**Chapter aim**

{bartholmai2013quantitative}

The purpose of the effort to develop software for analysis of thoracic HRCT is to provide a reproducible way to quantify and characterize the extent of diffuse parenchymal disease. We hypothesize that the results of parenchymal analysis can be correlated with radiologist visual assessment, accepted physiologic features of pulmonary disease, and clinical outcomes. The quantity, type and distribution of abnormalities by an automated tool should provide utility in clinical practice by aiding in non-invasive diagnostic determination, detecting change in disease over time, and stratifying risk of progression or mortality.

**Aim**

{maldonado2013automated}

Accurate assessment of prognosis in idiopathic pulmonary fibrosis remains elusive due to significant individual radiological and physiological variability.

However, the utility of such longitudinal physiological indices for individual patient management is still significantly limited by substantial intra-individual variability, perhaps with confounding by coexisting conditions, such as emphysema or pulmonary hypertension \citep{ cottin2009syndrome, nathan2007idiopathic}.

{bartholmai2013quantitative}

High-Resolution chest CT (HRCT) is essential in the characterization of interstitial lung disease (ILD). The HRCT features of some diseases can be diagnostic. Longitudinal monitoring with HRCT can assess progression of ILD; however, subtle changes in the volume and character of abnormalities can be difficult to assess. Accuracy of diagnosis can be dependent on expertise and experience of the radiologist, pathologist or clinician. Quantitative analysis of thoracic HRCT has the potential to determine the extent of disease reproducibly, classify the types of abnormalities and automate the diagnostic process.

However, these distinctive disease processes often have a very similar clinical phenotype, and can also have indeterminate pathologic and radiographic appearances. In addition, some patients can have mixed restrictive/fibrotic and destructive/obstructive processes, such as in combined pulmonary fibrosis and emphysema syndrome \citep{cottin2011combined}, which confound physiologic testing and for which biopsy results can vary wildly based on the location of sampling. The fundamental clinical problems of how to consistently detect, characterize, and differentiate the various ILDs remain diagnostic challenges. Yet, all of the various ILD and mixed parenchymal diseases have distinctly different prognoses and opportunities for therapy \citep{lynch2005high} and it is increasingly clear that specific therapy targeted on a particular pathological process will be key to altering the progression of these generally inexorable diseases. Furthermore, how even known processes can be consistently characterized and quantified over time and, more importantly, how these changes can predict prognosis all remain largely unanswered questions. The promise of early and appropriate application of therapy so as to significantly affect prognosis has not been achieved despite many efforts, and assessment of response to therapy has been difficult in even well-designed, randomized prospective trials – in part due to lack of robust non-invasive biomarkers for ILD.

{raghunath2014quantitative}

Individualized treatment strategies based on patient-specific disease manifestations are the ultimate goal of modern pulmonary medicine. In reality, this lofty goal is unachievable with our current medical knowledge. Obstacles include lack of established diagnostic, prognostic and predictive patient-specific biomarkers. Nevertheless, with present day technologies and innovative strategies the practice of stratified medicine is becoming more feasible. Patient populations can be stratified based on quantifiable characteristics that define the disease and characterize key physiologic, pathologic, anatomic or patient-reported factors that impact quality of life, morbidity and mortality. The goal of stratified medicine entails the ability to guide individualized patient care based on recognized disease characteristics and established prognostic features of matched group [6]. However, the definitions of disease phenotypes and prognostic and predictive biomarkers to guide patient management remain elusive for DPLD \citep{cesuroglu2012public}.

**Advantage of quantitative analysis**

{maldonado2013automated}

A number of other studies suggest that the extent of fibrosis assessed semiquantitatively on HRCT is a strong predictor of outcomes in IPF \citep{best2008idiopathic,wells2003idiopathic, sumikawa2008computed}. Correlation of quantitative estimates of pulmonary fibrosis by automated analysis to mortality has been the object of few reports \citep{best2008idiopathic, best2003quantitative, iwasawa2009assessment}.

The use of longitudinal quantitative indices of lung fibrosis as identified on HRCT could represent an appealing alternative to physiological measures. An accurate and reproducible method allowing monitoring of fibrosis progression on HRCT would be a valuable surrogate marker of disease. Unfortunately, assessment of fibrosis volumes by expert radiologists has been hampered by substantial intra- and interobserver variability, and quantitative CT indices using fractal analysis and global histogram-based methods have not been validated or found helpful in clinical practice thus far \citep{ best2008idiopathic, wells2003idiopathic, iwasawa2009assessment, coxson1997quantification, kauczor2000automatic, kazerooni1997thin, lynch2005high, uppaluri1999interstitial}.

