Artificial Intelligence and Chest Computational Tomography to predict prognosis in Pulmonary Hypertension and lung disease.

Dr Krit Dwivedi

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

in the

Department of Infection, Immunity and Cardiovascular Disease

14 April 2023

Acknowledgements

If I have seen further, it is by standing on the shoulders of giants.

This thesis would not be possible without the tireless help, support, and guidance of many.

Thank you to the Wellcome Trust and the 4ward North Scheme for believing me in and awarding me a Clinical Research Training Fellowship to undertake this thesis.

I could not have undertaken this journey without Andy and David. They are the very best supervisors, and not even a global pandemic could dampen their boundless enthusiasm and support. I feel fortunate and privileged to be a member of their team.

I would like to thank Smitha Rajaram, Nigel Hoggard, Paul Griffiths, Jim Wild, Tim Chico and Paul Morris for their continual guidance and support throughout my academic career, and Curtiz Langlotz for inviting me to his lab at Stanford. It is impractical to thank everyone at the Academic Unit of Radiology, Sheffield Pulmonary Vascular Disease Unit and Clinical Radiology department at Sheffield Teaching Hospitals NHS Trust, but I wish I could. This thesis and I are products of the environment created by thousands doing their very best for patients.

Finally, I would not be here today without the sacrifices, love, and unending support of my family – my parents, Sanjay and Swapna, my sister Kritica, and my loving wife Anna. All my achievements stand on their shoulders.

Declaration

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means. (www.sheffield.ac.uk/ssid/unfair-means). This work has not previously been presented for an award at this, or any other, university.

This is a publication format thesis. Multiple chapters have been published as peer-reviewed journal articles. Further detail is provided in the 'Structure of thesis' section.

Krit Dwivedi
15 March 2023

Table of Contents

Acknowledgements *2*

Declaration *3*

List of Figures *10*

List of Tables *12*

Synopsis *13*

Structure of thesis, hypothesis, aims and objectives *14*

1 Background *15*

1.1 Pulmonary Hypertension **15**

1.2 CT Imaging in Pulmonary Hypertension **19**

1.2.1 Quantitative CT and AI approaches to CT 20

1.2.2 CT Features of PH 20

1.3 Clinical dilemma between IPAH and PH-CLD **21**

2 Pulmonary Hypertension In Association With Lung Disease: Quantitative CT And Artificial Intelligence To The Rescue? State-Of-The-Art Review *24*

2.1 Abstract **26**

2.2 Introduction **26**

2.3 AI, Machine Learning and Deep Learning **31**

2.3.1 Definitions 31

2.3.2 Supervised vs unsupervised learning 33

2.3.3 General vs narrow intelligence 33

2.4 Machine Learning in Chest CT **34**

2.5 Promise of Quantitative Chest CT in PH **35**

2.5.1 Imaging 36

2.5.2 Diagnosis 38

2.5.3 Prognostication 39

2.6 Limitations, Challenges, and Solutions **41**

2.6.1 Limitations in Quantitative CT Research 41

2.6.2 Current challenges facing machine learning research 43

2.7 Conclusion **45**

3 Methodology: building clinico-radiological database 46

3.1 Raw ASPIRE registry data 46

3.1.1 Clinical data 46

3.1.2 CT imaging data 47

3.2 Merging imaging and clinical data 48

3.3 Quality control and manual radiological review 48

4 Computed tomography lung parenchymal descriptions in routine radiological reporting have diagnostic and prognostic utility in patients with idiopathic pulmonary arterial hypertension and pulmonary hypertension associated with lung disease 50

4.1 Abstract 52

4.2 Introduction 53

4.3 Methods 54

4.3.1 Patient cohort 54

4.3.2 CT analysis 54

4.3.3 Clinical and morality data 54

4.3.4 IPAH sub-group analysis 55

4.3.5 Statistics 55

4.3.6 Ethics 56

4.4 Results 57

4.4.1 Patient characteristics 57

4.4.2 Cox regression analysis 57

4.4.3 Prognostic effect of extent of emphysema and fibrosis 58

4.4.4 Kaplan-Meier survival analysis stratified by CT Features 58

4.4.5 IPAH sub-group analysis 59

4.5 Discussion 61

4.5.1 Routine radiological reports of the presence of emphysema or fibrosis have diagnostic and prognostic utility in patients diagnosed with IPAH 61

4.5.2 Routine radiological reports of the extent and nature of parenchymal lung disease have prognostic utility 62

4.5.3 Centrilobular ground glass in IPAH 62

4.6 Limitations 63

4.7 Conclusion 63

4.8 Acknowledgements 64

4.9 Tables 64

4.9.1 Table 1: Baseline characteristics. 64

4.9.2 Table 2: Univariate analysis 66
4.9.3 Table 3: Multivariate analysis 69
4.9.4 Table 4: Baseline characteristics of patients with initial diagnosis of IPAH 72

4.10 Figures 74

4.10.1 Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram showing selection of study cohort. 74
4.10.2 Figure 2: Kaplan-Meier survival curves stratified by CT features of CGG, emphysema and fibrosis for: a. All patients, b. Patients initially diagnosed with IPAH and c., Patients initially diagnosed with PH-CLD. 75
4.10.3 Figure 3: Kaplan-Meier survival curves for patients classified as IPAH-LD, IPAH-noLD and PH-CLD. 76
4.10.4 Figure 4: Kaplan-Meier curve comparing survival in IPAH-noLD (IPAH with no CT features of lung disease), IPAH-LD (IPAH with CT features of lung disease) and IPAH-CGG (IPAH with centrilobular ground glass on CT). 77

4.11 Appendix 78

4.11.1 Univariate analysis of patients with IPAH-noLD 78
4.11.2 Baseline characteristics of IPAH-noLD vs IPAH-LD vs PH-CLD 80
4.11.3 Survival of patients with IPAH-noLD, IPAH-LD and PH-CLD 82

5 Idiopathic pulmonary arterial hypertension with a lung phenotype 83

5.1 Abstract 84

5.2 Introduction 85

5.3 Methods 86

5.3.1 Databases 86
5.3.2 Patient selection 86
5.3.3 Imaging 87
5.3.4 Statistical analyses 87
5.3.5 Role of the funding source 88

5.4 Results 88

5.4.1 Patient characteristics of the study cohorts 88
5.4.2 Imaging (ASPIRE data only) 89
5.4.3 Changes from baseline to first follow-up (COMPERA data only) 89
5.4.4 Survival 90

5.5 Discussion 92

5.6 Research in context 95

5.6.1 **Evidence before the Subject** 95
5.6.2 **Added value of this study** 95
5.6.3 Implications of the available evidence 95

5.7 Tables 96
5.7.1 **Patient characteristics at baseline in COMPERA** 96
5.7.2 Table 2 Patient characteristics at baseline in ASPIRE 99
5.7.3 Table 3 Lung parenchymal abnormalities on chest computed tomography (ASPIRE) 101
5.8 Figures 102
5.8.1 STROBE diagram showing patient selection in COMPERA 102
5.8.2 STROBE diagram showing patient selection in ASPIRE 103
5.8.3 Grouped barplot showing age distribution of patients classified as classical IPAH, IPAH with a lung phenotype, and group 3 PH in COMPERA 104
5.8.4 Grouped barplot showing age distribution of patients classified as classical IPAH, IPAH with a lung phenotype, and group 3 PH in ASPIRE 105
5.8.5 Baseline and first follow-up measurement for (a) functional class (FC), (b) 6-minute walking distance (6MWD), (c) N-terminal fragment of pro-brain natriuretic peptide (NT-proBNP) and (d) mortality risk (as determined by the ESC/ERS 4-strata model) in COMPERA 106
5.8.6 Kaplan-Meier survival estimates for patients classified as classical IPAH, IPAH with a lung phenotype, and group 3 PH in COMPERA 110
5.8.7 Kaplan-Meier survival estimates for patients classified as classical IPAH, IPAH with a lung phenotype, and group 3 PH in ASPIRE 111
6 External validation, radiological evaluation, and development of deep learning automatic lung segmentation in contrast enhanced chest CT 112
6.1 Abstract 113
6.2 Introduction 114
6.3 Materials and methods 115
6.3.1 Study cohorts 115
6.3.2 Model development 116
6.3.3 Clinical segmentation scoring system: RadSeg 117
6.3.4 Statistical Analysis 118
6.3.5 Role of the funding source 118
6.4 Results 118
6.4.1 Technical results 118
6.4.2 Clinical segmentation scores 118
6.5 Discussion 119
6.6 Figures 123
6.6.1 Fig 1: STROBE flow diagram 123
6.6.2 Fig 2: Example of RadSeg scores 124
6.7 Tables 125
6.7.1 Table 1: Patient characteristics 125

6.7.2 Table 2: Clinical evaluation of radiological segmentation (RedSeg) results 126

6.8 Supplemental Material / Appendix 128

6.8.1 DICE and accuracy scores for each patient in performance testing 128

6.8.2 DICOM information for each cohort 129

7 Improved quantification and prognostication of lung disease on CT in pulmonary hypertension by combining the strengths of AI and radiologists: a retrospective multicentre validation study 131

7.1 Abstract 132

7.2 Research in context panel 134

7.2.1 Evidence before this study 134

7.2.2 Added value of this study 134

7.2.3 Implications of all the available evidence 135

7.3 List of abbreviations 136

7.4 Introduction 137

7.5 Methods 138

7.5.1 Study population 138

7.5.2 CT image acquisition, analysis and quantitative AI model 138

7.5.3 Statistics 139

7.5.4 Ethics 140

7.5.5 Role of the funding sources 140

7.6 Results 141

7.6.1 Patient and cohort characteristics 141

7.6.2 Correlations with existing methods of quantifying lung disease severity 141

7.6.3 Survival analysis 142

7.6.4 Additional prognostic value of CT AI features 143

7.7 Discussion 144

7.8 Figures 148

7.8.1 Fig 1 - STROBE diagram 148

7.8.2 Fig 2 - AI output 149

7.8.3 Fig 3 - Kaplan Meier curves stratified by significant lung disease thresholds in external cohort 150

7.9 Tables 151

7.9.1 Table 1 - demographics and clinical characteristics 151

7.9.2 Table 2 – Univariate and multivariate regression analysis for survival in internal cohort 153

7.9.3 Table 3 - Univariate and multivariate regression analysis for survival in external cohort 155

7.9.4 Table 4 – Significance and extent of AI quantified disease in those patients radiologically scored as 'no fibrosis' 156

7.9.5 Table 5 - Additive value of AI metrics on radiological fibrosis severity scoring....157

7.10 Supplementary material 158

7.10.1 Table S1a - Scanner types between cohorts 158

7.10.2 Association between AI CT disease and radiological reporting 159

7.10.3 Available data for each variable 159

7.10.4 Fig S1a - LOESS curves 160

8 Discussion 162

8.1 Limitations 164

9 Future Work 168

10 Conclusion 174

Publications, Presentations and Awards 175

Publications directly related to thesis 175

International presentations 175

Leadership and management roles related to thesis 176

Invited talks 176

Awards 177

Bibliography 178

List of Figures

Figure 1.1: PH clinical classification taken from the latest 2022 joint ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension[3]. 16

Figure 1.2 REVEAL (Registry to Evaluate Early and Long-Term PAH Disease Management) 2.0 risk score features and thresholds. Taken from original REVEAL 2.0 publication. [8] 19

Figure 2.1: Spectrum of lung disease severity within pulmonary hypertension and the diagnostic and treatment dilemma. Five-year survival figures quoted from REVEAL registry five year outcomes and recent studies [34,44,45] 28

Figure 2.2 CT features of Pulmonary Arterial Hypertension (PAH) on CT. A - dilated main pulmonary artery. B - right atrial and ventricular dilation with moderate right ventricular hypertrophy and flattening of the interventricular septum. C - centrilobular ground glass nodularity. These are a feature of PAH, but are also more commonly seen in another sub-phenotypes of Pulmonary Hypertension, such as Pulmonary Vascular Obstructive Disease (PVOD). In PVOD, they are often accompanied by interlobular septal thickening and mediastinal lymphadenopathy. D – zoomed in view of regions of centrilobular ground glass nodularity. 29

Figure 2.3 Patterns of lung disease in PH-Lung as visualised on CT. A - mild emphysema, localised predominantly to the upper lobe. B - widespread severe emphysema. C - combined emphysema and fibrosis. D - interstitial lung disease. 30

Figure 2.4 Layers of artificial intelligence approaches applied to medical imaging 32

Figure 2.5 Stages within a radiology diagnostic workflow, with potential AI applications at each stage. This review focuses on the image analysis stage - incorporating image perception and reasoning. Image reproduced with permission from original author Dr Hugh Harvey [73] 34

Figure 2.6 Demonstration of a Quantitative CT approach (adaptive multiple features method), acquired using PASS software. Different lung parenchymal disease patterns are identified and highlighted. Blue - emphysema/low attenuation pattern. Yellow - Fibrotic changes. Pink - Ground glass change. 35

Figure 2.7 Summary figure. Domains of AI application with corresponding advantages. Increasing clinical impact through clinically meaningful endpoints. 36

Figure 4.1: CONSORT (Consolidated Standards of Reporting Trials) flow diagram showing selection of study cohort. 74

Figure 4.2: Kaplan-Meier survival curves stratified by CT features of CGG, emphysema and fibrosis for: a. All patients, b. Patients initially diagnosed with IPAH and c., Patients initially diagnosed with PH-CLD. 75

Figure 4.3: Kaplan-Meier survival curves for patients classified as IPAH-LD, IPAH-noLD and PH-CLD 76

Figure 4.4: Kaplan-Meier curve comparing survival in IPAH-noLD (IPAH with no CT features of lung disease), IPAH-LD (IPAH with CT features of lung disease) and IPAH-CGG (IPAH with centrilobular ground glass on CT). 77

Figure 5.1 STROBE diagram showing patient selection in COMPERA 102

Figure 5.2 STROBE diagram showing patient selection in ASPIRE 103

Figure 5.3 Grouped barplot showing age distribution of patients classified as classical IPAH, IPAH with a lung phenotype, and group 3 PH in COMPERA **104**
Figure 5.4 Grouped barplot showing age distribution of patients classified as classical IPAH, IPAH with a lung phenotype, and group 3 PH in ASPIRE **105**
Figure 5.5 Baseline and first follow-up measurement for (a) functional class (FC), (b) 6-minute walking distance (6MWD), (c) N-terminal fragment of pro-brain natriuretic peptide (NT-proBNP) and (d) mortality risk (as determined by the ESC/ERS 4-strata model) in COMPERA **109**
Figure 5.6 Kaplan-Meier survival estimates for patients classified as classical IPAH, IPAH with a lung phenotype, and group 3 PH in COMPERA **110**
Figure 5.7 Kaplan-Meier survival estimates for patients classified as classical IPAH, IPAH with a lung phenotype, and group 3 PH in ASPIRE **111**
Figure 6.1 STROBE diagram showing patient selection in both internal reference datasets (Sheffield, ASPIRE cohort) and external datasets (Stanford ILD cohort). **123**
Figure 6.2 Examples of cases with corresponding RadSeg scores. **124**
Figure 7.1 STROBE diagram showing patient selection. **148**
Figure 7.2: Example outputs from the AI model. **149**
Figure 7.3 Kaplan-Meier survival estimates for patients in the external validation cohort (n=246) stratified into groups by thresholds for significant disease derived in the internal derivation cohort (n=275). **150**
Figure 7.4 LOESS (LOcally Estimated Scatterplot Smoothing) curves **160**

List of Tables

Table 4.1: Baseline characteristics 65
Table 4.2 Univariate analysis 67
Table 4.3: Multivariate analysis 71
Table 4.4 Baseline characteristics of patients with initial diagnosis of IPAH per subgroups 73
Table 4.5 Univariate analysis of patients with IPAH-noLD 79
Table 4.6 Baseline characteristics of IPAH-noLD vs IPAH-LD vs PH-CLD 81
Table 4.7 Survival of patients with IPAH-noLD, IPAH-LD and PH-CLD 82
Table 5.1 Patient characteristics at baseline in COMPERA 98
Table 5.2 Patient characteristics at baseline in ASPIRE 100
Table 5.3 Lung parenchymal abnormalities on chest computed tomography (ASPIRE) 101
Table 6.1 Patient characteristics for both internal reference and external cohorts 126
Table 6.2 Clinical evaluation of radiological segmentation (RedSeg) results and failure analysis for suboptimal performance in each cohort. 127
Table 6.3 DICE and accuracy scores against validation cases. 128
Table 6.4 Scan DICOM (Digital Imaging and Communications in Medicine) information for each cohort. 130
Table 7.1 Patient characteristics for the full, international derivation and external validation cohorts, with group comparison p-values for between cohort comparisons. 152
Table 7.2 Unadjusted univariate and adjusted multivariate cox proportional hazard's regression for each AI quantified CT feature in the internal derivation cohort (n=275 patients). 153
Table 7.3 Unadjusted univariate and adjusted multivariate cox proportional hazard's regression performed in the external validation cohort (n=246 patients) for each AI quantified CT feature grouped by thresholds for significant disease derived in the internal cohort 155
Table 7.4 For sub-group of patients scored by radiologists as having no fibrosis (n=300 patients), extent of and univariate cox proportional hazard's regression analysis for each significant AI quantified CT feature. 156
Table 7.5 Comparison of predictive strength of three multivariate cox proportional hazard's regression models in the external validation cohort (n=246 patients). 157
Table 7.6 Scanner types between internal derivation and external validation cohorts. 159
Table 7.8 Association between AI quantified CT features and radiologically reported semi-quantitative disease severity 159
Table 7.9 Available data for each variable. 160

Synopsis

Pulmonary hypertension (PH) is an incurable severe condition with poor survival and multiple clinically distinct sub-groups and phenotypes. Accurate diagnosis and identification of the underlying phenotype is an integral step in patient management as it informs treatment choice. Outcomes vary significantly between phenotypes. Patients presenting with signs of both PH and lung disease pose a clinical dilemma between two phenotypes - idiopathic pulmonary arterial hypertension (IPAH) and pulmonary hypertension secondary to lung disease (PH-CLD) as they can present with overlapping features. The impact of lung disease on outcomes is not well understood and this is a challenging area in the literature with limited progress. All patients suspected with PH undergo routine chest Computed Tomography Pulmonary Angiography (CTPA) imaging. Despite this, the prognostic significance of commonly visualised lung parenchymal patterns is currently unknown. Current radiological assessment is also limited by its visual and subjective nature. Recent breakthroughs in deep-learning Artificial Intelligence (AI) approaches have enabled automated quantitative analysis of medical imaging features.

This thesis demonstrates the prognostic impact of common lung parenchymal patterns on CT in IPAH and PH-CLD. It describes how this data could aid in phenotyping, and in identification of new sub-groups of patients with distinct clinical characteristics, imaging features and prognostic profiles. It further develops and clinically evaluates an automated CT AI model which quantifies the percentage of lung involvement of prognostic lung parenchymal patterns. Combining this AI model with radiological assessment improves the prognostic predictive strength of lung disease severity in these patients. The studies within this thesis represents the largest and most comprehensive analysis performed in this domain. They have been published in leading journals including *The Lancet Respiratory Medicine* and *European Respiratory Journal Open Access*.

Structure of thesis, hypothesis, aims and objectives

The thesis contains multiple chapters which have been published as peer-reviewed journal articles. Therefore, the thesis is styled in a 'publication format'. Each chapter has a detailed introduction, methods, results, discussion, and conclusion. My contribution to each piece of work is outlined at the beginning of the chapter. In addition to these published chapters (chapters 2, 4, 5, 6 and 7), there are brief background (chapter 1), methods (chapter 3), overarching discussion (chapter 8), future work (chapter 9) and conclusion (chapter 10) chapters.

The overarching thesis **hypothesis** is that in patients with Pulmonary Hypertension and lung disease (defined as those with Group 1 IPAH, Idiopathic Pulmonary Arterial Hypertension, and Group 3, Pulmonary Hypertension associated with Chronic Lung Disease, PH-CLD), radiological lung parenchymal patterns on routine Computational Tomography (CT) imaging are of prognostic significance.

The **primary** thesis aim is to identify lung parenchymal patterns of prognostic significance in these patients. The **secondary** thesis aim is to develop an automated quantitative CT AI model to better characterise these patterns.

The **objectives** are:

1. Perform a state-of-the-art literature review on artificial intelligence and quantitative chest CT (achieved in chapter 2)
2. Build a large clinic-radiological imaging database from the ASPIRE PH registry (achieved in chapter 3)
3. Investigate the prognostic impact of common CT lung parenchymal patterns (primary aim, achieved in chapter 4)
4. Establish if CT lung parenchymal patterns can be used to identify new phenotypes of pulmonary hypertension (primary aim, achieved in chapter 5)
5. Develop and externally validate a deep-learning automatic lung segmentation algorithm on contrast enhanced pulmonary angiography (CTPA, secondary aim, achieved in chapter 6).
6. Clinically evaluate the prognostic significance of automated AI quantified lung parenchymal disease severity (secondary aim, achieved in chapter 7)

1 Background

1.1 Pulmonary Hypertension

Pulmonary hypertension (PH) is a progressive, heterogenous, incurable condition with significant morbidity and mortality. It is defined as an elevation of mean pulmonary arterial pressure (mPAP) measured by right heart catheterisation (RHC). The threshold of ≥ 25 mmHg was defined at the 1st World Symposium on Pulmonary Hypertension in 1973[1]. The most recent 6th World Symposium on Pulmonary Hypertension Task Force in 2019 proposed a reduction of this threshold to > 20 mmHg, which would be above the 97.5th percentile for pulmonary arterial pressure[2]. The updated definition also considers pulmonary vascular resistance, to allow for differences in cardiac output, and should be ≥ 3 Wood Units in addition to mPAP > 20 mmHg. This was incorporated into the latest 2022 combined European Respiratory Society (ERS) and European Society of Cardiology (ESC) guidelines for Pulmonary Hypertension[3].

The underlying condition of pulmonary hypertension has multiple aetiologies, which the World Health Organisation (WHO) categorises into five groups with similar clinical and pathological characteristics. These are - Group 1: Pulmonary arterial hypertension (PAH), Group 2: PH due to left heart disease (PH-LHD), Group 3: PH due to lung disease and/or hypoxia (PH-chronic lung disease (PH-CLD)), Group 4: Chronic thrombo-embolic Pulmonary Hypertension (CTEPH), and Group 5: PH with unclear/multifactorial mechanisms. Each group has multiple sub-groups and sub-classifications, which are detailed in Figure 1.1.

GROUP 1 Pulmonary arterial hypertension (PAH)
1.1 Idiopathic
1.1.1 Non-responders at vasoreactivity testing
1.1.2 Acute responders at vasoreactivity testing
1.2 Heritable[a]
1.3 Associated with drugs and toxins[a]
1.4 Associated with:
1.4.1 Connective tissue disease
1.4.2 HIV infection
1.4.3 Portal hypertension
1.4.4 Congenital heart disease
1.4.5 Schistosomiasis
1.5 PAH with features of venous/capillary (PVOD/PCH) involvement
1.6 Persistent PH of the newborn
GROUP 2 PH associated with left heart disease
2.1 Heart failure:
2.1.1 with preserved ejection fraction
2.1.2 with reduced or mildly reduced ejection fraction[b]
2.2 Valvular heart disease
2.3 Congenital/acquired cardiovascular conditions leading to post-capillary PH
GROUP 3 PH associated with lung diseases and/or hypoxia
3.1 Obstructive lung disease or emphysema
3.2 Restrictive lung disease
3.3 Lung disease with mixed restrictive/obstructive pattern
3.4 Hypoventilation syndromes
3.5 Hypoxia without lung disease (e.g. high altitude)
3.6 Developmental lung disorders
GROUP 4 PH associated with pulmonary artery obstructions
4.1 Chronic thrombo-embolic PH
4.2 Other pulmonary artery obstructions[c]
GROUP 5 PH with unclear and/or multifactorial mechanisms
5.1 Haematological disorders[d]
5.2 Systemic disorders[e]
5.3 Metabolic disorders[f]
5.4 Chronic renal failure with or without haemodialysis
5.5 Pulmonary tumour thrombotic microangiopathy
5.6 Fibrosing mediastinitis

Figure 1.1: PH clinical classification taken from the latest 2022 joint ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension[3].

Given the complexity of the condition, management is recommended in highly specialist tertiary referral centres, of which there are nine in the United Kingdom (UK) [4]. The Sheffield Pulmonary Vascular Disease Unit (SPVDU) at the Royal Hallamshire Hospital, Sheffield Teaching Hospitals NHS Trust, is an internationally recognised centre of excellence, one of the largest PH centres in the world, responsible for 1 in 4 actively managed PH patients in UK[5].

A central goal of management in IPAH is continuous risk-stratification and assessment, with an aim to maintain a 'low-risk' status. The joint ESC/ERS guidelines defines risk groups as low, intermediate and high based on an estimated 1 year mortality of 5%, 5-10%, and >10%

respectively [6]. This is assessed holistically using a multidimensional approach. Clinical variables include signs of heart failure, progression of symptoms, and presence of syncope. Traditional biochemical biomarkers include include B-type natriuretic peptide (BNP) and N-terminal pro-brain natriuretic peptide, which non-specifically correlate with myocardial function and pulmonary haemodynamics [7]. Heamodynamic variables on RHC include right atrial pressure, cardiac index and mixed venous oxygen venous saturations. Functional assessment is performed using cardiopulmonary exercise testing and walk tests. Imaging variables currently include Right Ventricular End-Systolic Volume Index (RVESVI), Right Ventricular Ejection Fraction (RVEF) and Stroke Volume Index (SVI) on cardiac Magnetic Resonance Imaging (MRI), and right atrial area on echocardiography. These variables together are assessed to provide important prognostic information to both clinicians and patients.

Despite being routinely performed as part of the diagnostic algorithm, Chest Computational Tomography (CT) is not currently used for prognostication. Lung disease is not currently considered to be a prognostically important factor in PAH in the latest ERS/ESC guidelines[3]. A widely used prognostic risk-score in PAH is REVEAL 2.0 (Registry to Evaluate Early and Long-Term PAH Disease Management), which incorporates 11 diverse variables to provide a risk-score[8]. Variables include patient demographics, clinical features, biomarker, imaging, spirometric and haemodynamic data. Differing 'points' are given for specific thresholds for each feature and an overall risk-score is calculated; this is demonstrated in Figure 1.2. Lung disease severity is only assessed through one specific variable in pulmonary function testing (PFT) - diffusing capacity of the lungs for carbon monoxide (DLco). A <40% predicted DLco scores one risk point. REVEAL 2.0 has been validated in the Pulmonary Hypertension Society of Australia and New Zealand (PHSANZ) registry [9]. The original REVEAL 1.0 risk score has been validated and used in in different studies across multiple cohorts and registries. [10–14] Other risk scores have been derived by the French Pulmonary Hypertension Registry (FPHR)., Swedish Pulmonary Arterial Hypertension Register (SPAHR) and the Comparative, Prospective Registry of Newly Initiated Therapies for Pulmonary Hypertension (COMPERA). [15–18]

PFTs are recommended for the initial workup of patients and should include forced spirometry, DLCO and arterial blood gas sampling[3]. A strength of PFTs is their wide availability, relatively low cost and extensive established reference values and equations[19]. They represent an important investigation when lung disease is suspected and are routinely used to diagnose

obstructive or restrictive disease. However, specifically within PH, PFTs alone can be conflicting with patient reported symptoms and outcomes[3,20]. In Group 1 PAH, PFTs are usually normal or may show mild restive, obstructive, or combined abnormalities[21,22]. DLCO may be normal, but it usually mildly reduced[21]. In Group 3 PH-CLD, more significant abnormalities are seen secondary to the associated lung disease. However, the severity of PH does not correlate with the severity of lung disease as assessed by PFTs, and patients with preserved lung function can still have severe PH[20,23–26]. CT imaging offers the advantage of another means of assessing lung disease. In IPF, marginal changes in PFT values such as FVC can be challenging to interpret as they could reflect either genuine physiological deterioration or measurement variation, and CT features such as change in traction bronchiectasis have been shown to provide additional prognostic value[27]. Similarly, in PH, CT features could provide additional value, which when used in combination with PFTs, could be of prognostic significance.

