan are not significantly different from the pre-TBBx lesion. A post-TBBx CT scan may also be done to document any complications resulting from the procedure. This too is
a rare occurrence and is considered to not contribute significantly to the final data.

iv) Similarly it is possible that a patient with lung cancer may have had a TBBx with a positive pathology report done here at the Hershey Medical Center, but chose to go elsewhere for lung resection and further management. This limitation may reduce the true positive category of patients.

v) Since this study is retrospective, there may be some instances when multiple sites are biopsied during TBBx, but sent to pathology in one jar. If one of the many samples is reported positive for cancer, it is not possible to determine exactly from where that sample is taken. This limitation may affect the true and false positive categorization of data. However, since such occurrences are rare it is presumed that it did not significantly affect the final data.
IV: Results

Specific Aim 1: To define the current clinical process of care and to identify clinical datasets used for key elements of lung cancer diagnosis.

As per the guidelines suggested by Rubin et al. (Rubin, Pronovost et al. 2001; Rubin, Pronovost et al. 2001) quality and process of care are assessed;

i) The audience of this project is anticipated to be: lung cancer patients, pulmonologists, radiologists, pathologists, oncologists and others physicians involved in the care of lung cancer patients. Also, those interested in the quality of care assessment at the Hershey Medical Center. Finally, basic science researchers intending to study molecular markers for diagnostic, prognostic or other conditions relevant to utilizing histological samples from primary lung tumor tissue.

ii) The clinical area being evaluated is TBBx and its efficiency in diagnosing lung cancer. This procedure is being evaluated as a potential area where increased efficiency in diagnosing lung cancer may increase population diagnosed at an early stage thus improving long-term survival.

iii) The assessment team is primarily composed of the author and her thesis advisor, Rebecca Bascom MD. However, critical feedback, direction and advice from numerous physicians including pulmonologists, pathologists, radiologists, cardiothoracic surgeons, and oncologists guided the development and design of this project.

iv) The component of the TBBx process being measured is primarily the pre-procedure differential diagnosis of cancer and the final pathology report with regard to presence and absence of malignancy. However, this procedure is also being evaluated for feasibility to serve as a potential platform to integrate the routine use of biomarkers in the diagnosis of lung cancer. For this purpose and to assess the process of care for lung cancer patients, the aggregate volume of TBBx tissue, the frequency of insufficient tissue for evaluation, time intervals between CT scan, TBBx and lung resection and many other parameters are
being assessed. These are used as indicators to highlight the strengths and weakness in the current practice.

v) Preliminary tests and,

vi) Scoring and analytic specifications are mentioned in the specific aims that address them.

From participating with clinicians including bronchoscopists, oncologists, radiologists and other providers at the Hershey Medical Center an understanding of the process of care for patients with lung cancer is obtained. Initially we identified datasets where patient information is stored and got an estimate on the accessibility and quality of information available in each dataset.

Datasets: Patient information is usually stored in various databases for numerous purposes including, billing, research, surveillance, mandatory government reporting of certain conditions etc. Depending on the purpose for which individual databases are set up, the amount and details of information may differ. Here we provide some preliminary information on the databases where lung cancer patient information is stored.

Locally at the Hershey Medical Center: billing department, medical records department, Power Chart (Cerner), digital CT scan image repository, “Natural Language IIa search” a software available in the Department of Pathology to access pathology reports for specific specimens, CORI, Cardioaccess, Penn State Hershey Cancer Institute (PSHCl) Tissue Bank, and Oncore, are some of the databases in which information on lung cancer patients is stored.

Medical Records Department contains an exhaustive amount of all detailed information on all patients. However, access to patients’ files is strictly protected by privacy laws and regulated by the Institutional Review Board (IRB). However, because patient information is available as free text and without
specific structure, it was considered inefficient to utilize for this project. CORI is an electronic database developed by the pulmonary department which has information on all bronchoscopies done within the department. This database is recent and therefore currently does not have data of patients who underwent bronchoscopy more than 5 years ago. This database is not selective for lung cancer patients but includes information of all patients who underwent bronchoscopy. Cardioaccess is a similar electronic database developed by the cardio thoracic surgery department and has information on all patients that underwent any cardiothoracic surgery. PSHCI Tissue Bank has a rich reservoir of tissues obtained from various procedures and stored for research purposes. There is limited patient information available from this resource however. Oncore is among the most recent databases, still under development at the Hershey Medical Center and it includes all research clinical trials being conducted at the Hershey Medical Center.

The datasets finally used to acquire data for this study consist of:

i. The billing department database includes an exhaustive list of information on all patients seen and procedures conducted at the Hershey Medical Center. This database proved useful in identifying all transbronchial biopsies done at the Hershey Medical Center. It does not have detailed clinical data available. Therefore another dataset needed to be identified for that purpose.

ii. Power Chart (Cerner) is an electronic database with information on all patients undergoing care at the Hershey Medical Center. This resource too is strictly protected by confidentiality laws and IRB regulations. Although extensive patient information is available through this resource it appeared to be inconsistent in its format. For example some patient seen as out patient would have essential information in their outpatient letters; others who are admitted in hospital would have their information in a separate section. This made acquiring structured information cumbersome. However, TBBx pathology reports are consistently available in the same section for all patients and proved to be a
reliable source to acquire this information. Therefore Power Chart (Cerner) is utilized as a resource to acquire TBBx pathology reports for this study.

iii. Digital CT scan repository is a dataset where images of lesions under investigation can be acquired. Although this is a recent resource developed at the Hershey Medical Center available only since December 2002, it is considered useful to utilize in this study. Images of pulmonary lesions biopsied are obtained from this resource to assess the characteristic of malignant lesions accurately and inaccurately diagnosed by TBBx.

iv. “Natural Language IIa Search” a software available in the Department of Pathology is utilized to identify all patients who underwent lung resection between June 2000 and December 2006 at the Hershey Medical Center. This dataset is used to identify patients common to the TBBx dataset and lung resection pathology reports are used as gold standard to assess TBBx yield in diagnosing lung cancer.

v. Finally, AJCC guided lung cancer classification is used to develop the epidemiological codebooks to organize data collected from TBBx and lung resection.

Besides local resources, other datasets that may be used to acquire lung cancer patient information include:

State: Tumor Registry
National: NIH Tissue Banks, SEER, SPORE
International: WHO, IARC

These resources are not used in this study, but may be feasible for future studies.

A simplistic overview of the process of care of patients with lung cancer is shown below and schematically in figure 9.

1. Preclinical Detection (screening)
2. Clinical Presentation: Symptom of Disease
   - Disease Suspected
   - Diagnostic Studies Done:
Radiology
i. Chest X-ray
ii. Chest Computerized Tomography (CT) scan
iii. Positron Emission Tomography (PET)
iv. Magnetic Resonance Imaging (MRI)
Pathology
i. Bronchoscopy: brush, wash, forceps biopsy, TBBx, TBNA, BAL
ii. Trans-thoracic needle biopsy
iii. Biopsy from metastatic lesion, if any

Diagnosis and Staging
- Bronchoscopy: brush, wash, forceps biopsy, TBBx, TBNA, BAL
- Tran-thoracic needle biopsy
- Biopsy from metastatic lesion, if any
- Surgery: wedge biopsy, pneumonectomy, lobectomy, resection, other

Treatment
- Surgery: pneumonectomy, lobectomy, wedge resection, other
- Chemotherapy: cisplatin, carboplatin, paclitaxel, docetaxel, cetuximab, bevacizumab, gefitinib, erlotinib, vandetanib, vorinostat
- Radiation Therapy:
- Combination Treatment

3. Follow-up:
- Rehabilitation / Quality of Life
- Monitoring for Recurrence
  - Same as step 1 onwards
Figure 9: Algorithm showing elements of process of care, in the diagnosis and management of lung cancer. Variety of tissue may be available from the same patient at different stages of the process. BAL = Bronchoalveolar Lavage, TBNA = transbronchial needle aspiration, TBBx = transbronchial biopsy.

With the long term goal of this project in mind, it is identified that tissue for molecular markers may be acquired at various stages during process of care and that a variety of tissues may be acquired from the same patient. For example as shown in figure 9, tissue may be obtained from the same patient at the stage of diagnosis during procedures like bronchoscopy, tissue may again be acquired from same patient during staging and during surgical stage of treatment. Tissue
is obtained during diagnosis and staging via numerous procedures including trans-thoracic needle aspiration for peripheral pulmonary tissue. CT-guided biopsy, surgical wedge resections and other surgical specimens are some more methods in which tissue is acquired as routine process of care for patients with lung cancer and can serve as useful resources to assess the potential of molecular bio-markers.

Tissue obtained at first clinical encounter can serve as baseline tissue for molecular marker assays. These markers can serve as both diagnostic and prognostic markers. When tissue is acquired later during the process of care, any difference in the molecular markers can be useful to assay for response to treatment. Thus serial tissue acquired from same patient has the potential to individualize that specific patients’ line of treatment.

Besides availability of tissue from same patient during different stages of process of care, different types of tissue may be acquired from same patient for molecular marker assays, during the same procedure and can serve for initial feasibility and comparison studies. For example, during bronchoscopy at initial diagnosis tissue may be acquired via numerous procedures including bronchial brushing that gives proximal airway tissue for cytological evaluation, bronchial washes provide proximal airway mucosal tissue for fluid cytology, bronchoalveolar lavage provides distal alveolar airway tissue for fluid cytology, transbronchial needle aspiration provides lymph node tissue for cytology and histology, forceps biopsy provides airway mucosal and tracheobronchial mass tissue for histological evaluation, and finally transbronchial biopsy provides pulmonary parenchymal tissue for histological evaluation. Figure 10 gives an over view of various methods and types of tissues that can be acquired during bronchoscopy alone. TBBx provides intact primary tissue for histology and hence is the focus of this project in aim 2. Molecular marker use at this stage has tremendous potential to contribute to diagnostic, prognostic and development of individualized care for lung cancer patients.
Once the outline of process of care for lung cancer patient is defined, it guided the development of the remaining specific aims of this project. Thus, TBBx is identified as a critical step during the process when an initial histological confirmation of lung cancer is made and is chosen to be the focus of this study. Surgical lung resection is chosen to serve as the gold standard to assess TBBx diagnostic yield for lung cancer to identify true positive and false negative cases. Finally, digital CT scan images are chosen to identify pulmonary lesion characteristics that are accurately or inaccurately identified by TBBx. The outcomes of this study will identify gaps in the current process of care and guide future direction of research to continue finding methods to identify patients with lung cancer at an earlier stage, develop more efficient diagnostic and staging techniques, and better treatment regimens for individual patients to improve long term survival.
Specific Aim 2: To critically evaluate a retrospective cohort of bronchoscopically acquired TBBxs and the pre-test probability of diagnosing lung cancer.

Eight hundred and one transbronchial biopsy samples obtained from 653 procedures comprised the dataset for assessment. Figure 11 shows the distribution of number of TBBxs done and samples acquired each year for the duration of this study. Since the first and last few months of the study did not comprise a whole year, the corresponding bars in figure 11 give an apparent indication of low patient volume. To assess this apparent decrease, average monthly TBBxs and samples obtained for each year are calculated as shown in figure 12. It is this clear that the TBBx patient volume remains consistent in the last part of the study period. Overall the bronchoscopy practice has consistently increased over the period of this study.

Figure 11: Number of procedures done and samples acquired per year by bronchoscopic TBBx between August 1999 and June 2005 at the Hershey Medical Center. The first and last bars in figure show data for 6 months of clinical activity, while the middle 5 bars each represent one year data.
Patients’ age averaged 58yrs 2mths +/- 15yrs 2mths (mean +/- SD) and 360/653 (55.13%) are male. Procedures are performed by 10 different attending bronchoscopists (range 1-455 patients), demonstrating that within the pulmonary division itself this is a multi-individual practice. Clinical history is not mentioned in many clinical pathology reports and therefore could not be used as an outcome parameter. However, pre-procedure suspected diagnosis is mentioned and used as an outcome parameter. Table 7 gives an overview of the pre-procedure differential diagnosis. No differential diagnosis was specified in the pathology reports for about 60% of the procedures. Lung Cancer was suspected pre-procedure in 122 (18.7%) cases.
Table 7: An overview of the pre-procedure differential diagnosis mentioned in the TBBx pathology reports.

<table>
<thead>
<tr>
<th>DIFFERENTIAL DIAGNOSIS:</th>
<th>SUSPECTED</th>
<th>NOT SUSPECTED</th>
<th>NOT SPECIFIED</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUNG CANCER</td>
<td>122 (18.7%)</td>
<td>110 (16.9%)</td>
<td>421 (64.5%)</td>
<td>653 (100%)</td>
</tr>
<tr>
<td>METASTASIS</td>
<td>20</td>
<td>217</td>
<td>416</td>
<td>653</td>
</tr>
<tr>
<td>LYMPHOMA</td>
<td>13</td>
<td>226</td>
<td>414</td>
<td>653</td>
</tr>
<tr>
<td>SARCOIDOSIS</td>
<td>51</td>
<td>190</td>
<td>412</td>
<td>653</td>
</tr>
<tr>
<td>DIFFUSE LUNG DISEASE</td>
<td>41</td>
<td>198</td>
<td>414</td>
<td>653</td>
</tr>
<tr>
<td>HEMORRHAGE</td>
<td>6</td>
<td>233</td>
<td>414</td>
<td>653</td>
</tr>
<tr>
<td>INFECTION</td>
<td>94</td>
<td>146</td>
<td>413</td>
<td>653</td>
</tr>
<tr>
<td>OTHER</td>
<td>24</td>
<td>214</td>
<td>415</td>
<td>653</td>
</tr>
</tbody>
</table>

Billing errors are noticed during data acquisition. For example, forceps biopsies that are intended to acquire mucosal tissue and/or tissue from a tracheobronchial mass are occasionally billed as TBBx which is intended to acquire parenchymal tissue. Occasionally both parenchymal and mucosal tissue were acquired and submitted together in one jar during one procedure. Also, sometimes the pathology report mentions presence of only mucosal tissue in a TBBx. These conditions made it impractical to discern what is the intended tissue for biopsy, if intended tissue is obtained or not and if the sample is adequate or not. Thus only where specifically indicated in the pathology report that the sample was insufficient for evaluation is it documented as insufficient.