{raghunath2014quantitative}

Diffuse parenchymal lung diseases (DPLDs) are characterized by widespread pathological changes within the pulmonary tissue that impair the elasticity and gas exchange properties of the lungs. Clinical-radiological diagnosis of these diseases remains challenging and their clinical course is characterized by variable disease progression. These challenges have hindered the introduction of robust objective biomarkers for patient-specific prediction based on specific phenotypes in clinical practice for patients with DPLD. Therefore, strategies facilitating individualized clinical management, staging and identification of specific phenotypes linked to clinical disease outcomes or therapeutic responses are urgently needed. A classification schema consistently reflecting the radiological, clinical (lung function and clinical outcomes) and pathological features of a disease represents a critical need in modern pulmonary medicine.

**Advantage of HRCT**

{bartholmai2013quantitative}

High-Resolution chest CT (HRCT) is essential in the characterization of interstitial lung disease (ILD). The HRCT features of some diseases can be diagnostic. Longitudinal monitoring with HRCT can assess progression of ILD; however, subtle changes in the volume and character of abnormalities can be difficult to assess. Accuracy of diagnosis can be dependent on expertise and experience of the radiologist, pathologist or clinician. Quantitative analysis of thoracic HRCT has the potential to determine the extent of disease reproducibly, classify the types of abnormalities and automate the diagnostic process.

{bartholmai2013quantitative}

The general rationale for developing a method for quantitative analysis of the lung parenchyma derives from the knowledge that the generic term ‘ILD’ includes multiple different diseases with numerous different imaging features in variable distribution. In many cases, the patterns and distribution of disease may allow differentiation of these processes. Specifically, HRCT is able to demonstrate features of disease, with visual patterns such ground glass opacities, reticular infiltrates and honeycombing that enable the differential diagnosis to be narrowed \citep{mueller2007every} . The impact of HRCT on clinical management is reflected in the decreased need for surgical lung biopsy between the diagnostic workflow proposed in the ATS/ERS consensus statement in 2002 \citep{travis2002american} and more recent evidence-based guidelines for diagnosis and management in clinical practice \citep{raghu2011official, peikert2008assessment}. Even in cases where the HRCT images lack the specific features that reflect the cellular infiltration, fibrosis and architectural distortion with honeycombing typical of a process such as UIP, HRCT remains a useful non-invasive technique to reveal the abnormal parenchymal densities resulting from microscopic morphological changes \citep{raghu1999accuracy, kazerooni2001high, misumi2006idiopathic, diette2005high, lynch2005idiopathic, zerhouni1990high, costabel2007diffuse} and can provide guidance for the optimal site to obtain a biopsy of characteristic or active disease \citep{ diette2005high} . Similarly, it is generally accepted that the extent of visual abnormalities correlates with extent of pathological involvement as well as the severity of physiologic abnormalities \citep{ kazerooni1997thin} , and therefore longitudinal HRCT can be useful in the monitoring of disease progression and response to therapy \citep{ saketkoo2011developing, kim1999nonspecific, wells2003idiopathic} .

**Difficulties of diagnosis**

{bartholmai2013quantitative}

The complex morphological patterns of ILD that can change in extent and/or appearance over time can be challenging to assess. Similarly, manual classification and evaluation of extent is tedious and not reproducible. Evaluation is complicated by significant in inter- and intra-observer variation in the diagnosis of ILD \citep{ flaherty2007idiopathic, watadani2013interobserver}. Clearly, there are differences in the skill and experience of physicians involved in the process of determining a diagnosis, but there are more complex inherent differences in perception, interpretation of visual features of disease and ‘reader error’ which lead to variable description of findings and characterization of disease that may not be surmountable by any degree of training or advances in imaging technology \citep{ garland1949scientific, tuddenham1969roentgen, kundel2006history}. Specifically, the final clinical diagnosis of ILD has been shown to be variable based on the training and experience of the radiologist, clinician and pathologist working independently or involved in the multidisciplinary evaluation of disease despite the presence of accepted criteria and clinical pathways even in a well- controlled investigation with well-characterized subjects. These variations are even more prevalent in real-world patient care within- and across-specialty physicians from academic and community health centers \citep{ flaherty2004idiopathic} . Even diagnosis by consensus of multiple experts and utilization of continuous learning techniques, which do improve consistency to some degree, does not assure dependable results \citep{ sverzellati2011method}.

