Theses and Dissertations

Fall 2010

Development of human lung query atlas

Weichen Gao
University of Iowa

Recommended Citation

DEVELOPMENT OF A HUMAN LUNG QUERY ATLAS

by

Weichen Gao

A thesis submitted in partial fulfillment of the
requirements for the Master of Science
degree in Electrical and Computer Engineering
in the Graduate College of
The University of Iowa

December 2010

Thesis Supervisor: Professor Gary E. Christensen

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

MASTER'S THESIS

_____

This is to certify that the Master's thesis of

Weichen Gao

has been approved by the Examining Committee for the
thesis requirement for the Master of Science degree in
Electrical and Computer Engineering at the December
2010 graduation.

Thesis Committee: _____
                  Gary E. Christensen, Thesis Supervisor

                  _____
                  Joseph M. Reinhardt

                  _____
                  Milan Sonka

                  _____
                  Eric A. Hoffman

ACKNOWLEDGEMENTS

I would like to thank everyone who made this work possible.

I would like to thank my advisor Prof. Gary Christensen for providing me the opportunity to work on this project and for his patient support and guidance in this work. I would like to say that I feel fortunate to be one of his graduate students.

I would like to acknowledge the support and help from VIDA Diagnostics for providing PW2 software in this project and thank Rosanna for helping me learn PW2 and for all her help. I would like to thank Prof. Joseph Reinhardt for his patient support in this project. I would like to thank Prof. Gary Christensen and Prof. Milan Sonka for providing a productive lab environment and facilities. I would like to thank Nathan Burnette for helping me develop the database and implement the user interaction with the database. I would like to thank Janice Cook-Granroth for her patient guidance in using the human data for the atlas. I would like to thank my lab mate Matineh Shaker for helping me develop this human lung atlas. A special thank to Dr. Eric Hoffman for allowing me to participant this project and providing the human lung data.

I would also like to thank my roommates Panfang and Yunye, my friends Yao, Xiayu, Yunlong, Xiaoxin, Bo, Qi, Chichi, Wei, Yuwei and others for their support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure

CHAPTER 1
INTRODUCTION

1.1 Motivation and Background

An atlas plays an important role in daily life. For example, road atlases are widely used when people travel to an unfamiliar city. In this application, the atlas provides an accurate and adequate collection of information that exhaustively describes the city roads by using illustrations, informative tables or textual matters [17] and it provides standardized nomenclature for universal usability. Similarly, the atlas of human anatomy acts as a standard reference frame defining a standard nomenclature and provides a common coordinate system for relating images, models and semantic information. It reports structural and functional observation from different individuals and experiments with respect to a standard spatial coordinate.

Although these book atlases have been historically useful in illustrating human body, they still have many limitations. For example, the illustrations in book atlas were usually drawn by hand or from the photos taken from a single individual's anatomical structures. This limits its function for analyzing population information.

So far, the significance of anatomy studies are always shown in the form of population anatomical structures. For example, researchers investigate the sex differences in diseases where large populations are required. Electronic atlases can incorporate more anatomical segmentations and provide more quantitative analysis such as the measurements of anatomies. In comparison, this makes book atlas function ineffi-

ciently and insufficiently. Therefore, the demand of building more accurate and fast electronic atlases has increased.

In this study, we built an electronic human lung query atlas, which allows users to query the individual or population airway and volume information based on age, gender, ethnicity and race. It provides quantitative analysis among comparisons of population to population, population to individual and individual to individual. This human query atlas has user-friendly graphic user interface connected to a back-end MySQL database. Updating subject information in the back-end database does not change the front-end interface and ensures the atlas keeps up-to-date.

1.2   Related work

Since the constructions of human anatomy atlas are related to image segmentation and registration, considerable works have been done in this field by using different segmentation and registration methods.

With regard to the human anatomy, brain atlas has become more mature than other anatomical atlases. For example, Greitz *et al.* [8] developed a computerized brain atlas with a standard nomenclature and provided its future applications; Mazziotta *et al.* [15] introduced a probabilistic atlas of the human brain and they continued to develop a four dimensional probabilistic atlas of the human brain about 6 years later [16].

Less attention has been paid to human lung atlas compared to the human brain atlas. The reasons for this may result from the development of imaging technology, such as MRI, and the immobile structure of the brain. However, the new technique

MRI seemed to have more advantages against CT, such no radiation exposure and volumetric imaging. MRI can provide high tissue contrast and detained anatomy in brain. Researchers turned to paid more attention on the brain atlases due to the better brain images. Lung images are usually obtained by CT which exposes patients to radiation during scanning and may increase the risk of malignancy. But this has been improved significantly by developing low-dose CT scanners. Besides, the complexity of lung structure and the movement during scanning may account for less popularity in lung research .

However, despite the above difficulties, the need for acquiring an accurate and comprehensive electronic lung atlas has increased considerably. For example, the diagnosis of lung diseases has involved with lung imaging techniques, such as building 3D lung modelings [12]. A lung atlas can be used for early detection of lung disease. The accurate segmentation of the lungs, lobes, airways, and blood vessels in a human lung can provide important information for lung analysis. A lung atlas is a useful tool for teaching purpose, for example, the online navigators. The e-Anatomy is developed by IMAIOS [4] and assists medical imaging professionals and students in understanding human anatomy. More than 3,000 images of human body are provided including various types of images, such as CT and MRI, and are divided into different modules by parts of human body with standard nomenclatures [7]. Figure 1.1 shows a screenshot of IMAIOS cross-sectional 2D image in the atlas of thorax. The names are labeled outside of lung image. The set of image in the left side provides a good navigation for the right side lung image. This figure was obtained with the permission

of IMAIOS.
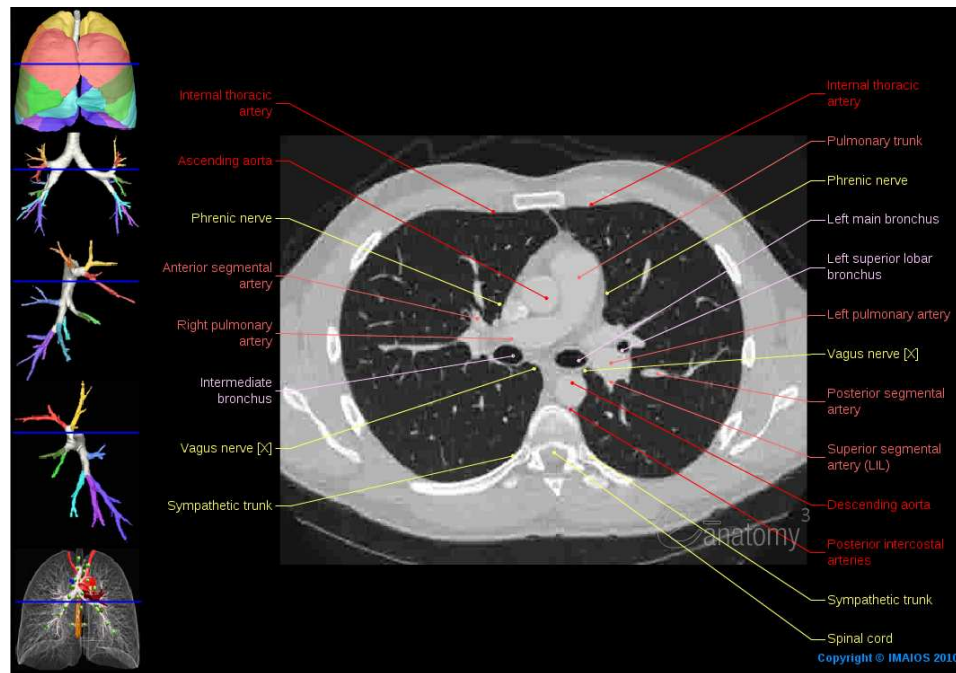
Figure 1.1: A screenshot of IMAIOS cross-sectional 2D image of human thorax.

These online navigators are a good tool for teaching and learning, but they also have limitations. For example, they only focus on a single individual anatomy and fail to provide population details, such as a mean shape and population variation. Most atlases cannot be used to detect abnormalities. Therefore, a more comprehensive lung atlas is needed.

Many kinds of lung models have been built with regards to lung shape, motion and diseases [9]. Li *et al.* established a normative human lung atlas by adopting inter-subject registration [13]. They applied a landmark and intensity based consistent

image registration method to average the transformations from one template to the rest of the database and deformed the template with the average transformation to obtain the atlas. Zhang *et al.* applied a simple lung atlas which only included lung boundaries and the lobar fissures to segment human lungs and lobes [26]. Zrimec *et al.* built a 3D modeling and visualization of the human lung [27]. Lee *et al.* presented a detailed review about the steps and algorithms of building lung models from lung images [12]. An advantage of an electronic lung atlas is that these models can be incorporated into the atlas.

1.3   Outline

This thesis consists the following parts.

Chapter 2 consists of five parts. The first part is to discuss the query atlas architecture (sections 2.1- 2.5). The atlas architecture contains five parts: the atlas database, atlas user interface, the connector ODBC, the programming and database query languages. The second part is to discuss the data acquisition. The third part is to introduce pulmonary structures and measurements. The fourth part is to introduce t-test performed in the atlas. The fifth part is to perform data accuracy assessment.

Chapter 3 consists of three parts. The first part is to show the atlas graphical user interfaces. The second part is to show the database demographics. The third part is to perform six studies using the human luang query atlas.

Chapter 4 consists of three parts. The first part is to discuss the advantages and disadvantages of using Microsoft Access in the atlas. The second part is to discuss how to make improvement and obtain more accurate quantitative analysis of human

lung. The third part is to discuss the future plan for the development of human lung atlas. Chapter 5 summarizes the work for building a human lung query atlas. Appendices A-D is to look up the tables and figures in Studies 2 to 5 in Chap 3.

CHAPTER 2
METHODS

This thesis describes the work that was done to construct a human lung query atlas. The human lung query atlas is a pulmonary application associated with a front-end graphic user interface (GUI) and a back-end database. It allows user to retrieve human pulmonary data and provides statistical analysis. The atlas database was implemented using MySQL and the GUI was created using Microsoft Access 2007. Open database connectivity (ODBC) was used to connect the database with GUI. Visual Basic Application (VBA) was used to construct queries on the forms. SQL was used to build atlas database and make queries.

2.1   Atlas Database

The atlas database is a normalized relational database containing airway and volume measurements of human lung. A relational database is a relational model in which a collection of tables are used to represent both data and the relationships among those data [19]. MySQL was used to implement the atlas database. MySQL is an open source relational database management system which uses structured query language (SQL) to query databases. In the atlas database, database schemas were structured to represent subject characteristics, DICOM standard, airway and volume names, airway and volume measurements. Database tables were used to store these corresponding information. Each table represents one relation in the database.

Figure 2.1 gives a relation example in the table format. The numbers in the

table are not the real numbers in the atlas in order to protect patient privacy. This example is used to illustrate what the relation table looks like in the atlas database. This table shows the relation of subject information with the relation name $subject\_t$. The letter t is short for table and can be used to indicate a relation in the table format. For instance, $dcmstudy\_t$ and $dcmseries\_t$ are tables for the DICOM study and series information, respectively.

The database schema for $subject\_t$ is $subject\_t = (\underline{subject\_id},\ patient\_id,$ $sex,\ ethnicity,\ race,\ smoker\_type,\ normality)$ with the primary key $subject\_id$. A database schema is the logical design of a relation. Each item in $subject\_t$ represents an attribute (column name) in the table. In each relational table, a primary key is chosen to uniquely identify each record in the table. In the $subject\_t$ schema, $subject\_id$ was chosen to be the primary key. It can uniquely identify $subject\_t$'s other attributes, such as sex, normality. Similarly, $dcmstudy\_id$, $dcmseries\_id$, $airway\_id$ and $airway\_avg\_id$ were chosen to be the primary keys for relations $dcmstudy\_t$, $dcmseries\_t$, $airway\_t$ and $airway\_avg\_measurements\_t$, respectively. We usually use a underline to represent the primary key in the relation schema.

The atlas database consists of 21 relations (tables) for airway and volume analysis. For airway analysis, we used five relations to construct the airway relation: $subject\_t$, $dcmstudy\_t$, $dcmseries\_t$, $airway\_t$ and $airway\_avg\_measurements\_t$. $subject\_t$ stores the subject information which may not change (see Figure 2.1). $dcmstudy\_t$ stores the subject DICOM information which may change along with the time, such as height, weight, age and body mass index (BMI) (see Figure 2.2). $dcmseries\_t$

subject_id	patient_id	sex	ethnicity	race	smoker_type	normality
100	H-100	M	Not Hispanic or Latino	White	non-smoker	normal
101	H-101	F	Hispanic or Latino	White	non-smoker	normal
102	H-102	M	Not Hispanic or Latino	White	smoker	abnormal
103	H-103	F	Not Hispanic or Latino	White	non-smoker	normal
104	H-104	M	Hispanic or Latino	White	smoker	normal
105	H-105	F	Not Hispanic or Latino	Black/African American	non-smoker	abnormal
106	H-106	M	Not Hispanic or Latino	Native Hawaiian/Pacific	smoker	normal
107	H-107	F	Not Hispanic or Latino	Asian	non-smoker	normal
108	H-108	M	Not Hispanic or Latino	American Indian	smoker	normal
109	H-109	F	Not Hispanic or Latino	Unknow or not reported	non-smoker	normal

Figure 2.1: A relation example: *subject_t*.

stores other DICOM information, such as lung capacity and reconstruction kernel (convolution kernel). *airway_t* stores the airway anatomical names, such as trachea. *airway_avg_measurements_t* stores all the airway measurements, such as average inner area.

Another relation *dcmstudy_t* is shown in Figure 2.2. The numbers and schema are not the real data in the atlas for the patient privacy and simplicity. The database schema is *dcmstudy_t* = (*dcmstudy_id*, *subject_id*, *study_date*, *dcmstudy_name*, *modality*, *height*, *weight*, *age*). *dcmstudy_id* is the primary key in this relation.

In the relational database, an attribute in one relation r_1 may appear as the primary key in another relation r_2. This attribute is called a foreign key from r_1, referencing r_2. The relation r_1 is called the referencing relation of the foreign key dependency, and r_2 is called the referenced relation of the foreign key. Foreign key is important in joining relations. Essentially, an inner join is used to combine information from two relations based on a common value. Usually, this common value is the

dcmstudy_id	subject_id	study_date	dcmstudy_name	modality	height	weight	age
800	100	12/1/2000	p2brp-10	CT	1.67	65	20
801	101	12/2/2000	p2brp-11	CT	1.68	55	23
802	102	12/2/2000	p2brp-12	CT	1.8	50	34
803	103	12/2/2000	p2brp-13	CT	1.65	52	56
804	104	12/2/2000	p2brp-14	CT	1.75	90	35
805	105	12/3/2000	p2brp-15	CT	1.78	67	54
806	106	12/2/2000	p2brp-16	CT	1.56	70	26
807	107	12/2/2000	p2brp-17	CT	1.78	78	29
808	108	12/6/2000	p2brp-18	CT	1.73	50	30
809	109	12/9/2000	p2brp-19	CT	1.65	77	45

Figure 2.2: A relation example: *dcmstudy_t*.

foreign key.

For example, suppose we want to know the age of the subject with *subject_id* = 100 in *subject_t*. However, relation *subject_t* does not provide the age information. We can only find the age information in relation *dcmstudy_t*. Therefore, we need to combine these two relations together. Notice that *subject_id* is the primary key in relation *subject_t* and it also appears in relation *dcmstudy_t*. Therefore, *subject_id* is a foreign key from relation *dcmstudy_t*, referencing relation *subject_t*. The relation *dcmstudy_t* is the referencing relation of the foreign key dependency, and the relation *subject_t* is the referenced relation of the foreign key. These two relations can be joined using *subject_id*. The details about joining tables is discussed in section 2.5.

In the airway analysis, five relations are joined by their foreign keys. The foreign key dependencies are shown in Figure 2.3 along with the database schemas and primary keys of each relation. Each relation appears as a box with attributes listed inside. The relation name is above the relation. A primary key is underlined

inside each relation. Each foreign key dependency appears as an red arrow from the foreign key attribute of the referencing relation to the primary key of referenced relation.
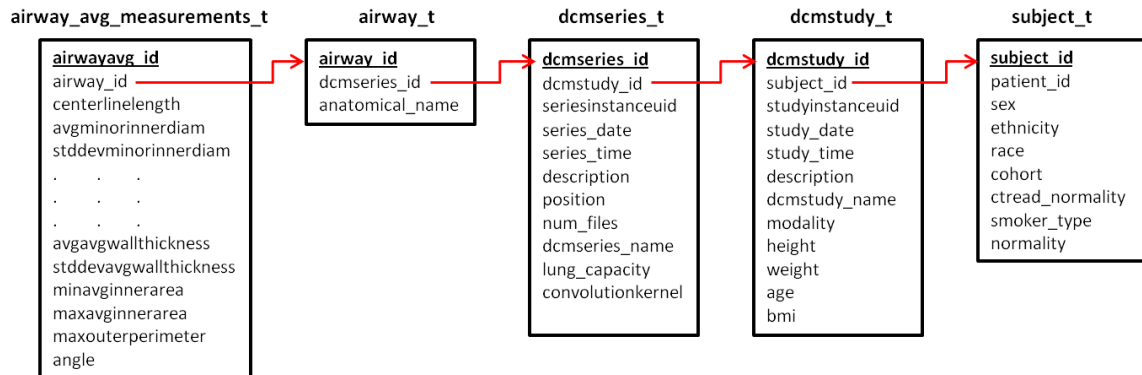
Figure 2.3: Schema diagram for airway analysis.

In the volume analysis, we use four relations to construct the volume relation: *subject_t*, *dcmstudy_t*, *dcmseries_t* and *lung_volume_t*. The foreign key dependencies are shown in Figure 2.4 along with the database schemas and primary keys of each relation.

The other relations in the atlas database are provided to enforce referential integrity and supply lookup values to the GUI front-end tool.

2.2 Atlas Graphic User Interface (GUI)

The atlas front-end graphic user interface was implemented using Microsoft Access 2007 under the Windows operating system. In this thesis, we use Access to refer to Microsoft Access 2007. Access is a Microsoft relational database management
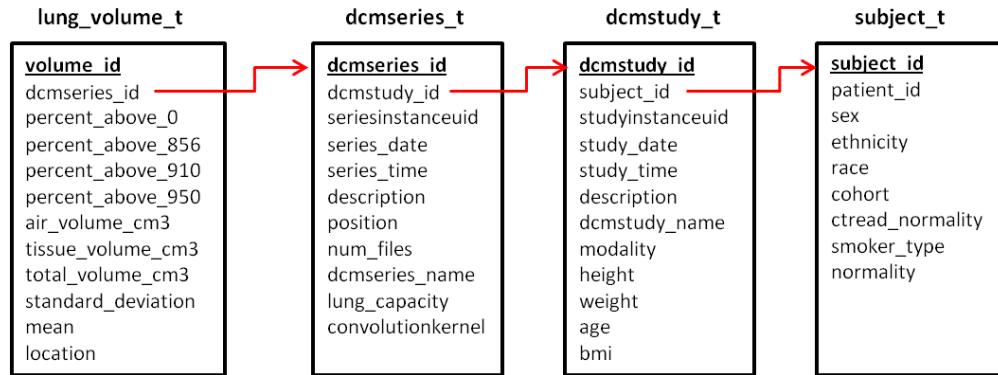
Figure 2.4: Schema diagram for volume analysis.

system which is used to create database solutions [25]. It provides wizards and intuitive user interfaces to create tables, forms, queries and reports step-by-step. It also provides easy conversion from form view to design view. It allows rapid data retrieval from tables or queries and stores the query results in the database. It has high integration with other Microsoft Office products such as Excel and Word. Access also offers several options for importing and exporting data. For example, we can import or link data from an external ODBC database, or export data into the Excel format. In the atlas, we linked to the data in MySQL server and used it only for viewing and querying. This allows the data in MySQL server to change and the Access platform to use the latest data.

The GUI in the human lung query atlas was created using forms in Access. A form is used to enter, edit or display data in a database. It is linked to a database source, such as a table or query. The source provides the structure for the form, such as attributes. In the atlas, the forms are linked to the queries which are made by user

actions.

The main workflow of human lung query atlas is shown in Figure 2.5. Airway and volume measurements are included in the atlas. For each part, there is one population statistics module and three comparisons modules: population-to-population, individual-to-population and individual-to-individual.

The population statistics module provides the statistics for the chosen population. For example, database demographics are acquired from this part (see section 3.2). The population-to-population comparison module provides the comparison between chosen populations. The comparison result contains the population names, branches or volume names, statistical measurement values (average and standard deviation), normalized statistical measurement values, t-statistics and p-values among populations. Many kinds of comparison studies can be performed such as how the lung volume changes from 20-29 population to 40-49 population. In this thesis, six studies were conducted and is discussed in Chapter 3.

The individual-to-population comparison module compares an individual to a population. It can be used to indicate how a subject is related to the chosen population. For example, comparing an abnormal subject to a normal population can be used to determine how abnormal the subject is. The individual-to-individual comparison module compares the measurements of two individuals.

The details about how to make these comparisons with the GUI are discussed in Chapter 3.
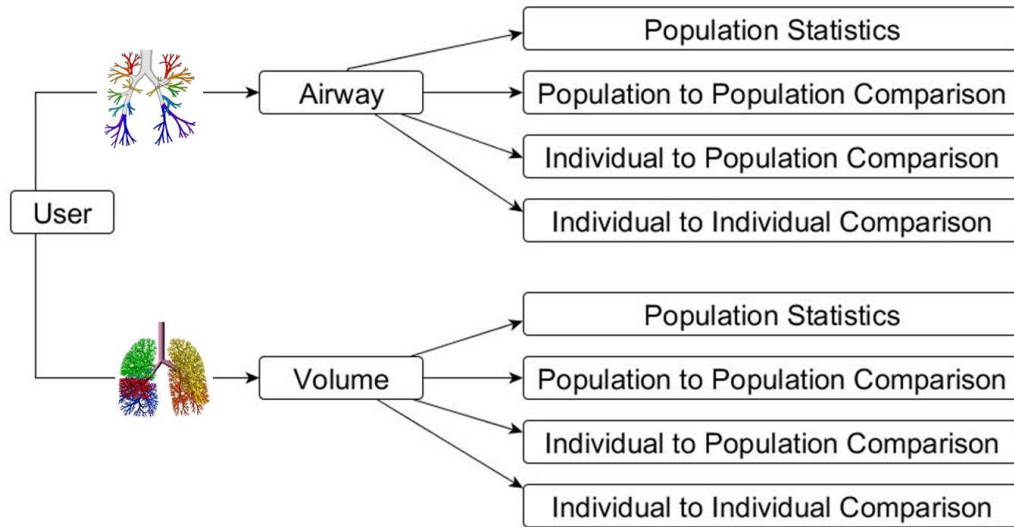
Figure 2.5: Main workflow of human lung atlas

2.3 Open Database Connectivity (ODBC)

ODBC is an application program interface to access information from database management systems. It helps the GUI application connect to a MySQL back-end database. Using ODBC, data can be easily imported (linked) from the database and stored locally in Access tables. In the human lung query atlas, we linked data to the database in a MySQL server by ODBC. This provides greater performance and deployment without the need of an Internet connection.

2.4 Visual Basic Application (VBA)

Visual Basic Application (VBA) is a programming language that is used with Microsoft Office applications, such as Access. It was used to construct SQL queries and provide user-defined functions on the forms. Each form is an object with many

event procedures inside, such as click event for a button control on a form.

Modules are used to store VBA code. VBA consists of two types of modules: standard modules and class modules. Standard modules are not associated with any object, while class modules are associated with a particular object, such as a form. In the atlas, there are 20 class modules which correspond to 20 existing forms and there are five standard modules consisting of the array module, control module, query module, statistics module, and public module.

The class module of the population-to-population airway comparison corresponds to population-to-population airway comparison form. The function cmdRun-Query_Click() is an event procedure and is called whenever the 'Run Query' button is clicked. This function inputs the query conditions from the user-input in the front-end GUI, constructs SQL queries from the back-end database and outputs the query results in the GUI. The function $cmdExportCSV\_Click()$ performs the query construction and prompts user to save the results in CSV format. The function $ClearForm$ resets all the controls in the GUI to their default values.

The statistics module is a standard module. It hardcodes the t-distribution table and provides statistical calculations. For example, whenever the Run Query button is clicked in the module of population-to-population airway comparison, it automatically calls the statistics module. The function $pValue$ calculates the t-statistics from the means and standard deviations from population 1 and 2. It looks up the table to find the 'right position' for this t-statistics. If the degree of freedom and this t-statistic value are both in the table, the $pValue$ function subsequently returns

the corresponding p-value in the table. Otherwise, linear interpolation is needed to calculate the p-value.

The definitions of t-test is discussed in the section 2.8.

