THE UNIVERSITY OF
WARWICK

University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

A Thesis Submitted for the Degree of PhD at the University of Warwick

http://go.warwick.ac.uk/wrap/74046

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

Quantification of Chronic Fibrosing Interstitial Pneumonia on Computed Tomography

by

Emma Jane Helm

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Medicine

University of Warwick, Warwick Medical School

May 2015

TABLE OF CONTENTS

INDEX OF TABLES AND FIGURES

Figure 5-11 Comparison of radiologists scores for total fibrosis, honeycombing, and reticulation for 40 axial CT slices

ACKNOWLEDGEMENTS

I would like to thank all my collaborators on the QUIC study including Dr Rhian Hughes, Dr Felix Woodhead, Dr David Parr, Dr Abhir Bhalerao, Dr Elke Thonnes, Eddie Ngandwe and Joanna Shakespeare.

I would also like to thank Dr Rachel Benamore of Oxford University Hospitals NHS Trust for visual scoring of CT scans.

I would especially like to thank my first supervisor, Professor Charles Hutchinson, for persuading me to write this thesis and for his persistent faith that, despite several signs to the contrary, I would produce a finished document.

Finally I would like to thank my husband, Dr Ed Nash for putting up with my long hours on the computer and for entertaining our children so that I was able to complete my thesis.

DECLARATION

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Medicine. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

Development of the automatic quantification software was performed by Dr Abhir Bhalerao and Dr Elke Thönnes of the Departments of Computer Science and Statistics, Warwick University.

Running of the automatic quantification software was performed by Dr Abhir Bhalerao, Department of Computer Science, Warwick University.

Selected figures were provided by Dr Abhir Bhalerao as detailed in the text.

ABSTRACT

The chronic fibrosing interstitial pneumonias (CFIPs) are diseases which cause progressive and often fatal progressive scarring of the lungs. The recent discovery of the first effective pharmacological therapies for this condition have increased interest in the monitoring of this disease.

Due to the complex appearance of the CFIPs on computed tomography, visual quantification of disease severity and extent is limited.

The purpose of this thesis was to develop and test a computer algorithm for the automated quantification of pulmonary fibrosis on CT using textural measures known as Minkowski functionals.

A computer algorithm was successfully developed and this thesis presents initial results of testing the algorithm on a series of normal scans and on 24 prospectively recruited patients who also underwent a series of other tests including pulmonary function tests and a patient reported symptom questionnaire. The computer output was also compared with the visual assessment of two radiologists.

Significant correlations were found between computer calculated lung volume and total lung capacity as measured on pulmonary function tests. We also found a significant correlation between computer calculated fibrosis volume and both gas transfer and forced vital capacity. The radiologists' visual assessment of fibrosis and the computer estimated fibrosis volume were highly correlated.

The novel computer algorithm represents a promising method for quantifying pulmonary fibrosis on CT with potential roles in monitoring disease progression and effects of therapeutic interventions.

ABBREVIATIONS

AMFM:	Adaptive multiple features method
COPD:	Chronic obstructive airways disease
CFIP:	Chronic fibrosing interstitial pneumonia
CFS:	Computer fibrosis score
CT:	Computed tomography
DIP:	Desquamating interstitial pneumonia
GGO:	Ground glass opacification
HRQL:	Health related quality of life
HU:	Hounsfield unit
ICC:	Intraclass correlation
IIP:	Idiopathic interstitial pneumonia
ILD:	Interstitial lung disease
MF:	Minkowski functional
MLD:	Mean lung density
NSIP:	Non-specific interstitial pneumonia
QCT:	Quantitative CT
RFS:	Radiologist fibrosis score
ROI:	Region of interest
STGRQ:	St George's Respiratory Questionnaire
UIP:	Usual interstitial pneumonia
VOI:	Volume of interest

CHAPTER 1: BACKGROUND

1.1 INTRODUCTION

The idiopathic interstitial pneumonias (IIPs) are a diverse group of diseases that are united in their ability to cause progressive fibrosis or scarring in the lungs, often leading to severe morbidity and ultimately death. The most important of the idiopathic pneumonias in terms of prevalence, morbidity and mortality is the clinical syndrome of idiopathic pulmonary fibrosis (IPF). Typically striking in the 6th to 7th decades, the disease causes dry cough and progressive breathlessness and has a median survival of only 3 years (Gribbin et al., 2006). Unfortunately there are many unknown quantities in this disease including the aetiology, the reason that some patients progress much faster than others and the lack of precision in measuring disease severity. This thesis will focus on one particular aspect of this disease, namely the need to precisely measure disease severity on computed tomography (CT). We will describe the development and testing of a novel computer algorithm designed to quantify disease severity on CT and compare it with the current clinical practice of a radiologist reading the scans. We will also compare the computer quantification with other measures of severity including pulmonary function tests (PFTs) and patient reported symptoms.

This first section of this introductory chapter will describe the current understanding of disease classification, clinical phenotypes, epidemiology, theories of aetiology, treatment and prognosis of the IIPs, focusing mainly on the clinico-pathological entity of idiopathic pulmonary fibrosis/usual interstitial pneumonia (IPF/UIP). The next section will examine the role of radiology in the assessment of IPF and the need to progress beyond a visual assessment of disease extent. We will also discuss non-

radiological methods of assessing disease severity and describe some of the pros and cons of the different methods.

1.2 THE IDIOPATHIC INTERSTITIAL PNEUMONIAS

1.2.1 Definition of the idiopathic interstitial pneumonias

The idiopathic interstitial pneumonias (IIPs) are a complex and heterogeneous collection of pathological conditions that cause significant morbidity and mortality. In 2002, the ATS/ERS international multidisciplinary consensus classification of idiopathic interstitial pneumonias was published, redefining the histological, clinical and radiological diagnosis of the IIPs (European and Society, 2002). In 2013 an update to the guidelines was issued which the authors specified should be treated as a supplement to the 2002 guidelines rather than as a 'stand-alone' document. This update introduced a number of alterations to the 2002 guidelines including the removal of the term 'cryptogenic fibrosing alveolitis', acceptance of idiopathic non-specific interstitial pneumonia (NSIP) as a distinct clinico-pathological entity and the use of the term 'chronic fibrosing interstitial pneumonia' (CFIP) to describe both idiopathic pulmonary fibrosis and idiopathic nonspecific interstitial pneumonia.

TABLE 1. REVISED AMERICAN THORACIC SOCIETY/EUROPEAN RESPIRATORY SOCIETY CLASSIFICATION OF IDIOPATHIC INTERSTITIAL PNEUMONIAS: MULTIDISCIPLINARY DIAGNOSES

Major idiopathic interstitial pneumonias
Idiopathic pulmonary fibrosis
Idiopathic nonspecific interstitial pneumonia
Respiratory bronchiolitis–interstitial lung disease
Desquamative interstitial pneumonia
Cryptogenic organizing pneumonia
Acute interstitial pneumonia
Rare idiopathic interstitial pneumonias
Idiopathic lymphoid interstitial pneumonia
Idiopathic pleuroparenchymal fibroelastosis
Unclassifiable idiopathic interstitial pneumonias*

* Causes of unclassifiable idiopathic interstitial pneumonia include (*1*) inadequate clinical, radiologic, or pathologic data and (*2*) major discordance between clinical, radiologic, and pathologic findings that may occur in the following situations: (*a*) previous therapy resulting in substantial alteration of radiologic or histologic findings (e.g., biopsy of desquamative interstitial pneumonia after steroid therapy, which shows only residual nonspecific interstitial pneumonia [153]); (*b*) new entity, or unusual variant of recognized entity, not adequately characterized by the current American Thoracic Society/European Respiratory Society classification (e.g., variant of organizing pneumonia with supervening fibrosis) (79); and (*c*) multiple high-resolution computed tomography and/or pathologic patterns that may be encountered in patients with idiopathic interstitial pneumonia.

Table 1-1 The 2013 updated ATS/ERS classification of idiopathic interstitial pneumonias. Reprinted with permission of the American Thoracic Society. Copyright © 2014 American Thoracic Society

The most common IIP is the clinical entity of idiopathic pulmonary fibrosis (IPF), which corresponds to the histological pattern described as Usual Interstitial Pneumonia (UIP). The distinction between subtypes of IIP is important because of the differences in prognosis associated with different histological patterns amongst patient cohorts (Mapel et al., 1998, Hubbard et al., 1998, Flaherty et al., 2002), with UIP having a significantly worse prognosis than most of the other IIPS. For the purpose of this thesis, we will not be discussing conditions such as desquamative interstitial pneumonia and hypersensitivity pneumonitis but will be confining our studies to patients with idiopathic disease.

1.2.2 Incidence/Prevalence

Estimating the incidence and prevalence of IPF is hampered by the fact that there is no readily available and specific test for the disease. However, available data from death certification and GP diagnostic databases suggest an incidence of between 4.6 (Gribbin et al., 2006) and 7.4 (Navaratnam et al., 2011) per 100,000 person years. This equates to approximately 5000 new diagnoses per year and is higher than the incidence of several cancers including thyroid cancer and lymphoma (Cancer Research UK, UK Cancer Incidence (2010) by Country Summary, April 2013).

It has been suggested that the incidence of IPF has increased over the last few decades (Navaratnam et al., 2011) but no specific reason for this has been identified. It may partly be explained by the increased use of CT scanning which is able to pick up early-stage disease. In addition, the introduction of a non-invasive test (CT) when the previous definitive test has been invasive and restricted to patients fit enough for surgery (open lung biopsy) is likely to lead to an increase in diagnosis. The endorsement of CT as a diagnostic test for IPF by professional societies is also likely to be a factor (Raghu et al., 2011).

1.2.3 Mortality

Studies of mortality in IPF have shown a median survival of between 3 and 4 years (Gribbin et al., 2006, Mapel et al., 1998, Hubbard et al., 1998). It should be noted that there is a higher median survival in incident cases than prevalent cases. This is due to survival bias, meaning that incident cases are representative of the whole spectrum of disease severity, whereas patients with aggressive disease who die quickly will be under-represented in the prevalence group. Because of this, it has been suggested that studies into prognostic variables in IPF are restricted to incident cases, although this is not always practical (Hubbard et al., 1998).

Unsurprisingly, mortality from IPF is increased in older age groups (Navaratnam et al., 2011). This may be due to co-morbidities in older patients, but may also reflect the fact that younger patients diagnosed with IPF are more likely to be affected by less aggressive histological forms of disease such as NSIP or connective tissue disease-associated pulmonary fibrosis.

1.2.4 Age and sex distribution

A British Thoracic Society study of 588 newly presenting patients found a median age at diagnosis of 67 years and a male:female ratio of 1.7:1 (Johnston et al., 1997). Other studies have suggested an earlier age at diagnosis but have been from specialist centres where younger patients with more aggressive disease are likely to be over-represented. Some authors have suggested that IPF is, in fact, a degenerative disease associated with ageing. This theory is supported by the finding of shortened telomeres in patients with familial pulmonary fibrosis and non-familial cases. Telomeres are non-coding areas of DNA nucleotide repeats found at the end of chromosomes which protect the chromosome during cell division. Each time a cell divides, chromosome replication occurs and a small amount of DNA is lost from the end of each chromosome. Loss of a small part of the non-coding telomere DNA from the end of the chromosome means that important coding DNA is not lost (Cowell, 2001). Telomere shortening is also thought to occur as a result of oxidative stress (Von Zglinicki, 2002). Telomeres were discovered by Elizabeth Blackburn in the 1970s (Blackburn and Gall, 1978) but it was only in the 2000s that the link with ageing was established and roles in the pathogenesis of multiple cancers and age-related diseases were proposed (Blasco, 2005, Brouilette et al., 2007, Torella et al., 2004, Ito and Barnes, 2009). A crucial factor in the maintenance of telomeres is the enzyme telomerase which is a ribonucleoprotein enzyme that catalyses the addition of

hexameric (6-base length) nucleotide repeats to the ends of chromosomes. This enzyme consists of 6 components - 3 pairs of each of the following molecules: telomerase reverse transcriptase (TERT), telomerase RNA (TERC) and dyskerin (DKC1). In 2007 two groups published evidence of mutations in the genes encoding telomerase in cases of familial pulmonary fibrosis (Tsakiri et al., 2007, Armanios et al., 2007). Tsakiri et al also found a mutation in one patient with no family history (Tsakiri et al., 2007). Following this, Cronkhite el al (Cronkhite et al., 2008) studied a cohort of patients with familial IPF and a cohort of patients with sporadic IPF but who did not have specific mutations of TERC or TERT genes. They found that even patients without a specific genetic mutation had shorter telomeres than a control group without IPF. They found that 14 of 59 patients (24%) with familial IPF had telomeres below the 10th percentile prediction line and 17 of 73 patients (23%) with non-familial IPF had telomeres in the bottom 10th centile. This was statistically significant in both familial ($p = 8.0 \times 10^{-6}$) and non-familial ($p = 2.6 \times 10^{-6}$) cases. The fact that the prevalence of IPF increases significantly with age also lends weight to the fact that telomere shortening may be an important co-factor in the aetiology of the disease. Copley et al studied CT scans of two groups of asymptomatic patients who were undergoing a CT of the abdomen or a CT of the head and did not have any history of respiratory disease. The first group consisted of 40 patients aged over 75 years (mean age 80.6 years) and the second group comprised 16 patients aged less than 55 (mean age 39.4 years). An extensive list of exclusion factors was designed to ensure that patients were unlikely to have a latent undiagnosed respiratory disease. This prospective study consented patients to undergo thin-slice prone inspiratory high-resolution CT. Two radiologists scored the CT scans for the presence and extent of reticular abnormality, cysts, bronchial dilatation, bronchial thickening, ground glass opacity, interlobular septal thickening and centrilobular emphysema. They found that

60% of the older group had a limited, predominantly subpleural, basal reticular pattern which was not seen in any of the younger group. Cysts were also seen in 25% of the older group but none of the younger group. The authors noted that the reticular pattern was not associated with traction bronchial dilatation, one of the main features of IIP. They concluded that a limited subpleural basal reticular finding may be a normal finding in older age groups and should not necessarily be interpreted as interstitial lung disease. Limitations of this study included the fact that the majority of patients were city dwellers and therefore there may be environmental factors which limit extrapolation to a non-urban population. Another limitation was that some subjects were ex-smokers. Perhaps the most significant limitation is the lack of histological confirmation, since biopsy could not be justified in this population.

1.2.5 Risk factors for idiopathic pulmonary fibrosis

No definitive cause for idiopathic pulmonary fibrosis has been identified but a number of risk factors have been described. The most powerful association is with smoking (odds ratio 2.9) (Iwai et al., 1994) but weaker associations have also been shown with exposure to metallic dusts and wood dust. Farming, raising birds, working with stone and exposure to animal dust have also been associated with IPF (Baumgartner et al., 1997). Another suggested aetiology is chronic viral infection with the largest amount of evidence for EBV and hepatitis C. Both DNA and protein from EBV have been detected in greater numbers of lung biopsies from patients with IPF compared to the general population (Egan et al., 1995, Stewart et al., 1999). However, definitive conclusions are hampered by the high prevalence of EBV in the normal population and by the fact that many patients with IPF have received immunosuppression at some point, which is likely to increase the incidence of EBV.

Evidence for the role of hepatitis C is also mixed. Ueda et al (Ueda et al., 1992) found that 19 of a group of 66 patients with IPF (28.8%) tested positive to hepatitis C compared with 3.7% of a control group whilst Meliconi et al (Meliconi et al., 1996) found a 13.3% prevalence of hepatitis C in 60 Italian patients with IPF compared with a 0.3% prevalence in a large control group of blood donors. However, they did not find a significant difference in prevalence amongst patients with IPF compared with a group with other mixed lung diseases (6.1%). Other implicated viruses are herpes viruses 7 and 8 and cytomegalovirus (Yonemaru et al., 1997). Several other medical conditions have been associated with IPF, including gastro-oesophageal reflux and diabetes mellitus (Tobin et al., 1998, Gribbin et al., 2009) but a causative relationship has not been proven.

1.2.6 Histology

The histological hallmark of IPF is a usual interstitial pneumonia (UIP) pattern of fibrosis. UIP is characterised by subpleural and paraseptal fibrosis and honeycombing, interspersed with areas of less severely affected or normal lung (spatial heterogeneity). There is a lack of transition zone between normal and affected lungs, in other words an abrupt change from normal to abnormal lung. Honeycombing consists of cystic, fibrotic airspaces with a bronchiolar epithelial lining. The honeycomb cysts may contain inflammatory cells and mucin. Within the abnormal lung are fibroblastic foci which reflect active fibrosis and suggest temporal heterogeneity (Figure 1-1). There is hyperplasia of type II pneumocytes and there may be a mild lymphocytic infiltrate, but this should not be a prominent feature.

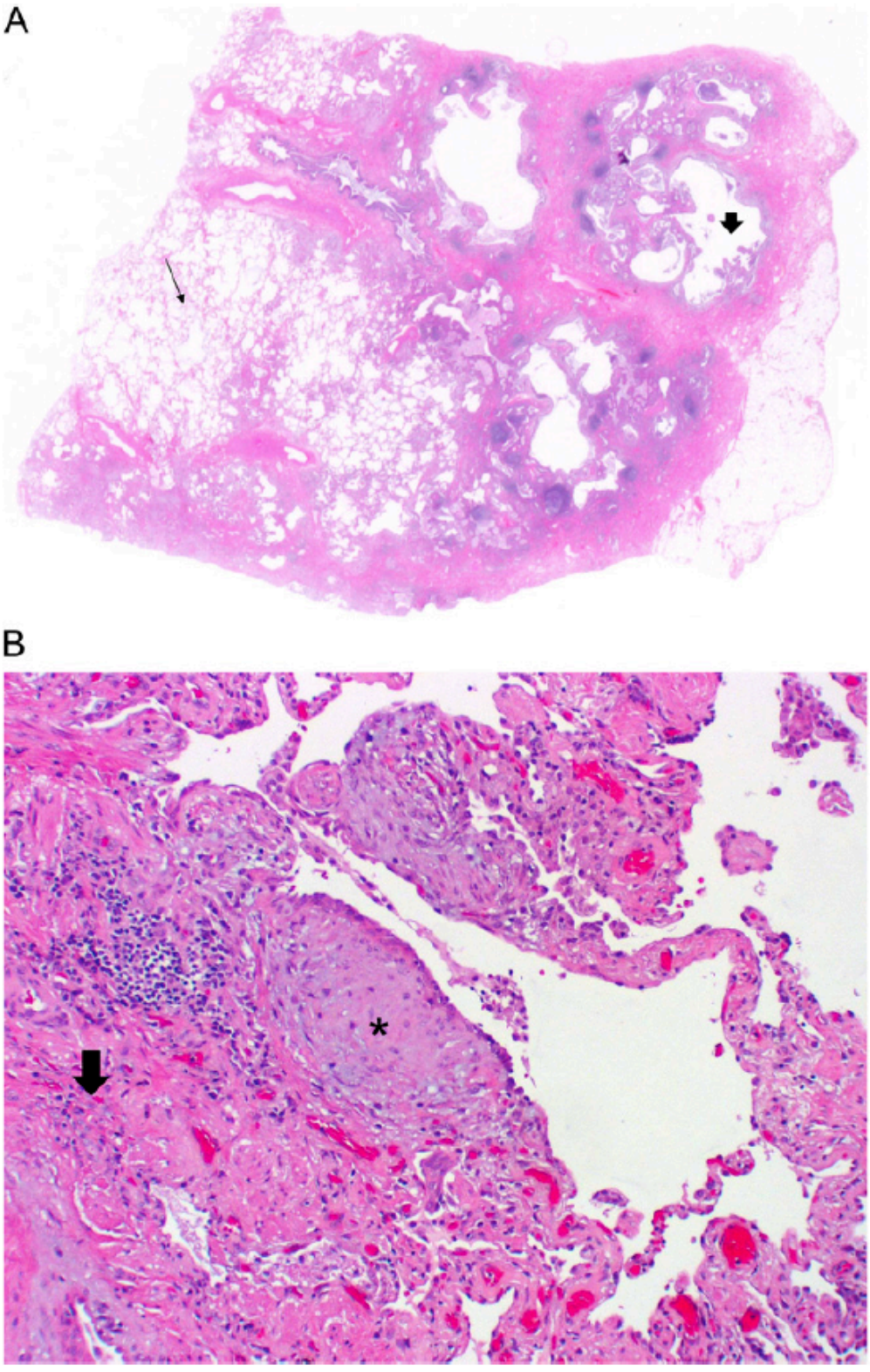

Figure 1-1 Surgical lung biopsy demonstrating UIP pattern. (A) Scanning power microscopy showing patchy process with honeycomb spaces (thick arrow), some preserved lung tissue regions (thin arrow), and fibrosis extending into the lung from the subpleural regions. (B) Adjacent to the regions of more chronic fibrosis (thick arrow) is a fibroblastic focus (asterisk), recognised by its convex shape and composition of oedematous fibroblastic tissue, suggestive of recent lung injury. Reproduced with permission from (Raghu et al., 2011)

1.2.7 Potential mechanisms of disease causation

Two main mechanisms have been proposed for the development and progression of IPF, namely the inflammatory pathway and the epithelial pathway (King Jr et al., 2011). Initial theories centred on the contribution of inflammatory mechanisms and demonstration of an increased population of lymphocytic cells in the broncho-alveolar lavage fluid of patients with IPF. Despite this, anti-inflammatory medication, including steroids, have shown consistently poor results in patients with IPF, leading investigators to question the role of inflammation in the disease. The epithelial pathway focuses on the role of epithelial-dependent activation of fibroblasts, which then leads to fibrosis. This is illustrated in Figure 1-2 below:

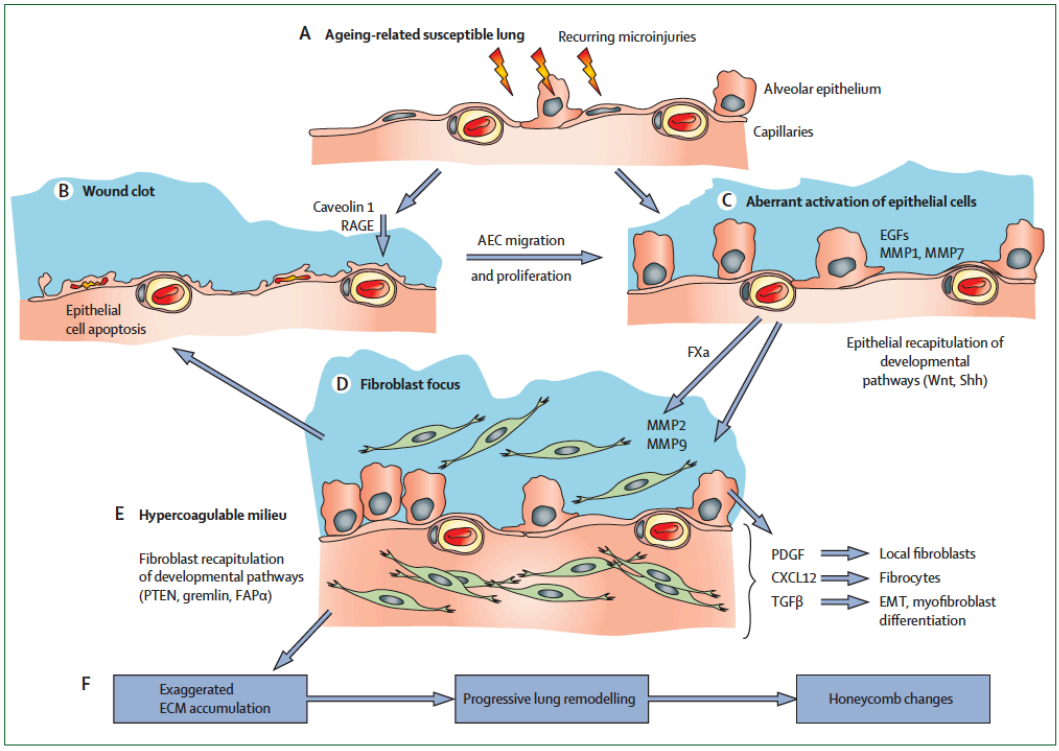

Figure 1-2 Schematic showing the potential causative mechanisms of pulmonary fibrosis in IPF. Reproduced with permission from (King Jr et al., 2011)

1.3 Role of Radiology in the Assessment of IPF

1.3.1 Role of the chest x-ray

The first imaging investigation in most patients with suspected lung disease is a chest x-ray and the majority of patients with IPF will have an abnormal chest radiograph at diagnosis. The classical appearance of IPF on chest x-ray is of peripheral reticulation, which is worse at the lung bases and often causes ill-definition of the heart borders. The advantages of chest x-ray are that it is inexpensive, readily available and has a relatively low radiation burden (typical dose equivalent to 2.4 days background radiation) (Hall, 2002). The disadvantages of chest x-ray are that technical factors such as degree of inspiration and obesity may limit interpretation, that early disease may be missed and that findings in interstitial lung disease are often non-specific. For example, in a study of 118 patients with diffuse interstitial lung disease where radiologists were asked to specify their first choice diagnosis and level of confidence in that diagnosis, chest x-ray was accurate in only 57% of cases compared to an accuracy of 76% for CT. Therefore, before the advent of HRCT, open lung biopsy was often recommended for definitive diagnosis (Mathieson et al., 1989). Nevertheless, chest x-ray remains the first imaging test in most patients, is often performed at routine clinic appointments to look for disease progression and is usually the first line investigation when patients present acutely with an exacerbation of disease.

1.3.2. Role of computed tomography

CT has a number of roles in the assessment of patients with suspected or known interstitial lung disease including: diagnosis, assessment of severity, follow-up and prediction of prognosis.

The typical HRCT findings of interstitial pulmonary fibrosis are of peripheral, subpleural reticulation and honeycombing. The widely accepted Fleischner Society guidelines define reticulation as 'a collection of innumerable small linear opacities that, by summation, produce an appearance resembling a net'. On CT imaging, reticulations are typically peripheral/subpleural and are usually composed of thickened interlobular and intralobular septa (Hansell et al., 2008) (Figure 1-3b). Honeycombing is defined as 'clustered cystic airspaces, typically of comparable diameters on the order of 3-10 mm' (Hansell et al., 2008). It is usually subpleural and characterised by well-defined walls and is considered to be a feature of established fibrosis and to be irreversible (figure 1-3a).

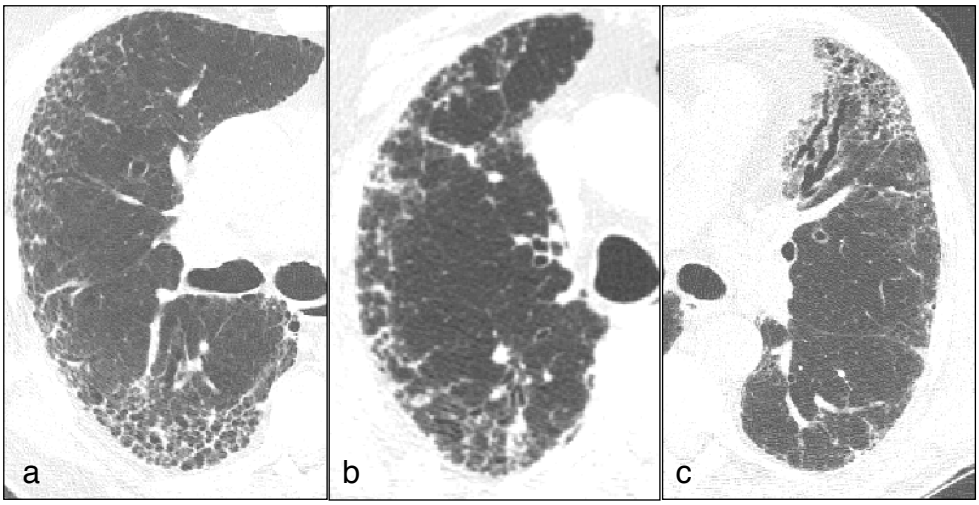

Figure 1-3 Selected axial CT images showing a) honeycombing b) reticulation and c) traction bronchial dilatation

Traction bronchiectasis or traction bronchiolectasis is another feature that is commonly seen on HRCT and is defined as 'irregular bronchial or bronchiolar dilatation caused by surrounding retractile pulmonary fibrosis' (Hansell et al., 2008). Dilated airways normally appear as tubular, air-filled structures that do not taper in the same way as normal airways but may appear as cysts or microcysts at the

periphery of the lung, in which case they may be difficult to distinguish from honeycombing (Figure 1-3c).

Traditionally, open lung biopsy has been considered to be the gold standard for diagnosis of IIP. Unfortunately, many patients with suspected ILD are elderly with frequent co-morbidities and surgical risk factors that may preclude open biopsy. In addition, patients often present with relatively severe disease which many make them unsuitable for surgical biopsy. With this in mind, the sensitivity and specificity of CT for diagnosis has been explored. Several studies have indicated that HRCT has a high sensitivity and specificity for diagnosis of IPF with a positive predictive value of between 90 and 100% (Mathieson et al., 1989, Hunninghake et al., 2001, Raghu et al., 1999, Grenier et al., 1991, Lee et al., 1994). As a result, several guidelines now recommend that if the CT appearances are typical of UIP, biopsy is not required to make a diagnosis (Raghu et al., 2011, Wells, 2013). Criteria for a typical/probable UIP pattern on CT and are described by Raghu et al in the 2011 ATS statement (Raghu et al., 2011) and are shown in the table below (table 1-2). It should be noted that all four features (subpleural basal predominance, reticular abnormality, honeycombing and absence of features inconsistent with UIP) should be present in order to make a confident CT diagnosis of UIP.

UIP Pattern (All Four Features)	Possible UIP Pattern (All Three Features)	Inconsistent with UIP Pattern (Any of the Seven Features)
• Subpleural, basal predominance • Reticular abnormality • Honeycombing with or without traction bronchiectasis • Absence of features listed as inconsistent with UIP pattern (*see* third column)	• Subpleural, basal predominance • Reticular abnormality • Absence of features listed as inconsistent with UIP pattern (*see* third column)	• Upper or mid-lung predominance • Peribronchovascular predominance • Extensive ground glass abnormality (extent > reticular abnormality) • Profuse micronodules (bilateral, predominantly upper lobes) • Discrete cysts (multiple, bilateral, away from areas of honeycombing) • Diffuse mosaic attenuation/air-trapping (bilateral, in three or more lobes) • Consolidation in bronchopulmonary segment(s)/lobe(s)

Table 1-2 High-resolution computed tomography criteria for a diagnosis of UIP. Reproduced with permission from (Raghu et al., 2011)

Fell et al recently studied the predictive power of several clinical, physiological and CT variables for diagnosis of IPF (Fell et al., 2010). They studied 97 patients with biopsy proven IPF and 38 patients with other IIPs and specifically excluded patients with honeycombing on CT. Clinical variables included age, sex, smoking status, pulmonary function tests (FVC and DL_{CO}) and 6-minute walk test (distance walked and whether or not the patient desaturated to <88%). CT scans were analysed by 2 experienced radiologists using a semi-quantitative method which assesses the percentage of lung with ground glass change (alveolar score) and the degree of interstitial changes (reticulation or honeycombing) and was previously described by Kazerooni et al (Kazerooni et al., 1997). Using a multiple logistic regression approach, they found that the two most powerful predictors of IPF on biopsy were age and extent of fibrosis on CT. Even without honeycombing they found that they could confidently predict a biopsy diagnosis of IPF based on age and degree of fibrosis on CT. For example, they found that for patients aged 55 and over with relatively minor fibrosis on CT, there was a positive predictive value of 100% for IPF at surgical biopsy. A grading system was proposed which integrated the patient's age and extent of CT fibrosis in order to predict the positive predictive value for IPF on biopsy. The grading formula is as follows: (0.084 x age + 2.346 x HRCT interstitial score) – 3.31/5.856.

1.3.3 Role of CT in assessing IPF prognosis

Flaherty et al examined the prognostic implication of a radiological diagnosis of UIP compared with a radiological diagnosis of NSIP. The study group comprised 76 patients with a histological diagnosis of UIP and 23 patients with a histological diagnosis of NSIP. Two radiologists read the scans and assigned them to one of three categories: 'definite/probable UIP'; 'definite/probable NSIP' or 'indeterminate'. Analysis of survival curves showed that there were significant differences in the

survival of the three categories, with the poorest survival seen in 'definite/probable UIP' and the best survival in 'definite/probable NSIP' (Flaherty et al., 2003a). The extent of fibrosis on CT as judged by a semi-quantitative scoring system has also been shown to be a powerful predictor of prognosis (Lynch et al., 2005).

1.3.4 Role of CT in quantification of pulmonary fibrosis

A number of approaches have been applied to the quantification of pulmonary fibrosis on HRCT. The most basic, but probably the most widely used in clinical practice, is simply to describe the disease as mild, moderate or severe. There are no specific definitions of these severity categories so the reporting radiologist will typically describe the disease severity in relation to other cases they have seen in the past and one reader's 'mild' may be another reader's 'moderate'. With this in mind, a number of attempts have been made to develop a more quantitative and reproducible approach to the visual estimation of disease severity.

Goh et al, in a study of 215 patients with systemic sclerosis referred to the Royal Brompton Hospital (UK), performed visual scoring at 5 defined anatomical levels on the HRCT scan (Goh et al., 2008). They firstly calculated a global extent score by estimating the amount of lung affected by interstitial lung disease on each slice to the nearest 5% and averaging this score over the 5 slices. They then assessed 'coarseness of reticulation' using a 3 point score as follows: ground glass (grade 1); microcystic honeycombing (air spaces less than or equal to 4 mm in diameter – grade 2); macrocystic honeycombing (airspaces greater than 4 mm in diameter – grade 3). The total 'coarseness score' was calculated by summating the score at each level for a total score of 0 to 15. They also introduced a multiplier whereby if the HRCT was completely normal on one section, they adjusted the score by multiplying by 5/4,

although this approach may lead to over-estimation of disease severity in patients with milder disease.

Edey et al, also working with the Brompton Hospital group, used a similar method to calculate disease severity in idiopathic pulmonary fibrosis (Edey et al., 2011). They calculated the global severity score in the same way as Goh et al although they analysed 6 sections per patient and assessed 5 features. The features analysed were: ground glass opacification, fine reticulation, coarse reticulation, microcystic and macrocystic honeycombing (lumped together), and consolidation. The final scores for each pattern were calculated as a percentage of abnormal lung then summed and a mean overall score for the 6 levels was calculated. They also gave a binary score for emphysema at each level (0 – present, 1 – absent) and produced a total score for each patient (0 to 6). Traction bronchiectasis was assessed in each section for each parenchymal pattern (i.e. fine reticulation, coarse reticulation etc.) according to a 3 point score: 0 = none, 1 = mild, 2 = moderate or 3 = severe/striking. A so-called summed traction bronchiectasis score was calculated for each parenchymal pattern over the 6 sections. A discrepancy between the two radiologist observers was defined as a > 15% difference in global disease scores, more than 1 grade difference in traction bronchiectasis scores and disagreement of whether or not bronchiectasis or emphysema were present. These were said to be resolved by consensus evaluation although details of the process to obtain consensus were not given in the paper.

1.4 COMPUTERISED METHODS FOR QUANTIFICATION OF PULMONARY FIBROSIS ON CT

1.4.1 The need for computerised methods

There are several disadvantages to the visual assessment of pulmonary fibrosis including lack of precision, inter/intra-observer variation and the tedious and time-consuming nature of the task. The computerised analysis of pulmonary fibrosis offers a potential solution to these problems. The next section of this introduction details the main steps of this process and some of the different approaches to lung analysis.