{raghunath2014quantitative}

Patients with DPLD represent a substantial health-care burden due to high disease prevalence, chronic nature of such processes and lack of curative therap semi-quantitative y with associated morbidity and premature mortality \citep{olson2007mortality, mannino2007global}. Despite advances in modern medicine, including medical imaging, pathology and genomics, diagnosis and treatment of DPLD remain difficult. The American Thoracic Society / European Respiratory Society \citep{raghu2011official, travis2013official} strongly recommends a multidisciplinary approach to diagnosis in DPLDs based on consensus of radiologists, clinicians and pathologists. Unfortunately, a successful classification scheme that allows recognition of disease groups or even consensus diagnosis of specific diseases identically across radiology, pulmonary and pathology disciplines remains ambiguous \citep{hansell2013classification}.

**Quantitative method**

{bartholmai2013quantitative}

The variability in clinical evaluation of ILD is an opportunity for automation, computer-aided detection and quantitative image analysis. Specifically, there are opportunities to use current image processing technology on CT data to optimize detection of the abnormalities recognized to represent pathological changes on HRCT and enable reproducible quantification and characterization of the manifestations of ILD. These quantitative results have promise as biomarkers and could result in more consistent diagnosis, more sensitive disease monitoring and accurate determination of prognosis \citep{lynch2007quantitative}. There have been many different efforts using different methods to robustly characterize the discernible patterns in chest x-rays and CT scans over the last several decades but the assessment of results from some early texture-based analysis remains generally unchanged: “…classification and quantification of interstitial lung disease is difficult, and even experienced chest radiologists frequently struggle with differential diagnosis. Automated schemes that indicate a percentage of affected lung or the probability of a certain disease would certainly be welcome, but require much more research” \citep{delorme1997usual}. Encouragingly, quantitative methods used to assess the severity of emphysema and other features of COPD on CT have become more robust over the last 20 years, with evolution and optimization of various techniques. These quantitative analysis results can play a significant role as biomarkers used in the diagnosis of COPD phenotypes, assessment of disease progression and prognosis \citep{ galban2012computed}. However, since ILD changes on HRCT are even more complicated than the density changes seen with emphysema, the development of quantitative CT-based measures for lung fibrosis has been more challenging \citep{ lynch2007quantitative, delorme1997usual, galban2012computed, depeursinge2010comparative} and results less encouraging. In general, simple methods based on pixel counting, first order features based on density masks or whole-lung histogram analysis, methods using multiple higher-order features or texture methods, and even more sophisticated classification techniques including continuous learning with physician-in-the-loop are only partially successful for evaluation of specific diseases or even simple determination of normal vs. abnormal regions \citep{ depeursinge2010comparative, uppaluri1999interstitial, shyu1999assert}. Furthermore, many of the current methods used in evaluation of ILD for research purposes are computationally intense and require processing times that may take hours or even days. These real-world limitations make those techniques difficult to translate into routine clinical practice.

{bartholmai2013quantitative}

Given that there are visually distinct normal anatomic and disease-specific morphological manifestations apparent on HRCT scans, parenchymal classification is typically approached though texture analysis, computer vision-based image understanding of volumetric histogram features and 3D morphology of the classified voxels. In general, the classifier results are determined by methods used to solve content-based information retrieval problems \citep{delorme1997usual, uppaluri1999computer, uchiyama2003quantitative, hoffman2003characterization, chabat2003obstructive, kim2005computer, zavaletta2007high}. Central to all these schemes is the selection of representative expert-labeled volumes of interest (VOIs) as a training set for a classifier. The features of these VOIs are then used by the classifier on subsequent input to reproduce the expert labels. Descriptors based on histogram statistics, co-occurrence matrices, run length parameters, and fractal measures are typically used to enumerate the features. Artificial neural networks, Bayesian classifiers, support vector machines and k-neighborclassifiers are used to classify the features \citep{ van2002automatic, xu2006computer, arzhaeva2007computer}.