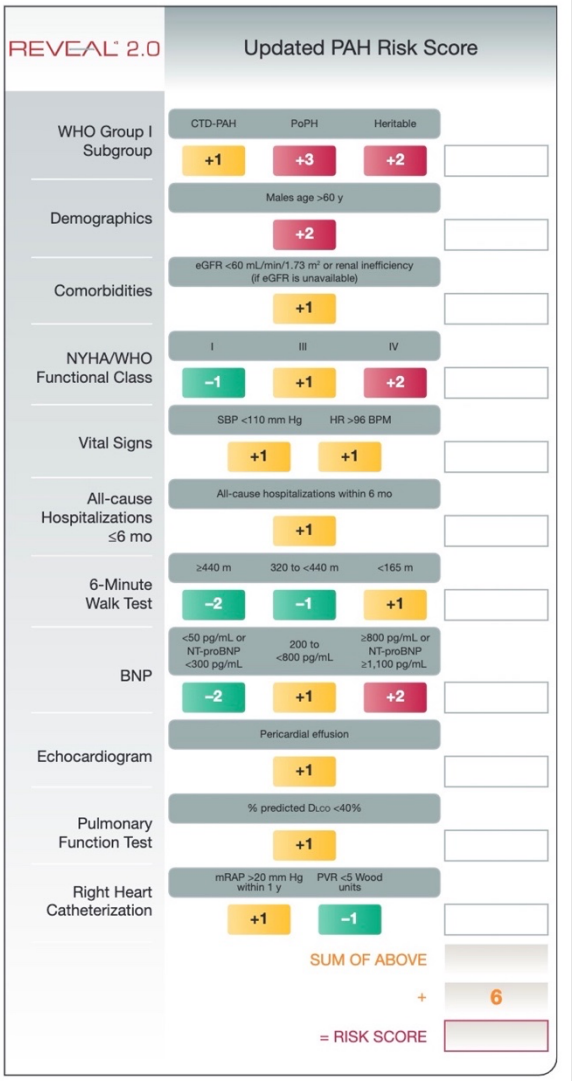


Figure 1.2 REVEAL (Registry to Evaluate Early and Long-Term PAH Disease Management) 2.0 risk score features and thresholds. Taken from original REVEAL 2.0 publication. [8]

1.2 CT Imaging in Pulmonary Hypertension

Computed Tomography (CT) is the gold standard for evaluating the extent and distribution of lung parenchymal disease, and provides an additional non-invasive assessment of vascular, cardiac and mediastinal structures [28,29]. It plays an integral role in PH management and is recommended by the latest joint European Respiratory Society/European Society of Cardiology PH guidelines and Pulmonary Vascular Research Institute (PVRI) statement on imaging in pulmonary hypertension[6,30]. Advantages of CT imaging are near ubiquitous availability, rapid time of acquisition and high spatial resolution in addition to the overall assessment of multiple structures and features. Whilst often used in adjunct to traditional PFTs in assessing lung

disease, quantitative CT features have been shown to provide additional clinically significant prognostic benefit in other diseases such as Interstitial Lung Disease (ILD)[27].

1.2.1 Quantitative CT and AI approaches to CT

Quantitative CT (QCT) is a rapidly developing field which extracts numerical information from CT imaging data. Artificial Intelligence (AI) approaches coupled with QCT have had several recent breakthroughs in a range of lung diseases. These have been extensively described in Chapter 3, which is a published peer-reviewed article which introduces, critically evaluates, and discusses the value of these concepts in context of PH imaging.

1.2.2 CT Features of PH

** Please refer to Figure 2.2 and Figure 2.3 in Introduction section of Chapter 3 (published article) for images that visualise the CT features described.*

There are multiple features of PH on CT. Mediastinal features include main pulmonary artery dilatation (MPA) and elevated MPA:aorta ratio. Cardiac features include atrial, right ventricular dilatation or hypertrophy and flattening of the intraventricular septum. Vascular features include pulmonary vessel remodelling such as dilatation, stenosis or tortuisity and neovascularity. The most associated finding in PH, common to all groups, on CT is MPA dilation, with MPA diameter >29 mm having 84% sensitivity, 75% specificity, and 97% positive predictive value (PPV) for PH defined as defined by the threshold of mPAP ≥ 25 mmHg[30,31].

The lung parenchymal features of CT are dependent on the underlying pathophysiology and subsequent WHO group. Therefore, these features are particularly helpful in ascertaining the aetiology of PH. Group 1 PH can often present with no parenchymal abnormalities. The most prevalent feature is ground glass opacification (GGO), particularly centrilobular ground glass (CGG) nodularity, and is seen in 41% of patients [28]. Within Group 1 PAH, an important sub-group to identify is pulmonary vascular obstructive disease (PVOD). It presents with a typical triad of interlobular septal thickening, ground glass nodularity and mediastinal lymphadenopathy. If confused for PAH, PAH-specific vasoactive therapy can lead to potentially fatal pulmonary oedema [32,33]. In Group 2 PH-LHD, common features include diffuse GGO, interlobular septal thickening and pleural effusions due to heart failure. Features in Group 3 PH-CLD depend on the underlying chronic lung disease, with emphysema and fibrosis

the most common sub-types. Fibrotic features include sub-pleural reticular ground glass change, traction bronchiectasis and overall anatomical architectural distortion. Group 4 CTEPH, features include a mosaic pattern of attenuation, peripheral scarring and GGO.

It is the role of the radiologist to evaluate the multiple mediastinal, vascular, cardiac and lung parenchymal features to provide an opinion in classification of PH. A particular challenge is attributing the role of lung disease in PH and assigning a diagnosis between Group 1 Idiopathic Pulmonary Arterial Hypertension (IPAH) and Group 3 PH due to lung disease and/or hypoxia (PH-CLD).

1.3 Clinical dilemma between IPAH and PH-CLD

** please note some of this is repeated in Chapter 4 and 5 – the published journal articles – but it is included in here to give brief context and introduction to the topic.*

Group 3 PH (PH-CLD) is a complex, highly heterogeneous group where pre-capillary PH most commonly complicates chronic obstructive pulmonary disease (COPD) and or emphysema, interstitial lung disease (ILD) and alveolar hypoventilation [23,26,34,35]. These conditions may also co-exist to varying degrees and patients with a combination of pulmonary fibrosis and emphysema (CPFE) with severe PH having a particularly poor prognosis [34]. Usually patients with PH in association with lung disease have a mild elevation of pulmonary artery pressure in keeping with the severity of the underlying lung disease, however, some patients present with severe PH with variable degrees of parenchymal involvement [26]. A review by Kovacs et al has highlighted the spectrum of pulmonary vascular involvement in patients with COPD [36]. In these patients it is important to exclude other forms of PH such as Group 1 PH - Pulmonary arterial hypertension (PAH) and Group 4 PH (Chronic Thromboembolic Pulmonary Hypertension (CTEPH) [37].

Idiopathic pulmonary arterial hypertension (IPAH) is a form of PAH where no other cause of PH is identified although in some cases it has a heritable component [38–40]. Current guidelines recommend that patients with severe pre-capillary PH with no other cause identified be classified as Group 1 IPAH [26,41] , whereas patients with severe lung disease and mild PH are classified as having Group 3 PH-Lung.

However, in clinical practice some patients do not neatly fit into either category and this creates a clinical dilemma of significant prognostic importance [42]. These patients have hemodynamically confirmed PH on Right Heart Catheterisation (RHC) and mild to moderate lung disease on CT. However, they have unremarkable spirometry, no thromboembolic disease and an absence of a relevant associated medical condition. Do they have PH-CLD or IPAH? Accurate diagnosis and identification of the underlying phenotype is vital as it informs treatment choice, response, risk-stratification, and prognosis. [A]

The recommended management of IPAH and PH-CLD are divergent [6]. The last decade has seen the introduction of multiple novel PAH specific drug therapies targeting the endothelin, nitric oxide and prostanoid pathways [43]. These therapies are recommended in patients diagnosed as Group 1 PAH. They have significantly improved the outlook for these patients, with mean five-year survival for IPAH reported between 50-65% [5,44]. In contrast, patients diagnosed with PH-CLD do not qualify for novel PAH therapy, and management is limited to optimisation of the lung disease and lung transplantation in exceptional cases. PH-CLD five-year survival is between 28-31% [5,34,45].

This dilemma understandably is one of active interest within the literature. The latest 6th World Symposium on PH (WSPH) recommended that patients with minor lung disease, who otherwise meet the criteria for IPAH (no other significant cause or risk factor identified), should be considered to have IPAH despite the lung disease [26]. This recommendation was based on unpublished subgroup analysis in patients with mild to moderate lung disease from randomised controlled studies [26,46–48]. However, a recent publication has demonstrated that patients diagnosed as IPAH using this criteria from the 6th WSPH with mild lung disease had significantly worse outcomes with poor survival and treatment response [41]. Furthermore, whilst PH-CLD patients are thought to represent an entirely different pathophysiological group compared to IPAH, recent studies have raised the possibility of a distinct phenotype that lies in between this spectrum - with mild to moderate lung disease [36,41,47].

[A] As discussed in later chapters and in the overarching discussion section, a phenotype is defined by its clinical characteristics, such as demographics, treatment response and prognosis. The work in this thesis is important as it shows a phenotype with differing characteristics. If recognised and incorporated into formal guidance, then this 'phenotype' definition can be subsequently used to inform patient treatment choice, response, risk-stratification, and prognosis in clinical practice.

The impact of commonly encountered lung parenchymal patterns on Chest CT in PH is unknown. As noted in the recent Pulmonary Vascular Research Institute (PVRI) statement on imaging and PH, significant parenchymal abnormalities may be seen on CT in the presence of simple screening test of lung function such as spirometry alone[30]. Can Chest CT serve as an imaging biomarker to help solve this diagnostic dilemma - can it predict survival and help aid in classification and prognostic risk-stratification? If so, can we automate this by developing a QCT AI package that can perform this analysis? These are the main questions this thesis aims to answer.

The overarching thesis **hypothesis** is that in patients with Pulmonary Hypertension and lung disease (defined as those with Group 1 IPAH, Idiopathic Pulmonary Arterial Hypertension, and Group 3, Pulmonary Hypertension associated with Chronic Lung Disease, PH-CLD), radiological lung parenchymal patterns on routine Computational Tomography (CT) imaging are of prognostic significance.

The **primary** thesis aim is to identify lung parenchymal patterns of prognostic significance in these patients. The **secondary** thesis aim is to develop an automated quantitative CT AI model to better characterise these patterns.

2 Pulmonary Hypertension In Association With Lung Disease: Quantitative CT And Artificial Intelligence To The Rescue? State-Of-The-Art Review

Peer-reviewed journal article published in *Diagnostics (2021) [49] Prepared in part from literature review undertaken for thesis.

Contribution: First author. Contributor Roles Taxonomy (CRediT) – lead for conceptualisation, investigation, methodology, project administration, resources, writing, review and editing).

I lead this study. I did all the literature searching, analysis and writing independently for this literature review. It was done as part of my confirmation review for this thesis. DGK and AS, my supervisors and senior authors, helped guide the project. All other authors had a small involvement with support, project administration and resources.

Krit Dwivedi[1], Michael Sharkey[1,5], Johanna M Uthoff[3], Samer Alabed[1], Peter Metherall[1,5], Haiping Lu[2,3], Robin Condliffe[1,4], Jim M Wild[1,2], Eric A Hoffman[6], Andrew J Swift[1,2,5 +] and David G Kiely[1,2,4 +]

1. Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, UK.
2. INSIGNEO, Institute for In Silico Medicine, University of Sheffield, Sheffield, UK.
3. Department of Computer Science, University of Sheffield, Sheffield, UK.
4. Sheffield Pulmonary Vascular Disease Unit, Royal Hallamshire Hospital, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK.
5. Radiology Department, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

[6] Advanced Pulmonary Physiomic Imaging Laboratory, University of Iowa, C748 GH, Iowa City, IA 52242, USA.

* Correspondence: Krit Dwivedi; krit.dwivedi1@nhs.net

+ David G Kiely and Andrew J Swift contributed equally to the manuscript.

2.1 Abstract

Accurate phenotyping of patients with pulmonary hypertension (PH) is an integral part of informing disease classification, treatment and prognosis. The impact of lung disease on PH outcomes and response to treatment remains a challenging area with limited progress. Imaging with Computed Tomography (CT) plays an important role in patients with suspected PH when assessing for parenchymal lung disease, however, current assessments are limited by their semi-qualitative nature. Quantitative Chest-CT (QCT) allows numerical quantification of lung parenchymal disease beyond subjective visual assessment. This has facilitated advances in radiological assessment and clinical correlation of a range of lung diseases including emphysema, interstitial lung disease and COVID-19. Artificial Intelligence approaches have the potential to facilitate rapid quantitative assessments. Benefits of cross-sectional imaging include ease and speed of scan acquisition, repeatability and the potential for novel insights beyond visual assessment alone. Potential clinical benefits include improved phenotyping and prediction of treatment response and survival. Artificial Intelligence approaches also have the potential to aid more focused study of PAH therapies by identifying more homogeneous subgroups of patients with lung disease. This state-of-the-art review summarises recent QCT developments and potential applications in patients with PH with a focus on lung disease.

2.2 Introduction

Pulmonary hypertension (PH) is a heterogeneous, life-limiting condition defined by an elevated pulmonary artery pressure and if untreated results in right ventricular failure and death. The current classification of PH identifies 5 groups each with shared pathophysiological characteristics [6] . However, increasingly patients are identified with overlapping features. Group 3 PH - Pulmonary hypertension in association with lung disease and/or hypoxia (PH-Lung) is a complex, highly heterogeneous group where pre-capillary PH most commonly complicates chronic obstructive pulmonary disease (COPD) and or emphysema, interstitial lung disease (ILD) and alveolar hypoventilation [23,26,34,35]. These conditions may also co-exist to varying degrees and patients with a combination of pulmonary fibrosis and emphysema (CPFE) with severe PH having a particularly poor prognosis [34]. Usually patients with PH in association

with lung disease have a mild elevation of pulmonary artery pressure in keeping with the severity of the underlying lung disease, however, some patients present with severe PH with variable degrees of parenchymal involvement [26]. A review by Kovacs et al has highlighted the spectrum of pulmonary vascular involvement in patients with COPD [36]. In these patients it is important to exclude other forms of PH such as Group 1 PH - Pulmonary arterial hypertension (PAH) and Group 4 PH (Chronic Thromboembolic Pulmonary Hypertension (CTEPH) [37]. Idiopathic pulmonary arterial hypertension (IPAH) is a form of PAH where no other cause of PH is identified although in some cases it has a heritable component [38–40]. Current guidelines recommend that patients with severe pre-capillary PH with no other cause identified be classified as Group 1 IPAH [26,41] , whereas patients with severe lung disease and mild PH are classified as having Group 3 PH-Lung. However, in clinical practice some patients do not neatly fit into either category and this creates a clinical dilemma (Figure 2.1).

The recommended management of IPAH and PH-Lung are divergent. The last decade has seen the introduction of multiple novel PAH specific drug therapies targeting the endothelin, nitric oxide and prostanoid pathways [43]. These have improved the outlook for PAH, particularly in younger patients with IPAH where the UK National Audit reports a 5 year survival in excess of 80% [5]. Previous data from the ASPIRE Registry highlighted the impact of lung disease on survival compared with other forms of PH. Patients with PH-Lung assessed at a PH referral centre had a 5 year survival of 31%, worse than all other forms of PH [50]. For patients with severe PH-Lung the survival is particularly poor, worse in ILD compared to COPD [50]. Severe PH occurs in a minority of patients with lung disease, however due to the very high prevalence of lung disease, estimates of severe PH in lung disease are seven times more common than group 1 PAH [37,47,51,52]. In contrast to Group 1 PAH, for patients with Group 3 PH-Lung guidelines recommend treatment of the underlying lung disease or in highly selected cases lung transplantation. Interestingly, the 6th World Symposium on PH (WSPH) recommended that patients with minor lung disease, who otherwise meet the criteria for PAH with no other causes identified, should be considered to have IPAH [26]. This recommendation was based on unpublished subgroup analysis in patients with mild to moderate lung disease from randomised controlled studies [46–48]. However, a recent publication has demonstrated that patients diagnosed as IPAH using criteria from the 6th WSPH with mild lung disease had significantly worse survival with no improvement in exercise capacity or quality of life [41]. This study highlights the challenge of classifying patients with PH and a need for improved disease phenotyping to ensure that the most appropriate patients receive treatment and where uncertainty exists that patients are entered into appropriate clinical trials.

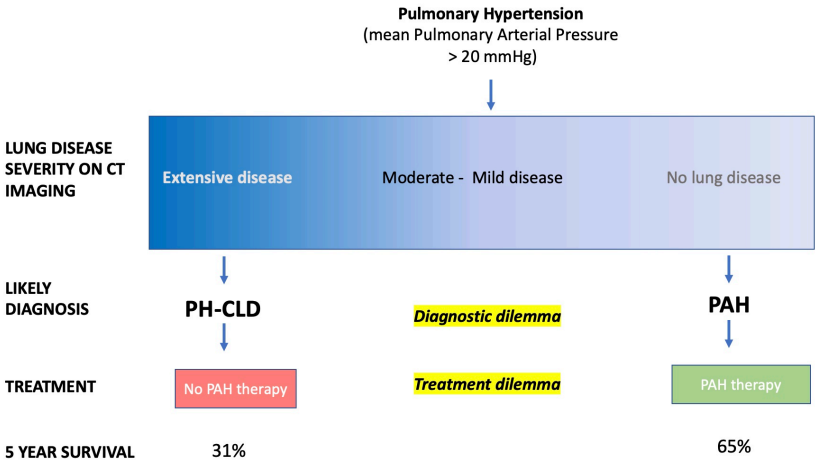


Figure 2.1: Spectrum of lung disease severity within pulmonary hypertension and the diagnostic and treatment dilemma. Five-year survival figures quoted from REVEAL registry five year outcomes and recent studies [34,44,45]

Traditionally the severity of lung disease is quantified using lung function tests using spirometric values, lung volumes and measurements of gas diffusion (DLco and Kco), with increasing evidence showing that DLco is strongly prognostic in a number of forms of PH [53–55]. Indeed DLco is part of a widely used risk score, REVEAL 2.0 [8]. Lung function tests, however, show significant variability, low reproducibility, and can be insensitive to change to disease progression [56–58]. [B]Computed Tomography (CT) is the gold standard for evaluating the extent and distribution of lung parenchymal disease, and provides an additional non-invasive assessment of vascular, cardiac and mediastinal structures [28,29]. CT is therefore recommended by both the latest joint European Respiratory Society/European Society of Cardiology PH guidelines and the Pulmonary Vascular Research Institute (PVRI) statement on imaging in pulmonary hypertension as an important part of the diagnostic strategy in suspected PH [6,30]. The features and patterns of disease in PAH and PH-Lung on CT are visualised in Figure 2.2 and Figure 2.3 respectively.

[B] The strengths and limitations of pulmonary function testing has been expanded on page 16 to provide more detail to this statement.

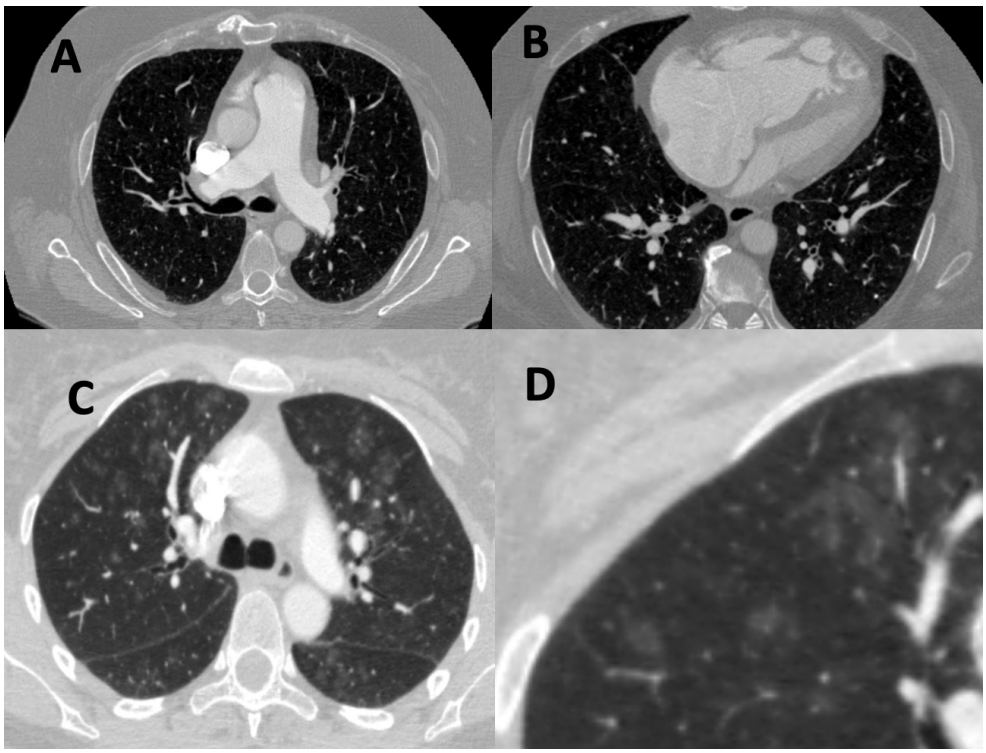


Figure 2.2 CT features of Pulmonary Arterial Hypertension (PAH) on CT. A - dilated main pulmonary artery. B - right atrial and ventricular dilation with moderate right ventricular hypertrophy and flattening of the interventricular septum. C - centrilobular ground glass nodularity. These are a feature of PAH, but are also more commonly seen in another sub-phenotypes of Pulmonary Hypertension, such as Pulmonary Vascular Obstructive Disease (PVOD). In PVOD, they are often accompanied by interlobular septal thickening and mediastinal lymphadenopathy. D – zoomed in view of regions of centrilobular ground glass nodularity.

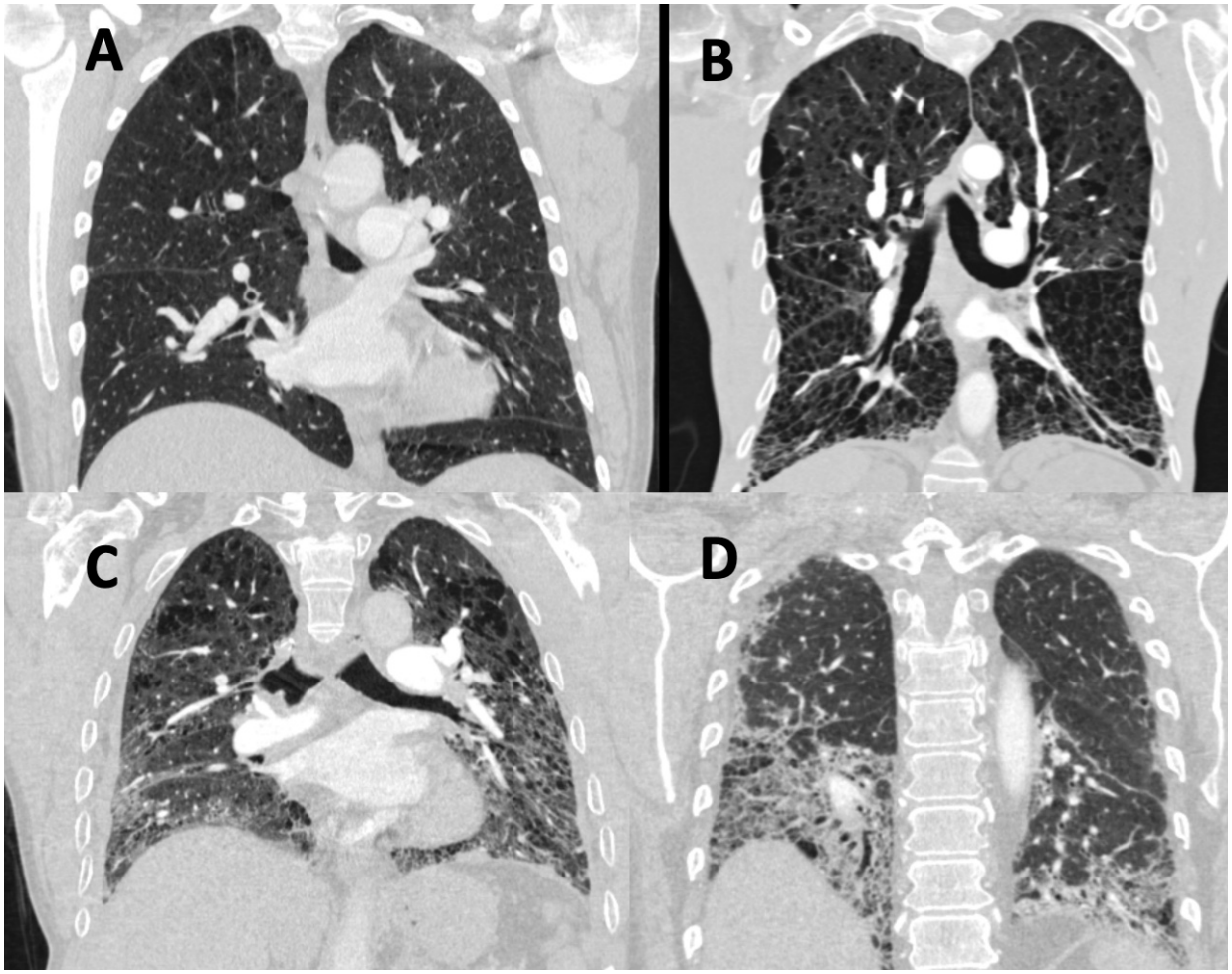


Figure 2.3 Patterns of lung disease in PH-Lung as visualised on CT. A - mild emphysema, localised predominantly to the upper lobe. B - widespread severe emphysema. C - combined emphysema and fibrosis. D - interstitial lung disease.

AI has seen a rise in prominence and performance, especially with recent machine learning (ML) approaches. It has been labelled as 'a major paradigm shift' and 'one of the most fundamental changes in medical care' [59,60]. Within Chest CT and respiratory medicine, there has been an explosion in AI research, with a 5x increase in studies between 2014-19 compared to 2010-14. Correspondingly, there are now >75 FDA and several CE approved AI software packages in Chest CT alone, compared to just 1 in 2014 [61]. We have reviewed the literature and found no studies directly involving AI, Chest CT and PH. There are however many studies involving AI, ML and Chest CT in other diffuse lung diseases such as emphysema, COPD, ILD and recently COVID-19. The discoveries, solutions and findings of these studies are directly applicable to development of AI models in PH.

Despite a number of studies highlighting the negative impact of PH on survival in PH-Lung [23,37,62], data is very limited for the use of PAH therapies. Studies have been performed primarily in COPD and ILD and have yielded conflicting results [26]. This reflects in part the small numbers, study designs and the heterogeneous nature of patients enrolled. Despite studies with ambrisentan [63] and riociguat [64] identifying safety concerns in patients with ILD a recent randomised controlled study, INCREASE, has demonstrated an improvement in 6 Minute Walk Test in patients with ILD [65]. This has re-invigorated the PH community to explore PAH therapies in PH-Lung. However, given the heterogenous nature of PH in Lung disease a technique to aid the identification of more homogeneous subgroups of PH-Lung would be very welcome.

Given the potential of AI approaches to assess the lung parenchyma, vessels and cardiac chambers the authors postulate that Quantitative CT (QCT) could 'come to the rescue' of investigators wishing to improve the outlook for patients with PH-Lung. By improving phenotyping AI approaches could aid the characterisation of disease and enrich for populations most likely to benefit from PAH therapies. This state-of-the-art review critically appraises the recent developments in the adjacent fields of ML in CT imaging and contextualises their potential impact on imaging, diagnosis, classification and assessment of prognosis in patients with PH with a focus on lung disease.

2.3 AI, Machine Learning and Deep Learning

2.3.1 Definitions

Artificial intelligence (AI) is a general term encompassing computer algorithms capable of performing tasks requiring human intelligence.Figure 2.4 summarises the relationship between different AI approaches. Historically, AI has been trained with a rules-based approach, where a programmer explicitly creates a set of conditions upon which the machine executes actions.