1.4.2 Computerised segmentation of the lungs

A prerequisite to quantitative analysis of the lung parenchyma is to separate the lung from the surrounding structures of the chest wall, since subsequent analysis needs to be applied to lung tissue only. A number of approaches to this have been developed. The most simplistic is termed 'thresholding'. In this technique, a density (Hounsfield Unit - HU) threshold is identified which is chosen to distinguish between lung tissue and other tissues of the chest wall and mediastinum. Since lung typically has Hounsfield Units values in the range -1000 HU to -500 HU and soft tissues typically have values greater than -50 HU, a threshold of approximately -200 HU effectively separates normal lung from adjacent soft tissue. This method also effectively segments emphysematous lungs from the chest wall since this disease typically lowers lung density. On the other hand, fibrotic or consolidated lung may have values of up to +100 HU and therefore segmentation of abnormal lung from adjacent soft tissue is problematic. Several approaches to overcoming this problem have been developed, all of which have advantages and disadvantages and include 'snake', 'rolling-ball' and region-growing algorithms. Other authors such as Hu et al (Hu et al., 2001) have developed algorithms which use a combination of methods such as

adaptive thresholding, region-growing and void filling. The adaptive thresholding method selects a density threshold which is determined by the individual scan and can therefore be varied according to differences in scan technique, patient and disease characteristics.

1.4.3 Automated quantification of lung density

Once the lungs have been segmented from the chest wall, it is then possible to analyse the characteristics of the lung parenchyma. One of the most simple quantitative metrics is the mean lung density (MLD), whereby the Hounsfield Unit value for each voxel of lung tissue is added together and divided by the total number of voxels. This approach is used in the analysis of emphysema, where diseased lung typically has lower density than normal lung (Müller et al., 1988). A number of other density measures have been used, such as the percentage of lung tissue below a defined density threshold. Commonly used thresholds for assessment of emphysema range between -950 and -900 HU (Coxson et al., 2013, Müller et al., 1988, Coxson, 2013).

Multiple studies have shown good correlation between quantitative emphysema indices, pulmonary function tests and patient reported symptoms and it has also been shown that MLD is increased in idiopathic pulmonary fibrosis (Hartley et al., 1994) . However, global indices of lung density are much less useful for the assessment of interstitial lung disease compared with emphysema. There are a number of reasons for this: firstly, areas of decreased attenuation such as the cystic spaces of honeycombing and traction bronchial dilatation are offset by the increased attenuation of reticulation, the borders of honeycomb cysts and areas of ground glass consolidation; secondly, IIP is by definition a spatially heterogenous disease (more so than emphysema) so global measures of lung density cannot reflect this.

Other problems with using density measurement are the variation with degree of inspiration, sensitivity to artefacts (e.g. beam-hardening) and dependence on scanner calibration (Parr et al., 2004, Coxson, 2013).

1.4.4 Analysis of the CT density histogram

Initial attempts to produce a more sophisticated measurement than MLD have focussed on analysis of the shape of the CT density histogram, a distribution representing the densities of all the pixels/voxels in a CT scan. Various metrics can then be used to describe the shape of the histogram, including kurtosis and skewness. Kurtosis describes how 'peaked' the histogram is compared with the normal distribution, which is considered to have a kurtosis of zero. A histogram which is more peaked than the normal distribution is said to have a positive kurtosis or to be leptokurtic, whilst a histogram which is flatter than the standard normal distribution is said to have a negative kurtosis or to be platykurtic. Skewness describes deviations in symmetry of a distribution compared with the symmetrical normal distribution. If the left tail of a distribution is longer than the right, it is said to be skewed to the left or negatively skewed. Positive skewness describes a distribution where the right tail is longer than the left. The CT density histogram of normal lung is strongly skewed to the left with a sharp peak around -800 HU (negative skewness and positive kurtosis). An illustration of the histogram distribution of a region of normal lung is shown below (Figure 1-4). This image was produced using the Osirix™ open source software (© Pixmeo Sarl).

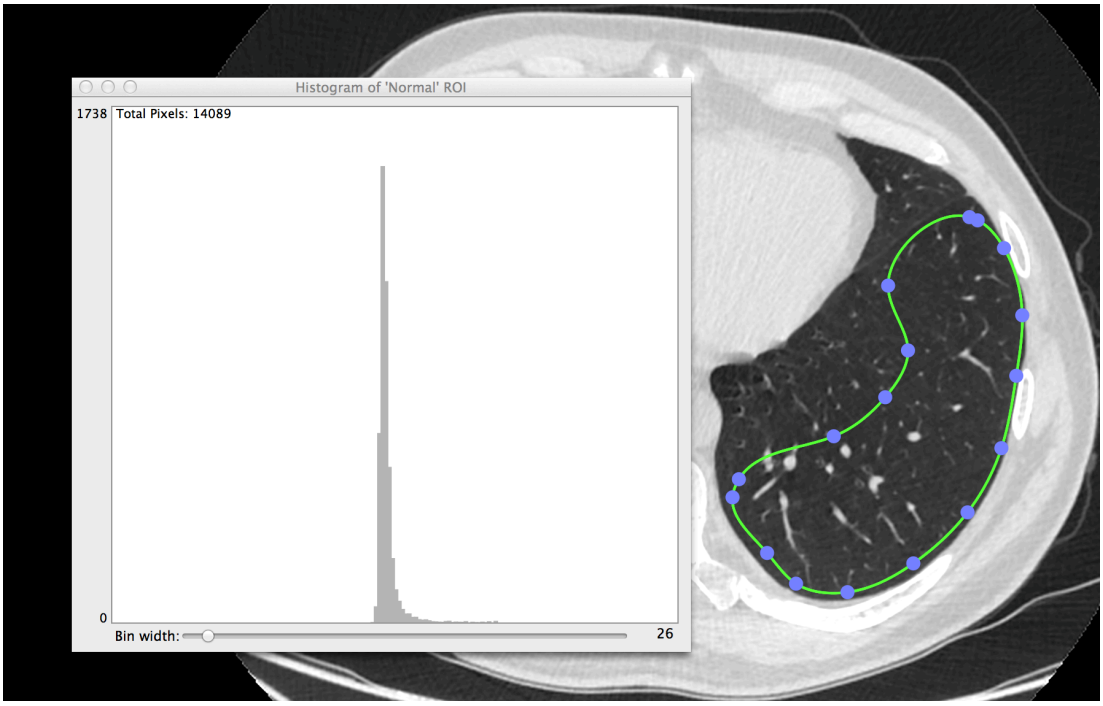

Figure 1-4 Region of interest (inside blue dots) of a slice of normal lung with its corresponding histogram. Note the sharp, narrow peak.

Pulmonary fibrosis typically causes an increase in the amount of soft tissue density (higher densities) in the lung and therefore causes an increase in mean lung density, a reduction in the peak and increased skewness to the left (Hartley et al., 1994). An example of the histogram from a patient with pulmonary fibrosis is shown in Figure 1-5.

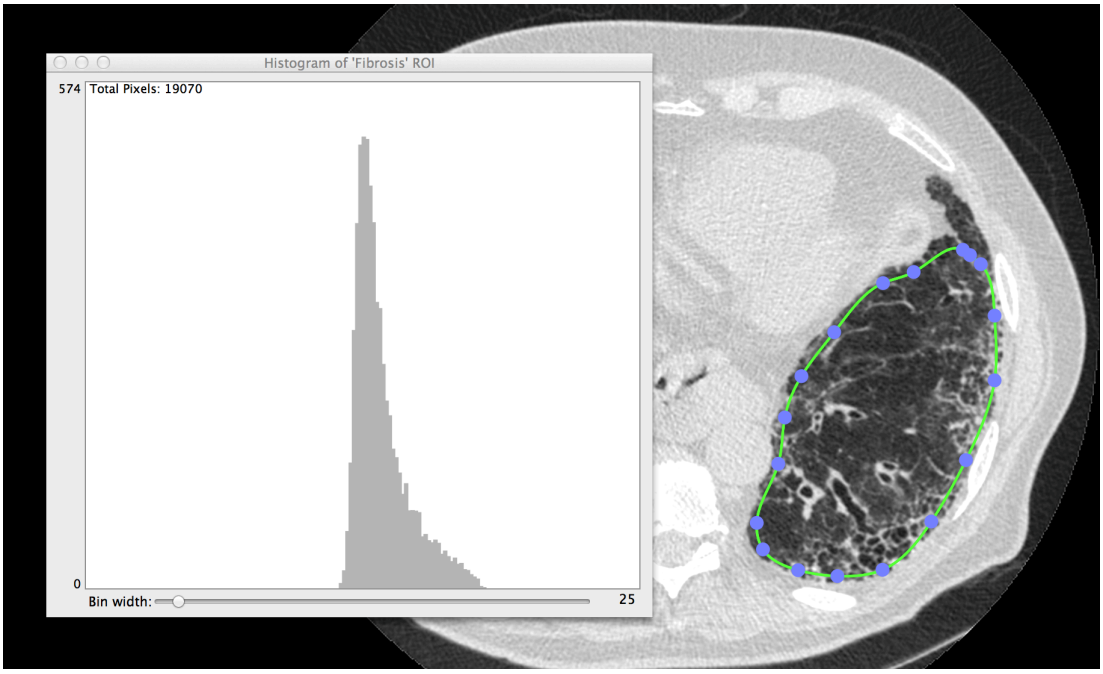

Figure 1-5 Region of interest (inside blue dots) of a slice of fibrotic lung with its corresponding histogram. There is a wider peak, larger right sided 'tail' and lower peak (NB: x-axis is automatically scaled on Osirix™).

Best et al (Best et al., 2003), in a study of 144 patients enrolled in a therapeutic trial, demonstrated moderate correlation between PFTS and kurtosis (r=0.53) but found relatively poor correlation between DL_{CO} and all histogram features, despite DL_{CO} being widely accepted as one of the most sensitive physiological measure of IIP. This study was limited by several factors including: retrospective design; CT scans obtained from 30 different institutions; CT scanners from 5 different manufacturers and lack of a standardised image acquisition protocol. Scans were typically acquired with a 2 cm interval between slices and manual correction was required to remove central airways and blood vessels.

Zavaletta et al (Zavaletta et al., 2007) used a more complex histogram-based method to analyse CT scans from patients with IPF. The method involved adaptive binning of the density histogram (using K-means clustering), followed by creation of a canonical signature for 5 sub-classes of lung pattern (normal, reticular, ground glass, honeycombing and emphysema). Fourteen scans were classified by 3 expert

radiologists into volumes of interest (VOIs) containing at least 70% of a lung pattern (reticulation, honeycombing, ground glass opacification, normal and emphysema) and these VOIs were used to train the classifier. The classifier was also trained on four whole (volumetric) scans. The algorithm was highly successful in distinguishing normal areas of lung from abnormal (sensitivity 93%, specificity 94%) but was less successful at distinguishing all 5 types of pattern with the following sensitivities and specificities: normal (92%, 95%), ground glass (75%, 89%), reticular (22%, 92%), honeycombing (74%, 91%), emphysema (94%, 98%). It is noteworthy that the least successful classification is in cases of ground glass opacification and reticular pattern. There may be several reasons for this. In terms of ground glass, this is a subtle and rather subjective density change which may be seen in normal lung in gravity dependent areas and when a scan is performed in relative expiration. In terms of reticulation, one of the major challenges is how to distinguish a linear 'reticulation' from a blood vessel. Both may be of similar width in cross-sectional diameter and of similar density. Whilst the authors attempted a semi-automatic segmentation and removal of blood vessels greater than one-third of the size of the VOI, they admitted that removal of smaller blood vessels is not yet reliably achievable in patients affected by fibrotic lung disease.

1.4.5 Analysis of lung texture - general approaches

A more sophisticated approach to computerised analysis in IPF is to look at textural features of the lung. This approach has been used in materials science (Mecke, 2000) and aims to quantify different visual patterns of disease. Since this is an important element of a radiologist's reading of a CT, automated textural analysis is a logical approach to the quantification of disease. Uppaluri et al developed an adaptive multiple feature method (AMFM) for the assessment of emphysema and then

extended this to the assessment of idiopathic pulmonary fibrosis (Uppaluri et al., 1999). The AMFM used in the emphysema study had the following stages: 1. Lung segmentation 2. Pre-processing of the scan using 'edgmentation' – a region-growing technique that merges adjacent pixels where the difference in grey level of the pixels is small and then assigns a single grey level to the whole region based on the average of the pixels within the region. 3. Regions of interest defined on the original and pre-processed image. 4. Feature extraction – five first-order features (mean, variance, skewness, kurtosis and grey-level entropy) and eleven second order features (five run-length features: short-run emphasis, long-run emphasis, grey-level non-uniformity, run-length non-uniformity, run percentage and six co-occurrence matrix features: angular second moment, entropy, inertia, contrast, correlation, inverse difference moment).

Whilst Uppaluri et al analysed multiple different features prior to selecting those which were most discriminating, Uchiyama et al (Uchiyama et al., 2003) pre-selected 6 features designed to address the specific task. Of the 6 pre-determined features, there were 3 grey-scale distribution measures including the mean CT value of an ROI, the standard deviation of CT values in an ROI and the fraction of lung with density between -910 HU and -1000 HU in an ROI. The other three features were shape measures including measures of nodularity, linearity and multi-loculation. An artificial neural network algorithm was used, with training based on 315 slices marked up by 3 radiologists and comprising the following textures: ground-glass opacity; reticular and linear opacities; nodular opacities; honeycombing; emphysematous change and consolidation. Two other mark-up labels were employed but not used for training the algorithm, namely 'non-specific/indeterminate' and 'other' which included any other abnormal feature such as atelectasis, bullae or artefact. The algorithm was able to reliably distinguish honeycombing from normal

and ground glass opacity, but ground glass opacity showed some overlap with normal lung. They also found it difficult to distinguish nodularity from normal lung, presumably partly due to the non-spiral nature of the CT scans and the fact that blood vessels may appear very similar to nodules in cross-section. Of 53 indeterminate slices, the algorithm classified 28 as normal and 25 as abnormal.

Sluimer et al (Sluimer et al., 2006) developed an algorithm for the textural analysis of diffuse parenchymal lung disease using two different sets of texture-analysis features, one that they had previously described and a set based on the Uchiyama method. They classified lung into one of six classes: normal, hyperlucent, fibrotic (including reticulation, honeycombing and traction bronchial dilatation in association with ground glass), ground glass, solid and focal (including solid and ground glass nodules, mucus-plugging, scars).

They did not find any significant difference in the performance of the system depending on which of the two texture-analysis feature sets was used. One interesting feature of their approach was the use of non-square regions of interest, designed to encompass more of an area of interest and ensure it contained a more homogenous texture. This approach improved the performance of the algorithm with up to a 10% increase in the area under the ROC curve.

In a more recent study, Yoon et al (Yoon et al., 2013) used a texture-based automated quantification system (AQS) to assess 89 patients with fibrotic interstitial lung disease (71 UIP and 18 fibrotic NSIP) each of whom had 2 CT scans performed one year apart. The AQS classified the lung on each CT slice into the following categories: normal, emphysema, ground glass opacification, reticular opacities, honeycombing or consolidation and calculated the percentage of lung affected by each category. Interestingly, they used round ROIs with a diameter of 10 pixels whereas most studies use square ROIs. They also calculated the total abnormal lung

fraction (sum of all abnormal lung categories) and a fibrosis score (sum of honeycombing and reticular opacity). The AQS was compared with visual readings by 2 radiologists who visually estimated the percentage of involved lung on each CT slice. The CT scans were not volumetric but rather 0.75 mm slices, which were obtained at 10 mm intervals. Interclass correlation coefficient (ICC) was used to assess the degree of inter-reader agreement and the degree of agreement between reader 1 and the AQS, reader 2 and the AQS and the mean of the two radiologist scores and the AQS. Scores were also compared with the patient's FVC and DL_{CO}. There was good agreement between both readers and the AQS in terms of fraction of emphysema and honeycombing (ICC 0.70 – 0.79 and 0.62 – 0.79 respectively) but the readers had less good agreement in terms of reticular opacity (ICC 0.49). Both readers had relatively poor agreement with the AQS in terms of ground glass opacification (R1 and AQS = 0.36, R2 and AQS = 0.44) and relatively poor agreement with the AQS in terms of reticular opacification (R1 and AQS = 0.32, R2 and AQS = 0.40). Reader 1 had relatively good agreement with AQS in terms of consolidation (ICC = 0.66) but reader 2 had less good agreement (ICC = 0.39). Agreement at the one-year follow up scan was not so good, with poor agreement between radiologists and the AQS in terms of whether there was interval change in the percentage of affected lung.

1.4.6 Analysis of lung texture – Minkowski functionals

Minkowski functionals (MFs) are a group of integral geometry measures that describe the geometry and topology of an image and have been used in both materials science and cosmology. In 3D, there are 4 different MFs which are proportional to volume, surface area, mean breadth and the Euler-Poincaré characteristic. A precursor to analysis of MFs is to binarise the image so that all pixels are either black or white. This can be done at a number of different pre-defined density (Hounsfield Unit)

thresholds or at an adaptive threshold which is selected depending on the image properties. The topological features are based on the connectivity of the pixels in an image. Connectivity can be expressed as either 4-connectivity, in which a square pixel is said to be connected to an adjacent pixel if it contacts one whole side of that pixel or as 8-connectivity, in which a pixel is said to be connected even if it only contacts a corner of the pixel (figure 1-6).

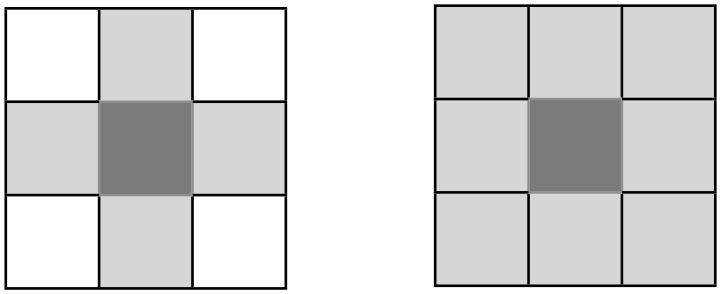

Figure 1-6 4-connectivity (left) and 8-connectivity (right)

The first report of MFs for the analysis of thoracic CT scans was proposed by Boehm et al in 2008 (Boehm et al., 2008). They used MFs to analyse a total of 275 volumes of interest from 7 patients with emphysema, 7 patients with pulmonary fibrosis and 7 patients with normal lungs. All these VOIs were cubic volumes with edge length of 40 pixels. Only a proportion of each scan was sampled. Computerised classification of pathological subtype (emphysema, fibrosis or normal) was compared with classification by 2 radiologists. They found that the computer algorithm agreed with radiologist classification in 98% of normal VOIs and in 86% of fibrotic VOIs. In 2010, this work was extended by Thonnes et al (Thönnes et al., 2010) with use of smaller VOIs (approximately 5 x 5 x 1 voxels). A total of 10 fibrotic, 10 normal and 8 emphysematous VOIs were analysed with a mis-classification rate of 7%. The advantage of the smaller VOIs is to increase the spatial accuracy of classification, although image noise may be more of a problem.

1.5 CONCLUSION AND THESIS STRUCTURE

In summary, there is a need for more accurate methods of assessing the severity of IIP and measuring change in disease over time. Methods of quantification have progressed from visual estimates, to single quantitative metrics to the assessment of multiple textural features. Further work needs to be done to assess the use of these metrics in a prospective study with strictly controlled scanning parameters and detailed clinical correlation. In this thesis we will report on the development and testing of a computer algorithm to analyse lung texture in CFIP, followed by its application to scans from a group of prospectively recruited patients. The computer output will be compared with radiologists visual estimation, with physiological measures of lung function and with a questionnaire-based assessment of patient symptoms and wellbeing. Chapter 2 will describe the methodology of the study, Chapter 3 will describe the testing of the computer algorithm on a number of pre-selected, retrospectively-obtained normal and abnormal CT scans, Chapter 4 will describe the results of the prospective study in terms of comparing the computer and radiologist output, Chapter 5 will describe the comparison of the computer output with lung function tests and the symptom and well-being questionnaire, Chapter 6 will comprise an overall discussion of the results and Chapter 7 will discuss future directions.

CHAPTER 2: MATERIALS AND METHODS

2.1 INTRODUCTION TO CHAPTER

In this chapter we will discuss the study design including patient recruitment, ethics approval, patient selection and sample size. We will then discuss the timeline of study visits and the variables measured at each visit, including a detailed description of the CT scanning protocol. This is followed by an account of the automatic computerised quantification algorithm, detailing its design and development. Finally we describe the procedure for radiologist visual scoring of CT scans and the planned statistical analysis. The computer analysis work was done in collaboration with Dr Abhir Bhalerao at the Department of Computer Science at Warwick University, who also kindly supplied Figures 2-2 to 2-5.

2.2 STUDY DESIGN

2.2.1. Study overview/aims

The work in this thesis forms part of a larger study, the Quantification of Interstitial lung disease on CT (QUIC) study. The QUIC study is a prospective longitudinal cohort study of patients with idiopathic interstitial pneumonia. The primary aim of the QUIC study is to see if change in fibrosis on CT, as assessed by an automated computer algorithm, is a better predictor of mortality at 5 years than change in pulmonary function tests. Therefore, the primary outcome measure was all cause mortality at 5 years. Secondary aims were to compare automated computerised estimation of fibrosis with radiologist estimation of fibrosis, pulmonary function tests and a patient-reported outcome measure.

Due to delays in starting the study and the fact that recruitment that was slower than expected, this thesis will focus on the baseline results from the first 24 patients which

were recruited and we will not therefore address the primary outcome measure in this thesis. Therefore the aim of the current thesis is to describe the development and testing of a novel computer algorithm based on Minkowski functionals for the quantification of CFIP on CT.

2.2.2 Ethical approval

Full ethical approval was obtained from the West Midlands Research Ethics Committee (study reference 11/WM/0387).

2.2.3 Study population

The study population consisted of patients with a diagnosis of chronic fibrotic interstitial pneumonia (CFIP) based on clinical findings and either a surgical lung biopsy showing UIP or NSIP and/or a CT showing a typical UIP pattern as described by Raghu et al (Raghu et al., 2011). All patients were discussed at the bi-weekly regional interstitial lung disease multi-disciplinary team meeting at University Hospital Coventry and Warwickshire NHS Trust, where a consensus diagnosis was documented. The table below shows the patient inclusion and exclusion criteria:

Inclusion criteria	Exclusion criteria
Diagnosis of CFIP based on clinical findings and either lung biopsy and/or CT scan	Clinical, biopsy or CT findings suggesting a secondary cause of interstitial lung disease
Age greater than or equal to 40 years.	Age less than 40 years
Ability to provide informed consent	Inability to provide informed consent
Ability to breath-hold or lie flat for the CT scan	Inability to breath-hold or lie flat for the CT scan
	Pregnancy

Table 2-1 Patient inclusion and exclusion criteria

2.2.4 Patient recruitment

Patients who were thought potentially suitable for the study were approached by their usual physician during their routine outpatient appointments. If they expressed an interest in participating in the study, they were given a copy of the patient information leaflet. They were invited to take this away, think about the study and contact the study co-ordinator if they decided they would like to participate.

2.2.5 Sample size calculation

We planned to recruit a total of 80 patients with idiopathic interstitial pneumonia (UIP or NSIP), averaging 1-2 patients a week. UIP has a mortality of 50-70% at 5 years whilst idiopathic NSIP has a mortality of approximately 20% at 5 years. Therefore we expected half the patients to be alive at 5 years. Assuming a linear regression model linking radiological change to time to death, for 4 predictor variables and a sample of 80 patients we would have approximately 80% power to detect a medium effect size (R2 = 0.13), at the 5% significance level (Cohen, 1988).

Previous studies have examined between 39 (Xaubet et al., 1998) and 109 patients (Flaherty et al., 2003b) with CFIP and have analysed the relationships between findings at HRCT and changes in DL_{CO} and FVC. Flaherty et al (Flaherty et al., 2003b) showed that a greater than 10% change in FVC over 6 months was an independent predictor of mortality. They did not find any predictive value of changes on CT, as assessed by a semi-quantitative visual scoring system but suggested that use of a computerised scoring system may be more sensitive. Xaubet et al (Xaubet et al., 1998) studied 39 patients who underwent 2 CT examinations at a mean interval of 7.5 months. With a semi-quantitative visual scoring system they showed an approximately 7% change in global disease score between scans and found that this change was significantly

correlated with DL_{CO} and FVC. We therefore concluded that 80 was an appropriate number of patients to recruit. At the time of writing, 24 patients have been recruited.

2.3 STUDY VISITS

2.3.1 Schedule of visits

The QUIC study protocol specified that patients should undergo a total of 5 study visits over a 24 month period. A detailed case record form (CRF) was completed at baseline (appendix A) and a shortened CRF was completed at each subsequent visit. The St George's Respiratory Questionnaire (Appendix B) and full pulmonary function tests were completed at each visit. Screening blood tests were performed at the first visit including: full blood count; urea and electrolytes; liver function tests; creatine kinase; rheumatoid factor; anti-CCP titre and nuclear antibodies. The latter four tests were to look for possible connective tissue disease.

CT scans were performed at 0, 3, 12 and 24 months. All other tests were performed at 0, 3, 6, 12 and 24 months. As previously mentioned, this thesis will be limited to analysis of the baseline investigations.

2.3.2 Clinical assessment/case record form

Patients were each assessed by one of the study respiratory physicians (DP or FW), both of whom have a subspecialty interest in interstitial lung disease. A diagnosis of CFIP was only made only once secondary causes of interstitial pneumonia had been excluded. Secondary causes included exposure to inorganic dusts (pneumoconioses), organic dusts (hypersensitivity pneumonitis) or therapeutic agents known to cause interstitial lung disease. Patients were questioned about symptoms associated with collagen vascular disease and examined for signs of these conditions. Patients were designated

to have a CFIP if they did not have a secondary cause and did not fulfill the criteria for a defined rheumatological condition.

The CRF was designed to capture information about all current and previous medical conditions, current medication and exposures to possible toxic agents (including a detailed occupational history). The CRF was also used to record lung function and blood test results.

2.3.3 St George's Hospital Respiratory Questionnaire

The St George's Hospital Respiratory Questionnaire (STGRQ) was used to assess the severity of patients' symptoms and their impact on their daily activities. The STGRQ is a 50 item disease-specific questionnaire which was originally designed to be used in patients with chronic obstructive airways disease (COPD) and asthma (Jones et al., 1992). The questionnaire has two parts. The first part asks the patient about their symptoms in the preceding 3 months and produces the 'symptoms' score. The second part is concerned with how patients are functioning currently, how their disease affects their physical functioning (the 'activities' domain), how it affects their psychological state and how it affects their social functioning (the 'impacts' domain). The full questionnaire is reproduced in Appendix B. The highest maximum total STGRQ score is 100 and the highest score for each of the domains is also 100. A higher score indicates a greater degree of limitation. Each response is individually weighted and the domain/total scores are calculated using a free custom-designed excel spreadsheet supplied by the St George's group. The questionnaire was developed by Professor Paul Jones at St George's Hospital and although originally validated for the assessment of patients with asthma and COPD (Jones et al., 1991, Jones et al., 1992), it has also been used in a number of studies of IPF. For example, in their 2005 systematic review of the use of health related quality of life (HRQL) questionnaires in IPF, Swigris et al (Swigris et al., 2005) found 7

studies which met their inclusion criteria and which enrolled between 10 and 330 patients into various studies (median 34 patients). All of these studies administered their questionnaires at a single time-point. Three of the studies used the STGRQ, three used a generic (not respiratory specific) HRQL questionnaire called the 'short form 36' (SF-36) (Ware Jr and Sherbourne, 1992), and two used a generic quality of life form developed by the World Health Organisation (WHOQOL-100) (group, 1995). The SF-36 and WHOQOL-100 both look at multiple aspects of wellbeing and health, whereas the STGRQ is specific to patients with respiratory disease. Five of the studies were cross-sectional in design and two were therapeutic trials. Only one of the studies was specifically designed to try to validate the use of the HRQL questionnaire in IPF. This cross-sectional study by Martinez et al (Martinez et al., 2000) compared 34 patients with IPF with 34 age and sex-matched controls. They administered the SF-36 questionnaire and another respiratory questionnaire, the Baseline Dyspnea Index (BDI) (Mahler et al., 1984), to IPF patients and controls; IPF patients also underwent pulmonary function tests and resting arterial blood gas measurements. They found that the IPF patients scored significantly lower than normal subjects on 7 out of 8 components of the SF-36. They also found that there was significant correlation between five of the SF-36 score components and the BDI and that there was significant correlation between two components of the SF-36 score (physical functioning and general health perceptions) and spirometry (FEV_1 and FVC).

More recently, Swigris et al (Swigris et al., 2010) looked to further validate the SF-36 and the STGRQ in a large group of patients with IPF and try to estimate the minimum important differences in scores for the two questionnaires. The data was collected as part of the BUILD-1 study of the use of Bosanten in IPF. The questionnaires were administered to 158 patients with IPF who were randomised to receive either Bosanten or placebo. They found that changes in both the SF-36 and STGRQ reflected changes in

patients disease progression (as assessed by the BDI, FVC and DL_{CO}). They found that the minimum important difference in scores was 2-4 for SF-36 and 5-8 for STGRQ.

2.3.4 Pulmonary Function Tests

All patients underwent full lung function tests at baseline including spirometry (FEV_1, FVC), lung volumes (TLC, RV, ERV and FRC) and transfer factor (TLCO, KCO, VA). TLC was measured using the helium dilution technique and DL_{CO} was measured using the single breath technique.

2.3.5 CT scans

The study protocol specified that scans should be performed within 2 weeks of pulmonary function tests. All patients were scanned on a state of the art 64-slice CT scanner (Discovery HD - GE Healthcare, Milwaukee, WIS). The following table details the CT scan parameters and the choice of parameters is described in more detail below:

Parameter	Setting
Number of detectors	64
Pitch	0.98
kV	120
mA	100
Slice thickness (mm)	0.625
Reconstruction algorithms	Bone, Lung and Standard

Table 2-2 CT scan parameters

2.3.5.1 Spiral vs non-spiral acquisition

In the early days of CT, high-resolution CT (HRCT) of the lungs was performed using a non-spiral technique, since acquiring a whole lung volume using thin slices on the

earlier generation of CT scanners would take much longer than a reasonably achievable breath-hold. Therefore, the percentage of lung scanned had to be sacrificed for the increased resolution achieved by using thin slices. With the advent of modern multi-detector scanners it has become possible to obtain whole lung coverage using thin slice spiral acquisitions. There are many advantages of this volumetric spiral scanning, including the possibility to acquire a dataset which represents the whole of the lungs and the ability to reconstruct images in multiple orthogonal planes. For situations where serial scans need to be compared volumetric scanning techniques make slice matching much easier. For quantitative analysis, spiral acquisition is necessary for 3D-analysis techniques to be performed. The potential downside of spiral acquisition is that it typically results in a higher patient radiation dose than non-spiral CT and therefore non-spiral CT continues to be used in certain circumstances such as in paediatric patients. In the older patients who are typically affected by IPF, radiation dose is not usually such an important consideration, since older patients are less sensitive to the effects of radiation and have less time to develop the potential side effects (Kleinerman, 2006).

Another important feature of the study protocol is that patients were scanned from the lung bases to the lung apices (caudo-cranially). This is the opposite direction to how most CT is obtained. The reason for starting the scan at the lung bases is that during normal breathing there is much greater excursion of the lower part of the chest compared with the apices. Thus if a patient is unable to hold their breath for the duration of the scan, any breathing movement is more likely to affect the apices, which are less mobile, thus leading to less breathing artefact.

2.3.5.2 Slice thickness

Choice of slice thickness on CT is a compromise between image noise, spatial resolution and partial volume effect. Thinner slices have greater image noise (due to less data) but have better spatial resolution and decreased partial volume effect. In order to facilitate accurate 3D reformatting and analysis of texture, we decided that the CT scan should be acquired with the thinnest possible slices achievable on our scanner, which for the GE Discovery HD is 0.625 mm. This results in near-isotropic voxels.

2.3.5.3 Dose parameters - mA and kV

In choosing appropriate levels of mA and kV, consideration must be given to patient dose, image noise and image contrast. An increase in mA leads to a linear increase in dose, whereas the relationship between kV and dose is non-linear with, for example, a 20% increase in kV leading to an approximately 30% increase in dose.

Since the lung shows high inherent contrast between aerated lung and adjacent structures, a low dose scan protocol was chosen with a fixed mA of 100 and a fixed kV of 120. Previous studies have shown that such low dose protocols do not lead to an important decrease in the ability of automated algorithms to detect lung nodules (Diederich et al., 1999) and that emphysema quantification is not significantly hampered by thresholds as low as 50 mAs (Zaporozhan et al., 2006). Sverzellati et al performed an analysis of histogram features in patients with idiopathic interstitial pneumonia and did not find any significant difference between scans performed at 50 mAs and scans performed at 100 mAs (Sverzellati et al., 2005).

Specific dose reduction post processing techniques such as adaptive statistical iterative reconstruction were not used because the effect on automated quantification is not yet known.

2.3.5.4 Pitch

The pitch of a multi-detector CT scanner can be defined as the distance moved by the table during one rotation, divided by the detector collimation (total length of detectors) (Schilham et al., 2010) and can be thought of in terms of how tightly a spring is coiled, imagining the beam of the x-ray tube tracing a spiral around the patient. A pitch of 1 means a spring where there are no gaps between the coils of the spring but no overlapping of coils. Therefore during a single 360 rotation of the x-ray tube, each element within the body is sampled twice. A pitch of less than 1 implies overlapping and a pitch of more than 1 implies 'gaps' between the springs and reduced sampling. For our study, a pitch as near as possible to 1 (0.98) was chosen as a compromise between the reduced amount of data acquired with a pitch of more than 1 and the oversampling and increased dose of a pitch less than 1.

2.3.5.5 Reconstruction algorithm

CT raw data can be reconstructed in multiple different ways using different mathematical algorithms. A number of studies have examined the effect of using different reconstruction algorithms on automated quantification of emphysema. They concluded that most accurate results were obtained with use of a 'soft tissue' algorithm. Studies of the effect of algorithm choice on automated quantification of IPF are less numerous and a number of different algorithms have been used in quantitative studies including bone (Maldonado et al., 2014, Sverzellati et al., 2005), lung (Yoon et al., 2013) and 'non-edge enhanced' (Bartholmai et al., 2013) . At the time of writing, we are not aware of any texture-based quantification study which compares more than one algorithm.

2.3.5.6 Window level and width

The CT attenuation of a voxel of tissue is defined by its CT number which is defined as the difference in attenuation of the contents of the voxel relative to water. The CT number is expressed in terms of the Hounsfield Unit (HU) where water is assigned a value of zero. The window level and width specify the CT numbers which define the midpoint and the range respectively of the gray-level scale used to display a CT image (Barnes, 1992). Radiologists viewed all images on standard lung windows (window level = - 500 HU, window width =1400 HU).

2.3.5.7 Level of inspiration/breathing instructions

The QUIC protocol specifies that the scans are performed at full inspiration. The most important factor in ensuring that a CT scan of the chest is suitable for quantitative analysis is making sure that the patient achieves the appropriate level of inspiration and expiration and that the patient holds their breath during the scan (Newell Jr et al., 2013). In order to ensure this, specific breathing instructions were created for the QUIC study and pre-programmed onto the scanner. The instructions are given in the table below:

Breathing instructions	"Breath in....breath out....breath in.... breath out....breath all the way in... and stop breathing"

Table 2-3 CT scan breathing instructions

2.4 DEVELOPMENT OF THE AUTOMATED COMPUTERISED ANALYSIS SOFTWARE

2.4.1 Steps required for automated analysis

The process of developing the algorithm was an iterative process building on a previous algorithm developed by the Department of Computer Science at Warwick University (Charemza et al., 2008). The process can be broken down into a number of steps as follows:

Step 1: Separation of the lung voxels from the chest wall and the air outside the thorax

Step 2: Segmentation of the major airways and their removal from the lung volume

Step 3: Training the algorithm using marked up data, analysis of the different lung textures and development of classifiers according to the different Minkowski functionals

Step 4: Texture analysis of the new lung volume using Minkowski functions.