{raghunath2014quantitative}

High-resolution computed tomography (HRCT) is the preferred radiologic imaging modality for evaluating lungs. In current clinical practice, HRCT adds tremendous value in its ability to diagnose and manage patients with DPLDs and may often obviate or direct specific targets for surgical lung biopsies \citep{hayhurst1984diagnosis, lynch2005idiopathic, xaubet1998pulmonary, akira2009usual, mooney2013radiographic, flaherty2003prognostic, macdonald2001nonspecific}. It can be both diagnostic and prognostic for some pathological processes. Several CT-based automated quantitative methods (quantitative CT - QCT) have been proposed to quantify and characterize parenchymal abnormalities \citep{bartholmai2013quantitative, maldonado2013automated, uppaluri1999interstitial, xu2006mdct, castaldi2013distinct}. QCT has demonstrated correlation of quantified parenchymal patterns with well-accepted clinical endpoints – physiologic indices, visual radiology scores, as well as prognostic outcomes for subsets of diseases \citep{maldonado2013automated, castaldi2013distinct, kim2013chronic, gietema2011quantifying}. Furthermore, correlation of quantified parenchymal abnormalities with pathologic features has resulted in confident interpretation and diagnosis of certain diseases (e.g., usual interstitial pneumonia) obviating the need for surgical lung biopsy \citep{raghu2011official}. The quantitative nature of radiological image-based biomarkers \citep{sullivan2008imaging} enables development of an automated and consistent image-based methodology to achieve objective population stratification.

My paper

In the past few years, there has been considerable effort to provide quantitative analysis on CT scans of lung parenchymal abnormalities. Renuka et al \citep{ uppaluri1999interstitial, uppaluri1999computer} was an early group to present a computer aided diagnosis (CAD) method to quantify lung tissues based upon HRCT. An adaptive multiple feature method (AMFM) which combined statistical texture measures and a fractal measure was developed to assess CT features for classifying a tissue pattern. Alan et al \citep{ best2003quantitative, best2008idiopathic} used mean lung attenuation (MLA), skewness (asymmetry) and kurtosis (peakedness) as quantitative CT indexes and furtherly used univariate and multiple correlation and regression statistical analyses to determine relationships between histogram features and results of PFTs. Hyun et al \citep{ kim2010computer, kim2011quantitative, kim2015comparison} published a series of papers presenting a texture-based CAD scoring system to assess quantitative lung fibrosis (QLF) as a measurement of lung disease severity and as a surrogate imaging marker. Brian et al [16-18] developed a software for analysis of thoracic HRCT through classifying CT imaging into texture patterns such as honeycomb, reticular and ground-glass, thus providing a reproducible way to quantify and characterize lung parenchymal disease. However, the current published methods mainly focus on the global analysis of each CT pattern or texture-based index as a whole lung (such as the percentile analysis or the correlation between indexes), but seldomly characterize the spatial distribution of each diseased region or the change extent of these abnormalities.

**CALIPER introduction**

{maldonado2013automated}

CALIPER (Computer-Aided Lung Informatics for Pathology Evaluation and Ratings) is an image analysis tool developed by the Biomedical Imaging Resource Laboratory at the Mayo Clinic Rochester (Rochester, MN, USA) for the characterisation and quantification of lung parenchymal findings on high-resolution computed tomography (HRCT). The detection and quantification of pulmonary parenchyma by CALIPER is based on histogram signature mapping techniques trained through expert radiologist consensus assessment of pathologically confirmed training sets obtained through the Lung Tissue Research Consortium (LTRC) \citep{zavaletta2007high}. We hypothesised that this novel computer-aided method of ana pairwise dissimilarity metrics lysing pulmonary parenchymal features on chest HRCT would provide a reproducible and accurate assessment of disease that may correlate with the semi-quantitative assessment of radiologists. We further postulated that the overall extent or short-term changes in parenchymal features as detected by CALIPER or radiologists may correlate with and be independently predictive of survival in patients with IPF.

{bartholmai2013quantitative}

Towards addressing the need for improved quantitative analysis tools that can be utilized for both research analysis and clinical practice, a multidisciplinary team of Mayo Clinic clinicians, scientists and engineers at the Mayo Biomedical Imaging Resource have developed CALIPER (Computer Aided Lung Informatics for Pathology Evaluation and Rating): a computational platform for the near-real-time characterization and quantification of lung parenchymal patterns on CT scans. The following briefly outlines the computational components and methods utilized to test the efficacy of CALIPER.