2.5   Database Query Language

A database query language is a computer language used to make requests to a database. The structured query language (SQL) is the query language that was used to build atlas database and make queries. In this section, we introduce some basic structures of SQL queries and give some atlas query examples.

The basic structure of SQL queries contains three clauses: **select**, **from** and **where**. The **select** clause is used to select the output attributes in the query result. The **from** clause is used to specify the relations scanned in the evaluation of the expression. The **where** clause is used to list the predicates involving attributes of the relations that appear in the **from** clause.

For example, suppose we want to make a query to find the patient id and sex of all the normal white subjects in *subject_t* (see Figure 2.1). The SQL query can be expressed as:

```
SELECT patient_id, sex

FROM subject_t

WHERE normality = 'normal' and race = 'White'
```

The result is a new relation consisting of two attributes with the headings *patient_id* and *sex*, shown in Figure 2.6.

patient_id	sex
H-100	M
H-101	F
H-103	F
H-104	M

Figure 2.6: Patient id and sex of all the normal white subjects in Figure 2.1

.

SQL provides aggregate functions that take a collection of values as input and return a single statistical value. SQL offers five built-in aggregate functions: **avg**(average), **max**(maximum), **min**(minimum), **sum**(total) and **count**(count). The atlas uses these aggregate functions in the population statistics form (Figure 3.4).

For example, suppose we want to find the dcmstudy name, modality, study date, height, weight of the youngest subject in the *dcmstudy_t* (Figure 2.2). The SQL query can be expressed as:

```
SELECT dcmstudy_id, modality, study_date, height, weight

FROM dcmstudy_t

WHERE age = (select min(age) from dcmstudy_t)
```

The result is a new relation consisting of attributes with the headings *dcmstudy_id*, *modality*, *study_date*, *height* and *weight* (see Figure 2.7). The **select** clause specifies the output attributes: *dcmstudy_id*, *modality*, *study_date*, *height* and *weight*. The **where** clause contains a subquery which uses the aggregate function **min** to find the youngest subject.

Observe that the attribute *subject_id* appears in both *subject_t* and *dcmstudy_t*.

dcmstudy_id	modality	study_date	height	weight
800	CT	12/1/2000	1.67	65

Figure 2.7: The dcmstudy name, modality, study date, height and weight of the youngest subject in Figure 2.2

.

Therefore, we can use an **inner join** operator to join these two relations together by their *subject_id*. The SQL query can be expressed as:

```
SELECT *

FROM subject_t inner join dcmstudy_t

on subject_t.subject_id = dcmstudy_t.subject_id
```

The star symbol (*) in the SQL query represents all the attributes of *subject_t* and *dcmstudy_t*. The operator **on** shows the join condition that the *subject_id* in *subject_t* equals to the *subject_id* in *dcmstudy_t*.

Figure 2.8 shows the inner join of *subject_t* and *dcmstudy_t*. The first seven attributes come from the relation *subject_t* and the latter eight attributes come from the relation *dcmstudy_t*. The **inner join** operator keeps the *subject_id* twice in the resulting relation, with the names *subject_t.subject_id* and *dcmstudy_t.subject_id*. The **natural join** operator keeps the common attributes once in the result.

In the airway and volume analysis, the schema diagrams (Figure 2.3 and Figure 2.4) show the joins among these relations. For example, in the airway analysis, we used **inner join** to join *subject_t* and *dcmstudy_t*, *dcmstudy_t* and *dcmseries_t*, *dcmseries_t* and *airway_t*, *airway_t* and *airway_avg_measurements_t*. This con-

subject_t. subject_id	patient _id	sex	ethnicity	race	smoker _type	normality	dcmstudy _id	study_date	dcmstudy _name	modality	height	weight	age	dcmstudy_t. subject_id
100	H-100	M	Not Hispanic or Latino	White	non-smoker	normal	800	12/1/2000	p2brp-10	CT	1.67	65	20	100
101	H-101	F	Hispanic or Latino	White	non-smoker	normal	801	12/2/2000	p2brp-11	CT	1.68	55	23	101
102	H-102	M	Not Hispanic or Latino	White	smoker	abnormal	802	12/2/2000	p2brp-12	CT	1.8	50	34	102
103	H-103	F	Not Hispanic or Latino	White	non-smoker	normal	803	12/2/2000	p2brp-13	CT	1.65	52	56	103
104	H-104	M	Hispanic or Latino	White	smoker	normal	804	12/2/2000	p2brp-14	CT	1.75	90	35	104
105	H-105	F	Not Hispanic or Latino	Black/ African American	non-smoker	abnormal	805	12/3/2000	p2brp-15	CT	1.78	67	54	105
106	H-106	M	Not Hispanic or Latino	Native Hawaiian /Pacific	smoker	normal	806	12/2/2000	p2brp-16	CT	1.56	70	26	106
107	H-107	F	Not Hispanic or Latino	Asian	non-smoker	normal	807	12/2/2000	p2brp-17	CT	1.78	78	29	107
108	H-108	M	Not Hispanic or Latino	American Indian	smoker	normal	808	12/6/2000	p2brp-18	CT	1.73	50	30	108
109	H-109	F	Not Hispanic or Latino	Unknow or not	non-smoker	normal	809	12/9/2000	p2brp-19	CT	1.65	77	45	109

Figure 2.8: The inner join of *subject_t* and *dcmstudy_t*

.

structed a new relation: airway analysis.

SQL provides the **like** operator to search for a specified pattern in an attribute. We use this operator to categorize measurements. There is a difference between standard SQL and the SQL in Access when expressing missing letters. In standard SQL, missing letters are expressed using % while * is used in Access.

The attribute *measurement_name* shown in Figure 2.9 has many records, including area, diameter, wall thickness, perimeter and center line length. We can make queries to find all the diameters and wall thickness measurements, respectively. The Access SQL queries can be expressed as:

```
SELECT measurement_name

FROM measurement_t
```

```
WHERE measurement_name LIKE *diam*
```

And

```
SELECT measurement_name

FROM measurement_t

WHERE measurement_name LIKE *wallthickness*
```


Figure 2.9: A relation example: *measurement_t*.

The query results are shown in Figure 2.10. The left figure shows all the wall thickness measurements and the right figure shows all the diameter measurements in Figure 2.9.

measurement_name
avgavgwallthickness
avgminwallthickness
avgmaxwallthickness
minavgwallthickness
maxavgwallthickness

measurement_name
avgminorinnerdiam
avgmajorinnerdiam
minavgminorinnerdiam
avgminorouterdiam

Figure 2.10: The queried wall thickness and diameter measurements for Figure 2.9.

2.6   Data Acquisition

The atlas data comes from CT volumetric image data and measurements of the CT data. Figure 2.11 shows a flow chart illustrating how the atlas data was collected.

3D CT data was collected and stored in MIFAR, which is a relational database used to store medical image data along with associated post-image processing files. MIFAR is a data management tool developed by the Iowa Comprehensive Lung Imaging Center (I-CLIC). PW2 software analyzes 3D CT image data and produces segmented airways and lobes, and comma-separated values (CSV) file of airway measurements and volume histogram. The back-end atlas database was built based on these CSV files.

3D CT data collection is discussed in section 2.6.1. PW2 measurements data is discussed in section 2.6.2.
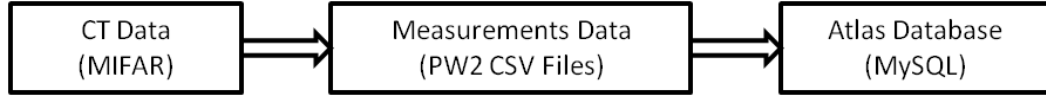
Figure 2.11: Data acquisition workflow.

2.6.1 Three-dimensional CT Data

The data used in the human lung atlas consist of 148 volumetric pulmonary CT image sets from 102 subjects. A Siemens Sensation 64 multi-detector row CT scanner (Florsheim, Germany) and a Philips MX 16 multi-detector row CT scanner were used to acquire these data during breath-holds near functional residual capacity (FRC) and total lung capacity (TLC) in the same scanning session. The volume of gas in the lung after a normal expiration is called functional residual capacity (FRC), and the volume of air in the lung after maximum inspiration is called total lung capacity (TLC) [23]. The data acquisition was approved by the University of Iowa Institutional Review Board.

Most of the volumetric data set contain between 300 and 600 contiguous image slices with a slice thickness $0.75mm$ and were reconstructed with a 512×512 matrix. The in-plane pixel spacing is approximately $0.6 \times 0.6mm^2$. Four reconstruction kernels were used in this study, where B31f represents a kernel for the Siemens scanner and B represents a kernel for the Philips scanner.

2.6.2 PW2 Data

The Pulmonary Workstation 2 (PW2) (VIDA Diagnostics, Inc, Iowa City, IA) software was used to analyze CT image volumes. It incorporates lobar segmentation,

airway analysis and parenchymal density analysis, and produces an airway segmentation, airway labeling, airway measurements, lung segmentation, lobe segmentation, vessel segmentation, histogram analysis and low attenuation cluster (LAC) analysis [1]. PW2 has been approved by Food and Drug Administration (FDA) for clinical use of pulmonary measurements. The analyzed data from PW2 were used to build the atlas database.

2.7   Pulmonary Structures and Measurements

2.7.1   Pulmonary Structures

The human lung is divided into left and right lungs. The left lung is divided into two lobes: left upper lobe and left lower lobe separated by the left oblique fissure. The right lung has three lobes: right upper lobe, right middle lobe and right lower lobe. The right oblique fissure and right horizontal fissure divide these right lobes.

The human airway tree is a tree-like structure consisting of the trachea and other branches. The trachea divides into two main branches: left main bronchus (LMB) and right main bronchus (RMB). The airway labeling reference is shown in Figure 2.12.

Figure 2.12 shows the 15 main branches of the left lung: LMB, left upper lobe (LUL), left lower bronchus 6 (LLB6), left bronchus 1+2 (LB1+2), left bronchus 3 (LB3), left bronchus 4+5 (LB4+5), left bronchus 6 (LB6), left lower bronchus (LLB), left bronchus 1 (LB1), left bronchus 2 (LB2), left bronchus 4 (LB4), left bronchus 5 (LB5), left bronchus 8 (LB8), left bronchus 9 (LB9) and left bronchus 10 (LB10).

The right lung contains 16 main branches: RMB, right upper lobe (RUL),

bronchus intermedius (BronInt), right bronchus 1 (RB1), right bronchus 2 (RB2), right bronchus 3 (RB3), right bronchus 4+5 (RB4+5), right bronchus 6 (RB6), right lower lobe 7 (RLL7), right bronchus 4 (RB4), right bronchus 5 (RB5), right bronchus 7 (RB7), right lower lobe (RLL), right bronchus 8 (RB8), right bronchus 9 (RB9), and right bronchus 10 (RB10).
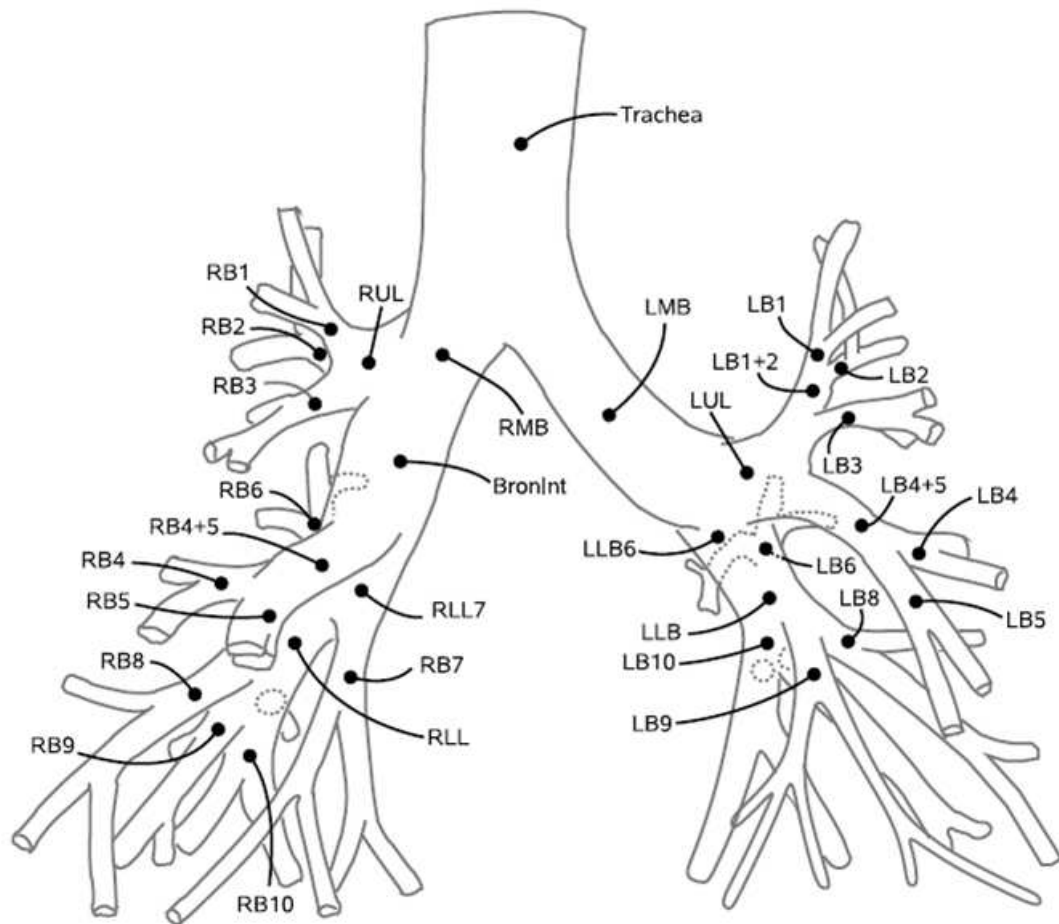
Figure 2.12: Airway labeling reference from VIDA.

The airway generation number is defined as the number of branches that it

takes to reach the trachea. VIDA has predefined 6 generations for these 32 main branches. The number 0 to 5 indicate the positions from top to bottom. Figure 2.13 shows the generation numbers for the airways. For example, the generation for trachea is 0 which means trachea starts the airway tree.

Generation Number	Airway Segment					
0	Trachea					
1	LMB	RMB				
2	LUL	RUL	LLB6	BronInt		
3	LB1+2	LB3	LB4+5	LB6	LLB	
	RB1	RB2	RB3	RB4+5	RB6	RLL7
4	LB1	LB2	LB4	LB5	LB8	
	LB9	LB10	RB4	RB5	RB7	RLL
5	RB8	RB9	RB10			

Figure 2.13: Airway segments and their generation numbers from VIDA.

2.7.2 Airway Measurements

Airway measurements provided by PW2 are categorized into five groups: area, diameter, perimeter, wall thickness and others. The others category includes the center line length, airway segment directions, generation number, lobe number and angles. The measurements from the first four categories are combined with the other extra four measurements (minor, major, inner and outer) and four statistics measurements (average, standard deviation, maximum, minimum). In all, there are 121

measurements in the human lung atlas.

Some measurements definitions are listed below. These definitions come from PW2 FDA version user manual [1].

The minor measurement refers to the smallest measurement of interest at each airway cross section. For example, the smallest diameter in an airway cross section is called the minor diameter. The major measurement refers to the largest measurement of interest at each airway cross section. The inner measurement refers to the measurement for the lumen area of the airway cross section. The outer measurement refers to the measurement for the lumen and airway wall area of the airway cross section. The center line length follows the curvature of the segment, measured between parent- and child-branch point, and represents the true path length.

The measurements used in the airway analysis (Chapter 3) are average major inner diameter, average minor inner diameter, average inner area, average outer area and average average wall thickness. According to the definitions in [1], the average major inner diameter represents the average of the greatest inner (lumen) diameters measured at each centerline-voxel position along the middle 1/3 of the segment, measured in mm. The average minor inner diameter represents the average of the smallest inner (lumen) diameters measured at each centerline-voxel position along the middle 1/3 of the segment, measured in mm. The average inner area represents the average of the lumen areas measured at each centerline-voxel position along the middle 1/3 of the segment, measured in mm^2. The average outer area represents the average of the area enclosed by the outer airway border at each centerline-voxel position along

the middle 1/3 of the segment, measured in mm^2. The wall thickness is measured along rays cast from the center of gravity of the airway lumen at every centerline voxel position. Rays are cast every half-degree, for a total of 720 measurements. The average average wall thickness represents the averaged 720 wall thickness measurements for each centerline voxel position along the middle 1/3 of the airway segment. It is measured in mm.

2.7.3   Volume Measurements

Volume measurements are provided by the histogram analysis module of PW2. This module is aimed to provide quantitative measurements for segmented lungs and lobes. The volume measurements are described in the PW2's user manual [1] and are listed below.

The mean density represents the average density value in Hounsfield Units (HU) of all voxels within the given region. The standard deviation represents the standard deviation of the density values of all voxels within the given region. Total volume measures the total volume for the anatomical region, such as any lobe or any lung. It is measured in cm^3. Tissue Volume measures the tissue volume for the anatomical region. Air volume measures the air volume for the anatomical region. The percent below -950 HU measurement refers to the percentage of tissue below and including -950 HU. The percent below -910 HU refers to the percentage of tissue below and including -910 HU. The percent below -856 HU refers to the percentage of tissue below and including -856 HU. The percent below 0 HU refers to the percentage of tissue below and including 0 HU.

The total volume v_{tot} of a voxel is defined by its spatial dimensions x_{dim}, y_{dim} and z_{dim}:

$$v_{tot} = x_{dim} \cdot y_{dim} \cdot z_{dim}$$

Given the Hounsfield Unit density value d of a voxel, its air volume is defined by

$$v_{air} = \begin{cases} v_{tot} & d < d_{air} \\ v_{tot} \cdot \frac{d_{blood} - d}{d_{blood} - d_{air}} & d_{air} \leq d \leq d_{blood} \\ 0 & d > d_{blood} \end{cases}$$

with $d_{blood} = 65HU$ and $d_{air} = -1000HU$.

The tissue volume v_{tissue} is defined by $v_{tissue} = v_{tot} - v_{air}$.

2.8   T-statistic and P-value

In Chapter 3, we use the t-statistic and p-value to determine the significant differences between chosen populations. In this section, the concepts of t-statistics and p-value are introduced.

For a statistical hypothesis, statements or claims are made about an unknown parameter in the population, such as the population mean μ. A statistical hypothesis test makes two hypotheses. The hypothesis which is assumed to be true before the experiment is called the *null hypothesis* and is denoted as H_0. The hypothesis which interests the researchers is called the *alternative hypothesis* and is denoted as H_A. The evidence in sample data is used to either reject the *null hypothesis* in favor

of the *alternative hypothesis*, or fail to reject the *null hypothesis*. If the sample evidence supports H_A and we reject H_0 in favor of H_A, we claim that the results are statistically significant. If the sample evidence does not support H_A and we fail to reject H_0, we claim that the results are not statistically significant.

A decision rule for determining when H_0 should be rejected or not is a pre-define probability value α. α is called the *significance level* of the test and it occurs when the *null hypothesis* is ture, yet based on sample data, we reject it anyway. The most common choices of α are 0.01, 0.05 and 0.10.

In the test, a probability is computed and compared with α. This probability is called the p-value of the test statistics. If the $p - value$ is smaller than α, then we reject H_0 in favor of H_A and claim that the results are statistically significant. The smaller the value of α, the more proof required to reject H_0 in favor of H_A.

Usually, a t-statistic is computed first and the p-value is calculated by looking it up in a t-distribution table. A t-distribution is bell-shaped distribution like a normal distribution. It has only one parameter called degree of freedom which depends on the sample numbers.

In the atlas, we perform a two-sample, two-sided, t-test with unequal variances to compare means from two independent populations. The sample mean from population 1 is denoted as μ_1, and the sample mean from population 2 is denoted as μ_2. We are interested to know whether the mean in population 1 is statistically different from that in population 2.

The hypotheses are: $H_0 : \mu_1 = \mu_2$ and $H_A : \mu_1 \neq \mu_2$

The test statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}}}$$

compared to $t(v)$ degree of freedom, where

$$v = \frac{(a+b)^2}{[(\frac{a^2}{n_1-1}) + (\frac{b^2}{n_2-1})]}$$

where

$$a = \frac{s_1{}^2}{n_1}, b = \frac{s_2{}^2}{n_2}$$

The corresponding p-value is computed by looking up t-distribution table, which was hardcoded in the atlas statistic module. The implementation of statistic module including calculating t-statistics and p-value is discussed in section 2.4.

2.9   Data Accuracy Assessment

PW2 from VIDA has obtained FDA 510(k) clearance for clinical use (K083227). This section summarizes the evaluations performed by VIDA to provide the accuracy assessment for the data used in the human lung query atlas.

2.9.1   Lung Segmentation Accuracy Assessment

The lung segmentation accuracy of Hu *et al.* 's segmentation algorithm was reported in Hu *et al.* 's paper [10] and in VIDA's testing documentations [2, 3]. These results show that the algorithm used in the PW2 lung segmentation module has a good performance and the data used in the human lung atlas are reliable.

The algorithm used in the PW2 lung segmentation module was proposed by

Hu *et al.* [10]. This algorithm contains three main steps. In the first step, intensity-based optimal thresholding was used to automatically extract the whole lung region. In addition, connectivity and topological analysis were adopted to eliminate disconnected regions and useless interior cavities. In the second step, dynamic programming was applied to locate anterior and posterior junctions in order to separate the left and right lungs. In the last step, binary mathematical morphological operations were used to smooth the lung boundary in order to be consistent with the one labeled by the human expert.

Hu *et al.* evaluated lung segmentation performance using the root mean square (RMS) error and the volume error. RMS error was computed between manually defined and computer defined lung borders. The volume error was the error from repeated scans of the same subject at TLC. For their experiment, they showed that the RMS error was 0.54 *mm* (0.8 *voxel*) and volume error was $2.7\% \pm 2.29\%$.

In order to compute RMS errors, the minimum distances from each pixel on the computer defined border to the pixels on the manually defined border were computed as

$$d_i = \min_j \{ \sqrt{(x_j^C - x_i^M)^2 + (y_j^C - y_i^M)^2} \}$$

The point (x_i^M, y_i^M) is defined as the voxel position on the manually defined lung borders and (x_i^C, y_i^C) is defined as the voxel position on the computer defined lung borders.

The RMS error is given by

$$d_{rms} = \sqrt{\frac{\sum_{i=1}^{N} d_i^2}{N}}$$

where N is the number of points on the manually defined lung borders.

VIDA evaluated the segmentation performance using 32 human CT scans from 27 subjects. A human expert created a reference segmentation by manually tracing the lung borders in both left and right lung. The RMS and volume errors were computed to evaluate the algorithm performance. The RMS error was $0.58 \pm 0.13 mm$ for left lung, and $0.57 \pm 0.23 mm$ for right lung. The volume error was $2.83\% \pm 1.72\%$ for left lung, and $2.08\% \pm 2.29\%$ for right lung.

Hu *et al.* and VIDA's results show high similarity. The numbers in VIDA's study contain a few outliers when the automatic segmentation failed. To avoid these outliers in the human lung atlas, all the data were inspected by human experts and any outlier was manually corrected. It is assumed that the accuracy of the segmentation measurement in the atlas is at least good as the numbers reported in the Hu *et al.* 's and VIDA's studies.

2.9.2   Lobe Segmentation Accuracy Assessment

The lobe segmentation accuracy was studied by Ukil *et al.* and also reported in the VIDA's testing documentations [2, 3]. The results show that the algorithm used in the PW2 lobe segmentation module has a good performance and reliable data were contained in the human lung atlas.

The algorithm used in the PW2 lobe segmentation was proposed by Ukil *et al.* [22]. In their algorithm, they proposed an automatic method to detect three lung fissures (left oblique fissure, right oblique fissure and right horizontal fissure) in order to segment five lobes in two lungs. Specifically, the watershed transformation was applied to segment the initial lobar boundaries provided by the segmentation and analysis of the pulmonary airways and vessels. To refine the initial segmentations, optimal surface detection using 3D graph search was applied. Ukil *et al.* also introduced an algorithm to detect incomplete fissures.

Ukil *et al.* used RMS error to evaluate lobe segmentation performance. For their experiment, they showed that the minimum RMS errors among two different reconstruction kernels at TLC were $1.81 \pm 0.58mm$ for left oblique, $1.57 \pm 0.17mm$ for right oblique and $1.43 \pm 0.57mm$ for right horizontal.

VIDA evaluated the lobe segmentation performance using 30 human CT scans for left oblique fissure and 32 human CT scans for right fissures. Three dividing fissures were traced manually by a human expert. The RMS and volume errors were computed after the deletion of statistical outliers. The RMS error was $0.21 \pm 0.70mm$ for left oblique, $0.89 \pm 0.38mm$ for right oblique, and $0.64 \pm 0.39mm$ for right horizontal fissure. The volume error was $0.21\% \pm 0.68\%$ for left upper lobe, $1.07\% \pm 0.61\%$ for left lower lobe, $-0.12\% \pm 0.45\%$ for right upper lobe, $0.23\% \pm 0.92\%$ for right middle lobe and $0.18\% \pm 0.18\%$ for right lower lobe.