One hundred and twenty two of 653 patients are suspected to have lung cancer pre-procedure. One hundred and forty seven TBBx samples were obtained from these 122 patients (figure 13). Of these, 1 sample is reported to be insufficient for evaluation, 113 samples are reported to be negative for malignancy, 2 samples are reported to be highly suspicious for malignancy but indeterminate and 31 samples are positive for malignancy. Pre-procedure suspicion of metastasis is evaluated separately. Most of these patients are also suspected to have lung cancer and have been reported in that group. Metastasis
and is suspected in 20/653 patients, from whom 27 TBBx samples are acquired. All 27 samples were considered adequate for evaluation. Twenty one of these 27 samples are negative for malignancy and 6 samples are positive. Lung cancer is not suspected in 110/653 patients from whom 143 TBBx samples are acquired. Of these, 4 samples are reported inadequate for evaluation, 138 samples are reported negative for malignancy and 1 sample is reported positive for malignancy. Five hundred and eleven samples from 421 patients did not have any differential diagnosis specified pre-procedure. Of these, 1 sample is reported insufficient for evaluation, 439 samples are negative for malignancy, 9 samples are suspicious of malignancy and 62 samples are reported positive for malignancy.

Figure 13: showing correlation between pre-test differential diagnosis and final pathology report with reference to presence or absence of malignancy. DDx= Differential diagnosis, pts= patients.
Smoking history is not provided at all in the pathology reports for TBBx 74% (482/653) of the times. Table 8 (a and b) show the classification and data for smoking history. Figure 14 gives a diagrammatic representation of smoking history as provided to the diagnosing pathologist. In 12% (78/653) of the cases history of smoking is mentioned, but no further details about pack years or current status of smoking or history of passive smoke exposure are provided. Thirty one cases (5%) have past history of smoking, but again information about pack years, time since patient quit smoking and other details are not mentioned. Twenty eight patients (4%) are reported to be never smokers. However, history of passive smoke exposure is not provided. Ten patients (2%) are reported to be currently not smoking, but no indication of past history of smoking, pack years or time since they quit smoking is provided. Eight patients (1%) are reported to have a history smoking more than 20 pack years. However, if they are current smokers, or not, history of secondary smoking etc is not provided. History of passive smoke exposure is mentioned in 2 cases, but details are not provided (Figure 14). In one patient only is a relatively detailed history of smoking mentioned. This patient is reported to be currently not smoking, has history of smoking less than 20 pack years and quit smoking more than 5 years ago. History of passive smoke exposure is not provided for this patient.

Table 8a (top) shows categorization of smoking history data acquired from TBBx reports. Table 8b (bottom) shows detail information acquired for the 16 reports that are categorized as others in table 8a.

<table>
<thead>
<tr>
<th>No. of procedures</th>
<th>Smoking History</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>h/o smoking +, &gt;20 pack years</td>
</tr>
<tr>
<td>10</td>
<td>currently not smoking.</td>
</tr>
<tr>
<td>28</td>
<td>never smoker, h/o passive smoke exposure not provided</td>
</tr>
<tr>
<td>31</td>
<td>past h/o smoking</td>
</tr>
<tr>
<td>78</td>
<td>h/o smoking +, details not provided</td>
</tr>
<tr>
<td>482</td>
<td>not provided</td>
</tr>
<tr>
<td>16</td>
<td>others</td>
</tr>
</tbody>
</table>
Table 8b:

<table>
<thead>
<tr>
<th></th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Quit smoking &lt; 5 years ago.</td>
</tr>
<tr>
<td>1</td>
<td>quit smoking &gt;5 years ago</td>
</tr>
<tr>
<td>1</td>
<td>currently not smoking, &lt;20 pack years, quit smoking &gt; 5 years ago</td>
</tr>
<tr>
<td>2</td>
<td>currently smoking, &gt; 20 pack years</td>
</tr>
<tr>
<td>2</td>
<td>never smoker, h/o passive smoke exposure +, details not provided</td>
</tr>
<tr>
<td>4</td>
<td>currently smoking, pack years not provided</td>
</tr>
<tr>
<td>5</td>
<td>currently smoking +</td>
</tr>
</tbody>
</table>

![Smoking History Pie Chart]

**Figure 14: Pie chart for smoking history obtained from TBBx pathology reports. No smoking history is provided in pathology reports about 74% of times. Even in the 12% of procedures where there is mention of history of smoking present, no further details are provided. Passive smoke exposure history is rarely provided.**

Billing record proves to be an unreliable source of information to distinguish between forceps biopsy and TBBx as some misclassification is observed. The epidemiological codebook does not code separately for forceps biopsy and TBBx, However, it is designed to document airway and tracheobronchial tree as intended tissue separately. Since tissue intended for biopsy is not specified in the pathology report pre-procedure, it is not possible to discern the difference most of the time the raw data is depicted in figures 15 for airway tissue and in figure 16 for tracheobronchial tree tissue. Many procedures
are coded in both since it could not be discerned what the intended tissue was. Based on data gathered it is seen that the right lung is biopsied 55.3% to 60% of the times. Whereas the left lung is biopsied 29.2% to 33.8%. The specimen site is specified to the level of the lobar bronchus in 619/801 samples (77.3%) samples. Specification of site to the level of segmental bronchus and beyond is specified for 24/240 samples (10%) for the left lung and 32/443 (7.2%) samples on the right lung as shown in figure 15 below. The ‘not provided’ category includes samples in which either specimen site was not provided at all (~20 samples, figure 16) or it was clearly a forceps biopsy.

![Specimen Site: Airway](image)

Figure 15: Distribution of parenchymal specimen site biopsied. RUL= right upper lobe, RML= right middle lobe, RLL= right lower lobe, LUL= left upper lobe, LL= left lingula, LLL= left lower lobe, post= posterior segment, ant= anterior, med= medial, lat= lateral, sup= superior, inf= inferior, apicopost= apico-posterior.
Formalin fixation is documented in 795/801 samples (99.25%). Two samples were submitted unfixed, one of these is reported to be 5cc of red fluid with no tissue, and the other is frozen for tumor bank. Of the remaining 4 samples, 3 are acquired from one patient during one procedure and it is not specified as how the samples were submitted to pathology. One sample acquired from another patient is submitted to pathology and it too is not specified if it was submitted in formalin.

The aggregate tissue dimensions are reported for 771/801 (96%) samples. Average aggregate volume calculated from these dimensions for 771
samples is 0.048 cm$^3$. Of these 771 samples 103 are reported to be malignant and 668 samples are reported to be non-malignant. Of the 30 samples that do not provide tissue dimensions 2 samples are reported to be malignant and 28 are reported to be non-malignant. The absence of this data does not significantly influence the result (Table 9). The probability of volume measurements being unavailable is the same for cancer and not cancer samples (chi test = 0.29).

Table 9: Showing that the probability of volume measurements being unavailable is not statistically significantly different for samples reported positive for cancer versus samples reported negative for cancer.

<table>
<thead>
<tr>
<th>Observed</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer +</td>
<td>Cancer -</td>
</tr>
<tr>
<td>Vol +</td>
<td>103 (98%)</td>
</tr>
<tr>
<td>Vol -</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>105</td>
</tr>
</tbody>
</table>

The average aggregate tissue volume for samples reported negative for neoplasia is 0.047 +/- 0.059 cm$^3$. The average aggregate tissue volume of samples reported positive for neoplasia 0.062 +/- 0.065 cm$^3$. The average aggregate tissue volume for samples reported negative for neoplasia is statistically significantly less than the average aggregate tissue volume of samples reported positive for neoplasia, p<0.03, unpaired two-tailed t-test with unequal variance) (Figure 17).
Figure 17: The aggregate volume of TBBx tissue obtained from 771/801 TBBxs (96%) performed between August 31st 1999 and June 30th 2005. Excluded are 30 samples for which volume is not provided. The average aggregate volume of 103 samples reported positive for malignancy is significantly more than the average aggregate volume of samples reported negative for malignancy.

This result indicates one of two possibilities. The higher average aggregate tissue volume of malignant tissue could be due to increased cellularity of the malignant tissue or secondary to sampling bias.

A histopathological examination is conducted of the reports positive for malignancy. As shown in figure 18, adenocarcinoma is the most common histopathological type of malignancy observed (44 samples, 39%). Seventeen samples (15%) are reported to be non-small cell lung cancer, but further detail on histopathological type of cancer is not reported, therefore these are categorized separately. Squamous cell carcinoma is reported in 16 (14%) samples. Bronchoalveolar carcinoma is seen in 12 (11%) samples and Carcinoma-in-site is seen in 6 (5%) samples. Carcinoid (1%) and atypical carcinoid (0%) tumors are very rare. Since the codebook developed for this project did not code for mixed type of tumors, they are coded by the histopathological type of cells mentioned in
the pathology report. For example, if a tissue is reported to have adeno-squamous type of carcinoma, this tissue is added to both the categories of adenocarcinoma and squamous cell carcinoma. Therefore the total in figures 18 and 19 is greater than 105, the total number of samples reported positive for malignancy.

Figure 18: Pie chart showing histopathological distribution of type of malignancies reported. One hundred and five samples are reported positive for malignancy. However, the total is 112 since samples reported to have mixed cellularity are counted more than once.

The different histopathological types of tissue are also categorized by differentiation. Adenocarcinoma is the most commonly reported carcinoma in all three categories of “well differentiated”, “poorly differentiated” and “differentiation not specified”. Overall, well differentiated adenocarcinoma is the most common type of malignancy reported, seen in 20 samples (figure 19). Among the non-small cell lung cancer tissue for which no further detail is provide, equal number of samples (7 samples each) are seen in the “differentiation not specified” and “poorly differentiated” categories. Among the 4 samples in the “not specified” category in figure 19, one is highly suspicious of non-small cell lung cancer but
not definitive, one is cellular atypia suggestive of carcinoma, one is reported as undifferentiated malignant neoplasia but no further detail on histological type, and one is suggestive of MALT (mucosal associated lymphoid tissue) lymphoma. This analysis demonstrates that this classification system provides a matrix of cell types and degree of differentiation.

Figure 19: Distribution of various histopathological types of malignancies based on differentiation.

The various histopathological types of malignancies are also categorized by gender as shown in figure 20. The 105 samples reported positive for malignancy are acquired from 92 patients, 64 (69.6%) male and 28 (30.4%) female. Adenocarcinoma is the most common histopathological type of malignancy seen in both genders; 25/64 (39.1%) in male, 10/28 (35.7%) in female. Of note here is that the total number of malignancies reported in figure 20 is 94, not 92. This is because some samples are reported highly suggestive of malignancy but no further detail about histopathological type is specified and therefore they are not included in this figure. While there are some samples that have mixed cell type and are included in two categories. For example, if a pathology report mentions adeno-squamous type of malignant tissue seen, it is documented once under adenocarcinoma and once under squamous cell carcinoma. Therefore this figure is an estimated representation of the type of malignancies seen in both genders.
Figure 20: Gender distribution of histopathological types of malignancies diagnosed by TBBx performed between August 31st 1999 and June 30th 2005. Included are all conclusive/definite pathological diagnoses; excluded are specimens diagnosed as suspicious. Samples reported to be of mixed cellularity are assigned to each cell type. E.g.: adeno-squamous carcinoma is counted once as adeno carcinoma and once as squamous carcinoma.

In order to assess the pre-test probability of diagnosing lung cancer the pre-procedure differential diagnosis mentioned in pathology reports (column I in table 4) is correlated with final impression of the pathologist (column Z in table 4). Figure 21 shows the distribution of pre-test differential diagnosis and final report with regard to presence or absence of malignancy. Lung cancer is suspected in 122 patients from whom 147 samples are acquired. Of these 147 samples one sample is insufficient for evaluation, 33 samples are reported positive for malignancy and the remaining 113 samples are reported negative for malignancy. The pretest probability is calculated as 33/147 (22.45%). Samples inadequate for evaluation are included in this calculation as this sample may have been positive if sufficient and therefore contributes to the final outcome. From a patient’s perspective, it is an opportunity lost and a cause of delay in diagnosis.

The pre-test probability is thus highest (22.45%) in patients suspected to have lung cancer and lowest (0.7%) in patients not suspected to have lung cancer. However, from this data, as seen in figure 21, the largest group of
patients is the ‘not specified’ category. This category also has the largest number of samples reported positive for malignancy (71).

![Pre-Test Probability](image)

**Figure 21:** The pre-test probability of pathology report positive for malignancy when it is suspected pre-procedure. It is highest for patients suspected to have malignancy and lowest in patients not suspected to have malignancy pre-bronchoscopy.

Although lung cancer is the main focus of the current project, non-malignant conditions are included in the codebook to assess for feasibility of doing future studies for other clinical conditions using the same or similar codebook. Cellular atypia is identified in 127 samples. Squamous metaplasia in 35 samples, type II pneumocyte hyperplasia in 49 samples, reactive hyperplasia in 3 samples, and other cellular atypia in 40 samples (table 10).

*Table 10: Overview of types of cellular atypia reported in pathology reports of all 801 samples.*

<table>
<thead>
<tr>
<th>Cellular Atypia</th>
<th># of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent</td>
<td>674</td>
</tr>
<tr>
<td>Squamous Metaplasia</td>
<td>35</td>
</tr>
<tr>
<td>Type II Pneumocyte Hyperplasia</td>
<td>49</td>
</tr>
<tr>
<td>Reactive Hyperplasia</td>
<td>3</td>
</tr>
<tr>
<td>Other Cellular Atypia</td>
<td>40</td>
</tr>
<tr>
<td>Sum</td>
<td>801</td>
</tr>
</tbody>
</table>
Table 11: Overview of the non-malignant findings noted in pathology reports of all 801 samples.

<table>
<thead>
<tr>
<th>Nonmalignant Diagnoses</th>
<th>Absent</th>
<th>Present</th>
<th>Not Specified</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>675</td>
<td>107</td>
<td>19</td>
<td>801</td>
</tr>
<tr>
<td>Non-Caseating Granuloma</td>
<td>688</td>
<td>104</td>
<td>9</td>
<td>801</td>
</tr>
<tr>
<td>Caseating Granuloma</td>
<td>792</td>
<td>2</td>
<td>7</td>
<td>801</td>
</tr>
<tr>
<td>Fibrosis</td>
<td>666</td>
<td>127</td>
<td>8</td>
<td>801</td>
</tr>
<tr>
<td>Sarcoid</td>
<td>759</td>
<td>31</td>
<td>11</td>
<td>801</td>
</tr>
<tr>
<td>Acute Inflammatory Cells</td>
<td>613</td>
<td>180</td>
<td>8</td>
<td>801</td>
</tr>
<tr>
<td>Chronic Inflammatory Cells</td>
<td>497</td>
<td>296</td>
<td>8</td>
<td>801</td>
</tr>
<tr>
<td>Vasculitis</td>
<td>784</td>
<td>5</td>
<td>12</td>
<td>801</td>
</tr>
<tr>
<td>Hemorrhage</td>
<td>719</td>
<td>75</td>
<td>7</td>
<td>801</td>
</tr>
<tr>
<td>Other- specify</td>
<td>607</td>
<td>186</td>
<td>8</td>
<td>801</td>
</tr>
</tbody>
</table>

Data for numerous non-malignant conditions including caseating and non-caseating granulomas, acute and chronic inflammation, fibrosis, vasculitis, hemorrhage and miscellaneous conditions are all documented as mentioned in

Table 12: Overview of staining data acquired from pathology reports of all 801 samples.