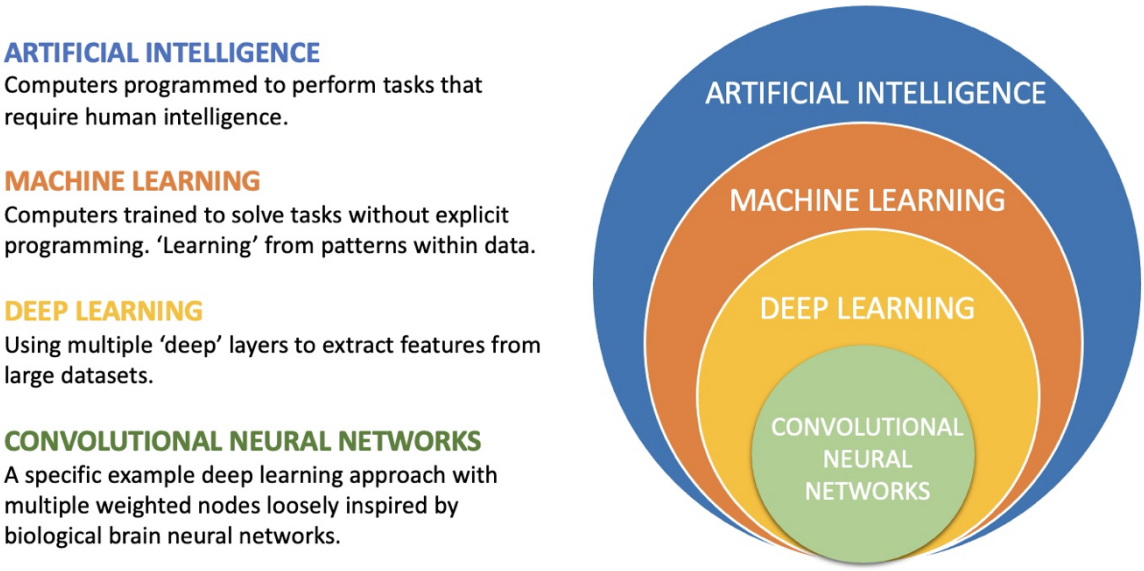


Figure 2.4 Layers of artificial intelligence approaches applied to medical imaging

Machine learning (ML) is a subset of AI in which algorithms are trained to solve tasks through feature learning instead of an explicit rules-based approach. When presented with a 'training' cohort, the algorithm identifies salient features, which are subsequently used to make predictions. Hence the 'machine' 'learns' from the data itself.

Deep learning (DL) is a subset of ML which uses multiple layers to extract features. Each 'layer' provides information to the algorithm. 'Convolutional neural networks' (CNN) is a subset and one approach to 'deep learning', which is loosely inspired by neurons in the human brain. Several 'nodes' exist within multiple layers, and each 'node' as a data point has a weighting which affects the whole 'network'. This approach in particular is most prevalent in the domain of image recognition and analysis and has demonstrated superior problem-solving capabilities. While most earlier AI methods have led to applications with subhuman performance, recent deep learning algorithms are able to match and even surpass humans in task-specific applications.[59,66–70]. Two recent examples of AI superiority in medical imaging include measuring wall thickness in hypertrophic cardiomyopathy on cardiac Magnetic Resonance Imaging (MRI) and diagnosing breast cancer in mammography [66,71].

2.3.2 Supervised vs unsupervised learning

Two broad categories of ML methods exist - supervised and unsupervised learning. Supervised learning requires explicit labelling of the training data with a ground truth, upon which the machine develops its model and makes predictions. Labelling lung parenchymal disease patterns as 'ground glass change' or 'emphysema' is an example of supervised learning. Performance is tested by how well it predicts these labels in a validation cohort. Unsupervised learning requires no explicit labelling of the training data, rather it learns to cluster/group together, thereby requiring very large datasets [72]. The algorithm makes its own inferences within the data space to learn its internal structure and uses that make predictions. Semi-supervised learning is a combination of two methods - where a smaller labelled dataset is combined with a larger unlabeled dataset. The labelled dataset is used to guide the algorithm before it is used in the unlabeled data. The vast majority of all medical imaging AI studies use supervised learning and is therefore reliant on accurate labelling.

2.3.3 General vs narrow intelligence

General intelligence is the ability to apply knowledge across a range of domains. A radiologist is trained to have 'general' intelligence across all modalities and age groups, from detecting congenital anomalies in fetal ultrasounds to complex neurodegenerative disease in brain MRIs. Methods and skills learnt in one domain are often transferable to other domains. In contrast, almost all state-of-the-art AI advances are currently limited to narrow intelligence in a defined domain [59]. The focus of such studies has been to match and occasionally surpass human radiologist performance in that narrow specific instance. How 'narrow' an algorithm is depends on how it was developed and what datasets were used; an algorithm trained entirely on ILD could not evaluate emphysema.

Figure 5 outlines the multiple stages within a diagnostic radiology workflow, highlighting the potential avenues for AI solutions. This review focuses on the image perception and reasoning stages, however each stage is a current active research topic of interest.

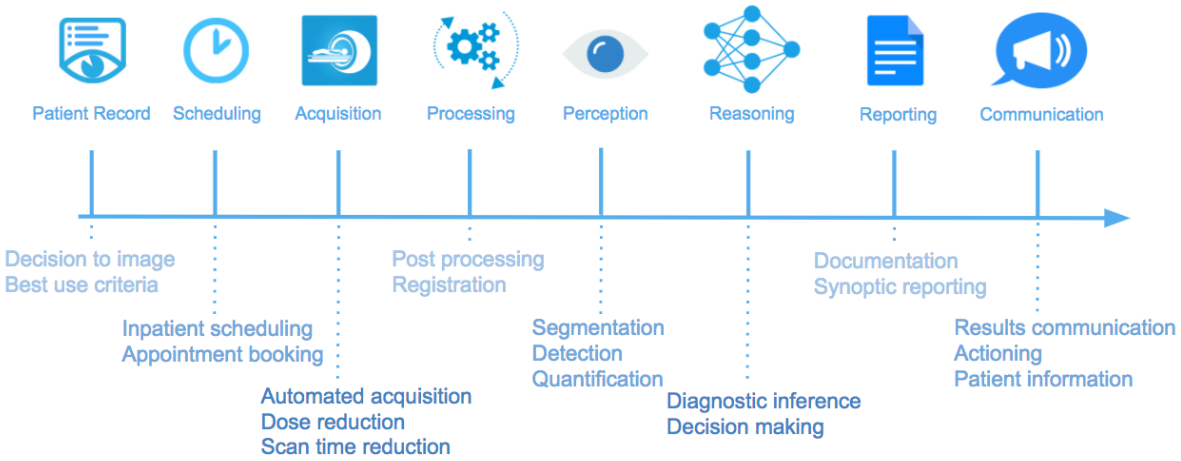


Figure 2.5 Stages within a radiology diagnostic workflow, with potential AI applications at each stage. This review focuses on the image analysis stage - incorporating image perception and reasoning. Image reproduced with permission from original author Dr Hugh Harvey [73]

2.4 Machine Learning in Chest CT

Within image analysis, machine learning in chest CT broadly has two domains: nodule detection and radiomics. Multiple AI solutions exist for lung cancer nodule detection and have shown to have a high level of accuracy, sensitivity and specificity [74–76]. Labelled public datasets for model development exist, the largest of which is the Lung Image Database Consortium Image Collection (LIDC-IDRI) [77]. Radiomics is the broad field of study which aims to extract information from imaging using computer-aided mathematical analysis that is not accessible through traditional visual inspection [78]. Initial use was predominantly in oncology settings to make predictions on disease course, survival and treatment response from tumour features [78,79]. Recently the field has expanded to non-oncological settings. In diffuse lung diseases, QCT has been the main application of ML algorithms.

Quantitative CT (QCT) is the principle of extracting quantitative information from standardised imaging data. This ranges from simple human hand-drawn manual measurements of anatomical structures such as the trachea or main pulmonary artery (mPA) to complex AI driven texture analysis of lung parenchymal disease patterns. It has been used in a range of diffuse lung diseases [80,81]. State-of-the-art applications of QCT use ML and DL to provide end-to-end solutions - where the entire CT scan is automatically analysed with an output provided. Such approaches require a suite of algorithms - from segmentation where the lung parenchyma is accurately identified and extracted - to quantification and incorporation into clinical models. A

recent example is CORADS-AI - which automatically quantifies and scores the extent of lung parenchymal in COVID-19 [82]. Figure 6 demonstrates such a QCT approach using an adaptive multiple features method, with different parenchymal patterns highlighted [83].

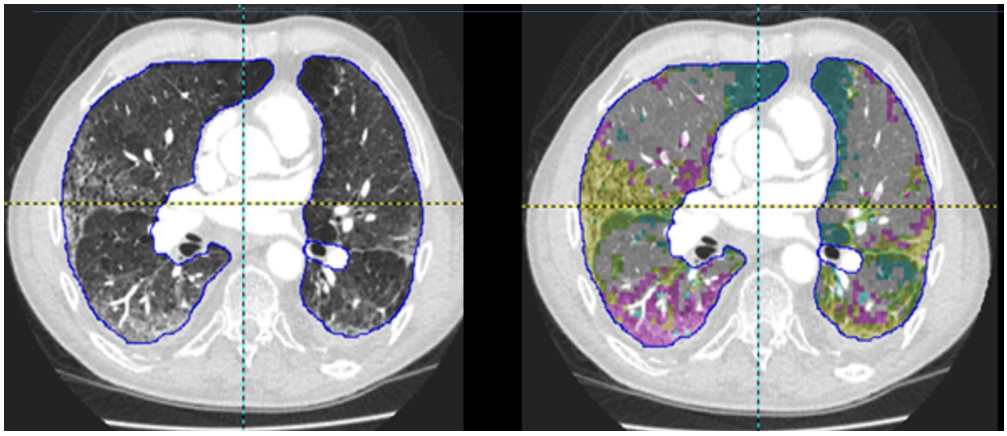


Figure 2.6 Demonstration of a Quantitative CT approach (adaptive multiple features method), acquired using PASS software. Different lung parenchymal disease patterns are identified and highlighted. Blue - emphysema/low attenuation pattern. Yellow - Fibrotic changes. Pink - Ground glass change.

2.5 Promise of Quantitative Chest CT in PH

The current standard for assessment of Chest CT is by an expert radiologist. This approach fundamentally treats scans as pictures for subjective visual assessment. QCT and AI approaches in contrast treat scans as imaging data, which can be processed and analysed.

AI models can be categorised by their domain of application and corresponding endpoints. Figure 7 is a summary diagram which visualises the promise and advantages of QCT in each domain - imaging, diagnosis, and prognostication.

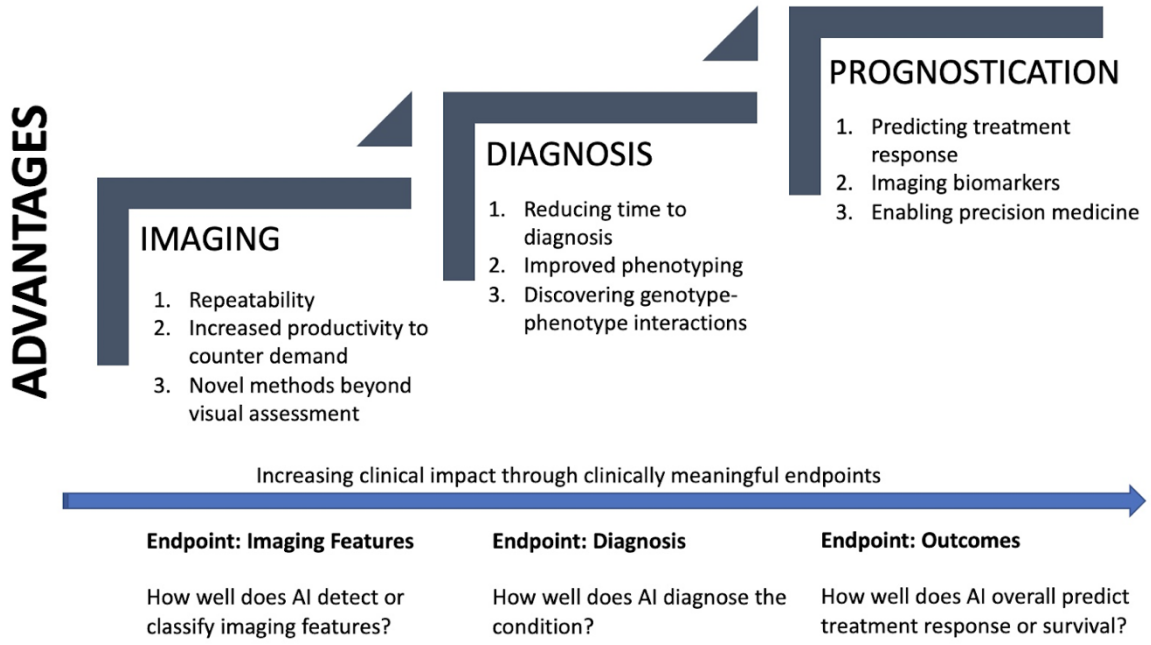


Figure 2.7 Summary figure. Domains of AI application with corresponding advantages. Increasing clinical impact through clinically meaningful endpoints.

2.5.1 Imaging

2.5.1.1 Repeatability

Radiological assessment is a subjective process. Reports between radiologists differ significantly in their style and content, and there exists significant inter-observer variability on visual assessment even between highly experienced radiologists [84–86]. Within Chest CT imaging, reports are often broad and give an overall impression of the disease process. It is either binary (disease present or absent) or a rough categorical assessment of a degree of severity (mild, moderate or severe). Clinically significant differences have been demonstrated even in assigning final diagnosis in ILD [87].

QCT in comparison provides reproducible data which numerically quantifies disease severity [59,88]. It therefore provides an objective measure which can be integrated into diagnostic and prognostic models. This principle of repeatability minimises the inherent variability from a visual approach to imaging.

2.5.1.2 Increased efficiency to counter rising demand.

The demand for medical imaging is rising far steeper than the available radiology workforce [59,89,90]. In 2015, it was estimated that an average radiologist must interpret an image every 3-4 seconds in a normal 8-hour workday to meet workload demands [91]. The demand is particularly steep for complex investigations that require longer to report - such as chest CT and cardiac MRI. These modalities are routinely used in PH and integral to assessment.

AI has been shown to improve productivity, reducing the time needed to review imaging. Within PAH, routine and repetitive tasks such as measuring pulmonary artery size or right atrial area could be automated, saving time over hand-drawing regions. Temporal subtraction is an AI method to highlight interval change between successive imaging [92,93]. In bone CT scans this method reduced reading time by 43% and increased sensitivity by 14.6% [94]. PH patients require regular reassessment with imaging to monitor treatment response, and such techniques could help increase the efficiency of repeat comparative reporting.

Triaging of scans is another area which has seen significant research [95]. Scans are first read by the AI model which then categorises them based on probability of disease. This would enable more efficient use of resources, with radiologists prioritising those scans that are likely to involve the disease, thereby reducing time to report and enabling clinical decision making.

2.5.1.3 Novel methodologies beyond visual assessment

A limitation of all current imaging modalities is the inability to visualise and assess the distal pulmonary artery vascular - the region of disease in PAH [30]. In PAH, large pulmonary arteries are known to demonstrate dilation, pruning and abrupt tapering or tortuosity [96]. Pulmonary vessel morphology is an example of an approach made possible only through quantitative AI analysis, as it would be unfeasible visually. The pulmonary vascular tree is segmented and features such as lumen size, tortuosity and tapering quantified. There exist several 3D vessel lumen segmentation techniques in both CT and MRI to enable this [97]. A deep learning CNN approach recently demonstrated 94% accuracy in segmentation of the vascular tree [98]. The findings from clinical applications of such approaches could have an impact on diagnosis and prognostication.

2.5.2 Diagnosis

2.5.2.1 Reducing time to diagnosis and error

A dual reader approach has been shown to reduce error and misses. However, this is prohibitively time and radiologist resource intensive. AI systems have therefore been suggested as an alternative 'second reader' and have shown promise in reducing errors and improving sensitivity [93].

Quantifying severity with a continuous index rather than a broad visually estimated category can increase potential to detect more subtle changes [80]. 48% of PH patients do not receive a diagnosis until one year after experiencing symptoms, and 40% see four or more health care providers prior to diagnosis [99]. Chest CT is commonly performed however radiologists assessing the lungs will not routinely evaluate for features of PH and some characteristic abnormalities of CTEPH may be subtle and require specialist radiology expertise limited to tertiary centres. AI models that automatically evaluate studies for signs of PH have the potential to reduce the time to diagnosis.

In emphysema, airway wall thickening visualised on incidental Chest CT was shown to be an independent predictor of COPD exacerbations that led to hospitalization or death in a large multicenter randomised controlled trial [100]. A similar simple approach can be applied to PH where QCT could automatically segment, measure and plot the main pulmonary artery size against a normative curve. This would provide the reporting radiologist with additional context to evaluate and consider PH as a potential diagnosis.

2.5.2.2 Improving phenotyping and classification

In COPD, new subtypes and phenotypes have been discovered through ML approaches [101,102]. These distinct patient subtypes characterized by imaging correlate with physiological parameters. In ILD, QCT metrics have been shown to correlate well with a range of clinical function tests such as lung function tests [81,88,103,104]. [C]

PH-Lung has multiple different phenotypes with distinct treatment responses [23,34]. A current clinical dilemma in PH is differentiating IPAH with mild lung disease from PH-Lung;

[C] To be more specific, some not all metrics have shown to correlate. For example in IPF, pulmonary vessel volume is an example of a novel quantitative metric with no visual scoring equivalent that correlates well, but other QCT metrics such as measures of emphysema extent did not correlate well.

therefore, there is a need to understand more about different phenotypes and why some patients with variable degrees of parenchymal disease do or do not develop PH [26,105]. New phenotypes have also been proposed in IPAH based on patterns of lung involvement on CT [36,41] that differ significantly in treatment response to PAH therapy and prognosis. Quantitative approaches could therefore aid identification of new phenotypes and improve on assessments based on traditional visual based assessment alone.

2.5.2.3 Discovering genotype-phenotype associations

Four large collaborative genomic and multi-omic programmes and biobanks are established for PAH - PVDOMICS, NBR, US PAH Biobank and UK national IPAH cohort [39,106–108]. Bone morphogenetic receptor type 2 (BMPR2) gene abnormalities are the most common cause of heritable PAH, comprising ~15% of all cases; but 20+ new genes have been identified [39,109]. Patients with BMPR2 mutations are unlikely to demonstrate vasoreactivity - which informs clinical management [110]. Advances in genetic understanding and targeting BMPR2 have developed novel therapies that are tested in clinical trials [111].

The distinct imaging appearances of these phenotypes is not currently well characterised. Cardiac MRI has demonstrated RV function to be more severely affected in BMPR2 patients, but CT features are currently an area of research [112]. Marrying genetic data with imaging data offers the potential to better phenotype patients.

2.5.3 Prognostication

2.5.3.1 Predicting treatment response and survival

In Idiopathic Pulmonary Fibrosis (IPF), texture based QCT of lung parenchymal disease patterns was superior to both visual scoring and traditional lung function tests in predicting outcomes [113]. QCT specific lung texture patterns were also found to be an independent predictor for survival when comparing short term interval changes between two scans [114]. For radiologist interpretation, only the overall disease progression was a predictor, and not specific lung features. Models have also demonstrated the ability to identify and select patients who would be steroid responders [115]. In chronic hypersensitivity pneumonitis, extent of fibrosis and reticulation independently predicted time to death or lung transplantation [116]. Another study found severity of traction bronchiectasis and honeycombing to predict mortality [117]. As an example of going beyond conventional metrics such as lung function tests, severity of traction

bronchiectasis on HRCT was found to be an independent predictor for mortality in those patients that had marginal annual forced vital capacity (FVC) declines [27].

In COPD, large longitudinal multicentre prospective trials such as COPDGene and SPIROMICS include QCT data to better understand the disease process [118,119]. The results of these studies have found the impact of QCT metrics on a vast range of outcomes [120,121]. In emphysema, lung volume reduction surgery is a treatment in which disease distribution pattern and fissure integrity are important predictors of success [122]. Regional quantification by QCT models here has shown to predict postoperative lung function, thereby aiding clinical decision making [123]. These findings demonstrate the prognostic potential of QCT in PH.

2.5.3.2 Imaging biomarkers as clinical endpoints

Traditional biochemical biomarkers in PAH include B-type natriuretic peptide (BNP) and N-terminal pro-brain natriuretic peptide, which non-specifically correlate with myocardial function and pulmonary haemodynamics [7]. These are routinely used in practice to inform clinical opinion.

Imaging biomarkers are imaging features of pathological conditions [124]. Traditional CT imaging biomarkers in PH include mPA size and secondary signs of heart failure such as inferior vena cava dilation, pleural effusions and septal lines [28]. A mPA size >29 mm has 97% positive predictive value for PH and a PA to ascending aorta ratio >1 is 92% specific for a raised mean arterial pressure >20 mmHg [30,31,125]. Quantitative CT enables more complex biomarkers that can combine multiple measurements or perform higher level textural analysis to create a model that could be then validated to be diagnostic or prognostic value. Whilst the majority of studies focus currently on the analytical performance of such models, Swift et al last year validated a CT model against clinical outcomes to demonstrate both diagnostic and prognostic value in suspected PH [29].

Current assessment tools in PH clinics and endpoints used in clinical trials such as the 6 minute walking distance and right heart catheterisation are limited in part by their insensitivity to change and invasive nature, respectively [30]. As highlighted by the Pulmonary Vascular Research Institute statement on imaging in PH, there is a need to identify new tools for both clinical use and for use as endpoints in studies [30]. There is particularly an unmet need for biomarkers that can help differentiate between PAH and PH-Lung [47]. Imaging biomarkers identified by QCT metrics may help solve this clinical dilemma.

Repeatable and quantifiable imaging biomarkers can measure treatment response and are currently being used as endpoints for clinical investigations and trials in emphysema and ILD

[30,126,127]. QCT metrics have been shown to predict outcomes better than lung function tests in ILD [27,113].

2.5.3.3 Enabling precision medicine and big data analysis

The numerical nature of QCT derived imaging biomarkers naturally integrates well with big data analysis in precision medicine. Precision medicine is a process of enabling targeted tailored therapies to patient groups through deep phenotyping of patients [128]. The goal is bringing together data from genetics, imaging, immunology/histology and traditional clinical assessment in a holistic manner to refine diagnosis and offer target therapies that improve outcomes. PH as a heterogeneous condition with distinct sub-phenotypes is well suited for precision medicine. A multi-domain and multimodality approach is already established for clinical assessment. The 10th biannual symposium of the International Society for Strategic Studies in Radiology recognised the implementation of quantitative imaging as critical to this goal [129]. It highlighted how imaging findings have strong yet currently untapped potential to guide patient care and influence outcome through imaging-based biomarkers.

2.6 Limitations, Challenges, and Solutions

The promise and potential of AI should be tempered by a realistic, pragmatic understanding of current limitations. We highlight three limitations to chest QCT and discuss three challenges faced by all ML, radiomic and imaging biomarker studies.

2.6.1 Limitations in Quantitative CT Research

2.6.1.1 Variations in data

A large number of steps exist within the imaging data pipeline, each of which can add variation and noise to the data. These include CT image acquisition parameters, reconstruction, segmentation, feature-extraction and post-processing algorithms. These factors reduce the robustness, performance, and generalisability of radiomics or imaging biomarker approaches, including quantitative CT.

Solutions are therefore an active area of research. The Image Biomarker Standardisation Initiative and groups such as the Quantitative Imaging Biomarker Alliance, and the Association of University Radiologists Research Alliance Quantitative Imaging Task Force are dedicated to such solutions [80,130,131]. These range from standardisation of imaging protocol to more recent

algorithms and data processing techniques that limit variance. The SPIROMICS study developed an imaging protocol specifically for Chest QCT [119]. Pyradiomics is a flexible open-source approach to feature extraction, allowing for more widespread reproducible feature analysis [132]. Differences in extracted features between different CT scanners can be tested with physical phantoms to understand the underlying variation [133]. These features can then be standardised amongst protocols to account for this invariance [124,134]. Differences in slice thickness, voxel sizes and convolutional kernels can be normalised using a range of approaches such as voxel-size resampling, batch effect correlation, and grey-level normalisation [135–137]. A predictive internal calibration approach was shown to improve performance of emphysema prediction in a COPD study [138]. Moving to an ML based automated approach for segmentation has higher accuracy and reduced variability compared to manual segmentation [139]. For DL approaches, domain adaptation and transfer learning are approaches insensitive to data heterogeneity [124,140,141]. Convolutional neural networks have been shown to dramatically improve the similarity of CT radiomic features obtained with different imaging reconstruction algorithms and kernels [142].

2.6.1.2 Inspiration, breathing and motion artefacts

Differences in lung volume secondary to inspiratory effort and artefacts from breathing and cardiac motion are inherent to chest CT [143]. They limit traditional visual asessment for radiologists and can lead to misinterpretation, as pulmonary density is influenced by respiratory phase. For a quantitative approach, these must be minimised to avoid errors propagating throughout the pipeline. Failure to do so can lead to errors where the disease is improperly quantified. Solutions include use of modern scanners with rapid acquisition times and larger detectors, clear instructions to patients to explain the importance of breath holding, and data techniques to adjust for variability. Parametric response mapping is a method where volumetric non-rigid registration of both inspiratory and expiratory scans are fused [144]. The overall lung volume can be quantified to alert for differences and used to normalise or weight measurements [145]. [D]

[D] A limitation of this approach is if there is only a single scan series at one time point (for example, baseline incident scan). Variations in lung inflation and motion artefact then cannot be accurately accounted for or corrected by the algorithm.

2.6.1.3 Lack of studies involving intravenous contrast

Most QCT applications to date use non-contrast CT. In PH, most studies involve intravenous contrast and are performed as a CT Pulmonary Angiogram (CTPA). Intravenous contrast increases the opacification of the lung parenchyma due to perfusion of a high density contrast media throughout the pulmonary vessels and lung parenchyma. Whilst the volume and speed of contrast material administered is routinely standardised, there can be variability between scans in the extent of contrast uptake within lung parenchyma. In pulmonary nodule characterization studies, this variability was reduced by acquiring images between 60-150 seconds post injection [124,146]. We found only one study assessing lung parenchymal patterns in contrast Chest CT, finding significant difference in mean lung density in patients with pulmonary embolism [147]. This is an area in need of further research. Can the existing algorithms trained on non-contrast CT be applied to CTPA? Can they be adapted using additional information such as density of contrast in main pulmonary arteries and cardiac chambers?

2.6.2 Current challenges facing machine learning research

2.6.2.1 High-quality training data is hard to obtain

Development of any algorithm requires robust training data - both in quantity and quality. The performance of ML models improves logarithmically with increased volume of training data available [148–150]. As the algorithm 'learns' through feature recognition, the quality of the training cohort fundamentally shapes its performance. An ideal training data set is contextual to the problem it seeks to address, expertly labelled, quality controlled against imaging artefacts and noise, and appropriately powered for its clinical use. It should follow the FAIR principles of scientific data management and stewardship - be Findable, Accessible, Interoperate and Reusable [151]. The lack of high-quality labelled training data is a limitation throughout all domains of ML research. Carefully preparing, validating and labelling training data often form the bulk of the development work [150].

All patient identifiable information needs to be carefully removed from any imaging data set prior to use. Although standards exist for medical imaging data such as DICOM, they are only loosely adhered to, with wide variation in the metadata [150]. Patient information can be difficult to remove, and at times hard coded into the imaging data. Clinical governance standards are stricter for medical data, and mandate secure data management and storage solutions.

To address these challenges, standardised and data validation systems have been proposed. Kohli et al proposed an extensive 16 point baseline metadata list to consider to catalog medical

imaging data [152]. A Medical imaging DAta Readiness (MiDAR) scale has been proposed as a four-point framework that assesses the readiness of medical imaging data for development [150]. All AI studies should have a data preparation and quality control framework that ensures training data robustness.

2.6.2.2 Lack of external multi centre validation and prospective studies

ML algorithms should be validated in external multi-centre cohorts to avoid overfitting. Overfitting describes the ML model being exceedingly narrow in its performance, such that it learns from noise and other specific quirks of the training data. Therefore, its performance degrades on external validation cohorts when other variances and variables are present. Overfitting is a major obstacle that hinders generalisability - the application of the clinical model in other similar cohorts or centres. This can be minimised by using a large, diverse training dataset and performing techniques such as augmentation, regularisation and dropout [153].

Of 82 studies describing 147 patient cohorts that compared AI performance vs health care professionals, Liu et al found only 25 performed external validation [70]. Of these, only 14 used the same sample for comparison. Another review found only 9 to be prospective and 6 to be tested in a real world clinical setting [154]. Both reviews found a high risk of potential bias in the validation procedures and poor methodology in study design. These findings highlight the need for further prospective studies designed with external multi-centre validation as a primary target. The need for generalisability inherently has the trade-off for poorer performance across those several centres over stronger performance at a single centre [155]. The retrospective nature of studies leads to large variations in imaging protocols, sample sizes and AI approaches. There is a need for standardisation in reporting practices [156]. In late 2020, SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials-Artificial Intelligence) and CONSORT-AI (Consolidated Standards of Reporting Trials-Artificial Intelligence) standards were published to help improve this in the context of interventions and clinical trials involving AI [157,158].