Ukil *et al.* and VIDA's results show high similarity. The numbers in VIDA's study contain a few outliers when the automatic segmentation failed. To avoid these

outliers in the human lung atlas, all the data were inspected by human experts and any outlier was manually corrected. It is assumed that the accuracy of the lobe segmentation measurement in the atlas is no worse than the numbers reported in the Ukil *et al.* 's and VIDA's studies.

2.9.3 Airway Segmentation and Labeling Accuracy

The algorithms for airway segmentation and labeling accuracy were proposed and studied by Tschirren *et al.* [20, 21] and were reported in the VIDA's testing documentations [2, 3].

VIDA used the algorithms proposed by Tschirren *et al.* and conducted airway segmentation and labeling accuracy using 32 human CT scans from 27 subjects. Since airway segmentation was the precursor for the lobe segmentation, it was required to detect five critical branches (LUL, LLB6, RUL, BronInt and RB4+5) in order to locate the five lobes. Another requirement was the detection of all seven branches at generation 0, 1, and 2 (Trachea, LMB, RMB, LUL, LLB6, RUL, BronInt). Human experts inspected all the data in the atlas and any wrong airway segmentation and labeling were manually corrected. Reliable data are assumed to be contained in the human lung atlas.

2.9.4 Airway Measurement Accuracy

VIDA conducted the testing for airway measurement accuracy using a physical phantom (shown in Figure 2.14). Seven tubes in the phantom were simulated as human airways and the dry potato flakes filled between these tubes were used as

simulated lung parenchyma. This phantom was scanned at four angles to mimic different orientations of imaging airways within the lung. The CT scan of this physical phantom is shown in Figure 2.15. In the calculation, the smallest tube 0 was excluded due to its invisibility in the CT scan.

Figure 2.14: The Plexiglas phantom used in the airway measurement accuracy testing

The algorithms to evaluate airway measurement accuracy were proposed by Reinhardt *et al.* and Palagyi *et al.* [11, 18]. Reinhardt *et al.* introduced a new method to estimate the inner and outer wall locations based on a model of scanning process. Palagyi *et al.* proposed a robust method which efficiently identify the centerline and bifurcation points in the pulmonary airway tree structure.

VIDA only evaluated the accuracy of inner and outer diameters of airway
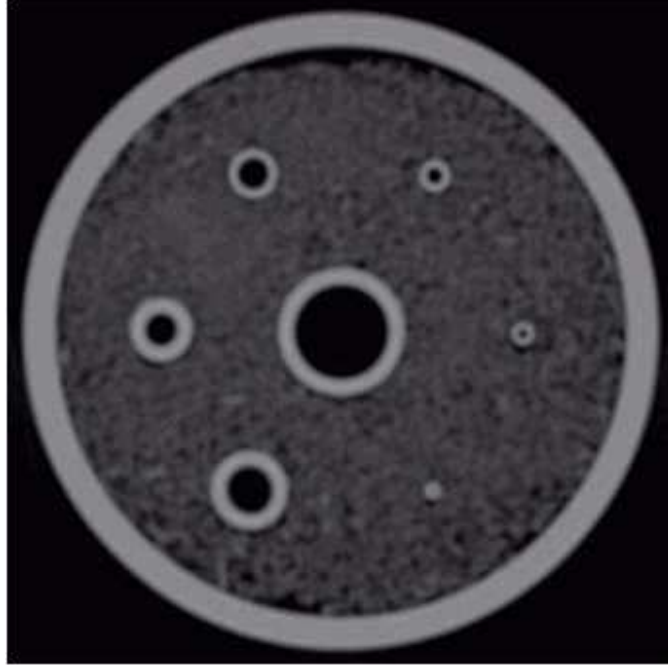
Figure 2.15: The CT scan of the Plexiglas phantom

walls but the accuracy of other measurements (area, perimeter and wall thickness) used in the human lung atlas were obtained from the diameter measurements. Since the inability of detecting sub-voxel structures for human expert, one voxel accuracy was used to evaluate the error between computer analysis and human analysis. These errors of inner diameter were $0.04 \pm 0.07mm$ for tube 1, $0.01 \pm 0.08mm$ for tube 2, $-0.01 \pm 0.04mm$ for tube 3, $-0.10 \pm 0.05mm$ for tube 4, $-0.02 \pm 0.01mm$ for tube 5, and $-0.04 \pm 0.06mm$ for tube 6. The errors of outer diameter were $-0.06 \pm 0.09mm$ for tube 1, $0.02 \pm 0.04mm$ for tube 2, $-0.04 \pm 0.09mm$ for tube 3, $0.19 \pm 0.20mm$ for tube 4, $-0.03 \pm 0.12mm$ for tube 5, and $0.02 \pm 0.18mm$ for tube 6. The voxel size was $0.39mm \times 0.39mm \times 0.6mm$. The errors in both inner and outer diameter were within one voxel dimension. Since wall thickness is half difference between outer diameter

and inner diameter, the errors of wall thickness for tube 1 to 6 were $0.05 \pm 0.08mm$, $0.02 \pm 0.06mm$, $0.03 \pm 0.06mm$, $0.15 \pm 0.13mm$, $0.03 \pm 0.1mm$, $0.02 \pm 0.18mm$, which fell within one voxel size as well. The errors for areas, perimeters can be calculated in the same way and it can be seen that these errors were much smaller than one voxel size. Therefore, it is assumed that the accuracy of the airway measurement in the atlas is high and the data in the atlas is reliable.

CHAPTER 3
RESULTS

3.1 Atlas Graphic User Interface

The design for human lung atlas GUI is discussed in Chapter 2. This section discusses more details in the GUI design.

When the human lung query atlas is loaded, a splash screen is displayed shown in Figure 3.1. This screen does not disappear until the 'Start' button in the middle bottom is clicked.
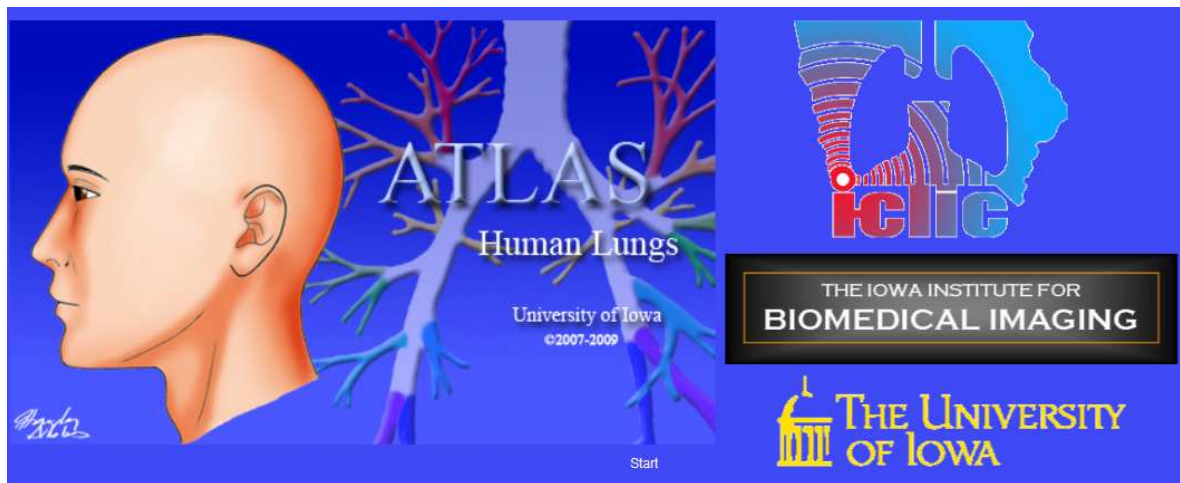
Figure 3.1: Screen splash of human lung atlas

The dispearance of splash screen takes the user to the main screen. The main screen consists of four parts: home page, airway measurements, volume measurements and contact page. The airway and volume measurements pages are shown in Figure

3.2 and Figure 3.3.
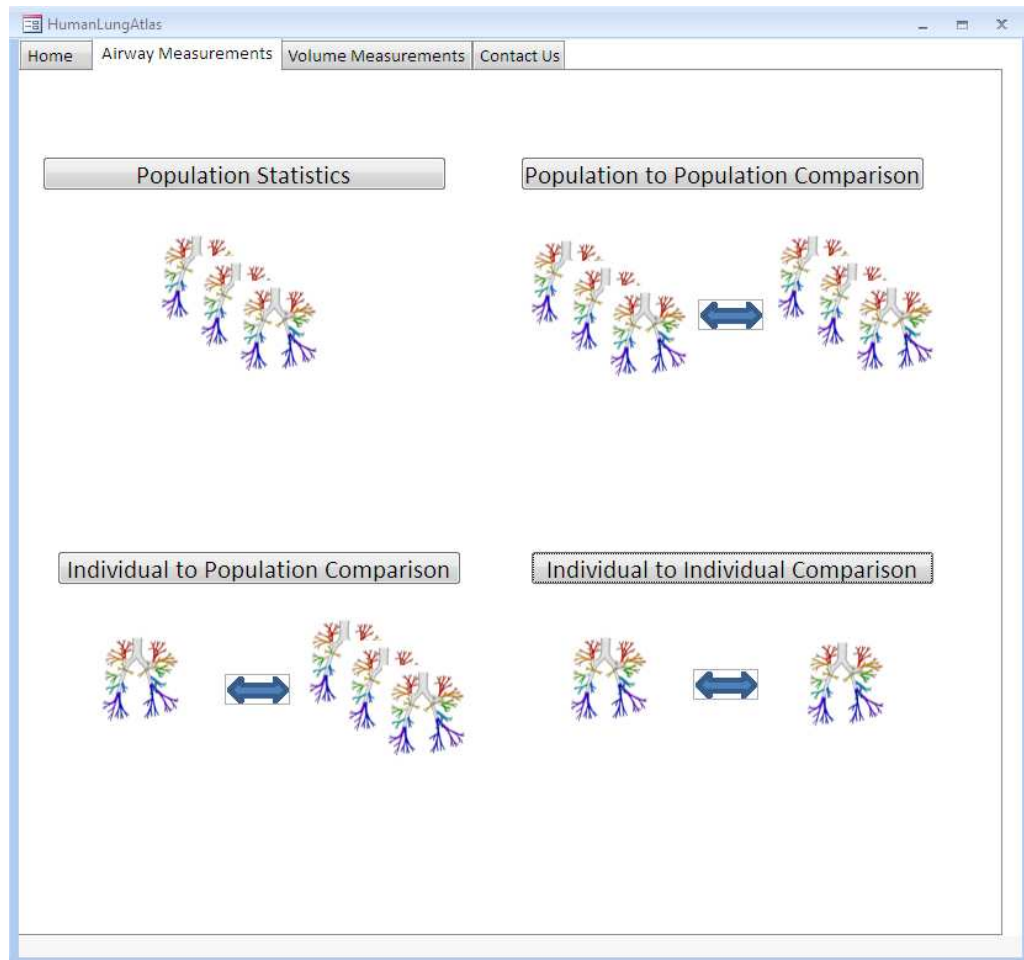
Figure 3.2: Airway measurement page.

Each measurement page contains four modules. Since two measurements have the similar interface, we only discuss about the four forms in the airway measurements, and the volume measurements GUI will not be discussed.

Figure 3.4 shows the population statistics form and corresponding execution steps. The interface reflects the back-end database information. For example, the
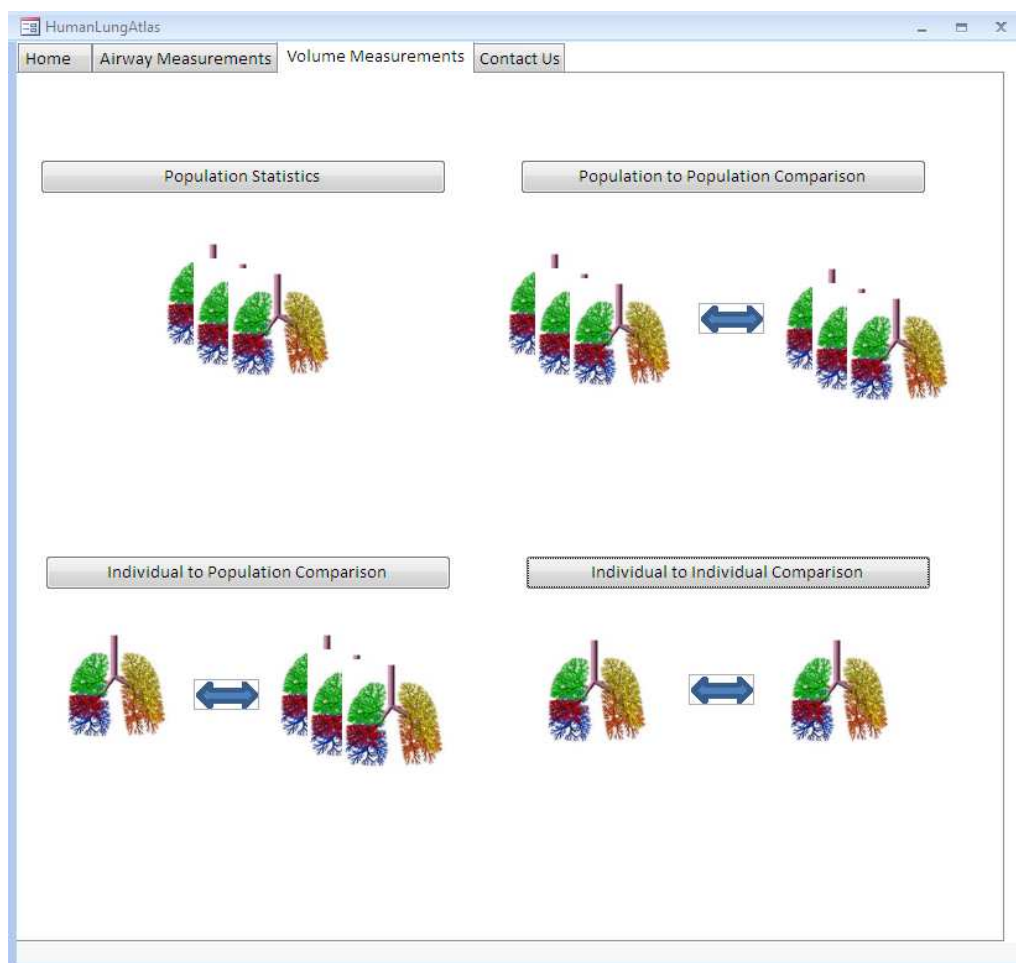
Figure 3.3: Volume measurement page.

subject information in step 1 comes from *subject_t* and *dcmstudy_t*, the branches information in step 2 are found in the *airway_t*, the measurements information in step 3 comes from *airway_avg_measurements_t*, and the lung capacities in step 4 are included in the *dcmseries_t*.

In the comparison forms, some terms need to be defined. The user must specify at least one gender: male, female or both, otherwise, an error window pops up. The default age is set up to be the maximum 100 and the minimum 0. Ethnicity contains two groups: Hispanic or Latino and not Hispanic or Latino. According to wikipedia, Hispanic or Latino Americans are defined as Americans with origins in the Hispanic countries of Latin America or in Spain [24]. PFT stands for pulmonary function test and is a group of procedures to measure lung functions. In the atlas, the PFT has two states: normal and abnormal. Smoking status contains smoker and non-smoker. Lung capacity consists of FRC and TLC. Convolution kernel represents CT reconstruction kernel. There are four kernels: B35f, B31f, B30f and B. Since one subject may have multiple convolution kernels, we define a default kernel in the sequence of B35f, B31f, B30f and B. For example, if one subject has two kernels: B31f and B30f, the default kernel will be B31f. Branch consists of 1115 branch labels. We mainly use the 32 major branches in Figure 2.12.

We provide two examples to illustrate different results of population statistics.

Population Statistics Part I:

Suppose we want to find the population statistics of 20-29 normal white non-smoking females at FRC. Figure 3.5 shows the result if no measurements and branches

Step 1:
Specify the population. It is not necessary to select all the options of Ethnicity, Smoking Status, PFT Results, and Racial categories.

Step 2:
Specify branches (optional). Press "Choose Branches" and select the desired branches

Human Lung Query Atlas Population Statistics

Subject Information [?]

Age [0] to [100]

Gender ☑ Male ☑ Female

Ethnicity
☑ Hispanic or Latino
☑ Not Hispanic or Latino
☑ Unknown or not reported

Smoking Status
☑ non-smoker
☑ smoker

PFT Result
☑ normal
☑ abnormal

Racial Categories
☑ American Indian/ Alaskan Native
☑ Asian
☑ Black or African Ameri
☑ More than one race
☑ Native Hawaiian or Pacific
☑ Other
☑ Unknown or not reported
☑ White

[Select All] [Deselect]

Branches Information [?]
[Choose Branches]

Lung Capacity [?]

☑ FRC ☑ TLC

Convolution kernel
[Default ▾]

Aggregate Functions

☑ Average ☑ Max

☑ Std Dev ☑ Min

Measurements Information [?]

| Area | Diameter | Perimeter | WallThickne | ◂ ▸ |

avginnerarea
stddevinnerarea
minavginnerarea
maxavginnerarea
avgouterarea
stddevouterarea
minavgouterarea
maxavgouterarea
innercrosssecarea_30
outercrosssecarea_30
innercrosssecarea_40
outercrosssecarea_40
innercrosssecarea_50
outercrosssecarea_50
innercrosssecarea_60
outercrosssecarea_60
innercrosssecarea_70
outercrosssecarea_70
minInnerArea
maxOuterArea
avgWallAreaFraction
minOuterArea

[Select All] [Deselect All]

[Submit] [Reset Form]

Step 3:
Specify measurements (optional). Select the desired measurements from the list boxes.

Step 4:
Select lung capacity and convolution kernel.
At least one of the "frc" and "tlc" must be selected.

Step 5:
Select aggregate functions (optional).

Step 6:
Click "Run Query" button to retrieve data from atlas.

Figure 3.4: Population statistics form with execution steps.

are selected. There are twenty-seven 20-29 normal white non-smoking females at FRC in the database. In this part, the aggregate functions are disabled since no measurement was selected.
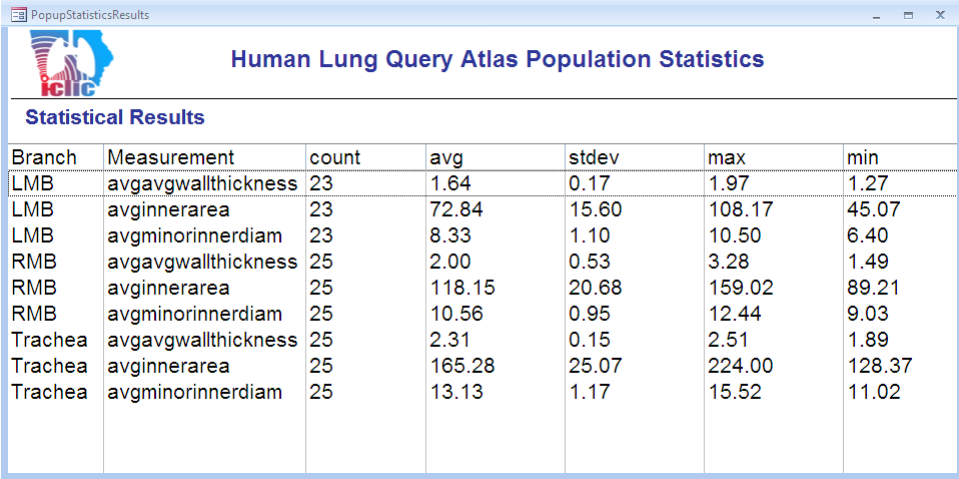
Figure 3.5: Population statistical results without measurement selection.

Population Statistics II:

Suppose we want to find the average average wall thickness, average inner area and average minor inner diameter measurement values of LMB, RMB and trachea from 20-29 normal white non-smoking females at FRC. Since there are measurement selections, the aggregation function are enabled and the other population statistical functions are disabled. Figure 3.6 shows the result. There are 23 subjects who have LMB information which indicates that there are four subjects who do not have trachea information in the database.

The measurements from two populations can be compared using the airway population-to-population comparison module. This module with its execution steps

Figure 3.6: Population statistical results with measurement selection.

is shown in Figure 3.7. Step 1 and 2 are similar to the step 1 in Figure 3.4. A new function was added in step 3, which calculates the normalized airway values by height, weight, BMI or none. People with larger height or weight may have a larger body shape, which may affect the mean value of the airways in one population. Therefore, it is necessary to compare normalized values. In Figure 3.7, all the females in the database are selected as population 1 and all the males composite population 2. The trachea and average inner area are chosen as the branch and measurement, respectively. In the Statistical Results part, several values are calculated: the normalized average values with standard deviations and the non-normalized average values with standard deviations. The t-statistics and p-value are automatically calculated to measure the significant differences between the selected populations. In statistics, the lower the p-value, the more significant the result is. For example, if the p-value is less than 0.05 or 0.01, it can be concluded that there is 5% or 1% chance of rejecting the

null hypothesis when it is true and there is significant difference between compared groups, respectively. In this case, all the comparison combinations from these two populations show the significant differences.

Figure 3.7: Airway population to population comparison form.

The other two comparison forms (individual-to-population and individual-to-individual) have a similar interface as that shown in Figure 3.7.

Figure 3.8 shows the airway individual-to-population comparison form. A

button called 'Choose Individual' is used to select an individual from the database. In this example, the comparison is between an abnormal white female and a 40-49 normal population.

Figure 3.8: Airway individual to population comparison form.

Figure 3.9 shows the airway individual-to-individual comparison form. A normal white non-smoking female and a normal white smoking male are compared at FRC.

This GUI is designed to be user-friendly. For example, the tab key can be used to move cursor from one option to the next option. A button called Search is used

Figure 3.9: Airway individual to individual comparison form.

to look up branches or measurements more easily. A button with a question mark is used to provide help information.

3.2 Database Demographics

At present, the human lung atlas consists of measurements from 148 CT image series, and 102 subjects. This includes data from 50 males and 52 females; 56 non-smokers and 46 smokers; and 2 American Indians, 3 Asians, 12 Black/African Americans, and 85 White/Caucasians of which 3 are Hispanic/Latino. 12 of the atlas subjects are designated as having abnormal pulmonary functional tests (PFT) of which 9 are smokers and 3 are non-smokers. The range of ages of the individuals stored in the atlas is 20-80 years old. 95 of the 148 CT series were imaged at FRC, and 53 were imaged at TLC. Thus, the atlas does not contain measurements at FRC and TLC for all individuals.

Table 3.1: Database Demographics Summary

Subject Information	Categories				
Age	$20-29$	$40-49$	$50-59$	$60-69$	>70
No. of subjects	49	41	1	10	1
Race	American Indians	Asians	Black/African Americans	White	
No. of subjects	2	3	12	85	
Ethnicity	Hispanic or Latino		Not Hispanic or Latino		
No. of subjects	3		99		
Gender	Male		Female		
No. of subjects	50		52		
Smoking Status	Smoker		Non-smoker		
No. of subjects	46		56		
PFT	Normal		Abnormal		
No. of subjects	90		12		

3.3   Quantitative Analysis of Human Airway and Volume

In the thesis, six studies are presented using data extracted from the human lung query atlas. These studies show the significance of human lung query atlas in helping learn the variations among individuals and populations. Figure 3.10 shows the six studies that were performed. Study 1 (section 3.3.1) investigates the differences between females and males in the age range of 20-29. Study 2 (section 3.3.2) investigates the differences between individuals in the age range of 20-29 and 40-49. Study 3 repeats the procedure of Study 1 to investigate the differences between females and males in the age range of 40-49. Study 4 and 5 repeats the procedure of Study 2, but compares female only and male only populations in the two age groups, respectively. Study 6 compares normal and abnormal populations.
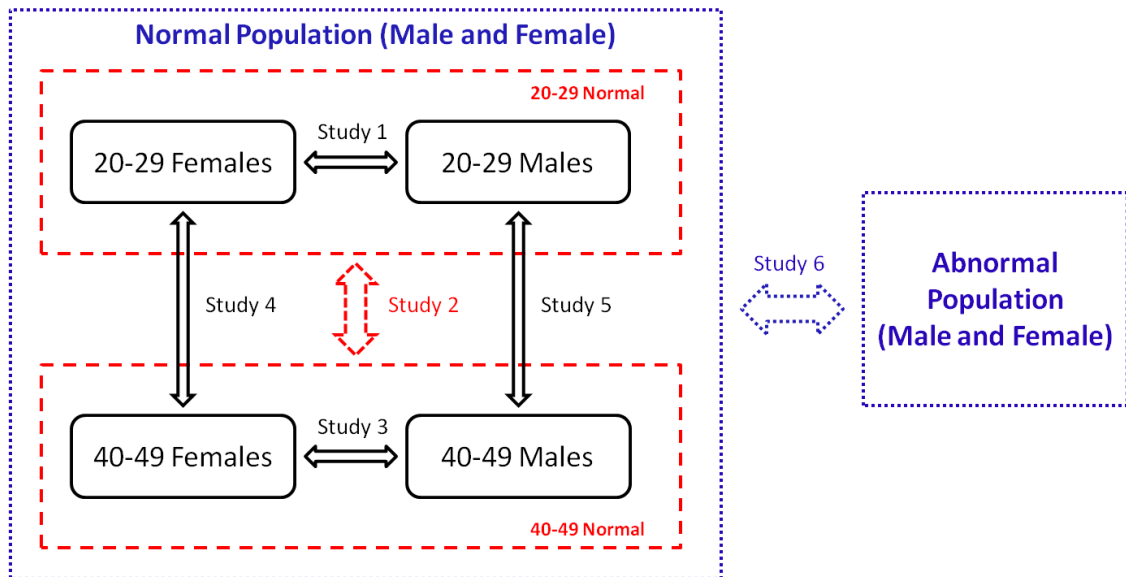
Figure 3.10: The overview of studies using the human lung query atlas

3.3.1 Study 1 : 20-29 Females and Males Comparison

According to Carey *et al.* [6], gender plays a significant role in influencing lung development and physiology and it may be an important factor affecting lung diseases, such as pulmonary fibrosis, asthma and lung cancer. Carey *et al.* described the development of the human lung in males and females and concluded that the airways of adult females were smaller than those of males. Another study from Martin *et al.* [14] found that the tracheal areas in female were significantly smaller than those of males and gender was a related determinant in airway size.