<table>
<thead>
<tr>
<th>Staining</th>
<th>Not Done</th>
<th>Positive</th>
<th>Negative</th>
<th>Equivocal</th>
<th>Not Specified</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>H&amp;E</td>
<td>11</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>782</td>
<td>801</td>
</tr>
<tr>
<td>MUCIN</td>
<td>3</td>
<td>1</td>
<td>11</td>
<td>1</td>
<td>785</td>
<td>801</td>
</tr>
<tr>
<td>LCA</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>791</td>
<td>801</td>
</tr>
<tr>
<td>CD20</td>
<td>4</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>783</td>
<td>801</td>
</tr>
<tr>
<td>CD43</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>793</td>
<td>801</td>
</tr>
<tr>
<td>SYNAP</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>787</td>
<td>801</td>
</tr>
<tr>
<td>GMS</td>
<td>2</td>
<td>3</td>
<td>171</td>
<td>0</td>
<td>625</td>
<td>801</td>
</tr>
<tr>
<td>S100</td>
<td>4</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>791</td>
<td>801</td>
</tr>
<tr>
<td>HMB 45</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>795</td>
<td>801</td>
</tr>
<tr>
<td>CYTO 7</td>
<td>4</td>
<td>14</td>
<td>4</td>
<td>1</td>
<td>778</td>
<td>801</td>
</tr>
<tr>
<td>CYTO 20</td>
<td>4</td>
<td>2</td>
<td>15</td>
<td>2</td>
<td>778</td>
<td>801</td>
</tr>
<tr>
<td>AFB</td>
<td>5</td>
<td>2</td>
<td>151</td>
<td>0</td>
<td>643</td>
<td>801</td>
</tr>
<tr>
<td>GRAM</td>
<td>2</td>
<td>1</td>
<td>39</td>
<td>0</td>
<td>759</td>
<td>801</td>
</tr>
<tr>
<td>OTHER</td>
<td>2</td>
<td>47</td>
<td>69</td>
<td>3</td>
<td>680</td>
<td>801</td>
</tr>
</tbody>
</table>
the pathology reports and data is represented in table 11. Chronic inflammation is the most common (296/801 samples, 36.95%) non-malignant finding reported. Sarcoidosis is specifically mentioned in 31 samples, although non-caseating granulomas are seen in 104 samples.

Similarly data for staining done is also acquired and analyzed, as summarized in table 12. Note here that from this data it appears that hematoxylin and eosin (H&E) staining is frequently not specified. This is most likely misleading information because it is routine practice to do this staining and therefore is not regularly specifically mentioned in the report. However, all other stains are special stains and are usually meticulously reported. It is thus evident that special stains are not routinely done. GMS (174, 21.72%) and AFB (153, 19.1%) are the most commonly used stains.

**Specific Aim 3:** To identify and integrate matched pairs of TBBx and lung resections in order to validate the performance of TBBx against lung resection as the gold standard.

Search results using “Natural Language Ila search” software identified 690 lung resection surgeries performed at the Hershey Medical Center between 2\textsuperscript{nd} June 2000 and 8\textsuperscript{th} December 2006 (6.5 years). Using patient name and MRN to identify patients common to both, the TBBx and lung resection datasets resulted in 86 patients (figure 22). Six lung resection surgeries are found to be misclassified and are actually TBBx, therefore excluded. One patient’s TBBx pathology report is not available on Power Chart, therefore excluded. Five more patients’ data are excluded since their TBBx was done post-operatively. Thus 74 procedures are common and available in both datasets for further analysis (figure 22).

The TBBx pathology reports of these 74 patients are analyzed with regard to presence or absence of neoplasia. Two of these procedures report insufficient
sample for evaluation therefore could not be categorized into test positive or negative categories. However, because from a process of care and patient perspective it is considered a missed opportunity to diagnose lung cancer, the lung resection results of these patients are documented. One of these patients is positive for malignancy on lung resection and the other is reported negative for malignancy. The remaining 72 cases are categorized based on TBBx pathology reports.

Of the 72 procedures included for further analysis, 15 (20.83%) TBBx pathology reports are positive for neoplasia and 57 (79.17%) are reported negative for neoplasia (figure 22).

The lung resection pathology reports of these 72 patients reported 42 (58.33%) specimens to be positive for neoplasia and 30 (41.67%) specimens to be negative for neoplasia (figure 22).
Of the 15 TBBx samples reported positive for neoplasia, 14 (93.33%) were also reported positive for neoplasia on subsequent lung resection surgery. This group is referred to as the true positive (+/+). One of the 15 TBBx samples (0.07%) reported positive for neoplasia, is reported negative for neoplasia on subsequent lung resection. This sample was reported to be carcinoma in situ on TBBx. This sample is referred to as false positive (+/-).
Of the 57 TBBx samples reported negative for neoplasia, 28 (49.12%) are reported positive for neoplasia on subsequent lung resection surgery. Three of these 28 samples were reported to have cellular atypia and otherwise indeterminate and finally considered negative for neoplasia. All three were subsequently reported to be positive for neoplasia on surgery. This group of 28 patients is referred to as false negative (-/+). henceforth in this study (table 13). The remaining 29 (50.88%) TBBx samples reported negative for neoplasia are also reported negative for neoplasia on subsequent surgery. This group is referred to as true negative (-/-) henceforth in this study.

Table 13: Correlation between TBBx and lung resection pathology reports with regard to presence or absence of malignancy. Two patients’ data is excluded, suboptimal for evaluation.

<table>
<thead>
<tr>
<th>TBBx</th>
<th>Surgery</th>
<th>Neoplasia Positive</th>
<th>Neoplasia Negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neoplasia Positive</td>
<td>14 (TP)</td>
<td>1 (FP)</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Neoplasia Negative</td>
<td>28 (FN)</td>
<td>29 (TN)</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>42</td>
<td>30</td>
<td>72</td>
<td></td>
</tr>
</tbody>
</table>

Test Performance Characteristics are calculated as follows:

Sensitivity = True positive / (True positive + False negative)
= 14/ (14+28) = 14/42 = 0.33

Specificity = True negative / (False positive + True negative)
= 29/ (1+29) = 29/30 = 0.967

False Negative Rate (FNR) = False Negative / (True positive + False negative)
= 28/ (14+28) = 28/42 = 0.67

False Positive Rate (FPR) = False positive / (False positive + True negative)
= 1/ (1+29) = 1/30 = 0.03

Predictive Value Positive (PVP) = True positive / (True positive + False positive)
= 14/ (14+1) =14/15 = 0.93

Predictive Value Negative (PVN) = True negative / (False negative + True negative)
= 29/ (28+29) = 29/57 = 0.51
A correlation between histopathological types of malignancy reported in the two datasets, revealed adenocarcinoma to be the common histopathological type of cancer seen (figure 23). Of note here is that the total number of TBBx samples in figure 23 is 22, not 15, and the total number of lung resection specimens is 49, not 42. This again is due the presence of tumors with mixed cellularity. There are a few cases where the histopathology reported on TBBx sample is different from the histopathology reported on lung resection specimen. A detailed analysis of lesion characteristics for each of the four groups is conducted in aim 5.

![Figure 23: Correlation between histopathological types of malignancy reported on TBBx and lung resections.](image)

**Specific Aim 4:** To benchmark TBBx pathology reports as a function of digitally acquired computerized tomography (CT) lesions using a novel classification scheme and identify characteristics of lesions commonly biopsied and lesions frequently reported to be positive for malignancy.

For the 653 TBBx procedures done between 31\textsuperscript{st} August 1999 and June 30\textsuperscript{th} 2005, 298 (45.64\%) digital CT scan images acquired between December
2002 and 2007, are available for retrieval and analysis. Parenchymal lesions, aka region of interest (ROI), are seen in 333 transverse slices. Two hundred and forty two of the 298 (81.21%) digital CT scans are acquired pre-TBBx, 9 (3.02%) are acquired on the same day as TBBx, and 47 (15.77%) are acquired post-TBBx (figure 24). A more detailed analysis evaluating the availability of digital CT scans with time, correlating the time interval between digital CT scan acquisition, ROI characteristics and TBBx report with reference to presence or absence of malignancy is conducted in aim 6.

Of the 333 visible ROIs available for further analysis, 180 (54.05%) are focal ROIs and 153 (45.95%) are diffuse ROIs. Among the focal lesions 33/180 (18.33%) are less than 1 cm, 69/180 (38.33%) are 1-3 cm, 65/180 (36.11%) are greater than 3 cm in length and 13/180 (7.22%) are multifocal. Among the diffuse ROIs 73/153 (47.71%) involve less than 25% of a lobe, 37/153 (24.18%) involve 25-50% of a lobe and 43/153 (28.1%) involve more than 50% of a lobe (figure 24).
A correlation between TBBx pathology report with respect to presence or absence of malignancy and digital CT scan acquired ROI classification is established. It is seen that 1/33 (3.03%) focal ROIs less than 1 cm is reported to be malignant, 12/69 (17.39%) of focal ROIs 1-3 cm in length, 24/65 (36.92%) of focal ROIs greater than 3 cm in size and 2/13 (15.38%) multifocal lesions biopsied are reported positive for malignancy (figure 25).

![Histogram demonstrating correlation between CT scan ROI classification and TBBx pathology report with regard to malignancy.](image)

Among the diffuse lesions biopsied, 9/73 (12.33%) ROIs involving less than 25% of the lobe, 2/37 (5.41%) of ROIs involving 25-50% of the lobe and 8/43 (18.6%) ROIs involving more than 50% of the lobe are reported to be positive for malignancy (figure 25).

An essential question raised here is to identify lesions characteristics that are either missed or inaccurately diagnosed (false negative) during TBBx. This question is addressed in aim 5.
Specific Aim 5: Assessment of pulmonary lesion characteristics accurately and inaccurately diagnosed to be malignant by TBBx.

Of the 74 patients common to the TBBx and lung resection datasets, 2 TBBx samples are reported inadequate for evaluation. Lung resection data for both these patients is available. One of them is reported to be positive for malignancy while the other was negative. Thus one can be considered to be false negative and the other true negative.

Of the 72 patients that have been categorized into true positive, false positive, true negative and false negative, a detailed assessment of their data is conducted to identify lesion characteristics for each group. The number of patients in each group is:

i. True Positive = 14
ii. True Negative = 29
iii. False Positive = 1
iv. False Negative = 28

Using patient name and MRN as linking criteria, 31 of the 74 patients also have digital CT scan images available, thus are common to all three datasets (figure 26). All outcome measures, lesion location, congruency of site, histopathology and CT scan lesion characteristics are presented by group in which the patients belong.
**True Negative:** Of the 29 true negative cases, digital CT scan data is available for 14 cases (figure 27). Five of these 14 (35.7%) true positive cases have diffuse lesions involving less than 25% of the lobe. Three are diffuse lesions involving more than 50% of the lobe, 2 in each focal category of lesions 1-3 cm and > 3cm in size. One is a focal lesion less than 1 cm and one a diffuse lesion involving 25-50% of the lobe.
Lesion locations biopsied and resected for the 29 true negative cases are shown in figure 28. The total number of TBBx samples acquired from these 29 patients is 34, and the total number of specimens acquired on resection is 55. Right lower lobe is most frequently biopsied (11 TBBx samples) and resected (18 specimens) in this group. Figure 29 shows the congruency in specimen sites biopsied and resected. The total here again is more than 29 because some patients have more than one site biopsied and resected. The right lower lobe has maximum congruency. Since these are all true negative, histopathology details for these samples is not relevant to this study thus not presented here.
Figure 28: Lesion location for 29 true negative TBBx and lung resection matched pairs.

Figure 29: Congruency of specimen site for 29 true negative TBBx and lung resection matched pairs.

True Positive: Of the 14 true positive matched pairs identified, digital CT scan data is available for 4 cases (figure 30). Two of these have focal lesions 1-3 cm in size and one each has a focal lesion >3cm in size and the other has a diffuse lesion involving less than 25% of the lobe. Figure 31 shows the
histopathological reports of these 14 true positive matched pairs. Adenocarcinoma is the most common type of malignancy in both groups. The total number of lung resections in this figure is 15 and not 14 because one patient is reported to have adenocarcinoma and large cell carcinoma therefore included in two categories.

Figure 30: CT scan lesion characteristics for 14 true positive TBBx and lung resection matched pairs.

Figure 31: Lesion histopathology data for 14 true positive TBBx and lung resection matched pairs.
Figure 32: Lesion location for 14 true positive TBBx and lung resection matched pairs.

Right upper lobe is the most common site biopsied and resected in this category. Seven of the 14 matched pairs are congruent in the site biopsied and site resected, as shown is figure 33. Right upper lobe has the maximum congruency of lesion site in this group (figure 34).
Figure 34: Lesion location and congruency of specimen site for 14 true positive TBBx and lung resection matched pairs.

False Positive: Of the 74 matched pairs, one sample is positive for neoplasia on TBBx but negative on lung resection. CT scan data for this sample is not available and the specimen site is left upper lobe for both TBBx and resection, thus congruent. The histopathology report of the TBBx sample stated it to be a carcinoma-in-situ, squamous cell carcinoma.

False Negative: Of the twenty eight cases in the false negative category, digital CT scan data are available for 13 cases. Six of these have focal lesions 1-3 cm in size, 4 have a focal lesion >3cm, one each has a focal lesion <1cm, a diffuse lesion involving <25% of the lobe, and a diffuse lesion involving more than 50% of the lobe (figure 35).

Histopathology of 15/28 (53.57%) false negative cases is reported to be adenocarcinoma (figure 36). Of the 3 cases that are reported to be indeterminate and thus categorized as test negative, all three are reported positive for
malignancy on lung resection. These three are categorized as “other ca” in the TBBx pathology report.