**CALIPER method**

{maldonado2013automated}

The data processing step included lung extraction and segmentation with identification and extraction of central airways and central vascular structures, manual segmentation of the anatomical lobes, and classification of the remaining pulmonary parenchyma. Segmentation of the lungs was achieved using an adaptive density-based morphological approach \citep{hu2001automatic}. Airways were segmented using iterative three-dimensional region growing and connected components analysis. Pulmonary vessels were extracted using an optimised multi-scale tubular structure enhancement filter \citep{shikata2009segmentation}. Additional sub-segmentation into anatomical lobes by manual tracing technique was performed by a thoracic radiologist (B.J. Bartholmai) and determination of central/peripheral regions was completed by three-dimensional volume erosion techniques, such that the peripheral zone represented ,50% of the total volume of each lung. Separate inclusion of the perihilar regions into the central zone of each lobe was performed by semi-automated determination of the hila through tracheal tree analysis and inclusion of a spherical region of 5 cm, bilaterally, from the hilar points. Parenchymal tissue type detection and quantification of the classes (normal, emphysema, ground glass, reticular and honeycombing) (fig. 1) was performed using a sliding window supervised classification scheme.

During the pre-processing supervised training phase, multiple (n5976) 15615615-pixel volumes of interest were selected through independent analysis by four subspecialty thoracic radiologists from HRCT scans of \citep{ zerhouni1990high} subjects with proven pathological diagnosis of diffuse pulmonary disease (interstitial lung disease (ILD) of emphysema) or known control subjects without ILD. The exemplar candidates with agreement on the class of abnormality by all four radiologists were used to determine canonical histogram signatures for each of the classes of visual abnormality with automatic cluster affinity techniques and those signatures of each of the visual classes were used for the volumetric classification of the HRCT data of test subjects.

{bartholmai2013quantitative}

Initial handling of data within CALIPER involves segmentation and extraction of anatomic regions as a pre-processing step towards the eventual characterization and quantification of the pulmonary parenchyma.

The detection and quantification of pulmonary parenchyma by CALIPER is based on histogram signature mapping techniques trained through expert radiologist consensus assessment of pathologically confirmed datasets obtained through the Lung Tissue Research Consortium (LTRC). The LTRC is a NIH/NHLBI sponsored, multi-site initiative dedicated to helping investigators develop a better understanding of ILD and COPD through development of a repository of clinical data (including demographics, questionnaire responses, medical history), physiologic data, pathological specimens, blood and tissue characterization, CT scan data. The LTRC also provides central expert review of the clan inical, radiologic, pathologic data. The data and tissue specimens on the well-characterized subjects in the LTRC are available at no cost to qualified investigators through a standardized process overseen by the NIH/NHLBI (requests can be made at http:// www.ltrcpublic.com/data\_requests.htm). For our training VOIs, HRCT scans from fourteen subjects were selected from the LTRC repository to create a set of 976 VOIs. The VOIs were selected through independent characterization by four subspecialty thoracic radiologists, with instructions to determine if visual appearance of 70% or more of the given VOI spanning 15×15×15 voxels was normal, contains emphysema or belonged to one of the characteristic ILD parenchymal CT patterns: ground glass opacities (GG), reticular infiltrates (RI), or honeycombing (HC). Based on this criterion 80, 150, 187, 265 and 294 VOIs were determined by consensus agreement selected to represent emphysema, ground glass, honeycombing, normal and reticular infiltrates, respectively.