In this study, we hypothesized that the airways and lung volumes in young females are smaller than those of males in the age range of 20-29. Study 1 was conducted to test this hypothesis.

Study 1 used the airway and volume population-to-population comparison forms (Figure 2.5). Airway measurements from eight representative bronchi were chosen for this study: trachea, LMB, RMB, RUL, LUL, BronInt, LLB6 and RB4 + 5. These bronchi have significant roles in airways and lobar structures. According to VIDA's documentation [2], LUL, LLB6, RUL, RB4+5 and BronInt are critical airways for the lobar segmentation. Trachea, LMB, RMB are the zero and first generation airways in the human lung and support other airway structures. Four representative measurements were used to analyze these eight bronchi at FRC and TLC: average inner area, average outer area, average major inner diameter, average of average wallthickness. Volume measurements of all eight lung objects were studied: both, left, left-lower, left-upper, right, right-lower, right-middle, right-upper. Total volume

was selected as the representative measurement to analyze these lungs and lobes at

FRC and TLC. The definitions of measurements can be found in section 2.7.

Table 3.2: Study1: t-statistics, p-values and population counts for Figure 3.11

		Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4 + 5
FRC									
t-stat/p-value	Population 1	5.43	3.71	4.74	2.5	1.64	2.96	1.52	0.58
	Population 2	0	0	0	0.02	0.11	0	0.15	N/A
Count	Population 1	23	23	23	23	23	23	23	23
	Population 2	22	22	22	22	22	22	22	22
TLC									
t-stat/p-value	Population 1	6.21	3.59	5.59	2.03	1.42	4.37	2.53	0.9
	Population 2	0	0	0	0.05	0.17	0	0.02	N/A
Count	Population 1	23	23	23	23	23	23	23	23
	Population 2	22	22	22	22	22	22	22	22

Inner area and outer area measurements were compared to obtain how area

measurements change in two populations (section 2.7). Similar trends can be observed

in Figures 3.11 and 3.12. In these two figures, Population 1 indicates the normal

females in the age range of 20-29 and Population 2 indicates the normal males in

the age range of 20-29. The vertical axis shows the average values of measurements

in populations normalized by height. The unit in vertical axis is mm^2/m where m

comes from height unit and mm^2 comes from area unit. The horizontal axis shows

the selected bronchi grouped by two populations. The error bars in the shape of "I"

show the standard deviation of measurements in the populations. Tables 3.2 and 3.3

provide the t-statistics, p-values and counts for Populations 1 and 2, where '0' means

Average of the Average Inner Area (FRC)

	Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4+5
Population1	89.59	37.72	65.22	24.44	30.02	37.97	26.97	13.45
Population2	113.89	46.16	80.88	36.82	37.02	44.28	34.79	14.49

Average of the Average Inner Area (TLC)

	Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4+5
Population1	106.78	49.43	80.27	35.19	41.37	45.36	36.22	17.7
Population2	135.18	59.32	103.58	47.99	49.63	55.58	46.85	18.96
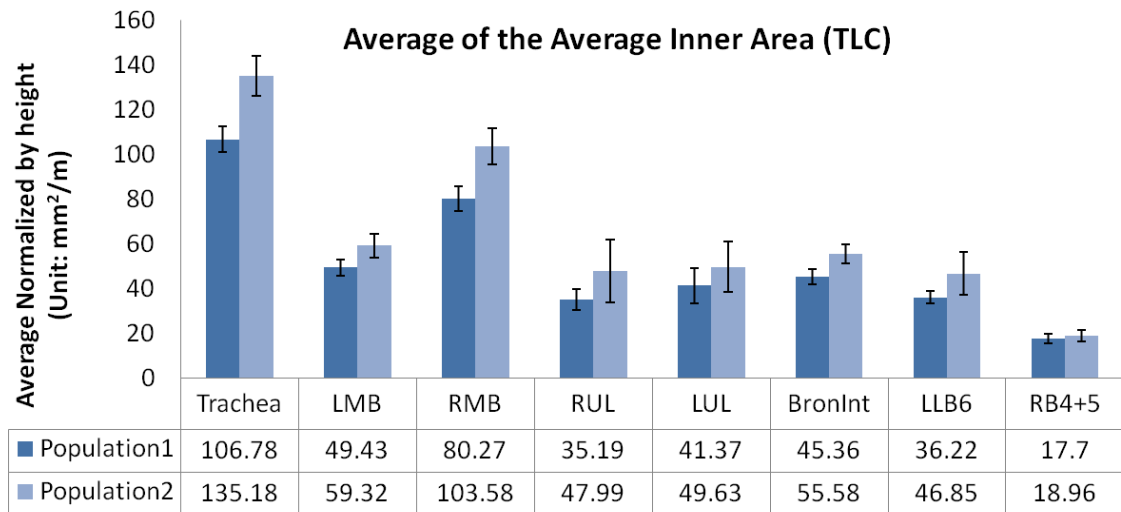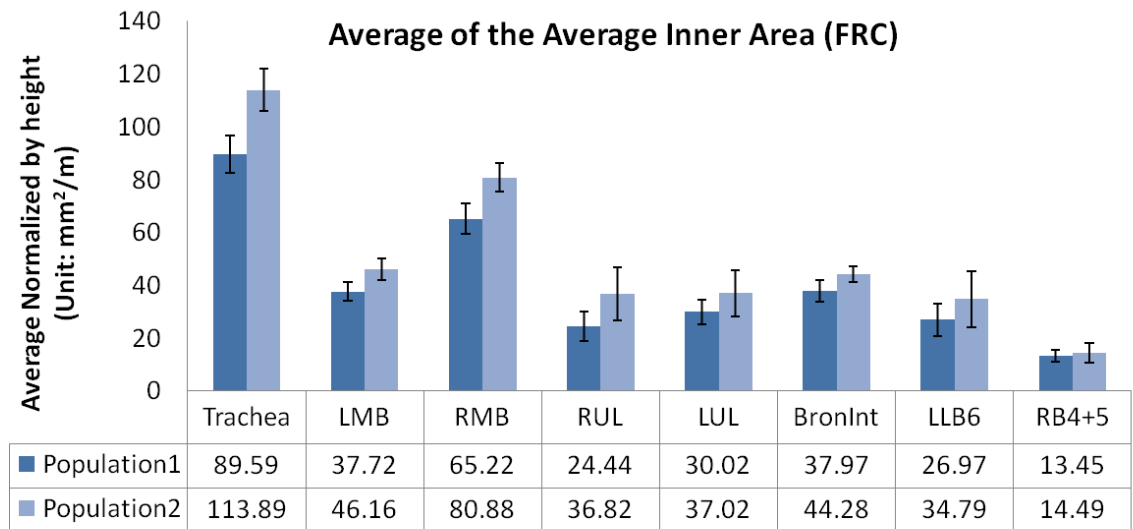
Figure 3.11: Study1: Population averages of the average inner area measurement normalized by height at FRC and TLC.

Average of the Average Outer Area (FRC)

	Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4+5
Population1	157.79	70.47	114.15	50.67	58.64	70.23	53.42	32.08
Population2	201.9	84.31	143.78	72.3	72.74	83.21	68.39	35.91

Average of the Average Outer Area (TLC)

	Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4+5
Population1	183.84	85.54	142.96	65.34	76.04	80.47	68.06	38.46
Population2	228.54	103.4	178.46	88.75	93.3	98.88	86.46	42.98
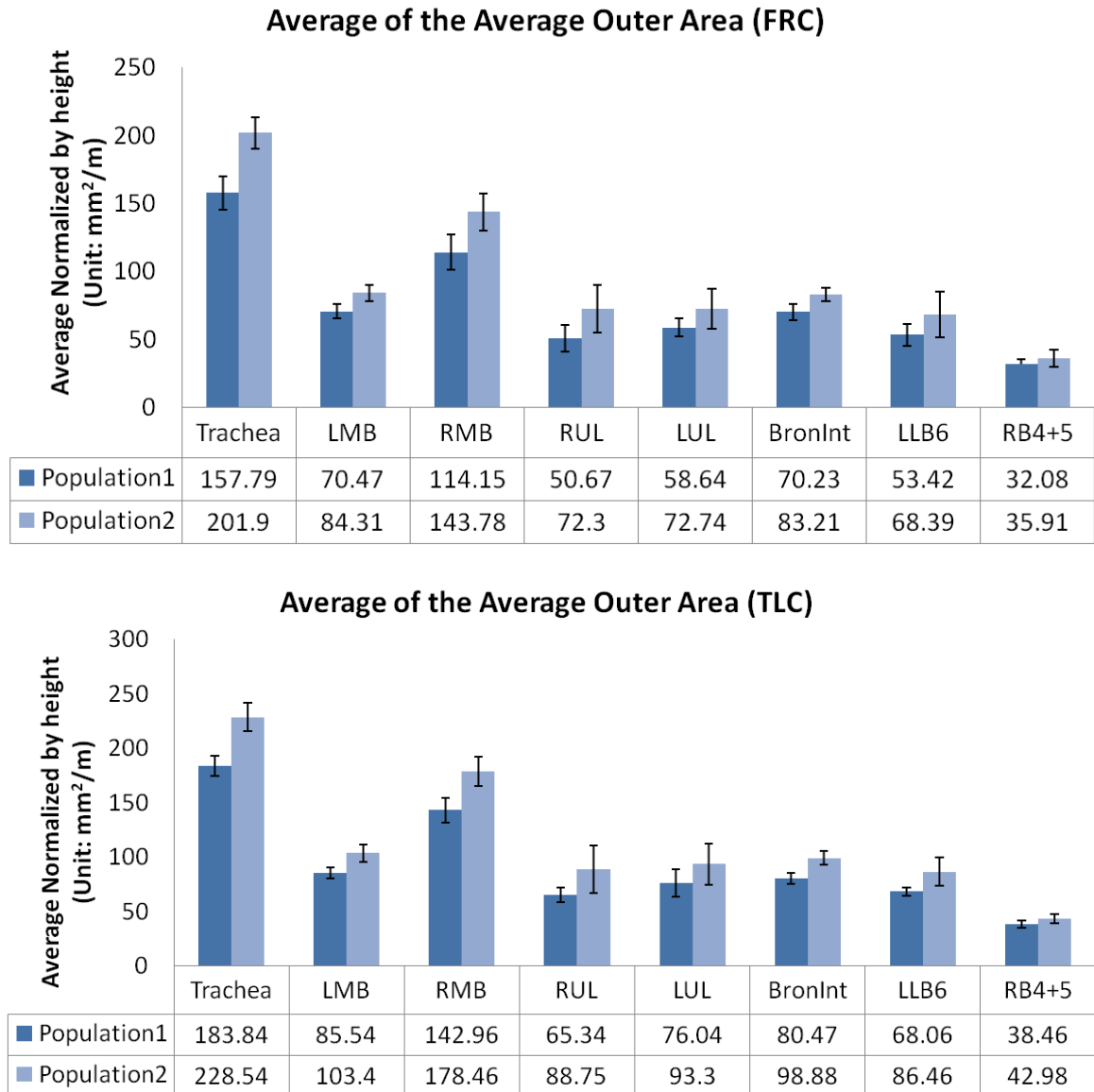
Figure 3.12: Study1: Population averages of the average outer area measurement normalized by height at FRC and TLC

Table 3.3: Study1: t-statistics, p-values and population counts for Figure 3.12

		Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4 + 5
FRC									
t-stat/p-value	Population 1	6.19	4.14	3.73	2.52	2.05	4.02	1.91	1.25
	Population 2	0	0	0	0.02	0.05	0	0.07	N/A
Count	Population 1	23	23	23	23	23	23	23	23
	Population 2	22	22	22	22	22	22	22	22
TLC									
t-stat/p-value	Population 1	6.6	4.53	4.78	2.42	1.8	5.47	3.19	1.97
	Population 2	0	0	0	N0.02	0.08	0	0	0.06
Count	Population 1	23	23	23	23	23	23	23	23
	Population 2	22	22	22	22	22	22	22	22

the calculated p-value is extremely small and approximately equal to zero and N/A indicates a p-value showing no significant differences. The graphs for Study 1 uses the same plotting format.

Comparing the measurements in Figures 3.11 and 3.12 shows that the trachea, LMB, RMB, RUL, BronInt measured by average inner area and outer area at FRC and TLC in young males are significantly larger than those of young females. LLB6 at TLC values for young males are significantly larger than those of young females, but the significant difference at FRC may depend on the chosen significant level. The values of young males and young females in RB4 + 5 at FRC and TLC have no significant differences. We choose significant level α to be 0.05.

In Figure 3.13, the unit in the vertical axis is mm/m, where mm comes from major inner diameter unit and m comes from height unit. Table 3.4 provides the t-statistics, p-values and counts for the Populations 1 and 2.

It can be observed that the trachea and LMB measured by average major inner

Average of the Average Major Inner Diameter (FRC)

	Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4+5
Population1	9.17	6.23	8.2	5.24	5.61	6.54	5.42	3.56
Population2	9.98	6.61	8.62	6.08	6.11	6.64	6.06	3.57

Average of the Average Major Inner Diameter (TLC)

	Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4+5
Population1	9.55	6.84	8.88	6.74	6.86	7.19	6.09	0.87
Population2	10.47	7.26	9.66	7.38	7.37	7.66	6.97	0.94

Figure 3.13: Study1:Population averages of the average major inner diameter measurement normalized by height at FRC and TLC

Table 3.4: Study1:t-statistics, p-values and population counts for Figure 3.13

		Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4 + 5
FRC									
t-stat/p-value	Population 1	3.79	2.26	1.5	1.75	0.96	0.4	0.98	0.04
	Population 2	0	0.03	0.15	0.09	N/A	N/A	N/A	N/A
Count	Population 1	23	23	23	23	23	23	23	23
	Population 2	22	22	22	22	22	22	22	22
TLC									
t-stat/p-value	Population 1	4.28	2.67	2.7	1.04	0.75	2.19	1.85	2.94
	Population 2	0	0.01	0	N/A	N/A	0.04	0.08	0
Count	Population 1	23	23	23	23	23	23	23	23
	Population 2	22	22	22	22	22	22	22	22

diameter at FRC and TLC in young males are significantly larger than those of young females. The values of RMB and RB4+5 at TLC in young males are significantlylarger than those of females, but there is no difference at FRC. The values of RUL, LUL, BronInt and LLB6 at FRC and TLC in young males and females have no significant differences. We choose significant level α to be 0.05.

Table 3.5: Study1:t-statistics, p-values and population counts for Figure 3.14

		Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4 + 5
FRC									
t-stat/p-value	Population 1	3.45	0.3	0.31	1.24	1.22	2.95	1.59	3
	Population 2	0	N/A	N/A	N/A	N/A	0	0.13	0
Count	Population 1	23	23	23	23	23	23	23	23
	Population 2	22	22	22	22	22	22	22	22
TLC									
t-stat/p-value	Population 1	1.83	2.53	1.56	2.01	0.89	2.9	1.95	2.94
	Population 2	0.08	0.02	0.14	0.05	N/A	0	0.06	0
Count	Population 1	23	23	23	23	23	23	23	23
	Population 2	22	22	22	22	22	22	22	22

Average of the Average Average Wallthickness (FRC)

	Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4+5
Population1	1.32	1.04	1.32	0.98	1.01	0.95	1.03	0.87
Population2	1.49	1.06	1.26	1.09	1.2	1.04	1.23	0.93

Average of the Average Average Wallthickness (TLC)

	Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4+5
Population1	1.38	0.96	1.67	0.93	1.1	0.96	1	0.87
Population2	1.44	1.03	1.29	1.05	1.23	1.04	1.15	0.94

Figure 3.14: Study1: Population averages of the average average wallthickness measurement normalized by height at FRC and TLC

In Figure 3.14, the unit in vertical axis is mm, where mm comes from wall thickness unit and m comes from height unit. Table 3.5 provides the t-statistics, p-values and counts for the Populations 1 and 2.

It can be observed that the values of the trachea measured by average wall thickness in young males at FRC are significantly larger than those of young females, but there is no significant difference at TLC. The values of BronInt and RB4+5 at FRC and TLC in young males are significantly larger than those of females. We choose significant level α to be 0.05.

Table 3.6: Study1: t-statistics, p-values and population counts for Figure 3.15

		both	left	LL	LU	right	RL	RM	RU
FRC									
t-stat/p-value	Population 1	1.3	1.13	0.52	1.67	1.45	0.49	2.74	1.64
	Population 2	N/A	N/A	N/A	0.11	0.16	N/A	0	0.11
Count	Population 1	20	20	20	20	20	20	20	20
	Population 2	17	17	17	17	17	17	17	17
TLC									
t-stat/p-value	Population 1	4.24	4.38	3.86	3.76	4.01	3.33	2.98	3.28
	Population 2	0	0	0	0	0	0	0	0
Count	Population 1	11	11	11	11	11	11	11	11
	Population 2	11	11	11	11	11	11	11	11

In Figure 3.15, the unit in vertical axis is cm^3/m, where cm^3 comes from total volume unit and m comes from height unit. Table 3.6 provides the t-statistics, p-values and counts for the Populations 1 and 2.

It can be observed that the values of both, left, left-lower (LL), left-upper

Average of the Total Volume (FRC)

	both	left	left-lower	left-upper	right	right-lower	right-middle	right-upper
Population1	1601.89	744.46	361.77	382.69	866.43	399.14	166.62	300.66
Population2	1823.95	837.88	384.03	453.85	986.07	420.93	211.42	353.73

Average of the Total Volume (TLC)

	both	left	left-lower	left-upper	right	right-lower	right-middle	right-upper
Population1	3013.46	1414.02	757.75	656.27	1599.44	820.56	244.35	534.53
Population2	3754.76	1773.03	950.57	822.45	1981.73	1017.52	305.44	658.76

Figure 3.15: Study1: Population averages of total volume measurement normalized by height at FRC and TLC

(LU), right, right-lower (RL), measured by total volume at TLC in young males are significantly larger than those of young females, but there is no difference at FRC. The values of right-middle (RU), right-upper (RU) at TLC and right-middle (RM) at FRC in young males are larger than those of females, but the significant differences depends on the chosen significant level. The values of right-upper at FRC in young males and females have no significant differences. We choose significant level α to be 0.05.

In Study 1 (and Study 3), gender was the main factor affecting study results. The average inner and outer area, average major inner diameter, and total volume measurements can effectively reflect the differences between female and male. The measured values show that the females have smaller airways and smaller lung volumes. There is no enough evidence to show what is the trend from younger to older population.

3.3.2   Study 2 : 20-29 and 40-49 Populations Comparison

In this study, two populations were compared in the airway population-to-population comparison form in Figure 3.7. One population (Population 1) represents individuals in the age range 20-29. The other population (Population 2) represents individuals in the age range of 40-49. Seven branches were selected: trachea, LMB, RMB, LB5, RB5, LB10 and RB10. The average major inner diameter, average minor inner diameter and average inner area measurements of these major seven airways were compared. The definitions of the above four measurements can be found in section 2.7. Height was selected as the value normalized.

Graphs in Figure A.1 show the population averages of the average major inner diameter measurement normalized by height at FRC and TLC for pre-selected seven branches. The average number of subjects used to calculate the population normalized average were 44, 44, 45 and 45 corresponding to Population1-FRC, Population2-FRC, Population1-TLC and Population2-TLC, respectively.

In these two figures, the vertical axis shows the average values of measurements in populations normalized by height, The unit in vertical axis is mm/m where m comes from height unit and mm comes from diameter unit. The horizontal axis shows the selected bronchi grouped by two populations. The error bars in the shape of "I" show the standard deviations of measurements in populations. The graphs in Study 2 uses the same plotting format.

Table A.1 provides the t-statistics, p value and population counts for Population 1 and 2 in Figure A.1.

The p-values for the trachea are less than 0.01 or smaller, which means the averages of the average major inner diameters at FRC and TLC between young and old generations are significant different. The values in the old population are all larger than the values in the young population. The deeper an airway is in the airway tree, the less significant differences exist. In the lower figure, for LB5, the graph shows that there are significant differences between the young and old populations.

Graphs in Figure A.2 show the population averages of the average minor inner diameter measurement normalized by height at FRC and TLC for pre-selected seven branches. The average number of subjects used to calculate the population normalized

average were 44, 44, 45 and 45 corresponding to Population1-FRC, Population2-FRC, Population1-TLC and Population2-TLC, respectively.

Table A.2 provides the t-statistics, p value and population counts for Population 1 and 2 in Figure A.2. We can see that for the trachea, LMB, RMB at FRC and TLC, significant population differences can be seen by their p-values.

Graphs in Figure A.3 show the population averages of the average inner area measurement normalized by height at FRC and TLC. The average number of subjects used to calculate the population normalized average were 44, 44, 45 and 45 corresponding to Population1-FRC, Population2-FRC, Population1-TLC and Population2-TLC, respectively.

Table A.3 provides the t-statistics, p value and population counts for Population 1 and 2 in Figure A.3. We can see that the average inner areas at FRC and TLC between the young and old populations show significant differences for the trachea, LMB in the graph. we can also notice that the average inner area for the older population are all larger than the values for the younger population.

In order to find the inner area measurement variation at FRC and TLC between young population and old population, we selected three main branches and plotted the histograms. The three main branches are trachea, LMB and RMB.

Graphs in Figure A.4 show the histogram of the average inner area of the Trachea for the young and old populations normalized by height at FRC and TLC. Figures A.5 and A.6 show the histograms of LMB and RMB normalized average inner area, respectively. These figures shows that the distribution of the older population

at FRC is shifted to right compared to the younger population. There is no apparent shift in the distributions at TLC. This suggests that the cross-section of Trachea, LMB and RMB may increase as a result of aging at FRC but not at TLC.

3.3.3 Study 3 : 40-49 Females and Males Comparison

In the study, we repeated the same procedure in section 3.3.1 and compare female and male populations in the age range of 40-49. The hypothesis was that the airways and lung volumes females are smaller than those of males in the age range 40-49.

The graphs in this section have the same plotting format. Population 1 indicates the normal females in the age range of 40-49 and Population 2 indicates the normal males in the age range of 40-49. The vertical axis shows the average values of measurements in populations normalized by height, The horizontal axis shows the selected bronchi grouped by two populations. The error bars in the shape of "I" show the standard deviations of measurements in populations.

In Figures B.1 and B.2, the unit in vertical axis is mm^2/m where m comes from height unit and mm^2 comes from area unit. Tables B.1 and B.2 provide the t-statistics, p-values and counts in Populations 1 and 2, where "0" means the calculated p-value is extremely small and approximately equal to zero and N/A indicates a p-value showing no significant differences.

We can observe that the values of trachea, LMB, RMB, BronInt and LLB6 at FRC and TLC measured by average inner and outer area in female population are significantly smaller that those of males. The values of RUL at FRC and TLC in

female population are smaller than those of males, but the significance may depend on the significant level. The values of LUL at TLC in female population are significantly smaller than those of males, but there is no significant difference at FRC. The values of RB4+5 at FRC and TLC show no significant differences between two populations.