2.6.2.3 'Black box' and interpretability

DL approaches by design have multiple hidden layers that obscure the decision-making process. This lack of transparency has been labelled a 'black box' problem. Approaches are being developed that offer more insight to improve interpretation, such as through visualisation of features[159]. Medical systems and workflows value interoperability as assumptions can be

checked and errors appropriately evaluated. There are debates from a legal, regulatory and accountability standpoint regarding the suitability of such systems. Explainable and interpretable AI therefore is a field which is getting more interest. [160]. In the near future, we expect QCT models to serve as an adjunct to visual assessment. Therefore, AI and ML systems which are more clinically comprehensible and better integrated into current clinical imaging workflows will be preferred.

2.7 Conclusion

Leveraging the power of ML has demonstrated significant breakthroughs in a range of lung diseases. These can be applied to improve both radiological assessment and clinical management in pulmonary hypertension. Quantitative imaging in particular can lead to a data-driven decision-making process which combines clinical, physiological, genetic and radiological data for better assessment. This would help answer the topical clinical and radiological problems in the field - helping with better phenotyping and assessment for treatment decisions. Once validated, imaging-based biomarkers can help assess treatment response and therefore can be used as endpoints in clinical trials. The growing use of AI will help reduce errors, increase productivity, and can enable a precision medicine approach to pulmonary hypertension.

It is, however, important to be realistic and pragmatic about the current state of medical imaging using AI, with a clear understanding of its limitations. AI has been shown to excel in the narrow domain sensory image perception and identification tasks. It however does not have the ability to make broader assessments away from its domain, recognise the larger context of its use or even appraise if it is being deployed appropriately. This role falls to humans. Current AI should best be viewed and used as a specific tool in a validated narrow clinical context. Going forward, further prospective, large multi-centre studies are required to better assess technical development in clinical settings. Studies using clinically meaningful endpoints for AI algorithms such as treatment response and survival are preferred over purely technical performance in image classification.

In conclusion, for physicians managing patients with PH and associated lung disease, it is hoped that the application of AI approaches to CT imaging, may “come to the rescue” by providing mechanistic insights and improved phenotyping and by doing so facilitate much needed therapy trials.

3 Methodology: building clinico-radiological database

This chapter outlines the steps taken to build the clinico-radiological database used throughout the thesis. Methods for specific analysis are provided in each in each individual chapter.

3.1 Raw ASPIRE registry data

3.1.1 Clinical data

The ASPIRE (Assessing the Spectrum of Pulmonary Hypertension Identified at a Referral Centre) is a large database of all consecutive patients referred to the Sheffield Pulmonary Vascular Diseases Unit (SPVDU), based in Sheffield Teaching Hospitals NHS Trust. As PH is a rare and complex disease, management for adults in the UK is limited to nine specialist tertiary referral centres. SPVDU is the largest PH referral centre in the UK, and sees approximately one in every four patients in the UK[5]. This makes it internationally one of the largest PH centres.

All patients referred undergo comprehensive clinical and radiological assessment. This includes a review with a specialist PH consultant and investigations such as echocardiogram (ECG), chest X-ray (CXR), right heart catheterisation (RHC), High resolution Chest CT (HRCT), CT pulmonary angiography (CTPA), cardiac MRI, MR angiography (MRA), and isotope perfusion scan (Q scan). All patients are discussed in a multidisciplinary team (MDT) involving clinicians from respiratory, cardiology, radiology and interventional radiology.

Data from the above investigations and MDT outcome is documented and stored electronically in multiple systems. Most clinical data is stored on ‘Infoflex’ and ‘ICE’; radiological data is stored on a Picture Archiving and Communication System (PACS), and mortality data is stored in the NHS Personal Demographics Service. Data such as patient smoking history is only available through manual review of clinical notes and letters. The ASPIRE database has been built, maintained, and added to through the contributions of several researchers over the years within SPVDU.

The database is hosted on encrypted secured shared University storage. All imaging is de-identified at point of transfer from NHS PACS system to the research POLARIS XNAT imaging database. The de-identification is performed by Clinical Scientists with appropriate training. Patients are given a unique 'PHD_' identifier which is then used for matching with the ASPIRE clinical database. A separate spreadsheet, stored in a different location on encrypted storage, contains the link between the patient's hospital numbers and unique 'PHD_' identifiers. There is a dedicated data management plan in place for this data.

3.1.2 CT imaging data

CT scans in Sheffield are typically acquired on a light-speed 64-slice MDCT scanner (GE Healthcare). Standard acquisition parameters are: 100 mA with automated dose reduction, 120 kV, pitch 1, rotation time 0.5s and 0.625 mm collimation. There can be variation in these parameters due to different pre-set settings and inbuilt algorithms which automatically adjust acquisition parameters based on factors such as patient body mass index. There can also be a variation in between scanner types and scanner manufacturers, each of which can have different default acquisition parameters and adjustment algorithms. A 400 mm × 400 mm field of view with an acquisition matrix of 512 × 512 is a standard parameter for all CT.

The routine scanning protocol for a PH patient in Sheffield involves a volumetric CT pulmonary angiogram (CTPA), with High Resolution CT (HRCT) reconstructions and additional sequential expiratory slices. Slice thickness is currently set at 0.5 mm. HRCT was historically a term used for highly collimated CT acquisition, to enable better spatial resolution for image analysis. Acquisition could either be volumetric (continuous acquisition with no gaps) or sequential (gaps in between slices). Modern CT scanners have made significant advantages in dose reduction through multiple different scientific and technical advantages, such as algorithmic iterative reconstruction, wide detector arrays and low tube potential[161]. CTPA scanning involves the administration of intravenous contrast to opacify the pulmonary vascular structures. Scans performed by the National Health Service (NHS) Sheffield Teaching Hospitals (STH) NHS Trust are stored in their Picture Archiving and Communication System (PACS) as a DICOM (Digital Imaging and Communications in Medicine) file. These scans are linked to the patient with multiple parameters such as a unique hospital number, date of birth and NHS number.

The POLARIS (Pulmonary, Lung and Respiratory Imaging Sheffield) XNAT database is a platform hosted at the servers of the Academic Unit of Radiology at the University of Sheffield. It stores pseudonymised medical imaging data, which is only accessible through appropriate authentication. Each patient has a unique identifier, which can be linked to their hospital number via a codex sheet. Download of data from STH PACS to POLARIS XNAT is done by Clinical Imaging Scientists to ensure it is appropriately anonymised, stored and secured.

3.2 Merging imaging and clinical data

I downloaded all DICOM CT scan series of patients identified from the ASPIRE registry. Each DICOM file contain metadata and technical information. This includes scanner type, manufacturer, number of series, individual series ID, kernel and reconstruction algorithms used, x and y spacing, slice thickness, kvp, number of slices, date, orientation, study and series ID. This information was extracted directly from the DICOM file into a .csv file using a python script. The script was written by Michael Sharkey, Clinical Scientist.

Data from all sources – ASPIRE, POLARIS XNAT, NHS Personal Demographics Service, and raw DICOM information, was merged used 'R' software (major version 4) into one database file. Multiple unique identifier (such as hospital number, date of birth, date of study) were used to match across databases to create the imaging database for the study. I wrote all the R scripts to perform the merge. I also maintained and updated the database with new data throughout the thesis.

3.3 Quality control and manual radiological review

The aim of data quality control and radiological review is to take raw NHS data and transform it to be appropriate for AI development. This stage is vital to success of the project, and is often a limiting step as it is the most time consuming, and requires specialist domain expertise[150]. All steps here were taken by me (speciality trainee radiologist with Fellowship of the Royal College of Radiologists), and supported by the supervisor (specialist Cardiothoracic Radiologist). All analysis was done in R, with scripts written by me.

The MIDaR scale is a standardised approach for preparing medical data for machine learning tasks in radiology[150]. It incorporates the FAIR guiding principles for scientific data management and stewardship of Findable, Accessible, Interoperable and Reusable[151]. The

MIDaR scale ranges from level D where data is inaccessible, un-anonymised and immeasurable in terms of quality and quantity to level A where data is contextual, annotated and appropriately powered for studies and AI development.

Each CT scan event is stored as a single 'study' DICOM file with a unique 'study ID'. Each 'study' can contain multiple scans, each termed a 'series'. Each 'series' is a distinct acquisition event. For example, a single PH Chest CT 'study' typically contain three different scan acquisitions ('series') – volumetric CTPA, volumetric HRCT construction, and sequential HRCT expiratory slices.

I manually reviewed each scan 'series' within each 'study'The parameters I evaluated were:

1. Scan acquisition type
2. Radiologically acceptable scan quality, defined as:
    a. No severe breathing or motion artefact
    b. No severe noise
    c. 'Complete' scan in which the entire chest is imaged in the study
3. Contrast phase (CTPA phase required)
4. Slice thickness <1.25 mm

Of 900 scans, 564 were thin slice CTPA scan with radiologically acceptable scan quality. The main reason for excluding scans was incorrect scan acquisition type and errors in the coded DICOM information. The DICOM file may have been labelled as 'CTPA' but upon review it was not performed in the right phase, did not contain contrast, or had thick slices. 12 were excluded due to breathing or motion artefact. Two cases were excluded due to lobectomy.

This information was added onto the database. For scans with multiple series, information for each series was added as separate columns. This ensured the database was at a patient level and represented 'tidy' data, with each row identifying a unique patient. 'Tidy data' refers to a structured database which is suited for programmatic statistical analysis[162]. Every column is a unique variable, every row is a unique observation, and every cell is a single value. Although time consuming, the manual review and data tidying stages were essential for establishing confidence in the data prior to any clinical analysis or AI development.

4 Computed tomography lung parenchymal descriptions in routine radiological reporting have diagnostic and prognostic utility in patients with idiopathic pulmonary arterial hypertension and pulmonary hypertension associated with lung disease

*** Peer-reviewed journal article published in *European Respiratory Journal Open Access (2021)*** [49]

Contribution: First author. Contributor Roles Taxonomy (CRediT) – conceptualisation, data curation, formal analyses, funding acquisition, methodology, project administration, resources, software, validation, visualisation, writing – original draft, review, and editing).

I lead this study. I performed all the data analysis and independently wrote the code for the regex string search algorithm that is used for extracting information for the radiology reports. I did almost all the writing for this paper and wrote the initial draft. RC contributed with editing my initial draft for publication, and with the changes made after reviewer comments. DGK and AS as primary supervisors and senior authors helped guide the project. All other authors had a small involvement with data gathering, support, project administration and resources.

Krit Dwivedi, Robin Condliffe, Michael Sharkey, Robert Lewis, Samer Alabed, Smitha Rajaram, Catherine Hill, Laura Saunders, Peter Metherall, Faisal Alandejani, Dheyaa Alkhanfar, Jim M Wild, Haiping Lu, David G Kiely and Andrew J Swift.

Take home message / shareable abstract: Routine radiological reports describing extent of CT lung parenchymal disease can identify groups of patients IPAH and CLD-PH with significantly different prognoses.

4.1 Abstract

Background: Patients with pulmonary hypertension (PH) and lung disease may pose a diagnostic dilemma between idiopathic pulmonary arterial hypertension (IPAH) and PH associated with lung disease (PH-CLD). The prognostic impact of common CT parenchymal features is unknown.

Study Design and Methods: 660 IPAH and PH-CLD patients assessed between 2001-19 were included. Reports for all CT scans one year prior to diagnosis were analysed for common lung parenchymal patterns. Cox regression and Kaplan-Meier analysis was performed.

Results: At univariate analysis of the whole cohort, centrilobular ground glass (CGG) changes (Hazard Ratio, HR 0.29) and GGO (HR 0.53) predicted improved survival while honeycombing (HR 2.79), emphysema (HR 2.09) and fibrosis (HR 2.38) predicted worse survival (p all <0.001). Fibrosis was an independent predictor after adjusting for baseline demographics, PH severity and DLco (HR 1.37, p<0.05). Patients with a clinical diagnosis of IPAH who had an absence of reported parenchymal lung disease (IPAH-noLD) demonstrated superior survival to patients diagnosed with either IPAH who had coexistent CT lung disease or PH-CLD (2-year survival of 85%, 60% and 46% respectively, p<0.05). CGG changes were present in 23.3% of IPAH-noLD and 5.8% of PH-CLD patients. There was no significant difference in survival between IPAH-noLD patients with or without CGG changes. PH-CLD patients with fibrosis had worse survival than those with emphysema.

Interpretation: Routine clinical reports of CT lung parenchymal disease identify groups of patients IPAH and CLD-PH with significantly different prognoses. Isolated CGG changes are not uncommon in IPAH and are not associated with worse survival.

4.2 Introduction

Pulmonary hypertension (PH) is a heterogenous, progressive and incurable condition associated with significant morbidity and mortality. Five classification groups with similar clinical and pathological characteristics are described including group 1 (Pulmonary Arterial Hypertension, PAH) and Group 3 (PH due to chronic lung disease and/or hypoxia, PH-CLD)[6]. PH-CLD most commonly complicates chronic obstructive pulmonary disease (COPD) and/or emphysema, interstitial lung disease (ILD) and alveolar hypoventilation[23,26,34,35]. Patients with PH-CLD typically present with mild to moderate PH, although a small proportion of patients have severe PH. In contrast patients with idiopathic PAH tend to have more severe PH at presentation. Accurate classification of the form of PH is important as it informs prognosis and impacts on treatment decisions. In PH-CLD the importance of haemodynamically characterising disease severity has also been highlighted[51,163,164]

In practice, distinguishing between IPAH and PH-CLD may be challenging [47]. Recommendations from the 6th World Symposium on PH recommended that patients with minor lung disease, who otherwise meet criteria for IPAH, may be considered to have IPAH [26]. More recently it has been suggested that those with mild lung disease who are haemdynamically similar to IPAH with a so-called "pulmonary vascular phenotype"[23,36] are not part of the IPAH continuum but a different entity closer to PH-CLD[165]. We have recently demonstrated that the presence of a reduced diffusion capacity for carbon monoxide (DLco) percent predicted (<45%) is associated with poorer survival and response to PAH therapy in patients diagnosed with IPAH [42], in a carefully phenotyped population with minor lung disease were excluded. Hoeper *et al* subsequently identified a cluster amongst patients diagnosed with IPAH characterised by older age, frequent comorbidities, a higher proportion of males and a reduced DLco [166]

In addition to pulmonary function assessment, many patients undergoing assessment for suspected PH also undergo chest CT imaging. We therefore hypothesised that descriptions of lung parenchyma at routine radiological reporting could be used to predict outcomes in patients with IPAH and PH-CLD.

4.3 Methods

4.3.1 Patient cohort

Patients assigned a diagnosis of IPAH, Hereditary PAH (HPAH), or PH-CLD between February 2001 and January 2019 were identified from the ASPIRE registry (a database of consecutive patients referred to the Sheffield Pulmonary Vascular Diseases Unit). All patients underwent comprehensive multimodality assessment including right heart catheterisation. Patients with PH-CLD associated with conditions other than COPD and/or emphysema or ILD were excluded. Patients with two or more radiological features of possible pulmonary veno-occlusive disease (centrilobular ground-glass opacities, significant mediastinal lymphadenopathy and interlobular septal lines) were also excluded [167].

4.3.2 CT analysis

All CT scans, including those performed externally, were reported by specialist cardiothoracic radiologists with expertise in pulmonary hypertension. Patients were assigned diagnoses following a multidisciplinary meeting. The clinical reports of CT scans performed in the year prior to diagnosis were retrieved. Reports were searched using a regular expression string-search function for mention of 6 specific lung parenchymal patterns - centrilobular ground glass (CGG) changes, ground glass opacification (GGO), honeycombing, consolidation, fibrosis, and emphysema. Each result was manually validated to ensure they represented a true positive. Reports containing false positive phrases such as ‘no evidence of emphysema’ were not counted. The radiologist (KD) who reviewed the CT reports was blinded to the results of other investigations. [E]

4.3.3 Clinical and morality data

Clinical data collected included age, sex, World Health Organization (WHO) functional class (FC), pulmonary haemodynamics obtained at right heart catheterisation and pulmonary function tests. Mortality data were obtained from systems linked to the National Health Service Personal Demographics Service (PDS), which is updated when a death is registered in the UK.

[E] Further detail discussed in the viva. ‘Minor’ disease was coded as ‘mild’ disease. ‘Mild to moderate’ was coded as ‘moderate’ and ‘moderate to severe’ was coded as ‘severe’.

Patients who emigrated (n=3) were excluded, as were patients without a record on the PDS (n=2). Patients undergoing lung transplantation were censored at the time of surgery, and mortality data were collected using a census date of May 31, 2019.

4.3.4 IPAH sub-group analysis

Patients with an initial diagnosis of IPAH but with a radiological report of a degree of emphysema or fibrosis were reclassified as 'IPAH-Lung Disease' (IPAH-LD). The remainder of the patients were classified as IPAH-no Lung Disease (IPAH-noLD). Separate analysis also compared the effect of patients with initial diagnosis of IPAH and CGG changes. Patients with no significant lung disease and CGG were reclassified as 'IPAH-CGG'. Survival and group comparison was performed between IPAH-CGG, IPAH-LD and IPAH-noLD. Those with both CGG and significant lung disease (n=8) were excluded from this sub-group analysis.

4.3.5 Statistics

Analysis was performed using R statistical software package using version 4.0.3, using packages 'tidyverse' and 'survminer', and SPSS version 26.0 (IBM Corp). Continuous data is presented as mean±SD (compared using paired/unpaired t-tests) or median (interquartile range) for nonparametric data (compared using Wilcoxon signed-rank/Mann–Whitney U-tests). Frequencies are compared using the Chi-square test. Categorical variables are shown in magnitude and percentage.

Cox proportional hazard's regression was used to determine association between different CT parenchymal features and survival. Hazard ratios are presented with 95% confidence intervals. Three separate multivariate models were created. Model 1 adjusted for patient demographics: age, gender and WHO functional class. Model 2 adjusted for all demographics and additionally mPAP. Model 3 adjusted further for diffusing capacity for carbon monoxide (DLco).[F] Kaplan–

[F] Model 3 includes DLCO, which is known to be prognostic and correlate with extent of emphysema and fibrosis. This analysis was performed to see if the CT features are prognostic in addition to the effect of DLCO. A limitation of this analysis is that a small prognostic effect of CT features will be 'subtracted' and therefore not meet statistical significance in emphysema and fibrosis. The three separate prognostic models were used in the methodology to allow us to investigate, and account for this effect, as model 2 does not include DLCO. This has been discussed in the discussion section of this paper.

Meier survival curves were compared using the log-rank test, truncated at 5 years. Group comparisons were made with two-tailed ANOVA with post-hoc Bonferroni correction.

4.3.6 Ethics

Ethical approval was granted by Sheffield Teaching Hospitals NHS Foundation Trust (STH14169) and approved by the National Research Ethics Service (16/YH/0352). All research was conducted in agreement with the Declaration of Helsinki and the European General Data Protection Regulation.

4.4 Results

4.4.1 Patient characteristics

From 5643 patients diagnosed with all forms of pulmonary hypertension, 660 patients met the inclusion criteria and formed the main study cohort (Figure 1). This included 335 patients diagnosed with IPAH and 325 with PH-CLD, who formed the sub-groups for analysis. Patients with PH-CLD were more frequently male (58% vs 39%, p<0.001), older (67 ±17 years vs 60 years ±17 years, p<0.001), had lower mPAP (42 ±10 mmHg vs 53 ±12mmHg, p<0.001) and DLco (28 ±14% vs 44 ± 20%, p<0.001) compared to those diagnosed with IPAH (**table 1**). Two hundred and eighty-three (43%) patients had imaging performed externally. RHC data was available in 100% and PFT data in 95% of patients.

4.4.2 Cox regression analysis

Univariate regression results for demographics, clinical parameters and CT features are shown in **Table 2**. Being older, male, and having a higher WHO Functional Class were significant univariate predictors of poor survival across all groups. **Table 3** shows the results for the different multivariate regression models performed on significant univariate variables.

4.4.2.1 Main Cohort

CGG changes (HR 0.29, 0.17-0.50) and GGO (HR 0.53, 0.38-0.74) were significant (p<0.001) predictors of improved survival at univariate analysis while honeycombing (HR 2.79, 1.57-4.99), emphysema (HR 2.09, 1.71-2.56) and fibrosis (HR 2.38, 1.94-2.91) were significant predictors of poor survival (p all <0.001). After adjusting for the impact of age, gender and WHO FC in multivariate model 1, CGG changes (HR 0.50, 0.28-0.89, p=0.01), emphysema (HR 1.48, 1.21-1.83, p<0.001) and fibrosis (HR 1.75, 1.42-2.15, p<0.001) remained significant predictors of outcome. These parameters were also significant predictors of mortality in model 2 after additionally adjusting for the severity of PH. In model 3, fibrosis (HR 1.37, 1.09-1.73, p=0.008) remained an independent predictor after additionally adjusting for DLco.

4.4.2.2 IPAH

In patients assigned a diagnosis of IPAH, the presence of CGG changes (HR 0.44, 0.25-0.78, p=0.005) or GGO (HR 0.52, 0.32-0.86, p=0.01) predicted improved outcomes while the presence of any degree of emphysema (HR 2.74, 1.96-3.81, p<0.001), fibrosis (HR 2.48, 1.76-3.50, p<0.001) or honeycombing (HR 3.72, 1.37-10.1, p=0.01) were significant univariate predictors of increased mortality. Subgroup univariate analysis of IPAH patients with no degree of parenchymal lung disease demonstrated that CGG changes and GGO no longer significantly predicted survival (**table A1**). In multivariate model 1 of the overall IPAH group, only the presence of emphysema (HR 1.72, 1.22-2.42, p=0.002) remained a significant predictor. Emphysema was also a significant prognostic factor in model 2 (HR 1.76, 1.24-2.49, p=0.002) but not in model 3 (p=0.3).

4.4.2.3 PH-CLD

Fibrosis (HR 1.83, 1.04-4.30, p<0.001) and honeycombing (HR 2.11, 1.04-4.30, p=0.039) were significant predictors of mortality at univariate analysis while the presence of emphysema or GGO did not predict mortality. Fibrosis was an independent predictor in all multivariate models, including after adjustment for DLco (Model 3, HR 1.46, 1.09-1.96, p<0.001). Honeycombing was not a significant predictor in any of the multivariate models.

4.4.3 Prognostic effect of extent of emphysema and fibrosis

Increasing extent of both emphysema and fibrosis derived from radiology reports was associated with increasing risk of mortality in the whole cohort at univariate analysis (Emphysema; mild: HR 1.78 (1.3-2.43), moderate: HR 2.18 (1.69-2.81), severe: HR 2.92 (2.15-3.97). Fibrosis; mild: HR 1.94 (1.46-2.58), moderate: 2.77 (1.99-3.85), severe: 3.19 (2.3-4.43), **table 2**). A similar pattern was observed in the IPAH group while in the PH-CLD group the presence of emphysema was not associated with significant increased mortality.

4.4.4 Kaplan-Meier survival analysis stratified by CT Features

The Kaplan-Meier survival curves for each cohort are presented in **figure 2**. In the full cohort, survival was significantly worse in patients with any form of parenchymal lung disease

(p<0.001 in all). Survival was also significantly worse in patients with reported fibrosis compared to emphysema (p=0.023).

In patients diagnosed with IPAH, the presence of any parenchymal lung disease was associated with significantly poorer survival (p all<0.001) with no significant difference in survival between patients with the different forms of lung disease. There was no significant difference in survival between patients with isolated CGG compared to those with no reported parenchymal abnormalities (p=0.57).

In patients diagnosed with PH-CLD, survival in patients with emphysema was superior to survival in patients with fibrosis (p=0.02). Although some evidence of improved survival was observed between patients without parenchymal lung disease compared to those with emphysema, differences between groups did not meet conventional levels of statistical significance (p=0.09).

4.4.5 IPAH sub-group analysis

4.4.5.1 Impact of lung disease in IPAH compared to PH-CLD

Patients with an initial clinical diagnosis of IPAH (n=335) were firstly reclassified as either IPAH-lung disease (IPAH-LD, n=138) or IPAH-no lung disease (IPAH-noLD, n=197), based on the presence or absence of emphysema and/or fibrosis. Survival in patients with IPAH-noLD was significantly superior to patients with IPAH-LD or PH-CLD (p<0.001, **figure** 3). Patients with IPAH-LD were distinct from those diagnosed with PH-CLD with less severely impaired spirometry but more severely abnormal pulmonary haemodynamics (**table A2)**. There was no statistically significant difference in survival (p=0.065) in patients with PH-CLD when compared to IPAH-LD. Two-year survival for patients with IPAH-noLD, IPAH-LD and PH-CLD was 85%, 60% and 46%, respectively (**table A3**).

4.4.5.2 Impact of CGG

The impact of CGG changes on characteristics and survival of patients with an initial clinical diagnosis of IPAH (n=335) was then assessed. CGG changes were uncommon in patients with co-exisiting lung disease (being present in 8/138 patients (5.8%) with IPAH-LD, **table A2**) compared with patients with no co-existing lung disease (being present in 46/197 patients

(23.3%), **table 4**). Eight patients who had both CGG changes and parenchymal lung disease were excluded from this analysis. One hundred and thirty patients with emphysema or fibrosis were therefore reclassified as IPAH-LD, 46 patients with CGG changes as IPAH-CGG and 151 patients with no parenchymal abnormalities as IPAH-noLD. Patients with IPAH-CGG had more severe PH but a similar DLco when compared with patients with IPAH-noLD (**table 4**). Patients with IPAH-LD were older, with a greater proportion in WHO FC IV and less severe PH but a lower DLco than patients with IPAH-noLD. There was no significant difference in survival between patients with IPAH-noLD and IPAH-CGG while survival in patients with IPAH-LD was significantly worse (**figure 4**).

4.5 Discussion

We have demonstrated that parenchymal abnormalities noted at routine radiological reporting of CT scans performed in patients with suspected IPAH or PH-CLD have diagnostic and prognostic utility. We have also found that isolated CGG changes are not uncommon in patients with IPAH and are not associated with lower DLco or worse survival.

4.5.1 Routine radiological reports of the presence of emphysema or fibrosis have diagnostic and prognostic utility in patients diagnosed with IPAH

Differentiating PH-CLD from IPAH is an important part of the PH diagnostic algorithm and may be straightforward in the presence of severe lung function abnormalities or severe parenchymal lung disease. The presence of less severe degrees of lung disease provides more of a diagnostic challenge as has been highlighted by others [47,165]. The 6th World Symposium on Pulmonary Hypertension (WSPH) task force suggested that patients with significant PH but only modest parenchymal abnormalities should be assigned a diagnosis of IPAH[26]. In the present study we found that radiological reports of emphysema or fibrosis in patients who had been clinically assigned a diagnosis of IPAH in keeping with the 6th WSPH recommendations was associated with significantly worse survival at Kaplan-Meier analysis. This supports the comments of Godinas *et al* who suggest that such patients represent a distinct group from those with IPAH being phenotypically closer to PH-CLD[165].

However, we also observed at Cox regression analysis that this prognostic importance of emphysema or fibrosis in patients diagnosed with IPAH was not independent of DLco. The prognostic importance of reduced DLco with or without co-existing parenchymal lung disease in patients assigned a diagnosis of IPAH has previously been reported. [41,168,169] We have previously demonstrated that response to PAH therapies in patients with IPAH-noLD and DLco <45% is lower than in patients with IPAH-noLD and DLco ≥45%, while on average patients with IPAH-LD exhibit a lack of response to PAH therapies in terms of exercise capacity and quality of life.[41] Therefore, the identification of emphysema or fibrosis on CT provides clinically-relevant information in addition to that provided by DLco alone.

4.5.2 Routine radiological reports of the extent and nature of parenchymal lung disease have prognostic utility

In addition to the *presence* of parenchymal lung disease identified at routine reporting being associated with worse survival, we also observed that the *extent* and *nature* of lung disease described at routine reporting provided additional prognostic information. In patients originally assigned a diagnosis of IPAH, the extent of emphysema or fibrosis described qualitatively at reporting was strongly associated with prognosis. In patients with PH-CLD, the nature of parenchymal lung disease impacted on survival with the presence of fibrosis conveying a worse prognosis than emphysema, as previously described.[50]

4.5.3 Centrilobular ground glass in IPAH

CGG changes in the absence of emphysema or fibrosis were not uncommon, being reported in 23.3% of those IPAH patients who had no emphysema or fibrosis but were only reported in only 5.8% of patients with co-existing emphysema or fibrosis. The nature of CGG changes in patients diagnosed with IPAH is not clear. It is possible they represent pulmonary veno-occlusive disease (PVOD), however we excluded patients with an additional radiological feature of PVOD (significant mediastinal lymphadenopathy or interlobular septal lines). Furthermore, there was no significant difference in DLco compared to IPAH-noLD patients without CGG changes. Nolan *et al* postulated that they represented cholesterol granulomas while Horton *et al* presented a case report of a patient with fenfluramine-induced PAH with diffuse micronodules on CT who had an extensive diffuse plexogenic arteriopathy at lung biopsy.[170,171] In addition to IPAH, CGG changes have also been commonly described in other forms of PAH including PAH associated with connective tissue disease and congential heart disease [28]. In Eisenmenger syndrome and IPAH it has been postulated that this may be a feature of lung neovasulcarity[172 173]. It is interesting to note that in our study there was no significant difference in survival compared with patients with IPAH-noLD at Kaplan-Meier analysis despite having significantly more severe pulmonary haemodynamics.