Step 5: Classification of the lung voxels according to similarity to the classifier.

In the first step, the lung voxels must be separated from surrounding structures. There have been many different methods described for automatically segmenting the lungs from the chest wall which are described in a recent review article by van Rikxoort et al (van Rikxoort and van Ginneken, 2013). These methods vary in their sophistication, computational efficiency and degree of user input required.

In people with normal lungs the task can be relatively easily accomplished since the lung is typically much lower attenuation than the structures of the chest wall. The density of normal lung is typically below -500 HU whereas normal chest wall structures are greater than -50 HU and there is therefore a large difference between the two. A process of simple thresholding can therefore be used to isolate the lungs using a threshold of, for example -100 HU. On the other hand, diseased lung often has increased density and may therefore be very similar in attenuation to the adjacent chest wall. This makes separation of abnormal lung and the chest wall difficult. Another challenge is that diseased lung (e.g. emphysema, honeycomb cysts) may have abnormally low density, creating 'holes' in the image which may be confused with airways. We chose an approach

which uses a combination of thresholding and a 'shrink-wrapping' technique. This technique has a number of steps, as follows:

1) Segmentation: a basic initial segmentation is performed using an adaptive thresholding technique as described by Hu et al (Hu et al., 2001). As shown below (Figure 2-1) , this may have some 'holes' i.e. fail to include some lung tissue, both at the periphery of the lungs and more centrally within the lung parenchyma.

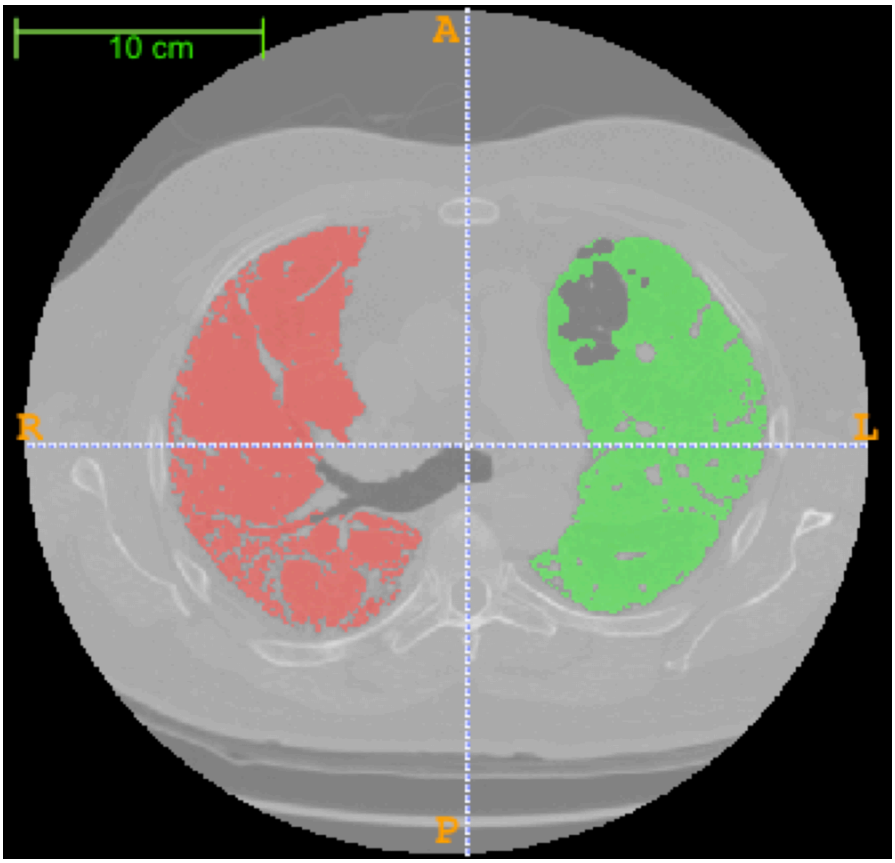

Figure 2-1 Initial segmentation of the lungs using thresholding

2) 'Shrink-wrapping': this technique involves generating multiple points on the surface of the lung volume and then casting linear rays through these points in order to identify points exterior and interior to the surface. A bounding, convex mesh is then initialised outside the surface points. Figure 2-2 provides a schematic representation of this using a cube and sphere as the object to be segmented. Next, the positions of the mesh are updated using a self organising map (SOM).

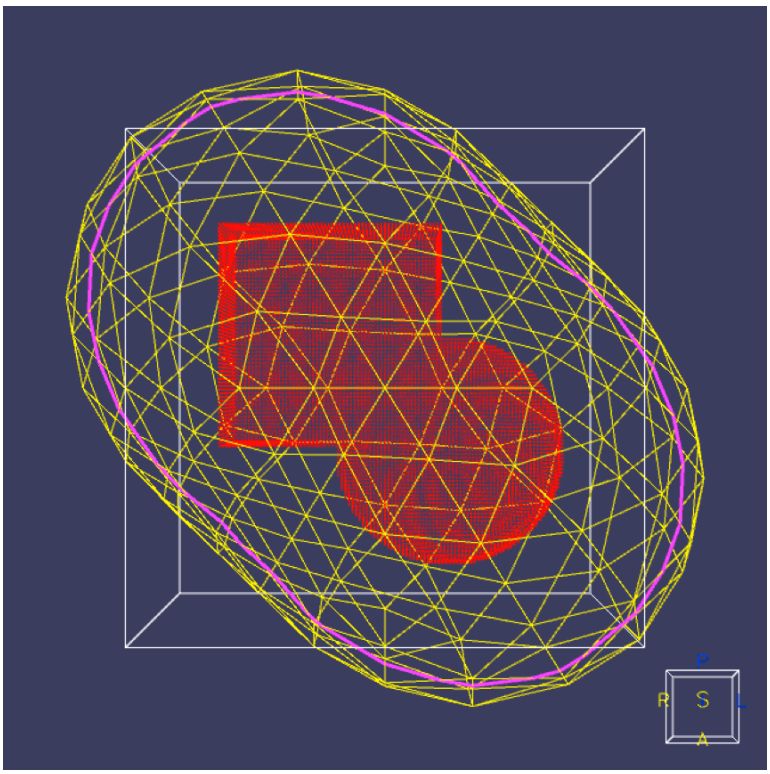

Figure 2-2 Convex mesh placed outside the desired volume

This is a type of artificial neural network where points (called 'nodes' or 'neurons') are assigned to an approximation of the desired structure and then moved in position according to their similarities with neighbouring nodes. Multiple iterations of this process gradually decrease the size of the polygons forming the mesh and allow it to move closer to the desired shape. This procedure was originally described by Kohonen et al and therefore SOMs are sometimes known as Kohonen maps (Kohonen, 1982). Figure 2-3 shows several iterations of the software as it approximates the lung surface and Figure 2-4 shows a 3D surface-rendered representation of the final segmented lung volume.

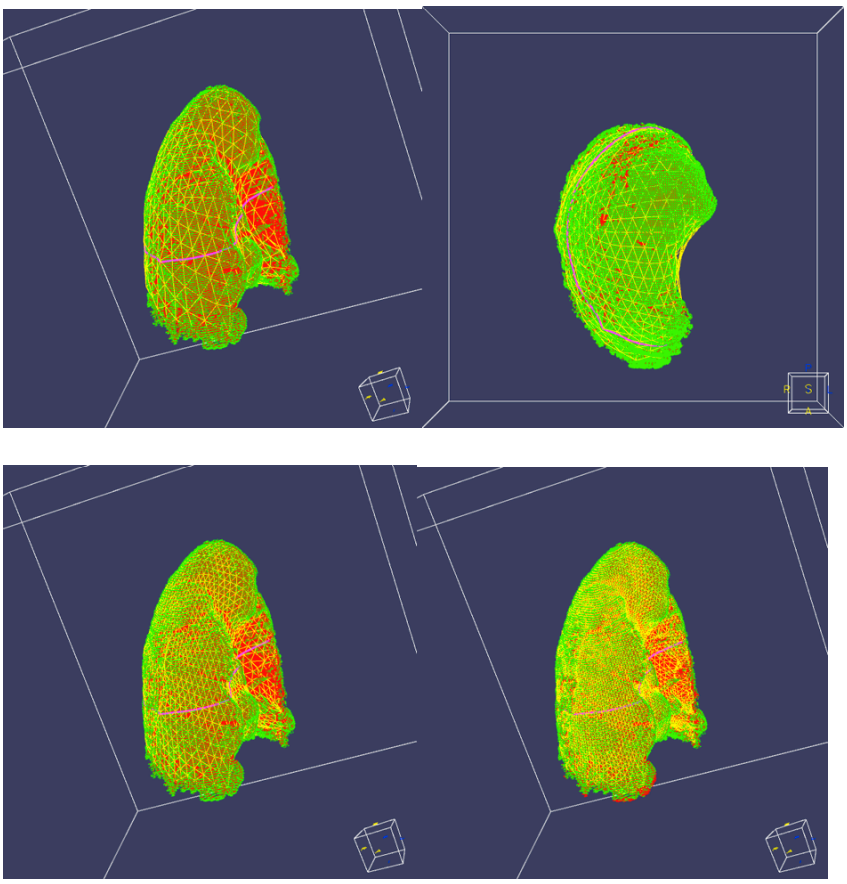

Figure 2-3 Several iterations of the self-organising map (SOM) showing gradually decreasing surface polygon size as the external mesh approximates to the true lung volume

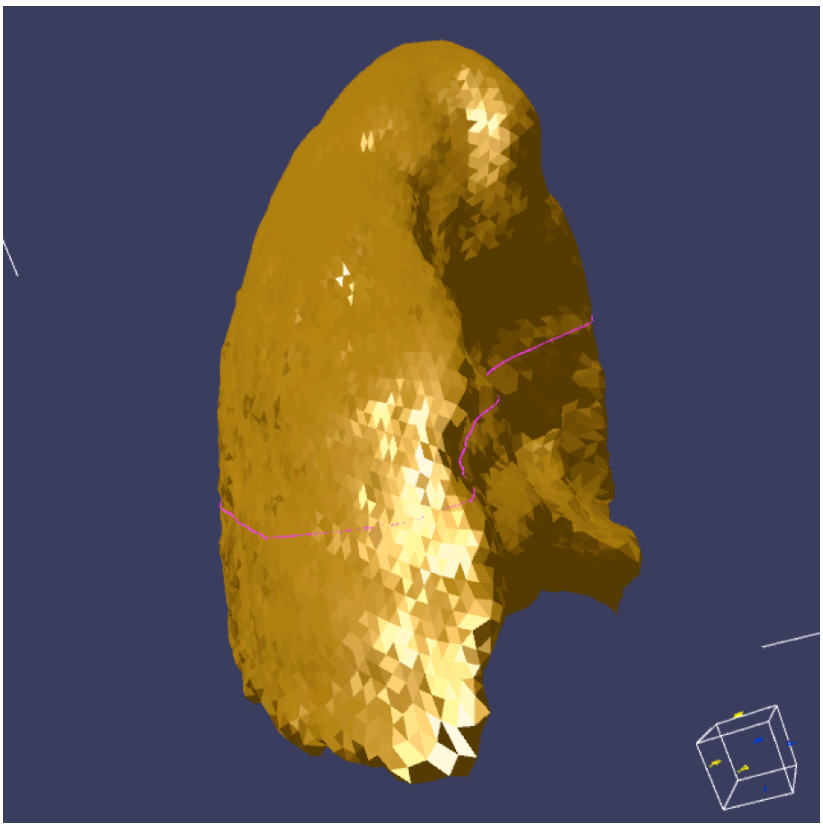

Figure 2-4 Surface rendered representation of the final 3D lung volume after it has been 'shrink-wrapped'

3) Removal of airways: the major airways must now be removed from the lung volume. The method chosen is based on the approach by Doel et al **(Doel et al., 2012)**. Firstly, a seed point is first manually placed in the trachea. Next, a region-growing algorithm is used whereby neighbouring voxels are grouped together if their density is below a threshold of -900 HU. An 'explosion-control' mechanism is used to prevent low attenuation lesions adjacent to airways being included within the airway volume. This mechanism relies on the fact that airways will generally decrease in size over successive generations and stipulates that, as the region growing algorithm progresses, the number of voxels joining the volume should gradually decrease (apart from small temporary increases at airway bifurcations). If there is a sudden increase in the number of voxels joining the airway tree, this implies 'leakage' into the surrounding lung and

the process is terminated. Once the airway volume has been 'grown', it is subtracted from the lung volume. Figure 2-5 (below) shows an example of the extracted airways.

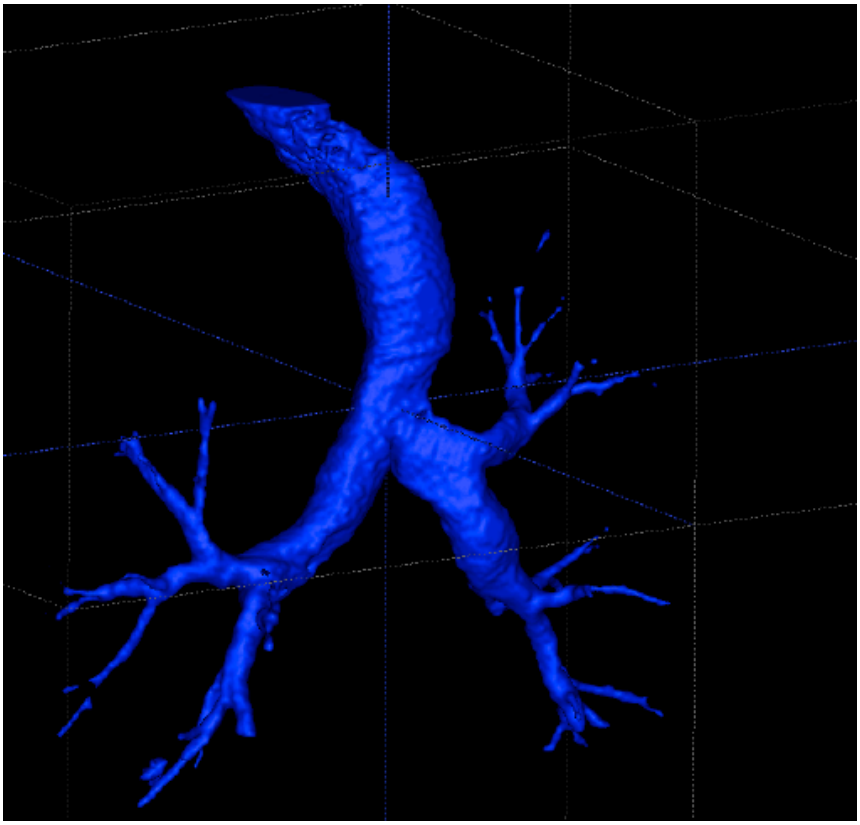

Figure 2-5 Extracted major airways

4) Training: the next stage in the process is to train the software using expert radiologist mark-up. In order to provide the algorithm with ground truth data, a Radiologist (EH - 'Radiologist 1'), marked up a scan of a patient with IPF. The scan was chosen to be representative of all the typical features of IIP and was of a patient with moderate severity disease. The algorithm was marked up using a software called ITK-snap (http://www.itksnap.org), an open-source software which allows the viewing and drawing of irregular ROIs on a CT scan saved in DICOM format. These ROIs can be colour-coded to indicate different types of abnormality and saved in several different formats **(Yushkevich et al., 2006)**. For the purpose of this study, the .NRRD format was chosen.

The labels used and colour codes were as follows:

Texture	Colour code	Degree of Confidence
Honeycombing	Red	Definitely abnormal
Reticulation	Green	Definitely abnormal
Emphysema	Orange	Definitely abnormal
Ground glass change	Yellow	Definitely abnormal
Indeterminate*	Cyan	Subtly abnormal
Normal	Purple	Definitely normal

Table 2-4 Labels for radiologist mark-up of scan 13

Category 5 (indeterminate) was selected when the region of lung did not fit into any of the other categories. Generally, this was when the region was only subtly abnormal but not abnormal enough to be coded as 1-4. Particular care was taken to only mark as normal areas which were definitely normal, since reliable categorisation of lung into normal or abnormal was felt to be one of the most important characteristics of any algorithm.

Initially, the whole right lung of a single scan was marked up. This involved marking up 409 slices using an average of 7 ROIs per slice producing a total of approximately 2800 irregular ROIs. This task took approximately 40 hours in total.

The whole of the lung volume was then divided up by the computer algorithm into non-overlapping ROIs of 5 x 5 x 5 pixels each. This resulted in 13,855 ROIs which had been assigned by the radiologist to the honeycomb class, 14767 reticulation ROIs, 4875 indeterminate ROIs and 5422 normal ROIs (total 38,919 ROIs - table 2-5). Since there were an uneven number of voxels in each class, and in order not to bias the training of the algorithm, a total of 3000 voxels of each class were used to train the algorithm.

Class	Number of ROIs
Honeycombing	13855
Reticulation	14767
Indeterminate	4875
Normal	5422

Table 2-5 Number of ROIs marked-up for each class

5) Calculation of Minkowski Functionals: the next stage in training the algorithm was to calculate the Minkowski functionals (MFs) for each of the voxels in order to try and separate out the different classes. The calculation of the MFs can only be performed on binary data so the pixel density had first to be converted to black or white (rather than the 256 shades of grey which are shown in a typical CT image). This can be done at a single threshold or at multiple thresholds. We chose to perform the calculations at 100 thresholds of 10 HU from -1000 to 0. The four MFs were calculated for each of the 100 thresholds giving 400 samples per ROI. The formula which is used to derive the MFs for convex sets is expressed in terms of the volume of a given set when dilated by a ball, Br of radius r. In 3D the formula is :

$$V(K \oplus Br) = V(K) + S(K)r + 2\pi B(K)r^2 + {}^4/_3\chi(K)r^3$$ (Arns et al., 2002)

Where K is a convex set and $\oplus$ is the dilation operation; V= volume; S = surface area, B = mean breadth and χ = the Euler-Poincare Characteristic (EPC)

Building a classifier: a neural network (NN) using a multi-layer perceptron (MLP) was used to build a classifier. This approach is similar to that adopted by Huber et al and is suited to dealing with the high-dimensionality of our multiply thresholded features.

Having designed a computer algorithm according to the steps listed above, we then tested it in a number of different situations, as detailed below.

2.4.2 Testing the computer algorithm – control scans

The computer algorithm was first tested on normal CT scans, designed to act as controls. For this purpose, 7 consecutive high resolution CT scans of the thorax which had been reported as normal were selected from the routine CT work-list. These were validated by EH as being normal, then fully-anonymised (no patient identifiable data) and the computer algorithm was run on these scans. Results of this experiment are presented in Chapter 3).

2.4.3 Testing the computer algorithm against a CT scan marked up by a radiologist

Next the algorithm was tested on the initial scan (patient 13) which the radiologist had marked up for training the algorithm. Although the radiologist had marked up only the right lung, the algorithm was applied to both the left and right lungs. Comparison of the radiologist mark-up with the computer output for the left lung is discussed in Chapter 3.

2.4.4 Testing the computer algorithm against selected slices marked up by a radiologist

The algorithm was tested on multiple scans from prospectively recruited patients. One subject (patient 19) was subsequently omitted due to their scan having normalised. The output of the computer algorithm was compared with 5 slices per patient which had been visually scored by a radiologist. For these scans, all 23 were marked up by at least one radiologist and 8 scans were marked up by 2 radiologists. Before being presented to the radiologists, all scans were fully anonymised using the GE advantage workstation™ . This viewing and post processing workstation has a feature called 'Anonymous Maker' which enables the removal of all patient identifier and demographic data but allows

metadata concerning the scan parameters (e.g. slice thickness, field of view, voxel dimensions) to be retained. Scans were provided to radiologists for scoring in a random order.

A total of five slices were selected for the radiologists to score, chosen at 5 pre-determined anatomical levels as follows: the top of the arch of the aorta, the carina, the right superior pulmonary vein, 1 cm above the dome of the diaphragm and 2 cm below the dome of the diaphragm. These 5 slices were presented to the radiologists in a separate folder for each patient.

A scoring sheet (Appendix C) was developed based on the work of Edey et al (Edey et al., 2011) and Goh et al (Goh et al., 2008). Radiologists were provided with specific scoring instructions for this purpose (Appendix D) and were blinded to the results of the computerised estimation of fibrosis. To summarise, the scoring process was as follows:

1. For each slice, radiologists were asked to give a visual estimate of the percent of lung involved with honeycombing, reticulation, ground glass opacification and consolidation. Estimates were to be given to the nearest 5%.
2. Radiologists were asked to report whether they thought there was emphysema on each slice, giving a simple yes or no answer.

The results of the radiologists' scoring are presented in Chapter 5.

2.5 Statistical analysis

Correlation between computer calculated lung volume and TLC was assessed by linear regression and was examined for systematic error using Bland-Altman plots (Bland and Altman, 1986) .

Linear regression was used to examine the correlation between the two radiologists, between radiologist and computer, and between computer scores, pulmonary function tests and respiratory questionnaire scores. Spearman's rank-order correlation was

performed to compare the ranking of scan severity by the computer and radiologist. Intraclass correlations were used to compare radiologists' visual scores. All analysis was performed using SPSS statistics software version 22.0.0.0 (SPSS, Chicago, IL). The results of these analyses will be presented in Chapter 5.

CHAPTER 3: RESULTS PART 1: INITIAL TESTING OF THE COMPUTER ALGORITHM

3.1 Overview of the results chapters

This is the first of three results chapters in which we will discuss the initial testing of the computer algorithm including testing on 7 normal (control) scans and a single pulmonary fibrosis scan, comparing the output of the computer algorithm on the fibrosis scan with radiologist manual segmentation.

3.2 Introduction to chapter 3

In this chapter (Chapter 3) we will discuss the outputs of the computer algorithm on 7 scans from patients with no lung disease and on a single patient with pulmonary fibrosis. We will use the normal scans to look specifically at different sources of error and variation in the behaviour of the computer algorithm including artefact at tissue boundaries, movement artefact and gravitational effects. We will also assess the important influence of CT reconstruction algorithm and degree of inspiration/expiration on the computer output. We will then look at the computerised classification of lung texture on a scan from a patient with pulmonary fibrosis, including the Minkowski functional characteristics of different lung textures, and compare the computer classification with radiologist manual segmentation.

3.3 Computer analysis of normal scans

Details of the selection of the normal scans is given in Chapter 2 (Materials and Methods). We will first describe examine how the computer algorithm classified the 7 normal scans when performed in full inspiration and reconstructed using the bone algorithm (the algorithm that the software was trained on), focussing specifically on accuracy of segmentation and classification. Several representative slices of the normal

scans will be presented, along with a table showing the variation in lung texture classification between the scans. We will then look at the effect of CT reconstruction algorithm on 5 of the 7 scans which had been reconstructed using three different algorithms (bone, standard and lung). Finally we will look at the effect of lung volume on two of the normal scans which were performed in both inspiration and expiration.

3.3.1 Computer segmentation and classification of 7 normal inspiratory scans

Normal scan 1:

The overall segmentation for this scan was good but it can be seen that there is mis-classification of some of the most peripheral voxels as honeycombing. This is seen as a universal phenomenon on all scans and is believed to be due to inclusion of some soft tissue of the chest wall in these voxels (partial volume effect). We can also see that there is some erroneous inclusion of the hilar vessels in the segmented lung component.

These two sources of error can be seen in the figure below:

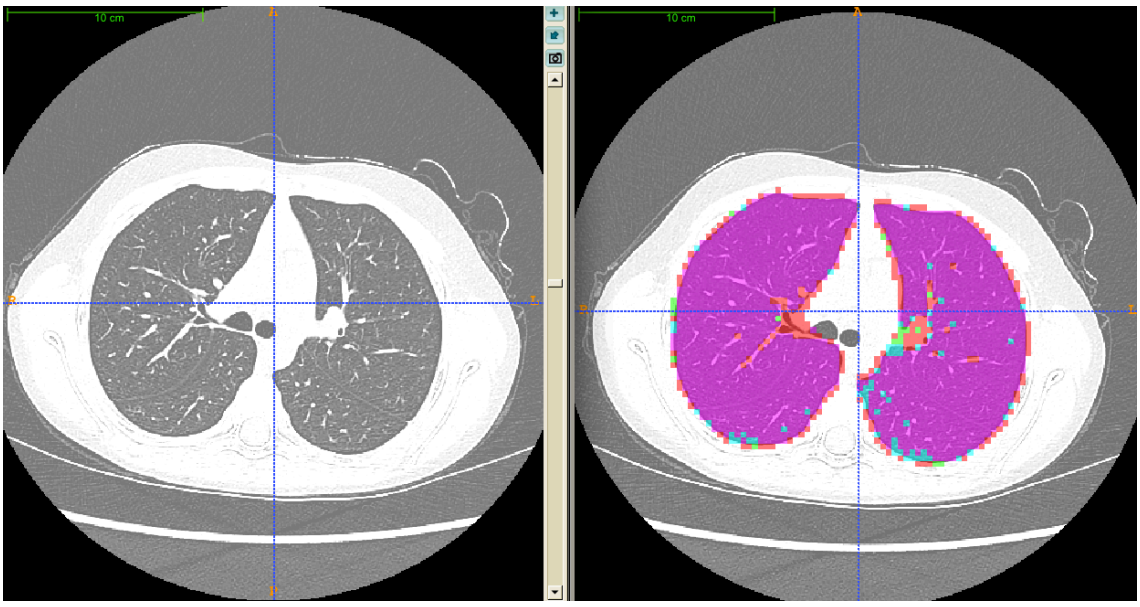

Figure 3-1 Computer output on axial slice of normal scan 1 at the level of the carina

This scan also suffered from some breathing artefact at the lung bases and some cardiac pulsation artefact. The breathing artefact led to some 'ghosting' of the diaphragm onto the lung parenchyma (Figure 3-2).

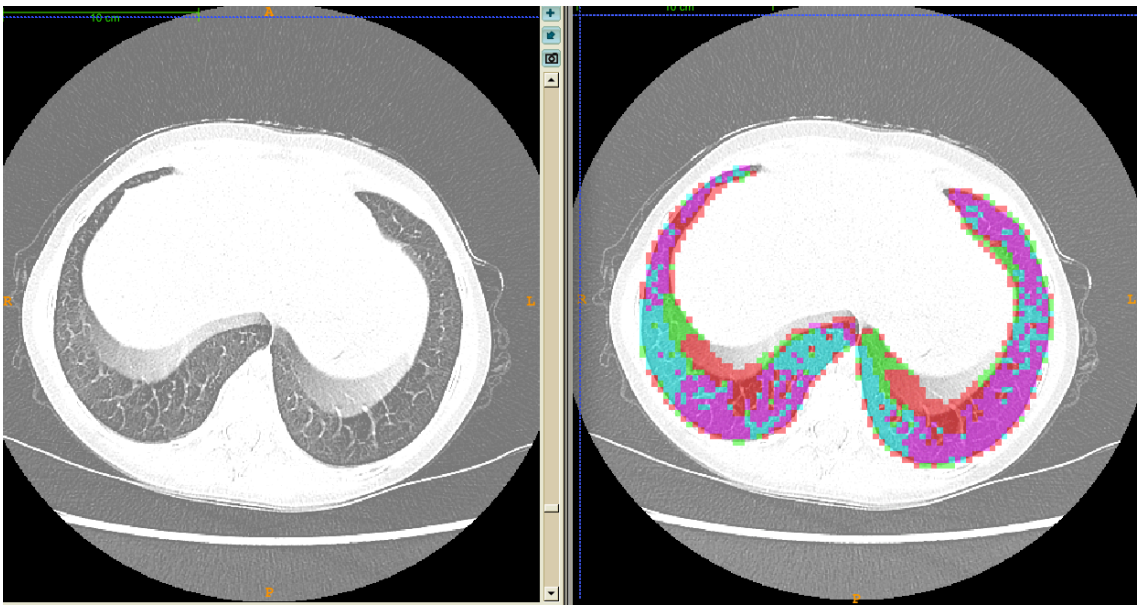

Figure 3-2 Computer output on axial slice of normal scan 1 at the level of the diaphragm

It is evident that this led to both segmentation issues and texture classification issues with some of the ghosted diaphragm included in the segmented lung and some omitted. The lung that was included was classified as either honeycombing or reticulation. Cardiac pulsation artefact was also present on this scan and led to some normal lung being classified as 'indeterminate' (Figure 3-3).

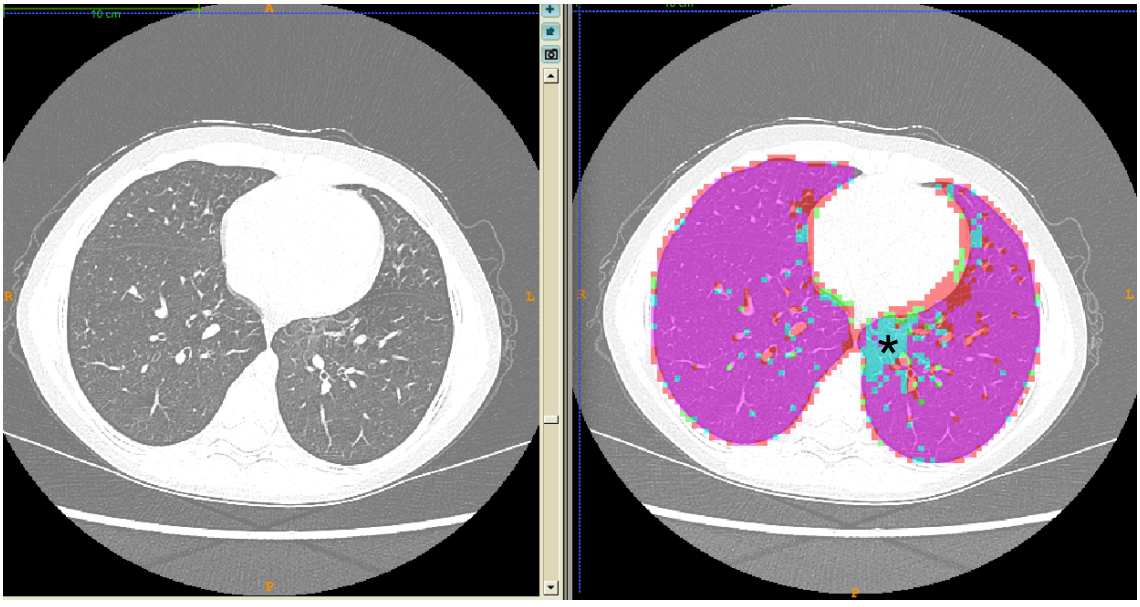

Figure 3-3 Computer output on axial slice of normal scan 1 at the level of the heart. Large area of blue colouration in the left lung is consistent with cardiac pulsation artefact (asterisk)

Normal scan 2:

This scan showed good segmentation but experienced problems with texture classification at the posterior basal parts of the lung (Figure 3-4)

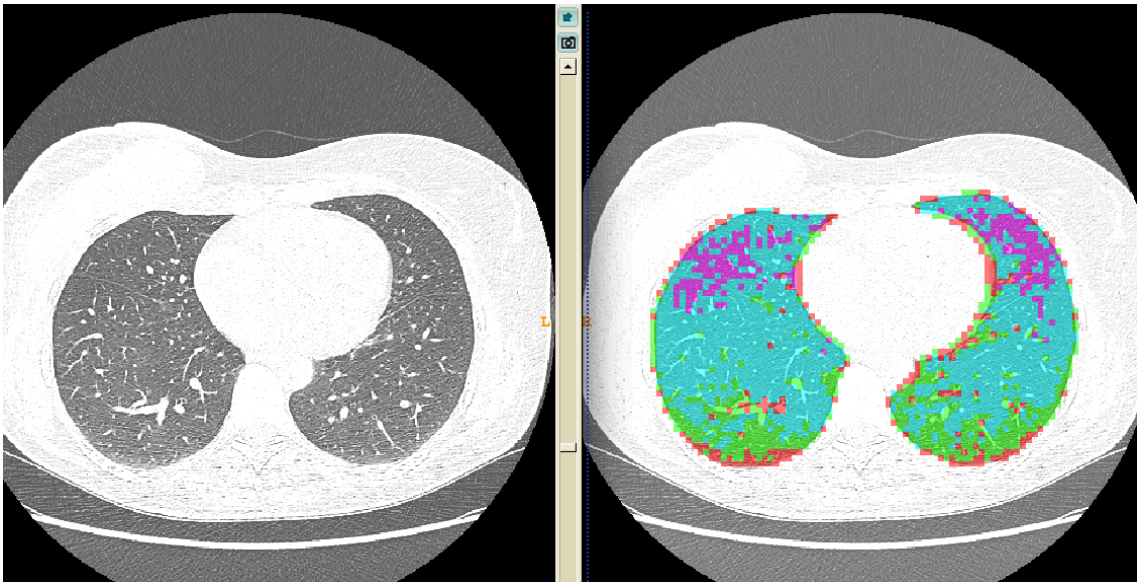

Figure 3-4 Computer output on axial slice of normal scan 2 at the level of the heart. Note the large volume of indeterminate lung (blue) and the erroneous reticulation classification at the lung bases (green).

This was felt to be due mainly due to the degree of noise affecting the scan with a lot of streak artefact posteriorly. This linear streaking was misinterpreted by the computer algorithm as being reticulation and illustrates the fact that it may not be possible to distinguish between different types of linear abnormality. This scan was also rather expiratory and this led to a large amount of the lung being classified as indeterminate. Further analysis of inspiratory/expiratory acquisitions is discussed in Section 3.3.3.

Normal scan 3:

This scan again demonstrates the edge artefact which gives a spurious band of honeycomb classification around the lung edges. It also shows areas of 'indeterminate' classification posteriorly which is thought to be due to a gravitational gradient in lung

density (figure 3-5). This phenomenon has previously been described in studies of CT lung density in normal patients who were imaged supine and prone at different degrees of inspiration (Verschakelen et al., 1993).