In Figure B.3, the unit in vertical axis is mm/m where m comes from height unit and mm comes from inner diameter unit. Table B.3 provides the t-statistics, p-values and counts in Populations 1 and 2.

We can observe that the values of of trachea, LMB and RMB at FRC and TLC measured by major inner diameter in female population are significantly smaller that those of males. The values of RUL, LUL, BronInt, LLB6 and RB4+5 at FRC and TLC in female and male populations have no significant differences.

In Figure B.5, the unit in vertical axis is mm^3/m where m comes from height unit and mm^3 comes from volume unit. Table B.5 provides the t-statistics, p-values and counts for Populations 1 and 2.

We can observe that the values of of right-lower and left-lower at FRC measured by total in female and male populations have no significant differences. The values of both, left, left-upper, right-middle, right-upper at FRC in female population is significantly smaller than those of males. The values of left-lower, right, right-lower, right-middle at FRC in female and male populations have no significant differences. The values of left-upper and right-upper at TLC in female population are significantly smaller than those of males. The values of both and left at TLC in female population are smaller than those of males, but the significance may depend on the

chosen significant level.

3.3.4   Study 4 : 20-29 and 40-49 Females Comparison

In this study, we repeated the procedure in section 3.3.2 and compared age groups of 20-29 and 40-49 in female only population.

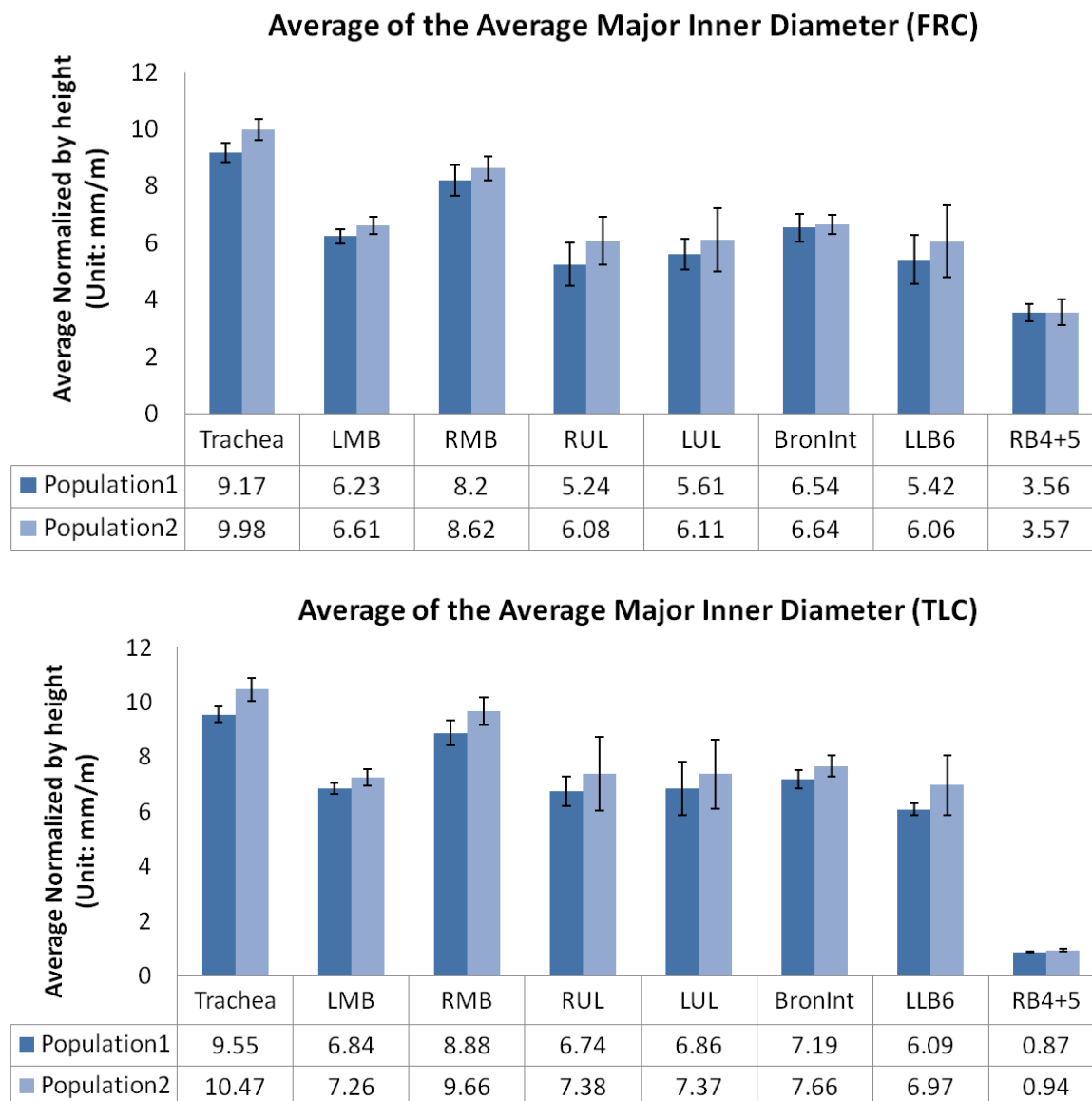Graphs in Figure C.1 show the population averages of the average major inner diameter measurement normalized by height at FRC and TLC for pre-selected 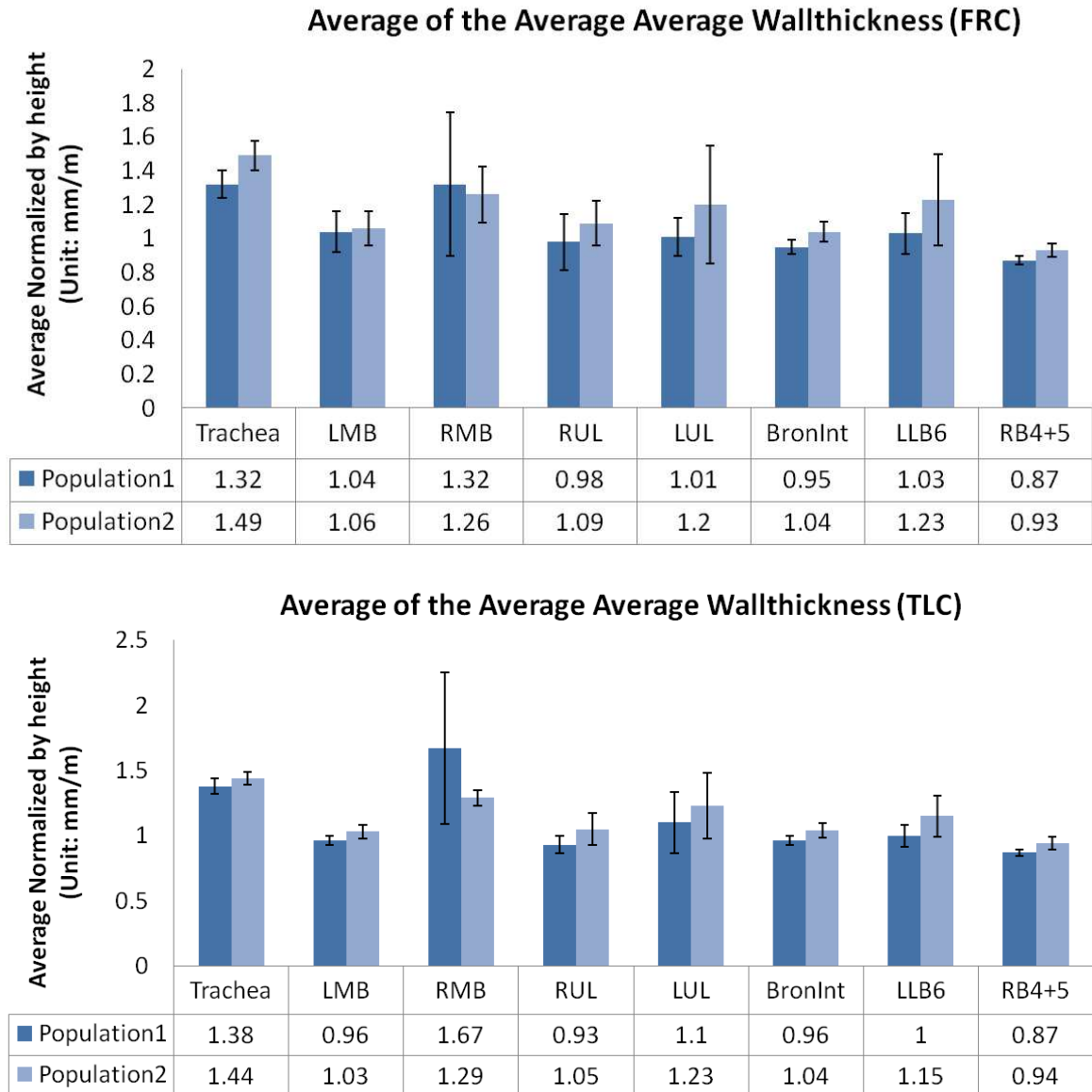seven branches. Table C.1 provides the t-statistics, p-values and counts for Figure C.1. The average number of subjects used to calculate the population normalized average were 23, 23, 23 and 23 corresponding to Population1-FRC, Population2-FRC, Population1-TLC and Population2-TLC, respectively. The averages of the average major inner diameters of the trachea at FRC and TLC in young population are significantly smaller than those of old generation.

Graphs in Figure C.2 show the population averages of the average minor inner diameter measurement normalized by height at FRC and TLC for pre-selected seven branches. Table C.2 provides the t-statistics, p-values and counts for Figure C.2. The average number of subjects used to calculate the population normalized average were 23, 23, 23 and 23 corresponding to Population1-FRC, Population2-FRC, Population1-TLC and Population2-TLC, respectively. The averages of the average minor inner diameters of the trachea at FRC and TLC in young population are significantly smaller than those of old generation.

Graphs in Figure C.3 show the population averages of the average inner area measurement normalized by height at FRC and TLC for pre-selected seven branches.

Table C.3 provides the t-statistics, p-values and counts for Figure C.3. The average number of subjects used to calculate the population normalized average were 23, 23, 23 and 23 corresponding to Population1-FRC, Population2-FRC, Population1-TLC and Population2-TLC, respectively. The averages of the average inner area of the trachea at FRC in young population are significantly smaller than those of old generation.

3.3.5   Study 5 : 20-29 and 40-49 Males Comparison

In this study, males in the age range of 20-29 and in the age range of 40-49 are compared.

Graphs in Figure D.1 show the population averages of the average major inner diameter measurement normalized by height at FRC and TLC for pre-selected seven branches. Table D.1 provides the t-statistics, p-values and counts for Figure D.1. The average number of subjects used to calculate the population normalized average were 22, 22, 15 and 16 corresponding to Population1-FRC, Population2-FRC, Population1-TLC and Population2-TLC, respectively. The averages of the average major inner diameters of the trachea and LB5 at FRC and TLC in young population are significantly smaller than those of old generation.

Graphs in Figure D.2 show the population averages of the average minor inner diameter measurement normalized by height at FRC and TLC for pre-selected seven branches. Table D.2 provides the t-statistics, p-values and counts for Figure D.2. The average number of subjects used to calculate the population normalized average were 22, 22, 15 and 16 corresponding to Population1-FRC, Population2-FRC, Population1-

TLC and Population2-TLC, respectively. The averages of the average minor inner diameters of LMB and LB5 at FRC and TLC in young population are significantly smaller than those of old generation.

Graphs in Figure D.3 show the population averages of the average inner area measurement normalized by height at FRC and TLC for pre-selected seven branches. Table D.3 provides the t-statistics, p-values and counts for Figure D.3. The average number of subjects used to calculate the population normalized average were 22, 22, 15 and 16 corresponding to Population1-FRC, Population2-FRC, Population1-TLC and Population2-TLC, respectively. The averages of the average major inner diameters of the trachea, LMB and LB5 at FRC and TLC in young population are significantly smaller than those of old generation.

In Study 2, 4, 5, age is the main factor affecting the results. The average inner, average major inner diameter and average minor inner diameter can effectively reflect the differences between young and old populations.

3.3.6 Study 6 : Normal and Abnormal Populations Comparison

In this study, a normal and an abnormal population are compared. The normal population contains all the normal subjects in the database including male and female. The abnormal population contains all the abnormal subjects. The number of abnormal subjects is small and may affect the comparison result. In this study, we chose eight representative bronchi: trachea, LMB, RMB, RUL, LUL, BronInt, LLB6 and RB4 + 5. Four representative measurements were used to analyze these eight bronchi at FRC and TLC: average inner area, average outer area, average major

inner diameter, average of average wallthickness.

Table 3.7 provides a comparison summary for average inner area, average outer area and average major inner diameter measurements. In the table, the measured values only for RB4+5 at FRC are listed. The values for other branches show no significant differences.

Table 3.7: Study6: Summary of Normal and Abnormal Populations Comparison for LLB6 at FRC.

		AvgInnerArea mm^2/m	AvgOuterArea mm^2/m	AvgMajorInnerDiameter mm/m
avgValue	Pop1	14.95	35.57	3.72
	Pop2	10.89	28.74	3.18
stdValue	Pop1	6.23	10.60	0.75
	Pop2	2.78	4.82	0.49
Tstat/Pvalue	Pop1	3.9	3.81	3.32
	Pop2	0	0	0
Count	Pop1	88	88	88
	Pop2	12	12	12

CHAPTER 4
DISCUSSION AND FUTURE WORK

4.1 Quality Assessment of Atlas Data

The atlas will be modified to perform data quality assessment in the future. This section describes the details of data quality assessment.

The overall atlas data acquisition follows four steps: obtain CT scans, upload images to MIFAR, import images to PW2 and build atlas database from PW2 measurement CSV files (see Figure 2.11). Two CT scanners were used to acquire lung CT images: a Siemens Sensation 64 multi-detector row CT scanner and a Philips MX 16 multi-detector row CT scanner. MIFAR is a relational database system management developed by the Iowa Comprehensive Lung Imaging Center (I-CLIC) and is used to store medical image data along with associated post-image processing files. The PW2 software analyzes 3D CT image data and produces segmented airways and lobes, and a comma-separated values (CSV) file containing airway measurements and volume histograms. The back-end atlas database was built based on these CSV files.

The core part of data acquisition is to use PW2. Getting started with PW2 includes five steps: scan image, transfer data, analyze scan quality, choose analysis method and run the analysis. For example, Figure 4.1 shows the PW2 airway analysis methodology. In the step of airway segmentation, airway skeletonization and airway coloring, the airway voxels are segmented from the CT scan and branches are assigned to the segmented airway automatically. Next, these automatic segmentations, skele-

tonizations and colorings are edited. In this step, airway tree branches can be added, deleted and edited, after which the segmentation, skeletonization and coloring algorithms must be repeated to update the segmented airways. Otherwise, human experts need to re-edit until the completion of airway segmentation, skeletonization and coloring processes. The following two steps are used to label segmented branches: airway tree branches are assigned labels and the labels can be added, deleted or edited. The last step is to measure airway segmentations: accurate measurements are extracted from the labeled airway tree.
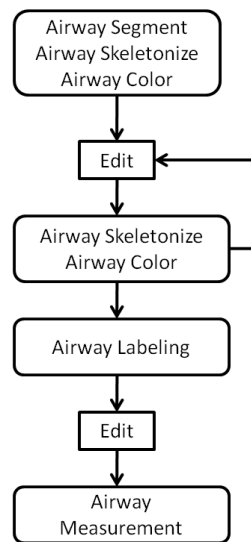
Figure 4.1: PW2 Airway Analysis Methodology.

In the PW2 airway analysis, 1161 branches in different airway generations can be segmented and labeled of which at least 32 major ones should be ready for analysis. However, in some cases, the segmented and labeled airways of a subject may

only include less than 32 branches. This situation is called missing airway labeling and several reasons can be responsible for this. The segmentation program may fail to segment a branch which directly leads to the missing labeling of this branch. For instance, the program may fail to segment LB5 from LB4+5, and it may label LB4+5 and LB5 as the single branch segmentation LB4+5. This leads to a small number of LB5 segmentations. In the left lung, LB4 and LB5 are near the heart. The heart beat affects lung CT images which leads to failures in the segmentation of LB4 and LB5. Patients feel more comfortable in TLC than in FRC and they may not hold their breath during FRC scanning. In this case, lung movement affects lung CT images and some lower branches may not be segmented.

In order to void missing airway labellings, the atlas will be modified to check the imported data. First, after airway segmentation, human experts need to visually go through airway trees of all the patients to make sure at least all 32 major branches are segmented. Second, after airway labeling, airway trees of all the patients need to be visually examined to make sure at least all major branches are labeled. After the database is imported from PW2, the atlas will start checking program for missing labels. The program will check airway labellings for each patient and ensure that the patient has all 32 major branches. If any missing branch is found, the atlas will give a warning to notify users to pay attention to this record, and provide users the option of deleting the patient's airway information. The name of the missing branch with patient information will be reported as feedback to human experts. Human experts will re-run the segmentation and labeling program to make sure the missing branch

is labeled. The updated data will be imported into the atlas and the same checking procedure is applied until no missing branch is found.

In the airway labeling program, mis-labellings may occur. For example, LMB and RMB labellings could be switched. In order to avoid airway mislabeling, human experts need to check several things for all the patients before atlas data is imported into atlas. First, parent-child branch relationship needs to be checked. For example, the parent and child branches of a left/right branch is also a left/right branch. Second, the branch generation number needs to be checked to make sure every branch has a correct generation number. For example, the trachea has generation number 0 and LMB, RMB have generation 1 (see Figure 2.13). Since PW2 software is case-sensitive while Microsoft Access is case-insensitive, every name of branches and measurements need to be checked to make sure the names in PW2 and Microsoft Access are consistent.

In some cases of airway analysis, two branches could be labeled as the same. For example, both LMB and RMB are labeled as LMB. This situation is called duplicate labeling. In order to avoid airway duplicate labeling, human experts need to check every patient's airway labellings by visual inspection before data is imported into atlas. This inspection will follow the airway tree and check all the major branches in the tree. This is aimed to make sure every patient has correct airway labellings without duplicates. After data imported from PW2, the atlas will start to check imported data to detect duplicate labellings. The program will go through all the patients' airway information to check they do not have duplicate labellings. If any

duplicate labels are found, a warning will be given to notify users that there might be an duplicate labeling the database. The atlas will report this labeling with patient information and send the report back to human experts as feedback. The updated data from human experts will be imported into the atlas and the same checking procedure is applied until no duplicate branch is found. In the latest PW2 program, duplicate labellings will not occur since attribute constrains in the database are set to ensure that no two branches share one name.

In the airway measurement analysis, there may occur that a branch may not be measured by all 122 measurements. For example, the trachea may only be measured by 121 measurements and it misses the average of average inner area measurement. This situation is called missing measurement. In order to detect this situation, the program will check all the patients' airway and measurement information and begin by checking the number of branches to make sure this patient has at least all the major branches. For each branch, the program will check the number of measurements to make sure each branch has all 122 measurements. If any missing measurement is found, the atlas will give a warning to notify users to pay attention to this record and provide users the option of deleting the airway measurement. The report of missing measurements will be sent back to human experts. The updated data will be imported into the atlas and the same checking procedure is applied until no missing measurement is found.

In some cases, airways are not measured correctly. For example, left main bronchus (LMB) is nearly $5cm$ long; however, the atlas may return the center line

length of LMB as $20cm$. This situation is called mis-measurement. The atlas will start to check airway measurements in order to make sure the measurements in the atlas have no outliers. For every branch of a patient, the program will first calculate the average and standard deviations of every single measurement. A threshold will be set as $\mu \pm 2\sigma$ to detect outliers. For each measurement, if its value is out of this range, it will be regarded as an outlier. the atlas will give a warning to notify users to pay attention to this record and provide users the option of deleting the airway measurement. This procedure will be applied to all the branches of all the patients. Human experts will examine the measurements, re-run the measurement analysis program and update the database. The atlas will import the updated database and repeat the checking procedure until no measurement outlier is detected.

4.2   Microsoft Access

The GUI development tool used in the atlas used Microsoft Access 2007. Access has many advantages in developing GUI. For example, it provides wizards and intuitive user interfaces to create tables, form, queries and reports step-by-step. In query design, four views are provides, such datasheet view, SQL view, design view. It is very easy to convert from one to another. Access has a high level of integration with other Microsoft Office products such as Excel and Word. It also offers several options for importing and exporting data. For example, we can import or link data from an external ODBC database or export data into the Excel format. In the atlas, we linked to the data in MySQL server and used it only for viewing and querying. This allows the data in MySQL server to change and the Access platform to use the

latest data. Access allows rapid data retrieval from tables or queries and stores the query results in the database.

However, Microsoft Access has some limitations as well. Access has difficulties dealing with large databases. The number of attributes in the tables is limited to 255, which is not enough for some larger databases. The query construction also meets the same problem. We cannot make long queries, otherwise, Access prompts an error message. For example, in the airway population-to-population comparison module, the maximum number of chosen measurements is limited to 9 given one chosen branch. To solve this problem, a threshold was set up and if the query exceeds the maximum query length in Access, a message box will pop up to notify the user to narrow down the query. Furthermore, only one developer at a time can modify the access front-end, which is error prone when merging code from multiple developers. These are the reasons why Microsoft Access is only used to create human lung atlas prototype, and will be replaced in the future development.

4.3   Quantitative Analysis of Human Lung

In section 3.3, only two populations were compared. The factors such as gender, age, race, ethnicity, PFT result, smoking status, lung capacity, branches, measurements and normalized options are all contained in the atlas. In the future, the atlas will be used to learn about many more differences and similarities between different populations.

According to Camp *et al.* [5], a good comparison study between sex should contain many factors and build multivariable models. For example, height, weight

and BMI should function simultaneously and a multivariable statistical model should be made to obtain more reliable results. PFT result may be divided into multiple levels and only normal or abnormal is not enough in the analysis.

Furthermore, the total number of subjects is an important factor as well. In this study, only 102 subjects are contained in the atlas consisting of 50 males and 52 females. Although this is a large number, it is not when selecting individuals in specific populations. For example, the number of subjects in a population may be only less than 20 after querying, which can not meet the statistical minimum number requirement in the central limit theorem. In order to obtain more significant results, more subjects will need to be added to the atlas in the future.

The atlas does not contain measurements for individuals at both FRC and TLC and this limits the intra-subject comparison between FRC and TLC.

4.4 Future Work

In the future, we will make improvements in the existing human lung atlas and incorporate lung images into the atlas. The existing atlas will be moved to a web-based application. We will not continue development using Microsoft Access to develop interfaces and may turn to find a more effective and efficient interface development tool.

In the future, we will provide visualization of 3D human lung with transverse, coronal and saggital views, as well as 2D display. The detailed labeling will be provided in the visualization part. A statistical part will exist as in the existing human lung atlas. The statistics can make image queries by selecting subject, branch

and measurement information. The results will be returned in the image format with numbers labeled rather than only numbers. The conversion between 2D and 3D display will be easily realized.

Furthermore, we will build a standard average shape of a normal human lung with variations. We will define a standard human lung coordinate system. The atlas will import external lung images and image registration will be used to warp individual lung images to this standard lung coordinate system. Statistical report will be provided to analyze these individual images given the standard normal lung.

CHAPTER 5
CONCLUSION

This thesis reports on our initial work constructing a human lung query atlas that provides clinically relevant population statistics for normal and abnormal individuals.

The atlas consists of a front-end user interface and a back-end database. The user interface was developed using Microsoft Access 2007 and the database was implemented using MySQL. ODBC was used to import database into Access and provide connection for database and interfaces. VBA was used to write SQL queries and realized the interaction with interfaces. SQL queries are written to extract the data which researchers may interest in.

The atlas provides measurements of the human airway tree and lung volumes from 148 CT images collected from a population of 102 individuals. It provides population statistics based on age, race, ethnicity, gender and other information. It provides functionality for comparing airway measurements between populations, individuals to a population, and individuals to individuals. Statistical significance, such as p-value, is provided to compare individuals and populations.