4.6 Limitations

This study has utilised clinical radiological reports involving qualitative descriptions from the time of PH diagnosis rather than subsequent quantitative analysis. By using this approach, we have, however, been able to demonstrate that features described in routine "real-world" radiological reports have the ability to refine diagnostic and prognostic processes. It may well be that more in-depth quantitative analysis utilising artificial intelligence algorithms may provide superior diagnostic and prognostic information and further studies are therefore warranted.[49]

4.7 Conclusion

CT lung parenchymal descriptions in routine radiological reporting have diagnostic and prognostic utility in patients with idiopathic pulmonary arterial hypertension and pulmonary hypertension associated with chronic lung disease. Chest CT features should therefore be considered in patient assessment and risk-stratification. Patients diagnosed with IPAH who have modest lung disease demonstrate unique clinical and survival characteristics and are likely to represent a distinct phenotype separate from IAPH. Centrilobular ground glass changes are not uncommon in patients with IPAH and are associated with more severe pulmonary haemodynamics but non-inferior survival.

4.8 Acknowledgements

This research was funded in whole, or in part, by the Wellcome Trust [Grant numbers Krit Dwivedi 222930/Z/21/Z and 4ward North 203914/Z/16/ ; Andrew Swift AJS 205188/Z/16/Z]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission

4.9 Tables

4.9.1 Table 1: Baseline characteristics.

	Full Cohort	**Subgroups**		
Characteristic	**N = 660**	**IPAH, N = 335**	**PH-CLD, N = 325**	**p-value**
Age at diagnosis	64 (15)	60 (17)	67 (11)	<0.001
Male Gender	318 (48%)	131 (39%)	187 (58%)	<0.001
WHO Functional Class				0.020
II	78 (12%)	44 (13%)	34 (10%)	
III	398 (61%)	213 (64%)	185 (57%)	
IV	181 (28%)	76 (23%)	105 (32%)	
CT – Centrilobular Ground Glass (CGG)	54 (8.2%)	54 (16%)	0 (0%)	<0.001
CT – Ground Glass Opacification (GGO)	93 (14%)	62 (19%)	31 (9.5%)	<0.001
CT – Honeycombing	15 (2.3%)	5 (1.5%)	10 (3.1%)	0.2
CT – Consolidation	31 (4.7%)	9 (2.7%)	22 (6.8%)	0.013
CT – Fibrosis	213 (32%)	72 (21%)	141 (43%)	<0.001
CT – Fibrosis (by severity)				<0.001
None	447 (68%)	263 (79%)	184 (57%)	
Mild	82 (12%)	54 (16%)	28 (8.6%)	
Moderate	53 (8.0%)	9 (2.7%)	44 (14%)	
Severe	48 (7.3%)	0 (0%)	48 (15%)	

Unknown	30 (4.5%)	9 (2.7%)	21 (6.5%)	
CT – Emphysema	302 (46%)	98 (29%)	204 (63%)	<0.001
CT – Emphysema (by severity)				<0.001
None	358 (54%)	237 (71%)	121 (37%)	
Mild	77 (12%)	48 (14%)	29 (8.9%)	
Moderate	129 (20%)	38 (11%)	91 (28%)	
Severe	69 (10%)	5 (1.5%)	64 (20%)	
Unknown	27 (4.1%)	7 (2.1%)	20 (6.2%)	
CT – CPFE	101 (15%)	32 (9.6%)	69 (21%)	<0.001
mPAP (mmHg)	47 (13)	53 (12)	42 (10)	<0.001
mRAP (mmHg)	10.2 (5.6)	11.4 (5.8)	9.0 (5.1)	<0.001
PAWP (mmHg)	11.3 (3.8)	10.8 (3.3)	11.8 (4.2)	0.002
Cardiac output (L/min)	4.65 (1.66)	4.31 (1.62)	5.00 (1.64)	<0.001
Cardiac index (L/min/m^2)	2.53 (0.87)	2.32 (0.81)	2.73 (0.87)	<0.001
PVR (Wood Units)	9.0 (5.1)	11.0 (5.2)	7.0 (4.1)	<0.001
SvO_2 %	63 (9)	61 (9)	65 (8)	<0.001
FEV_1 % predicted	72 (25)	83 (18)	60 (25)	<0.001
FVC % predicted	88 (26)	96 (20)	80 (28)	<0.001
FEV_1 / FVC Ratio	66 (15)	71 (10)	61 (18)	<0.001
DLco % predicted	37 (19)	44 (20)	28 (14)	<0.001

Table 4.1: Baseline characteristics

Data are presented as number (percentage) or mean (standard deviation). Abbreviations: CT – Computed Tomography, WHO – World Health Organisation, CPFE – Combined Pulmonary Fibrosis and Emphysema, mPAP – mean pulmonary arterial pressure, mRAP – mean right atrial pressure, PAWP – pulmonary arterial wedge pressure, PVR – pulmonary vascular resistance, SvO_2 – mixed venous oxygen saturation. FEV_1 – forced expiratory volume in 1 second, FVC – forced vital capacity, DLco – diffusing capacity of carbon monoxide

4.9.2 Table 2: Univariate analysis

	Full cohort (n= 660)			IPAH (n= 335)			PH-CLD (n= 325)		
Characteristic	**HR**	**95% CI**	**p-value**	**HR**	**95% CI**	**p-value**	**HR**	**95% CI**	**p-value**
CT – Centrilobular Ground Glass (CGG)	0.29	0.17, 0.50	**<0.001**	0.44	0.25, 0.78	**0.005**			
CT – Ground Glass Opacification (GGO)	0.53	0.38, 0.74	**<0.001**	0.52	0.32, 0.86	**0.010**	0.72	0.46, 1.14	0.2
CT – Honeycombing	2.79	1.57, 4.99	**<0.001**	3.72	1.37, 10.1	**0.010**	2.11	1.04, 4.30	**0.039**
CT – Consolidation	0.84	0.50, 1.40	0.5	1.10	0.45, 2.68	0.8	0.60	0.32, 1.13	0.11
CT – Fibrosis	2.38	1.94, 2.91	**<0.001**	2.48	1.76, 3.50	**<0.001**	1.83	1.42, 2.35	**<0.001**
CT – Fibrosis (any present, ref: none)									
None	—	—		—	—		—	—	
Mild	1.94	1.46, 2.58	**<0.001**	2.51	1.72, 3.66	**<0.001**	1.73	1.11, 2.71	**0.016**
Moderate	2.77	1.99, 3.85	**<0.001**	4.53	2.07, 9.92	**<0.001**	1.71	1.18, 2.48	**0.005**
Severe	3.19	2.30, 4.43	**<0.001**				1.98	1.40, 2.80	**<0.001**
CT – Emphysema	2.09	1.71, 2.56	**<0.001**	2.74	1.96, 3.81	**<0.001**	1.13	0.87, 1.47	0.4
CT – Emphysema (any present, ref: none)									
None	—	—		—	—		—	—	
Mild	1.78	1.30, 2.43	**<0.001**	2.90	1.89, 4.46	**<0.001**	0.89	0.56, 1.42	0.6
Moderate	2.18	1.69, 2.81	**<0.001**	3.16	2.01, 4.97	**<0.001**	1.13	0.82, 1.55	0.5
Severe	2.92	2.15, 3.97	**<0.001**	11.1	3.92, 31.6	**<0.001**	1.37	0.97, 1.93	0.075

CT – CPFE	2.20	1.72, 2.80	**<0.001**	2.09	1.33, 3.29	**0.001**	1.82	1.36, 2.44	**<0.001**
Age at diagnosis	1.05	1.04, 1.05	**<0.001**	1.06	1.05, 1.08	**<0.001**	1.02	1.01, 1.04	**<0.001**
Male Gender	1.66	1.36, 2.03	**<0.001**	1.59	1.16, 2.18	**0.004**	1.42	1.10, 1.83	**0.007**
WHO Functional Class III & IV (ref: I & II)	1.78	1.45, 2.18	**<0.001**	1.88	1.34, 2.64	**<0.001**	1.74	1.35, 2.25	**<0.001**
WHO Functional Class									
II	—	—		—	—		—	—	
III	2.57	1.70, 3.89	**<0.001**	3.00	1.51, 5.96	**0.002**	2.45	1.46, 4.13	**<0.001**
IV	5.08	3.31, 7.79	**<0.001**	5.49	2.69, 11.2	**<0.001**	4.80	2.81, 8.21	**<0.001**
mPAP (mmHg)	0.99	0.98, 1.00	**0.028**	0.98	0.97, 0.99	**0.006**	1.04	1.03, 1.05	**<0.001**
mRAP (mmHg)	1.01	0.99, 1.03	0.4	1.03	1.00, 1.06	**0.023**	1.03	1.00, 1.05	0.057
PAWP (mmHg)	1.02	0.99, 1.04	0.3	1.03	0.98, 1.09	0.2	0.98	0.95, 1.01	0.2
Cardiac output (L/min)	0.92	0.86, 0.98	**0.009**	0.88	0.78, 0.99	**0.035**	0.79	0.72, 0.86	**<0.001**
Cardiac index ($L/min.m^2$)	0.87	0.76, 0.98	**0.028**	0.80	0.63, 1.01	**0.063**	0.61	0.51, 0.74	**<0.001**
PVR (Wood Units)	1.04	1.00, 1.08	**0.036**	0.98	0.91, 1.05	0.5	1.10	1.05, 1.16	**<0.001**
SvO2 %	0.98	0.97, 0.99	**0.001**	0.97	0.95, 0.98	**<0.001**	0.96	0.94, 0.97	**<0.001**
FEV_1 % predicted	1.00	0.99, 1.00	**0.032**	0.99	0.98, 1.00	0.11	1.01	1.01, 1.02	**<0.001**
FVC % predicted	1.00	0.99, 1.00	**0.028**	1.00	0.99, 1.01	0.9	1.00	1.00, 1.01	0.4
FEV_1/FVC Ratio	0.99	0.99, 1.00	**0.038**	0.97	0.95, 0.98	**<0.001**	1.02	1.01, 1.03	**<0.001**
Dlco % predicted	0.95	0.95, 0.96	**<0.001**	0.95	0.94, 0.96	**<0.001**	0.96	0.95, 0.97	**<0.001**

Table 4.2 Univariate analysis

Abbreviations: HR - Hazard Ratio, CI - Confidence Interval , CT – Computed Tomography, WHO – World Health Organisation, CPFE – Combined Pulmonary Fibrosis and Emphysema, mPAP – mean pulmonary arterial pressure, mRAP – mean right atrial pressure, PAWP – pulmonary arterial wedge pressure, PVR – pulmonary vascular resistance, SvO2 – mixed venous oxygen saturation. FEV1 – forced expiratory volume in 1 second, FVC – forced vital capacity, Dlco – diffusing capacity of carbon monoxide.

4.9.3 Table 3: Multivariate analysis [G]

	Univariate			**Multivariate 1** (adjusted for demographics – age, gender, WHO FC)			**Multivariate 2** (adjusted for demographics & mPAP)			**Multivariate 3** (adjusted for demographics, mPAP & DLco)		
Full Cohort (n= 660)												
Characteristic	HR	95% CI	p-value	HR	95% CI	p-value	HR	95% CI	p-value	HR	95% CI	p-value
CT - Centrilobular Ground Glass (CGG)	0.29	0.17, 0.50	**<0.001**	0.50	0.28, 0.89	**0.010**	0.48	0.26, 0.86	**0.007**	0.76	0.39, 1.47	0.4
CT - Ground Glass Opacification (GGO)	0.53	0.38, 0.74	**<0.001**	0.82	0.58, 1.16	0.2	0.84	0.60, 1.19	0.3	0.80	0.54, 1.18	0.2
CT - Honeycombing	2.79	1.57, 4.99	**<0.001**	1.73	0.97, 3.11	**0.087**	1.74	0.97, 3.12	0.086	1.10	0.54, 2.24	0.8

[G] A discussion during the viva was the surprise that honeycombing is not prognostic model 3. This may be because of the relatively low prevalence (therefore sample size) of honeycombing in this cohort (only 2% of subjects had it). Another possible limitation is that the scans were not re-reviewed for honeycombing, so 'minor' disease may have not been reported by the original reporting radiologist.

CT - Emphysema	2.09	1.71, 2.56	**<0.001**	1.48	1.21, 1.83	**<0.001**	1.52	1.23, 1.88	**<0.001**	1.13	0.89, 1.44	0.3
CT - Fibrosis	2.38	1.94, 2.91	**<0.001**	1.75	1.42, 2.15	**<0.001**	1.77	1.43, 2.18	**<0.001**	1.37	1.09, 1.73	**0.008**
IPAH (n=335)												
CT - Centrilobular Ground Glass (CGG)	0.44	0.25, 0.78	**0.005**	0.79	0.43, 1.45	0.4	0.80	0.43, 1.48	0.5	0.92	0.47, 1.82	0.8
CT - Ground Glass Opacification (GGO)	0.52	0.32, 0.86	**0.010**	0.88	0.52, 1.47	0.6	0.91	0.54, 1.52	0.7	0.96	0.55, 1.69	0.9
CT - Honeycombing	3.72	1.37, 10.1	**0.010**	1.69	0.61, 4.65	0.3	1.68	0.61, 4.62	0.4	1.35	0.49, 3.74	0.6
CT - Emphysema	2.74	1.96, 3.81	**<0.001**	1.72	1.22, 2.42	**0.002**	1.76	1.24, 2.49	**0.002**	1.26	0.85, 1.86	0.2
CT - Fibrosis	2.48	1.76, 3.50	**<0.001**	1.42	0.99, 2.02	0.060	1.44	1.00, 2.07	0.056	1.23	0.84, 1.81	0.3
PH-CLD (n= 325)												
CT - Honeycombing	2.11	1.04, 4.30	**0.039**	1.80	0.88, 3.67	0.14	2.02	0.99, 4.14	0.081	1.06	0.39, 2.91	>0.9

CT - Fibrosis	1.83	1.42, 2.35	**<0.001**	1.73	1.34, 2.24	**<0.001**	1.63	1.26, 2.10	**<0.001**	1.46	1.09, 1.96	**0.011**

Table 4.3: Multivariate analysis

Abbreviations: HR, Hazard Ratio; CI, Confidence Interval; CT, computed tomography; IPA, idiopathic pulmonary arterial hypertension; PH-CLD, pulmonary hypertension associated with chronic lung disease.

4.9.4 Table 4: Baseline characteristics of patients with initial diagnosis of IPAH

Characteristic	IPAH-noLD, N = 151	IPAH-CGG, N = 46	IPAH-LD, N = 130	p-value
Age at diagnosis	56 (18) ‡	48 (19) ‡	70 (9) *†	<0.001
Male Gender	48 (32%)‡	11 (24%) ‡	69 (53%)*†	<0.001
WHO Functional Class	‡	‡	*†	0.011
II	25 (17%)	9 (20%)	9 (6.9%)	
III	98 (65%)	29 (64%)	81 (62%)	
IV	27 (18%)	7 (16%)	40 (31%)	
CT – Centrilobular Ground Glass	0 (0%) †	46 (100%) *‡	0 (0%) †	<0.001
CT – Ground Glass Opacification	9 (6.0%) †	39 (85%) *‡	8 (6.2%) †	<0.001
CT – Honeycombing	0 (0%) ‡	0 (0%) *‡	5 (3.8%) *	0.021
CT – Consolidation	5 (3.3%)	2 (4.3%)	2 (1.5%)	0.4
CT – Fibrosis	0 (0%) ‡	0 (0%) ‡	68 (52%) *†	<0.001
CT – Fibrosis (by severity)	‡	‡	*†	<0.001
None	151 (100%)	46 (100%)	62 (48%)	
Mild	0 (0%)	0 (0%)	52 (40%)	
Moderate	0 (0%)	0 (0%)	9 (6.9%)	
Unknown	0 (0%)	0 (0%)	7 (5.4%)	
CT – Emphysema	0 (0%) ‡	0 (0%) ‡	94 (72%) *†	<0.001
CT – Emphysema (by severity)	‡	‡	*†	
None	151 (100%)	46 (100%)	36 (28%)	
Mild	0 (0%)	0 (0%)	44 (34%)	
Moderate	0 (0%)	0 (0%)	38 (29%)	
Severe	0 (0%)	0 (0%)	5 (3.8%)	
Unknown	0 (0%)	0 (0%)	7 (5.4%)	
CT – CPFE	0 (0%) ‡	0 (0%) ‡	32 (25%) *†	<0.001
mPAP (mmHg)	54 (12) †‡	62 (13) *‡	49 (8) *†	<0.001
mRAP (mmHg)	12 (6)	10 (5)	11 (5)	0.3
PAWP (mmHg)	10.91 (3.10)	9.72 (2.77)	11.03 (3.21)	0.055
Cardiac output (L/min)	4.64 (1.87) †‡	3.85 (0.98) *	4.08 (1.44) *	0.014
Cardiac index (L/min/m^2)	2.47 (0.95) ‡	2.13 (0.52)	2.23 (0.73) *	0.031
PVR (Wood Units)	10.7 (5.2) †	14.6 (6.2) *‡	10.3 (4.4) †	<0.001
SvO_2 %	61 (10)	62 (7)	59 (8)	0.030

FEV_1 % predicted	82 (17)	88 (15)	83 (20)	0.124
FVC % predicted	93 (20)	101 (18)	99 (21)	0.016
FEV_1/FVC	74 (10) ‡	74 (8) ‡	67 (10) *†	<0.001
DLco % predicted	52 (20) ‡	56 (17) ‡	31 (14) *†	<0.001

Table 4.4 Baseline characteristics of patients with initial diagnosis of IPAH per subgroups

Data are presented as number (percentage) or mean (standard deviation). Between-group comparisons performed using one-Way ANOVA with Bonferroni Post-Hoc Correction. Difference between groups noted: * significant difference to IPAH-noLD, † significant difference to IPAH-CGG, ‡ significant difference to IPAH-LD. 8 patients with both CGG and LD not included. Abbreviations: WHO, World Health Organisation; CT, computed tomography; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity; DLco, diffusing capacity of carbon monoxide; IPAH, idiopathic pulmonary arterial hypertension; CGG, centrilobular ground glass; LD, lung disease; CPFE, combined pulmonary fibrosis and emphysema

4.10 Figures

4.10.1 Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram showing selection of study cohort.

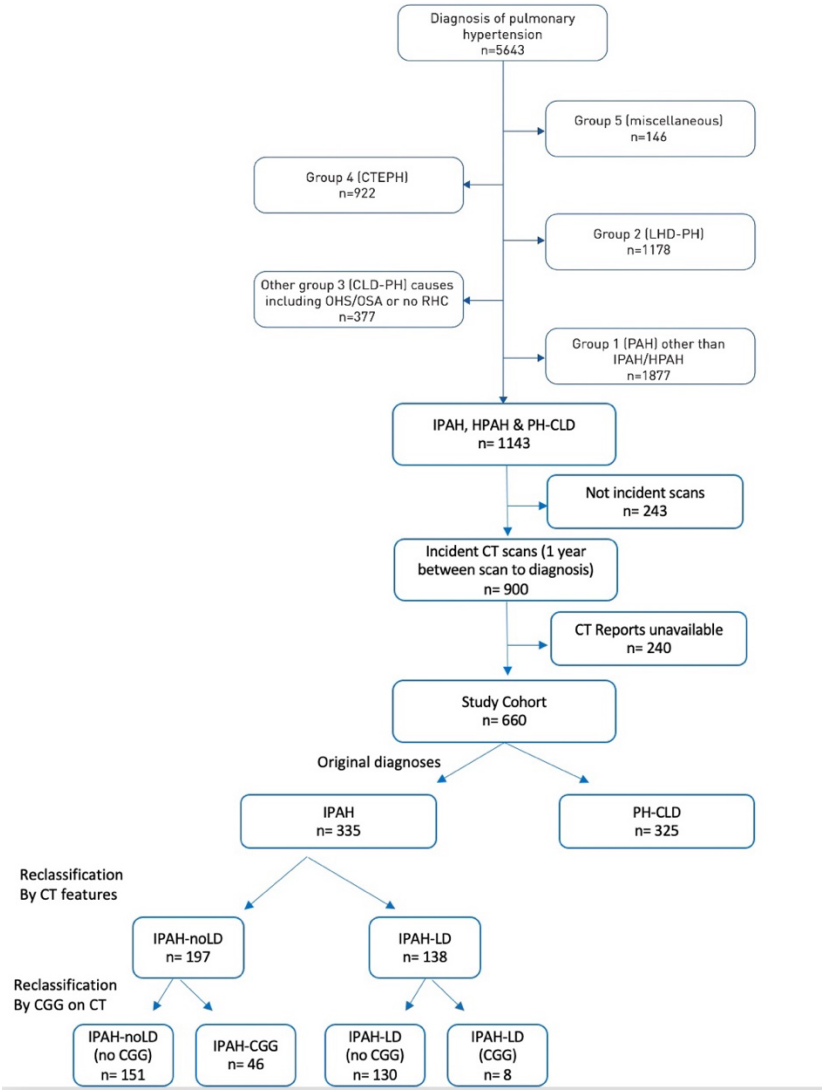


Figure 4.1: CONSORT (Consolidated Standards of Reporting Trials) flow diagram showing selection of study cohort.

Abbreviations. CTEPH , Chronic ThromboEmbolic Pulmonary Hypertension; LHD-PH, Pulmonary Hypertension with Left Heart Disease; PH-CLD, PH due to chronic lung disease and/or hypoxia; IPAH, Idiopathic Pulmonary Arterial Hypertension; HPAH, Hereditary Pulmonary Arterial Hypertension; CT, Computed Tomography; IPAH-noLD, Idiopathic Pulmonary Arterial Hypertension with no lung disease; IPAH-LD, Idiopathic Pulmonary Arterial Hypertension with Lung Disease

4.10.2 Figure 2: Kaplan-Meier survival curves stratified by CT features of CGG, emphysema and fibrosis for: a. All patients, b. Patients initially diagnosed with IPAH and c., Patients initially diagnosed with PH-CLD.

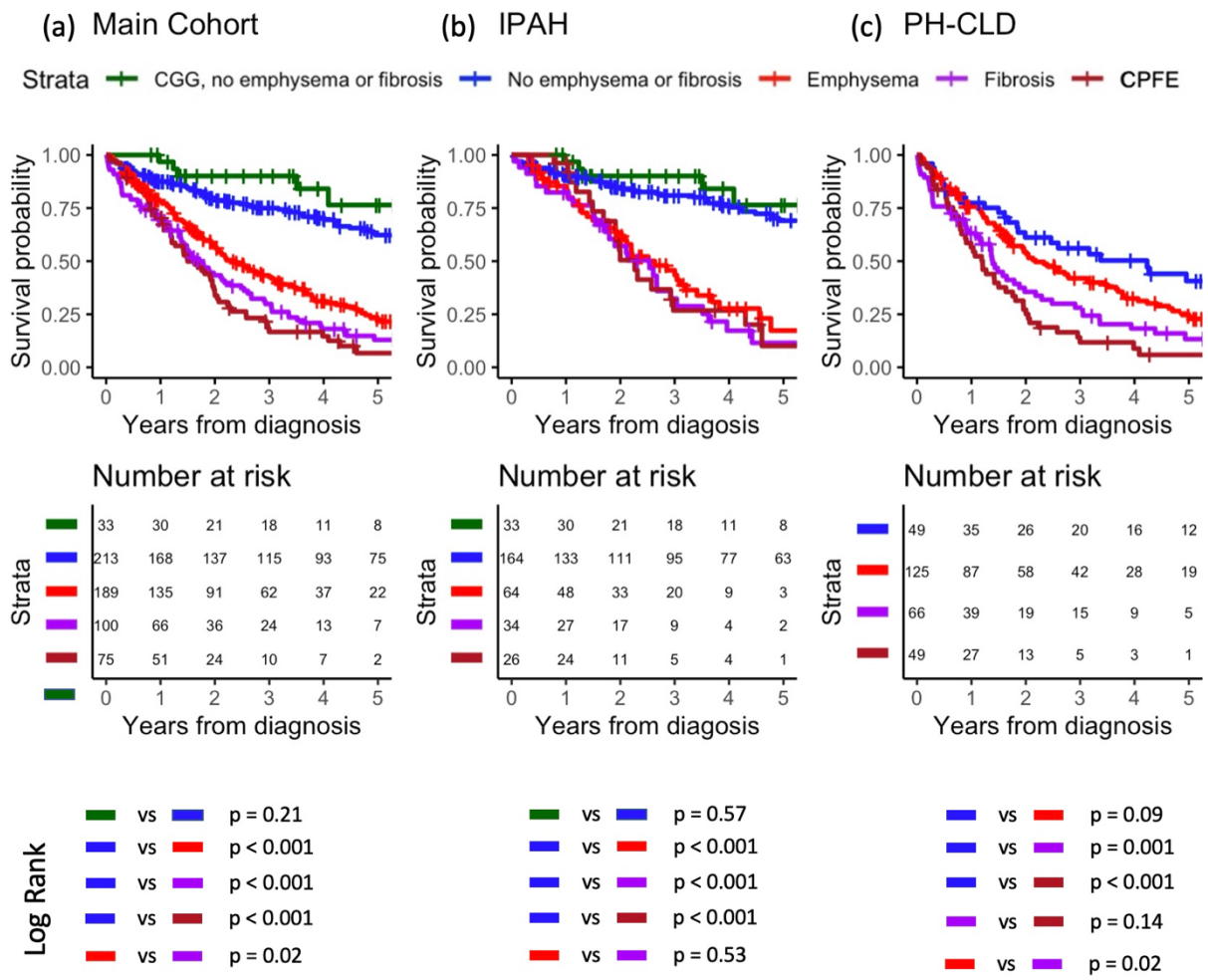


Figure 4.2: Kaplan-Meier survival curves stratified by CT features of CGG, emphysema and fibrosis for: a. All patients, b. Patients initially diagnosed with IPAH and c., Patients initially diagnosed with PH-CLD.

Abbreviations IPAH – Idiopathic Pulmonary Arterial Hypertension; PH-CLD – PH due to chronic lung disease and/or hypoxia.

4.10.3 Figure 3: Kaplan-Meier survival curves for patients classified as IPAH-LD, IPAH-noLD and PH-CLD.

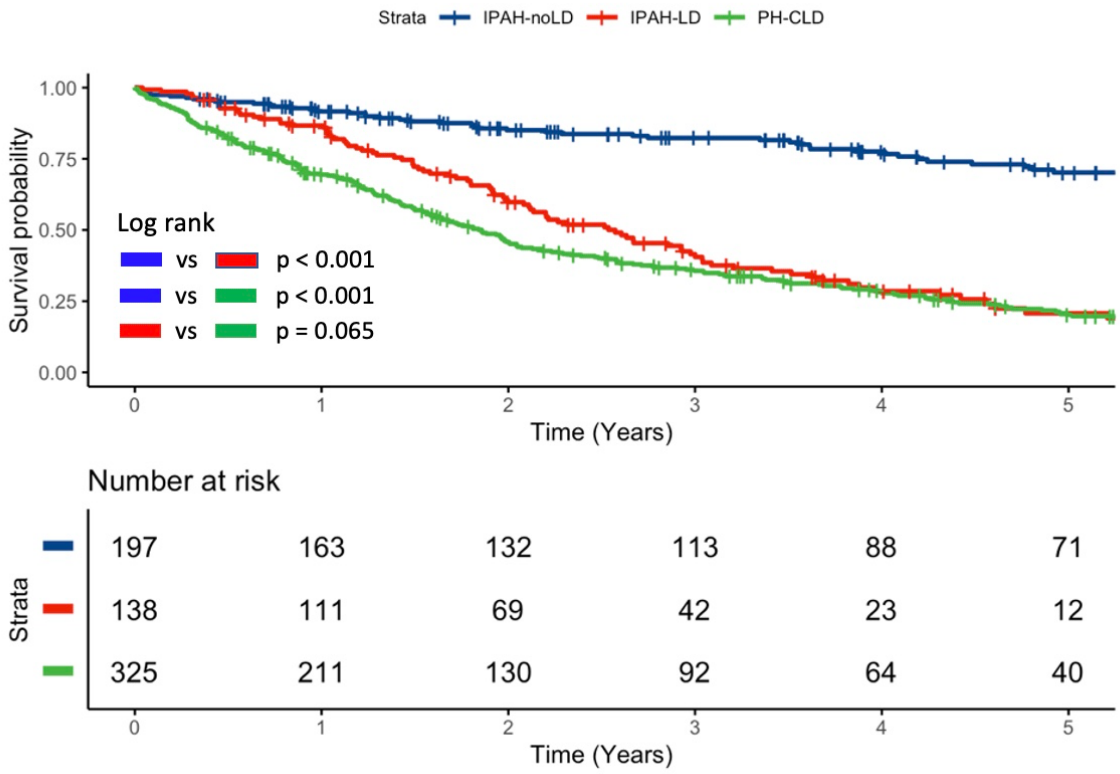


Figure 4.3: Kaplan-Meier survival curves for patients classified as IPAH-LD, IPAH-noLD and PH-CLD

Abbreviations. IPAH-noLD, idiopathic pulmonary arterial hypertension with no CT features of lung disease; IPAH-LD, idiopathic pulmonary arterial hypertension with CT features of lung disease; PH-CLD, pulmonary hypertension associated with chronic lung disease.