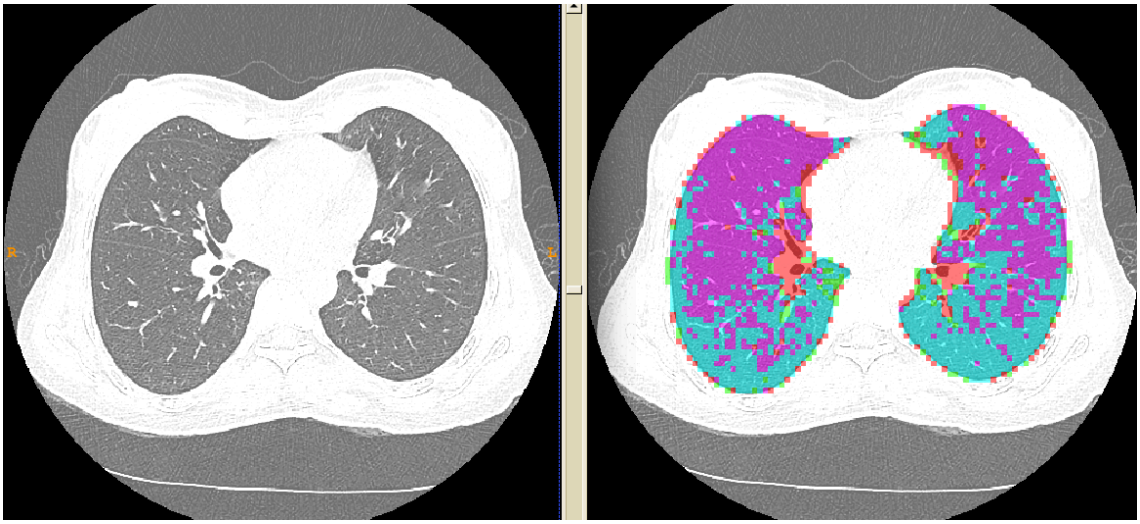

Figure 3-5 Computer output on axial slice of normal scan 3 at the level of the heart. Note the large volume of 'indeterminate' classification which is thought to be due to a gravitational change in lung density

Normal scan 4:

This shows similar findings to scan 3 but with less of the indeterminate classification (see figure below):

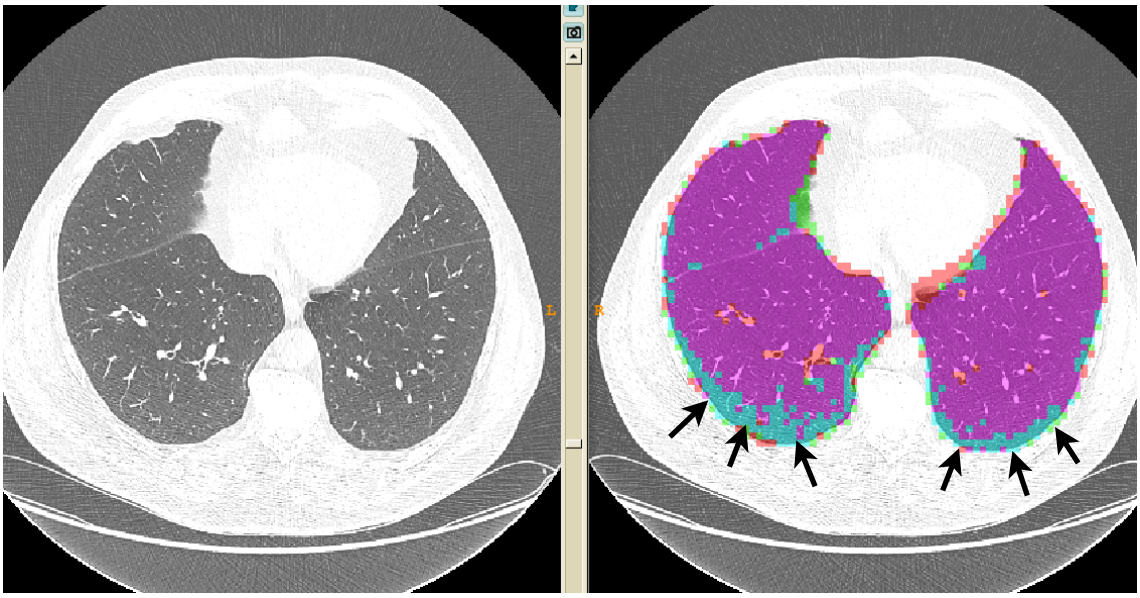

Figure 3-6 Computer output on axial slice of normal scan 4 at the level of the heart showing a small amount of indeterminate classification posteriorly (arrows)

Normal scan 5

The output from scan 5 was similar to that from scan 3, containing a moderate amount of indeterminate classification and some erroneous reticulation due to posterior streak artefact (Figure 3-7).

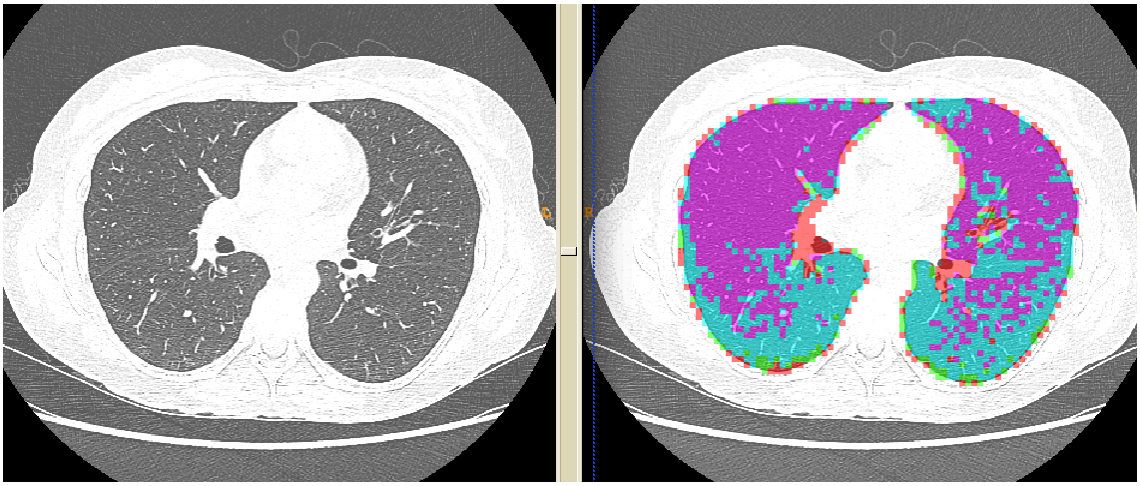

Figure 3-7 Computer output on axial slice of normal scan 5 at the level of the heart

Normal Texture Volumes:

The table below (Table 3-1) summarises the percentage of lung volume classified as each texture in the 7 normal inspiratory scans. This table also shows the percentage of voxels classified as honeycombing which were at the lung edge (surface voxels).

	Total Honeycomb (a)	Honeycomb surface voxels	Reticulation (b)	Total fibrosis (a+b)	Indeterminate (c)	Normal (d)	Non fibrosed (c+d)
Normal 1	11.4	8.4	2.8	14.1	6.7	79.2	85.9
Normal 2	13.1	9.0	15.3	28.4	54.9	16.8	71.6
Normal 3	11.1	8.0	5.5	16.6	45.7	37.7	83.4
Normal 4	9.2	6.8	4.3	13.5	14.0	72.5	86.5
Normal 5	9.0	7.2	4.4	13.4	33.8	52.7	86.6
Normal 6	9.6	7.1	2.3	11.9	5.9	82.2	88.1
Normal 7	11.1	7.6	1.4	12.4	2.9	84.7	87.6
Mean (SD)	10.6 (1.5)	7.7 (0.8)	5.1 (4.7)	15.8 (5.8)	23.4 (21.2)	60.8 (26.0)	84.2 (5.8)

Table 3-1 Classification of lung texture in 7 normal scans (bone algorithm). All figures are given as percentages

From this table it can be seen that there is a percentage of lung which is incorrectly classified as honeycombing on each of the normal scans (false positive honeycombing). This can be explained mainly by segmentation errors including at the interface between lung and chest wall, between lung and mediastinum and at the hila (incorrect inclusion of large airways and vessels in the lung volume). We can see that the percentage of lung classified as honeycombing which was at the edge of the segmentation has a mean volume of 7.7% (SD 0.78). The consistent nature of this error with a small standard deviation suggests that, if it is not correctable on future algorithms, it could effectively be subtracted from the final estimate of honeycombing.

Incorrect classification of normal lung as reticulation was generally less of a problem but also tended to occur at interfaces between lung and soft tissue. An exception was Normal scan 2 where excessive posterior streak artefact led to increased false positive reticulation. As previously described, indeterminate lung was thought to be a reflection of gravity dependent increased density or increased density due to an expiratory scan. This is further explored in Section 3.3.3.

3.3.2 Effect of CT reconstruction algorithm on normal scans

The table below (Table 3-2) illustrates the effect of the CT reconstruction algorithm on texture classification for five of the normal scans for which three separate reconstruction algorithms were available, all with the same slice thickness (lung, standard and bone algorithms).

		Lung	Standard	Bone
Normal 1	Honeycomb (%)	39.1	7.3	11.4
	Reticulation (%)	1.3	6.3	2.7
	Indeterminate (%)	1.8	5.6	6.7
	Normal (%)	57.8	80.8	79.2
Normal 2	Honeycomb (%)	51.3	6.9	13.1
	Reticulation (%)	7.2	17.0	15.3
	Indeterminate (%)	19.7	49.6	54.9
	Normal (%)	21.8	26.5	16.8
Normal 3	Honeycomb (%)	34.2	5.6	11.1
	Reticulation (%)	1.7	10.7	5.5
	Indeterminate (%)	18.0	31.2	45.7
	Normal (%)	46.2	52.6	37.7
Normal 4	Honeycomb (%)	26.1	5.9	9.1
	Reticulation (%)	2.5	7.1	4.3
	Indeterminate (%)	6.2	11.8	14.0
	Normal (%)	65.2	75.2	72.5
Normal 5	Honeycomb (%)	35.7	5.6	9.0
	Reticulation (%)	1.6	8.7	4.4
	Indeterminate (%)	13.2	24.3	33.8
	Normal (%)	49.5	61.4	52.7

Table 3-2 Effect of different reconstruction algorithms on classification of lung texture on normal scans

This table shows that the choice of CT algorithm has a significant effect on the classification of the lung.

Several conclusions can be drawn from the data, as follows:

1) The lung algorithm leads to significantly more lung being wrongly assigned to the honeycombing class. The mean percentage honeycombing for the 7 normal scans with

lung algorithm is 37.3% (SD = 9.2%) compared with 6.3% (SD = 0.8%) for the standard algorithm and 10.6% (SD = 1.7%) for the bone algorithm . The reason for this appears to be two-fold. Firstly, a large proportion of smaller vessels are mis-classified as honeycombing and secondly, there is an exaggeration of the tendency to classify voxels at the costal and mediastinal borders as honeycombing. This effect is demonstrated in figure 3-8 below.

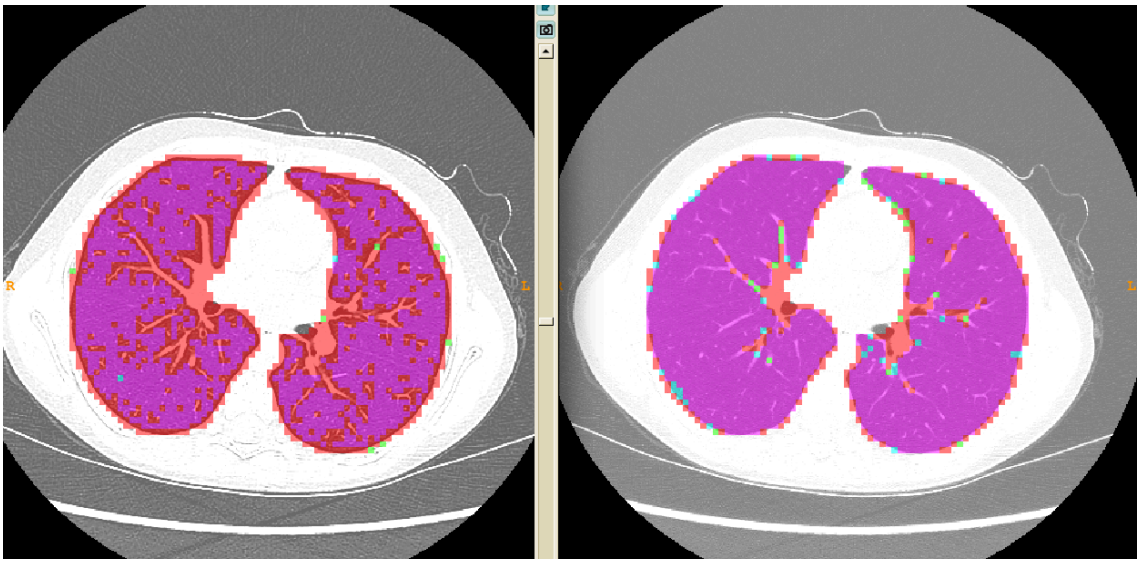

Figure 3-8 Selected axial slice from normal scan 1 showing vessels mis-classified as honeycombing (red) using the lung algorithm (left hand image) compared with the bone algorithm (right hand image)

2) The amount of lung which is mis-classified as reticulation is proportionally greater when using the standard algorithm and the amount of mis-classification as reticulation is least marked when using lung algorithm.

3) The bone algorithm gives results which are somewhere in between lung and standard algorithms.

4) The amount of lung which is classified as indeterminate is largest with the bone algorithm with progressively lower proportions of indeterminate classification on the standard and lung algorithms.

5) The standard algorithm consistently gives the greatest portion of normal classification with bone and standard in second or third place depending on the scan.

3.3.3 Effect of inspiration/expiration on normal scans

Two of the normal scans (normal 6 and 7) had both inspiratory and expiratory images available. These are discussed below:

Normal scan 6 – inspiration

This scan showed overall good segmentation of the inspiratory images (see figure 3-9 and 3-10) with just minor edge artefact and cardiac pulsation artefact.

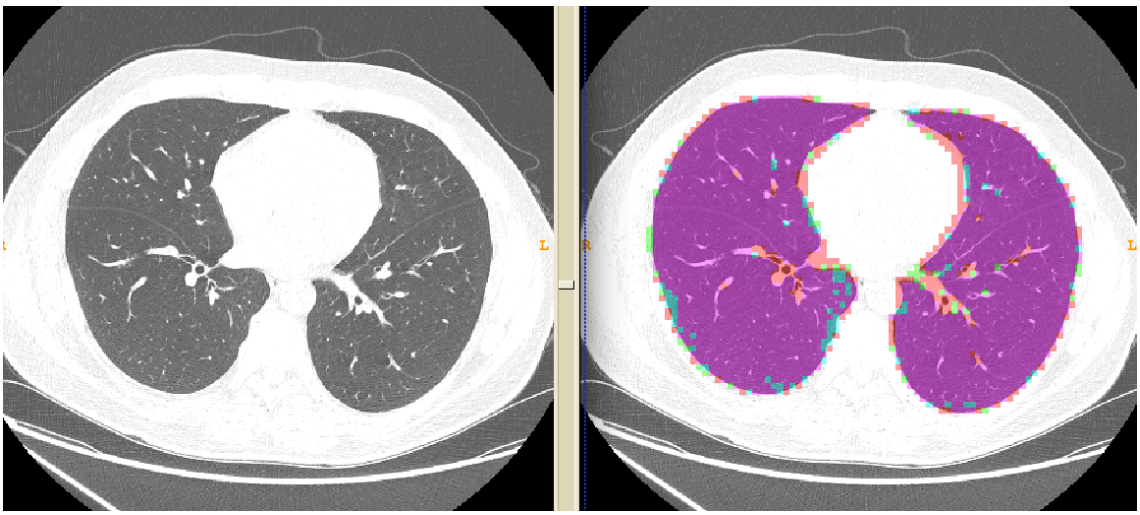

Figure 3-9 Normal scan 6 performed in inspiration (bone algorithm) showing minor edge artefact

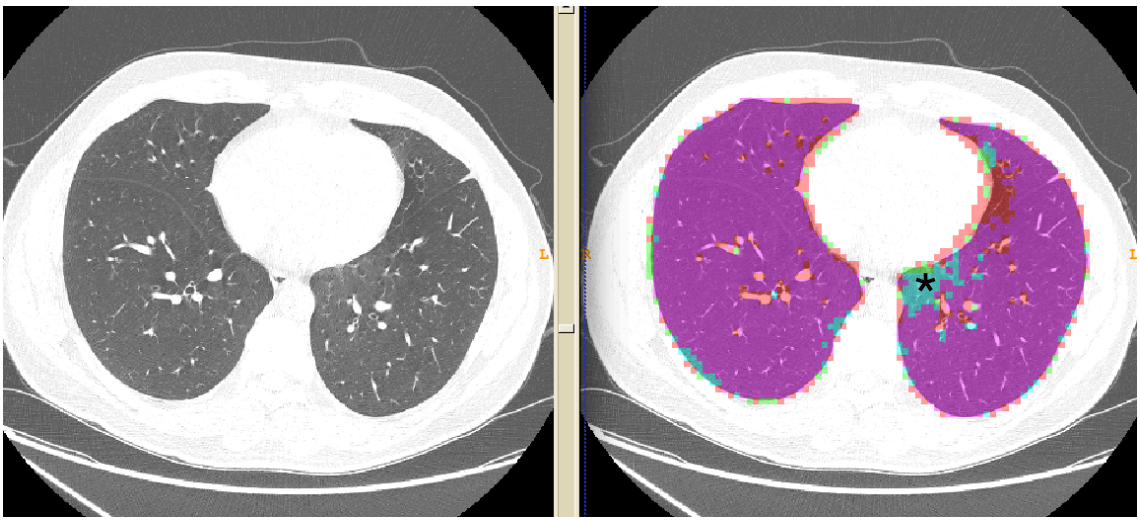

Figure 3-10 Normal scan 6 performed in full inspiration shows minor edge artefact and cardiac pulsation artefact (*)

Normal scan 6 – expiration

Expiration had a significant effect on the classification of the lung producing an increased percentage of lung incorrectly classified as indeterminate or reticulation. Figure 3-11 below shows a similar slice to figure 3-10 (note position of oblique fissures) and demonstrates that although the slice does not look very different visually, there is a dramatic difference in texture classification.

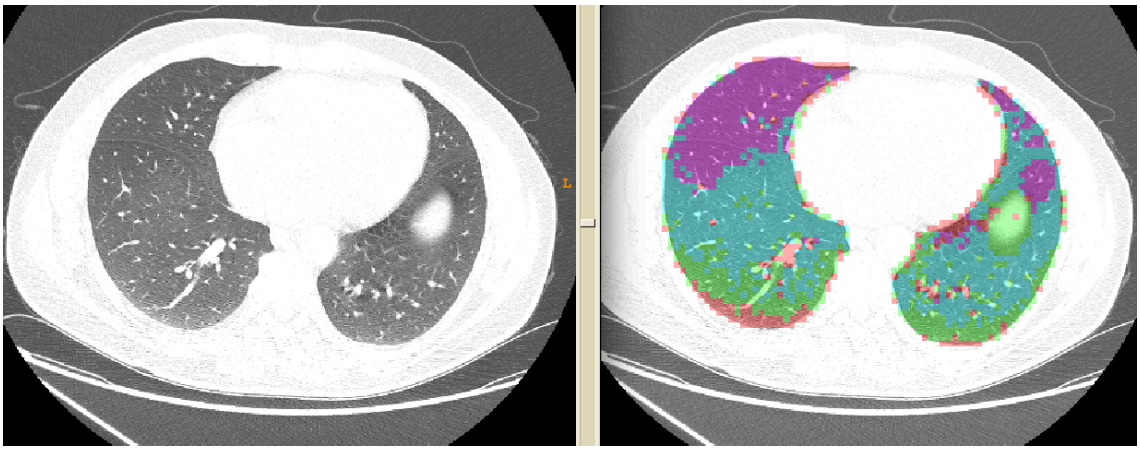

Figure 3-11 Normal scan 6 performed in expiration showing a large amount of indeterminate classification (blue)

Normal scan 7 – inspiration

The inspiratory images of normal scan 7 classified well, apart from some minor mis-classification of posterior lung, thought to be due to posterior linear streak artefact (Figure 3-12). The expiratory images showed a large amount of indeterminate lung as well as a small amount of spurious reticulation. Figure 3-12 and Figure 3-13 show a similar level slice on the same subject performed in inspiration and expiration.

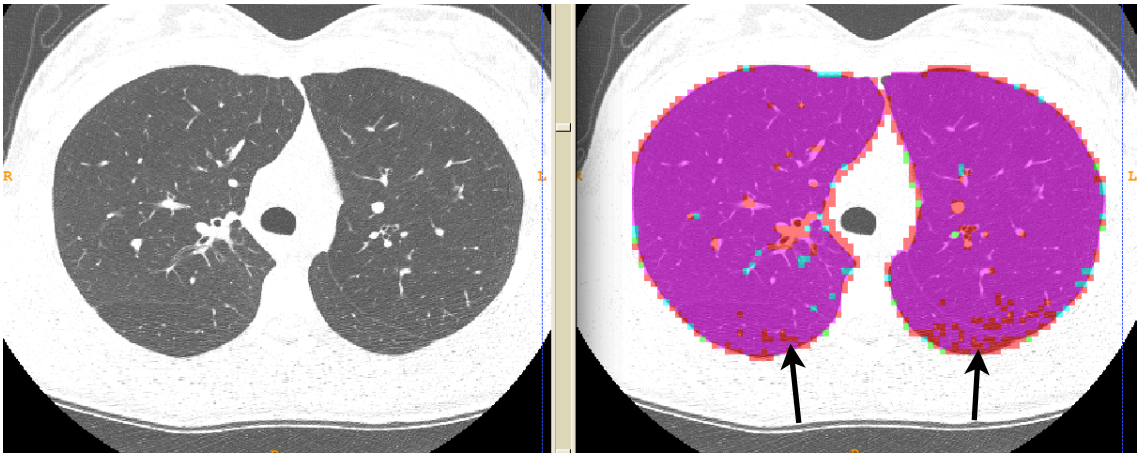

Figure 3-12 Normal scan 7 performed in full inspiration shows minor streak artefact posteriorly which has been mis-classified as honeycombing (arrows)

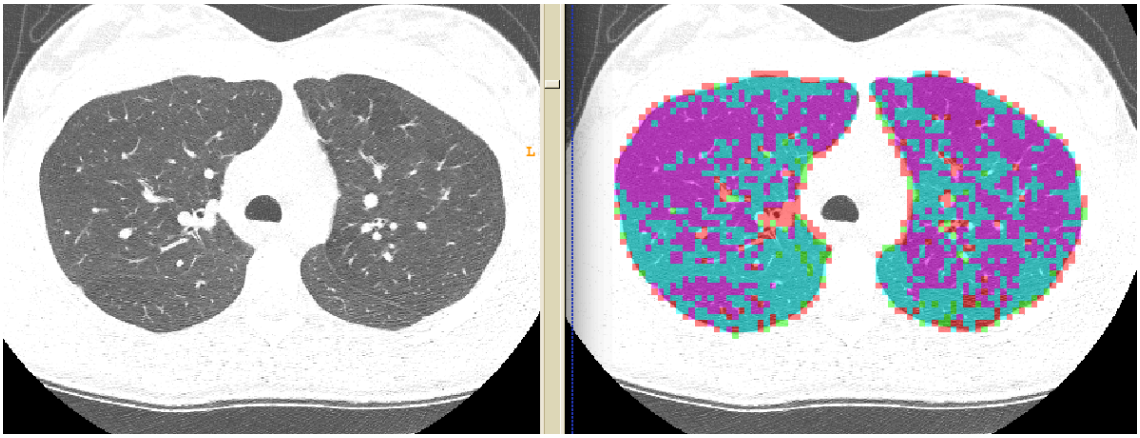

Figure 3-13 Normal scan 7 performed at full expiration shows a large amount of indeterminate lung (blue) as well as a small amount of erroneous reticulation (green)

Interestingly, the inspiratory/expiratory nature of the scan affected mis-classification of different lung textures differently and different algorithms were more sensitive to the degree of inspiration/expiration.

This can be seen in the table below (table 3-3). Note that the figures in brackets are the change in percentage volume on expiration compared with inspiration.

Normal 6	Lung insp	Lung exp	Standard insp	Soft tissue exp	Bone insp	Bone exp
Honeycombing (%)	38.8	42.5 (+ 3.6)	6.0	8.5 (+2.5)	9.6	13.7 (+4.1)
Reticulation (%)	1.2	16.9 (+ 15.7)	4.6	23.7 (+19.1)	2.3	26.9 (+24.6)
Indeterminate (%)	1.1	24.6 (+23.5)	4.7	49.6 (+44.9)	5.9	44.7 (+38.8)
Normal (%)	58.9	16.0 (−42.8)	84.6	18.2 (−66.4)	82.2	14.8 (−67.4)
Normal 7	Lung insp	Lung exp	Standard insp	Soft tissue exp	Bone insp	Bone exp
Honeycombing (%)	39.6	35.5 (−3.6)	5.3	5.1 (−0.2)	11.0	11.6 (+0.5)
Reticulation (%)	0.5	2.7 (+2.3)	4.5	11.0 (+6.6)	1.4	7.1 (+5.7)
Indeterminate (%)	0.3	23.5 (+23.1)	4.7	38.1 (+33.4)	2.9	48.3 (+45.4)
Normal (%)	60.1	38.3 (−21.8)	84.6	45.8 (−38.9)	84.7	33.0 (−51.7)

Table 3–3 Effect of reconstruction algorithm and inspiration/expiration on normal scans 6 and 7. Figures in brackets indicate change in percentage lung volume between inspiration and expiration

For both scans, the class which was least affected by inspiration/expiration was honeycombing, with only a small change in the percentage of lung classified as honeycombing across all the reconstruction algorithms. The reticulation class was more sensitive to expiration/inspiration with the affect again seen across all reconstruction algorithms. The most sensitive class was indeterminate with a large increase seen on the expiratory scans. These effects are reflected in the amount of lung classified as normal for each reconstruction algorithm. Both standard and bone algorithms show over 82% of lung as normal on inspiratory scans whereas this drops to between 14 and 46% for expiratory scans.

3.4 Testing the computer algorithm against an abnormal CT scan marked up by a radiologist

We will now present the results of testing the algorithm on the initial scan (patient 13) which was marked up by the radiologist for the purpose of training the algorithm. Although the radiologist had marked up only the right lung, the computer algorithm was used to analyse both the left and right lungs. A sample slice demonstrating the radiologist's original mark-up and the computer generated mark-up side by side is show in figure 3-14.

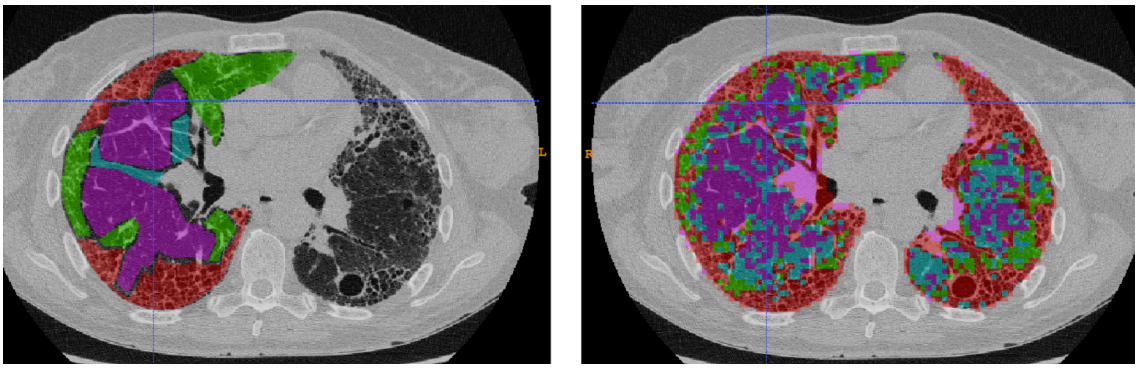

Figure 3-14 Radiologist training segmentation (left) compared with automated computerised segmentation (right) on a single identical slice. Courtesy of Dr A Bhalerao

This shows that there is a similar spatial distribution of texture classification for the computer and the radiologist but the computer algorithm shows more variability over a small area, consistent with it using smaller regions of interest (5 x 5 x 5 pixels) compared with the radiologist's larger and variable sized (freehand) ROIs. The previously described edge artefact is again seen around the mediastinal vessels.

Further comparison of the radiologist mark-up can be made by looking at principle components analysis (PCA) of the radiologist mark-up and computer classification. Principle components analysis provides a visual demonstration of the Minkowski functional features which best distinguish between the different textural classifications

on a VOI by VOI basis. Figure 3-15 shows the PCA analysis for the radiologist (left hand image) and computer (right hand image) for slice 266 (the same slice as in Figure 3-14).

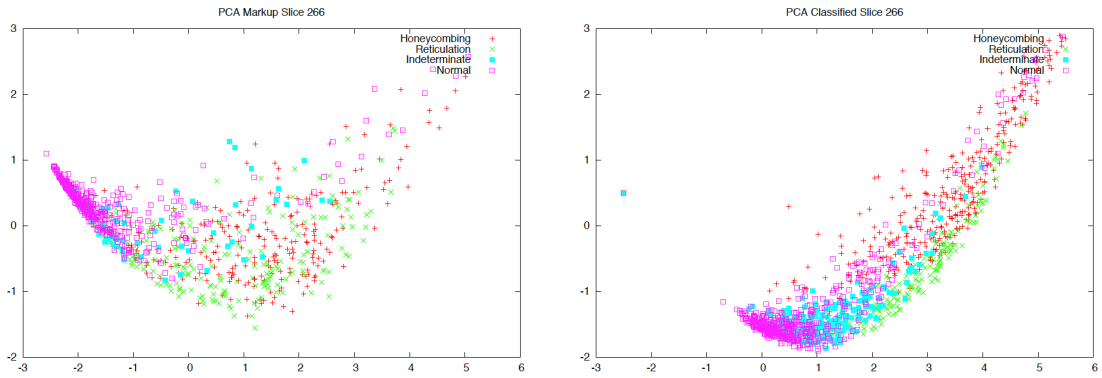

Figure 3-15 PCA analysis of the different textural classes in slice 266 of patient 13. The left hand image is from the radiologist mark up and the right hand image is from the computer mark up

From Figure 3-15 we can see the radiologist classification shows more overlap of the reticulation, honeycombing and indeterminate classes compared with the computer. The normal mark-up appears to be the most robust feature, with tight clustering of the normal VOIs at the left hand side of the graph. In comparison, the computer classification shows tighter clustering of the classes and less overlapping of the honeycombing and reticulation classes. It does show slightly more overlapping of the normal and honeycombed lung compared with the radiologist but this may be due to the edge segmentation artefact described earlier where normal lung at the edge of a slice may be mis-classified as honeycombing. There is also some uncertainty in the indeterminate class although this is less marked than for the radiologist mark-up.

Another way of comparing the radiologist and computer classification is to compare the classification of all 5 x 5 x 5 VOIs in the right lung of patient 13. This data is shown in the following 2 x 2 tables (table 3-4 and table 3-5) where 'true class' represents the radiologist classification, training set P represents the 3000 original training VOIs per classification (total 12,000) and the testing set Q represents the remaining 26,919 ROIs which were not used for training. All figures are given as percentages.

		Computer classification			
	Class	Honeycombing	Reticulation	Indeterminate	Normal
Radiologist Classification	Honeycombing	81.2	8.5	5.5	4.8
	Reticulation	9.3	44.9	16.3	29.6
	Indeterminate	2.7	6.5	42.5	48.3
	Normal	0.9	2.9	9.4	86.8

Table 3-4 Classification of the computer output compared with the radiologist classification on the training set (P)

		Computer classification			
	Class	Honeycombing	Reticulation	Indeterminate	Normal
Radiologist Classification	Honeycombing	65.2	18.4	8.5	7.9
	Reticulation	17.4	27.7	23.2	31.7
	Indeterminate	4.0	10.9	26.6	58.5
	Normal	2.2	4.5	18.3	75.0

Table 3-5 Classification of the computer output compared with the radiologist classification on the testing set (Q)

We can see that for the training set (P), there is an 81.2 % sensitivity for honeycombing, a 44.9% sensitivity for reticulation, a 42.5% sensitivity for indeterminate lung and an 86.8% sensitivity for normal lung. As would be expected, the performance for the testing set (Q) is less good with 65.2% sensitivity for honeycombing, 27.7% sensitivity for reticulation, 26.6% sensitivity for indeterminate and 75.0 % sensitivity for normal lung. For both data sets, it is evident that honeycombing and normal lung show best agreement between radiologist and computer.

3.4.1 Minkowski functional output

The figure below illustrates the Minkowski functional outputs from the whole of the right lung of subject 13 across the complete range of Hounsfield Unit thresholds. It

shows that normal and indeterminate categories do not show much separation in any of the MFs. Honeycombing is well separated across all the MF distributions but this separation varies according to the threshold.

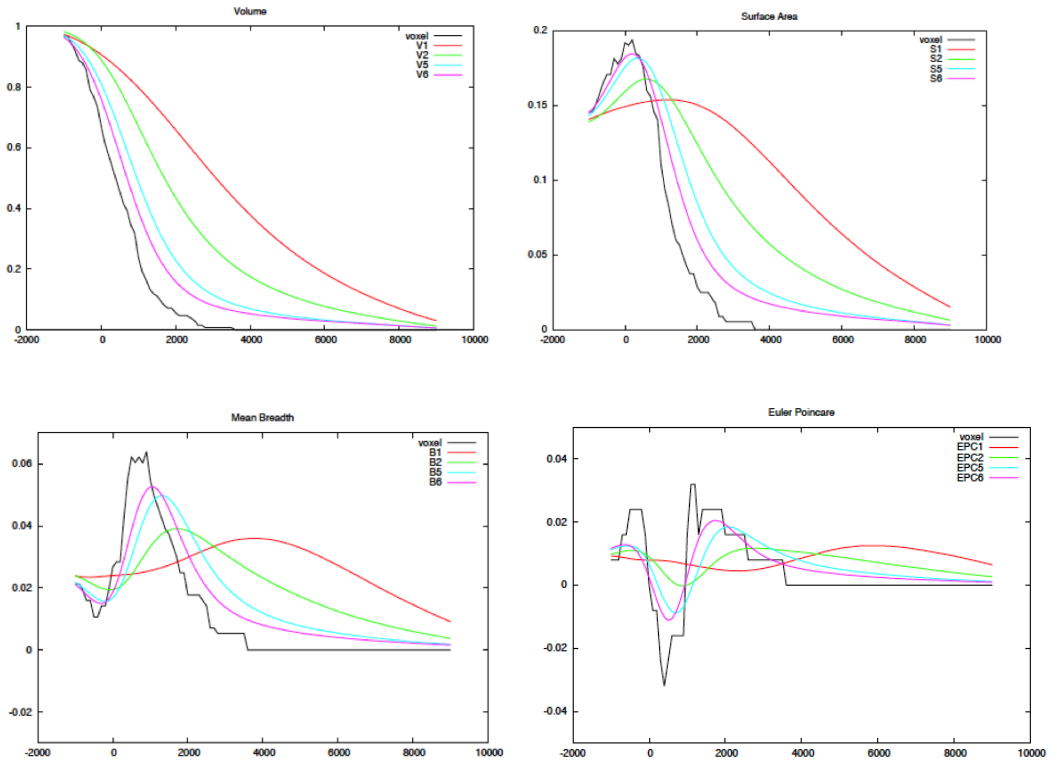

Figure 3-16 Minkowski functional distributions from the right lung of subject 13. Coloured curves represent the MF distributions for each texture classification averaged over all VOIs in the right lung. Red = honeycombing, green = reticulation, blue = indeterminate, purple = normal. The black line represents the MF distribution for a single voxel of normal lung

3.5 Conclusion

In this chapter, we have discussed the output of the computer algorithm both in qualitative and quantitative terms. We have looked at some of the sources of variation in the classification of lung texture by the computer algorithm including artefacts at the interface of different tissues, artefacts due to movement and artefacts due to gravitational effects. We have also looked at the influence of reconstruction algorithm and depth of breathing on classification. Specifically, we have looked at outputs produced when the algorithm is tested on scans from patients with normal lungs and a single scan from a patient with pulmonary fibrosis. We have looked at typical Minkowski

functional outputs from a whole scan and have used principle components analysis to study the power of discrimination of Minkowski functionals on a single slice. Classification of the pulmonary fibrosis scan has been compared with radiologist classification over a total of 26,919 VOIs.

CHAPTER 4 – RESULTS PART 2: TESTING OF THE COMPUTER ALGORITHM ON PROSPECTIVELY RECRUITED PATIENTS

4.1 Introduction to Chapter 4

In this chapter we will look at the characteristics of the 24 prospectively recruited patients with CFIP including demographics, pulmonary function tests, respiratory questionnaire scores and computerised analysis of the CT scans.

4.2 Demographics

4.2.1 Age

The subjects ages ranged from 62 years to 84 years. The mean age at enrollment to the study was 73 years and 8 months and the median age was 74 years 6 months.

4.2.2 Gender

The majority of patients were male (18 male patients and 6 female patients).