Figure 35: CT scan lesion characteristics for 28 false negative TBBx and lung resection matched pairs.

Figure 36: Lesion histopathology data for 28 false negative TBBx and lung resection matched pairs. Three of these are reported as indeterminate on TBBx, categorized as negative. Total is 29 since one specimen is mixed type carcinoma.
Figure 37: Lesion location for 28 false negative TBBx and lung resection matched pairs. Total number of TBBx is 30 because 2 patients had 2 sites biopsied.

Figure 38: Overall congruency of sampling location for the 28 false negative TBBx and lung resection matched pairs.
Figure 37 shows the site from where the biopsy and resection specimens are acquired. Right upper and lower lobes are the more frequently biopsied and resected than other sites in this group. Eighteen of the 28 sites are congruent (figure 38). Right upper lobe has the maximum congruency of site biopsied and resected (figure 39).

![Specimen Site Congruency in Samples Negative on TBBx but Positive for Malignancy on Resection](image)

*Figure 39: Lesion location and congruency of specimen site for 28 false negative TBBx and lung resection matched pairs.*

Table 14 shows the CT scan lesion characteristics for each of the groups. The largest number of false negative TBBx reports is for focal lesions 1-3 cm in size. However, the largest number of true positive is also in this category, although the numbers are small for any statistical assessment. The most common lesion in the true negative category is diffuse lesion involving less than 25% of the lobe. Digital CT scan data is not available for the one false positive TBBx sample which is reported to be carcinoma-in-situ.
Table 14: Digitally acquired CT scan pulmonary lesion characteristics of the 31 matched pairs categorized as a function of TBBx and lung resection matched pathology report with regard to presence or absence of malignancy. One patient in false positive category is reported separately. Digital CT scan data for this patient is not available.

<table>
<thead>
<tr>
<th>Group</th>
<th>Digital CT Scan acquired lesion characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Focal (SA)</td>
</tr>
<tr>
<td></td>
<td>&lt;1cm</td>
</tr>
<tr>
<td>True positive (+/+</td>
<td>0</td>
</tr>
<tr>
<td>True Negative (-/-</td>
<td>1</td>
</tr>
<tr>
<td>False Negative (-/+</td>
<td>1</td>
</tr>
<tr>
<td>False Positive (+/-</td>
<td>Digital CT scan not available. TBBx pathology report carcinoma-in-situ.</td>
</tr>
</tbody>
</table>

**Specific Aim 6:** To estimate the time intervals between CT scan, TBBx and lung resection in order to identify time gaps in the current process of care for lung cancer patients at the Hershey Medical Center.

**Availability:** Availability of data is quantified into 7 categories depending on time period in which the TBBx is done as shown in table 15. Maximum TBBxs included in this study (195/653, 29.86%) are done in 2004. This is also the year for which the largest number of matched digital CT scan images is available (139/195, 71.28%). Availability of lung resection data is comparatively low (74/653, 11.33% of TBBx), with the largest number being available in 2004. However, the largest proportion (8/50, 16%) is available in 2001. Of note here is that the availability is categorized by the time frame in which the corresponding TBBx is done, irrespective of when the CT scan is acquired or when the lung resection is done. For example if a patient had a TBBx done in the year 2000, but the CT scan is done in 1999 and the lung resection is done in 2001, this patients’ data is categorized in 2000.
Table 15: Availability of digital CT scan and lung resection data for all 653 TBBx done; categorized by date of service for TBBx.

<table>
<thead>
<tr>
<th>Time Period</th>
<th>Number of TBBx</th>
<th>Number of paired CT scans available (% of TBBx)</th>
<th>Number of paired Lung Resections available (% of TBBx)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>27</td>
<td>0 (0%)</td>
<td>1 (3.7%)</td>
</tr>
<tr>
<td>2000</td>
<td>67</td>
<td>0 (0%)</td>
<td>5 (7.46%)</td>
</tr>
<tr>
<td>2001</td>
<td>50</td>
<td>0 (0%)</td>
<td>8 (16%)</td>
</tr>
<tr>
<td>2002</td>
<td>79</td>
<td>6 (7.59%)</td>
<td>10 (12.66%)</td>
</tr>
<tr>
<td>2003</td>
<td>138</td>
<td>91 (65.94%)</td>
<td>16 (11.59%)</td>
</tr>
<tr>
<td>2004</td>
<td>195</td>
<td>139 (71.28%)</td>
<td>19 (9.74%)</td>
</tr>
<tr>
<td>2005</td>
<td>97</td>
<td>62 (63.92%)</td>
<td>15 (15.46%)</td>
</tr>
<tr>
<td>Total</td>
<td>653</td>
<td>298 (45.64%)</td>
<td>74 (11.33%)</td>
</tr>
</tbody>
</table>

Figures 40 and 41 respectively show in logarithmic scale the availability of lung resection and digital CT scan data with respect to time period in which TBBx is done. These figures clearly demonstrate that availability of lung resection data although low, is relatively consistent throughout the period of this study. On the other hand, digital CT scan data is not available for TBBxs done in the first three years of the study, but a good quantity is available for the last 3 years of the study.
Figure 40: Time trend demonstrating availability of the 74 matched lung resections.

Figure 41: Time trend demonstrating availability of 298 digital CT scan images for matched TBBx procedures.
Data are then analyzed by re-categorization depending on the date of service of digital CT scan and lung resections, irrespective of when the corresponding TBBx is done. Figure 42 shows that the 74 paired lung resections available are done between the years 2000 and 2006.

![Graph: Availability of 74 Matched Lung Resections With Time](image)

*Figure 42: Availability of 74 matched lung resections with time.*

Similarly availability of the 298 digital CT scan images with time is shown in figure 43. As mentioned earlier, due to unavailability of digital CT scan facility at the Hershey Medical Center prior to December 2002, digital CT scan data is available only after that time. Also, since the TBBx study period ends in June 2005, very few matched digital CT scan images are available after that time period. Of note here is that due to unavailability of digital CT scan scans closer to the date of TBBx, 2 scans are acquired in the year 2007.
Time to diagnosis (TDx): Digital CT scan data are categorized into pre and post-TBBx. As seen in figure 24, 242 of the 298 (81.21%) matched digital CT scans included in this study are done pre-TBBx. Nine CT scans (0.03%) are done on the same day as TBBx and 47 (15.77%) scans are done post-TBBx. Pre and post-TBBx CT scans are further categorized depending on the time interval between the date when the scans are acquired and the date when TBBx is done (figure 44). Maximum number of pre-TBBx CT scans (64) is acquired between 3-7 days before TBBx.

To assess the average time to diagnosis from these data, all post-TBBx are excluded (47/298). The average time to diagnosis calculated from the 251 pre-TBBx CT scan dates (CT scan done on same day included) available is 11.06 days (median = 7 days). The range is from 0 to 57 days pre-TBBx.
Time to treatment (TRx): Time period between TBBx and lung resection is then established for the 74 matched pairs. All lung resections included in the study are post-TBBx. Majority of the lung resections (33/74, 44.59%) are done between 31-90 days (3 months) post-TBBx (figure 45) with the median time to treatment being 63.5 days (average = 130 days). The range is from 4 to 1493 days (4.15 years) post-TBBx.
Of the 74 matched pairs in the TBBx and lung resection database, 31 patients have CT scan images available in digital format. Of these 31 patients, 23 underwent CT scans pre-TBBx (Table 16) and 8 post-TBBx. These 31 patients' data are evaluated to estimate time to diagnosis and time to treatment.

Table 16: Evaluation of process of care for lung cancer diagnosis. Time interval between CT scan and TBBx, and time interval between TBBx and lung resection.

<table>
<thead>
<tr>
<th>CT Scan to TBBx time for 31 matched pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days included</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>0 days</td>
</tr>
<tr>
<td>1-2 days</td>
</tr>
<tr>
<td>3-7 days</td>
</tr>
<tr>
<td>8-14 days</td>
</tr>
<tr>
<td>15-30 days</td>
</tr>
<tr>
<td>31-90 days</td>
</tr>
<tr>
<td>91-180 days</td>
</tr>
<tr>
<td>&gt;180 days</td>
</tr>
<tr>
<td>Days included</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>1-2 days</td>
</tr>
<tr>
<td>3-7 days</td>
</tr>
<tr>
<td>8-14 days</td>
</tr>
<tr>
<td>15-30 days</td>
</tr>
<tr>
<td>31-90 days</td>
</tr>
<tr>
<td>91-180 days</td>
</tr>
<tr>
<td>&gt;180 days</td>
</tr>
</tbody>
</table>

**Total post-TBBx scans** 8

**Grand total of all scans** 31

<table>
<thead>
<tr>
<th>Days included</th>
<th>Time period</th>
<th># of resections</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 days</td>
<td>same day</td>
<td>0</td>
</tr>
<tr>
<td>1-2 days</td>
<td>48 hrs</td>
<td>0</td>
</tr>
<tr>
<td>3-7 days</td>
<td>1 week</td>
<td>1</td>
</tr>
<tr>
<td>8-14 days</td>
<td>2 weeks</td>
<td>0</td>
</tr>
<tr>
<td>15-30 days</td>
<td>1 month</td>
<td>1</td>
</tr>
<tr>
<td>31-90 days</td>
<td>3 months</td>
<td>15</td>
</tr>
<tr>
<td>91-180 days</td>
<td>6 months</td>
<td>8</td>
</tr>
<tr>
<td>&gt;180 days</td>
<td>&gt; 6 months</td>
<td>6</td>
</tr>
</tbody>
</table>

**Total post-TBBx resections** 31

Figure 46 is a graphical representation for individual patient data with regard to time to diagnosis and time to treatment. As seen in this figure, the variance for time to diagnosis is smaller than the variance for time to treatment.
Figure 46: Time intervals between CT scan date, TBBx date and Lung Resection data among the 31 patients common in three datasets. An estimate of process of care in diagnosis of lung cancer.

Data acquired for these 31 patients is correlated with the categories established in table 6 (true positive, true negative, false positive, false negative) to assess differences in the time to diagnosis and time to treatment between each of these categories. Four patients are in the true positive category, 14 in the true negative and thirteen in the false negative category. The one patient who is in the false positive category does not have a digital CT scan available. This
patients’ information is reported separately. Figure 47 shows the time to diagnosis and time to treatment for each group, with time to treatment in logarithmic scale. The highest variation in both axes is seen in the true negative group.

Data of 31 Patients Common to all Three Datasets
Categorized by Group

Figure 47: Graphical representation of data for time to diagnosis and time to treatment for the groups of patients common to all three datasets.

Data for individual groups are shown separately in figures 48-50.
Figure 48: Time to diagnosis and time to treatment for four true positive TBBx samples.

Figure 49: Time to diagnosis and time to treatment for fourteen true negative TBBx samples.
Figure 50: Time to diagnosis and time to treatment for thirteen false negative TBBx samples.

The one false positive TBBx sample is reported to have carcinoma in situ, squamous cell carcinoma. Time to treatment for this sample is 217 days. Digital CT scan image for this patient is not available.

In order to assess significance of differences in TDx and TRx between each group a two tailed, unpaired, student’s t-test with unequal variance is conducted. P value <0.05 is considered statistically significant. Data are shown in figure 51 and 52 for median and standard error of mean for each group. There is no statistically significant difference in TDx and TRx between any of the three groups. However, it should be kept in mind that the total number of patients in these groups is small (4 to 14) thus may be compromising the power of such analysis.
Figure 51: Median and SEM of time to diagnosis for the three groups of patients showing that the difference between the groups is not statistically significant (p>0.05).

Figure 52: Median and SEM of time to treatment for the three groups of patients showing that the difference between the groups is not statistically significant (p>0.05).
V: Discussion

The Institute of Medicine's publication, “Crossing the Quality Chasm” (Corrigan, Donaldson et al. 2004), challenges us to re-engineer healthcare processes. The Institute of Medicine also focuses on the need for “error free hand-offs” in patient care. Lung cancer diagnosis, staging and treatment are highly interdisciplinary activities requiring repeated patient “hand-offs” among practice partners and members of interdisciplinary health care teams.

Lung Cancer patients continue to have a poor prognosis with the 5-year survival of 13-16% (ACS 2007). One of the reasons for poor outcome is that about 72% of patients are diagnosed at an advanced stage, when survival is low (Fry, Phillips et al. 1999). Factors contributing to this include lack of a feasible screening method, lack of efficient diagnostic procedures to detect lung cancer at an early stage, low accuracy in obtaining desired tissue during bronchoscopy, lack of diagnostic and prognostic biomarkers to identify sensitivity or resistance to drugs thus lack of guideline for appropriate line of treatment. Another factor contributing to poor long term outcome is lack of response and development of resistance to treatment (Rosell, Cecere et al. 2006). Proactive measures addressing all these areas are being pursued to improve the long term outcome for these patients.

The results obtained in this study represent the lung cancer patient population care because the Hershey Medical Center is a tertiary health care centre. HMC is currently in the process of becoming a cancer institute, however, the time period of this study (1999-2005) is a good representation of the process of care most lung cancer patients are subjected to across the nation and also other countries with similar health care standards. The distribution of lung cancer population (age, gender, histological type) is representative of lung cancer population characteristics in other similar studies, thus the results obtained in this
study may be generalizable to process of care at other similar health care centers.

As defined in the first aim of this study, management of lung cancer patients requires coordination between numerous disciplines and medical specialists. Patient information is stored and available in numerous electronic datasets. Although individual datasets do not contain comprehensive patient information, this study demonstrates that developing links between these datasets can provide an excellent resource to assess process of care and identify strengths and weakness in the current process of care provided at the Hershey Medical Center, and in other similar health care centers. Defining the process of care for lung cancer patients and identifying the disciplines of medicine involved in their care helped identify numerous resources where patient information is available. This guided identification of specific resources and integrating data acquired from them to critically evaluate the diagnostic yield of TBBx in diagnosing lung cancer. This approach of using links between disparate datasets to evaluate the diagnostic efficiency of TBBx in detecting lung cancer is unique to this study. Other similar studies in literature include a cohort of all patients diagnosed with lung cancer.

Bronchoscopically acquired TBBx tissue is obtained from only a limited portion of the patients clinically suspected to have lung cancer, as shown in figure 53. It is essential to identify this limitation of the current project. TBBx yield is a candidate quality benchmark but yield varies with lesion size in prospective studies (Rai and Arroliga 1995). Improving the accuracy and yield of acquiring TBBx tissue, although critical, may have limited impact on the overall process of care and long term survival of patients with lung cancer.
Molecular markers hold the promise to engineer innovative methods to manage lung cancer patients. However, their full potential can only be harnessed if the desired tissue is obtained in a timely, accurate, efficient manner along with sufficient clinical data provided to the diagnosing clinical pathologist. A molecular medicine program for lung cancer translational research requires development of updated clinical tools used for the collection and processing of diagnostic samples and issuing clinical reports.