APPENDIX A
FIGURES FOR STUDY 2

Table A.1: Study2: t-statistics, p-values and population counts for Figure A.1

		Trachea	LMB	RMB	LB5	RB5	LB10	RB10
FRC								
t-stat/p-value	Population 1	3.48	3.05	1.74	2.21	1.56	0.39	0.58
	Population 2	0	0	0.09	0.03	0.13	N/A	N/A
Count	Population 1	45	45	45	40	43	45	44
	Population 2	33	33	33	30	31	30	32
TLC								
t-stat/p-value	Population 1	3.17	1.91	0.28	2.73	0.55	1.22	1.55
	Population 2	0	0.06	N/A	0	N/A	N/A	0.13
Count	Population 1	45	45	45	45	44	45	45
	Population 2	34	34	34	34	32	34	33

Table A.2: Study2: t-statistics, p-values and population counts for Figure A.2

		Trachea	LMB	RMB	LB5	RB5	LB10	RB10
FRC								
t-stat/p-value	Population 1	3.77	4.16	2.82	1.9	1.01	0.1	1.43
	Population 2	0	0	0	0.07	N/A	N/A	0.16
Count	Population 1	45	45	45	40	43	45	44
	Population 2	33	33	33	30	31	30	32
TLC								
t-stat/p-value	Population 1	2.15	3.34	2.48	2.4	0.86	1.16	0.88
	Population 2	0.04	0	0.02	0.02	N/A	N/A	N/A
Count	Population 1	45	45	45	45	44	45	45
	Population 2	34	34	34	34	32	34	33

Table A.3: Study2: t-statistics, p-values and population counts for Figure A.3

		Trachea	LMB	RMB	LB5	RB5	LB10	RB10
FRC								
t-stat/p-value	Population 1	3.58	3.95	2.46	2.05	1.21	0.04	0.34
	Population 2	0	0	0.02	0.05	N/A	N/A	N/A
Count	Population 1	45	45	45	40	43	45	44
	Population 2	33	33	33	30	31	30	32
TLC								
t-stat/p-value	Population 1	2.6	2.8	1.09	2.77	0.56	1.4	1.5
	Population 2	0.01	0	N/A	0	N/A	0.17	0.15
Count	Population 1	45	45	45	45	44	45	45
	Population 2	34	34	34	34	32	34	33

Average of the Average Major Inner Diameters (FRC)

	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
■ Population1	9.56	6.41	8.4	2.33	2.62	2.88	2.78
■ Population2	10.31	6.89	8.78	2.6	2.8	2.94	2.67

Average of the Average Major Inner Diameters (TLC)

	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
■ Population1	9.99	7.04	9.25	2.68	3.04	3.4	3.12
■ Population2	10.69	7.35	9.19	2.97	3.1	3.55	3.41
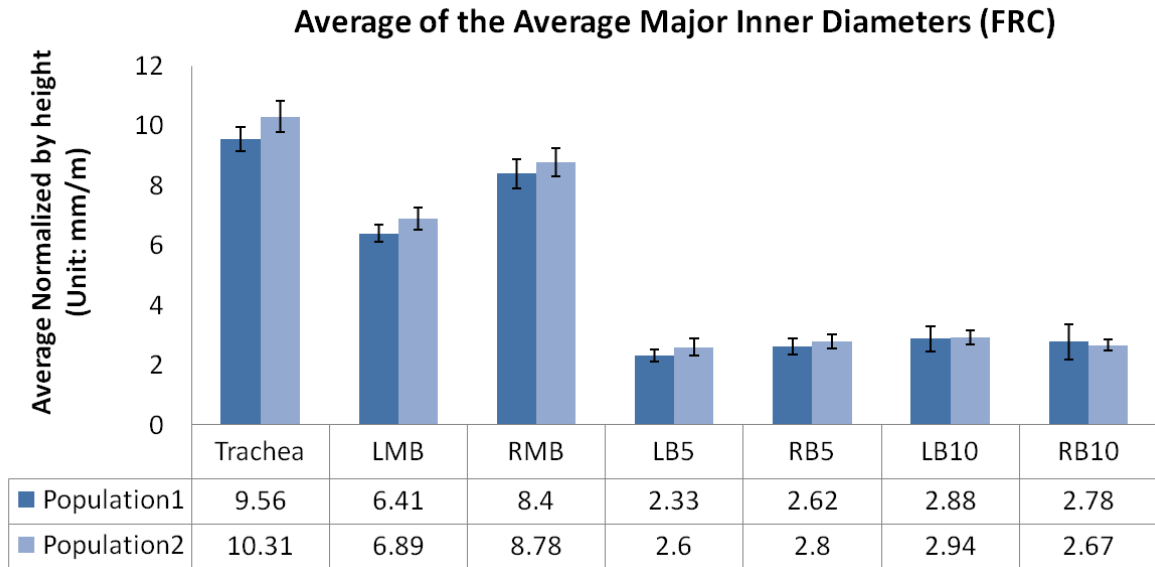
Figure A.1: Study2: Population averages of the average major inner diameter measurement normalized by height at FRC and TLC

Average of the Average Minor Inner Diameters (FRC)

	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
■ Population1	7.67	4.71	6.16	1.72	1.98	2.17	2.02
■ Population2	8.44	5.35	6.65	1.92	2.08	2.18	2.14

Average of the Average Minor Inner Diameters (FRC)

	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
■ Population1	7.67	4.71	6.16	1.72	1.98	2.17	2.02
■ Population2	8.44	5.35	6.65	1.92	2.08	2.18	2.14
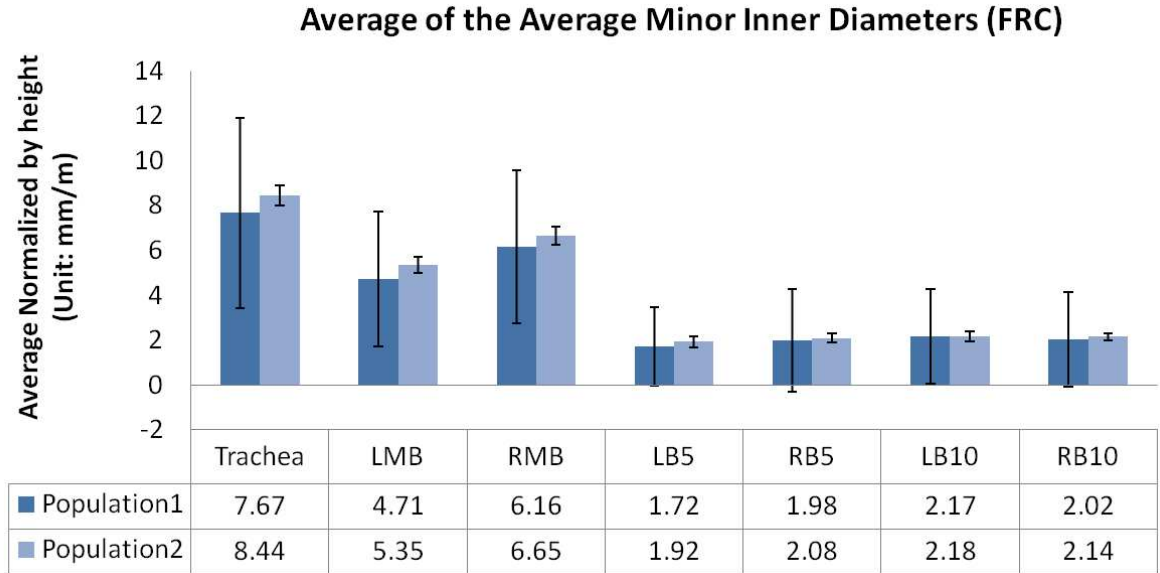
Figure A.2: Study2: Population averages of the average minor inner diameter measurement normalized by height at FRC and TLC.

Average of the Average Inner Areas (FRC)

	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
Population1	101.19	41.75	72.7	5.67	7.35	8.97	8.25
Population2	119.65	51.84	81.54	7.21	8.17	9	7.93

Average of the Average Inner Areas (TLC)

	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
Population1	120.34	54.15	91.39	7.65	9.76	12.4	10.95
Population2	135.94	62.86	95.97	9.6	10.16	13.76	12.56
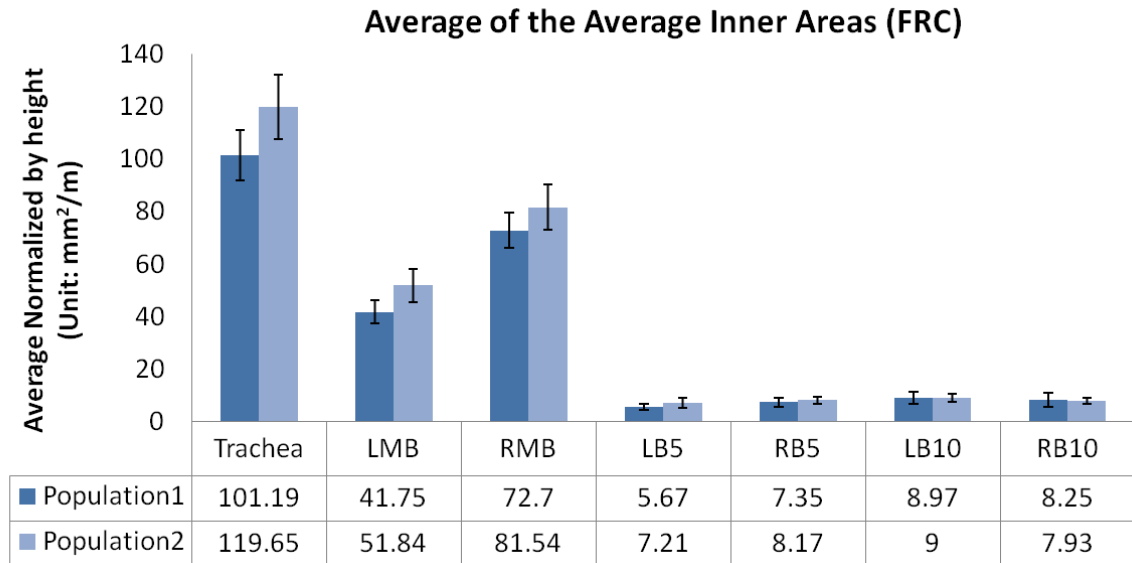
Figure A.3: Study2: Population averages of the average inner area measurement normalized by height at FRC and TLC.

Histogram of the Trachea Average Inner Area at FRC normalized by height (mm2/m)

	80	90	100	110	120	130	140	150	160	170	180
Population1	6	5	7	6	3	7	4	0	0	0	0
Population2	0	2	5	3	5	4	5	2	0	3	1

Histogram of the Trachea Average Inner Area at TLC normalized by height (mm^2/m)

	80	90	100	110	120	130	140	150	160	170	180
Population1	0	1	2	3	8	4	2	2	1	1	0
Population2	0	1	1	1	3	3	3	2	3	0	0
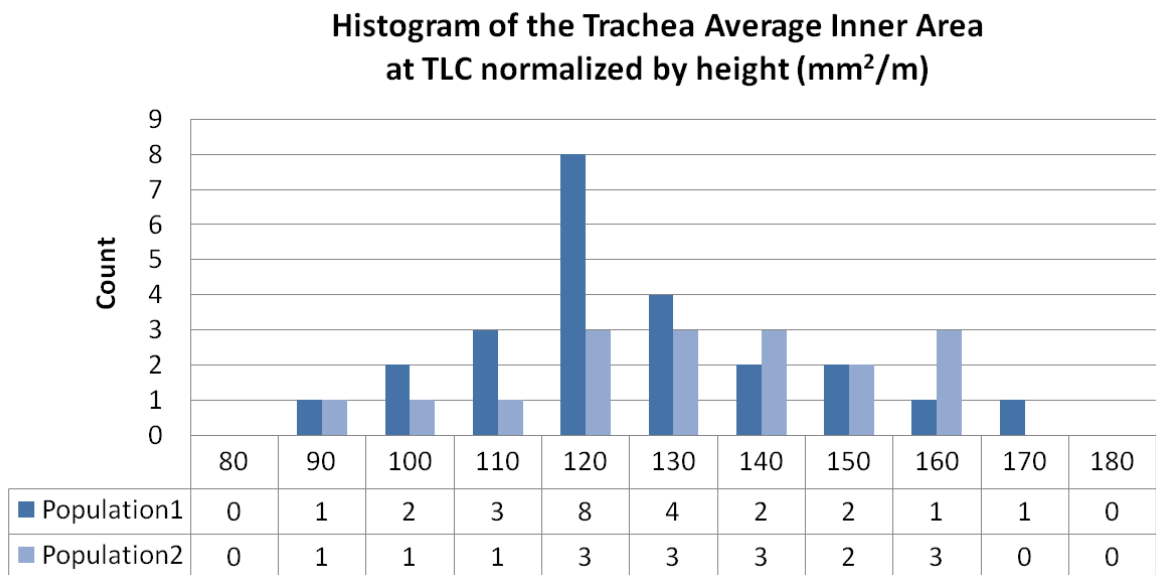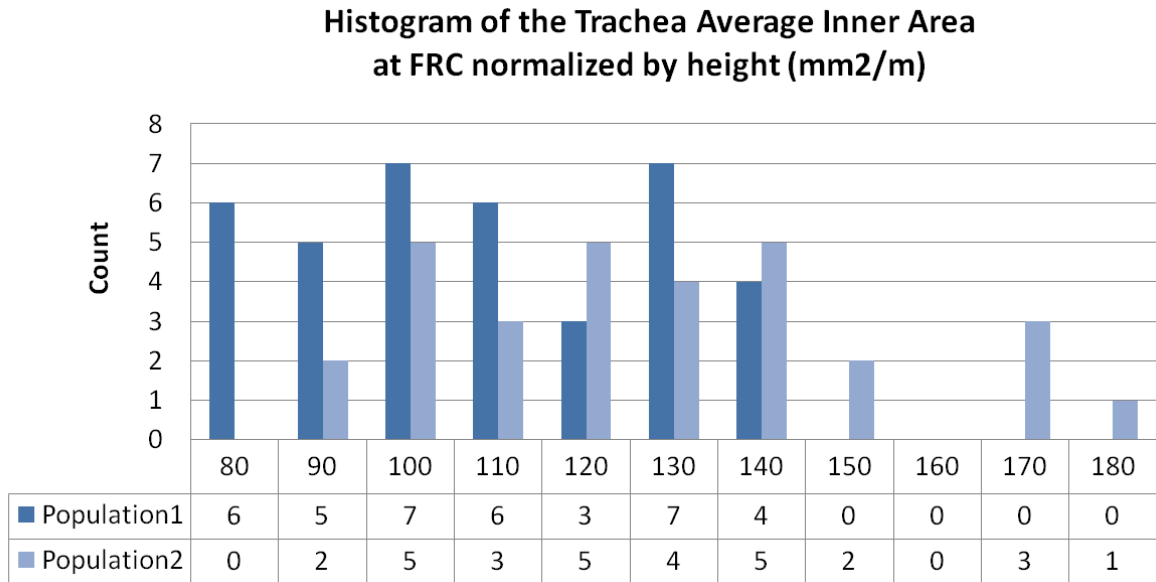
Figure A.4: Study2: Histograms of trachea average inner area measurement normalized by height at FRC and TLC.

Histogram of the LMB Average Inner Area at FRC normalized by height (mm²/m)

	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
Population1	3	4	8	11	2	6	2	2	0	0	0	0	0	0	0
Population2	0	2	1	6	1	5	6	3	4	1	0	1	0	0	0

Histogram of the LMB Average Inner Area at TLC normalized by height (mm²/m)

	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
Population1	0	1	1	2	3	6	2	5	2	0	1	1	0	0	0
Population2	0	0	2	1	2	0	3	3	1	1	2	0	1	1	1
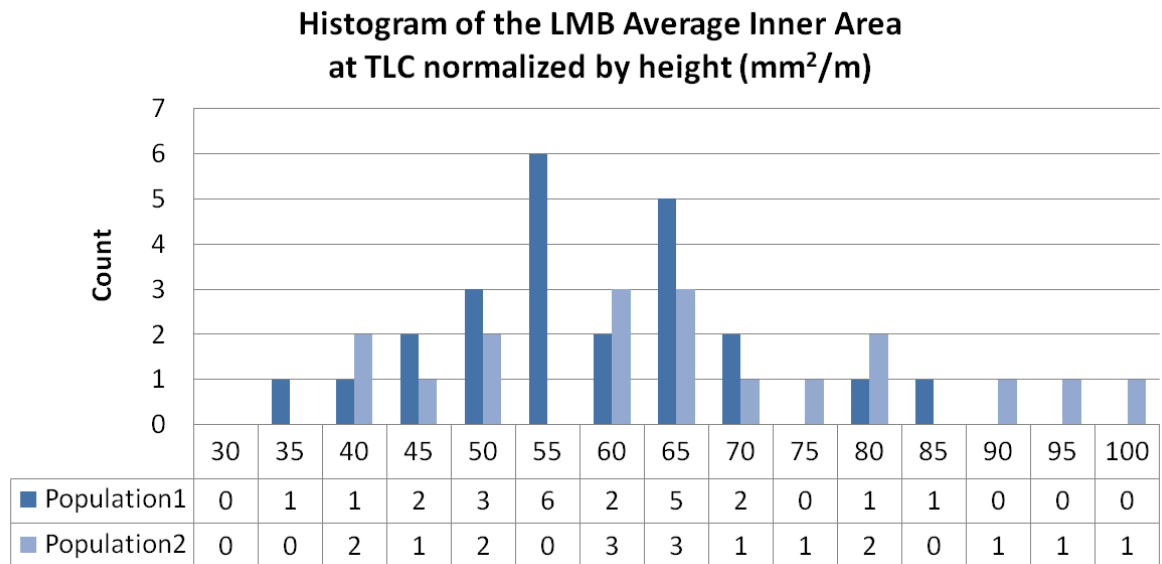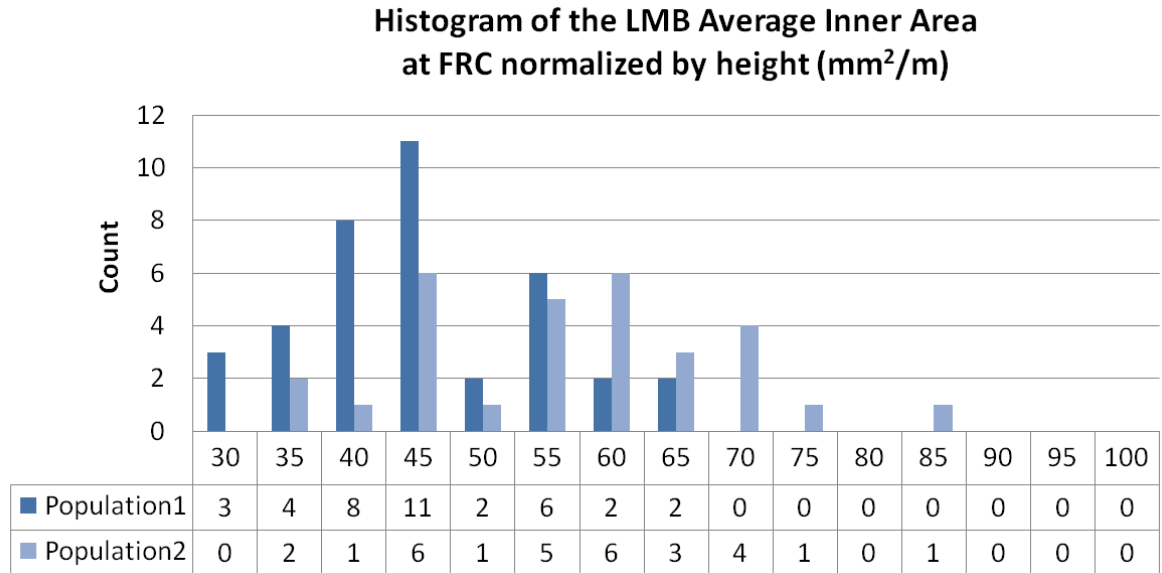
Figure A.5: Study2: Histograms of LMB average inner area measurement normalized by height at FRC and TLC.

Histogram of the RMB Average Inner Area at FRC normalized by height (mm²/m)

	50	55	60	65	70	75	80	85	90	95	100	105	110	115	120	125	130	135
Population1	1	2	1	6	4	5	7	6	4	1	2	2	0	0	0	0	0	0
Population2	0	0	1	4	2	2	4	4	3	3	2	3	0	0	1	1	0	0

Histogram of the RMB Average Inner Area at TLC normalized by height (mm²/m)

	50	55	60	65	70	75	80	85	90	95	100	105	110	115	120	125	130	135
Population1	0	0	1	0	2	2	2	1	3	3	1	4	1	1	2	0	1	0
Population2	0	0	0	0	1	1	0	2	1	1	2	3	2	2	1	0	0	2
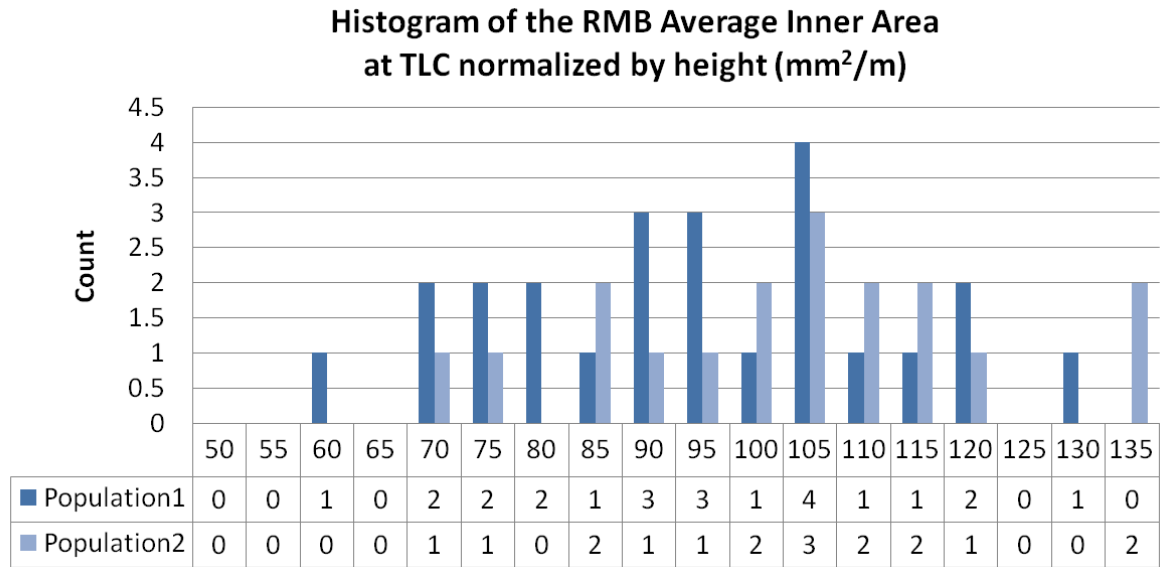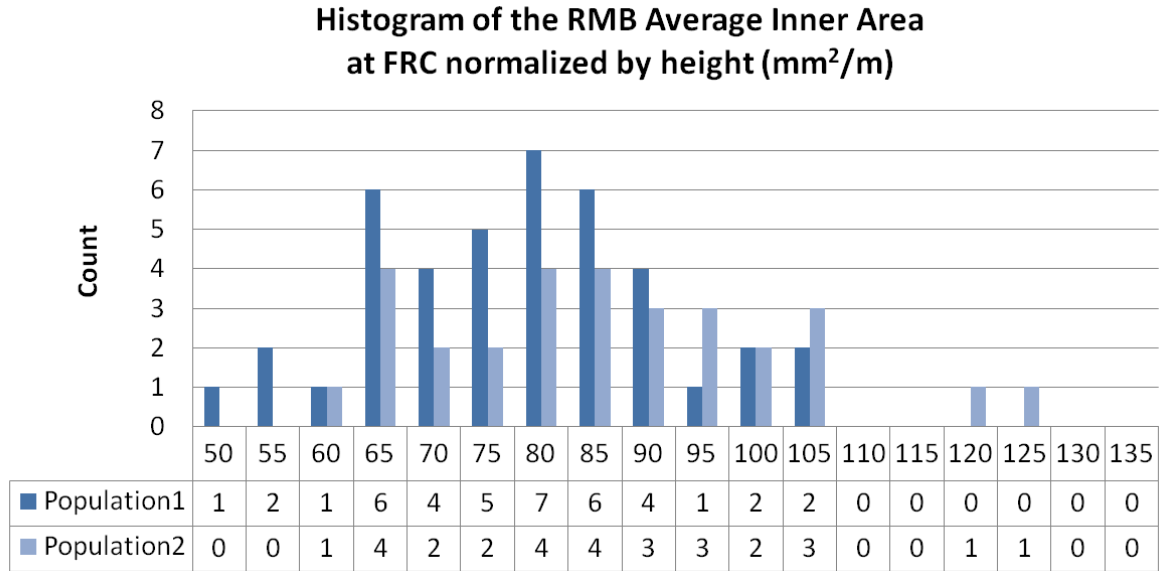
Figure A.6: Study2: Histograms of RMB average inner area measurement normalized by height at FRC and TLC.