4.2.3 Smoking status

The majority of patients were ex-smokers with only 8 of 24 patients classifying themselves as 'never smokers'. The mean exposure to cigarette smoke was 20.4 pack years (SD = 15.4 years). At the time of enrollment to the study none of the patients were current smokers or had smoked in the month before enrollment.

4.2.4 Sub-type of IIP

Five patients had undergone a lung biopsy for diagnosis. In one patient this showed desquamating interstitial pneumonia (DIP). For this patient, although a previous clinical scan had showed ground glass shadowing, the first study scan was completely normal and therefore this patient was excluded from further analysis. Biopsy showed a UIP

pattern in two patients and fibrotic NSIP in two patients. For the remaining patients without a lung biopsy, the conclusion of the regional interstitial lung MDT was used as the working diagnosis. Of the 19 patients who did not have a biopsy, eleven showed a typical UIP pattern on CT (according to the criteria in table 1-2); two were felt to have possible UIP; five were felt to have findings most compatible with NSIP and one was felt to have features of both NSIP and organising pneumonia.

4.3 Pulmonary Function tests

Two patients were unable to perform baseline pulmonary function testing: one patient was too breathless and one patient found it too unpleasant an experience. This left 21 patients with full pulmonary function and CT scan for analysis.

4.3.1 Spirometry

4.3.1.1 Forced expiratory value in 1 second (FEV_1)

Baseline FEV_1 ranged from 1.73 L to 2.55 L with a mean of 2.13 L and a standard deviation of 0.27 L. The percent predicted FEV_1 ranged from 70.8% to 107.8% with a mean of 85.6 % and a standard deviation of 11.2 %

4.3.1.2 Forced vital capacity (FVC)

Baseline FVC ranged from 1.97 L to 3.82 L with a mean of 2.71 L and a standard deviation of 0.40 L. The percent predicted FVC ranged from 62.3% to 103.2% (mean 83.4%, standard deviation 11.0%).

4.3.2 Gas transfer (DL_{CO})

The values for DL_{CO} at baseline ranged from 1.83 $mm.min^{-1}.kPa^{-1}$ to 6.51 $mm.min^{-1}.kPa^{-1}$ (mean = 3.58, SD = 1.14). The percent predicted DL_{CO} ranged from 23.8% (absolute

value = 1.83) to 85.1% (absolute value = 4.74) with a mean of 48.1% and a standard deviation of 16.6%.

4.4 St George's Respiratory Questionnaire (STGRQ) Scores

As previously discussed in Chapter 2, the highest possible total STGRQ score is 100 and the highest score for each of the domains (symptoms, activities or impacts) is also 100 with a higher score indicating a greater degree of limitation.

Total scores in the STGRQ ranged from 10 to 98 (mean = 47, SD = 23). Scores in the symptoms domain ranged from 11 to 95 (mean = 49, SD = 25). Scores in the impacts domain ranged from 4 to 100 (mean = 37, SD = 25) and scores in the activities domain ranged from 6 to 100 (mean = 59, SD = 28). This indicates a large spread in the distribution of STGRQ scores across all the domains and a wide range in the severity of patients symptoms, physical limitations and impairment in quality of life.

4.5 Computerised classification of CT scans

The table below (table 4-1) shows the complete computerised classification data for the 23 patients including the CT calculated total lung volume and the CT calculated volume of honeycombing, reticulation, indeterminate and normal lung, both in millilitres and in percentage of the total lung volume.

Study ID	CT total lung volume (ml)	CT honeycombing (ml)	CT honeycombing (%)	CT reticulation (ml)	CT reticulation (%)	CT indeterminate (ml)	CT indeterminate (%)	CT normal (ml)	CT normal (%)
1	6400	663	10	390	6	685	11	4663	73
2	4696	708	15	410	9	1013	22	2565	55
3	4620	456	10	507	11	1339	29	2318	50
4	4073	785	19	800	20	1788	44	700	17
5	4259	1346	32	429	10	975	23	1509	35
6	2037	901	44	507	25	544	27	85	4
7	4975	1130	23	412	8	752	15	2681	54
8	4637	1286	28	288	6	532	11	2531	55
9	3565	665	19	1195	34	1492	42	213	6
10	5356	2116	40	615	11	1122	21	1502	28
11	3499	720	21	615	18	862	25	1203	34
12	4272	502	12	990	23	1080	25	724	17
13	4015	1673	42	734	18	1084	27	525	13
14	4760	509	11	714	15	1644	35	1803	38
15	3680	489	13	684	19	510	14	1425	39
16	3730	1452	39	427	11	882	24	969	26
17	5240	669	13	392	7	1239	24	2941	56
18	4138	630	15	1179	28	1876	45	190	5
20	5903	3356	57	128	2	1098	19	3383	57
21	4003	1253	31	627	16	627	16	1829	46
22	4509	1017	23	444	10	745	17	2303	51
23	4117	1314	32	327	8	513	12	1963	48
24	4487	572	13	330	7	687	15	2898	65
Mean (SD)	4390 (884)	1088 (432)	26 (11)	519 (224)	13 (7)	1004 (408)	24 (10)	1779 (1149)	38 (20)

Table 4-1 Total lung volumes and computer classification of the 23 prospectively recruited patients (note that there is no patient 19 as this subject was excluded due to a normal CT)

From this table we can see that the CT calculated lung volume ranged from 2037 ml to 6400 ml (mean 4390 ml, SD 884 ml).

The CT calculated volume of honeycombing ranged from 456 ml to 3356 ml (mean = 1088 ml, SD = 432 ml) with the percentage of honeycombing ranging from 10 % to 57 % (mean = 26%, SD = 11%).

The estimated volume of reticulation ranged from 128 ml to 1195 ml (mean = 519 ml, SD = 224 ml) with the percentage of reticulation ranging from 2% to 34 % (mean = 13%, SD = 7%).

The estimated volume of indeterminate lung ranged from 532 ml to 1788 ml (mean = 1004 ml, SD = 408 ml) with the percentage of indeterminate lung ranging from 11 % to 45 % (mean = 24%, SD = 10%).

The estimated volume of normal lung ranged from 85 ml to 4663 ml (mean =1779 ml, SD = 1149 ml) with the percentage of normal lung ranging from 4% to 73% (mean = 38%, SD = 20%).

Examples of the automatically classified scans are given below, showing examples of a patient with mild disease, a patient with moderately severe disease and a patient with advanced disease (Figures 4-1 to 4-3).

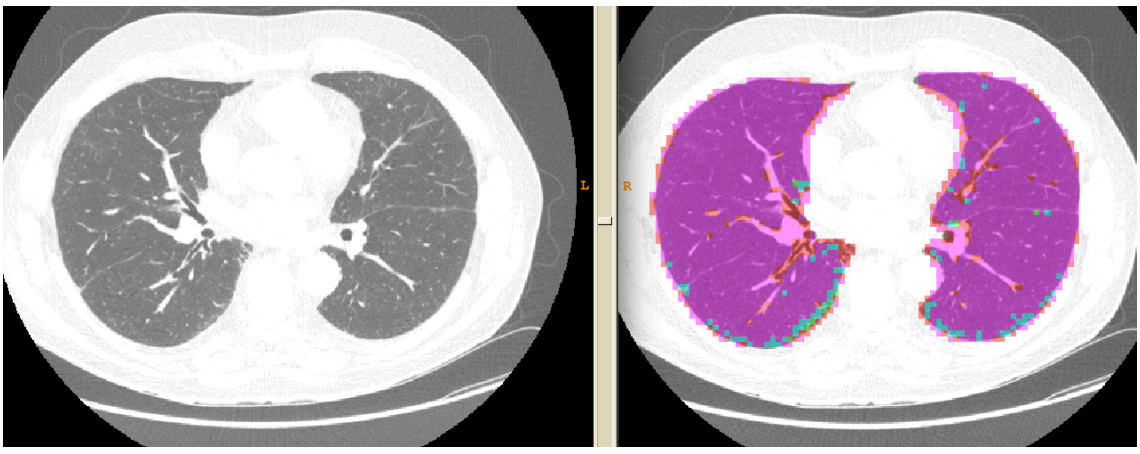

Figure 4-1 An example of an axial slice from the least severely affected patient showing that the computer has correctly identified a tiny area of honeycombing in the para-vertebral region of the right lung (red) but has also erroneously classified central vessels as honeycombing

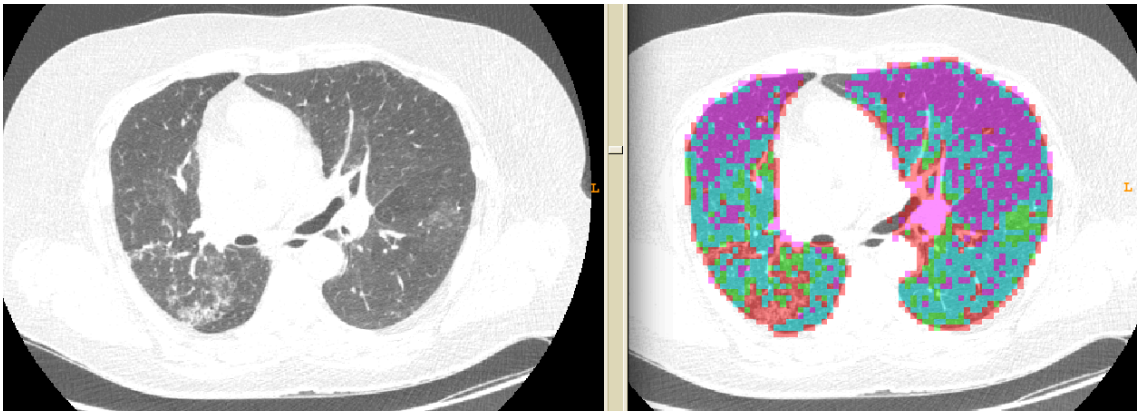

Figure 4-2 An example of an axial slice from a moderately severely affected patient showing an area of established honeycombing in the posterior segment right upper lobe adjacent to the oblique fissure (red) and areas of subtle reticulation (green). The large amount of indeterminate classification (blue) is thought to be be due to the scan being relatively expiratory

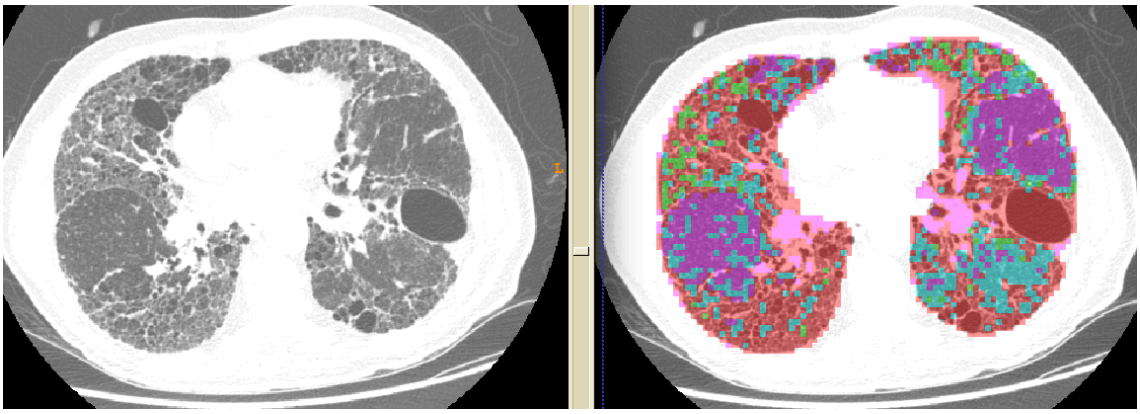

Figure 4-3 An example of an axial slice from a patient with advanced disease showing extensive honeycombing (red) admixed with reticulation (green) with only small amounts of normal lung remaining

4.6 Radiologist visual scoring of CT scans

The range of radiologists scores will be discussed in this section and comparisons between radiologists' scoring and the computer fibrosis score will be presented in the next chapter. Radiologist 1 (EH) scored all the scans, scoring 5 axial slices per patient (total of 115 slices). The range of honeycombing scores on a single slice for radiologist 1 was between 0% and 85%. One patient was scored as having no honeycombing on any slices whilst the patient scored as having most honeycombing had percentage involvement of 15%, 50%, 50%, 70% and 85% on their five slices (cranial to caudal). The range of reticulation scores on a single slice was between 0% and 45%. The most mildly

affected patient had a reticulation score of 5% on each slice and the most severely affected patient had scores of 35%, 20%, 15%, 10% and 60% (cranial to caudal). Radiologist 2 (RB) scored a subset of 40 slices and had a range of 0% to 60% for honeycombing and 5% to 50% for reticulation. Figure 4.4 shows an example slice which has was scored by the two radiologists. Radiologist 1 scored it as having 20% honeycombing whilst Radiologist 2 scored it as having 10% honeycombing. For the same slice, reticulation score was 10% for Radiologist 1 and 25% for Radiologist 2. This illustrates how it may be difficult for the radiologists to decide whether an abnormal pattern should be classed as honeycombing or reticulation.

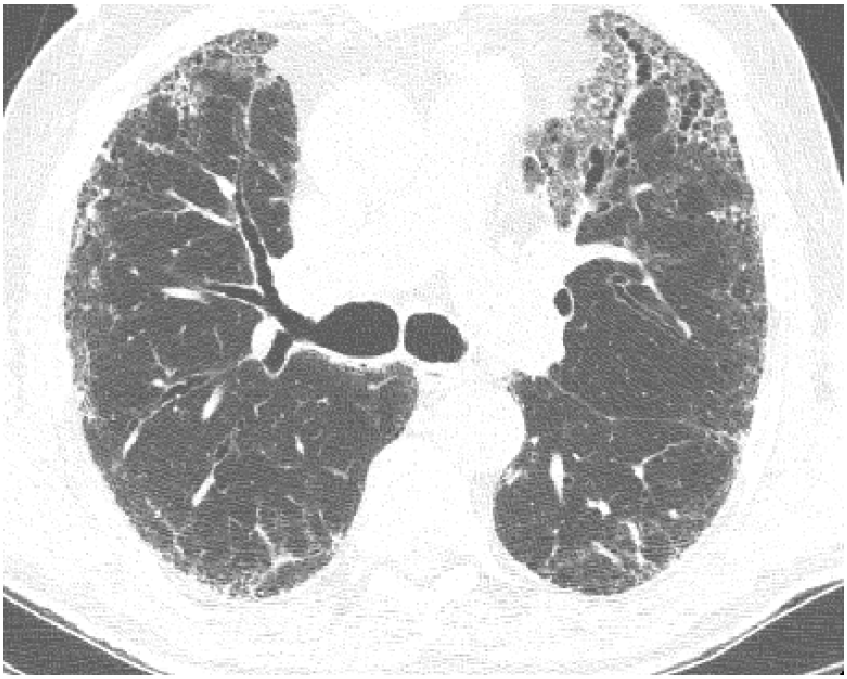

Figure 4-4 An example axial slice at the level of the carina where Radiologist 1 and Radiologist 2 scored the amount of honeycombing at 20% and 10% respectively and the amount of reticulation at 10% and 25% respectively

4.6 Conclusion

In summary, our group of prospectively recruited patients were largely male, ex-smokers with an average age of 74, most of whom had a diagnosis of UIP or NSIP. They had relatively mild impairment of their FEV1 (mean = 85.6 % predicted) and relatively

mild impairment of FVC (mean = 83.4% predicted). Patients generally had more severe impairment of DL_{CO} than FVC with a mean value of 48% predicted. The DL_{CO} also showed a much larger range of values than either FEV1 or FVC. Scores in the St George's Respiratory Questionnaire showed a wide variation between patients, both in terms of the total score and the individual domains. Equally, both the computerised and radiologist assessment of abnormal lung textures suggest a wide range in disease severity between patients.

CHAPTER 5 – RESULTS PART 3: RELATIONSHIPS BETWEEN THE MEASURED VARIABLES

5.1 INTRODUCTION

In this chapter we will explore the relationships between the computer fibrosis score (CFS), the pulmonary function tests (TLC, FVC and DL_{CO}) and the St George's respiratory questionnaire. We will also compare the computer fibrosis score and the radiologist fibrosis score (RFS) and the inter-observer variability between two radiologists. Each set of results will be followed by a brief discussion of the meaning of the results with a full discussion presented in Chapter 6.

5.2 COMPARISON OF COMPUTER ESTIMATED LUNG VOLUME AND TLC

The computer calculated total segmented lung volume was compared with the total lung capacity as measured on pulmonary function tests. The graph below (Figure 5-1) shows the correlation between the two measurements. The data for this graph excludes the two patients who were unable to perform lung function tests.

We can see that there is a very good correlation between the two measurements with an R-value of 0.85 ($p<0.001$).

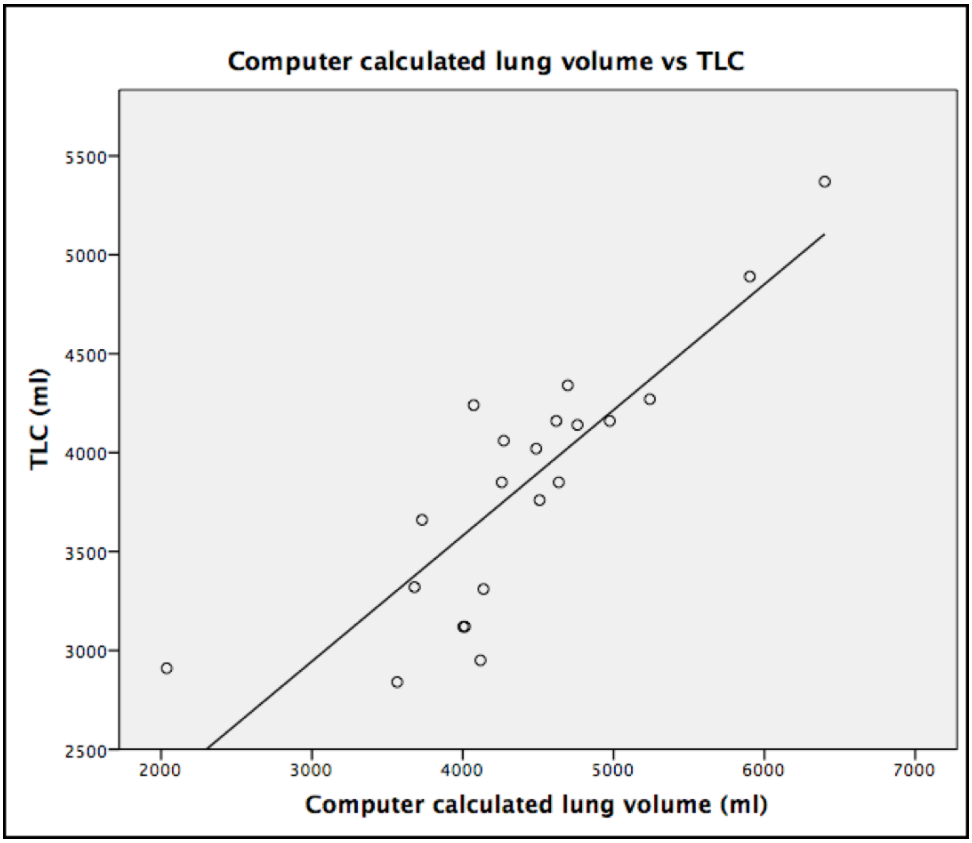

Figure 5-1 Linear regression showing correlation between computer calculated lung volume and TLC

A Bland-Altman plot was created to examine whether there was any systematic difference between the computer calculated lung volume and the TLC (Figure 5-2). Please note that the data are the same as in Figure 5-1.

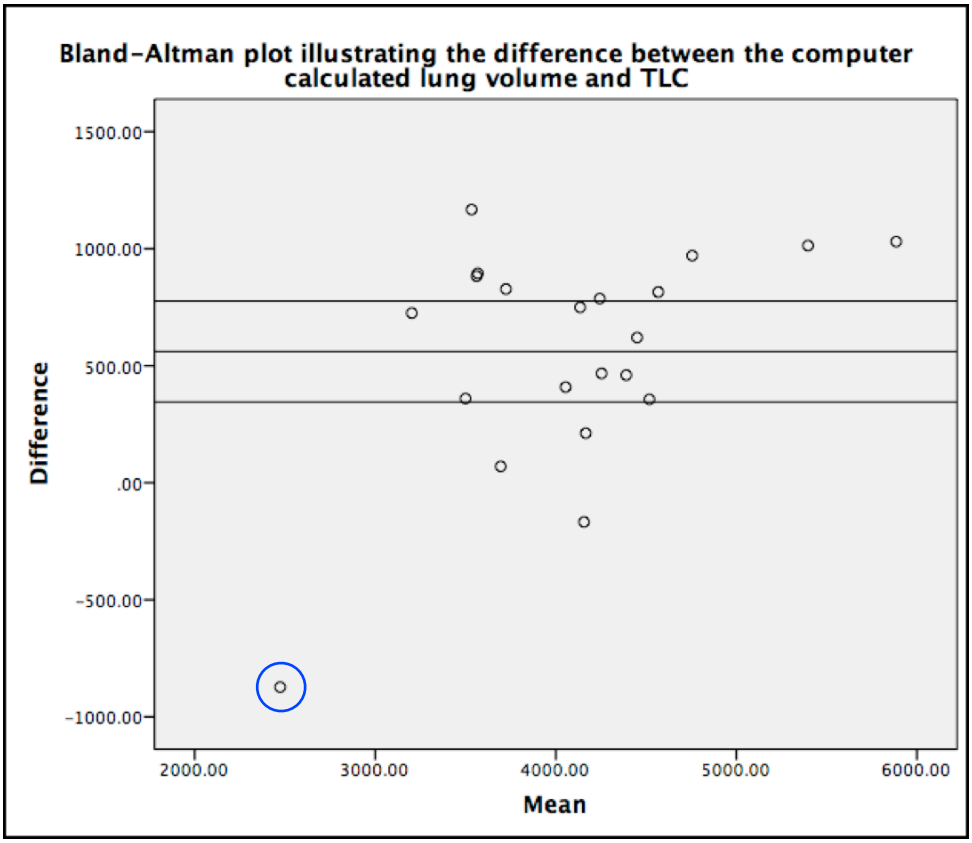

Figure 5-2 Bland-Altman plot illustrating the difference between the computer calculated lung volumes and TLC

This plot illustrates the fact that the mean difference between the computer estimated lung volume and the TLC is + 561 ml i.e. the computer tends to estimate the lung volume as larger than the TLC measurement. The confidence interval for the mean is +345 ml to +777 ml. The most likely explanation for the discrepancy is that the computer calculated CT lung volume includes the pulmonary interstitium and blood vessels which typically make up approximately 10% of the pulmonary parenchyma (Cressoni et al., 2013, van Rikxoort and van Ginneken, 2013). Therefore it would be expected that the CT calculated volume is approximately 10% larger than the TLC. There are several other variables that may contribute to the difference between these measurements including: areas of honeycombing which do not ventilate will reduce the TLC; areas of air-trapping will reduce the TLC; depth of inspiration will affect the CT calculated lung volume; errors in segmentation will

affect the CT calculated lung volume. There is an outlier in the bottom left hand corner of the plot (circled in blue). This was a patient with TLC of 2910 ml and CT calculated lung volume of 2037 ml and is point situated furthest to the left on Figure 5-1. The discrepancy is thought to be due to the fact that the patient's CT was quite expiratory, reducing the apparent lung volume on CT.

5.3 The relationship between computer estimated CT fibrosis and DL_{CO}

Linear regression was used to compare the CT calculated fibrosis score (honeycombing plus reticulation) and the percentage predicted DL_{CO}. The patients are the same as in Section 5.2. The graph below (Figure 5-3) illustrates that there is a moderately strong negative correlation between the two variables with an R-value of -0.65 (p=0.001).

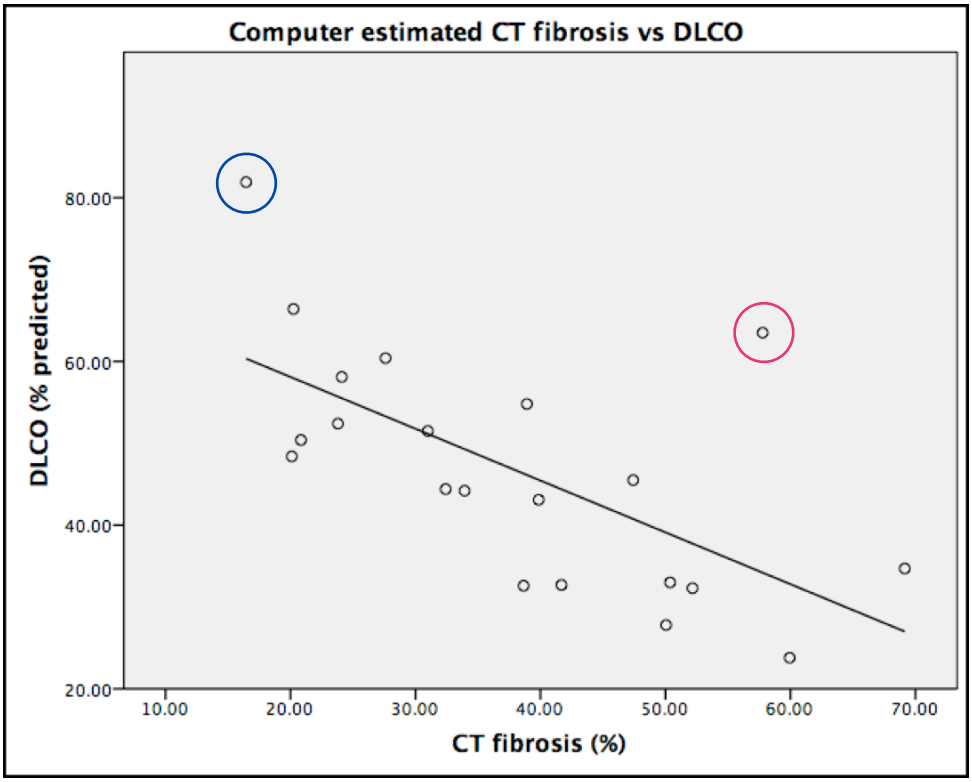

Figure 5-3 Linear regression showing the correlation between computer calculated percentage fibrosis and DL_{CO}. Outliers are circled.

We can see that there are two outliers who have a higher percent-predicted DL_{CO} than would be expected from the percentage of fibrosis measured on CT. The first outlier (Patient 1 - circled in blue) is the subject with both the lowest percentage of fibrosis on CT and the highest DL_{CO}. Since we have a lack of any other data-points with DL_{CO} more than 66%, there is more uncertainty about the shape of the graph near the beginning of the x-axis. It may be that the best fit-line for the plot is actually a logarithmic curve and patients with lower amounts of fibrosis on CT have proportionally greater values of DL_{CO} than patients with higher amounts of fibrosis. Further investigation of this relationship would require confirmation by testing in a larger patient group with more mildly affected patients. Alternatively, either the CT fibrosis or the DL_{CO} may have been under or over-estimated due to error in either measurement. Visual inspection of the classification of this patient shows that classification is good but due to this patient having very little honeycombing, the erroneous classification of edge voxels as honeycombing will contribute a significant proportion of this patients total fibrosis and may partly explain the higher DL_{CO} than would be expected for the estimated amount of fibrosis. The second outlier (Patient 12 - outlined in red) had a large amount of lung which was erroneously classified as honeycombing on CT. This is thought to be due to the scan being considerably noisier than the other scans due to the patient's relatively large body habitus. To test this theory, the algorithm was re-run on thicker slice reconstructions from the same patient (1.25 mm vs 0.625 mm). This reduced the amount of lung wrongly classified as honeycombing from 44% to 34%. An example slice is shown below (Figure 5-4).

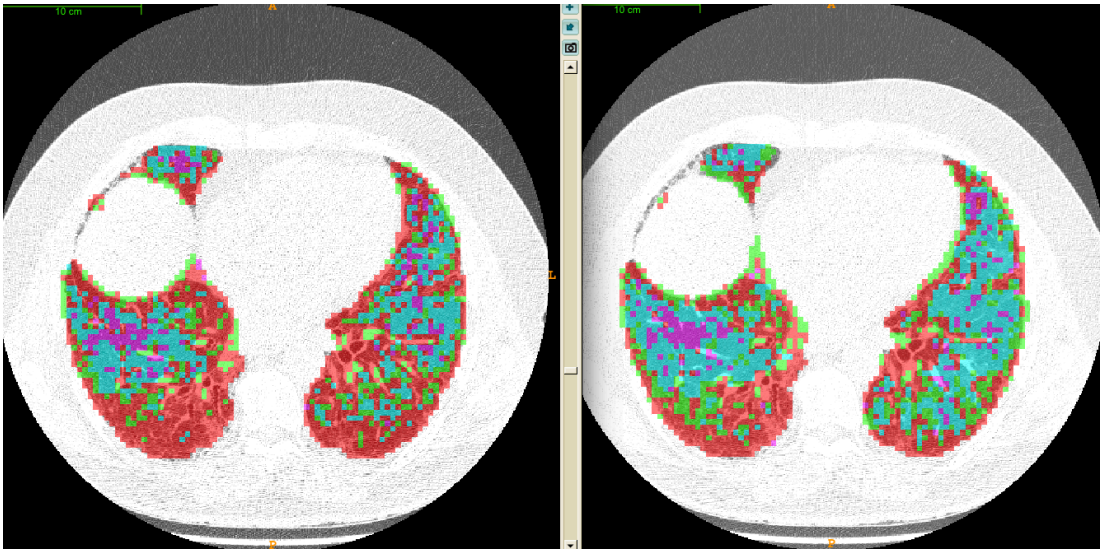

Figure 5-4 Example from Patient 12 showing a greater percentage of honeycomb classification (red) in the left-hand thin (0.625 mm) slice compared with the thicker (1.25 mm) slice on the right

5.4 The relationship between computer estimated CT fibrosis and FVC

Linear regression was used to compare computer estimated CT fibrosis and percent predicted FVC (Figure 5-5). The patients are the same as in the previous section.

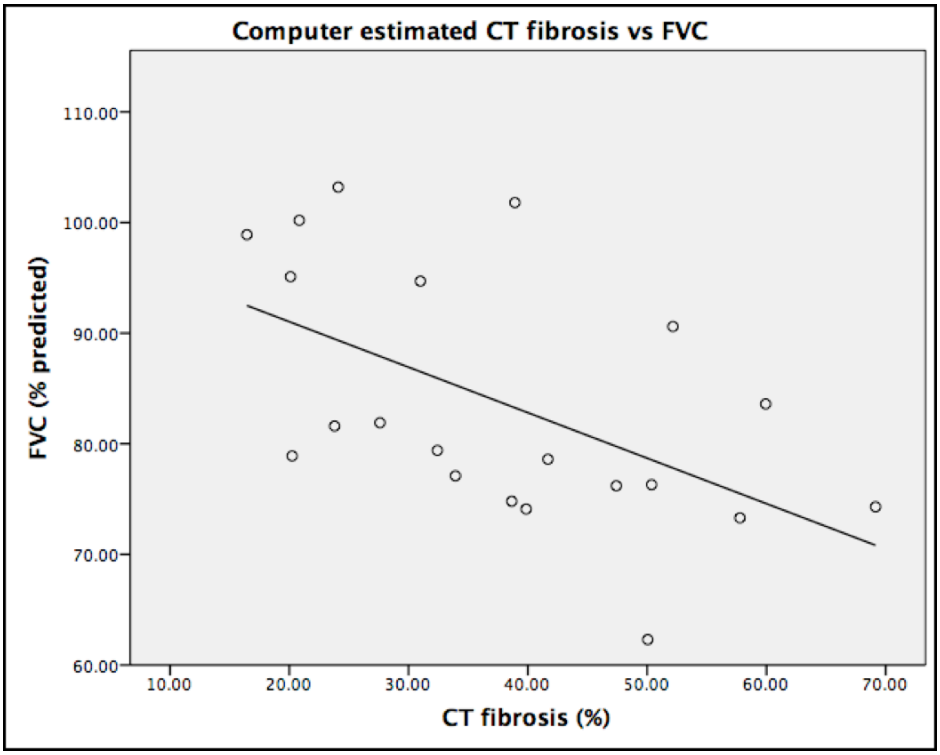

Figure 5-5 Linear regression showing the correlation between computer calculated percentage fibrosis and FVC

We found a moderate correlation between the two variables with an R-value of -0.54 (p=0.01). The strength of correlation is not quite as good as for CT fibrosis and DL_{CO} (Section 5.3). There are a number of possible explanations for this. Firstly, our patients had quite a narrow range of percent predicted FVC values, ranging from 62% to 103%. In fact, nearly half the patients had values above 80% which can be considered within the normal range. This means that FVC is likely to be a less sensitive measure of lung disease in our patient group and may be more affected by random variation (noise). Ideally more subjects with a greater range of FVC values would be needed to better understand the relationship between the CT fibrosis score and FVC.

5.5 Relationship between CT fibrosis and the St George's Respiratory Questionnaire

Data was available for all 23 patients. Linear regression analysis was performed to assess the correlation between CT fibrosis and the total St George's score and between CT fibrosis and the individual domains of the St George's score, namely 'symptoms', 'impacts' and 'activities'. As shown in Figure 5-6, there was no significant correlation between CT fibrosis and either the total St George's score (R=0.31, p = 0.17) or the individual components relating to symptoms (R = 0.27, p = 0.24), impacts (R = 0.26, p = 0.25) or activities (R = 0.36, p = 0.11). There are several possible reasons for this: the St George's Respiratory Questionnaire was originally designed for use in patients with COPD and therefore some of the questions, for example about wheeze, are less relevant to patients with CFIP; the questionnaire relies on patients recall of their symptoms over the previous 3 months and therefore their answers may be affected by recall bias; for the questions in the activity and

impacts domains patients may find it difficult to judge whether their limitations are due to their respiratory disease or affected by other co-morbidities – this is likely to be especially relevant in our relatively elderly patient group. Also, as a general rule, questionnaire studies normally need large numbers of patients to show meaningful results due to the subjective nature of this type of measurement tool.