Quantitative discriminability of a number of pairwise dissimilarity metrics based on the VOI histograms was examined using multi-dimensional scaling (MDS). Parametric and non- parametric dissimilarity metrics tested to evaluate the optimal dissimilarity included first and second order statistics and measures of effectiveness \citep{ sadjadi2003measures} such as Fechner-Weber contrast measure, target-reference inference ratio, Fisher distance, correlation coefficient, scale invariant normalized mean square error and normalized mutual information. Non-parametric dissimilarity metrics were based on pairwise histogram distances such as Manhattan, Euclidean, Bhattacharya, Kolmogrov-Smirnoff and Cramer Von Mises Distance (CVM), chi squared distance, Kullback-Liebler divergence, Jeffrey divergence, and histogram intersection \citep{ deza2006dictionary}. Of all the metrics, MDS representation of CVM (the squared L 2 –metric between cumulative density functions) was found to be most consistent with the expert groupings and, consequently, was chosen as the dissimilarity metric used by CALIPER in the automated classification. Figure 1 shows the three dimensional MDS projection for the pairwise CVM dissimilarity measure of the expert-characterized VOI histograms, revealing the natural orderliness with which the MDS representation of CVM metric projects the VOIs to align with expert consensus. Concordance between the consensus radiology labeling (columns) and the affinity propagation based, unsupervised clustering of the pairwise Cramer Von Mises (CVM) dissimilarity metric (rows) for the 976 Volumes of Interest used to train CALIPER (Computer Aided Lung Informatics for Pathology Evaluation and Rating) was assessed using the Kappa statistic. Multi-category analysis such as this requires the evaluation of the degree of agreement category by category and, therefore, a K × K tables method for agreement as described by Agresti \citep{ agresti2011categorical} was utilized to assess the agreement index for each category (i.e. the proportion of agreements observed).

Having established the qualitative equivalence of CVM and expert groupings and verified the results with expert visual validation, quantitative equivalence was evaluated using automatic clustering of CVM similarities. This process was utilized to group the VOIs into natural clusters. To create an unbiased stratification of VOIs into natural clusters, affinity propagation 48 was used. Affinity propagation uses message passing to iteratively find clusters, given pair-wise similarities of n-dimensional data. The clustering based on affinity propagation yielded ten natural clusters. In addition to resolving the clusters, this methodology identifies the exemplar that is most ‘central’ to each of the clusters in feature-space. The most ‘central’ within type of each cluster was identified as the fundamental type of its exemplar class. Since more natural clusters were found than classes of visual abnormalities described for ILD, multiple clusters were shown to correspond to some visual classes: 1, 2, 2, 2, and 3 respectively for emphysema, ground glass, honeycombing, normal and reticular.

The local histograms computed from the 15×15×15 neighborhood of each of the parenchymal voxel were compared against the histogram of the 34 exemplars identified in the training phase. CVM dissimilarity measure was used in the comparison and the fundamental type of the exemplar with the least CVM distance was assigned as the parenchymal class of the underlying voxel. The number of voxels belonging to each of the parenchymal classes was calculated across the whole lung and the individual lungs. The voxels identified as vessels were included as normal to account for the total lung volume.

{raghunath2014quantitative}

The input data for this study is the CALIPER-based quantification of parenchymal CT patterns. Briefly, CALIPER processes and characterizes the CT dataset by isolating the lung parenchyma and classifying every parenchymal voxel into one of the following characteristic CT patterns: normal (N), reticular (R), honeycomb (HC), ground-glass (GG), mild low attenuation areas (LAA), moderate LAA and severe LAA. The reliability of CALIPER-based classification of parenchymal patterns was evaluated for presence of artifacts or other image quality deficiencies such as respiratory motion or segmentation inaccuracies by a thoracic radiologist (BJB) as part of LTRC protocol. The efficacy of quantitative characterization of paren-chymal patterns by CALIPER was ascertained and reported outside this study \citep{bartholmai2013quantitative, maldonado2013automated}.

**CALIPER analysis**

{maldonado2013automated}

CALIPER analysis involved algorithmic identification and volumetric quantification of five radiological parenchymal features: normal lung, emphysema, ground-glass density, reticular abnormalities and honeycombing measured in total litres for the whole lung. Total ILD was defined as the volumetric sum of total ground-glass density, reticular abnormalities and honeycombing. Percentage ILD was defined as the ratio of the sum of the total ILD divided by the CALIPER segmented total lung parenchymal volume.

Changes in CALIPER measures between the two CT scans of each subject were calculated for total volume of normal lung, ground-glass density, reticular abnormalities, honeycombing, total lung volume, total ILD and percentage ILD.

**Discussion**

{maldonado2013automated}

CALIPER represents an automated volumetric quantification tool for assessing specific parenchymal radiological features on HRCT. In our study of IPF patients, CALIPER measured short-term (3–15 months) reticular changes, and percentage and total ILD changes were predictive of survival. Correlation between radiologists was moderate to substantial regarding ground glass and reticular findings when estimating to the nearest 5%, and mild to moderate between radiologists and CALIPER for those same ILD findings. No specific parenchymal estimates of ILD were predictive of survival with radiologist assessment, although overall global assessment of disease progression or change by radiologists was predictive in our study.