4.10.4 Figure 4: Kaplan-Meier curve comparing survival in IPAH-noLD (IPAH with no CT features of lung disease), IPAH-LD (IPAH with CT features of lung disease) and IPAH-CGG (IPAH with centrilobular ground glass on CT).

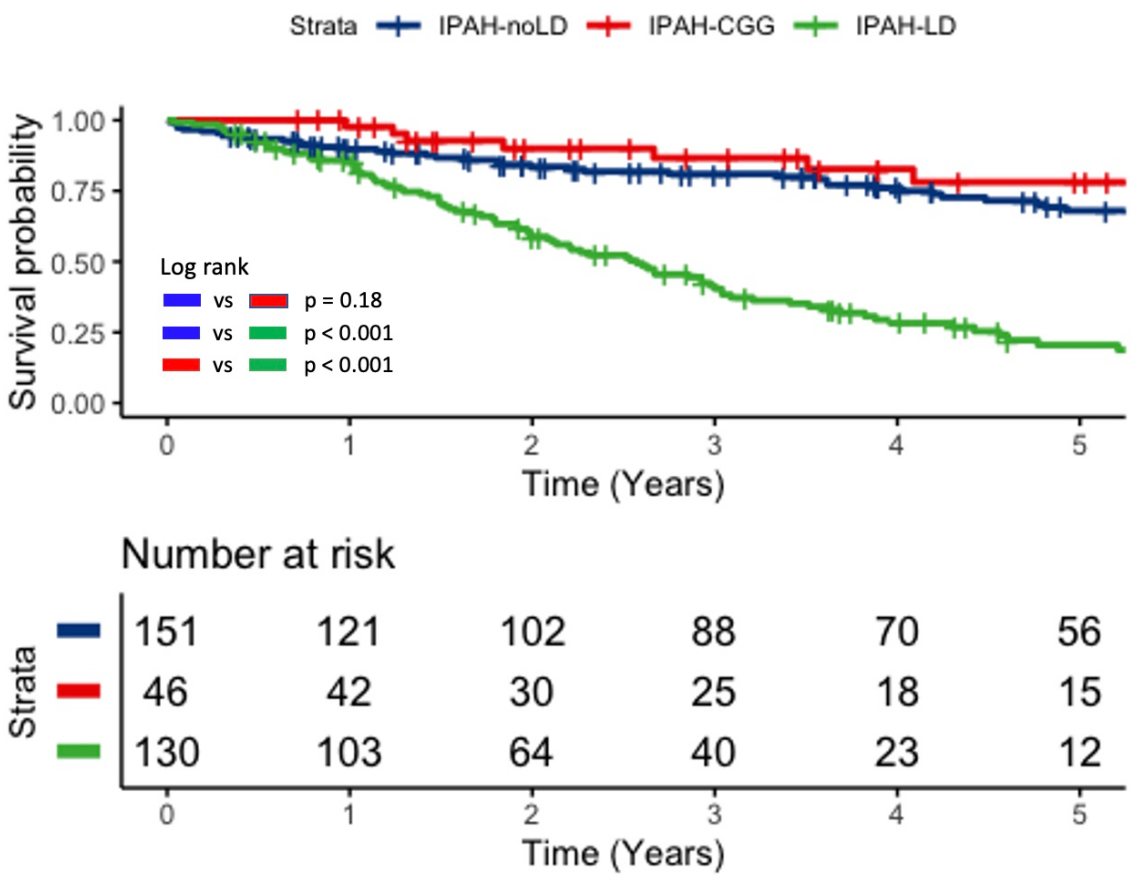


Figure 4.4: Kaplan-Meier curve comparing survival in IPAH-noLD (IPAH with no CT features of lung disease), IPAH-LD (IPAH with CT features of lung disease) and IPAH-CGG (IPAH with centrilobular ground glass on CT).

Abbreviations. IPAH-noLD, idiopathic pulmonary arterial hypertension with no CT features of lung disease; IPAH-LD, idiopathic pulmonary arterial hypertension with CT features of lung disease; IPAH-CGG, idiopathic pulmonary arterial hypertension with centrilobular ground glass changes.

4.11 Appendix

4.11.1 Univariate analysis of patients with IPAH-noLD

Characteristic	**HR**[1]	**95% CI**[1]	**p-value**
CT - Centrilobular Ground Glass (CGG)	0.62	0.30, 1.26	0.2
CT - Ground Glass Opacification (GGO)	0.64	0.32, 1.26	0.2
CT - Honeycombing			
CT - Consolidation	1.55	0.48, 4.96	0.5
Age at diagnosis	1.05	1.03, 1.07	<0.001
Male Gender	1.50	0.90, 2.52	0.12
WHO Functional Class III & IV (ref: I & II)	1.78	1.12, 2.81	0.014
WHO Functional Class			
II	—	—	
III	2.61	1.02, 6.67	0.046
IV	5.62	2.07, 15.2	<0.001
mPAP (mmHg)	0.99	0.97, 1.01	0.3
mRAP (mmHg)	1.04	1.00, 1.08	0.085
PAWP (mmHg)	1.10	1.00, 1.20	0.040

Cardiac output (L/min)	1.00	0.84, 1.18	>0.9
Cardiac index (L/min/m^2)	0.98	0.70, 1.38	>0.9
PVR (Wood Units)	0.97	0.92, 1.02	0.2
SvO_2 %	0.96	0.94, 0.99	0.011
FEV_1 % predicted	0.98	0.97, 0.99	0.008
FVC % predicted	0.99	0.98, 1.00	0.11
FEV_1/FVC Ratio	0.96	0.94, 0.98	<0.001
DLco % predicted	0.95	0.93, 0.97	<0.001

Table 4.5 Univariate analysis of patients with IPAH-noLD

Abbreviations: HR - Hazard Ratio, CI - Confidence Interval, IPAH-noLD, Idiopathic Pulmonary Arterial Hypertension with no lung disease, CT – Computed Tomography, WHO – World Health Organisation, CPFE – Combined Pulmonary Fibrosis and Emphysema, mPAP – mean pulmonary arterial pressure, mRAP – mean right atrial pressure, PAWP – pulmonary arterial wedge pressure, PVR – pulmonary vascular resistance, SvO2 – mixed venous oxygen saturation. FEV_1 – forced expiratory volume in 1 second, FVC – forced vital capacity, DLco – diffusing capacity of carbon monoxide.

4.11.2 Baseline characteristics of IPAH-noLD vs IPAH-LD vs PH-CLD

Characteristic	**IPAH-noLD,** N = 197	**IPAH-LD,** N = 138	**PH-CLD,** N = 325	**p-value**
Age at diagnosis	54 (18) †‡	70 (10) *	67 (11) *	<0.001
Male Gender	59 (30%)†‡	72 (52%)*	187 (58%)*	<0.001
WHO Functional Class	†‡	*	*	<0.001
2	34 (17%)†‡	10 (7.2%)*	34 (10%)	
3	127 (65%)	86 (62%)	185 (57%)	
4	34 (17%)	42 (30%)	105 (32%)	
CT - Centrilobular Ground Glass (CGG)	46 (23%)†‡	8 (5.8%)*	0 (0%)*	<0.001
CT - Ground Glass Opacification (GGO)	48 (24%)†‡	14 (10%)*	31 (9.5%)*	<0.001
CT - Honeycombing	0 (0%)	5 (3.6%)	10 (3.1%)	0.013
CT - Consolidation	7 (3.6%)	2 (1.4%)‡	22 (6.8%)†	0.031
CT - Fibrosis	0 (0%)†‡	72 (52%)*	141 (43%)*	<0.001
CT - Fibrosis (by severity)	†‡	*‡	*†	<0.001
Mild	0 (0%)	54 (39%)	28 (8.6%)	
Moderate	0 (0%)	9 (6.5%)	44 (14%)	
None	197 (100%)	66 (48%)	184 (57%)	
Severe	0 (0%)	0 (0%)	48 (15%)	
Unknown	0 (0%)	9 (6.5%)	21 (6.5%)	
CT - Emphysema	0 (0%)†‡	98 (71%)*	204 (63%)*	<0.001
CT - Emphysema (by severity)	†‡	*‡	*†	<0.001
Mild	0 (0%)	48 (35%)	29 (8.9%)	
Moderate	0 (0%)	38 (28%)	91 (28%)	
None	197 (100%)	40 (29%)	121 (37%)	
Severe	0 (0%)	5 (3.6%)	64 (20%)	
Unknown	0 (0%)	7 (5.1%)	20 (6.2%)	
CT - CPFE	0 (0%)†‡	32 (23%)*	69 (21%)*	<0.001
mPAP (mmHg)	56 (13) †‡	49 (9)* ‡	42 (10)* †	<0.001
mRAP (mmHg)	11.4 (6.1) ‡	11.3 (5.3) ‡	9.0 (5.1)* †	<0.001
PAWP (mmHg)	10.6 (3.1) ‡	11.1 (3.6)	11.8 (4.2)*	0.004
Cardiac output (L/min)	4.46 (1.74) ‡	4.10 (1.41)	5.00 (1.64)	<0.001
Cardiac index (L/min x m^-2)	2.39 (0.87) ‡	2.23 (0.71) ‡	2.73 (0.87)* †	<0.001

PVR (Wood Units)	11.6 (5.7) †‡	10.2 (4.4)* ‡	7.0 (4.1)* †	<0.001
SvO2 %	62 (10) ‡	59 (8) ‡	65 (8)* †	<0.001
PFT - FEV1 % predicted	83 (17)* ‡	83 (20) ‡	60 (25)* †	<0.001
PFT - FVC % predicted	95 (20) ‡	98 (21) ‡	80 (28)* †	<0.001
PFT - FEV1 / FVC Ratio	74 (10) †‡	67 (10)* ‡	61 (18)* †	<0.001
PFT - DLCO % predicted	53 (19) †‡	32 (15)* ‡	28 (14)* †	<0.001

Table 4.6 Baseline characteristics of IPAH-noLD vs IPAH-LD vs PH-CLD

Data are presented as number (percentage) or mean (standard deviation). Between-group comparisons performed using one-Way ANOVA with Bonferroni Post-Hoc Correction. Difference between groups noted: * significant difference to IPAH-noLD, † significant difference to IPAH-LD, ‡ significant difference to PH-CLD. Abbreviations: IPAH-LD, Idiopathic Pulmonary Arterial Hypertension with lung disease, PH-CLD, PH due to chronic lung disease and/or hypoxia, CT – Computed Tomography, WHO – World Health Organisation, CPFE – Combined Pulmonary Fibrosis and Emphysema, mPAP – mean pulmonary arterial pressure, mRAP – mean right atrial pressure, PAWP – pulmonary arterial wedge pressure, PVR – pulmonary vascular resistance, SvO_2 – mixed venous oxygen saturation. FEV_1 – forced expiratory volume in 1 second, FVC – forced vital capacity, DLco – diffusing capacity of carbon monoxide.

4.11.3 Survival of patients with IPAH-noLD, IPAH-LD and PH-CLD

Characteristic	1 Year	2 Year	3 Year	4 Year	5 Year
IPAH-noLD	92%	85%	82%	78%	70%
IPAH-LD	87%	60%	41%	29%	21%
PH-CLD	69%	46%	36%	28%	20%

Table 4.7 Survival of patients with IPAH-noLD, IPAH-LD and PH-CLD

Abbreviations: IPAH-noLD, Idiopathic Pulmonary Arterial Hypertension with no lung disease; IPAH-LD, Idiopathic Pulmonary Arterial Hypertension with lung disease; PH-CLD, PH due to chronic lung disease and/or hypoxia.

5 Idiopathic pulmonary arterial hypertension with a lung phenotype

Marius M Hoeper, **Krit Dwivedi**, Christine Pausch, Robert A Lewis, Karen M Olsson, Doerte Huscher, David Pittrow, Ekkehard Grünig, Gerd Staehler, Carmine Dario Vizza, Henning Gall, Oliver Distler, Christian Opitz, J Simon R Gibbs, Marion Delcroix, Da-Hee Park, H Ardeschir Ghofrani, Ralf Ewert, Harald Kaemmerer, Hans-Joachim Kabitz, Dirk Skowasch, Juergen Behr, Katrin Milger, Tobias J. Lange, Heinrike Wilkens, Hans-Jürgen Seyfarth, Matthias Held, Daniel Dumitrescu, Iraklis Tsangaris, Anton Vonk-Noordegraaf, Silvia Ulrich, Hans Klose, Martin Claussen, Stephan Eisenmann, Kai-Helge Schmidt, Andrew J. Swift, A A Roger Thompson, Charlie A Elliot, Stephan Rosenkranz, Robin Condliffe, David G Kiely*, Michael Halank*

*These authors contributed equally

*** Peer-reviewed journal article published in *The Lancet Respiratory Medicine (2022)***

Contribution: Second author. Contributor Roles Taxonomy (CRediT) – data curation, formal analyses, methodology, project administration, resources, software, validation, visualisation, writing – original draft, review, and editing).

This is an important multi-centre multi-registry paper which I contributed significantly to, and my contribution to it makes for an important part of both the paper and this thesis. The first author is the PH lead professor in Hannover, Germany who conceptualised and lead the multi-centre effort. I am performed all data analysis on the ASPIRE registry, liaised with the team at COMPERA (European registry) and provided the novel CT imaging data, which is only available due to the work done directly for this thesis. I wrote and edited the ASPIRE sections with DGK.

5.1 Abstract

Background: Among patients diagnosed with idiopathic pulmonary arterial hypertension (IPAH), there is an emerging lung phenotype characterised by a low diffusion capacity for carbon monoxide (DLCO) and a smoking history. These patients need to be further characterised.

Methods: We analysed data from two European pulmonary hypertension (PH) registries, COMPERA and ASPIRE, to identify patients diagnosed with IPAH who presented with a lung phenotype defined by a DLCO <45% predicted and a smoking history. We compared these patients with a cohort of patients with classical IPAH, defined by the absence of cardiopulmonary co-morbidities and DLCO ≥45% predicted and with a cohort of patients with PH due to lung disease (group 3 PH).

Findings: The COMPERA analysis included 128 patients with classical IPAH, 268 patients with IPAH and a lung phenotype, and 910 patients with PH due to lung disease. The corresponding numbers from ASPIRE were 185, 139, and 375, respectively. In both cohorts, patients with IPAH and a lung phenotype and patients with group 3 PH showed comparable age and sex distribution, exercise limitation, response to therapy, and survival. In contrast, patients with IPAH and a lung phenotype differed substantially in all these categories from patients with classical IPAH.

Interpretation: A cohort of patients diagnosed with IPAH suffers from a distinct, presumably smoking-related form of pulmonary hypertension accompanied by a low DLCO. Phenotypically, these patients resemble those with PH due to lung disease rather than classical IPAH. These observations have pathogenetic, diagnostic, and therapeutic implications, which must be further explored.

5.2 Introduction

The current clinical classification of pulmonary hypertension (PH) consists of 5 major groups: Group 1, pulmonary arterial hypertension (PAH); group 2, PH associated with left heart disease; group 3, PH associated with lung disease; group 4, chronic thromboembolic PH; and group 5, PH due to systemic or multifactorial conditions[2,6]. The criteria for the diagnosis and classification of PH have been outlined in recent guidelines,[6] but in some patients, the individual classification is not always straightforward. This problem is frequently encountered in patients with idiopathic PAH (IPAH), the most common form of PAH. Originally, IPAH, formerly called primary pulmonary hypertension, was described as a disease occurring mostly in younger, otherwise healthy individuals, predominantly women.[174] Such patients represent the classical phenotype of IPAH. However, registries from Europe and the US have demonstrated that IPAH is now more frequently diagnosed in elderly patients, many of whom have cardiac and/or pulmonary comorbidities.[15,175,176] In such patients, it is not always easy to distinguish IPAH from group 2 or group 3 PH. Several disease phenotypes have been reported, including a subtype of patients diagnosed with IPAH who present with a lung phenotype, mainly characterized by a history of smoking and a low lung diffusion capacity for carbon monoxide (DLCO), but otherwise no or only subtle signs of parenchymal lung disease. In accordance with current guidelines, these patients are classified as IPAH rather than group 3 PH.[41,168,177]

In a recent cluster analysis from the *Comparative, Prospective Registry of Newly Initiated Therapies for Pulmonary Hypertension* (COMPERA), a European PH registry, only 12.6% of 846 patients diagnosed with IPAH presented with the classical phenotype while 35.8% had a left heart phenotype and 51.6% a lung phenotype, respectively.[166] The high proportion of patients with a lung phenotype came as surprise. To further characterize these patients, we used the COMPERA database to identify those with IPAH and a lung phenotype and to compare them with patients with classical IPAH and those classified as PH associated with lung disease, i.e., group 3 PH, focussing on demographics, disease characteristics at diagnosis, response to PH therapy, and survival. Data obtained from the *Assessing the Spectrum of Pulmonary hypertension Identified at a REferral centre* (ASPIRE) registry were utilized for independent validation.[50]

5.3 Methods

5.3.1 Databases

Details of COMPERA (www.COMPERA.org; registered at Clinicaltrials.gov under the identifier NCT01347216) have been reported previously.[166,176] COMPERA is an ongoing PH registry launched in 2007 that prospectively collects baseline, follow-up, and outcome data of newly diagnosed patients who receive targeted therapies for any form of PH. PH centres from several European countries participate (Austria, Belgium, Germany, Greece, Hungary, Italy, Latvia, Lithuania, Netherlands, Slovakia, Switzerland, United Kingdom), with about 80% of the enrolled patients coming from Germany. COMPERA has been approved by the responsible ethics committee, and all patients provided written, informed consent prior to inclusion.

Details of the ASPIRE registry have been previously reported.[41,50] The ASPIRE Registry includes data on patients undergoing investigation for suspected PH at the Sheffield Pulmonary Vascular Disease Unit, a PH centre with a referral population of 15-20 million, based in Sheffield UK, from 2001 onwards. During their assessment, patients undergo systematic evaluation including multimodality imaging and right heart catheterisation, in accordance to annually audited national standards of care. Ethical approval was granted by the Institutional Review Board and approved by the National Research Ethics Service (16/YH/0352). Analyses were conducted in accordance with General Data Protection Regulation.

5.3.2 Patient selection

All analyses from COMPERA and ASPIRE were performed separately and the data were not combined. From COMPERA, patients were selected to form three cohorts: (i) patients with classical IPAH (PH group 1.1), defined by the absence of risk factors for left heart disease (body mass index (BMI) $\geq$ 30 kg/m^2, hypertension, diabetes mellitus, and coronary heart disease), and a DLCO $\geq$ 45%; (ii) patients diagnosed with IPAH and a lung phenotype, defined by a smoking history and a DLCO < 45% of the predicted value; and (iii) patients classified by their physicians as group 3 PH with the underlying conditions being either COPD (PH group 3.1) or ILD (PH group 3.2). The same selection criteria were used for

ASPIRE, except for risk factors for left heart disease not being considered as these data were not available.

The DLCO cut-off value of <45% versus ≥45% was derived from previous studies that have determined the prognostic value of this threshold.[41,53,166,168]

For all cohorts, further inclusion criteria were age ≥18 years, PH diagnosis made between Jan 2009 and Dec 2020 in COMPERA, and between Feb 2001 and Jan 2019 in ASPIRE, and data from right heart catheterization available at baseline showing mean pulmonary arterial pressure (mPAP) ≥ 25 mmHg, pulmonary artery wedge pressure (PAWP) ≤ 15 mmHg, and pulmonary vascular resistance (PVR) > 3 WU. Furthermore, only incident patients with at least one follow-up documentation were considered for COMPERA and incident patients for ASPIRE.

5.3.3 Imaging

Chest computed tomography (CT) data were available only from ASPIRE. CT scans were evaluated by experienced radiologists for the presence of fibrotic or emphysematous changes, which were graded as absent, mild, moderate, or severe as previously described.[41,178]

5.3.4 Statistical analyses

This was a post-hoc analysis of prospectively collected data. Analysis was performed using R software major version 4. Categorical data are presented as number and percentage, continuous data as median and first and third quartile [Q1, Q3]. First follow-up was defined as the first assessment within 3 to 12 months after treatment initiation. Vital status was ascertained by on-site visits or phone calls to the patients or their caregivers. Patients who underwent lung transplantation and patients who were lost to follow-up were censored at the date of the last contact.

The focus of the present study was the identification of similarities and differences between patients diagnosed with IPAH who present with a lung phenotype and group 3 PH. To compare the cohort of patients with IPAH and a lung phenotype with each of the two other cohorts, two-sample Welch t-tests were computed for continuous data. For non-normally distributed data, the Wilcoxon rank sum test was used. Categorical data were compared by

Pearson's Chi-squared test or by Fisher's exact test. Response to therapy was determined by changes from baseline to first follow-up in WHO functional class (FC), 6-minute walking distance (6MWD), N-terminal fragment of pro-brain natriuretic peptide (NT-proBNP), and mortality risk using the ESC/ERS 4-strata model.[16] These data were available only from COMPERA. Survival estimates from the time of enrolment were done by Kaplan-Meier analyses, log-rank test, and Cox proportional hazard regression models to adjust for age and sex.

5.3.5 Role of the funding source

COMPERA is funded by unrestricted grants from Acceleron, Bayer, GSK, Janssen and OMT. These companies were not involved in data analysis or the writing of this manuscript. The ASPIRE Registry is supported by Sheffield Teaching Hospitals NHS Foundation Trust.

5.4 Results

5.4.1 Patient characteristics of the study cohorts

In COMPERA, a total of 128 patients with classical IPAH, 268 patients with IPAH and a lung phenotype, and 910 patients with group 3 PH fulfilled the eligibility criteria and were included in the present analysis. The corresponding numbers from ASPIRE were 185, 139, and 375. Patient selection is shown in Figures 1a and 1b. The patient characteristics at baseline are shown in Tables 1a and 1b. The number of missing values for each variable is shown in the supplementary tables S1a and S1b. Histograms showing the age distribution of the cohorts are depicted in Figures 2a and 2b. The baseline characteristics of patients with IPAH who were excluded from the analyses are shown in supplementary tables S2a and S2b.

Patients with classical IPAH were mostly young with a median age of 45 and 52 years, respectively (although some patients were in the seventies and eighties as shown in Figures 2a and 2b) and predominantly female. About one third of these patients had a smoking history with a median of 14 and 20 pack years. Lung function was preserved while the DLCO was mildly reduced, and blood gas analyses (data available from COMPERA only) showed a

near-normal PaO_2 and a low $PaCO_2$. Haemodynamic assessment at time of diagnosis showed severe pre-capillary PH and most had a moderately impaired exercise capacity.

Compared to patients with classical IPAH, patients with IPAH and a lung phenotype were older (mean age of about 70 years) and more often male. Per inclusion criteria, all patients were smokers, and the median tobacco exposure was 40 (COMPERA) and 30 (ASPIRE) pack years. Forced vital capacity (FVC) and forced expiratory volume in 1s (FEV_1) were mostly normal. However, the DLCO was severely reduced (30% and 27% of the predicted value, respectively), and the patients were more hypoxaemic than patients with a classical phenotype. Severity of PH as determined by mPAP and PVR was comparable to patients with the classical phenotype, but exercise capacity was substantially lower.

Patients with group 3 PH had a similar age to patients with IPAH and a lung phenotype and had nearly the same age distribution as well as a comparable male-to-female ratio (Figures 2a and 2b). Eighty-one percent had a smoking history with a median of 40 pack years (data available for COMPERA only). FVC and FEV_1 were lower than in patients with IPAH and a lung phenotype, but most patients did not have severely impaired pulmonary function, except for a very low DLCO (26% and 25%, respectively, of the predicted value). Blood gas analyses showed marked hypoxaemia, comparable to patients with IPAH and a lung phenotype. mPAP and PVR were lower than in the other cohorts but still much elevated. The degree of exercise limitation was similar to patients with IPAH and a lung phenotype.

5.4.2 Imaging (ASPIRE data only)

The chest CT studies from ASPIRE showed absence of parenchymal lung disease in most patients with classical IPAH. The majority of patients with IPAH and a lung phenotype had normal CT findings. Although more parenchymal abnormalities were present in this group, most were mild. In contrast, almost all patients with group 3 PH had parenchymal abnormalities, mostly moderate or severe. Details are shown in Table 2.

5.4.3 Changes from baseline to first follow-up (COMPERA data only)

The first follow-up visit took place 4·7 [3·5, 6·6] months after baseline. FC, 6MWD, NT-proBNP and risk at baseline and first follow-up are shown in Figures 3a-d. In all categories,

patients with classical IPAH improved most, whereas there were less and quantitatively similar changes in the two other cohorts.

5.4.4 Survival

In COMPERA, the median observation time was 3·9 [1·8, 6·6] years for patients with classic IPAH, 2·0 [1·2, 3·4] years for patients with IPAH and a lung phenotype, and 1·7 [0·7, 3·3] years for patients with group 3 PH. In the cohort of patients with classic IPAH, 23 (18%) patients died, 5 (4%) underwent lung transplantation, and 8 (6%) were lost to follow-up. The corresponding numbers for patients with IPAH and a lung phenotype were 138 (52%), 5 (2%) and 13 (5%), respectively. Among the patients with group 3 PH, 583 (64%) died, 22 (2%) underwent lung transplantation and 46 (5%) were lost to follow-up.

In ASPIRE, the median observation time was 4·5 [2·1, 7·8] years for patients with classic IPAH, 1·7 [0·9, 2·8] years for patients with IPAH and a lung phenotype, and 1·4 [0·6, 3·1] years for patients with group 3 PH. No patients were lost to follow-up. In the cohort of patients with classic IPAH, 42 (23%) patients died and 7 (4%) underwent lung transplantation. The corresponding numbers for patients with IPAH and a lung phenotype were 90 (65%) and 0 respectively. Among the patients with group 3 PH, 286 (76%) died and 5 (1%) underwent lung transplantation.[H]

In both registries, the survival rates of patients with idiopathic PAH with a lung phenotype and of patients with group 3 PH were comparable and both much inferior to the survival rate of patients with classical IPAH (Figures 4a and b).

In COMPERA, the Kaplan-Meier estimated survival rates of patients with classical IPAH at 1, 3 and 5 years were 95%, 90%, and 84%, respectively. In patients with IPAH and a lung phenotype, the corresponding numbers were 89%, 49%, and 31%. In patients with group 3 PH, the respective survival rates were 78%, 43%, and 26%. The unadjusted survival rates differed significantly between patients with classical IPAH and IPAH with a lung phenotype

[H] A limitation of the high mortality rate in Group 3 is that few patients reach the 5-year figures follow-up reported. Furthermore, all patients, regardless of their date of diagnosis were included in the analysis (we did not exclude those with <5 years follow-up). Therefore, there are patients who will only have been observed for less than 5 years, which explains the drop off in total numbers at the yearly time points.

(p=<0·0001) and between the latter group and patients with group 3 PH (p=0·0159; Figure 4a). When adjusted for age and sex, the risk of death remained lower for patients with classical IPAH than for patients with IPAH and a lung phenotype (HR 3·48; 95% confidence interval 2·04 to 5·95, p<0·0001). The survival difference between patients with IPAH and a lung phenotype and patients with group 3 PH was smaller albeit still statistically significant (HR 0·79; 95% confidence interval 0·66 to 0·96, p=0·0150).