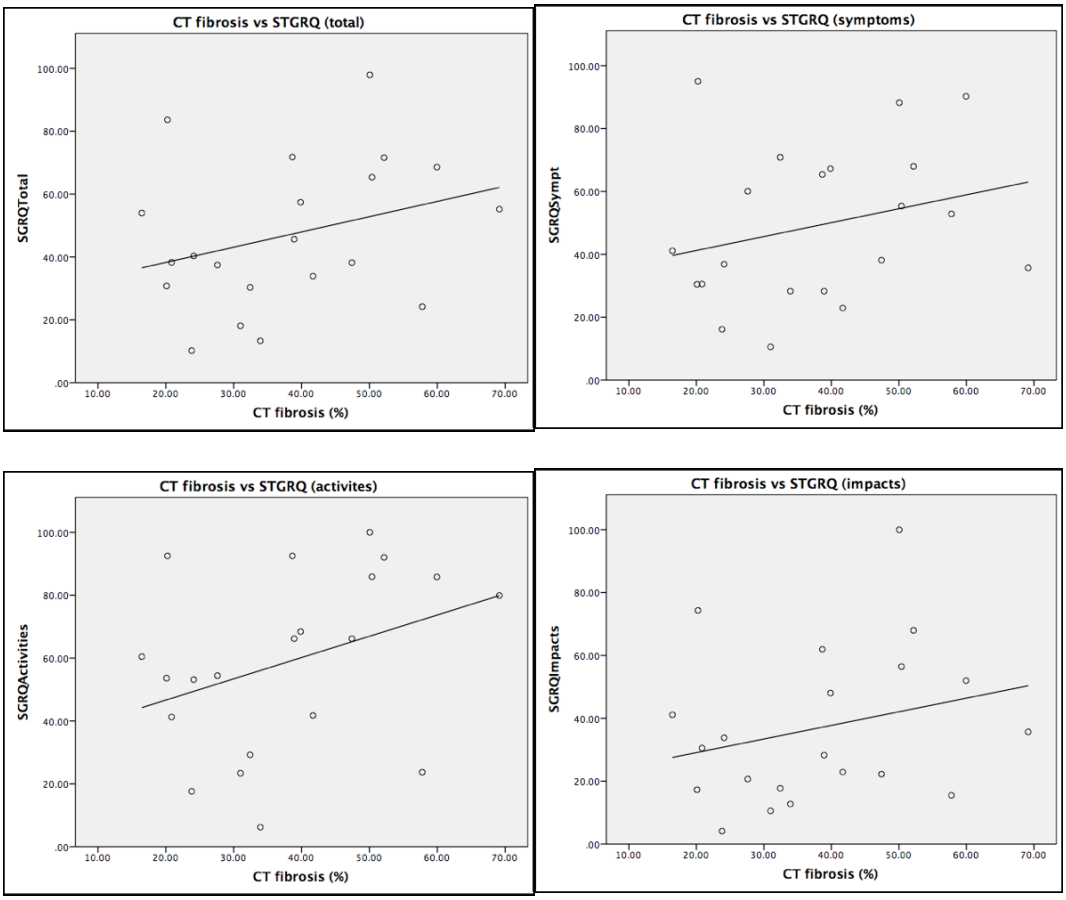

Figure 5-6 Relationship between CT fibrosis and the St George's Respiratory Questionnaire

5.6 Relationship between DL_{CO} and the St George's Respiratory Questionnaire

Linear correlation did not show any significant relationship between DL_{CO} and either the total score in the St George's Respiratory Questionnaire ($R = 0.33$, $p = 0.14$) or the individual St George's domains of symptoms ($R = 0.24$, $p = 0.29$), activities ($R =$

0.37, p = 0.10) or impacts (R = 0.35, p = 0.12) as shown in Figure 5-7 below. It is likely that this lack of correlation is due to similar reasons as the lack of correlation between CT fibrosis and STGRQ.

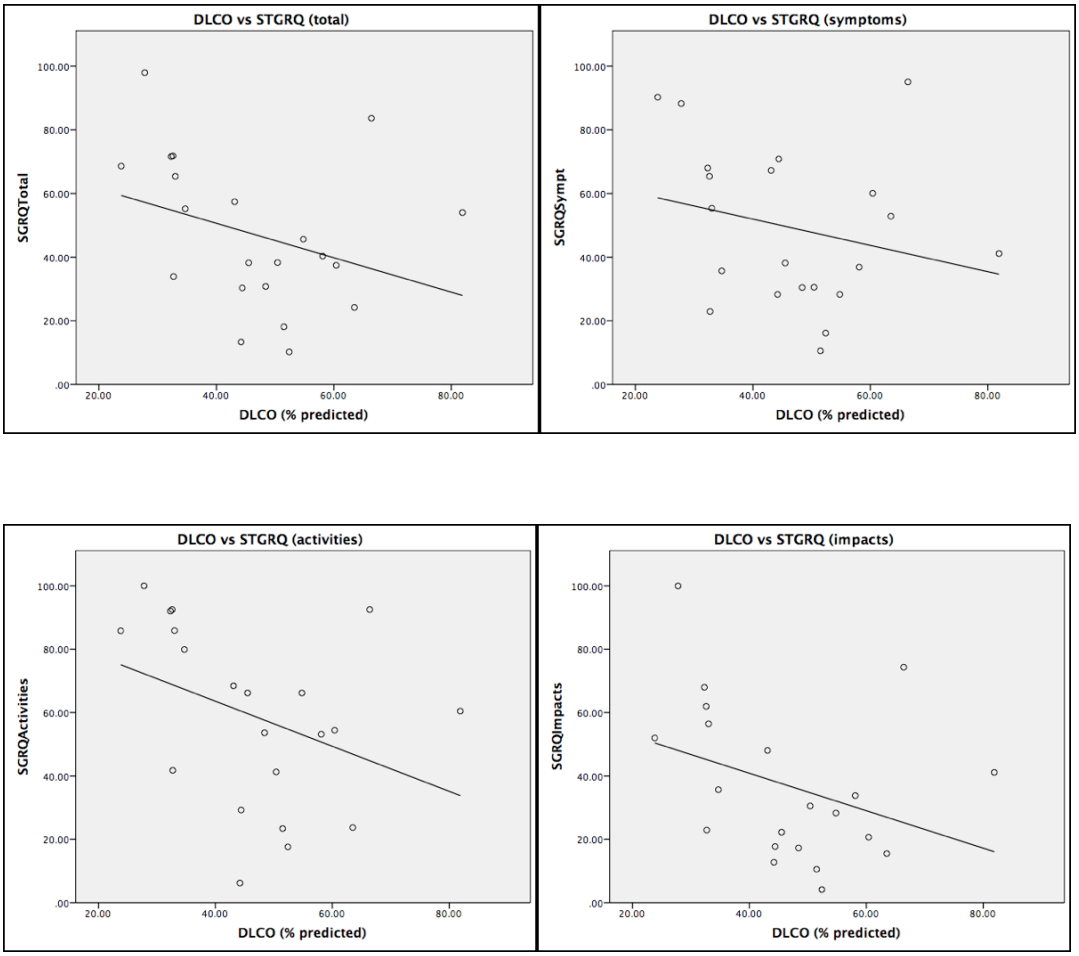

Figure 5-7 Relationship between DL_{CO} (percent-predicted) and the St George's Respiratory Questionnaire

5.7 Relationship between FVC and the St George's Respiratory Questionnaire

Linear correlation did not show any significant correlation between FVC and either the total STGRQ score (R = 0.30, p = 0.19) or the activities (R= 0.21, p = 0.36) or impacts (R = 0.32, p = 0.15) domains. There was a weak association shown between the FVC and the symptoms domain (R = 0.46, p = 0.04) but this result should be interpreted with caution, due to the small numbers of patients. Similar reasons to those

described in section 5.5 and section 5.6 are likely to explain the lack of association between FVC and the responses to the STGRQ.

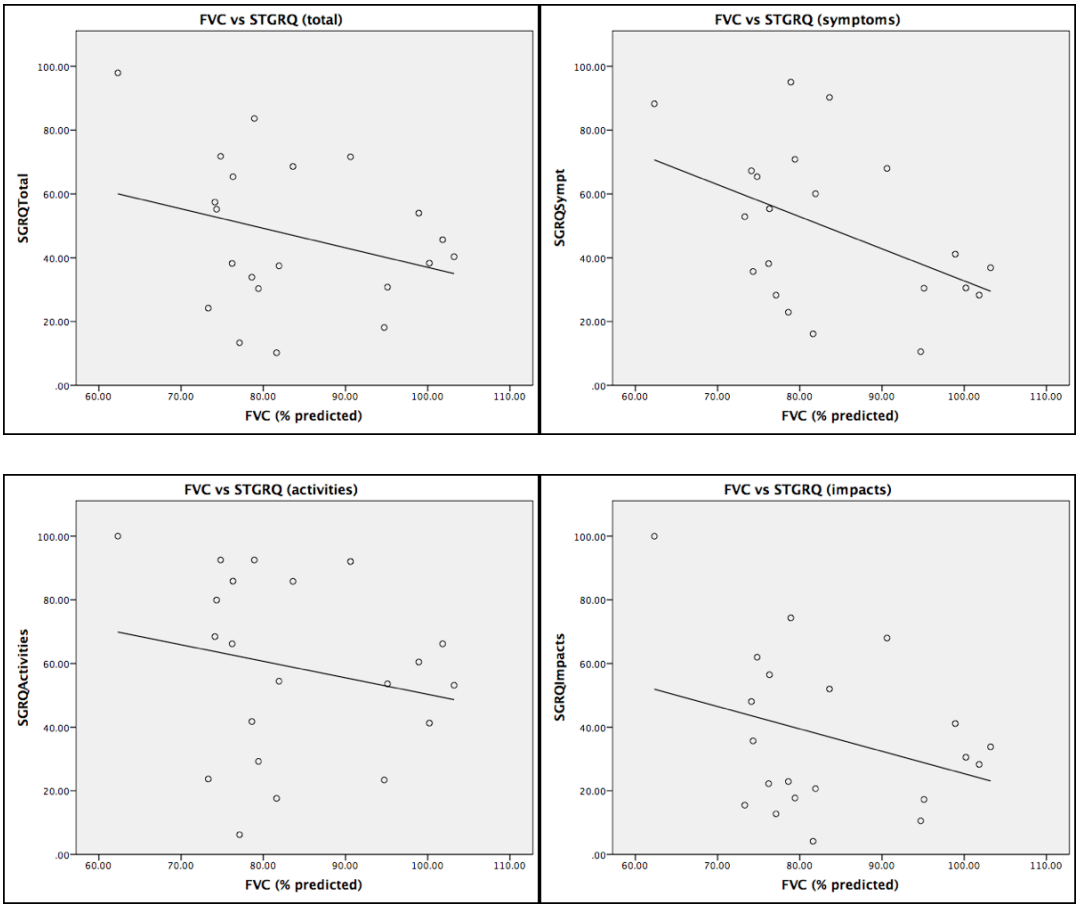

Figure 5-8 Relationship between FVC (percent-predicted) and the St George's Respiratory Questionnaire

5.8 Relationship between radiologist fibrosis score and computer fibrosis score

The correlation between radiologist visual score and computer calculated fibrosis score was analysed for the 23 patients scored by Radiologist 1 (EH). Five slices per patient were visually scored as previously described in the Methods section. This produced a total of 115 slices for which a radiologist visual score and computer fibrosis score was available. There was a moderate correlation between the radiologist fibrosis score (reticulation plus honeycombing) and the computer fibrosis score (reticulation plus honeycombing) with an R-value of 0.61 ($p<0.0001$) as shown in Figure 5-9.

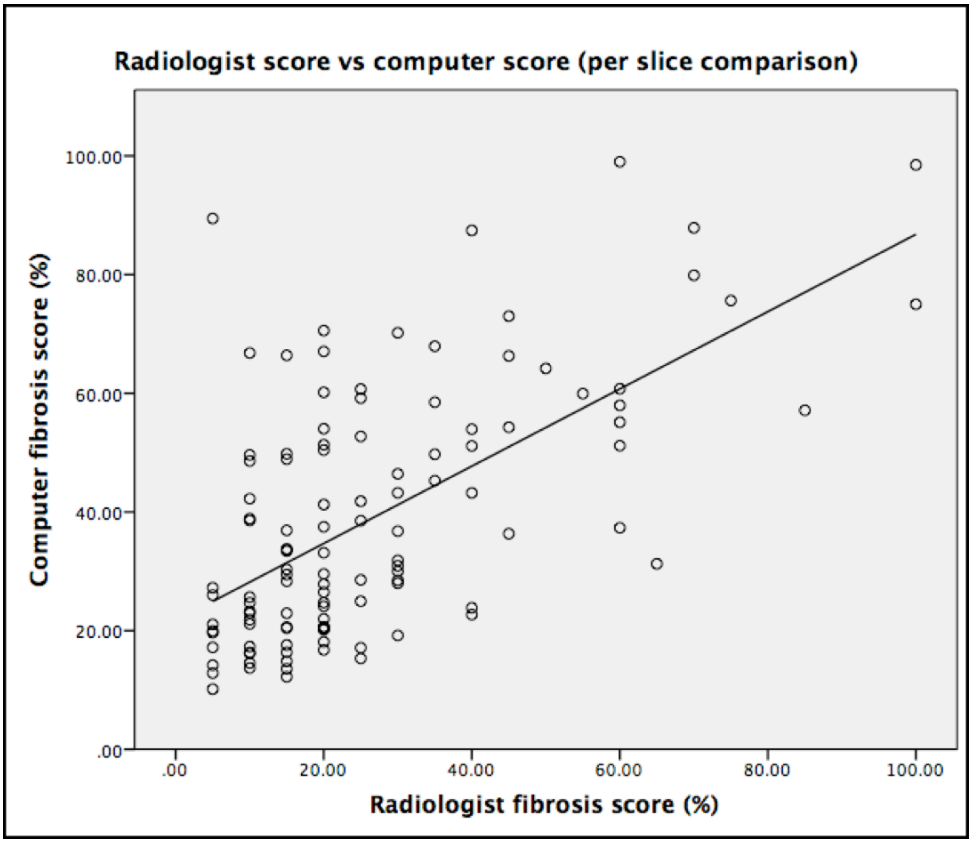

Figure 5-9 Relationship between the radiologist fibrosis score and the computer fibrosis score

Figure 5-9 illustrates a number of aspects of the data. Firstly, the radiologist score has been performed to the nearest 5% since it is not feasible to score more accurately than this and therefore there is clustering of the data along the x-axis. Secondly, because most of the patients in the study have relatively mild disease, there is clustering of data points towards the origin. Apart from the inherent differences between the continuous numbers provided by the computer and the discrete bins of the radiologist scoring, there are several other reasons for potential discrepancies between the radiologist and computer scoring including: the radiologist visual 'guestimate' of percentage involvement is prone to error which would be expected to be at least 5% either way; the radiologist is able to mentally dismiss artefact such as breathing or image noise which the computer frequently mis-classifies as pathology; the computer has not been trained to identify certain patterns such as ground glass opacity or consolidation which it must then assign to another category. Such

alternative patterns may be classified as fibrosis by the computer leading to a falsely inflated fibrosis result. In order to assess whether there was a systematic over- or under-estimation of the computer compared with the radiologist, we generated a Bland-Altman plot (Figure 5-10 below).

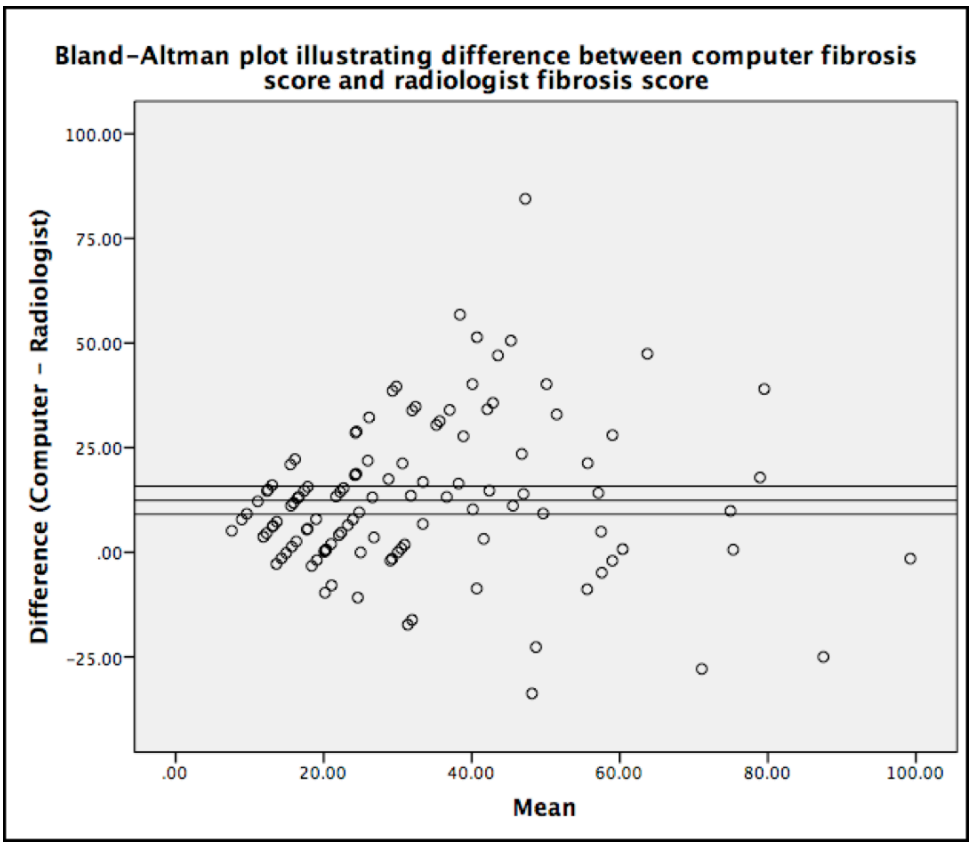

Figure 5-10 Bland-Altman plot illustrating the difference between the computer calculated fibrosis score and the radiologist estimated fibrosis score

As can be seen from figure 5-10, the computer fibrosis score tends to be higher than the radiologist fibrosis score with a mean difference of +12.4% (95% confidence interval = +9.1 to +15.8%). Most of this over-estimation can be accounted for by the edge artefact described earlier whereby the computer falsely assigns pixels at the periphery of the lung to the honeycombing class (accounting for a mean of 7.7% of erroneous honeycombing on the control scans) and by blood vessels which are erroneously classified as honeycombing. Other potential reasons include other artefacts such as breathing artefact and classification of vessels or airways as fibrotic

lung; mis-classification of emphysema or atelectasis as fibrosis and visual under-estimation by the radiologist.

Some clustering of data points is seen near the origin of the graph on the left. This may partly be explained by the fact that the potential to under-estimate disease is more limited when there are smaller amounts of fibrosis since it is not possible to have a fibrosis score of less than 0%. Whilst a score of more than 100% is also impossible, very few slices had such large amounts of fibrosis.

In order to assess whether the computer ranking of severity is similar to the radiologist ranking of severity, even if there was a systematic over-estimation of fibrosis by the computer or under-estimation by the radiologist, a Spearman's rank-order correlation was performed. This showed a correlation co-efficient of 0.568 which was significant at the 0.01 level and indicates a moderate correlation between the radiologist and computer fibrosis scores.

5.9 Inter-observer variability between radiologist's visual scoring

For 8 of the CT scans, scoring was performed by two radiologists - Radiologist 1 who performed the scoring on all 23 patients as presented in section 5.8 and Radiologist 2 who performed scoring on 8 patients. Therefore, at 5 slices per patient, there were 40 slices which were scored by both radiologists.

Linear regression analysis was performed to compare the two radiologists scoring for honeycombing, reticulation and total fibrosis (honeycombing plus reticulation). As shown in Figure 5-11, there was excellent agreement between the radiologists for honeycombing ($R=0.86$; $p < 0.001$) but less good agreement in terms of reticulation ($R = 0.44$; $p < 0.01$). This is not surprising as honeycombing usually has discrete margins, often lying adjacent to normal lung and is therefore fairly straightforward to visually assess in terms of area affected. However, linear reticulation typically affects

the sub-pleural lung with linear densities interspersed with more normal lung and with an ill-defined boundary between normal and abnormal lung. Interestingly, total fibrosis scores showed even better agreement than honeycombing with an R-value of 0.90 ($p<0.001$) suggesting that some lung which was classified as honeycombing by Radiologist 1 was classified as reticulation by Radiologist 2 and vice versa. Intraclass correlations were also used to compare Radiologist 1 and Radiologist 2. This gave an ICC of 0.94 for total fibrosis, 0.90 for honeycombing and 0.60 for reticulation which according to Landis et al, represents 'near perfect' agreement for total fibrosis and honeycombing and 'moderate' agreement for reticulation (Landis and Koch, 1977).

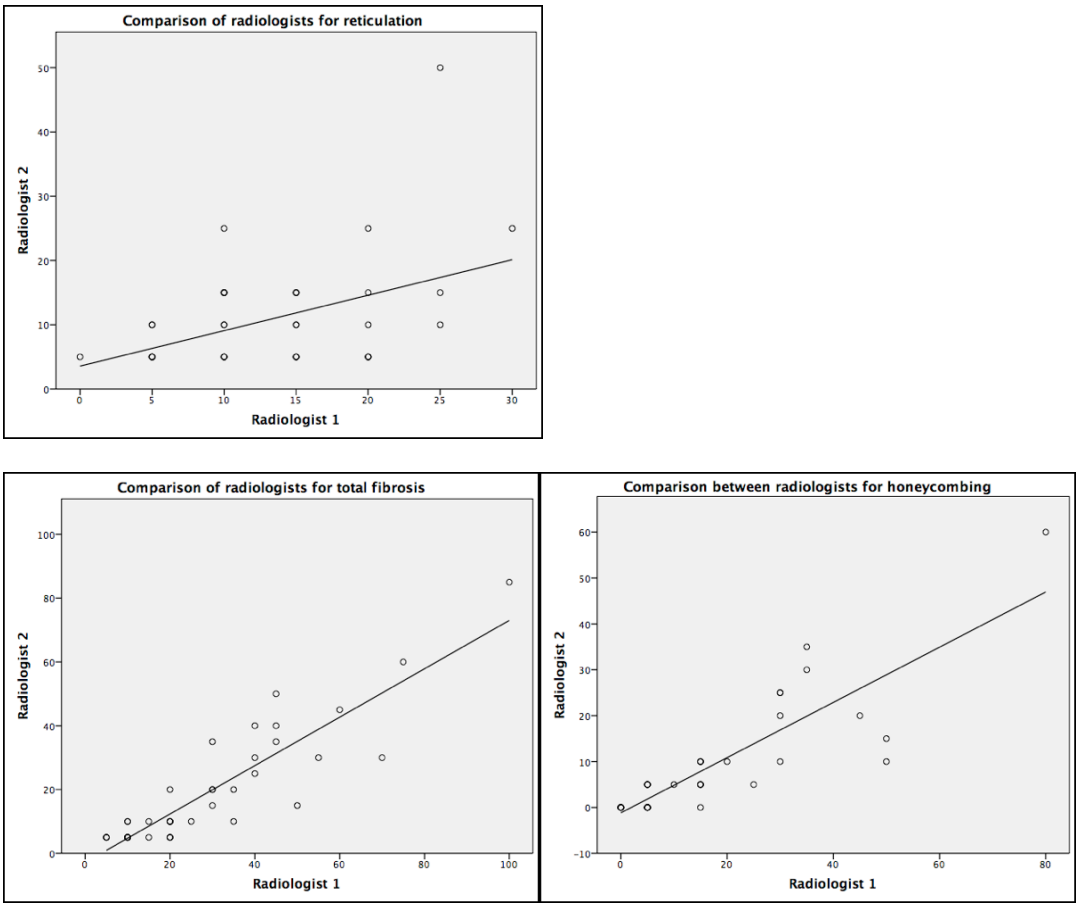

Figure 5-11 Comparison of radiologists scores for total fibrosis, honeycombing, and reticulation for 40 axial CT slices

5.10 CHAPTER SUMMARY

In summary, we have now looked at the correlations between the computer estimated lung volume and the total lung capacity (TLC); between the CFS and pulmonary function tests (DLCO and FVC); between the CFS and scores on the St George's Respiratory Questionnaire; between pulmonary function tests and the St George's Respiratory Questionnaire; between the visual scoring of radiologist 1 and the computer fibrosis score and between the two radiologists' scoring.

CHAPTER 6: DISCUSSION

6.1 INTRODUCTION TO CHAPTER

As previously discussed, the chronic fibrosing idiopathic pneumonias (CFIPs) are challenging diseases with no definite aetiology and with a poor prognosis. The commonest CFIP, which also has the worst prognosis, is idiopathic pulmonary fibrosis. Recently, new pharmacological treatments for this disease have become available (King Jr et al., 2014, Noble et al., 2011, Richeldi et al., 2014) but still have a relatively small effect on disease progression. They are also very expensive and therefore it is important to ensure they are being used on patients who are likely to benefit and to consider stopping treatment if it is ineffectual. Measuring disease severity at diagnosis and over time is important for determining treatment decisions and for assessing the effect of new therapeutic agents. Because new treatments are likely to have relatively small effects on disease severity an accurate and reproducible measurement is essential. A recent observational study of patients with CFIP who had two CT scans performed one year apart and were not receiving any interventional treatment (Yoon et al., 2013) showed that there was only a 1.6% change in disease severity over the year, as assessed by two radiologists, which is within the range of measurement error. Another potential role for CT is in predicting prognosis in people with CFIP – both visual assessment and quantitative indices of fibrosis have shown promise in this area (Best et al., 2008, Edey et al., 2011, Oda et al., 2014, Maldonado et al., 2014). In terms of measuring severity, CT has the advantage of providing a non-invasive and quick method for demonstrating the volume of disease as a percentage of the total lung volume. Whilst patients with significant lung disease may struggle to perform pulmonary function tests, almost all patients are able to tolerate a CT. CT also has the advantage of being able to provide spatial information on disease extent and can quantify the contribution of different

disease pathologies such as UIP and emphysema which often co-exist (Jankowich and Rounds, 2012).

To date, visual assessment has been the mainstay of assessing disease extent on CT and in the clinical situation is usually limited to a verbal description of disease severity such as mild, moderate or severe. Attempts to develop a more formal method of visually assessing disease severity have been made, typically using a semi-quantitative scoring system or a visual estimation of affected lung (Goh et al., 2008, Edey et al., 2011). However, such techniques are time consuming, prone to imprecision and limited by inter and intra-observer variability. Moreover, they are not practical for assessing the hundreds of slices generated by multi-detector volumetric CT. As a consequence, visual quantification techniques have not been adopted outside of the research setting.

An automated computerised approach to quantifying disease therefore has several potential advantages, particularly in terms of speed and reproducibility. Various attempts have been made to develop such methods but most studies have been retrospective and therefore have several biases including patient selection, use of different models of CT scanner and use of different CT algorithms, to which quantitative analysis is highly susceptible (Rosas et al., 2011, Uchiyama et al., 2003, Maldonado et al., 2014, Bartholmai et al., 2013, Yoon et al., 2013).

The purpose of the current thesis was to develop and test a new computer algorithm for the assessment of disease severity on CT using a specific measure of textural geometry known as Minkowski Functionals. In previous chapters we have discussed the need for novel methods of measuring lung abnormality on CT in the CFIPs and described the methodology for developing our computer software including its training and testing. We have examined the performance of the computer algorithm on a series of normal scans and on scans from 24 prospectively recruited patients with CFIP. The performance of the computer algorithm has been assessed by comparing the computer estimated

fibrosis score (CFS) with a radiologist estimated computer score (RFS) and by comparing the CFS with pulmonary function tests and with a symptom and activities questionnaire (St George's Respiratory Questionnaire). We will now discuss our results in the light of previous studies and draw conclusions about the meaning of our results, including the limitations of our findings.

6.2. Development of the computer algorithm

The first challenge when developing an automated lung analysis algorithm is to separate the lung tissue from other structures which do not need to be analysed such as the chest wall, mediastinum, large airways and blood vessels. The separation of the lung from the chest wall, mediastinum and large blood vessels is a relatively straightforward task in patients with normal lung owing to the large difference in density between aerated lung and other soft tissue structures. However, it has proven much more challenging in interstitial lung disease. For example Meng et al (Meng et al., 2012) tested a basic adaptive thresholding segmentation method on 2768 CT scans obtained from a number of databases including the Lung Image Database Consortium which includes patients with interstitial lung disease (Armato III et al., 2007). They found that 16% of the interstitial lung disease scans encountered segmentation problems, typically excluding the diseased lung from the calculated lung volume. Wang et al (Wang et al., 2009) developed a more successful method using a combination of thresholding and textural analysis. The method was tested on 45 scans with moderate to severe interstitial lung disease in which an expert manually traced the lung edge, on three slices per patient, to provide the reference standard. They measured the percentage overlap between the area segmented by the radiologist and the automatically segmented area and found a mean agreement of 96% (range of 90 – 99%). Although this is an impressive success rate, it should be noted that only three slices were assessed and the diaphragmatic lung was

not assessed, which is often the most difficult area to segment. In addition, the maximum discrepancy of 10% is not insignificant. Such a level of discrepancy would be enough to make a considerable impact on the estimation of the extent of abnormal lung and may limit assessment of small degrees of change. We found that our segmentation method was robust and reliable even when assessing scans with severe interstitial lung disease but encountered problems when there was breathing artefact, particularly at the lung bases, and when there was excessive image noise such as in larger patients.

It should also be mentioned that we have only segmented out the larger, more central airways and blood vessels from the lung volume. This was a conscious decision as there is a risk when segmenting the smaller peripheral airways and blood vessels that adjacent lung parenchyma will also be removed. This is a particular concern in interstitial lung disease which preferentially tends to affect the outer, pleural-based lung where smaller vessels are found. The downside of our approach is that it will lead to the classification of some smaller airways and blood vessels as diseased lung.

Another important aspect of our software development was the training of the algorithm. We trained our algorithm using regions of interest marked up by a single radiologist on a single CT scan with the identification of 4 categories of abnormality - honeycombing, reticulation, normal and indeterminate lung. The training scan was chosen to be representative of the four lung classifications and to have a range of disease extent (mild, moderate and severe) within the same scan. A total of 38,919 ROIs were generated from the radiologist's manual segmentation and a sample of 12,000 5 x 5 x 5 pixel VOIs were randomly selected from this total to train the algorithm. In contrast, in the paper by Zavaletta et al (Zavaletta et al., 2007), the CALIPER (Computer-Aided Lung Informatics for Pathology Evaluation and Rating) software was trained using 14 scans with representative VOIs which were determined by 'experts' to contain at least 70% of one of the following categories: honeycombing, reticulation, ground glass,

emphysema and normal lung. It is not specified in the paper how many radiologist readers were used or why they were considered to be experts. Despite training the algorithm to detect ground glass and emphysema, their test CT datasets (n=4) did not contain either of these abnormalities and therefore performance in this area could not be assessed. They did not have an indeterminate classification. The use of this indeterminate or 'near-normal' category in our study was designed to allow the categorisation of areas of subtle disease which could represent a pre-clinical or potentially reversible stage of pathology. Honeycombing and reticulation in UIP are believed to be irreversible and not amenable to pharmaceutical treatment (Müller et al., 1987) although reticular opacities in NSIP do have the potential to resolve (Nishiyama et al., 2000). Therefore when assessing new potential therapeutic treatments, identification of a pre-cursor to reticulation or honeycombing is desirable. The current study is insufficient to establish whether this 'near-normal' lung is important but future studies could be performed to address this.

Our algorithm has a number of limitations. Firstly, we did not train the algorithm to identify GGO or emphysema due to a lack of representative ROIs in our patient population. Therefore, our algorithm may mis-classify emphysema or ground glass as other abnormalities such as honeycombing or reticulation, falsely increasing the volume of these abnormalities. Most of our patients did not have significant emphysema and therefore the size of error is probably small but this may not be the case in other patient groups.

The fact that the training VOIs were provided by a single radiologist is a potential limitation of our study. Zavaletta et al (Zavaletta et al., 2007), in their study using the CALIPER software, describe how their test scans, which were marked up by 3 different radiologists, showed considerable variation in the areas considered to be normal or reticulation. It is also evident that three of their CT scans (datasets 1-3) contained very

little honeycombing and would therefore be considered to have relatively mild disease. Only the scan labelled 'dataset 4' contained a significant percentage of honeycombing and much of this looks to have been mis-classified, judging by the example of segmentation presented in their paper. In 2014, Maldonado (Maldonado et al., 2014) presented further work using an updated version of CALIPER in a study looking at interval progression of interstitial lung disease in a retrospective study of 55 patients with at least 2 scans spaced between 3 and 15 months apart. The CALIPER algorithm was trained using VOIs from the scans of 14 patients from the Lung Tissue Research Consortium (LTRC) (Armato III et al., 2007). From these scans 4 radiologists independently labelled ROIs as emphysema, ground glass, honeycomb, normal or reticular densities with the proviso that at least 70% of the VOI should be composed of a single class. VOIs were only selected for training purposes if there was consensus on the classification from all 4 radiologists. This resulted in the following number of training VOIs: emphysema (80), ground glass (150), honeycomb (187), normal (265), reticular densities (294). The CALIPER software uses VOIs of 15 x 15 x 15 pixels whereas we use VOIs of 5 x 5 x 5 pixels. We chose the smaller VOIs because we found that it allowed greater accuracy on our thin-slice acquisitions. We therefore used a considerably larger number of smaller ROIs in training our algorithm. Interestingly, it appears that texture-based methods may not need huge numbers of training ROIs/VOIs. The texture-based method of Lee et al (Lee et al., 2009) used a training dataset comprising 63 severe emphysema ROIs, 65 mild emphysema, 70 bronchiolitis obliterans and 67 normal lung. Despite not being trained on interstitial lung disease, this method formed the basis of the AQS software which was successfully used by Yoon et al (Yoon et al., 2013) to assess for serial changes in CFIP on scans performed 1 year apart.

6.3 LESSONS LEARNT FROM THE ANALYSIS OF NORMAL SCANS

Several important points can be learnt from the analysis of the normal scans, mostly in terms of sources of error in the computer algorithm, as follows:

1) Edge artefact: it is apparent that there is a mis-classification at the edge of the lung volume, where the lung abuts the chest wall, and where normal lung is sometimes mis-labelled as honeycombing or reticulation. This is consistent in that it affects all scans but is much more pronounced when using the lung algorithm. The 2014 paper by Maldonado et al (Maldonado et al., 2014) also shows a similar artefact in their Figure 2 where the strip of lung at the boundary between lung and chest wall has been mis-classified as normal whereas the rest of the lung in the slice of interest shows honeycomb change. This artefact is not alluded to in the paper but is consistent with a boundary phenomenon.

2) Choice of algorithm: we have demonstrated that the choice of reconstruction algorithm has a significant influence on the classification of lung texture. For example we found that on average, 37% of a normal scan was mis-classified as honeycombing on the lung algorithm whilst only 6% on average was mis-classified on the standard algorithm. A phantom study looking at the best reconstruction algorithm for assessing ILD, performed as a pre-cursor to the Lung Tissue Research Consortium project, was performed by Zhang et al using the American College of Radiologist's 'CT Image Quality Phantom Model 464' (Zhang et al., 2008). This phantom is designed to measure multiple aspects of the CT image including CT number accuracy, high and low contrast resolution and image noise. They found that the GE 'bone' algorithm and the Siemens 'B46f' algorithm were the best algorithms in terms of providing high enough spatial resolution for clinical use as well as preserving CT number accuracy. Maldonado et al (Maldonado et al., 2014) used the bone algorithm for their study using the CALIPER software, presumably because the software had been trained on LTRC scans. We conclude that it is

essential that all studies of automated quantification in IIP provide explicit details of the algorithm used, something which is not always done.

3) Effect of inspiration: we examined the effect of inspiratory volume on texture analysis of two normal scans which had been obtained in both full inspiration and full expiration and found that degree of inspiration had a dramatic effect on the amount of lung classified as indeterminate and a significant effect on the amount of lung incorrectly classified as reticulation but little effect on the amount of lung incorrectly classified as honeycombing. This highlights the importance of ensuring scans are performed in full inspiration.

The important lessons learnt from the analysis of normal scans should inform the interpretation of the results from the pathological scans. To our knowledge, there have been no previous studies of the use of a texture-based lung analysis algorithm on CT scans of normal lungs which we would argue is a fundamental aspect of assessing any new algorithm.

6.4 SCANS FROM PROSPECTIVELY RECRUITED PATIENTS

6.4.1. General observations

As expected, it was evident that issues that caused problems with normal scans also caused problems with pathological scans, including 'image noise' (particularly related to increased patient BMI) and 'breathing artefact'. Linear streak artefact which is seen in low dose scans are particularly problematic for texture based quantification as they can mimic reticulation (Coxson, 2013). This was more of a problem with the scans of the prospectively recruited patients than with the normal scans due to the reduced dose protocol used in the QUIC scans.