The absence of a gold standard to validate accuracy of CALIPER volumetric measurements to specific regions in a pathological specimen led us to attempt to correlate longitudinal changes in quantitative estimates of radiological fibrosis to patient outcomes, specifically mortality. As the quantitative assessment of fibrosis by CALIPER is not influenced by confounding conditions such as emphysema or pulmonary hypertension that may affect pulmonary function measures, we felt radiological features may reflect more directly the progression of fibrotic processes and could represent a novel and promising tool in the assessment and management of patients with IPF.

Interestingly, radiologists’ global assessment of short-term ILD progression was also predictive of survival in our study. We know, in general, that radiologists’ quantitative assessments and diagnostic interpretation are often inconsistent with one another, although our degree of radiologist correlation was higher. Perhaps this may be explained by our reviewing radiologists being subspecialists in thoracic radiology at the same institution, and both having reviewed and standardised cases and terminology as training for other ILD research studies. Nonblinding to time-points 1 and 2 may have increased radiologist vigilance for expected change between interval scans and biased interpretation of disease progression. Nonetheless, this conclusion is gratifying in regards to experienced radiologists detecting progression over shorter time intervals as being statistically predictive of survival, although detection of such subtle changes may not be reproducible or consistent across all practices and institutions. A quantitative method such as CALIPER may provide this consistency, particularly as detection of subtle progression over shorter time intervals may be valuable in estimating survival and perhaps be used as a marker of treatment response in future clinical trials. We are encouraged that correlation of our reproducible quantitative assessments of disease with outcomes is significant, independent of correlation with the subjective descriptions of disease features by a radiologist.

CALIPER is based on a texture-sensitive volumetric analysis that allows automated classification of lung parenchyma according to a database of HRCT volumes of interest validated by radiologists using data from the LTRC \citep{zavaletta2007high}. The majority of existing expert systems and associated quantitative tools depend on strictly controlled image acquisition protocols to provide consistent results. We believe careful selection of training sets for CALIPER enables more reproducible classification, whose greyscale local histogram-based algorithms are less affected by image noise, reconstruction kernel and other scan parameters.

While our preliminary results are promising, we recognise the limitations of our study. First, the CALIPER technology was developed based on ‘‘lung pattern signatures’’ derived from standardised CT analyses and acquisition protocols used in the LTRC database \citep{zavaletta2007high}. Scan parameters used for CT in this retrospective study were different to those used in the LTRC database and were not always identical for the scans at the two time-points. Despite differences in acquisition parameters for some of the data sets, the mild-to-moderate correlation between radiologist semiquantitative assessment and CALIPER analysis supports the validity of our results. We postulate that our histogram signature-based method may be more robust and less sensitive to specific slice thickness or image reconstruction parameters than other texture-based or pixel counting techniques. Secondly, the requirement of two serial HRCT obtained solely for follow-up purposes led us to exclude patients with HRCT obtained for coexisting illnesses (heart failure, infection and acute exacerbation) and others with only one HRCT available. This exclusion criteria arguably limits the external validity of our study, as included patients were more likely to represent a subset of IPF patients with gradual decline rather than stable (less likely to have repeat HRCT) or unstable patients (more likely to be lost to follow-up, die or experience acute exacerbation). However, the fact that the median survival of included patients was similar to that typically seen for patients with IPF is reassuring in this regard. Finally, the number of patients included was small, due to stringent eligibility criteria. While small, the numbers are comparable to those used in prior studies evaluating the value of longitudinal trends in physiological measures. We recognise these limitations and believe that further validation of our preliminary results is warranted, including prospective analysis of standardised HRCT with equivalent time intervals, comparison of IPF abnormalities to those of other ILD and application of short-term changes found with CALIPER to predicting acute exacerbation or the presence of related complications such as pulmonary hypertension.

{bartholmai2013quantitative}

Quantitative analysis results were provided in regional volumetric quantities for statistical analysis as well as a graphical representation. Analysis suggests that quantitative HRCT analysis can serve as a biomarker with physiologic, pathologic and prognostic significance.