In ASPIRE, the Kaplan-Meier estimated survival rates of patients with classical IPAH at 1, 3 and 5 years were 98%, 91%, and 80%, respectively. In patients with IPAH and a lung phenotype, the corresponding numbers were 79%, 35%, and 21%. In patients with group 3 PH, the respective survival rates were 64%, 32%, and 18%. The unadjusted survival rates differed significantly between patients with classical IPAH and IPAH with a lung phenotype (p<0·0001) and between the latter group and patients with group 3 PH (p=0·0450; Figure 4b). When adjusted for age and sex, the risk of death remained much higher for patients with IPAH and a lung phenotype than for patients with classical IPAH (HR 3·61, 95% confidence interval 2·35 to 5·54). The survival difference between patients with IPAH and a lung phenotype and patients with group 3 PH was smaller albeit still statistically significant (HR 0·74; 95% confidence interval 0·58 to 0·94, p=0·010).

5.5 Discussion

The key finding of this analysis was that patients diagnosed with IPAH and a lung phenotype defined by a smoking history and a low DLCO had little in common with classical IPAH patients, with the exception of severe pre-capillary PH, having similar baseline characteristics, treatment response and survival as patients with group 3 PH. These observations challenge the current classification of PH.

In the present cohorts, patients categorized as classical IPAH resembled those initially described as primary pulmonary hypertension, i.e., predominantly young, otherwise healthy females[174]. These patients had an 80% survival rate 5 years after diagnosis, which is about twice as high as in historical controls,[179] presumably owing to therapeutic advances. However, the classical form has become the least common phenotype of IPAH, at least in most European countries, where IPAH is now being diagnosed predominantly in elderly patients with co-morbidities[15,166]. These patients continue to have a high mortality risk.[166] In these patients, the diagnostic classification can be challenging. This problem is illustrated by our cohorts of patients diagnosed with IPAH who presented with a lung phenotype. Most of these patients had normal or near-normal static and dynamic lung function parameters, and, where available, the majority had normal or near-normal chest CT findings, but severe pre-capillary PH. Hence, the diagnosis of IPAH was in accordance with current guidelines.[6,26]

When we compared patients with IPAH and a lung phenotype with patients classified as group 3 PH (PH associated with either COPD or ILD, 81% of whom were smokers as well), we found striking similarities. Age distribution and male-to-female ratio were comparable as were FC and 6MWD. The same was true for the prevalence of risk factors for left heart disease, which may have contributed to the development of PH. Patients with IPAH and a lung phenotype and patients classified as group 3 PH had a similar response to medical therapy, i.e., comparable changes from baseline to first follow-up in FC, 6MWD, NT-proBNP and mortality risk. Taken together, patients with IPAH and a lung phenotype resembled those of patients with group 3 PH, while they had little in common with classical IPAH, except for the presence of severe pre-capillary PH.

As in previous studies,[41,53,168] a DLCO ≥45% or <45% of the predicted value discriminated between patients with classical IPAH and patients with IPAH and a lung phenotype. It is unknown whether the low DLCO in the latter group of patients is caused by parenchymal

abnormalities or by a distinct pulmonary vasculopathy involving the loss of small pulmonary vessels, for which the term *vanishing pulmonary capillary syndrome* has been proposed.[180] In animal models, prolonged exposure to tobacco smoke causes endothelial cell apoptosis in pulmonary capillaries, which precedes the development of emphysema,[181] and most of the patients diagnosed with IPAH and a low DLCO are elderly individuals with a history of heavy smoking (which may also explain the male predominance of this phenotype). We therefore speculate that in these patients, smoking may have been a contributor to the development of PH, or even its main cause. In addition, it is possible that the pulmonary vasculopathy of patients with IPAH and a lung phenotype and patients with group 3 PH is similar, yet distinct from classical IPAH.

Our findings have implications not only for the diagnostic classification but also for therapeutic considerations. We have insufficient data on the safety and efficacy of PAH drugs in patients diagnosed with IPAH who present with a lung phenotype. None of the pivotal trials of globally approved PAH drugs reported the DLCO of their participants.[182–190] This lack of data is particularly worrisome when considering a recent study showing that PAH drugs may further impair gas exchange in patients with a low DLCO.[191] Moreover, the response to therapy in patients with IPAH and a lung phenotype was blunted compared to patients with classical IPAH, but it is unclear if this was due to a distinct pulmonary vasculopathy, less aggressive therapy, or co-morbidities leaving little room for functional improvement.

It is important to note that IPAH with a low DLCO may also be found in patients who have never smoked. Such patients may suffer from various conditions such as unrecognized pulmonary veno-occlusive disease or connective tissue disease. A similar disease phenotype has been reported in patients who have been exposed to organic solvents,[192] and in certain forms of heritable PAH.[193]

Limitations of the present study include its post-hoc nature, missing values, lack of imaging data in COMPERA, and heterogeneities between the two registries. Even though all patients were evaluated at referral centres, we cannot fully exclude the possibility that misclassification bias may have interfered with our analysis, especially as a small proportion of patients diagnosed as IPAH had more than mild lung function test or CT abnormalities. Furthermore, for the present analysis, patients with IPAH were highly selected to ensure a proper phenotypic characterization, and the results may not be generalizable to patients with mixed phenotypes.

In conclusion, patients diagnosed with IPAH who present with a lung phenotype have much more features of group 3 PH rather than classical IPAH. These observations challenge the current diagnostic classification of PH, and we propose to add a phenotypic component to the classification of unexplained pre-capillary PH. In addition, further data is needed on the safety and efficacy of PAH drugs in these patients, and future clinical trials on PAH should collect and report data on smoking status and DLCO of their participants. Finally, our observations support the hypothesis that there is a distinct smoking-related pulmonary vasculopathy, which needs to be further investigated.

5.6 Research in context

5.6.1 Evidence before the Subject

Idiopathic pulmonary arterial hypertension (IPAH), originally observed mainly in young, otherwise healthy individuals, is increasingly diagnosed in elderly patients with co-morbidities. Among these patients, a distinct lung phenotype is emerging, characterized by a history of smoking and a low diffusion capacity for carbon monoxide (DLCO, <45% of the predicted value) without overt signs of parenchymal lung disease. This disease phenotype is not well characterized. When we searched PubMed on Oct 19, 2021, and on Dec 17, 2021, using the search terms “pulmonary arterial hypertension” AND “smoking” AND “diffusion capacity”, we found only three case series describing patients with this phenotype.

5.6.2 Added value of this study

This study demonstrates that patients diagnosed with IPAH who present with a lung phenotype share many features with patients suffering from pulmonary hypertension (PH) associated with lung disease including sex and age distribution, functional impairment at diagnosis, response to PH medications, and survival. At the same time, these patients have very little in common with patients who present with a classical IPAH phenotype, i.e., patients without cardiopulmonary co-morbidities and a DLCO ≥45% of the predicted value.

5.6.3 Implications of the available evidence

We expect our findings to lead to a re-classification of some forms of pulmonary hypertension. A better characterization of patients with IPAH and a lung phenotype will also allow an evaluation of the safety and efficacy of PAH medications in this cohort. Finally, our data support the hypothesis that there is a distinct pulmonary vasculopathy, seemingly related to extensive tobacco exposure, which adds another component to the spectrum of smoking-related lung injury.

5.7 Tables

5.7.1 Patient characteristics at baseline in COMPERA

	Classical IPAH (i) n=128	P-value (i) vs. (ii)	IPAH with a lung phenotype (ii) n=268	P-value (ii) vs. (iii)	Group 3.1 or 3.2 PH n=910 (iii)
Age, years	45 [32, 60]	<0·0001	72 [65, 78]	0·886	71 [65, 77]
Female	99 (77%)	<0·0001	95 (35%)	0·712	336 (37%)
BMI, kg/m^2	24 [22, 27]	<0·0001	27 [24, 32]	0·0002	26 [23, 29]
WHO FC I II III IV	 2 (2%) 30 (24%) 85 (67%) 10 (8%)	<0·0001	 0 (0%) 16 (6%) 184 (73%) 51 (20%)	0·055	 0 (0%) 32 (4%) 612 (71%) 223 (26%)
6MWD, m	410 [320, 476]	<0·0001	234 [167, 310]	0·926	238 [159, 318]
NT-proBNP, ng/L BNP, ng/L	1,027 [360, 2,058] 127 [73, 249]	0·0002 0·1117	1,871 [583, 4,348] 304 [120, 441]	0·042 0·004	1,423 [462, 3,380] 120 [59, 276]
Pulmonary function		0·0011		<0·0001	
TLC, % pred	98 [87, 110]	<0·0001	93 [79, 103]	<0·0001	85 [67, 100]
FVC, % pred	92 [78, 103]	<0·0001	80 [66, 94]	<0·0001	68 [53, 84]
FEV_1, % pred	85 [74, 96]	<0.0001	71 [60, 85]	0.0003	59 [44, 74]
FEV_1/FVC (%)	80 [76, 85]	<0·0001	71 [63, 79]	0·769	68 [52, 81]
DLCO, % pred	69 [59, 76]	<0·0001	30 [24, 36]	0·791	26 [20, 35]

PaO_2, mmHg	78 [71, 84]	<0·0001	56 [50, 63]	<0·0001	57 [49, 64]
$PaCO_2$, mmHg	33 [30, 35]		35 [31, 39]		37 [33, 43]
Smoking history					
Ever	40 (34%)	<0·0001	268 (100%)	<0·0001	212 (81%)
Never	76 (66%)		0 (0%)		50 (19%)
Pack years	14 [10, 30]	<0·0001	40 [21, 50]	0·167	40 (30, 60]
Comorbid conditions					
BMI >30 kg/m^2	0 (0%)	<0·0001	86 (32%)	0·0023	194 (23%)
Hypertension	0 (0%)	<0·0001	183 (70%)	0·526	506 (68%)
Coronary heart disease	0 (0%)	<0·0001	110 (42%)	0·171	270 (37%)
Diabetes mellitus	0 (0%)	<0·0001	94 (36%)	0·011	206 (27%)
Atrial fibrillation	7 (6%)	0·033	36 (14%)	0·576	106 (12%)
Haemodynamics					
RAP, mmHg	6 [4, 9]	0·126	7 [5, 10]	0·0011	6 [4, 9]
mPAP, mmHg	48 [40, 57]	0·0016	43 [36, 51]	<0·001	39 [33, 46]
PAWP, mmHg	8 [5, 10]	0·0003	10 [7, 12]	0·0148	9 [6, 11]
CI, L/min/m^2	2·1 [1·7, 2·7]	0·680	2·0 [1·6, 2·4]	0·051	2·1 [1·8, 2·6]
PVR, WU	10·9 [7·8, 15·6]	0·0005	8·7 [6·5, 12·0]	<0·0001	7·4 [5·9, 10·1]
SvO_2, %	66 [59, 70]	0·0011	62 [55, 66]	<0·0001	65 [59, 57]
Risk (4-strata model)[a]					
Low	16 (12%)	<0·001	5 (2%)	0·966	16 (2%)
Intermediate-low	42 (33%)		34 (13%)		108 (12%)
Intermediate-high	57 (45%)		139 (52%)		463 (52%)
High	13 (10%)		88 (33%)		311 (35%)

PH medications					
CCB	26 (20%)	<0·0001	10 (4%)	0·032	13 (1%)
ERA	56 (44%)	0·0007	70 (26%)	<0·0001	59 (6%)
PDE5i	82 (64%)	<0·0001	223 (83%)	<0·0001	852 (94%)
sGCs	11 (9%)	0·217	13 (5%)	0·005	15 (2%)
PPA	7 (5%)	0·166	6 (2%)	0·341	11 (1%)
Monotherapy	81 (63%)	<0·0001	220 (82%)	<0·0001	871 (96%)
Combination therapy	47 (37%)		48 (18%)		37 (4%)

Table 5.1 Patient characteristics at baseline in COMPERA

Categorical data are shown as n and (%) of the respective population. Continuous data are depicted as median [Q1, Q3].

[a]Risk was determined as published elsewhere **(14 ref)**

Definition of abbreviations: BMI, body mass index; IPAH, idiopathic pulmonary arterial hypertension; PH, pulmonary hypertension; WHO FC, World Health Organization Functional Class; 6MWD, 6-minute walking distance; NT-proBNP, N-terminal fragment of pro-brain natriuretic peptide; TLC, total lung capacity; FVC, forced vital capacity; FEV_1, forced expiratory volume in 1 s; DLCO, diffusion capacity of the lung for carbon monoxide; RA, right atrial pressure; mPAP, mean pulmonary arterial pressure; PAWP, pulmonary arterial wedge pressure; CI, cardiac index; PVR, pulmonary vascular resistance; SvO_2, mixed-venous oxygen saturation; CCB, calcium channel blocker; ERA endothelin receptor antagonists; PDE5i, phosphodiesterase-5 inhibitors; sGCs, stimulator of soluble guanylate cyclase; PPA, prostacyclin pathway agents.

5.7.2 Table 2 Patient characteristics at baseline in ASPIRE

	Classical IPAH (i) n=185	P-value (i) vs. (ii)	IPAH with a lung phenotype (ii) n=139	P-value (ii) vs. (iii)	Group 3.1 or 3.2 PH n=375 (iii)
Age, years	52 [38, 64]	<0.0001	71 [65, 76]	0.0486	69 [63, 74]
Female	133 (72%)	<0.0009	75 (54%)	0.0032	148 (39%)
BMI, kg/m^2	28 [25, 34]	0.431	28 [25, 31]	0.0561	27 [23, 31]
WHO FC I II III IV	 0 (0%) 47 (25%) 119 (64%) 19 (10%)	<0.0001	 0 (0%) 10 (7%) 80 (58%) 49 (35%)	0.939	 0 (0%) 29 (8%) 208 (56%) 135 (36%)
ISWD, m	260 [140, 400]	<0.0001	90 [30, 150]	0.199	70 [30, 140]
Pulmonary function FVC, % pred FEV$_1$, % pred FEV$_1$/FVC (%) DLCO, % pred	 97 [84, 110] 87 [75, 97] 75 [69, 80] 62 [52, 73]	 0.0114 0.259 <0.0001 <0.0001	 103 [91, 112] 88 [74, 99] 70 [63, 76] 27 [22, 34]	 <0.0001 <0.0001 <0.0001 0.0498	 82 [62, 102] 62 [44, 80] 63 [48, 76] 25 [19, 32]
Smoking history Ever Never Pack years	 76 (45%) 92 (55%) 20 [10, 30]	 <0.0001	 139 (100%) 0 (0%) 30.0 [20, 40]	 n/a	 n/a

Haemodynamics					
RAP, mmHg	9 [7, 14]	0.328	10 [7, 14]	0.0002	8 [5, 12]
mPAP, mmHg	54 [46, 64]	<0.0001	49 [43, 56]	<0.0001	41 [34, 49]
PAWP, mmHg	10 [8, 12]	0.636	10 [8, 13]	0.3722	11 [8, 13]
CI, L/min/m^2	2.3 [1.8, 2.9]	<0.0001	2.0 [1.6, 2.4]	<0.0001	2.6 [2.0, 3.1]
PVR, WU	10.5 [7.2, 14.8]	0.5	11.1 [7.8, 14.6]	<0.0001	6.5 [4.2, 9.9]
SvO_2, %	64 [58, 69]	<0.0001	58 [53, 66]	<0.0001	66 [60, 71]
Treatment*					
None	2 (1.1%)		2 (1.4%)		180 (48%)
CCB	17 (10%)		0 (0%)		1 (0.3%)
Oral monotherapy	40 (24%)		43 (31%)		165 (44%)
Oral combination	79 (47%)	0.0004	72 (52%)	<0.0001	22 (6%)
PPA ± oral therapy	29 (19%)		21 (15%)		7 (2%)

Table 5.2 Patient characteristics at baseline in ASPIRE

Categorical data are shown as n and (%) of the respective population. Continuous data are depicted as median [Q1, Q3].

Definition of abbreviations: BMI, body mass index; IPAH, idiopathic pulmonary arterial hypertension; PH, pulmonary hypertension; WHO FC, World Health Organization Functional Class; ISWD, incremental shuttle walk distance; FVC, forced vital capacity; FEV_1, forced expiratory volume in 1 s; DLCO, diffusion capacity of the lung for carbon monoxide; RA, right atrial pressure; mPAP, mean pulmonary arterial pressure; PAWP, pulmonary arterial wedge pressure; CI, cardiac index; PVR, pulmonary vascular resistance; SvO_2, mixed-venous oxygen saturation; CCB, calcium channel blockers; PPA, prostacyclin pathway agents.

*Oral monotherapy includes PDE5i or ERA or SGCs; oral combination includes ERA in combination with PDE5i or SCGs; PPA +/- oral therapy includes prostanoids either alone or in combination with PDE5i or sGCs +/- ERA.

5.7.3 Table 3 Lung parenchymal abnormalities on chest computed tomography (ASPIRE)

	Classical IPAH (i) n=185	P-value (i) vs. (ii)	IPAH with a lung phenotype (ii) n=139	P-value (ii) vs. (iii)	Group 3.1 or 3.2 PH n=375 (iii)
CT available	109 (79%)	0·1145	86 (70%)	0·3147	219 (75%)
CT – Fibrosis (any present)	9 (8%)	<0·0001	26 (30%)	0·0093	102 (47%)
CT – Fibrosis (by severity)					
None	100 (93%)		60 (71%)		117 (57%)
Mild	6 (6%)	<0·0001	21 (25%)	<0·0001	21 (10%)
Moderate	1 (1%)		4 (5%)		33 (16%)
Severe	0 (0%)		0 (0%)		36 (17%)
CT – Emphysema (any present)	15 (14%)	<0·0001	42 (49%)	0·0694	132 (60%)
CT – Emphysema (by severity)					
None	94 (89%)		44 (52%)		87 (41%)
Mild	11 (10%)		22 (26%)		21 (10%)
Moderate	1 (1%)	<0·0001	16 (19%)	<0·0001	62 (30%)
Severe	0 (0%)		3 (4%)		40 (19%)

Table 5.3 Lung parenchymal abnormalities on chest computed tomography (ASPIRE)

Data are shown as n (%). Statistical comparisons were made by Pearson's Chi-squared test or Fisher's exact test.

5.8 Figures

** Please note the numbering for the figures differences between the overall thesis and the paper. Both (eg, 1a and 1b) have been included, and please use the original paper figures to refer to the manuscript body.*

5.8.1 STROBE diagram showing patient selection in COMPERA

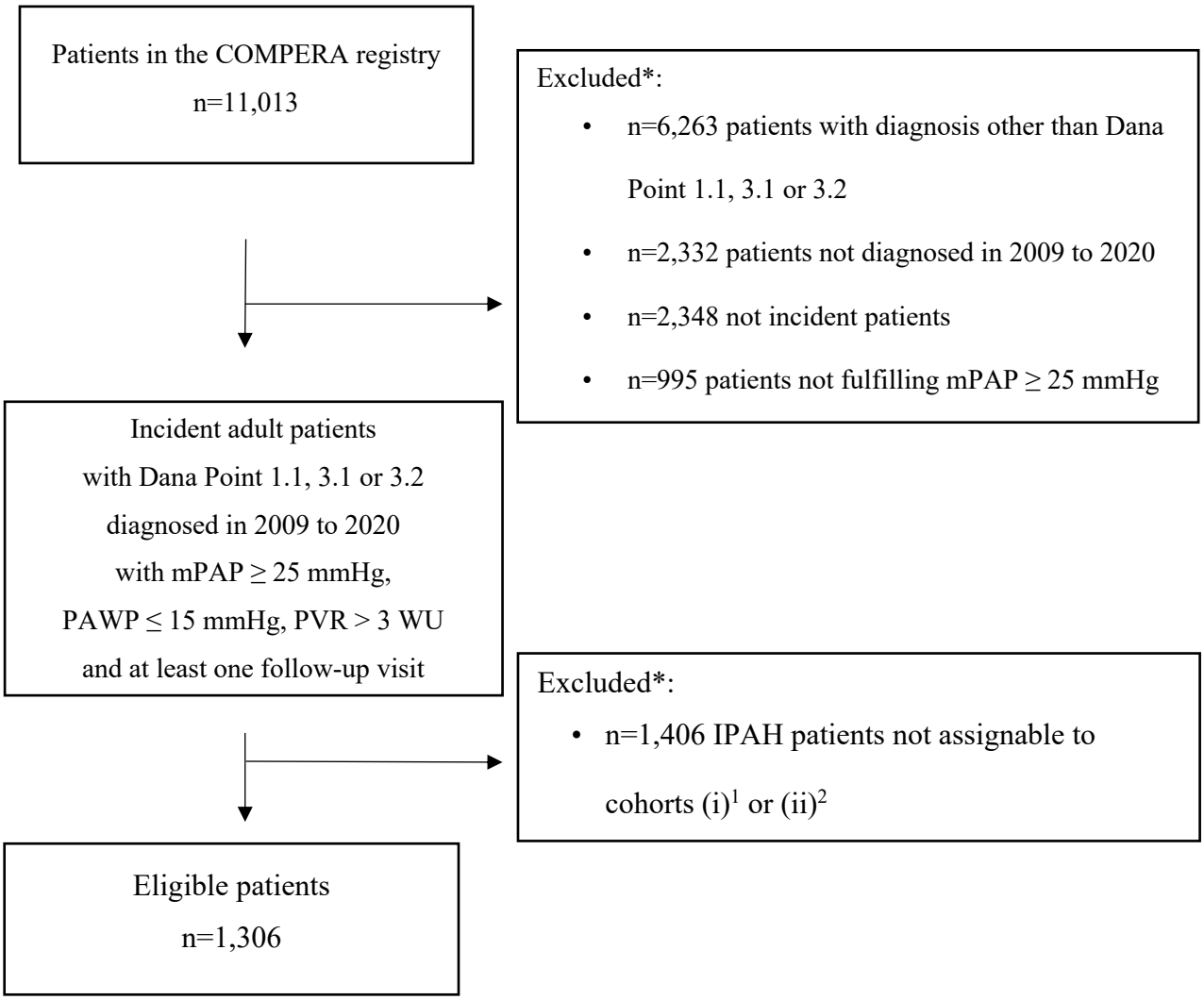


Figure 5.1 STROBE diagram showing patient selection in COMPERA

*more than one reason for exclusion could apply

[1] (i) Patients with classical IPAH, defined by the absence of risk factors for left heart disease (body mass index (BMI) ≥ 30 kg/m^2, hypertension, diabetes mellitus, and coronary heart disease), and a DLCO ≥ 45%

[2] (ii) Patients diagnosed with IPAH and a lung phenotype, defined by a smoking history (i.e., current or former smoker) and a DLCO < 45% of the predicted value

5.8.2 STROBE diagram showing patient selection in ASPIRE

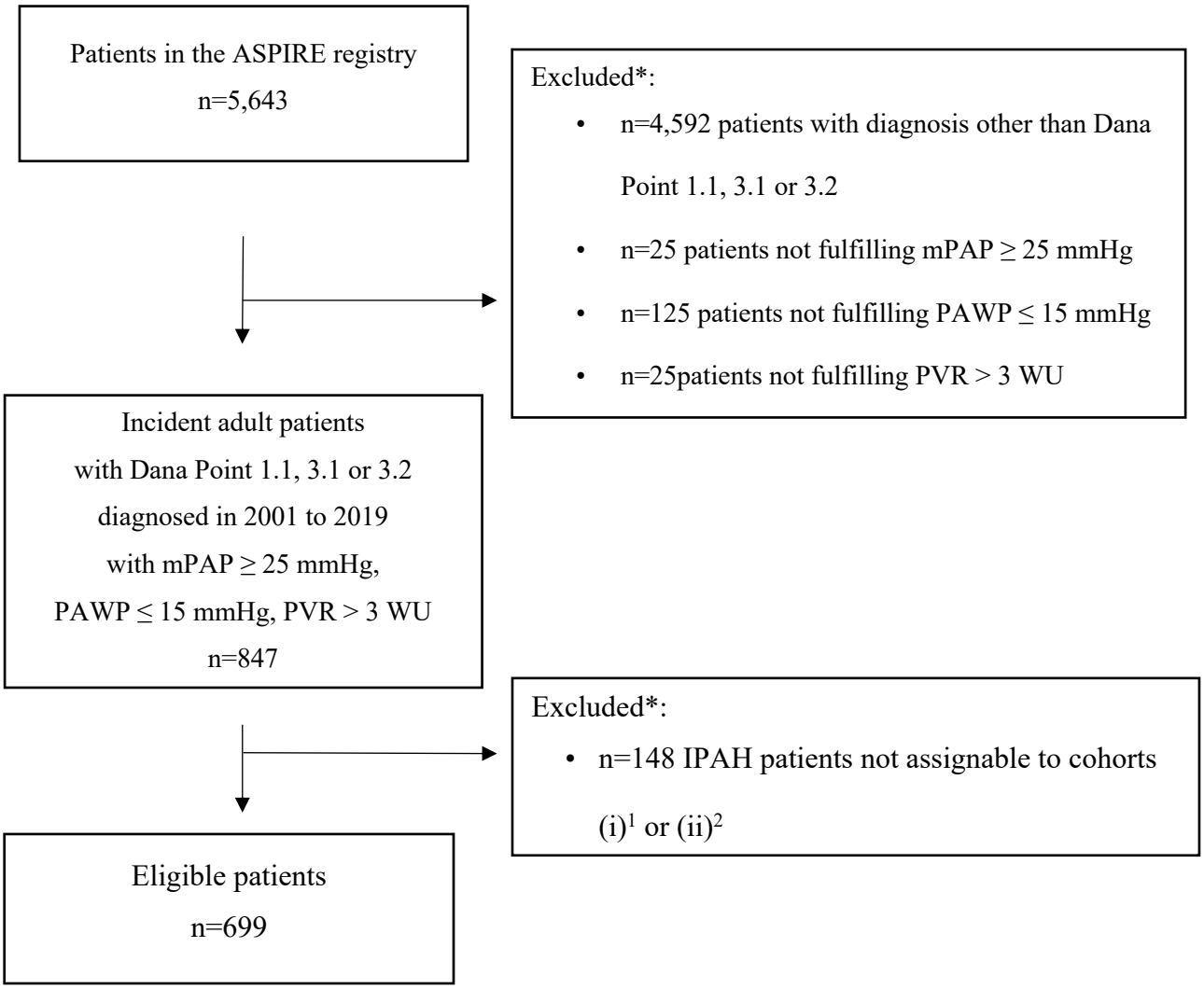


Figure 5.2 STROBE diagram showing patient selection in ASPIRE

*more than one reason for exclusion could apply

[1] (i) Patients with classical IPAH and a DLCO ≥ 45%

[2] (ii) Patients diagnosed with IPAH and a lung phenotype, defined by a smoking history (i.e., current or former smoker) and a DLCO < 45% of the predicted value

5.8.3 Grouped barplot showing age distribution of patients classified as classical IPAH, IPAH with a lung phenotype, and group 3 PH in COMPERA

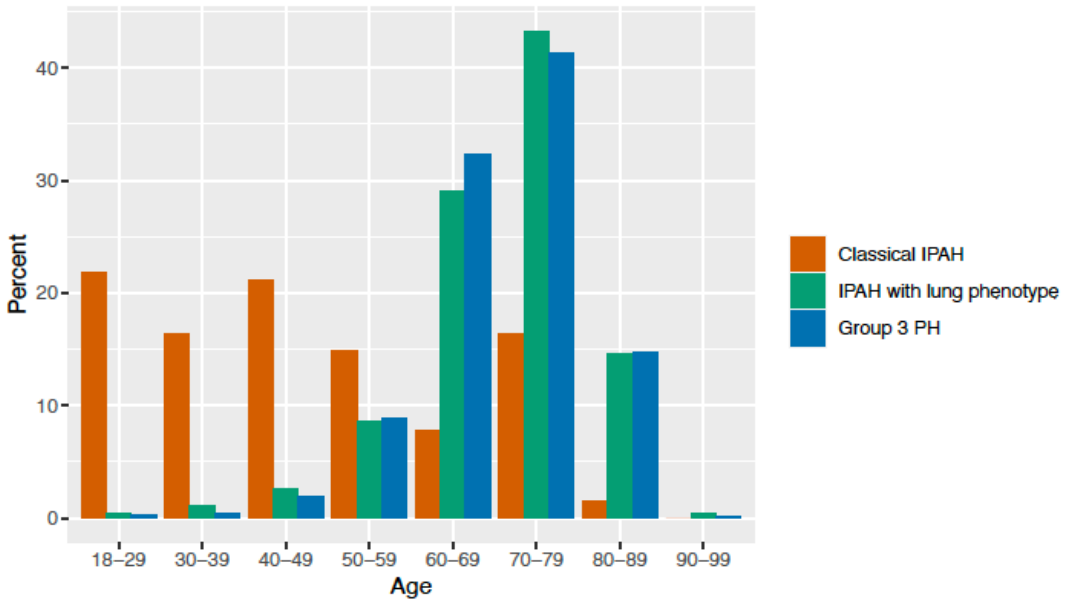


Figure 5.3 Grouped barplot showing age distribution of patients classified as classical IPAH, IPAH with a lung phenotype, and group 3 PH in COMPERA