6.4.2 Comparison of computer-estimated fibrosis and radiologist-estimated fibrosis

We used two methods to assess the accuracy of the computer algorithm compared with radiologist assessment, as follows:

1) In the first experiment, we compared the computer output with radiologist manual segmentation on a region by region basis using a single CT scan. This data was presented in Chapter 3 and showed that on a training set of 12,000 ROIs, the computer had a sensitivity of 81.2% for identification of honeycombing, 86.8% sensitivity for identification of normal lung, 44.9% sensitivity for reticulation and 42.5% sensitivity for indeterminate lung. The computer performed less well with classification on the testing set of 26,919 ROIs with sensitivity of 65.2% for honeycombing, 75.0 % sensitivity for normal lung, 27.7% sensitivity for reticulation and 26.6% sensitivity for indeterminate. Therefore, the computer is most accurate at distinguishing normal lung from abnormal lung and within the abnormal lung, is most accurate at classifying honeycombing compared with the other textures. There are several potential reasons for this: a) the results are likely to reflect the radiologist's confidence for the different classes during the preparation of the training data; b) whereas honeycombing is easily distinguished from normal lung by both radiologist and computer, the boundary between coarse reticulation and honeycombing is often less clear; c) the boundary between indeterminate and normal lung, by the very definition of the indeterminate class, is likely to be blurred. Of note, most of the honeycombing that was wrongly classified was classified as reticulation, which is likely to represent a milder stage of the same fibrotic process. It could therefore be argued that total fibrosis is a more important metric than individual features of fibrosis, such as honeycombing and reticulation.

Only a few previous studies have directly compared computer classification of lung texture with radiologist classification and only one (Uchiyama et al) has looked at the

sensitivity of computer versus radiologist on a region-by-region basis rather than on a slice-by-slice or lobar basis (Uchiyama et al., 2003). Uchiyama examined 315 single non-spiral slices from 105 patients with 3 radiologists labeling irregular ROIs as reticulation, nodular, honeycomb, emphysema, consolidation, non-specific or other (includes bulla, pleural thickening and atelectasis). They compared radiologist classification with a computer algorithm based on an artificial neural network. For the ROI analysis, they looked first at regions where all 3 radiologists agreed on the texture. In this case, the computer had 100% sensitivity for honeycombing, 100% sensitivity for reticulation and 88% sensitivity for normal lung. It should be noted that there were only 15 reticulation ROIs. They also performed a separate analysis where they looked at whether an ROI was classified as abnormal or normal. For this analysis they found that the computer had 97% sensitivity for abnormal lung if all 3 radiologists agreed the ROI was abnormal but this dropped to 85% if there was discrepancy between radiologists. They performed a third analysis on a slice-by-slice basis, this time looking at whether the radiologist and computer classified the whole slice as normal, abnormal or indeterminate. For slices classified by the radiologist as normal the computer classified that slice as normal 84% of the time and for slices classified by the radiologist as abnormal, the computer classified them as abnormal 90% of the time. On the other hand, if the radiologist classified the slice as indeterminate, then the computer classified it as normal 53% of the time and abnormal 47% of the time. Therefore, from both our findings and those of Uchiyama et al, we can conclude that: honeycombing and normal lung are most accurately identified by the computer; that computers using artificial neural network methods are highly sensitive to training data; and that both computers and radiologists struggle with 'indeterminate' lung which may represent an overlap between other categories.

2) The second method we used to validate the computer output was to prospectively compare it with radiologist visual scoring of multiple scans on a slice by slice basis with radiologist scoring of lung abnormalities to the nearest 5%. This was done using 5 slices for each of 23 patients (total of 115 slices). This method of comparison cannot be as spatially precise as the VOI-based method but it would be impractical to expect radiologists to manually segment the whole of a volumetric scan into all its different textural classifications. For comparison of the radiologist and computer in this method we decided to compare total fibrosis (reticulation plus honeycombing) rather than looking at honeycombing and reticulation separately. This decision was made due to the overlap that these classes had demonstrated in the first method described and because we felt that it was more important to distinguish diseased lung from non-diseased lung rather than to distinguish between different features of fibrosis.

Linear regression analysis was performed to examine the agreement between the radiologist total fibrosis score and the computer total fibrosis score. This showed a moderate correlation between the two scores with an R-value of 0.61 ($p<0.0001$) and Spearman's rank-order correlation co-efficient of 0.568 ($p < 0.01$). It is inevitable that there will never be perfect correlation with this method since the radiologist is scoring to the nearest 5% and the computer is measuring on a continuous scale. In addition, the radiologist visual scoring will always involve a margin of error due to the nature of the method, which is a value judgement. We have also identified a number of inaccuracies in the computer method caused by problems such as image noise; artefact from patient breathing, diaphragm motion and cardiac motion; problems segmenting airways and blood vessels and the edge artefact that was described earlier. It is also worth noting that the computer processes the data in terms of 3D volumes (5 x 5 x 5 pixel VOIs) whereas the radiologist is making their assessment on 2D slices. Nevertheless the degree of correlation is reasonable, particularly given that the computer has only been

trained on data from a single subject and by a single radiologist. Iterative training with new datasets could improve this performance.

6.4.3 Comparison of computer estimated fibrosis and pulmonary function tests

In the previous chapter we looked at the correlation between the computer fibrosis score (CFS) and pulmonary function tests, specifically the forced vital capacity (FVC) and the diffusion capacity of carbon monoxide (DL_{CO}). These lung function measures were chosen since they are the most frequently used clinical measures of CFIP and because changes in these measures are the most frequent end-points in clinical trials involving CFIP patients (Raghu et al., 2012). We found that there was a good correlation between CFS and DL_{CO} with an R-value of -0.65 ($p = 0.001$) despite two outliers. We also found a slightly lesser correlation between CFS and FVC with an R-value of -0.54 ($p = 0.01$).

As previously discussed, several factors may influence the CT fibrosis score including segmentation errors, image noise and movement artefact. Some of these errors, such as the erroneous honeycombing seen at the edge of scans, will have a proportionally larger effect on the more mildly affected patients. Factors which may affect the DL_{CO} measurement include the patient's haemoglobin concentration, current smoking status, cardiac output and ability to expire and inspire appropriately for the single breath-hold technique (Macintyre et al., 2005). Equally, a number of factors can affect the FVC measurement including patient effort, fitness levels and co-existent emphysema (Miller et al., 2005).

In a similar study to ours, Xaubet et al (Xaubet et al., 1998) prospectively studied 39 untreated patients with IPF who underwent HRCT and full pulmonary function tests at baseline. Two radiologists in consensus performed a semi-quantitative visual scoring of the amount of ground glass opacification and 'reticular patterns' at each of 6 pre-defined anatomical levels. They did not specify quantification of honeycombing and

therefore it is assumed that this was included within the reticular pattern. Scoring was performed to the nearest 10% and then scores at the six levels were averaged to give a single overall score – the 'global disease extent'. The same method was employed to quantify emphysema. In 34 of the patients there was both ground glass opacification and a reticular pattern and in 30 of these cases, the reticular pattern was greater in extent that the ground glass opacification. They found that there was a significant correlation between global disease extent and both DL_{CO} ($R = -0.40$, $p = 0.03$) and FVC ($R = -0.46$, $p = 0.003$). They found that patients with co-existent emphysema had a significantly higher FVC but that this did not have a significant effect on DL_{CO}. The fact that we have used a volumetric measure of fibrotic lung and that we have used a continuous scale rather than measuring to the nearest 10% may explain the greater correlation we saw in our study. An earlier study by Staples et al (Staples et al., 1987) also used a visual assessment of overall disease extent estimated to the nearest 10% and compared this with spirometric lung volumes and DL_{CO}. They did not find any significant correlation between CT score and TLC, FVC or FEV1 but did find a significant correlation with DL_{CO} ($R = -0.64$, $p < 0.001$). More recently, a number of studies using computerised methods of quantifying ILD have correlated their computer scoring with pulmonary function tests. For example, Yoon et al assessed the scans of 71 UIP and 18 NSIP patients with their AQS (automated quantification score) which examines multiple textural features. They studied 2 interval CTs performed one year apart. They found that baseline 'fibrosis score', defined as honeycombing plus reticulation, had a moderate correlation with DL_{CO} ($R = -0.47$, $p < 0.05$) and 'total abnormal' (which also included ground glass opacification, emphysema and consolidation) also showed a moderate correlation with DL_{CO} ($R = -0.52$, $p < 0.05$). This correlation is slightly less strong than in our study and there are several potential reasons for this, including patient related factors and factors to do with the software algorithm. However, the fact that the AQS was

not trained on patients with pulmonary fibrosis may be particularly relevant.

Interestingly, although their study reports correlations of change in FVC with change in AQS, they do not report correlation between the AQS score at baseline and baseline FVC, as they do with DL_{CO}. The reason for this is not specified. A separate study by Bartholmai et al (Bartholmai et al., 2013) described the use of the CALIPER software for CT quantification, an algorithm which uses a histogram signature mapping technique and multidimensional scaling (see also Section 6.2). The software was designed to measure honeycombing, reticulation, ground glass opacification and emphysema. They studied a group of 119 subjects from the Lung Tissue Research Consortium (LTRC) which is a multi-institutional database of CT scans, lung function and pathology information. The subjects were described as having 'interstitial lung disease' but a specific pathological diagnosis was not given. Although standardised protocols are provided by the LRTC for GE and Siemens CT scanners, it is likely that the patients were scanned on a number of different scanners from these manufacturers. The authors found that the percentage of reticulation showed significant correlation with FVC ($R = -0.63$) and DL_{CO} ($R = -0.65$) and that percentage normal lung correlated with FVC ($R = 0.66$), DL_{CO} ($R = 0.59$) and TLC ($R = 0.56$). Similar results were shown by Rosas et al (Rosas et al., 2011) in a group of patients with rheumatoid arthritis associated interstitial lung disease. Their texture-based quantification method with 25 vectors, including co-occurrence and run-length vectors, showed significant correlations with FVC ($R = -0.48$) and DL_{CO} ($R = -0.53$). In summary, our findings of correlation between CFS and pulmonary function tests (FVC: $R= -0.54$; DL_{CO}: $R= -0.65$) are similar to those achieved in previous studies which have used both visual scoring methods and computerised methods.

Ultimately, it may be that the fundamental differences between what is being measured on pulmonary function tests and what is being measured on computer fibrosis scores is such that linear correlation between the two measures with an R-value of more than 0.6

to 0.7 is not possible. It also raises the question of what the reference standard for measurement of pulmonary fibrosis should be. Whilst both spirometry and DL_{CO} provide functional measures of pulmonary capacity and DL_{CO} provides an integrated assessment of the process of transferring carbon monoxide (and by inference oxygen) from the alveolus to the blood (Cotton and Graham, 2005), CT provides regional information about disease severity which is potentially less affected by non-respiratory impairments that may affect lung function tests such as general fitness, musculoskeletal problems and haemoglobin levels. A CT scan may also be less onerous for patients than pulmonary function tests and can be successfully performed in patients who are unable to perform pulmonary function tests.

6.4.4 Comparison of computer-estimated total fibrosis, pulmonary function tests and the St George's Respiratory Questionnaire

As described in the previous chapter, we found no significant or clinically meaningful correlation between either the total STGRQ score or the individual domains of the STGRQ and either the computer fibrosis score or any of the pulmonary function tests in our patient group.

There are several possible reasons for this, as follows:

1) Most questionnaire studies require many more patients than we had in our study due to the subjective nature of this type of measurement.

2) It was noted that the mainly elderly patients in our study often had difficulty with the three-month recall elements of the questionnaire, particularly if they had had intercurrent illnesses.

3) Patients have had difficulty distinguishing between impairments due to respiratory disease and impairments due to other co-morbidities such as cardiac disease or musculoskeletal problems.

4) As previously discussed, the STGRQ was not primarily designed for patients with interstitial lung disease and therefore is likely to be a less good tool for quantifying IIP compared with quantifying the airways diseases for which it was designed.

To our knowledge, no other studies have compared automated computerised textural quantification of interstitial lung disease with quality of life questionnaires although Camiciottoli et al compared lung density features (MLA, kurtosis and skewness) and the baseline dyspnoea index (BDI) in 48 patients with systemic sclerosis associated interstitial lung disease (Camiciottoli et al., 2007) and found a correlation between one of the BDI domains (magnitude of task) and all three density features (R = −0.39 to −0.44; $p < 0.05$). In a separate study on patients with systemic sclerosis, Kim et al described the used of their texture-based quantitative lung fibrosis ('QLF') computer algorithm with radiologist scoring and with the baseline dyspnoea index. The QLF, which measured reticulation only, showed a small correlation with the BDI domains of 'magnitude of task' (R-value = 0.16, p =0.02) and 'magnitude of effort' (R-value = 0.17, p = 0.01) (Kim et al., 2010).

6.4.5 Radiologist agreement

We found excellent agreement between radiologists for total fibrosis (ICC = 0.94) and honeycombing (ICC = 0.90) with lesser agreement for reticulation (ICC = 0.60). As discussed in the previous chapter, honeycombing is likely to be easier to score accurately due to its generally well-defined borders, relative homogeneity and typical appearance. Reticulation tends to have less well-defined borders and reticular abnormalities often interdigitate with normal appearing lung, making it difficult to segment manually or visually. Other studies which have compared radiologist agreement for measures of pulmonary fibrosis include the study by Yoon et al (Yoon et al., 2013) which compared the scoring of 89 scans (71 UIP and 18 NSIP) by two

radiologists. Each radiologist scored the scans to the nearest 5% for the following classes: normal, emphysema, ground glass opacification, reticular opacities, honeycombing and consolidation. Although not explicitly stated in the paper, it is assumed that all slices were scored as they used non-spiral scans which would each have contained only approximately 20 slices with lung parenchyma on them. They found ICCs of 0.63 for honeycombing and 0.49 for reticulation which were less good than our correlations of 0.90 for honeycombing and 0.60 for reticulation. Maldonado et al (Maldonado et al., 2014), in the electronic supplement to their article, quoted ICCs for radiologist scoring of 12 different anatomical zones of the lung, dividing the lung into upper, middle and lower zones vertically and into 'rind' and 'core' regions from outside to inside. They quoted ICCs for honeycombing between 0.33 and 0.72 and for reticulation between 0.46 and 0.77 depending on the anatomical zone. Interestingly, there was better agreement between the radiologists for reticulation scores in all the rind sections compared with the corresponding core sections but there was no difference between rind and core with respect to agreement on honeycombing. Again, the agreement between the radiologists on the Maldonado paper is less good than in our study although clearly our findings are limited by the relatively small numbers of comparison slices.

6.4.6 Strengths and weaknesses of the study

There are several strengths to our study, as follows:

1) One of the main strengths is that the study was performed prospectively, whereas the majority of studies using QCT of ILD are retrospective. A potential criticism of retrospective studies is that only a small proportion of available scans were suitable for analysis with quantitative CT (QCT). For example, Maldonado et al were only able to identify 55 patients with at least 2 scans suitable for analysis over a 10 year period. It is

likely that many more patients than this were scanned but the scans were not suitable for analysis. The authors do not state how many scans were reviewed in order to identify the 55 paired scans for analysis. Issues which may make scans unsuitable for QCT include excessive image noise, breathing/cardiac motion, expiratory or non-spiral acquisition and super-added disease such as infection. Breathing artefact is a particular problem in patients with ILD who may often struggle to sustain a prolonged breath-hold. This sampling bias makes it difficult to know how generalisable the QCT technique is and how frequently it might work on routine scans. Other studies have scanned patients prospectively as part of pharmaceutical trials but the CT analysis is often done in a post-hoc manner (Best et al., 2003, Kim et al., 2010).

2) Another strength of our study is the spiral nature of our CT scans. Until relatively recently, the standard high-resolution CT protocol was a non-contiguous acquisition with slices obtained every 10 to 30 mm meaning that typically only 10% of the lung volume was scanned. Comparison of this non-volumetric technique with a global measure such as DL_{CO} or FVC is inherently flawed due to the sampling error produced by only assessing part of the lung volume with CT. Using a non-spiral technique also risks underestimating disease extent which is typically worse at the lung bases. The lower lobes may contribute most of the disease extent in terms of the total lung volume but if only a few lower lobe slices are obtained then their contribution may be under-estimated. A recent study of cyclophosphamide vs placebo in scleroderma lung disease showed a 12% difference between the two groups when the most severely affected parts of the lung were compared but only a 5% difference when the whole lung was assessed (Kim et al., 2011). Even recently, studies of quantitative CT in interstitial lung disease have used non-spiral acquisitions (Rosas et al., 2011, Yoon et al., 2013). This likely reflects the retrospective nature of these studies which analyzed scans which had been

acquired before spiral CT was routinely performed in IIP but limits how generalisable their findings are to modern scanning techniques.

3) As one of the benefits of the prospective study design, we took great care to optimise the CT parameters and ensure that these were accurately reproduced for all the scans. All scans were performed on the same scanner, using the same research protocol with dedicated breathing instructions and identical reconstruction algorithms.

4) A strength of our experimental approach was the use of the control group of normal scans. This informed our understanding of potential sources of error in the computer algorithm and the effects of varying CT acquisition parameters such as reconstruction algorithm and degree of inspiration.

Our study has a number of limitations. Firstly, we did not employ spirometric control to ensure that all the scans were performed in a fixed percentage of inspiration. It is known that depth of inspiration has a marked effect on lung attenuation (Newell Jr et al., 2013) and we have demonstrated significant effects of depth of respiration on the classification of normal lung using our texture-based method. However, whilst the use of spirometric gating allows more precision for determining depth of inspiration, it can be technically challenging for patients with lung disease (Madani et al., 2010) and is not routinely available in clinical practice.

Another potential criticism of our study is that we did not use a 'noise index' to increase or decrease the mA in order to produce a target noise level. As previously discussed, this can cause problems with classification of textural abnormalities, often generating false positive abnormalities. Whilst acknowledging this will have led to a degree of inaccuracy in our results, the decision to used fixed exposure parameters was made in order to ensure that there was a more consistent range of dose to our research patients, who may not directly benefit personally from participation in the study.

We have also acknowledged the fact that the training of the algorithm by a single radiologist and the use of a single training scan are potential limiting factors (section 6.2). Finally we recognise that the small numbers of patients in our study limits the conclusions which can be drawn.

6.5 Conclusion

In summary, we have developed and tested a novel computer algorithm based on Minkowski functionals for the analysis of lung texture on CT scans of the thorax. Testing of the algorithm on normal scans and a group of prospectively recruited patients with pulmonary fibrosis has shown that, despite some minor limitations, the algorithm can successful segment the fibrotic lung from the surrounding tissues and separate it into the main types of fibrosis normally assessed by radiologists. A moderate correlation was shown between the radiologist and computer scoring, which was comparable with previous studies. Significant correlations were also shown between the computer estimated fibrosis score and the pulmonary function tests DL_{CO} and FVC. Our study is the first prospective study to use Minkowski functionals for the assessment of pulmonary fibrosis and the standardised nature of our CT protocol and relative lack of selection bias makes our results likely to be more robust than similar retrospective studies. We also showed that training of the algorithm improved its performance. We believe that our computer algorithm has the potential for assessing severity of chronic fibrosing interstitial lung disease on individual scans as well as the potential for assessing change in lung disease with time.

CHAPTER 7: FUTURE DIRECTIONS

7.1 INTRODUCTION

Having demonstrated initial promising results with our new computer algorithm for the assessment of CFIP, there are several areas that need further work and several potential future applications for the algorithm. We will now discuss potential future work, concentration on improving the computer algorithm, testing of the algorithm in larger patient numbers, analysis of serial scans, potential use as a predictive classifier or surrogate endpoint and use in other lung pathologies.

7.2 DEVELOPING THE COMPUTER ALGORITHM

We have already highlighted a number of areas where the computer algorithm could be improved and these can be summarised as follows:

1) We would like to try and reduce or eliminate the edge artefact which is seen at the surface of the lung and erroneously increases the percentage of honeycomb or reticulation classifications. Several methods could be used to overcome this including using overlapping VOIs at the edge of the lung to increase the amount of sampling in this area and reduce any partial volume effect. Another approach is to 'reflect' or 'project' the outer rind of the lung outside the lung surface so that this lung is no longer seen by the algorithm as being at the edge. The excess projected lung could then be trimmed by the algorithm following the classification step.

2) We would like to improve the removal of more distal airways and blood vessels from the lung volume since these can be erroneously classified as fibrosis. There are several different potential methods for removing these structures and several different methods may need to be tested before finding the most successful method. Challenges to improving this aspect of the algorithm include whether/how to 'fill' the holes left by

small vessels and avoiding removal of lung parenchyma during the segmentation process.

3) It would be beneficial to train the algorithm to detect other pathological textures such as emphysema or ground glass opacification. This would need an appropriately selected training set of scans with these abnormalities and appropriate radiologist mark-up.

4) We believe the accuracy of the computer algorithm could be improved by training on a larger number of scans from patients with CFIP with a wider range of disease severity and by training the algorithm with larger numbers of radiologists.

7.3 Testing of the computer algorithm in larger numbers of patients

We have already acknowledged that we have tested the algorithm in a relatively small number of patients. Although it is encouraging that we have shown statistically significant agreement between the output of the computer algorithm and radiologist visual scoring and significant correlations with pulmonary function tests, we recognise that validation of the software in a larger patient group is necessary. We plan to do this in the near future.

7.4 Testing the algorithm on serial scans

We also plan to test the algorithm on serial CTs performed at defined intervals in order to see how well the algorithm can quantify change in disease and whether it is more sensitive to change than pulmonary function tests or patient questionnaires. These patients have already been recruited into the QUIC study and we expect the data will be available for analysis in 2016.

If we can show that the algorithm is able to detect change in disease over time, there is potential for it to be used in the future as a surrogate endpoint for clinical trials of therapeutic agents. However, ultimate validation of the algorithm for this purpose is

likely to need a multi-centre study with sufficient patients in mild, moderate and severe groups and a follow-up period long enough to reach defined endpoints of either mortality or a significant change in lung function. In future studies we would also consider amending the CT protocol to use an automated mA modulation with a fixed noise level in order to ensure consistent image noise in all patients.

7.5 Use of the algorithm for assessment of other lung pathologies

With appropriate training, we believe the algorithm could be adapted for use in other lung pathologies. The most obvious examples would be other fibrotic lung disease such as scleroderma lung disease or rheumatoid associated lung disease. There is also potential for use in other lung diseases which cause distortion of the normal lung architecture and scarring such as sarcoidosis.

7.6 Conclusion

We believe that further technical improvement/training of the computer algorithm would improve its performance and that it has potential for use as a surrogate endpoint or predictive biomarker if it's validity can be proved in a large, longitudinal, prospective study. We also believe that textural analysis of the lung using algorithms such as ours has future use in several other lung diseases which affect the lung parenchyma.

BIBLIOGRAPHY

ARMANIOS, M. Y., CHEN, J. J.-L., COGAN, J. D., ALDER, J. K., INGERSOLL, R. G., MARKIN, C., LAWSON, W. E., XIE, M., VULTO, I. & PHILLIPS III, J. A. 2007. Telomerase mutations in families with idiopathic pulmonary fibrosis. *New England Journal of Medicine,* 356**,** 1317-1326.

ARMATO III, S. G., MCNITT-GRAY, M. F., REEVES, A. P., MEYER, C. R., MCLENNAN, G., ABERLE, D. R., KAZEROONI, E. A., MACMAHON, H., VAN BEEK, E. J. & YANKELEVITZ, D. 2007. The Lung Image Database Consortium (LIDC): an evaluation of radiologist variability in the identification of lung nodules on CT scans. *Academic radiology,* 14**,** 1409-1421.

ARNS, C. H., KNACKSTEDT, M. A. & MECKE, K. R. 2002. Characterising the morphology of disordered materials. *Morphology of Condensed Matter.* Springer.

BARNES, J. 1992. Characteristics and control of contrast in CT. *Radiographics,* 12**,** 825-837.

BARTHOLMAI, B. J., RAGHUNATH, S., KARWOSKI, R. A., MOUA, T., RAJAGOPALAN, S., MALDONADO, F., DECKER, P. A. & ROBB, R. A. 2013. Quantitative computed tomography imaging of interstitial lung diseases. *Journal of thoracic imaging,* 28**,** 298-307.

BAUMGARTNER, K. B., SAMET, J. M., STIDLEY, C. A., COLBY, T. V. & WALDRON, J. A. 1997. Cigarette smoking: a risk factor for idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine,* 155**,** 242.

BEST, A. C., LYNCH, A. M., BOZIC, C. M., MILLER, D., GRUNWALD, G. K. & LYNCH, D. A. 2003. Quantitative CT indexes in idiopathic pulmonary fibrosis: Relationship with physiologic impairment. *Radiology,* 228**,** 407-414.

BEST, A. C., MENG, J., LYNCH, A. M., BOZIC, C. M., MILLER, D., GRUNWALD, G. K. & LYNCH, D. A. 2008. Idiopathic pulmonary fibrosis: Physiologic tests, quantitative CT indexes, and CT visual scores as predictors of mortality. *Radiology,* 246**,** 935-940.

BLACKBURN, E. H. & GALL, J. G. 1978. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in< i> Tetrahymena</i>. *Journal of molecular biology,* 120**,** 33-53.

BLAND, J. M. & ALTMAN, D. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet,* 327**,** 307-310.

BLASCO, M. A. 2005. Telomeres and human disease: ageing, cancer and beyond. *Nature Reviews Genetics,* 6**,** 611-622.

BOEHM, H., FINK, C., ATTENBERGER, U., BECKER, C., BEHR, J. & REISER, M. 2008. Automated classification of normal and pathologic pulmonary tissue by topological texture features extracted from multi-detector CT in 3D. *European radiology,* 18**,** 2745-2755.

BROUILETTE, S. W., MOORE, J. S., MCMAHON, A. D., THOMPSON, J. R., FORD, I., SHEPHERD, J., PACKARD, C. J. & SAMANI, N. J. 2007. Telomere length, risk of coronary heart disease, and statin treatment in the West of Scotland Primary Prevention Study: a nested case-control study. *The Lancet,* 369**,** 107-114.

CAMICIOTTOLI, G., ORLANDI, I., BARTOLUCCI, M., MEONI, E., NACCI, F., DICIOTTI, S., BARCAROLI, C., CONFORTI, M. L., PISTOLESI, M. & MATUCCI-CERINIC, M. 2007. Lung CT densitometry in systemic sclerosis: correlation with lung function, exercise testing, and quality of life. *CHEST Journal,* 131**,** 672-681.

CHAREMZA, M., THONNCS, E., BHALCRAO, A., PARR, D. & TRUST, C. Integral Geometry Descriptors for Characterizing Emphysema and Lung Fibrosis in

HRCT Images. The First International Workshop on Pulmonary Image Analysis, 2008. Lulu. com.
COHEN, J. 1988. *Statistical power analysis for the behavioral sciences*, Psychology Press.
COTTON, D. J. & GRAHAM, B. L. 2005. Single-breath carbon monoxide diffusing capacity or transfer factor. *physiological basis of respiratory disease. Quebec: BC Decker*, 6596-6569.
COWELL, J. K. 2001. *Molecular genetics of cancer,* Bios, Oxford.
COXSON, H. O. 2013. Sources of variation in quantitative computed tomography of the lung. *Journal of thoracic imaging,* 28**,** 272-279.
COXSON, H. O., DIRKSEN, A., EDWARDS, L. D., YATES, J. C., AGUSTI, A., BAKKE, P., CALVERLEY, P., CELLI, B., CRIM, C. & DUVOIX, A. 2013. The presence and progression of emphysema in COPD as determined by CT scanning and biomarker expression: a prospective analysis from the ECLIPSE study. *The Lancet Respiratory Medicine,* 1**,** 129-136.
CRESSONI, M., GALLAZZI, E., CHIURAZZI, C., MARINO, A., BRIONI, M., MENGA, F., CIGADA, I., AMINI, M., LEMOS, A. & LAZZERINI, M. 2013. Limits of normality of quantitative thoracic CT analysis. *Crit Care,* 17**,** R93.
CRONKHITE, J. T., XING, C., RAGHU, G., CHIN, K. M., TORRES, F., ROSENBLATT, R. L. & GARCIA, C. K. 2008. Telomere shortening in familial and sporadic pulmonary fibrosis. *American journal of respiratory and critical care medicine,* 178**,** 729.
DIEDERICH, S., LENZEN, H., WINDMANN, R., PUSKAS, Z., YELBUZ, T. M., HENNEKEN, S., KLAIBER, T., EAMERI, M., ROOS, N. & PETERS, P. E. 1999. Pulmonary Nodules: Experimental and Clinical Studies at Low-Dose CT 1. *Radiology,* 213**,** 289-298.
DOEL, T., MATIN, T. N., GLEESON, F. V., GAVAGHAN, D. J. & GRAU, V. Pulmonary lobe segmentation from CT images using fissureness, airways, vessels and multilevel B-splines. Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on, 2012. IEEE, 1491-1494.
EDEY, A. J., DEVARAJ, A. A., BARKER, R. P., NICHOLSON, A. G., WELLS, A. U. & HANSELL, D. M. 2011. Fibrotic idiopathic interstitial pneumonias: HRCT findings that predict mortality. *European radiology,* 21**,** 1586-1593.
EGAN, J. J., STEWART, J. P., HASLETON, P. S., ARRAND, J. R., CARROLL, K. B. & WOODCOCK, A. A. 1995. Epstein-Barr virus replication within pulmonary epithelial cells in cryptogenic fibrosing alveolitis. *Thorax,* 50**,** 1234.
EUROPEAN, R. S. & SOCIETY, A. T. 2002. American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. *American journal of respiratory and critical care medicine,* 165**,** 277.
FELL, C. D., MARTINEZ, F. J., LIU, L. X., MURRAY, S., HAN, M. K., KAZEROONI, E. A., GROSS, B. H., MYERS, J., TRAVIS, W. D. & COLBY, T. V. 2010. Clinical predictors of a diagnosis of idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine,* 181**,** 832.
FLAHERTY, K., THWAITE, E., KAZEROONI, E., GROSS, B., TOEWS, G., COLBY, T., TRAVIS, W., MUMFORD, J., MURRAY, S. & FLINT, A. 2003a. Radiological versus histological diagnosis in UIP and NSIP: survival implications. *Thorax,* 58**,** 143.

FLAHERTY, K. R., THWAITE, E. L., KAZEROONI, E. A., GROSS, B. H., TOEWS, G. B., COLBY, T. V., TRAVIS, W. D., MUMFORD, J. A., MURRAY, S., FLINT, A., LYNCH, J. P. & MARTINEZ, F. J. 2003b. Radiological versus histological diagnosis in UIP and NSIP: survival implications. *Thorax,* 58**,** 143-148.

FLAHERTY, K. R., TOEWS, G. B., TRAVIS, W. D., COLBY, T. V., KAZEROONI, E. A., GROSS, B. H., JAIN, A., STRAWDERMAN, R. L., PAINE, R., FLINT, A., LYNCH, J. P. & MARTINEZ, F. J. 2002. Clinical significance of histological classification of idiopathic interstitial pneumonia. *European Respiratory Journal,* 19**,** 275-283.

GOH, N. S., DESAI, S. R., VEERARAGHAVAN, S., HANSELL, D. M., COPLEY, S. J., MAHER, T. M., CORTE, T. J., SANDER, C. R., RATOFF, J. & DEVARAJ, A. 2008. Interstitial lung disease in systemic sclerosis: a simple staging system. *American journal of respiratory and critical care medicine,* 177**,** 1248-1254.

GRENIER, P., VALEYRE, D., CLUZEL, P., BRAUNER, M. W., LENOIR, S. & CHASTANG, C. 1991. Chronic diffuse interstitial lung disease: diagnostic value of chest radiography and high-resolution CT. *Radiology,* 179**,** 123-132.

GRIBBIN, J., HUBBARD, R. & SMITH, C. 2009. Role of diabetes mellitus and gastro-oesophageal reflux in the aetiology of idiopathic pulmonary fibrosis. *Respiratory medicine,* 103**,** 927-931.

GRIBBIN, J., HUBBARD, R. B., LE JEUNE, I., SMITH, C. J., WEST, J. & TATA, L. J. 2006. Incidence and mortality of idiopathic pulmonary fibrosis and sarcoidosis in the UK. *Thorax,* 61**,** 980-5.

GROUP, W. 1995. The World Health Organization quality of life assessment (WHOQOL): position paper from the World Health Organization. *Social science and medicine,* 41**,** 1403-1409.

HALL, E. J. 2002. Lessons we have learned from our children: cancer risks from diagnostic radiology. *Pediatric radiology,* 32**,** 700-706.

HANSELL, D. M., BANKIER, A. A., MACMAHON, H., MCLOUD, T. C., MULLER, N. L. & REMY, J. 2008. Fleischner Society: glossary of terms for thoracic imaging. *Radiology,* 246**,** 697.

HARTLEY, P. G., GALVIN, J. R., HUNNINGHAKE, G. W., MERCHANT, J. A., YAGLA, S. J., SPEAKMAN, S. B. & SCHWARTZ, D. A. 1994. High-resolution CT-derived measures of lung density are valid indexes of interstitial lung disease. *Journal of Applied Physiology,* 76**,** 271-277.

HU, S., HOFFMAN, E. A. & REINHARDT, J. M. 2001. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *Medical Imaging, IEEE Transactions on,* 20**,** 490-498.

HUBBARD, R., JOHNSTON, I. & BRITTON, J. 1998. Survival in patients with cryptogenic fibrosing alveolitis: a population-based cohort study. *Chest,* 113**,** 396-400.

HUNNINGHAKE, G. W., ZIMMERMAN, M. B., SCHWARTZ, D. A., KING JR, T. E., LYNCH, J., HEGELE, R., WALDRON, J., COLBY, T., MULLER, N. & LYNCH, D. 2001. Utility of a lung biopsy for the diagnosis of idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine,* 164**,** 193-196.

ITO, K. & BARNES, P. J. 2009. COPD as a disease of accelerated lung aging. *CHEST Journal,* 135**,** 173-180.

IWAI, K., MORI, T., YAMADA, N., YAMAGUCHI, M. & HOSODA, Y. 1994. Idiopathic pulmonary fibrosis. Epidemiologic approaches to occupational exposure. *American journal of respiratory and critical care medicine,* 150**,** 670.

JANKOWICH, M. D. & ROUNDS, S. I. 2012. Combined pulmonary fibrosis and emphysema syndrome: a review. *CHEST Journal,* 141**,** 222-231.