Improving the yield of TBBx in diagnosing lung cancer has the potential to contribute significantly to the process of care. This is a procedure routinely done, usually at an early stage of diagnosis and may be repeated for staging, follow-up, recurrence etc. Although TBBx is identified for its potential, it has its limitations that need to be recognized. The advantages, disadvantages and possible sources of error that can compromise the yield are listed below and figure 54 shows the practical processes that may compromise TBBx diagnostic yield.
Transbronchial Biopsy

- **Advantages**
  - Fits process of care
  - Obtained at time of initial diagnosis
  - Easily accessible tissue
  - Relatively non-invasive
  - Repeat/Follow up tissue obtainable
  - Could use molecular markers to guide treatment plan, including sequence/timing of surgery and neoadjuvant therapy

- **Concerns**
  - Frequently performed at HMC
  - Accuracy of obtaining tissue from desired site
  - Quantity of tissue
  - Diagnostic yield
  - Tumor heterogeneity, hence not representative
  - Accuracy of returning to same site for follow-up

- **Possible Sources of Error**
  - Wrong lobe
  - Wrong Segment
  - Edge effect
  - Tumor heterogeneity (multiple cell type, necrosis)
  - Sample too small for assay method
  - Tissue damage during procedure (crushing, hemorrhage)
  - Wrong tissue acquired (mucosal vs. parenchymal)
Figure 54: A Schematic showing the possible sources of error in acquiring tissue during bronchoscopic transbronchial biopsies. T= tumor, E1= Tissue acquired from wrong lobe, E2= Tissue acquired from wrong segment, E3= Tissue acquired form the edge of the tumor, may not be representative sample, E4= Tumor heterogeneity may result in incorrect interpretation of cell type. Other errors that not shown in schematic but listed in figure E5= Sample obtained may be too small therefore insufficient for evaluation, E6= tissue obtained may be damaged due to manipulations during procedure, E7= TBBx is done to acquire parenchymal tissue, however, sometimes mucosal tissue is obtained and the parenchymal lesion is not sampled. RUL= right upper lobe, RML= right middle lobe, RLL= right lower lobe, LUL= left upper lobe, LL= left lingula, LLL= left lower lobe.
Not only is acquiring tissue accurately at initial suspicion of cancer essential but returning to sites of previous abnormalities during follow up would rely on recall by the performing bronchoscopist, a systems limitation in a multi-individual practice.

Tissue obtained at first clinical encounter can serve as baseline tissue for molecular marker assays. These markers can serve as both diagnostic and prognostic markers. When tissue is acquired later during the process of care, any difference in the molecular markers can be useful to assess response or resistance to treatment. Thus serial tissue acquired from same patient has the potential to guide development of individualized patient treatment regimens.

The choice of using TBBx as the procedure for critical evaluation of diagnostic yield is made for two reasons. First, because it is often the first step in the process of care for histological confirmation of lung cancer therefore considered a critical stage. Secondly, in this study particularly, it proved to be a good choice because it is done frequently at the Hershey Medical Center (653 procedures in 5yrs 10 months, average 9.57 procedures a month) and thus provided a large population for evaluation. This is in contrast to other studies like Koh et al. that evaluated the outcome of 38 patients over a 10 month period (Koh, Tee et al. 2008) and included patients with both malignant and benign conditions. Roth et al. evaluated the diagnostic yield of bronchoscopy in 363 patients over a 2 years period (Roth, Hardie et al. 2008). Patients included in this study were all patients with a final diagnosis of lung cancer “through all possible methods”. In these respects our study is unique since we establish a pre-test probability of diagnosing lung cancer by correlating pre-procedure suspicion of lung cancer with TBBx pathology report. We then later establish true positive and false negative outcome by comparing TBBx pathology reports with lung resection pathology report. This design of our study includes patients in early stages of disease when surgical resection is part of routine process of care and
focuses on lesion characteristics in this group of patients. Including all patients would shift the population into later stage disease with many patients without histological information of the type of disease and therefore not provide information as what types of lesions are providing false negative results.

Data obtained in specific aim 2 of this study quantifies the various gaps where effective transfer of information during the multiple patient hand-offs is apparently lacking. The diagnosing pathologist is not provided with critical clinical patient information. As seen in table 7 although lung cancer is the most frequently mentioned pre-procedure differential diagnoses for TBBx (122/653, 18.7%), no differential diagnosis is provided for over 60% of the patients. Similarly smoking history is not provided in almost 75% of cases (figure 14). Lung cancer being predominantly a disease of smokers, this is critical information not provided to the diagnosing pathologist. No correlation between smoking status and malignancy could be made in this study because of lack of this information thus precluding us from comparing our outcome with published literature. This is critical since recent literature highlights the occurrence of lung cancer among life-time non-smokers. There is also growing evidence of smoking as a confounding factor with other factors including genetic susceptibility, occupational exposure to risk factors, environmental exposure to second-hand smoke etc.

With the long term vision of this study being the incorporation of molecular markers into routine practice of care for lung cancer patients, it becomes even more critical that the diagnosing pathologist have comprehensive clinical information on the patient in order to diagnose presence of malignancy and discern molecular effects involved in lung cancer including field carcinogenic exposure, genetic polymorphism, genetic mutations, loss of heterozygosity, oncogene activation, TSG inactivation, apoptosis gene deregulation, epigenetic aberrations, generalized DNA hypomethylation, promoter hypermethylation, telomerase activation etc.
Figures 15 and 16 demonstrate that the exact site of biopsy and intended tissue for pathological evaluation was difficult to discern in this study. Distinction between forceps biopsy intended for mucosal tissue and TBBx intended for parenchymal tissue could not be made. Specimen site is specified to the level of lobar bronchus in 619/801 samples (77.3%). Lack of detailed information about specimen site prevented us from analyzing occurrence of different histological types of malignancies with respect to location. As reported by Brooks et al. squamous cell and small cell carcinoma is most frequently seen in the central tracheo-bronchial tree while adenocarcinoma frequently occurs in the peripheral pulmonary parenchyma (Brooks, Austin et al. 2005).

Aggregate tissue volume information is not reported in the literature. This study is unique in this respect and demonstrates a statistically significant higher tissue volume acquired from samples reported positive for malignancy compared to aggregate tissue volume acquired from samples negative for malignancy. Larger tissue volume provides more tissue for the pathologist to do special staining including molecular marker profiles.

Adenocarcinoma is the most frequently seen histological type of malignancy in this study (44/112, 39%). This is similar to the incidence of adenocarcinoma reported by Wahbabh et al. of about 40% as seen in figure 3 (Wahbah M, Boroumand N et al. 2007). However, incidence of squamous cell carcinoma, small cell carcinoma and large cell carcinoma in this study (14%, 4%, 4% respectively) is lower than that reported by Wahbah et al. (30%, 15%, and 15%). This could be due to the difference in classification of lung cancer histology in the two studies. Wahabhb et al. do not classify Bronchoalveolar carcinoma, carcinoid tumors, carcinoma-in-situ, non-small cell cancer not otherwise specified and miscellaneous types of cancers, as is done in this study.

In this study we classified histological types of malignancy by differentiation too, as seen in figure 19. This is also an analysis method unique to
this study, not seen in literature. This analysis is done to provide a matrix of information about histological types of malignancies. Poorly differentiated carcinomas tend to be more aggressive in their clinical course and resistant to therapy. However, such an analysis is lacking in literature. In our study we see that 19/112 (17%) of the malignancies are poorly differentiated. Adenocarcinoma and non-small cell carcinoma being the most frequently occurring poorly differentiated types of malignancies.

Of the 653 TBBx procedures included in this study 360 (55.13%) are males. Ninety two of the 653 procedures (14.09%) are reported to be positive for malignancy. Of the 92 patients whose TBBx pathology report is positive for malignancy, 64 (69.6%) are male and 28 (30.4%) are female. Roth et al. report a similar gender distribution among the lung cancer population in their study, 60.9% males and 39.1% females (Roth, Hardie et al. 2008). Koh et al. included almost an equal number of males and females (19 and 18 respectively) in their study, but do not report the gender distribution of patients found to have lung cancer, since their study included patients with non malignant lesions (Koh, Tee et al. 2008).

Another analysis of this data reveals that 64/360 (17.78%) men that underwent TBBx are reported to be positive for malignancy. While 28/293 (9.56%) women that underwent TBBx are reported to be positive for malignancy. Such an analysis is not reported in literature. It is uncertain what this information suggests. Since the proportion of patients suspected to have lung cancer pre-procedure (122.653, 18.7%) is almost similar to the proportion of patients reported positive for lung cancer (92/653, 14.09%), this data could imply that men have a higher probability of being diagnosed with lung cancer than women, or that TBBx is done on women for other clinical conditions and not mainly lung cancer.
Histological type of malignancy among the two genders is shown in figure 20. Adenocarcinoma is the most commonly seen histological type of malignancy in both genders, and has an almost equal distribution 25/64, 39.1% men and 10/28, 35.7% females. Our data differs from the analysis reported by Belani et al. that women constitute a larger proportion of patients with adenocarcinoma than men (Belani, Marts et al. 2007).

A correlation between pre-procedure differential diagnosis and TBBx pathology report to estimate pre-test probability of diagnosing lung cancer is a unique method of analysis in this study and has not been reported in literature. Most of the studies in literature only focus on patients already diagnosed with lung cancer. Our study shows that pre-test probability of diagnosing lung cancer is highest (33/147, 22.45%) among patients clinically suspected to have lung cancer pre-procedure. However, the largest number of patients diagnosed to have lung cancer did not have a pre-procedure differential diagnosis specified (71/511 13.89%). Only one of 143 samples (0.7%) collected from patients not suspected to have lung cancer is reported positive for malignancy thus having the lowest pre-test probability.

Occurrence of cellular atypia is also unique in this study. This histological finding is not commonly reported in literature. It is considered important in this study because it contributes to the clinical picture and management of patients suspected to have lung cancer. For example if a young woman with history of never smoking undergoes TBBx for an incidental finding of a single small peripheral pulmonary nodule and the pathology report returns as some cellular atypia seen but not definitive of malignancy, this patient is likely to have low probability of having lung cancer is unlikely to undergo a repeat TBBx. She will most likely be followed up after 6 months to a year to detect any change in the lesion. However, if a 75 years old male patient with a history of smoking 20 packs for the last 40 years is found to have a similar small peripheral pulmonary nodule with a TBBx report of cellular atypia, the clinical suspicion of lung cancer in this
patient is higher and therefore he is more likely to undergo a repeat TBBx to rule out lung cancer.

Among the non-malignant findings reported in this study, chronic inflammation is the most commonly seen condition. Very few studies report non-malignant conditions along with occurrence of malignancy in their study. Koh et al. reported 4/18 patients who had a positive diagnostic yield on endobronchial ultrasound guided TBBx to have fungal infection (Koh, Tee et al. 2008).

Data for special stains on TBBx tissue (table 12) suggests that H&E staining is frequently not specified. This is most likely misleading information because it is routine practice to do this staining and therefore is not regularly specifically mentioned in the report. However, all other special stains are meticulously reported. It is thus evident that special stains are not routinely or frequently done. GMS (174/801, 21.72%) and AFB (153/801, 19.1%) are the most commonly used special stains. One of the reasons for the rare use of special stains is that volume of tissue acquired via TBBx is very small (average= 0.04cm³), therefore after processing, fixing, embedding in paraffin wax and sectioning for slides, there is barely any tissue left for special stains. Another challenge with staining of TBBx tissue is that if a lesion of interest is seen in one section on H&E stain, it may not be seen in any other section therefore making it impractical to use special stain to better understand the characteristics of the lesion seen.

Validation of TBBx pathology results with lung resection reports as gold standard provide an estimate of the sensitivity and specificity of the performance of TBBx in diagnosing lung cancer. It is important to note here that only patients in early stages of the disease (stage I, II and III a) and otherwise in good physical condition are eligible for surgical resection. The design of this study excludes a large population of lung cancer patients since about 70% of patients are diagnosed in advanced stages of disease (stage III and IV). However, since an
effort is being made to detect patients in early stage, there is an advantage in segregating this group from the large majority and study lesions characteristics that may be unique to this group. The method design of this study using lung resection pathology results as gold standard to validate performance of TBBx, includes mainly patients in early stage of disease thus focuses on the lesion characteristics in this group. From the data acquired in aim 3 TBBx has a sensitivity of 33.33% and specificity of 96.7%. Since this method of analysis is unique to this study it is impractical to compare with published literature, which usually includes all patients diagnosed with lung cancer.

Overall a large number of TBBxs are false negative (28/72, 38.89%) while only 14/72 (19.44%) are true positive. This outcome is much lower than the biopsy diagnostic yield reported by Roth et al. of 60.7% (Roth, Hardie et al. 2008). Roth et al. determined diagnostic yield for TBBx from a patient population that is proven to have malignancy by “all possible methods”. Therefore their study did not have a true negative group (Roth, Hardie et al. 2008). Koh et al. assessed the yield of diagnosing lung lesions using endobronchial ultrasound guided transbronchial lung biopsy (EBUS-TBLB) and report a diagnostic yield of 62% in a population of 38 patients (Koh, Tee et al. 2008). They however, do not compare the yield of these processes without EBUS guidance. Also, their definition of yield includes non-malignant findings.

A high false negative yield is clearly a gap in the process of care that needs to be remedied in order to increase lung cancer detection at an early stage, thus improve patient survival. Patient survival is better with early diagnosis. In order to detect patients at an early stage the diagnostic yield of TBBx, which is often the first step for histological confirmation of lung cancer, needs to improve significantly. Efforts are ongoing towards this goal. Endobronchial ultrasound guided TBBx (Koh, Tee et al. 2008), virtual navigation system (VNS) guided TBBx (Higgins, Ramaswamy et al. 1998; Kiraly, Helferty et al. 2004; Dolina, Cornish et al. 2008; Higgins, Helferty et al. 2008),
electromagnetic guided TBBx (Gildea, Mazzone et al. 2006), and fluorescent TBBx (Gilbert, Luketich et al. 2004; Stanzel 2004) are some of the advances being made in the field to improve TBBx diagnostic yield.