APPENDIX B
FIGURES FOR STUDY 3

Table B.1: Study3:t-statistics, p-values and population counts for Figure B.1

		Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4 + 5
FRC									
t-stat/p-value	Population 1	5.09	4.38	4.04	1.94	1.67	2.8	2.55	1.24
	Population 2	0	0	0	0.07	0.11	0	0.02	N/A
Count	Population 1	17	17	17	17	17	17	17	17
	Population 2	16	16	16	16	16	16	16	16
TLC									
t-stat/p-value	Population 1	4.87	4.34	4.3	1.99	3.15	3.71	2.99	1.34
	Population 2	0	0	0	0.07	0	0	0	0.19
Count	Population 1	18	18	18	18	18	18	18	18
	Population 2	16	16	16	16	16	16	16	16

Table B.2: Study3: t-statistics, p-values and population counts for Figure B.2

		Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4 + 5
FRC									
t-stat/p-value	Population 1	5.31	4.48	3.82	2.05	2.03	2.74	2.69	1.41
	Population 2	0	0	0	0.05	0.05	0.01	0.01	0.18
Count	Population 1	17	17	17	17	17	17	17	17
	Population 2	16	16	16	16	16	16	16	16
TLC									
t-stat/p-value	Population 1	4.97	3.75	4	2.03	2.83	3.28	3.13	1.55
	Population 2	0	0	0.06	0	0	0	0	0.14
Count	Population 1	18	18	18	18	18	18	18	18
	Population 2	16	16	16	16	16	16	16	16

Table B.3: Study3: t-statistics, p-values and population counts for Figure B.3

		Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4 + 5
FRC									
t-stat/p-value	Population 1	4.57	2.47	3.33	0.65	0.4	1.26	1.61	0.72
	Population 2	0	0.02	0	N/A	N/A	*N/A*	0.13	N/A
Count	Population 1	17	17	17	17	17	17	17	17
	Population 2	16	16	16	16	16	16	16	16
TLC									
t-stat/p-value	Population 1	4.01	2.85	2.73	1.35	1.71	2.24	1.65	0.78
	Population 2	0	0	0.01	0.2	0.1	0.04	0.11	N/A
Count	Population 1	18	18	18	18	18	18	18	18
	Population 2	16	16	16	16	16	16	16	16

Figure B.1: Study3: Population averages of the average inner area measurement normalized by height at FRC and TLC

Average of the Average Outer Area (FRC)

	Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4+5
Population1	181.82	80.45	124.42	56.32	61.14	78.7	51.03	34.23
Population2	233.04	105.38	161.12	67.56	74.33	91.79	72.69	38.4

Average of the Average Outer Area (TLC)

	Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4+5
Population1	201.76	96.23	149.43	68.35	71.86	88.86	68.18	40.13
Population2	255.08	125.51	186.4	101.14	89.79	107.92	83.37	43.85

Figure B.2: Study3: Population averages of the average outer area measurement normalized by height at FRC and TLC

Average of the Average Major Inner Diameter (FRC)

	Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4+5
Population1	9.69	6.6	8.31	5.61	5.93	6.85	4.91	3.71
Population2	10.97	7.2	9.28	5.8	6.14	7.13	5.88	3.87

Average of the Average Major Inner Diameter (TLC)

	Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4+5
Population1	10.12	7.01	8.85	6.42	6.22	7.34	5.92	4.13
Population2	11.34	7.73	9.57	7.8	6.7	7.8	6.31	4.27

Figure B.3: Study3: Population averages of the average major inner diameter measurement normalized by height at FRC and TLC

Average of the Average Average Wallthickness (FRC)

	Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4+5
Population1	1.41	1.01	1.21	1.01	0.99	1.03	1.15	0.91
Population2	1.48	1.06	1.3	1.02	1.2	1	1.24	0.98

Average of the Average Average Wallthickness (TLC)

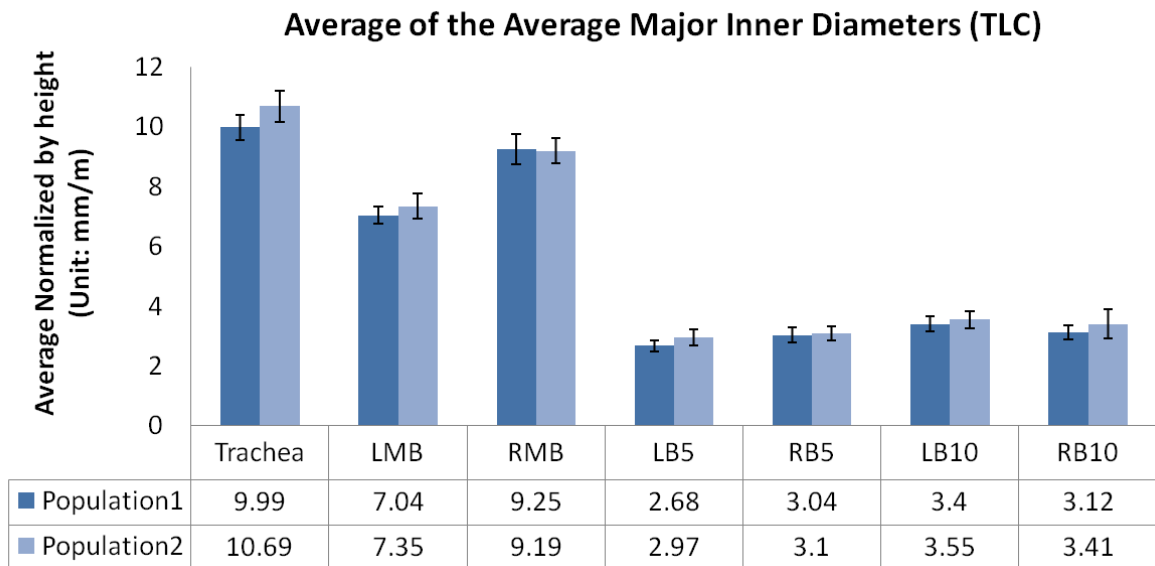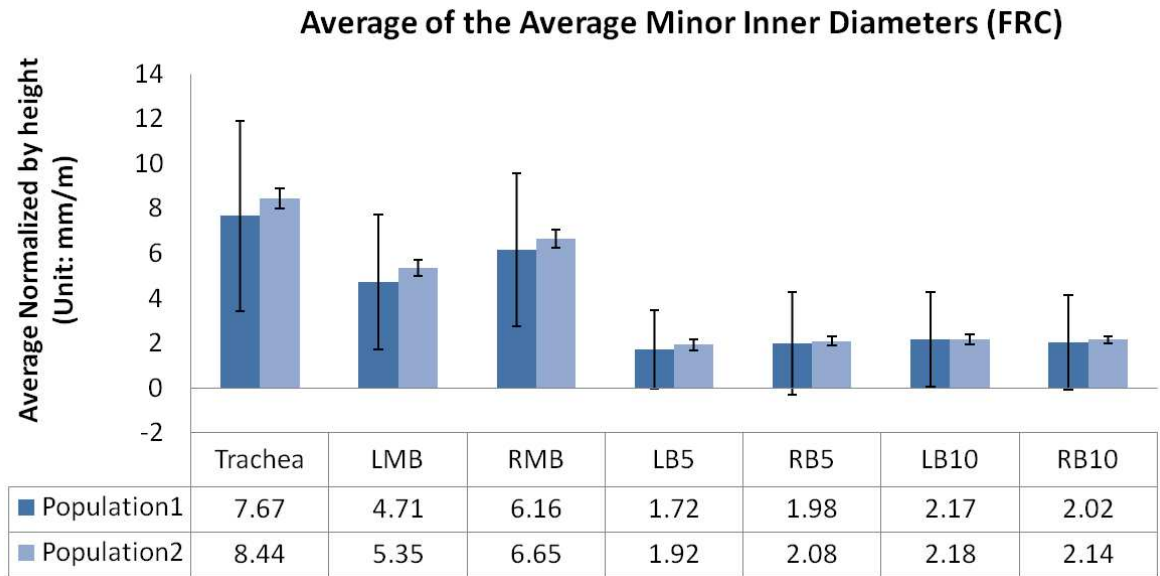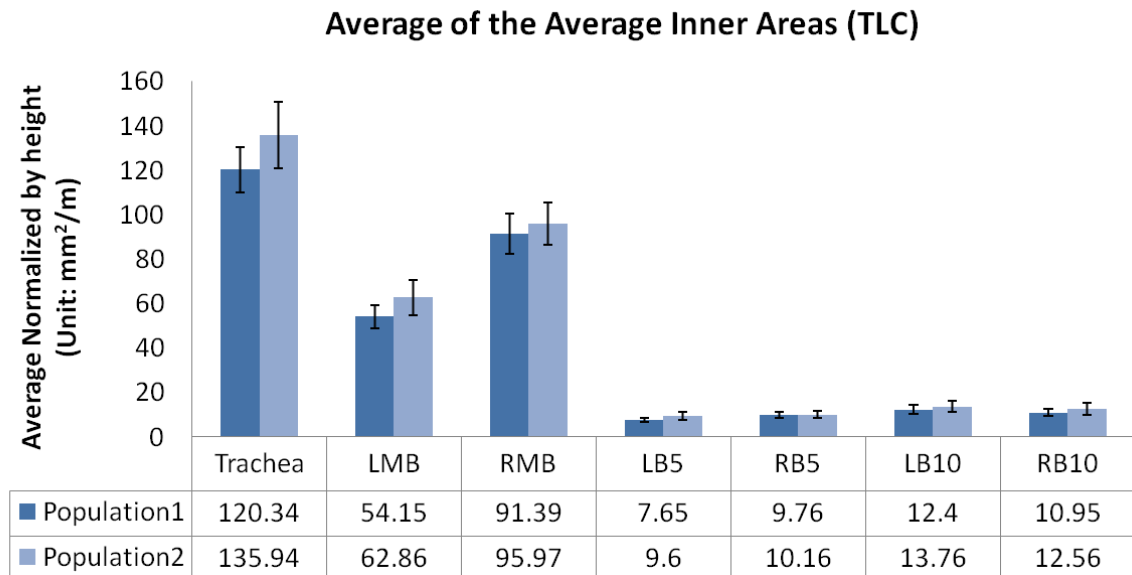|	Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4+5
Population1	1.44	1.07	1.27	1.02	1.02	1.08	1	0.9
Population2	1.42	1.1	1.36	1.2	1.28	1.01	1.15	0.89

Figure B.4: Study3: Population averages of the average average wallthickness measurement normalized by height at FRC and TLC

Table B.4: Study3: t-statistics, p-values and population counts for Figure B.4

		Trachea	LMB	RMB	RUL	LUL	BronInt	LLB6	RB4 + 5
FRC									
t-stat/p-value	Population 1	1.34	1.35	1.22	0.17	1.03	0.81	0.46	0.88
	Population 2	0.19	0.19	N/A	N/A	N/A	N/A	N/A	N/A
Count	Population 1	17	17	17	17	17	17	17	17
	Population 2	16	16	16	16	16	16	16	16
TLC									
t-stat/p-value	Population 1	0.72	0.6	1.42	1.05	1.24	0.78	1.02	0.42
	Population 2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Count	Population 1	18	18	18	18	18	18	18	18
	Population 2	16	16	16	16	16	16	16	16

Table B.5: Study3: t-statistics, p-values and population counts for Figure B.5

		both	left	LL	LU	right	RL	RM	RU
FRC									
t-stat/p-value	Population 1	2.51	2.63	1.68	3.27	2.32	0.88	3.23	2.72
	Population 2	0.03	0.02	0.12	0	0.04	N/A	0	0.02
Count	Population 1	10	10	10	10	10	10	10	10
	Population 2	10	10	10	10	10	10	10	10
TLC									
t-stat/p-value	Population 1	2.75	3.05	1.94	3.4	2.22	0.38	0.7	4.12
	Population 2	0.02	0	0.08	0	0.05	N/A	N/A	0
Count	Population 1	8	8	8	8	8	8	8	8
	Population 2	9	9	9	9	9	9	9	9

Average of the Total Volume (FRC)

	both	left	left-lower	left-upper	right	right-lower	right-middle	right-upper
Population1	1590.85	738.87	327.59	411.28	851.98	358.95	166.17	326.86
Population2	2076.39	994.32	416.36	577.96	1082.07	400.74	233.87	447.46

Average of the Total Volume (TLC)

	both	left	left-lower	left-upper	right	right-lower	right-middle	right-upper
Population1	3169.7	1488.06	770.31	717.75	1681.64	855.58	275.59	550.46
Population2	3728.26	1806.45	897.59	908.86	1921.81	887.83	255.35	778.63

Figure B.5: Study3: Population averages of total volume measurement normalized by height at FRC and TLC

APPENDIX C
FIGURES FOR STUDY 4

Table C.1: Study4: t-statistics, p-values and population counts for Figure C.1

		Trachea	LMB	RMB	LB5	RB5	LB10	RB10
FRC								
t-stat/p-value	Population 1	2.51	1.96	0.39	0	1.26	0	0.32
	Population 2	0.02	0.06	N/A	N/A	N/A	N/A	N/A
Count	Population 1	23	23	23	22	22	23	23
	Population 2	17	17	17	15	16	15	17
TLC								
t-stat/p-value	Population 1	2.75	0.88	0.12	0.71	0.2	0.55	0.92
	Population 2	0	N/A	N/A	N/A	N/A	N/A	N/A
Count	Population 1	23	23	23	23	22	23	23
	Population 2	18	18	18	18	18	18	18

Table C.2: Study4: t-statistics, p-values and population counts for Figure C.2

		Trachea	LMB	RMB	LB5	RB5	LB10	RB10
FRC								
t-stat/p-value	Population 1	2.85	2.36	2.36	0.5	0.7	0.4	1.66
	Population 2	0	0.03	0.03	N/A	N/A	N/A	0.11
Count	Population 1	23	23	23	22	22	23	23
	Population 2	17	17	17	15	16	15	17
TLC								
t-stat/p-value	Population 1	1.59	1.66	2.12	1.04	0.35	0.63	1.2
	Population 2	0.13	0.11	0.04	N/A	N/A	N/A	N/A
Count	Population 1	23	23	23	23	22	23	23
	Population 2	18	18	18	18	18	18	18

Table C.3: Study4: t-statistics, p-values and population counts for Figure C.3

		Trachea	LMB	RMB	LB5	RB5	LB10	RB10
FRC								
t-stat/p-value	Population 1	2.99	2.45	1.74	0.1	0.92	0.47	0.22
	Population 2	0	0.02	0.09	N/A	N/A	N/A	N/A
Count	Population 1	23	23	23	22	22	23	23
	Population 2	17	17	17	15	16	15	17
TLC								
t-stat/p-value	Population 1	2.33	1.43	1.25	0.75	0.23	0.61	1.31
	Population 2	0.03	0.17	N/A	N/A	N/A	N/A	N/A
Count	Population 1	23	23	23	23	22	23	23
	Population 2	18	18	18	18	18	18	18

Average of the Average Major Inner Diameter (FRC)

	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
Population1	9.17	6.23	8.2	2.31	2.55	2.83	2.5
Population2	9.69	6.6	8.31	2.31	2.75	2.83	2.55

Average of the Average Major Inner Diameter (TLC)

	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
Population1	9.55	6.84	8.88	2.62	3.02	3.43	3.05
Population2	10.12	7.01	8.85	2.7	2.99	3.51	3.17
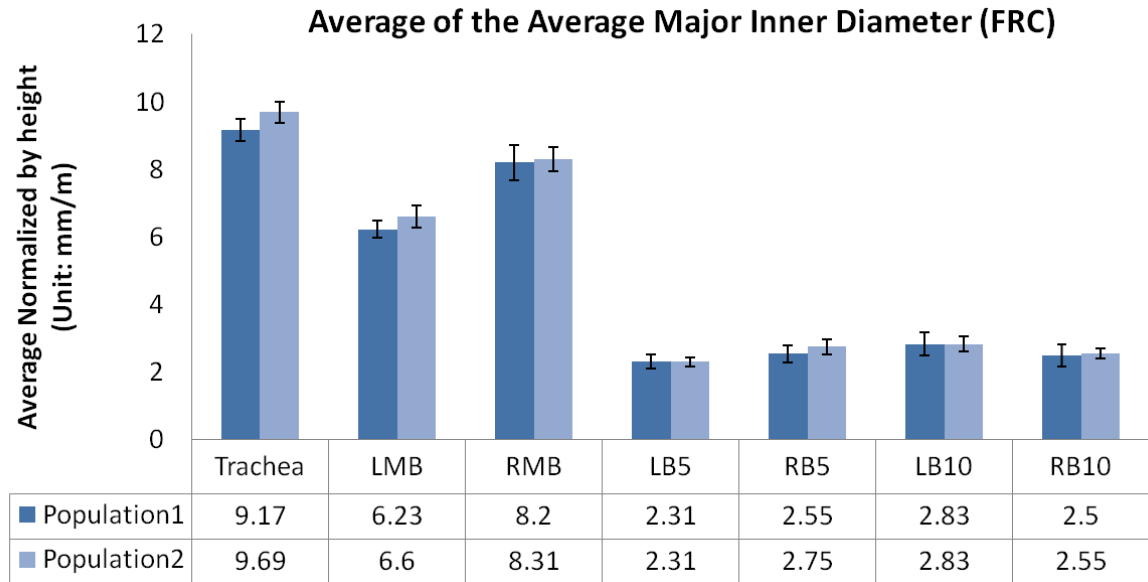
Figure C.1: Study4: Population averages of the average major inner diameter measurement normalized by height at FRC and TLC

Average of the Average Minor Inner Diameter (FRC)

	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
Population1	7.34	4.53	5.91	1.63	1.88	2.16	1.89
Population2	8.06	5.01	6.44	1.68	1.97	2.12	2.04

Average of the Average Minor Inner Diameter (TLC)

	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
Population1	8.5	5.45	6.82	1.95	2.21	2.61	2.41
Population2	8.83	5.78	7.24	2.04	2.25	2.7	2.55
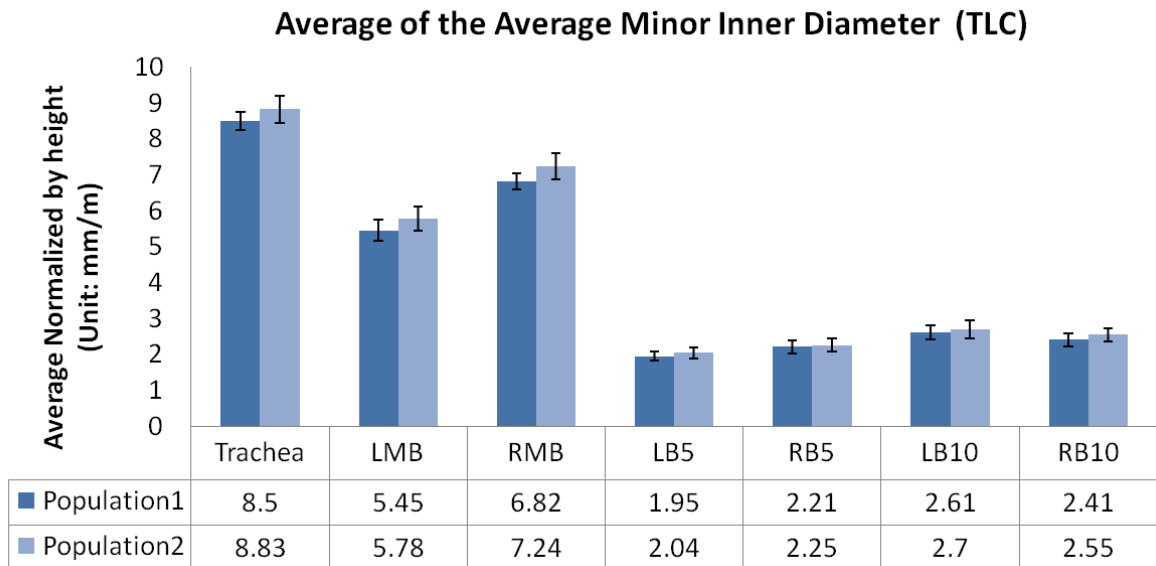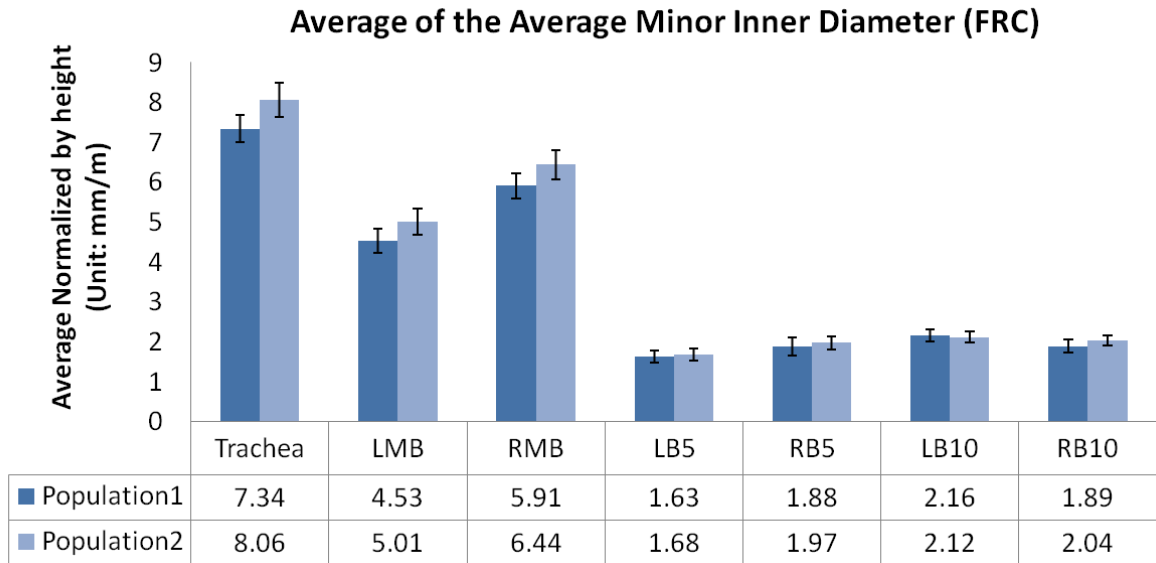
Figure C.2: Study4: Population averages of the average minor inner diameter measurement normalized by height at FRC and TLC

Average of the Average Inner Area (FRC)

	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
Population1	89.59	37.72	65.22	5.17	6.58	8.38	6.67
Population2	103.7	44.33	71.94	5.11	7.3	7.98	6.84

Average of the Average Inner Area (TLC)