It is likely that quantitative analysis of HRCT can be used in clinical practice as a means to aid in identifying probable diagnosis, stratifying prognosis in early disease, and consistently determining progression of disease or response to therapy. Further optimization of quantitative techniques and longitudinal analysis of well-characterized subjects would be helpful to validate these methods.

The potential for quantitative analysis to reliably characterize and quantify parenchymal abnormalities of HRCT in the setting of ILD is enormous. Ideally, computational tools can yield an objective biomarker thatmay allow for more consistent characterization of disease, with a mapping of specific characteristics and parenchymal abnormalities. Initial investigations on well-characterized data from the LTRC and retrospective analysis of clinical data in subjects with ILD demonstrate correlation with other biomarkers of pulmonary disease such as physiologic testing and also show promise as an independently significant predictor of outcomes.

Demonstration of progression of disease or response to therapy that may not be obvious to a radiologist would be helpful in clinical management. Similarly, determination of a quantity of disease or characteristic pattern that is independently predictive of mortality could be used to triage patients for transplant or other therapy. Assessment of results for multiple time points in a prospective longitudinal study or examination of previously characterized subjects involved in a longitudinal study will likely be necessary for additional validation of the CALIPER tool. Adequate visualization of quantitative results is also extremely important to real-world utility and acceptance of these results for patient management. A summary that provides unambiguous pathological, spatial and temporal information enables multivariate, multidimensional data to be intelligible. The ability to evaluate the results in context to the original CT data such as demonstrated in the CALIPER glyphs adds to the utility of these results.

With the optimizations in software and employment of modern computational hardware, CALIPER is able to complete segmentation, classification and calculate results for a volumetric HRCT of more than 600 slices in approximately 1 minute. Transitioning these tools from primarily a research role to clinical practice remains a challenge, however. As with all quantitative analysis tools, this transition will require recognition of the legitimacy of results, creation of new workflows and dataflows within clinical systems, and optimization of the technology so that results are available in clinically relevant time. It is hoped that the more robust future validation, acceptance of these results and those of other automated tools being developed throughout the community of medical imaging scientists will soon substantiate the utility of quantitative analysis in clinical practice.

{raghunath2014quantitative}

A fundamental problem in the diagnosis, severity assessment, individualized management and outcome analysis for DPLD is the lack of a robust and objective biomarker \citep{schuster2007opportunities}. Although HRCT is integral to clinical diagnosis and management of DPLDs, its use in clinical trials is sparse \citep{goldin2013computed}. The opportunities of QCT as a viable biomarker is being explored \citep{schuster2007opportunities} and several techniques for characterization of lung parenchymal disease with validated correlations of classified parenchymal patterns with physiology, visual radiology scores and patient survival have been proposed \citep{maldonado2013automated, castaldi2013distinct, best2008idiopathic, lynch2005high}. However, a recent editorial summarizes the present situation, ‘‘it is technically challenging to efficiently extract information on these patterns from CT scans … and there still seems to be a long way to go before computer software can automatically detect distinct and intuitively meaningful phenotypes’’ \citep{dirksen2013search}. Additionally, the challenges involved in optimization and standardization of acquisition and reconstruction protocols has limited the use of CT / QCT in multi-center clinical trials \citep{\citep{goldin2013computed, goldin2013computed}.

The study reported in this paper could be strengthened using an independent, even smaller DPLD population. LTRC study does not have patient follow-up and consequently, the efficacy of quantitative stratification could not be assessed with survival outcome. The methodology could also be investigated to investigate the stratification effects in response to an intervention or longitudinal disease progression. Notwithstanding the aforementioned limitations, the proposed stratification methodology can be extended to sub-stratify and identify radiological hetero-geneity within the grouped population. This could be useful to assess the radiological phenotypes and possibly different prognostic and therapeutic implications in patients. There is a need for reliable and sensitive measures to evaluate clinical significance and track efficacy of treatments in clinical trial cohorts \citep{lederer2013clinical}. The CT-based quantitative stratification could be an objective step to address this unmet need. We believe that, with further validation, meaningful information can be objectively interpreted based on the proposed quantitative stratification tool, just-in-time automated quantification software such as CALIPER and efficient glyph- based visualization. This can enable futuristic objective of physician-in-the-loop interpretation and evaluation of lung parenchymal disease that can reduce technical burden to the end user and facilitate clinical translation.