Other test characteristics evaluated in this study are:
False Negative Rate (FNR) = 0.67
False Positive Rate (FPR) = 0.03
Predictive Value Positive (PVP) = 0.93
Predictive Value Negative (PVN) = 0.51

Similar test outcome analysis have not been reported for TBBx, but plethora of information is available for transbronchial needle aspiration (TBNA) another technique commonly employed during bronchoscopy to obtain lymph node tissue for cytological and histological evaluation. TBNA is mainly useful for staging purposes. Therefore cannot be compared with TBBx.

Digital CT images offer an opportunity to include key CT images in pulmonologists’ reports and to use them for bronchoscopy planning and benchmarking. This is also helpful in developing simple valid measures to benchmark bronchoscopy performance quality, and to have readily available, accurate diagnostic summaries to ensure quality patient hand-offs. Although digital CT scan images were not available for the first half of the study period, they became increasingly available for the last three years of this study with an overall availability of 298/653 (45.64%).

The simple CT image classification scheme used in this study shows feasibility of establishing a valid correlation between lesion characteristic and pathology report outcome but could be improved with use of an image dictionary and further refinement of the classification guidelines. Stratification by lesion size is important for prospective outcome studies of image guidance systems.
Data in figure 24 suggests that focal lesions are biopsied more frequently (180/333, 54.05%) than diffuse lesions (153/333, 45.95%), and among focal ROIs lesions between 1-3 cm in short axis are biopsied most frequently (69/180, 38.33%), although lesions larger than 3 cm in size are biopsied almost as frequently too (65/180, 36.11%). However, when correlating with pathology reports it is seen that 12/69 (17.39%) ROIs 1-3 cm are reported to be positive for malignancy (figure 25). Whereas, 24/65 (36.92%) of ROIs larger than 3 cm are reported to be malignant. This suggests that lesions more than 3cm in size are most frequently found to be positive for malignancy. But it also raises the question as to how many lesions less than 3 cm are probably not accurately biopsied and therefore missed during TBBx, thus contributing to a delay in diagnosis. If this is true most of these lesions should fall into the false negative category. As seen in table 14, this is true. The highest false negative rate is seen in the focal ROIs 1-3cm in size. This clearly identifies the need to improve accuracy of acquiring tissue from smaller lesions in order to improve TBBx diagnostic yield for detecting lung cancer.

Among diffuse ROIs, lesions involving less than 25% of the lobe are most frequently biopsied (73/153, 47.71%). Highest diagnostic yield for malignancy is reported in diffuse lesions involving more than 50% of the lobe (8/43, 18.6%). The highest number of malignant samples from diffuse lesions are in the <25% of lobe category (9/73, 12.33%). An assessment to evaluate if smaller lesions are missed or inaccurately diagnosed by TBBx (false negative) demonstrates this to be not true. Most of the diffuse ROIs involving <25% of the lobe (5/14, 35.71%) are true negative (table 14).

The true negative group consists mainly of diffuse lesions involving <25% of the lobe and biopsied more frequently from the right lung. Sixty percent (18/30) of sites biopsied from the 29 cases in this group are congruent, suggesting the validity of outcome data.
In the true positive group only 3/14 (21.43%) digital CT scan images are available making it difficult to draw any conclusions about lesions characteristics. Two of these are focal ROIs 1-3 cm in size and one each is focal ROI >3cm and diffuse ROI involving <25% of lobe. Since data for only 3 digital CT scans is available for this group, any comparison with Roth et al. study is difficult (Roth, Hardie et al. 2008). They established that diagnostic yield is directly related to tumor size. A large number (10/15 samples from 14 cases, 66.67%) of the true positive cases are adenocarcinomas. However only 7/14 (50%) show congruency between site biopsied and site resected. A largest number of these samples are resected (6/14, 42.86%) from the right upper lobe. This is important to note because clinically this is the most challenging lobe to access during bronchoscopy due to its anatomy. However, 4/6 (66.67%) of these samples demonstrate congruency.

The one sample in the false positive group is reported to be a carcinoma-in-situ, squamous cell carcinoma. Digital CT scan for this patient is not available. This case is interesting because it almost suggests that the TBBx itself may have served as a treatment and excised the malignancy which may be small in size. However, it does raise the question why this patient underwent lung resection at a later date. Without these details it is inconclusive to make any suggestions from these data.

The most critical information drawn from this study is characteristics of lesions from the false negative group. This is critical because it identifies malignant lesion characteristics that are missed or inaccurately diagnosed during TBBx especially in patients at an early stage of lung cancer, therefore provides guidance for future direction. Digital CT scans are available for 13/28 (46.43%) of the cases in this category. Six of these 13 (46.15%) lesions are focal ROIs 1-3 cm in size. These data quantify lesions characteristics more accurately than the design used by Roth et al. (Roth, Hardie et al. 2008). Although they show that
tumor size is a predictor of diagnostic yield for biopsy, they did not study the characteristics of false negative cases.

Lung resection histological reports from these 28 cases showed that 15 of them are adenocarcinomas. This is important because well-differentiated adenocarcinoma may be difficult to diagnose, especially if examined by a pathologist with limited experience. This may be part of the reason why the TBBx reports are negative for malignancy. The right lung is biopsied and resected more frequently in this group too. Congruency is seen in 18/28 (64.29%) of cases, thus validates the strength and outcome of data presented.

It is also important to note that 3 cases that are reported to be indeterminate on TBBx, and thus categorized as test negative, all 3 are reported to be malignant on lung resection. One is histologically categorized as other type of cancer and has a focal lesion >3cm on CT scan. The other 2 are adenocarcinomas, one of which is well differentiated and CT scan image shows a focal lesion 1-3 cm in size. For the third patient adenocarcinoma differentiation is not specified and digital CT scan image is not available.

This method of data analysis and characterization of lesion as seen on CT scan with respect to TBBx pathology report correlated with lung resection pathology report is unique to this study. Although Roth et al. evaluated diagnostic yield with relation to tumor size, distance from carina, endobronchial visibility and lobe, they did not establish any correlation with histopathological type of malignancy, nor do they study characteristics of cases that are false negative. Our study fills this critical gap in literature.

Specific aim 6 evaluates process of care from a different perspective, time to diagnosis and time to treatment. This is important especially when studying a disease like lung cancer which is not only diagnosed at a late stage but kills a large number of these patients within one year of diagnosis. Therefore timing of
care is critical to identify gaps where delay may be occurring and needs to be remedied in order to improve patient survival.

Figure 40 shows that lung resection data are available consistently throughout the period of study therefore the outcome is probably not influenced by any changes in practice over the period of the study. Digital CT scan repository, on the other hand, was established in late 2002 and therefore the data acquired from this source is representative of the practice in the later half of the study period (figure 41). Availability with time is established to rule out changes in practice with time that may influence the TDx and TRX. Since lung resection data is almost equally distributed across all time periods of this study it can be ruled out to have a significant effect on TRx. Digital CT scan data however reflect TDx during the latter 3 years of the study and does not reflect TDx for the first 3 years of the study. However, it is safe to assume that it would not be significantly different.

The median time interval between digital CT scan acquisition and TBBx is 7 days pre-TBBx. However, the median time interval between TBBx and lung resection is 63.5 days (more than 2 months), with most patients in the 31 to 90 days category. This large time interval may reflect any of the following situations. First, due to the high false negative rate and low diagnostic yield of TBBx, it is possible that many of these patients had to undergo a repeat TBBx before a final diagnosis of lung cancer is made. Secondly, even if and when a final diagnosis of lung cancer is made, a patient may undergo repeat bronchoscopy for staging purposes to determine if surgical resection is the treatment of choice. Third, if a final diagnosis of lung cancer is made and the patient is determined to be in an early stage of disease (stages I, II, IIIa) when surgical resection is one of the treatment options, patients in stage II may need chemotherapy prior to surgery. Patients in stage IIIa may require chemotherapy and radiation prior to surgery. Finally, the patient needs to be evaluated for physical fitness to undergo such a major surgical procedure, especially if the patient is elderly, has poor lung
function from many years of smoking, and has undergone chemo and or radiation therapy. All these factors may contribute to the differences seen in TDx and TRx (figure 46).

Figure 47 shows the differences in TDx and TRx for individual groups of patients (true positive, false positive, true negative, false negative). The highest variation in both TDx and TRx is seen in the true negative group. This could be due to the difference in indication for surgery in this group, and the lack of urgency to do surgery since these are probably not as life threatening conditions as lung cancer.

Although variations are seen between each group for both TDx and TRx, none of these are found to be statistically significant. Thus implying that the process of care is not significantly different or delayed for any particular group. However, due to the small number of patients in each of these groups (4 to 14) the power of such statistical analysis is questionable. Similar study and analysis with larger number of patients is needed to confirm this outcome.

Schultz et al. report a median time to treatment of 63 days (Schultz, Powell et al. 2008) almost exactly the same as that reported in our study. However their definition of time to treatment included time interval between first suspicious chest X-ray to first treatment. If we had similar criteria for defining time to treatment, our median TRx would likely be longer.

Olsson et al. report that median time to diagnosis and treatment are longer than recommended (Olsson, Schultz et al. 2008). However, the factors considered in their study include teaching hospital, curative (vs. palliative) radiotherapy, the presence of comorbidities, and initial referral to a non-respiratory physician. Our study evaluates factors including CT scan lesion characteristics, TBBx pathology report and lung resection pathology report, categorization of patients into groups of true and false positive and negative.
The design and evaluation of TDx and TRx with regard to CT scan image characteristic and pathological outcome is also unique to this study. Schultz et al. recently presented data from a survey suggesting beliefs among experts about natural history of malignant solitary pulmonary nodule with regard to tumor growth, progression and prognosis (Schultz, Silvestri et al. 2008). They concluded that differences in experts’ beliefs may contribute to the variations seen in clinical practice. Our study provides tangible data along these lines and not beliefs. Our data suggest that at the Hershey Medical Center, there is no significant difference in management of patients irrespective of weather they are true positive or false negative. This is encouraging information. However, more studies need to be conducted including multiple centers and larger populations to confirm our findings.
VI: Conclusions

In conclusion this study provides the following information:

1) Proactive involvement of numerous medical disciplines is essential to build research projects for quality assurance, process of care assessment, and critical evaluation of diagnostic techniques used in current practice.

2) Electronic datasets offer an excellent resource for data acquisition in order to evaluate numerous aspects involved in process of care.

3) The diagnosing pathologist is not provided with critical clinical information including:
   - Clinical history
   - Smoking history
   - Occupational history
   - Differential Diagnosis
   - Pulmonary lesion characteristics
   - Exact site of biopsy.

4) Bronchoscopy acquired TBBx is feasible procedure chosen for performance evaluation in diagnosing lung cancer at the Hershey Medical Center since it is frequently done 653 procedures in 5 years 10 months.

5) Lung cancer is the most frequently mentioned pre-procedure differential diagnosis for TBBx.

6) The average aggregate volume of tissue acquired during TBBx is 0.048 cm$^3$. The average volume of TBBx samples reported positive for lung cancer by pathologist (0.061 cm$^3$) is significantly more than the average volume of samples reported negative for malignancy (0.047 cm$^3$).

7) Adenocarcinoma is the most frequently seen histological type of malignancy in TBBx pathology reports.

8) The pre-test probability of a positive pathology report for lung cancer is higher (22.45%) among samples obtained from patients clinically suspected to have lung cancer, intermediate (13.89%) from samples collected from patients
whose differential diagnosis is not specified, and least (0.7%) among samples from patients not suspected to have lung cancer.

9) Validation of TBBx performance in diagnosing lung cancer against lung resection as gold standard provided the following outcomes:

- Sensitivity = 0.33
- Specificity = 0.97
- False Negative Rate (FNR) = 0.67
- False Positive Rate (FPR) = 0.03
- Predictive Value Positive (PVP) = 0.93
- Predictive Value Negative (PVN) = 0.51

10) Digital CT scan acquired lesion characteristics indicate that focal ROIs 1-3 cm in size are most frequently biopsied, with focal ROIs >3cm biopsied with almost similar frequency. Among diffuse ROIs lesions involving <25% of the lobe are most frequently biopsied.

11) Malignancy is most frequently reported for focal lesions >3cm in size. Among diffuse ROIs malignancy is most frequently reported for lesions involving >50% of the lobe.

12) True negative cases frequently have diffuse lesions involving <25% of the lobe and most frequently occur in the right lower lobe.

13) True positive cases are most frequently adenocarcinoma and consist of focal lesions 1-3 cm in size. Most of these lesions are seen in the right upper lobe.

14) One false positive case is reported to be carcinoma-in-situ, squamous cell carcinoma on TBBx and found to be negative for malignancy on resection. Digital CT scan information on this case is not available. This lesion is biopsied from the left upper lobe which is also the lobe that is resected.

15) Twenty eight cases are categorized as false negative in this study. Most of these are focal lesions 1-3 cm in size. Adenocarcinoma is the most frequent histological type of malignancy reported in this group. Right upper lobe is the most frequent site of the lesions biopsied and resected.

16) The median time to diagnosis is 7 days pre-TBBx (range 0-57 days).
17) The median time to treatment is 63.5 days post-TBBx (range 4-1493 days).

18) When analyzing TDx and TRx between groups, maximum variation is seen in the true negative group. However there is no statistical difference in TDx and TRx between any of the four groups. However, since the total number of patients in each group is small (4-14) similar analysis on a larger group is needed to confirm this outcome.

Molecular markers hold the promise to engineer innovative methods to manage lung cancer patients. They can be critical in understanding tumor biology in different patient populations and thus in designing treatment regimes tailored for specific tumor and patient profiles. However, their full potential can only be harnessed if the desired tissue is obtained in a timely, accurate, efficient manner along with sufficient clinical data provided to the diagnosing clinical pathologist. The results of this study suggest that the current practice of TBBx in diagnosing lung cancer provides insufficient clinical data to the diagnosing pathologist, has low sensitivity and high false negative rate. Therefore is not ready to incorporate molecular markers into routine practice.

A molecular medicine program for lung cancer translational research requires development of updated clinical tools used for the accurate collection of diagnostic samples and issuing clinical reports. Navigations systems can be utilized in order to guide the bronchoscopist to the exact site of tumor; fluorescent markers can be used to accurately identify tissue to be biopsied. If tissue can thus be obtained in a more efficient and reliable manner then molecular markers will be useful in identifying tumor profiles and compared with characteristics of adjacent normal tissue from same patient or normal tissue obtained from patients without lung cancer.
VII: Applications, Implications and Future Directions

1) Other electronic datasets available at the Hershey Medical Center including CORI, Oncore, carcioaccess, and PSHCI tissue bank can be utilized for similar studies to evaluate quality, performance and process of care for any clinical condition.