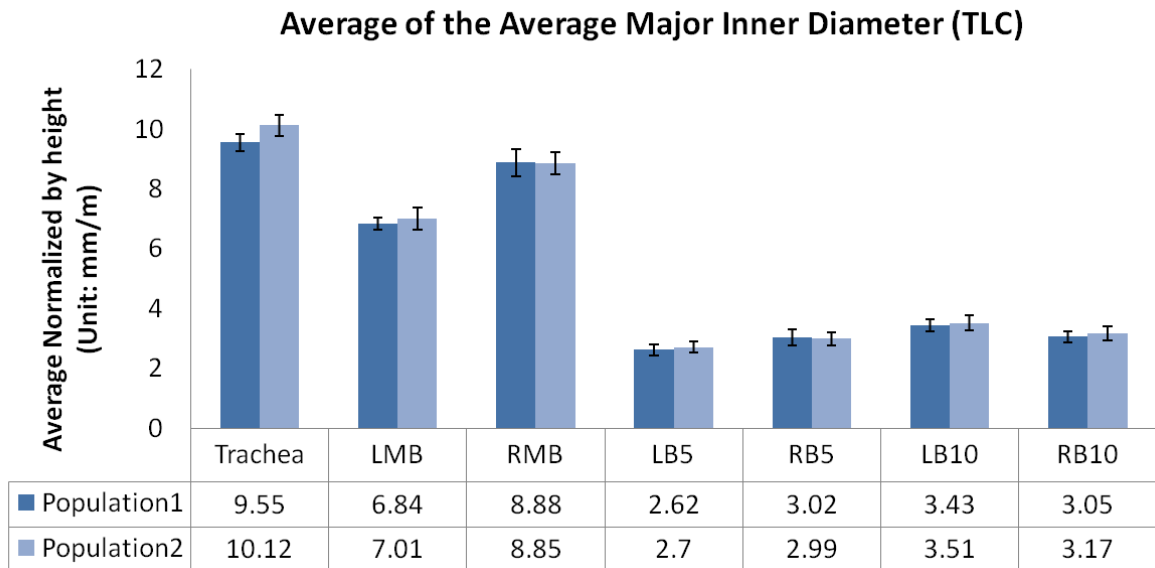|	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
Population1	106.78	49.43	80.27	6.94	9.1	11.88	9.73
Population2	117.45	53.87	85.35	7.37	8.91	12.58	10.79
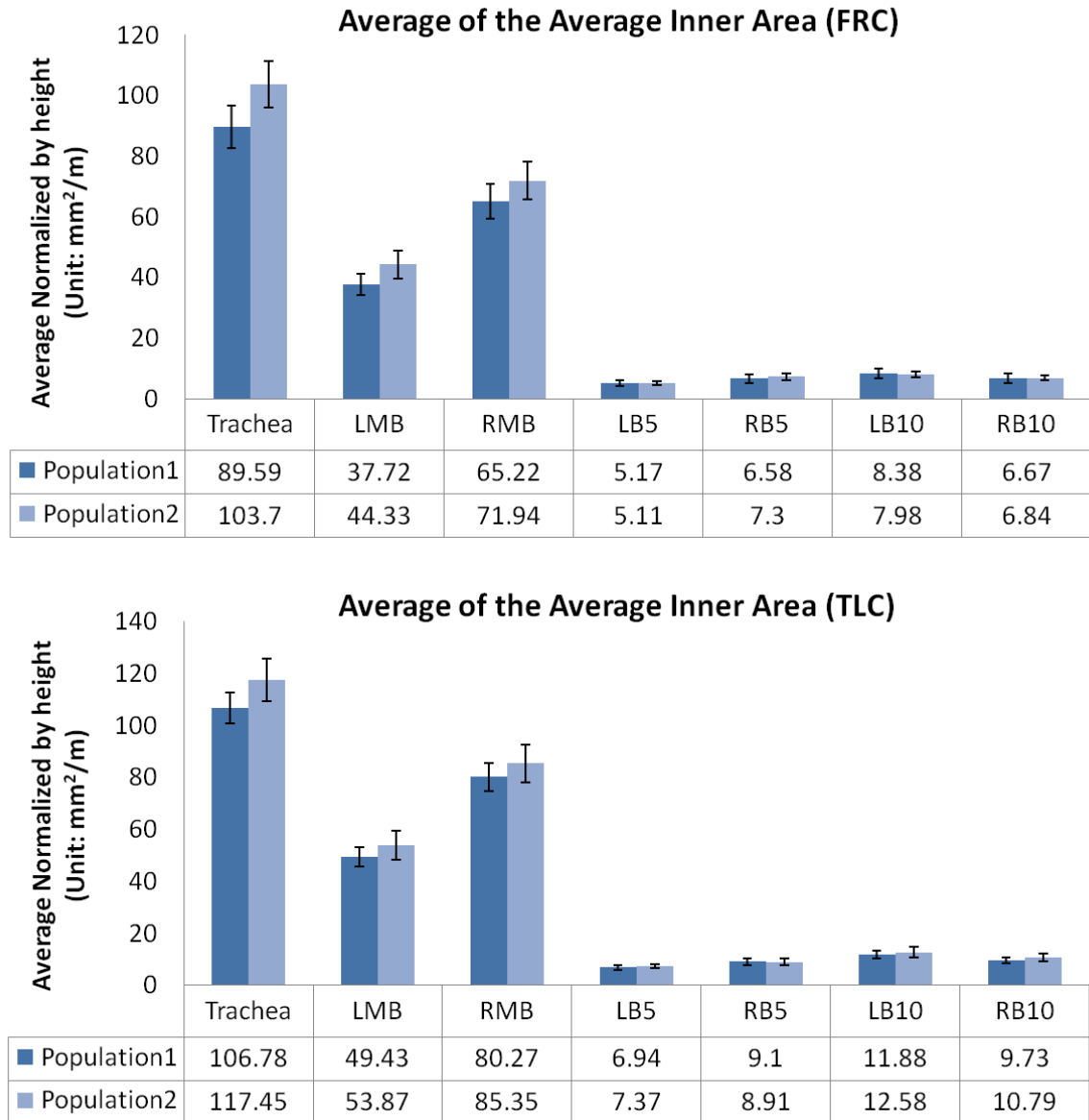
Figure C.3: Study4: Population averages of the average inner area measurement normalized by height at FRC and TLC

APPENDIX D
FIGURES FOR STUDY 5

Table D.1: Study5: t-statistics, p-values and population counts for Figure D.1

		Trachea	LMB	RMB	LB5	RB5	LB10	RB10
FRC								
t-stat/p-value	Population 1	3.47	2.59	2.26	2.8	0.94	0.47	0.78
	Population 2	0	0.02	0.03	0.01	N/A	N/A	N/A
Count	Population 1	22	22	22	18	21	22	21
	Population 2	16	16	16	15	15	15	15
TLC								
t-stat/p-value	Population 1	2.81	2.07	0.31	3.41	1.16	1.06	1.34
	Population 2	0	0.05	N/A	0	N/A	N/A	0.2
Count	Population 1	22	22	22	22	22	22	22
	Population 2	16	16	16	16	14	16	15

Table D.2: Study5: t-statistics, p-values and population counts for Figure D.2

		Trachea	LMB	RMB	LB5	RB5	LB10	RB10
FRC								
t-stat/p-value	Population 1	2.97	4.14	1.68	2.13	0.75	0.38	0.67
	Population 2	0	0	0.11	0.04	N/A	N/A	N/A
Count	Population 1	22	22	22	18	21	22	21
	Population 2	16	16	16	15	15	15	15
TLC								
t-stat/p-value	Population 1	1.9	3.54	1.75	2.56	1.04	0.95	0.17
	Population 2	0.07	0	0.09	0.02	N/A	N/A	N/A
Count	Population 1	22	22	22	22	22	22	22
	Population 2	16	16	16	16	14	16	15

Table D.3: Study5: t-statistics, p-values and population counts for Figure D.3

		Trachea	LMB	RMB	LB5	RB5	LB10	RB10
FRC								
t-stat/p-value	Population 1	3.45	4.42	2.15	2.41	1.84	0.72	0.04
	Population 2	0	0	0.05	0.03	0.08	N/A	N/A
Count	Population 1	22	22	22	18	21	22	21
	Population 2	16	16	16	15	15	15	15
TLC								
t-stat/p-value	Population 1	1.75	2.23	0.85	1.46	0.67	0.01	1.83
	Population 2	0.1	0.04	N/A	0.17	N/A	N/A	0.09
Count	Population 1	22	22	22	22	22	22	22
	Population 2	16	16	16	16	14	16	15

Average of the Average Major Inner Diameter (FRC)

	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
Population1	9.98	6.61	8.62	2.35	2.69	2.94	3.09
Population2	10.97	7.2	9.28	2.89	2.85	3.06	2.81

Average of the Average Major Inner Diameter (TLC)

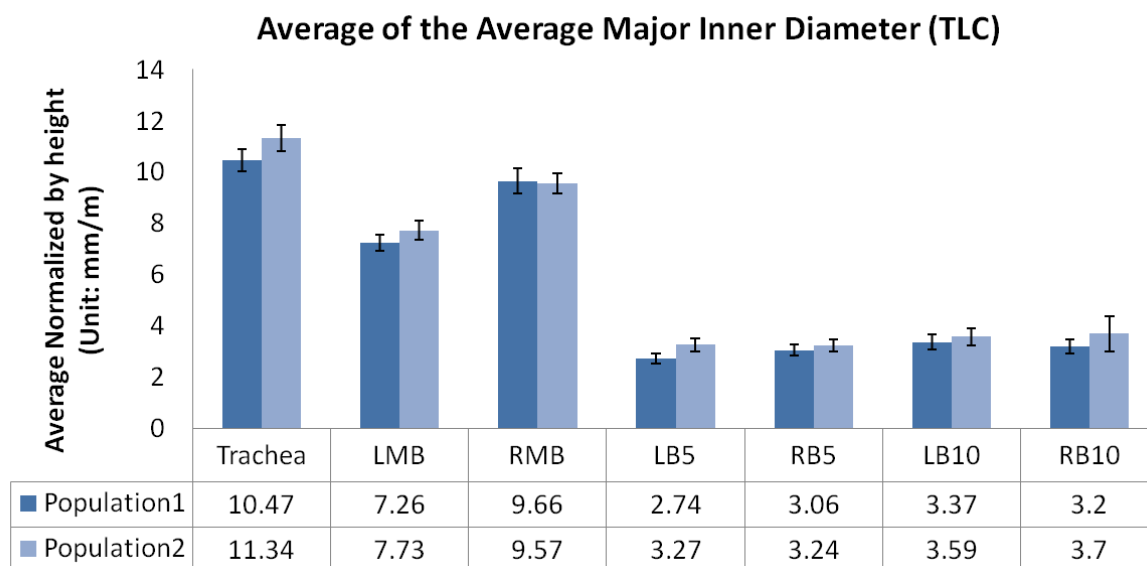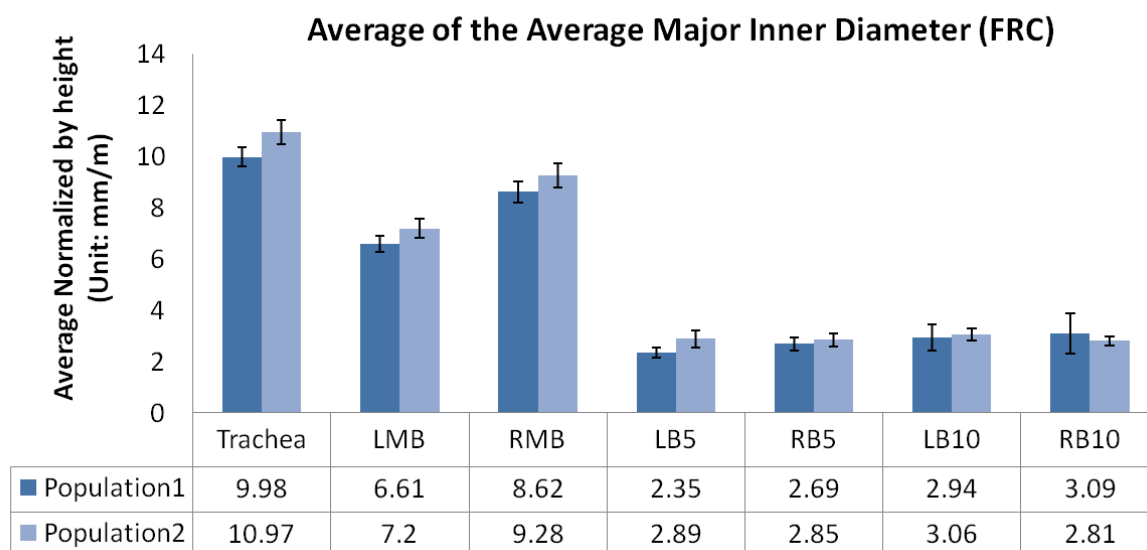|	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
Population1	10.47	7.26	9.66	2.74	3.06	3.37	3.2
Population2	11.34	7.73	9.57	3.27	3.24	3.59	3.7

Figure D.1: Study5: Population averages of the average major inner diameter measurement normalized by height at FRC and TLC

Average of the Average Minor Inner Diameter (FRC)

	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
■ Population1	8.04	4.9	6.44	1.82	2.09	2.18	2.16
■ Population2	8.86	5.7	6.86	2.17	2.2	2.25	2.25

Average Normalized by height (Unit: mm/m)

Average of the Average Minor Inner Diameter (TLC)

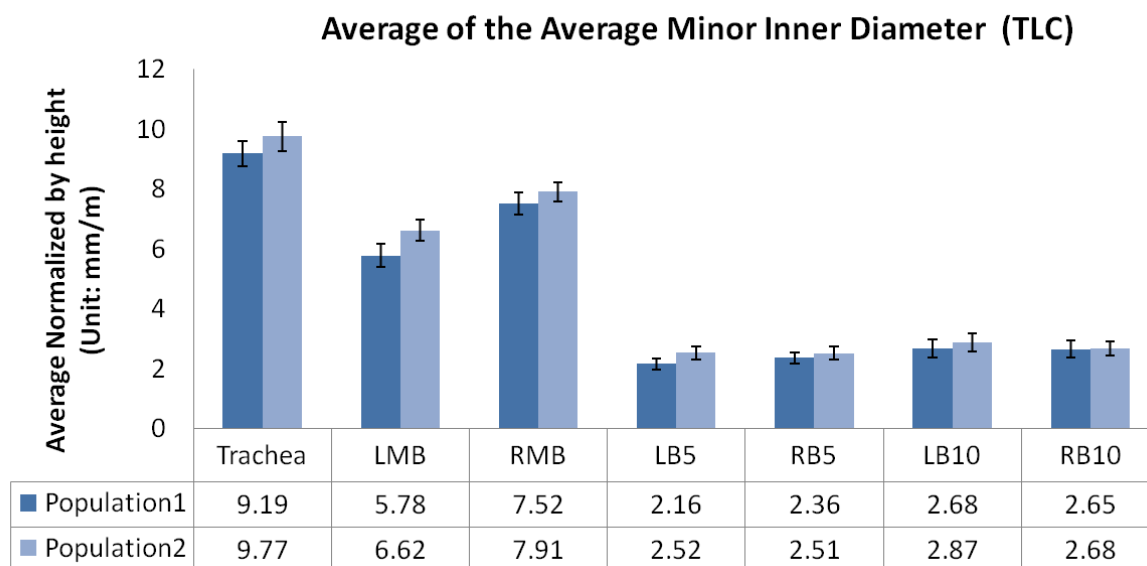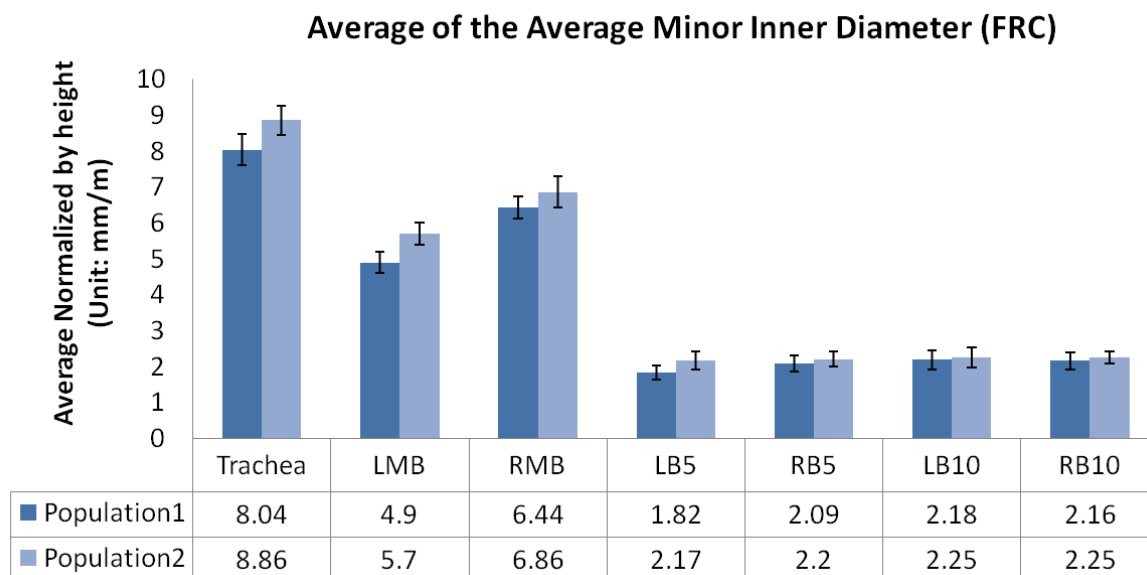|	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
■ Population1	9.19	5.78	7.52	2.16	2.36	2.68	2.65
■ Population2	9.77	6.62	7.91	2.52	2.51	2.87	2.68

Average Normalized by height (Unit: mm/m)

Figure D.2: Study5: Population averages of the average minor inner diameter measurement normalized by height at FRC and TLC

Average of the Average Inner Area (FRC)

	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
Population1	113.89	46.16	80.88	6.28	8.19	9.62	10.06
Population2	136.61	59.83	91.74	9.31	9.09	10.02	9.16

Average of the Average Inner Area (TLC)

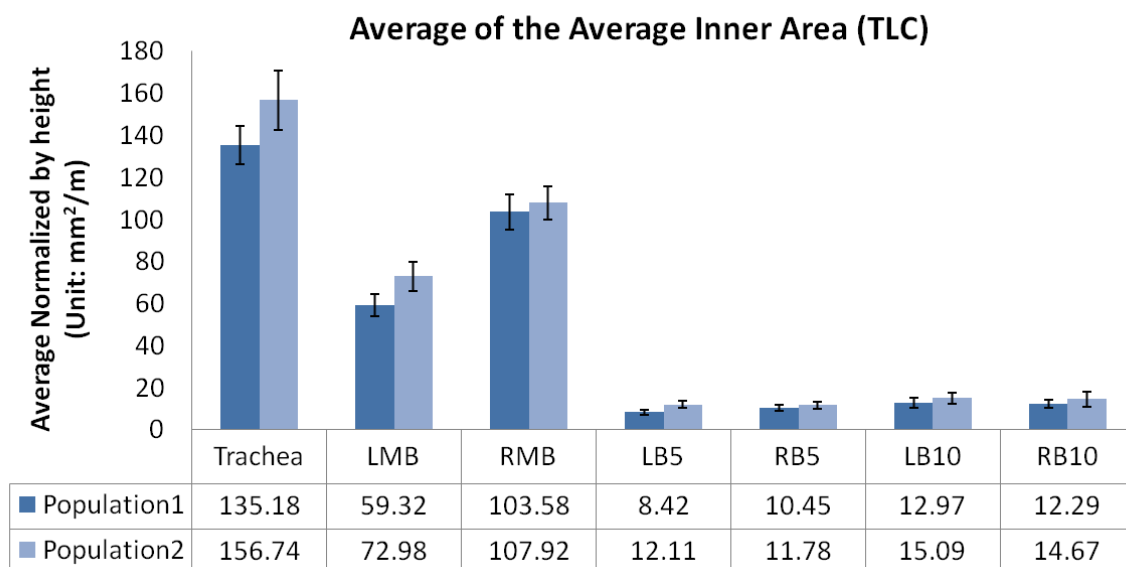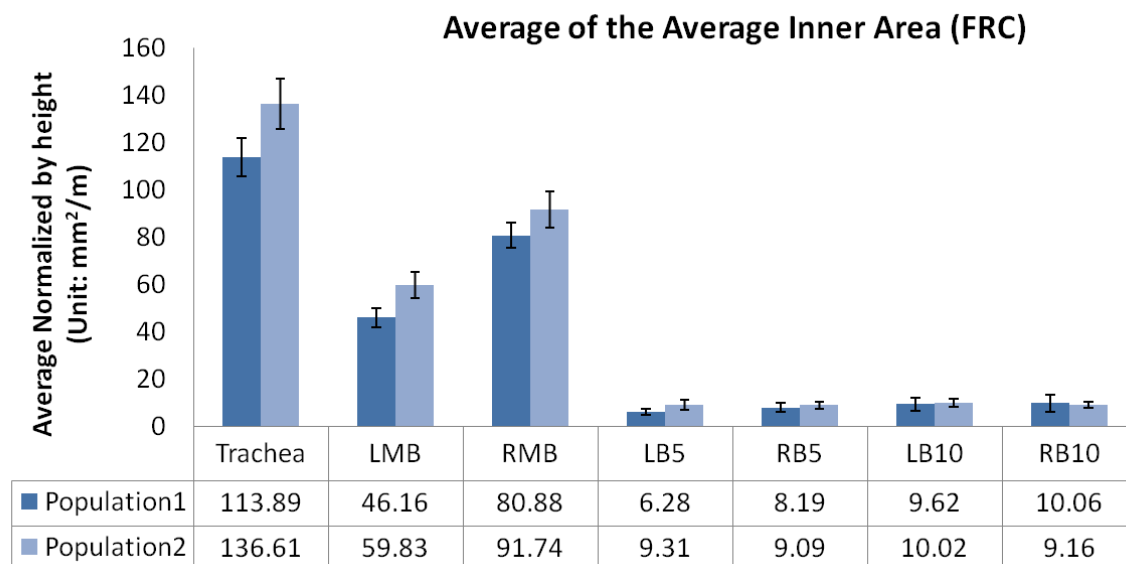|	Trachea	LMB	RMB	LB5	RB5	LB10	RB10
Population1	135.18	59.32	103.58	8.42	10.45	12.97	12.29
Population2	156.74	72.98	107.92	12.11	11.78	15.09	14.67

Figure D.3: Study5: Population averages of the average inner area measurement normalized by height at FRC and TLC

REFERENCES

[1] *Pulmonary Workstation 2 FDA Version User Manual.*

[2] *VIDA Diagnostics 510(K) Submission Clearance number K083227 Attachment J.*

[3] *VIDA Diagnostics 510(K) Submission Clearance number K083227 results.*

[4] Micheau A. and Hoa D. e-anatomy, August 2010.

[5] P.G. Camp, H.O. Coxson, R.D. Levy, S.G. Pillai, W. Anderson, J. Vestbo, S.M. Kennedy, E.K. Silverman, D.A. Lomas, and P.D. Paré. Sex differences in emphysema and airway disease in smokers. *Chest*, 136(6):1480, 2009.

[6] M.A. Carey, J.W. Card, J.W. Voltz, S.J. Arbes Jr, D.R. Germolec, K.S. Korach, and D.C. Zeldin. It's all about sex: gender, lung development and lung disease. *Trends in Endocrinology & Metabolism*, 18(8):308–313, 2007.

[7] John J Ellis. E-anatomy. *Journal of the Medical Library Association*, 98(3):267, July 2010.

[8] T. Greitz, C. Bohm, S. Holte, and L. Eriksson. A computerized brain atlas: Construction, anatomical content, and some applications. *Journal of Computer Assisted Tomography*, 15(1):26–38, Janurary/February 1991.

[9] E.A. Hoffman, A.V. Clough, G.E. Christensen, C. Lin, G. McLennan, J.M. Reinhardt, B.A. Simon, M. Sonka, M.H. Tawhai, EJ van Beek, et al. The comprehensive imaging-based analysis of the lung. *Academic radiology*, 11:1370–1380, 2004.

[10] Shiying Hu, Eric A. Hoffman, and Joseph M. Reinhardt. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Trans. on Medical Imaging*, 20:490–498, June 2001.

[11] Eric A. Hoffman Joseph M. Reinhardt. Accurate measurement of intrathoracic airways. *IEEE TMI*, 6:820–827, 1997.

[12] SLA Lee, AZ Kouzani, and E.J. Hu. From lung images to lung models: A review. In *IEEE International Joint Conference on Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence)*, pages 2377–2383, 2008.

[13] Baojun Li, Gary E. Christensen, Geoffrey McLennan, Eric A. Hoffman, and Joseph M. Reinhardt. Establishing a normative atlas of the human lung: Inter-subject warping and registration of volumetric CT. *Academic Radiology*, 10(3):255–265, March 2003.

[14] T.R. Martin, R.G. Castile, J.J. Fredberg, ME Wohl, and J. Mead. Airway size is related to sex but not lung size in normal adults. *Journal of Applied Physiology*, 63(5):2042, 1987.

[15] J.C. Mazziotta, A.W. Toga, A. Evans, P. Fox, and J. Lancaster. A Probabilistic Atlas of the Human Brain: Theory and Rationale for Its Development. *Neuroimage*, 2:89–101, 1995.

[16] J.C. Mazziotta, A.W. Toga, A. Evans, P. Fox, J. Lancaster, and et al. A four-dimensional probabilistic atlas of the human brain. *Journal of the Ameriacn Medical Information Association*, 8(5):401–430, 2001.

[17] Merriam-Webster. Atlas. Hypertext Document, August 2010.

[18] Kalman Palagyi, Juerg Tschirren, Eric A. Hoffman, and Milan Sonka. Quantitative analysis of pulmonary airway tree structures. *Computers in Biology and Medicine*, 36:974–996, October 2006.

[19] A. Silberschatz, H.F. Korth, and S. Sudarshan. *Database system concepts.* McGraw-Hill New York, 2006.

[20] Juerg Tschirren, Eric A. Hoffman, Geoffrey McLennan, and Milan Sonka. Intrathoracic airway trees: Segmentation and airway morphology analysis from low-dose ct scans. *IEEE Transactions on Medical Imaging*, 24(12):1529–1539, 2005.

[21] Juerg Tschirren, Geoffrey McLennan, Kalman Palagyi, Eric A. Hoffman, and Milan Sonka. Matching and anatomical labeling of human airway tree. *IEEE Transactions on Medical Imaging*, 24(12):1540–1547, 2005.

[22] Soumik Ukil and Joseph M. Reinhardt. Anatomy-guided lung lobe segmentation in x-ray ct images. *IEEE TMI*, 28(2):202–214, 2009.

[23] John B West. *Respiratory Physiology: The Essentials.* Lippincott Williams and Wilkins, Maryland and Philadelphia, 8th edition, 2008.

[24] Wikipedia. Hispanic and Latino Americans. Hypertext Document, November 2010.

[25] Wikipedia. Microsoft Access. Hypertext Document, October 2010.

[26] Li Zhang, Eric A. Hoffman, and Joseph M. Reinhardt. Atlas-driven lung lobe segmentation in volumetric x-ray ct images. *IEEE Transactions on Medical Imaging*, 25(1):1–16, 2006.

[27] T. Zrimec and S. Busayarat. 3d modelling and visualization of the human lung. In *2nd International Symposium on 3D Data Processing Visualization and Transmission.* IEEE Computer Society, 2004.