JOHNSTON, I. D., PRESCOTT, R. J., CHALMERS, J. C. & RUDD, R. M. 1997. British Thoracic Society study of cryptogenic fibrosing alveolitis: current presentation

and initial management. Fibrosing Alveolitis Subcommittee of the Research Committee of the British Thoracic Society. *Thorax,* 52**,** 38-44.
JONES, P. W., QUIRK, F. & BAVEYSTOCK, C. 1991. The St George's respiratory questionnaire. *Respiratory medicine,* 85**,** 25-31.
JONES, P. W., QUIRK, F. H., BAVEYSTOCK, C. M. & LITTLEJOHNS, P. 1992. A self-complete measure of health status for chronic airflow limitation: the St. George's Respiratory Questionnaire. *American Review of Respiratory Disease,* 145**,** 1321-1327.
KAZEROONI, E. A., MARTINEZ, F. J., FLINT, A., JAMADAR, D. A., GROSS, B. H., SPIZARNY, D. L., CASCADE, P. N., WHYTE, R. I., LYNCH 3RD, J. & TOEWS, G. 1997. Thin-section CT obtained at 10-mm increments versus limited three-level thin-section CT for idiopathic pulmonary fibrosis: correlation with pathologic scoring. *AJR. American journal of roentgenology,* 169**,** 977-983.
KIM, H., TASHKIN, D., CLEMENTS, P., LI, G., BROWN, M., ELASHOFF, R., GJERTSON, D., ABTIN, F., LYNCH, D. & STROLLO, D. 2010. A computer-aided diagnosis system for quantitative scoring of extent of lung fibrosis in scleroderma patients. *Clinical and experimental rheumatology,* 28**,** S26.
KIM, H. J., BROWN, M. S., ELASHOFF, R., LI, G., GJERTSON, D. W., LYNCH, D. A., STROLLO, D. C., KLEERUP, E., CHONG, D. & SHAH, S. K. 2011. Quantitative texture-based assessment of one-year changes in fibrotic reticular patterns on HRCT in scleroderma lung disease treated with oral cyclophosphamide. *European radiology,* 21**,** 2455-2465.
KING JR, T. E., BRADFORD, W. Z., CASTRO-BERNARDINI, S., FAGAN, E. A., GLASPOLE, I., GLASSBERG, M. K., GORINA, E., HOPKINS, P. M., KARDATZKE, D. & LANCASTER, L. 2014. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *New England Journal of Medicine,* 370**,** 2083-2092.
KING JR, T. E., PARDO, A. & SELMAN, M. 2011. Idiopathic pulmonary fibrosis. *The Lancet,* 378**,** 1949-1961.
KLEINERMAN, R. A. 2006. Cancer risks following diagnostic and therapeutic radiation exposure in children. *Pediatric radiology,* 36**,** 121-125.
KOHONEN, T. 1982. Self-organized formation of topologically correct feature maps. *Biological cybernetics,* 43**,** 59-69.
LANDIS, J. R. & KOCH, G. G. 1977. The measurement of observer agreement for categorical data. *biometrics***,** 159-174.
LEE, K. S., PRIMACK, S. L., STAPLES, C. A., MAYO, J. R., ALDRICH, J. E. & MÜLLER, N. 1994. Chronic infiltrative lung disease: comparison of diagnostic accuracies of radiography and low-and conventional-dose thin-section CT. *Radiology,* 191**,** 669-673.
LEE, Y., SEO, J. B., LEE, J. G., KIM, S. S., KIM, N. & KANG, S. H. 2009. Performance testing of several classifiers for differentiating obstructive lung diseases based on texture analysis at high-resolution computerized tomography (HRCT). *Computer methods and programs in biomedicine,* 93**,** 206-215.
LYNCH, D. A., GODWIN, J. D., SAFRIN, S., STARKO, K. M., HORMEL, P., BROWN, K. K., RAGHU, G., KING JR, T. E., BRADFORD, W. Z. & SCHWARTZ, D. A. 2005. High-resolution computed tomography in idiopathic pulmonary fibrosis: diagnosis and prognosis. *American journal of respiratory and critical care medicine,* 172**,** 488-493.
MACINTYRE, N., CRAPO, R., VIEGI, G., JOHNSON, D., VAN DER GRINTEN, C., BRUSASCO, V., BURGOS, F., CASABURI, R., COATES, A. & ENRIGHT, P. 2005. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. *European Respiratory Journal,* 26**,** 720-735.

MADANI, A., VAN MUYLEM, A. & GEVENOIS, P. A. 2010. Pulmonary Emphysema: Effect of Lung Volume on Objective Quantification at Thin-Section CT 1. *Radiology,* 257**,** 260-268.
MAHLER, D., WEINBERG, D., WELLS, C. & FEINSTEIN, A. 1984. The measurement of dyspnea. Contents, interobserver agreement, and physiologic correlates of two new clinical indexes. *CHEST Journal,* 85**,** 751-758.
MALDONADO, F., MOUA, T., RAJAGOPALAN, S., KARWOSKI, R. A., RAGHUNATH, S., DECKER, P. A., HARTMAN, T. E., BARTHOLMAI, B. J., ROBB, R. A. & RYU, J. H. 2014. Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis. *European Respiratory Journal,* 43**,** 204-212.
MAPEL, D. W., HUNT, W. C., UTTON, R., BAUMGARTNER, K. B., SAMET, J. M. & COULTAS, D. B. 1998. Idiopathic pulmonary fibrosis: survival in population based and hospital based cohorts. *Thorax,* 53**,** 469-76.
MARTINEZ, T. Y., PEREIRA, C. A., DOS SANTOS, M. L., CICONELLI, R. M., GUIMARAES, S. M. & MARTINEZ, J. A. 2000. Evaluation of the short-form 36-item questionnaire to measure health-related quality of life in patients with idiopathic pulmonary fibrosis. *CHEST Journal,* 117**,** 1627-1632.
MATHIESON, J., MAYO, J., STAPLES, C. & MÜLLER, N. 1989. Chronic diffuse infiltrative lung disease: comparison of diagnostic accuracy of CT and chest radiography. *Radiology,* 171**,** 111-116.
MECKE, K. R. 2000. Additivity, convexity, and beyond: applications of Minkowski Functionals in statistical physics. *Statistical Physics and Spatial Statistics.* Springer.
MELICONI, R., ANDREONE, P., FASANO, L., GALLI, S., PACILLI, A., MINIERO, R., FABBRI, M., SOLFOROSI, L. & BERNARDI, M. 1996. Incidence of hepatitis C virus infection in Italian patients with idiopathic pulmonary fibrosis. *Thorax,* 51**,** 315-317.
MENG, X., QIANG, Y., ZHU, S., FUHRMAN, C., SIEGFRIED, J. M. & PU, J. 2012. Illustration of the obstacles in computerized lung segmentation using examples. *Medical physics,* 39**,** 4984-4991.
MILLER, M. R., HANKINSON, J., BRUSASCO, V., BURGOS, F., CASABURI, R., COATES, A., CRAPO, R., ENRIGHT, P., VAN DER GRINTEN, C. & GUSTAFSSON, P. 2005. Standardisation of spirometry. *Eur Respir J,* 26**,** 319-38.
MÜLLER, N., STAPLES, C., MILLER, R. & ABBOUD, R. 1988. " Density mask". An objective method to quantitate emphysema using computed tomography. *Chest,* 94**,** 782-787.
MÜLLER, N., STAPLES, C. A., MILLER, R. R., VEDAL, S., THURLBECK, W. M. & OSTROW, D. N. 1987. Disease activity in idiopathic pulmonary fibrosis: CT and pathologic correlation. *Radiology,* 165**,** 731-734.
NAVARATNAM, V., FLEMING, K., WEST, J., SMITH, C., JENKINS, R., FOGARTY, A. & HUBBARD, R. 2011. The rising incidence of idiopathic pulmonary fibrosis in the UK. *Thorax,* 66**,** 462-467.
NEWELL JR, J. D., SIEREN, J. & HOFFMAN, E. A. 2013. Development of quantitative computed tomography lung protocols. *Journal of thoracic imaging,* 28**,** 266-271.
NISHIYAMA, O., KONDOH, Y., TANIGUCHI, H., YAMAKI, K., SUZUKI, R., YOKOI, T. & TAKAGI, K. 2000. Serial high resolution CT findings in nonspecific interstitial pneumonia/fibrosis. *Journal of computer assisted tomography,* 24**,** 41-46.
NOBLE, P. W., ALBERA, C., BRADFORD, W. Z., COSTABEL, U., GLASSBERG, M. K., KARDATZKE, D., KING, T. E., LANCASTER, L., SAHN, S. A. & SZWARCBERG, J. 2011. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *The Lancet,* 377**,** 1760-1769.

ODA, K., ISHIMOTO, H., YATERA, K., NAITO, K., OGOSHI, T., YAMASAKI, K., IMANAGA, T., TSUDA, T., NAKAO, H. & KAWANAMI, T. 2014. High-resolution CT scoring system-based grading scale predicts the clinical outcomes in patients with idiopathic pulmonary fibrosis. *Respiratory research,* 15**,** 10.

PARR, D. G., STOEL, B. C., STOLK, J., NIGHTINGALE, P. G. & STOCKLEY, R. A. 2004. Influence of calibration on densitometric studies of emphysema progression using computed tomography. *American journal of respiratory and critical care medicine,* 170**,** 883-890.

RAGHU, G., COLLARD, H. R., ANSTROM, K. J., FLAHERTY, K. R., FLEMING, T. R., KING JR, T. E., MARTINEZ, F. J. & BROWN, K. K. 2012. Idiopathic pulmonary fibrosis: clinically meaningful primary endpoints in phase 3 clinical trials. *American journal of respiratory and critical care medicine,* 185**,** 1044-1048.

RAGHU, G., COLLARD, H. R., EGAN, J. J., MARTINEZ, F. J., BEHR, J., BROWN, K. K., COLBY, T. V., CORDIER, J.-F., FLAHERTY, K. R. & LASKY, J. A. 2011. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *American Journal of Respiratory and Critical Care Medicine,* 183**,** 788-824.

RAGHU, G., MAGETO, Y. N., LOCKHART, D., SCHMIDT, R. A., WOOD, D. E. & GODWIN, J. D. 1999. The accuracy of the clinical diagnosis of new-onset idiopathic pulmonary fibrosis and other interstitial lung diseasea prospective study. *CHEST Journal,* 116**,** 1168-1174.

RICHELDI, L., DU BOIS, R. M., RAGHU, G., AZUMA, A., BROWN, K. K., COSTABEL, U., COTTIN, V., FLAHERTY, K. R., HANSELL, D. M. & INOUE, Y. 2014. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *New England Journal of Medicine,* 370**,** 2071-2082.

ROSAS, I. O., YAO, J., AVILA, N. A., CHOW, C. K., GAHL, W. A. & GOCHUICO, B. R. 2011. Automated quantification of high-resolution CT scan findings in individuals at risk for pulmonary fibrosis. *CHEST Journal,* 140**,** 1590-1597.

SCHILHAM, A., VAN DER MOLEN, A. J., PROKOP, M. & DE JONG, H. W. 2010. Overranging at Multisection CT: An Underestimated Source of Excess Radiation Exposure 1. *Radiographics,* 30**,** 1057-1067.

SLUIMER, I. C., PROKOP, M., HARTMANN, I. & VAN GINNEKEN, B. 2006. Automated classification of hyperlucency, fibrosis, ground glass, solid, and focal lesions in high-resolution CT of the lung. *Medical Physics,* 33**,** 2610-2620.

STAPLES, C., MÜLLER, N., VEDAL, S., ABBOUD, R., OSTROW, D. & MILLER, R. 1987. Usual interstitial pneumonia: correlation of CT with clinical, functional, and radiologic findings. *Radiology,* 162**,** 377-381.

STEWART, J. P., EGAN, J. J., ROSS, A. J., KELLY, B. G., LOK, S. S., HASLETON, P. S. & WOODCOCK, A. A. 1999. The detection of Epstein-Barr virus DNA in lung tissue from patients with idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine,* 159**,** 1336.

SVERZELLATI, N., ZOMPATORI, M., DE LUCA, G., CHETTA, A., BNÀ, C., ORMITTI, F. & COBELLI, R. 2005. Evaluation of quantitative CT indexes in idiopathic interstitial pneumonitis using a low-dose technique. *European journal of radiology,* 56**,** 370-375.

SWIGRIS, J., KUSCHNER, W., JACOBS, S., WILSON, S. & GOULD, M. 2005. Health-related quality of life in patients with idiopathic pulmonary fibrosis: a systematic review. *Thorax,* 60**,** 588-594.

SWIGRIS, J. J., BROWN, K. K., BEHR, J., DU BOIS, R. M., KING, T. E., RAGHU, G. & WAMBOLDT, F. S. 2010. The SF-36 and SGRQ: Validity and first look at minimum important differences in IPF. *Respiratory Medicine,* 104**,** 296-304.

THÖNNES, E., BHALERAO, A. & PARR, D. G. Classification of lung disease in HRCT scans using integral geometry measures and functional data analysis. Proceedings of Medical Image Understanding and Analysis, 2010. University of Warwick, 25-30.
TOBIN, R. W., POPE, C. E., PELLEGRINI, C. A., EMOND, M. J., SILLERY, J. & RAGHU, G. 1998. Increased prevalence of gastroesophageal reflux in patients with idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine,* 158**,** 1804.
TORELLA, D., ROTA, M., NURZYNSKA, D., MUSSO, E., MONSEN, A., SHIRAISHI, I., ZIAS, E., WALSH, K., ROSENZWEIG, A. & SUSSMAN, M. A. 2004. Cardiac stem cell and myocyte aging, heart failure, and insulin-like growth factor-1 overexpression. *Circulation research,* 94**,** 514-524.
TSAKIRI, K. D., CRONKHITE, J. T., KUAN, P. J., XING, C., RAGHU, G., WEISSLER, J. C., ROSENBLATT, R. L., SHAY, J. W. & GARCIA, C. K. 2007. Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proceedings of the National Academy of Sciences,* 104**,** 7552-7557.
UCHIYAMA, Y., KATSURAGAWA, S., ABE, H., SHIRAISHI, J., LI, F., LI, Q., ZHANG, C.-T., SUZUKI, K. & DOI, K. 2003. Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography. *Medical Physics,* 30**,** 2440-2454.
UEDA, T., OHTA, K., SUZUKI, N., YAMAGUCHI, M., HIRAI, K., HORIUCHI, T., WATANABE, J., MIYAMOTO, T. & ITO, K. 1992. Idiopathic Pulmonary Fibrosis and High Prevalence of Serum Antibodies to Hepatitis C Virus1• 2. *Am Rev Respir Dis,* 1**,** 268-268.
UPPALURI, R., HOFFMAN, E. A., SONKA, M., HUNNINGHAKE, G. W. & MCLENNAN, G. 1999. Interstitial lung disease: a quantitative study using the adaptive multiple feature method. *American journal of respiratory and critical care medicine,* 159**,** 519-525.
VAN RIKXOORT, E. M. & VAN GINNEKEN, B. 2013. Automated segmentation of pulmonary structures in thoracic computed tomography scans: a review. *Physics in medicine and biology,* 58**,** R187.
VERSCHAKELEN, J., VAN FRAEYENHOVEN, L., LAUREYS, G., DEMEDTS, M. & BAERT, A. 1993. Differences in CT density between dependent and nondependent portions of the lung: influence of lung volume. *AJR. American journal of roentgenology,* 161**,** 713-717.
VON ZGLINICKI, T. 2002. Oxidative stress shortens telomeres. *Trends in biochemical sciences,* 27**,** 339-344.
WANG, J., LI, F. & LI, Q. 2009. Automated segmentation of lungs with severe interstitial lung disease in CT. *Medical physics,* 36**,** 4592-4599.
WARE JR, J. E. & SHERBOURNE, C. D. 1992. The MOS 36-Item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical care,* 30**,** 473-483.
WELLS, A. U. 2013. The revised ATS/ERS/JRS/ALAT diagnostic criteria for idiopathic pulmonary fibrosis (IPF)-practical implications. *Respiratory research,* 14**,** S2.
XAUBET, A., AGUSTI, C., LUBURICH, P., ROCA, J., MONTON, C., AYUSO, M. C., BARBERA, J. A. & RODRIGUEZ-ROISIN, R. 1998. Pulmonary function tests and CT scan in the management of idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine,* 158**,** 431-436.
YONEMARU, M., KASUGA, I., KUSUMOTO, H., KUNISAWA, A., KIYOKAWA, H., KUWABARA, S., ICHINOSE, Y. & TOYAMA, K. 1997. Elevation of antibodies to cytomegalovirus and other herpes viruses in pulmonary fibrosis. *European Respiratory Journal,* 10**,** 2040.

YOON, R. G., SEO, J. B., KIM, N., LEE, H. J., LEE, S. M., LEE, Y. K., SONG, J. W., SONG, J. W. & KIM, D. S. 2013. Quantitative assessment of change in regional disease patterns on serial HRCT of fibrotic interstitial pneumonia with texture-based automated quantification system. *European radiology,* 23**,** 692-701.

YUSHKEVICH, P. A., PIVEN, J., HAZLETT, H. C., SMITH, R. G., HO, S., GEE, J. C. & GERIG, G. 2006. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage,* 31**,** 1116-1128.

ZAPOROZHAN, J., LEY, S., WEINHEIMER, O., EBERHARDT, R., TSAKIRIS, I., NOSHI, Y., HERTH, F. & KAUCZOR, H.-U. 2006. Multi-detector CT of the Chest: influence of dose onto quantitative evaluation of severe emphysema a simulation study. *Journal of computer assisted tomography,* 30**,** 460-468.

ZAVALETTA, V. A., BARTHOLMAI, B. J. & ROBB, R. A. 2007. High resolution multidetector CT-aided tissue analysis and quantification of lung fibrosis. *Academic radiology,* 14**,** 772-787.

ZHANG, J., BRUESEWITZ, M. R., BARTHOLMAI, B. J. & MCCOLLOUGH, C. H. 2008. Selection of appropriate computed tomographic image reconstruction algorithms for a quantitative multicenter trial of diffuse lung disease. *Journal of computer assisted tomography,* 32**,** 233-237.

APPENDICES

APPENDIX A: CASE RECORD FORM

NB: pages 8–16 of the CRF have not been reproduced as they relate to later study visits and this thesis deals only with the first (baseline) visit.

QUIC Case Record Form (CRF)

PLEASE FILL IN STUDY NUMBER AT THE BOTTOM OF EACH SHEET

1st. Study visit t = 0 months D D / M M / Y Y Y Y

Date of informed consent	D D / M M / Y Y Y Y
Date of Assessment	D D / M M / Y Y Y Y
Year of birth	Y Y Y Y
Age at enrolment	Y Y
Gender	☐ Male ☐ Female
Consent form signed	☐ Yes ☐ No

Section 1-1 Inclusion/Exclusion Criteria

Diagnosis of IIP based on histological, clinical and/or CT appearances	☐ Yes	☐ No
Age over 40 years.	☐ Yes	☐ No
Pregnant or breastfeeding	☐ Yes	☐ No
Able to provide informed consent	☐ Yes	☐ No
Able to hold breath for five seconds	☐ Yes	☐ No
Able to perform pulmonary function tests	☐ Yes	☐ No
Able to perform exercise tests	☐ Yes	☐ No
Able to lie flat	☐ Yes	☐ No

(tick box – exclude from study if any of the bold boxes are ticked)

Section 1-2 Previous Medical History

CONDITION (list one per line)	MONTH/YEAR DIAGNOSED	Tick box if condition still active
.....	M M / Y Y Y Y	☐
.....	M M / Y Y Y Y	☐
.....	M M / Y Y Y Y	☐
.....	M M / Y Y Y Y	☐
.....	M M / Y Y Y Y	☐

☐☐☐☐ Study no 1 of 17 1st. Study visit

Section 1-3 Current Medication

DRUG	DOSE	FREQUENCY	START DATE
			M M / Y Y Y Y
			M M / Y Y Y Y
			M M / Y Y Y Y
			M M / Y Y Y Y
			M M / Y Y Y Y
			M M / Y Y Y Y
			M M / Y Y Y Y
			M M / Y Y Y Y
			M M / Y Y Y Y
			M M / Y Y Y Y
			M M / Y Y Y Y
			M M / Y Y Y Y
			M M / Y Y Y Y
			M M / Y Y Y Y
			M M / Y Y Y Y
			M M / Y Y Y Y
			M M / Y Y Y Y
			M M / Y Y Y Y

Section 1-4 Drugs discontinued in last five years

DRUG	DATE STARTED	DATE STOPPED	REASON STOPPED
	M M / Y Y Y Y	M M / Y Y Y Y	
	M M / Y Y Y Y	M M / Y Y Y Y	
	M M / Y Y Y Y	M M / Y Y Y Y	
	M M / Y Y Y Y	M M / Y Y Y Y	
	M M / Y Y Y Y	M M / Y Y Y Y	
	M M / Y Y Y Y	M M / Y Y Y Y	

Section 1-5 Duration of disease, and smoking

When did you first experience breathlessness, cough or other breathing symptoms? Please estimate month and year — M M / Y Y Y Y

Date of last normal CXR — M M / Y Y Y Y

Date of first abnormal CXR or CT — M M / Y Y Y Y

Do you smoke currently, or have you in the last month? ☐ Yes ☐ No

If you answered no, have you ever smoked? ☐ Yes ☐ No ☐ N/A

If you have ever smoked, what was your age when you started? Y Y ☐ N/A

If you have ever smoked, what was your age when you stopped? Y Y ☐ N/A

How many cigarettes a day did you get through on average (even if you did not smoke them all the way)? ☐☐ ☐ N/A

Total pack years ☐☐

Have you ever smoked drugs or agents other than tobacco? ☐ Yes ☐ No

If so, please list

Have you ever injected drugs? ☐ Yes ☐ No

If so, which ones?

Section 1-6 Occupation and exposures

At what age did you leave school? Y Y

Please list all the jobs you have had, including the company you worked for if a factory etc and the approximate ages you started and ended employment

JOB	COMPANY	DATES OR AGES
		-
		-
		-
		-
		-

Have you ever kept, or cared for, birds including caged birds or racing birds?	☐ Yes	☐ No
If so, please describe the type of bird.		
Have you worked on a farm, kept large animals (eg horses) or been exposed to moulds?	☐ Yes	☐ No
Have you ever worked with or been exposed to asbestos?	☐ Yes	☐ No
Have you ever worked in coal mining?	☐ Yes	☐ No
Have you ever been exposed to rock dust or particles of sand, such as fettling or foundry work?	☐ Yes	☐ No
Have you worked machining wood?	☐ Yes	☐ No
Have you worked machining metals?	☐ Yes	☐ No
Have you ever worked as a welder?	☐ Yes	☐ No
If so, what type or types of welding have you done (MIG/TIG etc)		
If so, what types of metal have you welded?		
Have you ever sprayed paints?	☐ Yes	☐ No
Have you ever worked in hairdressing?	☐ Yes	☐ No
Have you worked with coolant oil?	☐ Yes	☐ No

Section 1-7 Exposure to specific medications

Have you ever taken chemotherapy, or other tablets or medicines to treat cancer? ☐ Yes ☐ No

If so, please list the drugs or agents and the type of cancer they were for

Have you ever taken drugs prescribed by a doctor to treat arthritis? ☐ Yes ☐ No

If so please list them

Have you ever taken amiodarone (cordarone)? ☐ Yes ☐ No

Section 1-8 Rheumatological features

Have you ever been seen by a rheumatologist (joint specialist) in hospital or in clinic? ☐ Yes ☐ No

If so, what diagnosis were you given?

Do you consider yourself to suffer from any of the following conditions: rheumatoid arthritis, systemic sclerosis, scleroderma, myositis, polymyositis, dermatomyositis, Sjögren's syndrome, SLE, lupus or Raynaud's? ☐ Yes ☐ No

If so, please list

Do your joints hurt or ache regularly? ☐ Yes ☐ No

If so, which joints?

If so, is this usually worse when you wake up, or at the end of the day? ☐ When you wake ☐ End of day

Are any joints stiff when you wake up in the morning? ☐ Yes ☐ No

If so, on average how many minutes does it take before they loosen up?

Do any joints swell? ☐ Yes ☐ No

If so, which ones?

Do you fingers ever change colour in the cold (go white)? ☐ Yes ☐ No

If so, do they then go blue and/or red? ☐ Yes ☐ No ☐ N/A

Do you ever get a rash (other than sunburn) in areas of the skin exposed to sunlight? ☐ Yes ☐ No

Have you lost weight without wanting to or dieting? ☐ Yes ☐ No

If so, how much weight have you lost?		
Do you experience dryness or grittiness of the eyes?	☐ Yes	☐ No
Do you experience troublesome dryness of the mouth?	☐ Yes	☐ No
Do you get frequent or troublesome mouth ulcers?	☐ Yes	☐ No
Have you experienced unexpected hair loss?	☐ Yes	☐ No
Do you experience pain or other difficulty swallowing?	☐ Yes	☐ No
Do you experience a foul taste, acid or bile in the mouth, or other features of 'reflux'	☐ Yes	☐ No
Do you suffer with a rash or rashes?	☐ Yes	☐ No
If so, where on the body?		
Do you suffer from fevers?	☐ Yes	☐ No
Do you experience weakness of the muscles?	☐ Yes	☐ No
Do you struggle to get up from a chair without using your arms, or difficulty climbing the stairs without using the banister?	☐ Yes	☐ No

Section 1-9 Examination features

Height (m)	M . cm cm
Weight (kg)	kg kg kg
Oxygen saturations on air (%)	% %
Digital clubbing?	☐ Yes ☐ No
Deforming arthritis of the hands?	☐ Yes ☐ No
Non-bony joint swelling?	☐ Yes ☐ No
Rashes over knuckles 'Gottren's papules'?	☐ Yes ☐ No
Sclerodactyly? (tight, tethered skin over fingers)	☐ Yes ☐ No
Scleroderma (tight, tethered skin over the skin proximal to MCPs)	☐ Yes ☐ No
Calcinosis? (white spots in the skin of the fingers)	☐ Yes ☐ No
Digital infarcts?	☐ Yes ☐ No

Rash or discolouration around eyes (Heliotrope rash)	☐ Yes	☐ No
'Shawl' rash? (over nape of neck and upper back)	☐ Yes	☐ No
V-rash? (rash over front of neck)	☐ Yes	☐ No
Telangiectasiae	☐ Yes	☐ No
Basal crackles?	☐ Yes	☐ No
Cor pulmonale?	☐ Yes	☐ No

Section 1-10 Laboratory investigations and diagnosis

ANA	☐ Positive	☐ Negative			
ENA profile if ANA positive					
Rheumatoid factor	☐ Positive	☐ Negative	☐ Value		
anti-CCP titre	☐ Positive	☐ Negative	☐ Value		
Creatinine kinase	☐ Raised	☐ Not raised	☐ Value		
Full blood count taken? (report low Hb, neutrophils or platelets or eosinophilia only)	☐ Yes, normal	☐ No	☐ Abnormal Value		
Urea and electrolytes taken?	☐ Yes, normal	☐ No	☐ Abnormal Value		
Liver function tests taken?	☐ Yes, normal	☐ No	☐ Abnormal Value		
PAXgene samples taken?	☐ Yes	☐ No			
Research DNA samples taken?	☐ Yes	☐ No			
Research serum samples taken?	☐ Yes	☐ No			
Research plasma samples taken?	☐ Yes	☐ No			
Has patient had a surgical lung biopsy?	☐ Yes	☐ No			
If so, what is the diagnosis?					
ILD MDT consensus diagnosis					

Section 1-11 Quality of Life

St George's Respiratory Questionnaire completed? ☐ Yes ☐ No

Section 1-12 Physiology

Full PFTs performed? Please fill in results in appendix	☐ Yes	☐ No
Shuttle walk test?	☐ Yes	☐ No
Cardiopulmonary exercise test performed?	☐ Yes	☐ No

Section 1-13 CT

CT performed?	☐ Yes	☐ No

Section 1-4 Additional Information

Any additional information? ☐ Yes ☐ No

If yes, please describe

Appendix Serial PFTs

Visit no	1		2		3		4		5	
Date	D D / N N / Y Y Y Y		D D / N N / Y Y Y Y		D D / N N / Y Y Y Y		D D / N N / Y Y Y Y		D D / N N / Y Y Y Y	
	Value	% pred	Value	% pred	Value	% pred	Value	% pred	Value	% pred
FEV_1	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □
FVC	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □
TLC	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □
RV	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □
TLco	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □
Kco	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □	□ . □□	□□ . □

APPENDIX B: ST GEORGE'S RESPIRATORY QUESTIONNAIRE (STGRQ)

St George's Respiratory Questionnaire – first page

ST. GEORGE'S RESPIRATORY QUESTIONNAIRE
ORIGINAL ENGLISH VERSION

ST. GEORGE'S RESPIRATORY QUESTIONNAIRE (SGRQ)

This questionnaire is designed to help us learn much more about how your breathing is troubling you and how it affects your life. We are using it to find out which aspects of your illness cause you most problems, rather than what the doctors and nurses think your problems are.

Please read the instructions carefully and ask if you do not understand anything. Do not spend too long deciding about your answers.

Before completing the rest of the questionnaire:

Please tick in one box to show how you describe your current health:

Very good	Good	Fair	Poor	Very poor
☐	☐	☐	☐	☐

Copyright reserved
P.W. Jones, PhD FRCP
Professor of Respiratory Medicine,
St. George's University of London,
Jenner Wing,
Cranmer Terrace,
London SW17 ORE, UK.

Tel. +44 (0) 20 8725 5371
Fax +44 (0) 20 8725 5955

St. George's Respiratory Questionnaire
PART 1

Questions about how much chest trouble you have had over the past 3 months.

Please tick (✓) *one box for each question*:

	most days a week	several days a week	a few days a month	only with chest infections	not at all
1. Over the past 3 months, I have coughed:	☐	☐	☐	☐	☐
2. Over the past 3 months, I have brought up phlegm (sputum):	☐	☐	☐	☐	☐
3. Over the past 3 months, I have had shortness of breath:	☐	☐	☐	☐	☐
4. Over the past 3 months, I have had attacks of wheezing:	☐	☐	☐	☐	☐

5. During the past 3 months how many severe or very unpleasant attacks of chest trouble have you had?

Please tick (✓) *one*:

- more than 3 attacks ☐
- 3 attacks ☐
- 2 attacks ☐
- 1 attack ☐
- no attacks ☐

6. How long did the worst attack of chest trouble last?
(Go to question 7 if you had no severe attacks)

Please tick (✓) *one*:

- a week or more ☐
- 3 or more days ☐
- 1 or 2 days ☐
- less than a day ☐

7. Over the past 3 months, in an average week, how many good days (with little chest trouble) have you had?

Please tick (✓) *one*:

- No good days ☐
- 1 or 2 good days ☐
- 3 or 4 good days ☐
- nearly every day is good ☐
- every day is good ☐

8. If you have a wheeze, is it worse in the morning?

Please tick (✓) *one*:

- No ☐
- Yes ☐

continued…

St. George's Respiratory Questionnaire
PART 2

Section 1

How would you describe your chest condition?

Please tick (✓) *one*:

- The most important problem I have ☐
- Causes me quite a lot of problems ☐
- Causes me a few problems ☐
- Causes no problem ☐

If you have ever had paid employment.

Please tick (✓) *one*:

- My chest trouble made me stop work altogether ☐
- My chest trouble interferes with my work or made me change my work ☐
- My chest trouble does not affect my work ☐

Section 2

Questions about what activities usually make you feel breathless <u>these days</u>.

Please tick (✓) in ***each box*** that applies to you ***these days***:

	True	False
Sitting or lying still	☐	☐
Getting washed or dressed	☐	☐
Walking around the home	☐	☐
Walking outside on the level	☐	☐
Walking up a flight of stairs	☐	☐
Walking up hills	☐	☐
Playing sports or games	☐	☐

continued…

f:\institut\cultadap\project\gsk1881\question\final versions\sgrqoriq.doc 14/03/03

St. George's Respiratory Questionnaire
PART 2

Section 3

Some more questions about your cough and breathlessness <u>these days</u>.

	Please tick (✓) in ***each box*** that applies to you ***these days***: True	False
My cough hurts	☐	☐
My cough makes me tired	☐	☐
I am breathless when I talk	☐	☐
I am breathless when I bend over	☐	☐
My cough or breathing disturbs my sleep	☐	☐
I get exhausted easily	☐	☐

Section 4

Questions about other effects that your chest trouble may have on you <u>these days</u>.

	Please tick (✓) in ***each box*** that applies to you ***these days***: True	False
My cough or breathing is embarrassing in public	☐	☐
My chest trouble is a nuisance to my family, friends or neighbours	☐	☐
I get afraid or panic when I cannot get my breath	☐	☐
I feel that I am not in control of my chest problem	☐	☐
I do not expect my chest to get any better	☐	☐
I have become frail or an invalid because of my chest	☐	☐
Exercise is not safe for me	☐	☐
Everything seems too much of an effort	☐	☐

Section 5

Questions about your medication, if you are receiving no medication go straight to section 6.

	Please tick (✓) in ***each box*** that applies to you ***these days***: True	False
My medication does not help me very much	☐	☐
I get embarrassed using my medication in public	☐	☐
I have unpleasant side effects from my medication	☐	☐
My medication interferes with my life a lot	☐	☐

continued…

f:\institut\cultadap\project\gsk1881\question\final versions\sgrqoriq.doc 14/03/03

St. George's Respiratory Questionnaire
PART 2

Section 6

These are questions about how your activities might be affected by your breathing.

Please tick (✓) in ***each box*** that applies to you ***because of your breathing***:

	True	False
I take a long time to get washed or dressed	☐	☐
I cannot take a bath or shower, or I take a long time	☐	☐
I walk slower than other people, or I stop for rests	☐	☐
Jobs such as housework take a long time, or I have to stop for rests	☐	☐
If I walk up one flight of stairs, I have to go slowly or stop	☐	☐
If I hurry or walk fast, I have to stop or slow down	☐	☐
My breathing makes it difficult to do things such as walk up hills, carrying things up stairs, light gardening such as weeding, dance, play bowls or play golf	☐	☐
My breathing makes it difficult to do things such as carry heavy loads, dig the garden or shovel snow, jog or walk at 5 miles per hour, play tennis or swim	☐	☐
My breathing makes it difficult to do things such as very heavy manual work, run, cycle, swim fast or play competitive sports	☐	☐

Section 7

We would like to know how your chest <u>usually</u> affects your daily life.

Please tick (✓) in ***each box*** that applies to you ***because of your chest trouble***:

	True	False
I cannot play sports or games	☐	☐
I cannot go out for entertainment or recreation	☐	☐
I cannot go out of the house to do the shopping	☐	☐
I cannot do housework	☐	☐
I cannot move far from my bed or chair	☐	☐

continued…

St. George's Respiratory Questionnaire

Here is a list of other activities that your chest trouble may prevent you doing. (You do not have to tick these, they are just to remind you of ways in which your breathlessness may affect you):

Going for walks or walking the dog
Doing things at home or in the garden
Sexual intercourse
Going out to church, pub, club or place of entertainment
Going out in bad weather or into smoky rooms
Visiting family or friends or playing with children

Please write in any other important activities that your chest trouble may stop you doing:

. . . . . .
. . . . . .
. . . . . .
. . . . . .

Now would you tick in the box (one only) which you think best describes how your chest affects you:

It does not stop me doing anything I would like to do ☐
It stops me doing one or two things I would like to do ☐
It stops me doing most of the things I would like to do ☐
It stops me doing everything I would like to do ☐

Thank you for filling in this questionnaire. Before you finish would you please check to see that you have answered all the questions.

APPENDIX C: RADIOLOGIST SCORING SHEET

1 cm above the dome of the right hemi-diaphragm (image 2)	**Right lung**	**Left Lung**	**Total lung**
			(slice only)
honeycombing (%)			
reticulation (%)			
ground glass (%)			
consolidation (%)			
emphysema (Y/N)			
2 cm below the dome of the right hemidiaphragm (image 1)			
honeycombing (%)			
reticulation (%)			
ground glass (%)			
consolidation (%)			
emphysema (Y/N)			
Other comments:			
TIME FINISHED			

STUDY ID:	**Right lung**	**Left Lung**	**Total lung**
TIME STARTED:			**(slice only)**
Arch of the aorta (image 5)			
honeycombing (%)			
reticulation (%)			
ground glass (%)			
consolidation (%)			
emphysema (Y/N)			
Carina (image 4)			
honeycombing (%)			
reticulation (%)			
ground glass (%)			
consolidation (%)			
emphysema (Y/N)			
Right superior pulmonary vein (image 3)			
honeycombing (%)			
reticulation (%)			
ground glass (%)			
consolidation (%)			
emphysema (Y/N)			

APPENDIX D: RADIOLOGIST SCORING INSTRUCTIONS

Radiologist scoring instructions:

At the start of each patient scoring, please put study ID (eg QUIC 2) and start time at top of sheet.

For each level, estimate the percent of each texture (eg honeycombing) for the right and left lung separately (eg % honeycombing as fracture of right lung volume).

All percentage estimation should be to the nearest 5%.

All lung not described by one of the textures is assumed to be normal. If there is another significant abnormal texture eg atelectasis, nodules, please record this in the comments section.

Record whether emphysema is present or not (Y/N).

Now try and estimate the percent of each texture for each whole slice. This estimate should reflect the differences in volumes of the left and right lungs at that level eg if left lung is half volume of right and right lung has no honeycombing, 50% involvement of the left lung may give an overall involvement of 25%.

At the end, record the finish time.