2) Other stages and procedures in the process of care for lung cancer patients may be evaluated similarly, including TBNA, bronchial washing, brushing, BAL, trans-thoracic needle aspiration, chemotherapy, radiation therapy or combination of these.

3) The epidemiological codebook developed in this study may be used to design an intake check list with clinical information to be provided to the diagnosing pathologist to over come the lack of information transfer from the bronchoscopist.

4) A prospective study that provides comprehensive clinical information to the diagnosing pathologist may provide information on impact of the current lack of information transfer to the pathologist.

5) This study demonstrates the feasibility of conducting a performance validation study. A similar prospective study with all essential clinical information and exact site of biopsy and resection meticulously documents can be done to confirm the outcomes of this study.

6) Similar study design can be used to evaluate performance of other procedures or diagnostic yields of other clinical conditions including sarcoidosis, tuberculosis, interstitial lung fibrosis, cystic fibrosis etc.

7) This study only used lung resection pathology report as gold standard to assess performance of TBBx in diagnosing lung cancer, thus limited the population included to the patient with early stage cancer. Including other methods of confirming lung cancer clinically will include patients in all stages of the disease and provide a wider assessment of TBBx in diagnosing lung cancer.
8) The current study did not have data on stage of disease. Acquiring stage of disease will provide a different perspective on lesion characteristics and pathological report in diagnosing lung cancer for different stages of disease.

9) Digital CT scan data acquired from the repository is limited because this resource became available only after December 2002. In order to get more information on these patients another source may be utilized to confirm the outcomes of this study. Alternatively a prospective study may also serve the same purpose.

10) Time to diagnosis and time to treatment in this study consists of a small population (4-14) and is restricted to patients who underwent surgical resection, therefore outcomes are inconclusive. A similar study including a larger population of lung cancer patients at multiple centers may provide more conclusive data.

11) Since the long-term goal of this project is to incorporate molecular markers in the standard process of care for patients with lung cancer, a next step could be to assess the potential of a novel biomarker as a diagnostic and prognostic tool in lung cancer. To this effect, the author has submitted a proposal for a retrospective study on lung cancer tissue to the Penn State Hershey Cancer Institute. This proposal has been accepted and approved by the scientific review committee. A copy of this proposal titled “Assessing the Potential of Sphingosine Kinase-1 as a Diagnostic and Prognostic Marker for Lung Cancer: a Pilot Study” is provided in appendix B.
Appendix A

An example of TBBx pathology report
Surgical

*Final Report*

HMCSP
Penn State Milton S. Hershey Medical Center
Department of Pathology and Laboratory Medicine
500 University Drive
Hershey, PA 17033
Phone: 717-531-8245 Fax: 717-531-7741

SURGICAL PATHOLOGY REPORT

Patient Name: 
MRN#: 
Physician: Milos Tucakovic Service Date: 9/3/02

Accession #: S02-22028 Billing #: 000002705994

FINAL PATHOLOGIC DIAGNOSIS:
Lung, right lower, biopsy:
- No pathologic diagnosis

***Electronically Signed Out***
Siloo B. Kapadia, M.D.

CLINICAL HISTORY
Patient with cough, hilar and mediastinal adenopathy. Rule out sarcomiosis. Previous bronch 3/00 non-diagnostic. Specimen site:
right lower lobe. Diagnosis Code 796.6.

SPECIMEN(S) RECEIVED
A: "Right lower lobe"

GROSS DESCRIPTION
A. The specimen is labeled with patient name and number and designated "Right lower lobe": Received in formalin are multiple tiny fragments of soft pale tan tissue measuring 0.4x0.4x0.2 cm. in aggregate. Entirely submitted. (ED/LEC)

MICROSCOPIC DESCRIPTION
Sections show benign lung parenchyma without any evidence of granulomatous inflammation or tumor. A thick-walled vessel is seen in one of the fragments.

SUMMARY OF SECTIONS

Printed by: Sinha, Anju
Printed on: 03/22/2006 13:01

Page 1 of 2
(Continued)
Appendix B

Assessing the Potential of Sphingoskine Kinase-1 as a Diagnostic and Prognostic Marker for Lung Cancer: a Pilot Study

R. Bascom M.D., M.P.H., A. Sinha M.B.B.S., M.S., J. Yung Ph.D., D. Zander M.D., S. Safaee M.D., David Mauger Ph.D., Shantu Amin Ph.D.

I. Introduction:

Lung Cancer Burden: Lung cancer continues to be the leading cause of death by cancer in the United States of America. Lung cancer is estimated to be the cause for 29% of all cancer deaths, with about 213,380 estimated new cases and 160,390 estimated deaths in 2007 (Jemal, Siegel et al. 2007). One of the challenges is the rarity of diagnosing lung cancer at an early stage when surgical resection and treatment can result in longer survival of the patients. Current 5-year survival rate of patients with lung cancer is only 16% (Jemal, Siegel et al. 2007). For stages III b and IV disease chemotherapy is the treatment of choice but has a poor initial response rate of only 30%. Furthermore, the initial responses last for only 4-6 months, after which patients typically develop resistance (Rosell, Cecere et al. 2006). Thus, there is a pressing need to identify therapeutic targets in order to develop more efficient methods of diagnosis, prognosis and treatment in order to enhance response and survival in lung cancer patients.

Sphingolipid Metabolism: Sphingolipid metabolism has been known to influence cell survival, proliferation and apoptosis. Sphingomyelin, present at the cell membrane, is converted to ceramide by the enzyme sphingomyelinase (fig1). Ceramide is converted to sphingosine by the enzyme ceramidase. Ceramide and sphingosine have been implicated in triggering apoptosis by up-regulating the death kinase p-38 and down-regulating the survival MAPK (mitogen activated protein kinase) ERK1/2 (extracellular signal-regulated kinase) (Haimovitz-
Friedman, Kolesnick et al. 1997; Basu and Kolesnick 1998; Chan and Goldkorn 2000; Zhang, Liu et al. 2007). The enzyme sphingosine kinase (SK) converts sphingosine to sphingosine-1-phosphate (S1P). S1P is the only substrate in this metabolic pathway that stimulates cell survival and proliferation by enhancing ERK1/2 and inhibiting p-38. SK enzyme has two known isoforms 1 and 2. SK1 is identified as an oncogene, since its up-regulation has been associated with carcinogenesis (Xia, Gamble et al. 2000).

**Review of Sphingolipid Pathway**

![Diagram of Sphingolipid Pathway](image)

*Figure 1: Sphingosine Kinase (SK) Regulates Apoptosis and Proliferation Balance*

**Sphingosine Kinase as a Possible Diagnostic Marker for Cancer (Aim 1):** Sphingolipid metabolism has been examined and found to be altered in many cancers. For example, French et al and Johnson et al showed that various human tumors demonstrated about a two fold higher expression of SK1 mRNA than adjacent normal tissue from same patients (French, Schrecengost et al. 2003; Johnson, Johnson et al. 2005).
Limited research has been done to determine SK1 levels in lung cancer tissue. Johnson K.R. et al. demonstrated by immunohistochemistry that SK expression was higher in non-small-cell human lung cancer tissue than in adjacent normal lung tissue from same subjects (Johnson, Johnson et al. 2005). They also demonstrated, by $^{32}$P labeled SK1 cDNA probe, that SK1 mRNA expression was markedly elevated in various human carcinomatous tissues compared to the adjacent normal tissue from the same patient. Lung cancer tissue demonstrated about two fold increase in SK1 mRNA levels compared to adjacent normal lung tissue from same subjects.

Many areas to characterize the potential role of SK in lung cancer remain unexplored. For example, although SK1 mRNA levels were determined, protein level was not evaluated nor was SK enzyme activity assessed in lung cancer tissue. Furthermore, there is no literature comparing SK1 expression in apparently normal lung tissue from patients with lung cancer and normal lung tissue from subjects without lung cancer. Spatial patterns of SK1 in normal tissue adjacent to lung cancer have also not been described as a function of distance from the pathologic margin or as a function of orientation (hilar side vs. pleural side). Finally, SK1 has not been evaluated for its potential as a diagnostic marker using receiver operating characteristics (ROC) analysis. This research project aims to explore these areas.

**Sphingosine Kinase as a Possible Prognostic Marker for Sensitivity to Chemotherapy (Aim 2):** Research on prostate cancer cell (PCC) lines suggests this as a promising area of investigation for lung cancer. Akao et al. chose PCC lines previously characterized as being sensitive or resistant to chemotherapy drug induced cell death *in vitro*. They demonstrated that in comparison to sensitive PCC, resistant PCC expressed higher SK mRNA, protein and enzyme activity (Akao, Banno et al. 2006). Resistant PCC also expressed higher S1P receptors 1, 2 and 3 mRNA levels.
SK has also been identified to play a significant role in the *in vitro* response of PCC to chemotherapy. Camptothecin (CPT) is a commonly used chemotherapeutic agent to treat prostate cancer. Resistant PCC demonstrated an increase in SK enzyme activity, expression and mRNA levels in response to CPT treatment in a time and dose dependent manner. Similarly, mRNA levels for S1P receptors 1, 2 and 3 also increased in response to CPT (Akao, Banno et al. 2006). These findings strongly suggest the significance of sphingolipid metabolism in determining sensitivity of cancer cells to chemotherapy.

Pchejetski, D. *et al.* elegantly elucidated the role of SK1 as an indicator of PCC sensitivity (Pchejetski, Golzio et al. 2005). They demonstrated that differential sensitivity of PCC to chemotherapy was paralleled by difference in expression of SK1 and ceramide/S1P ratio in response to chemotherapeutic agents. PCC demonstrated higher ceramide/S1P ratio and lower SK1 expression in response to chemotherapeutic agents the cells were more sensitive to. They further strengthened their hypothesis by demonstrating that inhibition of SK1 mimicked the effect of antineoplastic agents and that overexpression of SK1 rendered sensitive cells resistant to chemotherapy. They thus robustly established that ceramide/S1P ratio can serve as a “sensitivity marker” for response to chemotherapy.

**Sphingolipid Metabolism and Therapeutic Agents:** Nave V. E. *et al.* demonstrated the role of sphingolipid metabolism in radiation-induced apoptosis in prostate cancer cells (PCC). They proved that pre-treatment with sphongomyelinase, which results in an increase in ceramide level, sensitizes the otherwise resistant LNCaP cells to γ-irradiation. Pre-treatment with sphingosine, SK inhibitor DMS and TNF-α also sensitized the LNCaP cells to γ-irradiation (Nava, Cuvillier et al. 2000).

French et al developed five SK inhibitors and treated mouse breast cancer models with three of these drugs. They demonstrated that these drugs had anti-
tumor properties and were well tolerated by intraperitoneal injections. One of their drugs, SKI-II, was also well tolerated orally and maintained its antineoplastic effect (French, Upson et al. 2006). It is noteworthy that these mice were not subjected to any conventional chemotherapy or irradiation.

Some researchers have recently demonstrated that treatment of human lung adenocarcinoma A549 cell line with exogenous ceramide can induce apoptosis (Newton, Hart et al. 2000; Kurinna, Tsao et al. 2004; Zhang, Liu et al. 2007). Min J. et al. showed that overexpression of S1P lyase enzyme in A549 human lung cancer cell line increased their sensitivity to cisplatin, carboplatin and doxorubicin, (Min, Van Veldhoven et al. 2005) further enhancing the hypothesis that sphingolipid metabolism may regulate the sensitivity of cells to specific chemotherapeutic agents.

Summary: Data are accumulating suggesting that sphingolipid metabolism is altered in various cancers and influences response to chemotherapy and prognosis. Because of the importance of lung cancer to overall cancer morbidity and mortality, we believe it is important to test the hypotheses that tissue SK1 enzyme level is a diagnostic and prognostic marker for lung cancer. The protocols described below will provide key pilot and feasibility data. In the long term, answers obtained by this line of research could contribute in improving the long-term survival of lung cancer patients. There is a continuing need for improved lung cancer diagnostic tools, approaches to individualized drug therapy, and to find sensitizing agents to improve the efficacy of existing radiation and chemotherapies.

II. Specific Aims:

1. To assess the potential of SK1 as a diagnostic marker of lung cancer.

   The hypotheses are:

   i) SK1 enzyme mRNA levels and protein expression are highest in human lung cancer tissue, next highest in adjacent normal lung tissue from same subjects.
(2cms from tumor margin), and lowest in distant normal tissue from same subjects (4cms from tumor margin). SK1 enzyme expression in tissue derived from lung cancer patient is correlated with distance and orientation from centre of tumor tissue.

ii) SK1 enzyme expression is higher in normal lung tissue from subjects with lung cancer than from subjects without lung cancer.

iii) SK1 expression is higher in lung cancer tissue than in normal tissue from patients without lung cancer. This aim has three parts:

a) To quantify SK1 enzyme mRNA level, protein expression and localization in surgical lung tissue samples.

b) To evaluate three possible confounding factors: histological type, stage at tissue sampling and tobacco exposure history. The hypothesis is that SK1 expression will be higher in tissues sampled at higher stages of the disease, higher histopathological grade e.g.; poorly differentiated vs. well differentiated, small cell vs. non small cell, and current tobacco smoking (cumulative >20 pack years) vs. current smoking (< 20 pack years), ex-or non-smoking status.

c) To evaluate the receiver operating characteristics of SK1 as biomarker of lung cancer using pathologic diagnosis as a gold standard.

2. To assess the potential of SK-1 as a prognostic marker of lung cancer chemotherapy response.

The hypothesis is that SK1 enzyme expression is higher in lung cancer tissue of patients with worse outcomes than in patients with better outcomes.

The current proposal aims to determine the feasibility of stratifying patients’ lung cancer outcomes based on a medical record review. In this pilot study we will review medical records to determine the availability of clinical information in three time frames (3-6 months post biopsy, 6-12 months post biopsy, and survival (time from diagnosis). We will evaluate 4 possible outcomes, local recurrence, lesion specific growth or regression, metastases and
mortality. Finally, we will consider treatment regimens (radiation, chemotherapy and/or resection).

A case control study will be proposed associating lung cancer prognosis and tissue SK1 level. For this; access to medical records of all lung cancer patients whose tissue is available in the tissue bank is requested at this time. A structured dataset will be created based on the medical records review of the demographic and clinical features of patients. These records will be used to develop several potential indices of clinical outcome and to define a distribution of clinical response. Cases and controls will be proposed, with the criteria for matching identified. This characterization will be performed blinded to the SK1 level in the patient’s tissue. Sample size calculation will then be done to select the sensitive and resistant patients. This will then be presented to the Tissue Bank Committee for their review and request for specific patient’s lung cancer tissue for the next part of this aim. At the completion of this project, we offer to contribute the results of this work to the tissue bank repository for use by other investigators.

III. Objectives and Methodology:

A Study Design:

Specific Aim I: Diagnostic marker: The approach will be to measure sphingosine kinase in a random sample of 30 cases of lung cancer from the PSCI tissue bank. Comparison samples will include adjacent normal tissue in the cancer samples, and 30 cases of normal lung tissue (Table 1). This will allow us to: 1) estimate the distribution of sphingosine kinase in the PSCI tissue bank lung cancer population, 2) evaluate any correlation, if present, of Sk1 expression as a function of distance and orientation from tumor, 3) evaluate the effect of three possible confounders on SK expression: histologic type, stage at tissue sampling and tobacco exposure history; 4), using pathologic diagnosis of lung
cancer as a gold standard to determine the receiver operating characteristics of sphingosine kinase tissue levels as a diagnostic marker of lung cancer.

Table 1: Overview of study design for specific aim 1: Evaluating Sphingosine Kinase-1 as a Diagnostic Marker for Lung Cancer.

<table>
<thead>
<tr>
<th>Specific Aim 1</th>
<th>Lung Tissue From Patients with Lung Cancer</th>
<th>Normal Lung Tissue From Patient Without Lung Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>SK1 mRNA and protein</td>
<td>Lung cancer tissue (n=30)</td>
<td>Adjacent normal lung tissue from same patient at 2cms and 4 cms from tumor margin. (n=30)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>normal lung tissue from patients without cancer (n=30)</td>
</tr>
</tbody>
</table>

Specific Aim II: Prognostic marker: The future aim will be to perform a case control study associating lung cancer prognosis and tissue sphingosine kinase level. We are not requesting tissue for this aim at this time. We are, however, requesting identification of the patients for medical record review as an initial phase of this aim (hereafter called part IIa). The approach to part IIa will be to create a structured dataset based on medical record review of the clinical features of patients from whom the tissue bank samples have been derived. We will use these records to develop several potential indices of clinical outcome and attempt to define a distribution of clinical response and outcome. We request access to all the lung cancer tissue donor names because of the range of outcomes we are considering. This characterization will be performed blinded to the sphingosine kinase level in the patient’s tissue. Our plan is to present these data in the future to the Tissue Bank Committee for their review. Dr. D. Mauger has advised us to defer that sample size calculation until we have performed the chart review and clinical classification.
B. Experiments:

1. To evaluate the potential of SK1 enzyme as a diagnostic marker for lung cancer by quantifying SK1 enzyme mRNA, protein expression and distribution in lung cancer tissue, adjacent and distant normal tissue from same subject, and normal lung tissue from subjects without lung cancer.

Detailed characterization of the spatial features of SK levels will be determined for n=30 samples when permitted by the available tissue blocks. Since it is a well established standard to resect at least 2cms of surrounding normal tissue (Ginsberg and Rubinstein 1995), we plan to evaluate any correlation of Sk1 expression as a function of distance and/or orientation with respect to tumor tissue. Distance (D) will be defined as the distance in centimeters from the pathologic tumor margin. Normal adjacent tissue will be sampled at D=2 cms and D=4 cms, if possible. Orientation (O) will be defined on an arbitrary x-y grid at four 90° intervals (as shown in fig 2). A two way ANOVA analysis will be used to determine whether orientation or distance influence SK levels.

i. RT PCR for SK mRNA from human lung cancer tissue, adjacent normal tissue of same subject (at distances 2cms and 4cms from tumor margin) and normal lung tissue of patient without lung cancer.

ii. Western blot analysis for SK enzyme to quantify protein expression in human lung cancer tissue, adjacent normal tissue of same subject (at distances 2cms and 4cms from tumor margin) and normal lung tissue of patient without lung cancer.

iii. Immunohistochemistry to co-stain for SK enzyme and various other cell-types like endothelial cells, epithelial cells and alveolar cells to identify distribution and localization of SK enzyme in various cells of lung tissue. Tissues stained will be lung cancer tissue, adjacent normal lung tissue from patients with lung
cancer (at distances 2cms and 4cms from tumor margin) and normal lung tissue from patients without lung cancer.

Figure 2: Diagram representing planned spatial characterization to determine SK1 expression as a function of distance and orientation to the tumor.

2. **To assess the potential of SK1 enzyme as a prognostic marker for lung cancer by collecting clinical data from medical records of patients with lung cancer whose tissues are available in the PSCI tissue bank.**

No tissue is requested at this point for aim 2. Patient data from medical records of lung cancer patients whose tissue samples are available in the tissue bank will be collected. A structured dataset will be developed to define clinical outcome and response to therapy. If case controls pairs can be established from the tumor bank data, then along with SK enzyme assay we plan to also compare ceramide.S1P ratios in the various tissues.
C. Specific Methods:

Aim 1: Figure 3 shows the approach to sample identification.

Figure 3: Algorithm showing approach for sample identification for Aim 1: Evaluating Sphingosine Kinase-1 as a Diagnostic Marker for Lung Cancer.

Aim 2: Figure 4 shows the approach for population identification:

Figure 4: Algorithm for Population identification of Patients for Clinical Characterization.

SK1 mRNA expression analysis: Protocol excerpted from (French, Schrecengost et al. 2003)*.
SK mRNA levels will be examined using the Cancer-Proﬁling Array from Clontech, which consists of 241 paired samples of cDNA made from the tumor and adjacent normal tissue from individual patients. The array will be hybridized with a 32P-labeled cDNA probe corresponding to a 1-kb fragment of the open reading frame of human SK. The membrane will then be washed to remove nonspeciﬁc binding, and bound probe will be visualized using a phosphorimager. The relative intensities of bound probe will be quantiﬁed using Pathways software, and the ratios of the relative intensities of tumor samples and the corresponding non-tumor samples will be determined (French, Schrecengost et al. 2003).

**SK1 protein expression by Western blot analysis:** Protocol excerpted from (Pchejetski, Golzio et al. 2005)*.

Mouse anti-FLAG (Sigma) and mouse anti-ß-actin (Sigma) will be used as primary antibodies. Proteins will be visualized by enhanced chemiluminescence detection system (Pierce, Brabieres, France) using anti-mouse horseradish peroxidase–conjugated IgG (Bio-Rad, Marnes La Coquette, France) (Pchejetski, Golzio et al. 2005).

**SK1 protein expression and localization by Immunohistochemistry:** Protocol excerpted from (Johnson, Johnson et al. 2005)*

Paraffin blocks of normal and tumor tissue from lung cancer patients will be obtained from the Pennsylvania State Cancer Institute Tissue Bank at Milton S. Hershey College of Medicine. After sectioning onto microscope slides, tissue sections will be deparaffinized in xylene and rehydrated in a series of ethanol dilutions. Sections will be incubated 10 min in 3% hydrogen peroxide to quench endogenous peroxidase. To improve antigen retrieval, sections will be incubated in Vector Antigen Unmasking Solution (Vector Laboratories Inc.; Burlingame, CA) for 30 min in a warm humidifier chamber. After washing in PBS, sections will be blocked for 30 min in 2% normal goat serum in PBS. Sections will then be incubated for 1 hr at room temperature with primary rabbit anti-hSK1 antibody (7
µg/ml in PBS) alone or combined with 1 µg/ml of immunizing synthetic peptide (incubated 1 hr before adding to the sections), followed by 30 min incubation with diluted biotinylated secondary antibody and then 30 min incubation with VECTASTAIN ABC Reagent (Vector Laboratories Inc.). Diaminobenzidine/H$_2$O$_2$ will be used as a substrate for the immunoperoxidase reaction. Sections will be lightly counterstained with hematoxylin, rehydrated, and mounted for analysis by bright field microscopy. In the case of immunoabsorbed antibody incubations, 1 µg of synthetic oligopeptide against which the antibody was raised against will be incubated with 1 ml of diluted rabbit anti-SK1 antibody (7 µg/ml in PBS) for 1 hr at room temperature (Johnson, Johnson et al. 2005).

D. Statistical Analysis:

For specific aim 1 analysis of variance will be evaluated whether tissue source is a determinant of SK levels. If so, then paired comparisons will be made: (Table 2); two-tailed, paired analysis will be conducted when comparing lung cancer tissue with adjacent normal lung tissue from same patient. (p=0.05 will be used as cut off for this primary outcome specified in advance.) When comparing normal lung tissue from patients with lung cancer and those without lung cancer; unpaired two-tailed test will be used to evaluate statistical significance in SK1 expression. Parametric tests will be chosen for normally distributed data and non-parametric tests for data that are not normally distributed. (p=0.025 will be used as cut off to account for multiple comparisons.)
Table 2: Overview of study groups and statistical analysis for Specific Aim 1.

<table>
<thead>
<tr>
<th>Specific Aim 1</th>
<th>Lung Tissue From Patients with Lung Cancer</th>
<th>Normal Lung Tissue From Patient Without Lung Cancer</th>
<th>Statistic</th>
<th>Parametric ?</th>
</tr>
</thead>
<tbody>
<tr>
<td>SK1 mRNA and protein level</td>
<td>Lung cancer tissue (n=30)</td>
<td>Adjacent normal lung tissue from same patient (n=30)</td>
<td>Paired two-tailed test</td>
<td>to be determined</td>
</tr>
<tr>
<td></td>
<td>Adjacent normal lung tissue from patient with lung cancer (n=30)</td>
<td>normal lung tissue from patients without lung cancer (n=30)</td>
<td>unpaired two-tailed test</td>
<td>to be determined</td>
</tr>
<tr>
<td></td>
<td>Lung cancer tissue (n=30)</td>
<td>normal lung tissue from patients without lung cancer (n=30)</td>
<td>unpaired two-tailed test</td>
<td>to be determined</td>
</tr>
</tbody>
</table>

Possible Outcomes and Interpretations:

Figure 5 demonstrates two possible outcomes for aim 1, showing two possible receiver operating characteristics (ROC) of SK1 as a good or poor diagnostic marker of lung cancer.
For aim 2 we plan to develop an approach to test the primary hypothesis that tumor sphingosine kinase level is a predictor of tumor growth or regression after chemotherapy treatment. The secondary hypothesis is that SK levels relate to metastatic potential and mortality. We will likely propose a case-control study, matching for key confounders such as age, tobacco history, stage at tissue sampling, histological type of disease and treatment regimen. We will use our review of the tissue bank clinical data set to assess the feasibility of using these outcomes and matching on these factors.

**IV. Tissues Requested:**

1) Lung cancer tissue (N=30): Fresh frozen and formalin fixed paraffin embedded.
2) Adjacent normal lung tissue from same subject (N=30): Fresh frozen and formalin fixed paraffin embedded.
3) Normal lung tissue from subjects without lung cancer (N=30): Fresh frozen and formalin fixed paraffin embedded.

V. Sample Size:

Review of data in literature (French, Schrecengost et al. 2003; Johnson, Johnson et al. 2005; French, Upson et al. 2006) suggests that mRNA levels of SK1 are higher in tumor tissue when compared to adjacent normal tissue from same subjects. Johnson et al had an n=20 and French et al had an n of between 14 and 50. A sample size of N=30 will provide 85% power to detect a difference of 0.5 random units between cancerous and adjacent tissue. In addition a sample size of 30 will support a 3-factor analysis of variance (ANOVA) using 10 observations per factor. (Table 3) We thus request for 30 lung cancer tissues, 30 adjacent normal lung tissues from same subjects and 30 normal lung tissues from subject without lung cancer.

Table 3: Sample Size Analysis (two tailed test)

<table>
<thead>
<tr>
<th>alpha</th>
<th>Mean $^1$</th>
<th>Mean $^2$</th>
<th>Power</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>2+/-0.75</td>
<td>1.5+/-0.5</td>
<td>0.9</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.7</td>
<td>21</td>
</tr>
</tbody>
</table>

Footnotes:

$^1$: Anticipated mean SK1 enzyme expression in lung cancer tissue (Johnson, Johnson et al. 2005)(Figure 3 A and B)

$^2$: Anticipated mean SK1 enzyme expression in adjacent normal lung tissue from same subject (Johnson, Johnson et al. 2005)(Figure 3 A)
VIII: References


* Figures and material excerpted from cited references are considered "fair use" based on the purpose, nature, amount, and market effect of the use.
Vita

Anju Sinha

The author is born and raised in India. She received her medical training and completed her residency in Ophthalmology from Jawaharlal Nehru Medical College, Dharwad University, Karnataka, India. Her dissertation as a resident was on the ocular manifestations in HIV positive and AIDS patients. She was nominated for the gold medal award for her academic excellence. After her residency she volunteered as an ophthalmologist for the National Blindness Prevention Program in India. Later she moved to Canada to join her husband. There she gained experience at the University of Toronto, Department of Neurology, Neuro-Ophthalmology Division, initially as a visual field technician and research associate and later as a clinical fellow and post-doctoral research fellow. She was successful in obtained funding from the Positive Action Fund, AIDS Bureau Ministry of Health and Vision and Science Research. Her research there was on ocular movement abnormalities to detect early HIV encephalopathy. She also pursued courses towards a Master's degree at the Medical Sciences Institute, University of Toronto. Amidst her work, she was blessed with her son, Sushruth, and she re-located to State College, Pennsylvania, United States, again to be with family and was a stay-at-home mother for the following three years. She joined Pennsylvania State University in fall 2003 in the Biomedical engineering program. She moved to Hershey, College of Medicine and joined the IBIOS program to pursue research on diabetic retinopathy. Her work led to a poster presentation at ARVO, San Diego, in May 2005. She opted for Molecular Medicine program as her major and continued research on lung cancer in the Division of Pulmonary Allergy and Critical Care Medicine. Her research work lead to two poster presentations at ATS, Toronto in 2008. During her years as a graduate student in Penn State she became aware of the need for equal opportunity based on race and gender. She accepted numerous invitations from the Dean’s Council on Diversity, Women’s Commission, Chaplin Residents and local schools as a guest speaker to share her religion, culture and experiences. She is also actively involved with Boy Scouts of America, Keystone council, Pack 201 from Hummelstown. She is currently the director of Medicontrivers India Ltd. and is on the medical advisory board for Axzons Healthcare Services, NY.