5.8.4 Grouped barplot showing age distribution of patients classified as classical IPAH, IPAH with a lung phenotype, and group 3 PH in ASPIRE

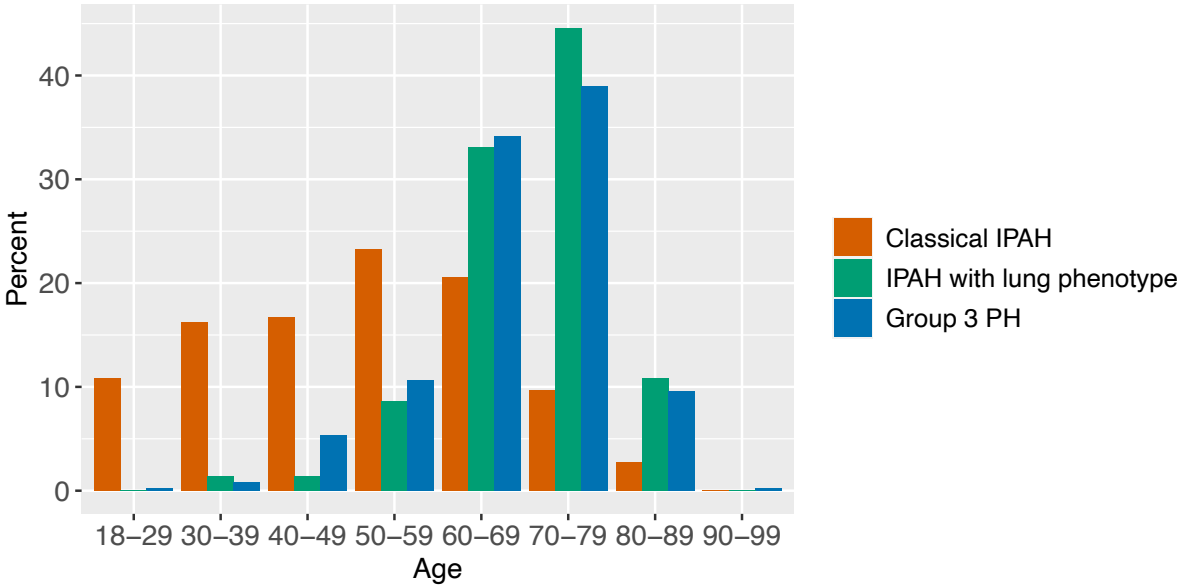
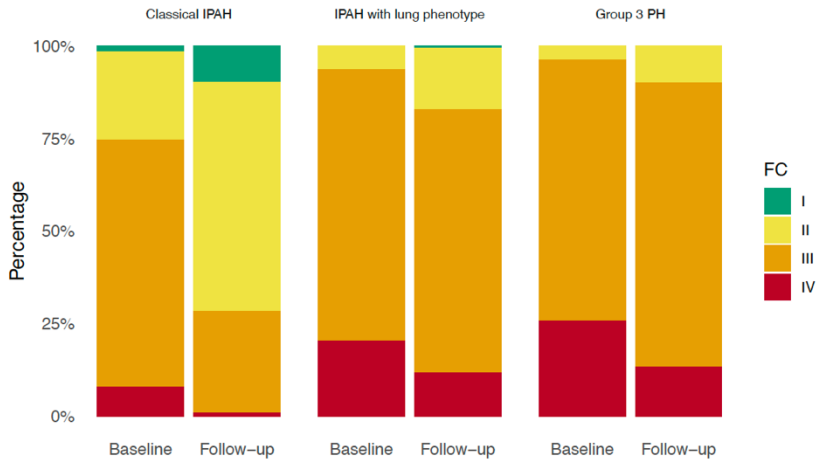


Figure 5.4 Grouped barplot showing age distribution of patients classified as classical IPAH, IPAH with a lung phenotype, and group 3 PH in ASPIRE

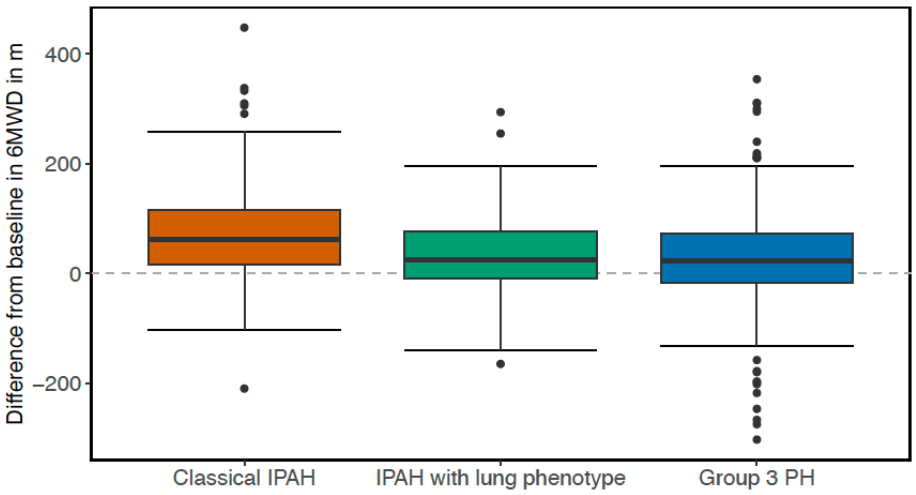
5.8.5 Baseline and first follow-up measurement for (a) functional class (FC), (b) 6-minute walking distance (6MWD), (c) N-terminal fragment of pro-brain natriuretic peptide (NT-proBNP) and (d) mortality risk (as determined by the ESC/ERS 4-strata model) in COMPERA

a)



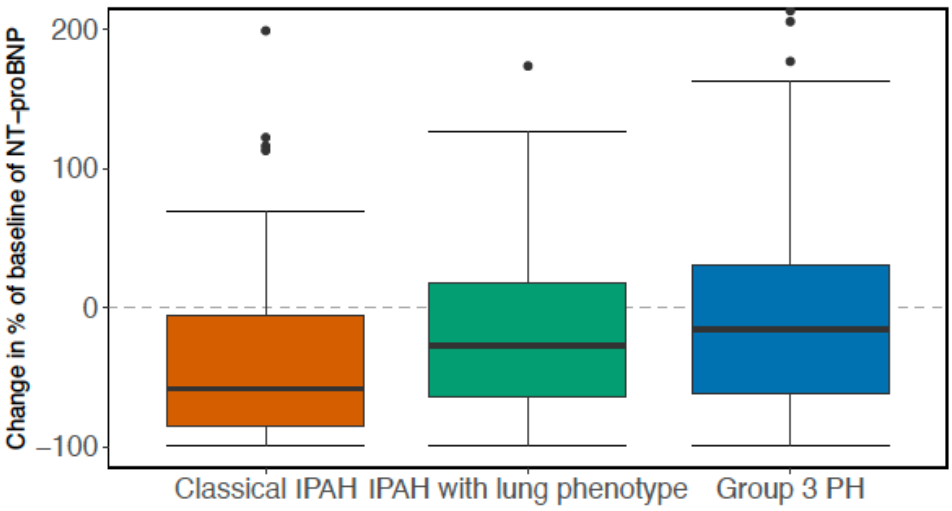
WHO FC improved from baseline to first follow-up in 54% of the patients with classical IPAH, 26% of patients with IPAH and a lung phenotype and 22% in patients with group 3 PH ($p<0{\cdot}0001$ for classical IPAH versus IPAH and a lung phenotype, and $p=0{\cdot}194$ for IPAH and a lung phenotype versus group 3 PH).

b)



6MWD improved from baseline to first follow-up by 83 ± 111 m in patients with classical IPAH, by 31 ± 82 m of patients with IPAH and a lung phenotype, and by 27 ± 89 m in patients with group 3 PH (p=0·0015 for classical IPAH versus IPAH and a lung phenotype, and p=0·64 for IPAH and a lung phenotype versus group 3 PH).

c)



NT-proBNP decreased from baseline to first follow-up by 58 [-85, -6] % in patients with classical IPAH, by 27 [-64, 18] % of patients with IPAH and a lung phenotype, and by 16 [-62, 30] % in patients with group 3 PH (p=0·0043 for classical IPAH versus IPAH and a lung phenotype, and p=0·142 for IPAH and a lung phenotype versus group 3 PH).

d)

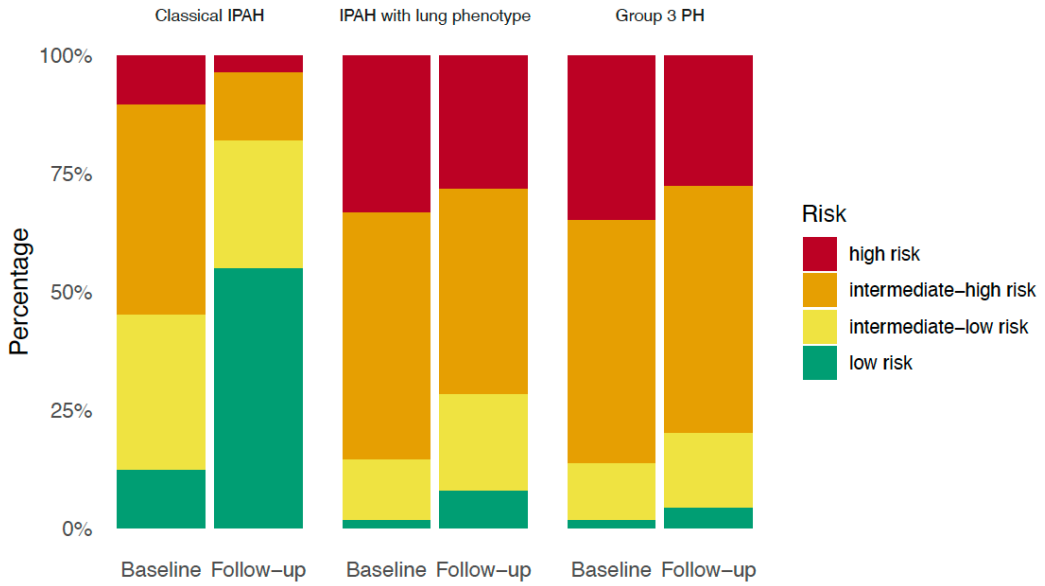


Figure 5.5 Baseline and first follow-up measurement for (a) functional class (FC), (b) 6-minute walking distance (6MWD), (c) N-terminal fragment of pro-brain natriuretic peptide (NT-proBNP) and (d) mortality risk (as determined by the ESC/ERS 4-strata model) in COMPERA

Risk improved from baseline to first follow-up in 64% of the patients with classical IPAH, 32% of patients with IPAH and a lung phenotype and 29% in patients with group 3 PH ($p<0{\cdot}0001$ for classical IPAH versus IPAH and a lung phenotype, and $p=0{\cdot}343$ for IPAH and a lung phenotype versus group 3 PH).

5.8.6 Kaplan-Meier survival estimates for patients classified as classical IPAH, IPAH with a lung phenotype, and group 3 PH in COMPERA

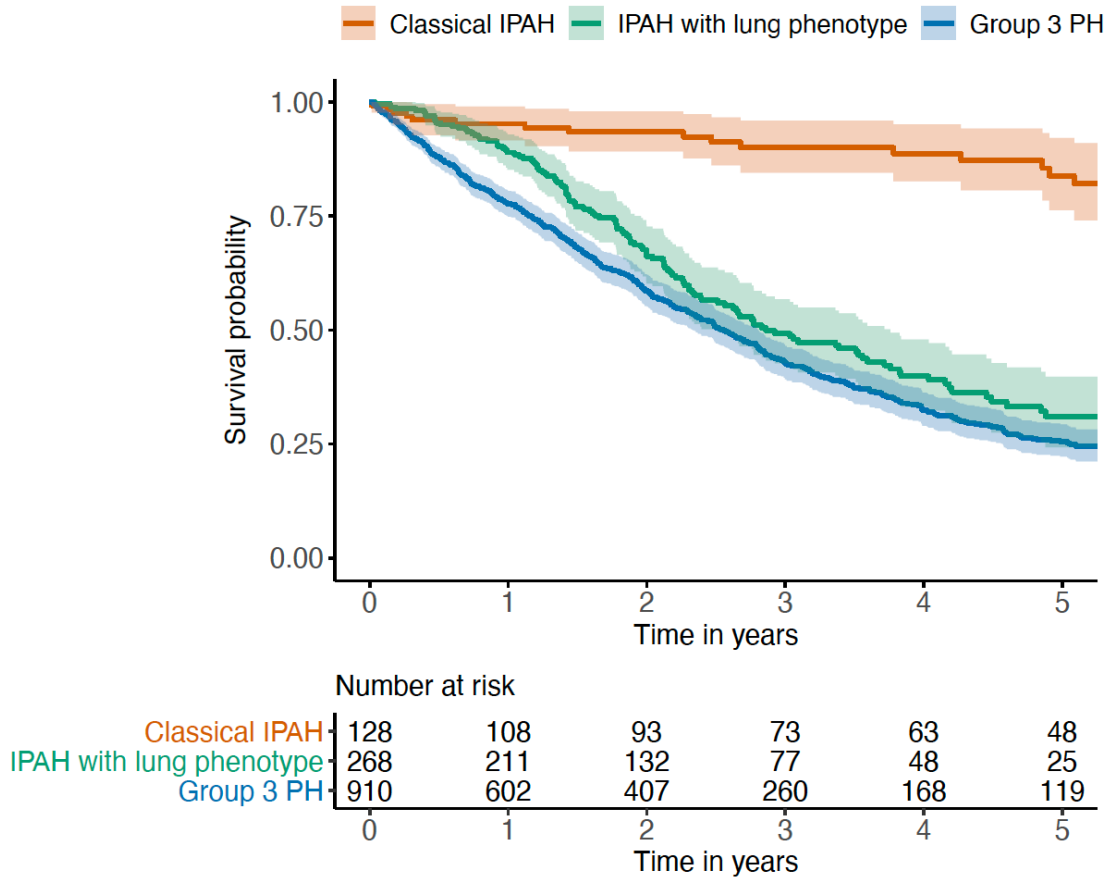


Figure 5.6 Kaplan-Meier survival estimates for patients classified as classical IPAH, IPAH with a lung phenotype, and group 3 PH in COMPERA

5.8.7 Kaplan-Meier survival estimates for patients classified as classical IPAH, IPAH with a lung phenotype, and group 3 PH in ASPIRE

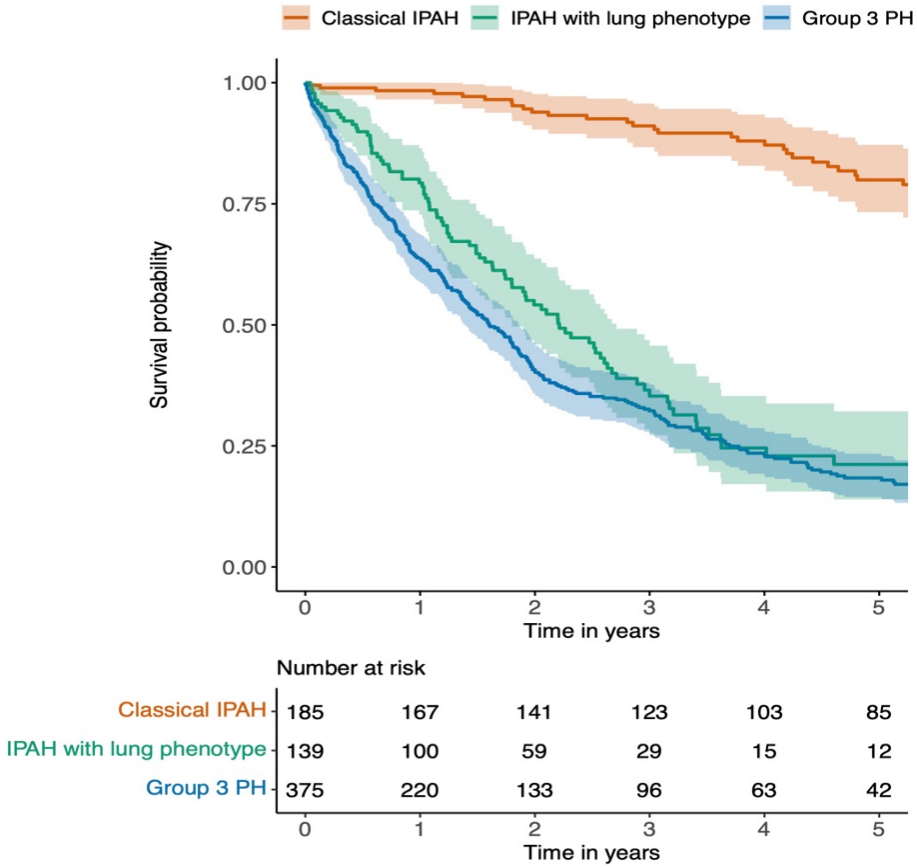


Figure 5.7 Kaplan-Meier survival estimates for patients classified as classical IPAH, IPAH with a lung phenotype, and group 3 PH in ASPIRE

6 External validation, radiological evaluation, and development of deep learning automatic lung segmentation in contrast enhanced chest CT

Krit Dwivedi BMedSci, BMBS (MD), FRCR[*1], Michael Sharkey MEng MSc*[2], Samer Alabed MD, FRCR[1], Curtis P. Langlotz MD, PhD, FACMI, FSIIM [3], Andy J Swift MD, PhD, FRCR*[1], Christian Bluethgen, MD, MSc*[3]

**Submitted to European Radiology journal.*

* These authors contributed equally to this work
1 - Department of Infection, Immunity & Cardiovascular Disease, University of Sheffield, Medical School, Beech Hill Rd, Sheffield S10 2RX
2- 3DLab, Sheffield Teaching Hospitals NHS Trust, Sheffield S5 7AU/
3- Stanford Center for Artificial Intelligence in Medicine and Imaging (AIMI), Stanford University, 1701 Page Mill Road, Palo Alto, CA 94304, United States

Contribution: First author. Contributor Roles Taxonomy (CRediT) – conceptualisation, data curation, formal analyses, funding analysis, methodology, project administration, resources, software, validation, visualisation, writing – original draft, review, and editing).

I lead this multicentre study at a placement at the Stanford AIMI Centre. I did all the data analysis and independently wrote the first draft. This was reviewed by the collaborators CB and supervisor AS. MS and CB helped write the code for performing the lung segmentation. All other authors had a small involvement with data gathering, support, project administration and resources.

6.1 Abstract

Objectives: For clinical translation of segmentation studies, there is a need to radiologically evaluate outputs, understand limitations and identify failure points. This study develops an accurate lung segmentation model in contrast enhanced CT pulmonary angiography (CTPA), and clinically radiologically evaluates outputs in two diverse patient cohorts with pulmonary hypertension (PH) and interstitial lung disease (ILD).

Materials and Methods: This retrospective study develops an nnU-net-based segmentation model using data from two specialist centres (UK & USA). Model was trained (n=36), tested (n=12) and clinically evaluated (n=177) on a diverse 'real-world' cohort of 225 PH patients with volumetric CTPAs. Ground truth was segmented by a radiologist. Dice score coefficient (DSC) and Normalised Surface Distance (NSD) were used for testing. Clinical evaluation of outputs was performed by two radiologists who assessed clinical significance of errors. External validation was performed on CTPA, HRCT, arterial and non-contrast scans from 28 ILD patients. All analysis was performed at patient level.

Results: Both patient cohorts had diverse demographic and clinical characteristics. Mean accuracy, DSC, NSD scores were 0.998 (95% CI 0.9976, 0.9989), 0.990 (0.9840, 0.9962), and 0.983 (0.9686, 0.9972) respectively. There were no segmentation failures. 82% and 71% of internal and external cases respectively were segmented without error. 18% and 25% respectively had clinically insignificant errors. Failure analysis demonstrated peripheral atelectasis and consolidation as common causes for suboptimal segmentation. One external case with patulous oesophagus had a clinically significant error.

Conclusion: 3D CTPA lung segmentation model provides accurate outputs with few clinically significant errors on validation across two diverse cohorts with PH and ILD.

Key words: Computational Tomography, Contrast enhanced CT, Artificial Intelligence, Segmentation, Deep-learning, Clinical evaluation, External Validation, Pulmonary Hypertension, Interstitial Lung Disease

6.2 Introduction

Lung segmentation in Computational Tomography (CT) imaging is the detection and extraction of the anatomical lung boundary on each slice of the study. Automated lung segmentation is an important and necessary step in almost all chest CT clinical artificial intelligence (AI) applications such as lung nodule detection or lung parenchymal disease severity quantification. Accurate segmentation allows for detection of the region of interest (lung parenchyma) and removal of confounders (cardiac and mediastinal structures) and is important as any segmentation error propagates throughout the rest of the data analysis pipeline. Manual segmentation is time consuming, arduous, and has significant inter and intraobserver variability [194,195]. It is impractical for radiologists to perform segmentation during routine clinical reading.

For translation into the clinical domain, there is a need to better evaluate segmentation outputs from a clinical perspective to understand model limitations and common failure points. A vast range of lung segmentation techniques exist for non-contrast chest CT, from traditional computer vision methods approaches, to newer machine and deep learning approaches using convolutional neural networks, which have surpassed the performance of older methods [194,196–199]. However, the vast majority of studies in this domain focus on technical developments with advances in underlying network architectures or processing techniques [196].

Computed tomography pulmonary angiography (CTPA) involves administration of intravenous contrast which enables assessment of the pulmonary vasculature in addition to other structures on chest CT. It is performed routinely for patients with suspected pulmonary embolism (PE) and pulmonary hypertension (PH). A known limitation of current deep-learning segmentation algorithms is failure to segment high density objects such as subpleural consolidation [196,200,201]. There are significant differences in the attenuation of the lung parenchyma between CTPA and non-contrast imaging due to parenchymal uptake of contrast. To the best of our knowledge, no study has trained a lung segmentation algorithm in PH or CTPA imaging and externally tested it in a heterogeneous mixed cohort of chest CT protocols.

This study develops a 3D deep-learning CTPA lung segmentation algorithm using the state-of-the-art nnU-net architecture [202]. nnU-net has shown breakthrough performance and has become the new standard in a variety of medical imaging segmentation tasks [203]. The study aims are to:

1. To develop a novel state-of-art nnU-net based 3D lung segmentation algorithm in CTPA imaging.
2. To clinically evaluate and score segmentation outputs by review from expert subspeciality thoracic radiologists, then perform failure analysis on cases with suboptimal performance.
3. To deploy and externally validate the algorithm in a heterogeneous patient cohort at another centre.

6.3 Materials and methods

This retrospective study uses data from two patient cohorts - Sheffield, UK and Stanford, USA. All patient data was de-identified as per GDPR and HIPAA compliant local guidelines. Ethical approval was granted by the Institutional Review Board at both centres and approved by the UK National Research Ethics Service (16/YH/0352). All analysis is done at the patient level, with each patient having a single corresponding CT scan.

6.3.1 Study cohorts

6.3.1.1 Sheffield (reference dataset)

Patients were selected from the ASPIRE registry, the details of which have been previously reported [41,50]. The registry prospectively includes comprehensive clinical and radiological data on patients referred with suspected pulmonary hypertension (PH) to a tertiary referral centre.

226 patients with a heterogeneous mix of normal and abnormal chest CT findings formed the internal reference dataset. 48 random cases were used for model development, randomly divided into 36 for model training (8 for tuning) and 12 for technical performance testing. Training cohort sample size was based on recent studies in which similarly sized cohorts utilising 3D segmentation approaches demonstrated high technical performance [196,204].

Radiological clinical validation was performed on all remaining unseen 177 cases. All scans were performed on two General Electric (GE) scanners with the patient in a supine position. Scanning parameters included multiple doses and kernels. All patients had thin-slice volumetric scans with contrast in the CTPA protocol. Patients had a diagnosis of either Idiopathic Pulmonary Arterial Hypertension (IPAH) or Pulmonary Hypertension secondary to Chronic Lung Disease (PH-CLD).

6.3.1.2 Stanford (external dataset)

To evaluate the model's generalisation performance on external data, an additional test set of CT scans was selected from a mixed cohort of patients evaluated for interstitial lung disease at Stanford Hospital and Clinics in a tertiary care setting. From a total of 2300 CT scans from 1330 patients, 28 scans were randomly selected. Scans were acquired using Siemens (n=21), GE (n=6) or Toshiba CT scanners (n=1) and reconstructed with a variety of convolutional kernels. All scans were performed with the patient in a supine position. 25 (89.3%) CT scans were non-enhanced, two (7.1%) of the scans were obtained using a CTPA protocol, and one scan (3.5%) was obtained using a CTA protocol. The underlying ILD diagnostic groups were connective tissue disease-related ILD (11 (39.3%)), exposure-related ILD (7 (25%)), idiopathic interstitial pneumonia (6 (21.4%)), post-infectious scarring of the lung parenchyma (1 (3.5%)) and IPAH (1 (3.5%)). No signs of ILD were seen in two scans (7.1%).

6.3.2 Model development

6.3.2.1 Ground truth manual segmentation

MIM (MIM Software, Cleveland, USA) was used to label and generate ground truth lung segmentation masks. A workflow was developed which used built-in operators to create an initial mask, and each step was continually reviewed by a certified radiologist. The RegionGrow tool was used to segment the lungs and airways from the trachea, manually checking for errors. Thresholds were -350 to -4000 Hounsfield Units (HU). Tendril diameter was 4.0, with 'fill holes' set to none and smoothing enabled. The workflow was continuously manually evaluated, with appropriate technical parameters adjusted to improve performance. After achieving this baseline result, the scan was then manually adjusted and contoured on a

slice-by-slice level by the radiologist to ensure correct segmentation of the lung border. This output was used as the ground truth for model development.

6.3.2.2 Deep-learning model development

Pre-processing steps included truncating the HU range to -1024 to 2500 then normalisation using SD (199HU) and mean (-761HU) of the segmented lung region across all training images. A single fold training approach was used, consisting of 1000 epochs with 250 mini-batches per epoch. Data augmentation during training was used with random rotation (-30^{o} to 30^{o} about 3 axes, p=0.2) , scaling (0.7 to 1.4, p=0.2), gamma correction (0.7 to 1.5, p=0.3) and mirroring about 3 axes (p=1). A two stage cascade was used where the output segmentation from the first stage was passed to the second stage. During training the output from the first stage has augmentations to randomly remove connected pixels and conduct morphological operations to enlarge/shrink the output to reduce co-adaptation. For the first stage the training images are resampled to a mean pixel size of 1.01x1.01x0.84mm and a 380x380x346 matrix, with a mean pixel size of 0.75x0.75x0.625 and 512x512x467 matrix for the second stage. A patch size of 256x224x224 with a batch size of 2 was used with a learning rate of $1x10^{-3}$. Dice's Similarity Coefficient (DSC), accuracy, and normalised surface distance (NSD) were calculated by comparing manual to deep learning segmentations for each case in the test cohort. Hardware used was a NVIDIA A6000 48GB GPU, 32 Core 64 thread processor, 256GB RAM. Post processing steps involved removal of regions <250 ml in volume.

6.3.3 Clinical segmentation scoring system: RadSeg

Segmentation outputs were clinically reviewed by two radiologists (KD and CB) with 4 and 7 years of experience in interpreting thoracic CT scans. An ordinal score (RadSeg score) was given to each segmentation output, scored as:

- 0 - failed to output a segmentation.
- 1 - lung segmented with clinically significant error.
- 2 - lung segmented with minor clinically insignificant error.
- 3 - full lung segmented without any clinically significant errors.

One radiologist reviewed all 177 cases from the 'Site 1' test cohort. The 28 external 'Site 2'

test set cases were reviewed by two radiologists. Cases with differing scores were reviewed together and a consensus RadSeg score was given.

6.3.4 Statistical Analysis

All analyses from Sheffield and Stanford were performed separately; the data were not combined. Analyses were performed using R software major version 4. Categorical data are presented as number and percentage, continuous data as median and interquartile range. Segmentations were compared using an overlap-based-metric (Dice Similarity Coefficient (DSC)), and a boundary-based-metric (Normalised Surface Distance (NSD). For NSD a threshold of 1.5 mm was used, as it was felt to be a clinically appropriate threshold. A p value of 0·05 or less was considered statistically significant.

6.3.5 Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

6.4 Results

Patient clinical characteristics of both cohorts are shown in Table 1. Patients had a diverse range of hemodynamic, spirometric and demographic factors.

6.4.1 Technical results

Mean accuracy, DSC score, and NSD across the internal validation cases was 0.998 (95% CI 0.9976 to 0.9989), 0.990 (95% CI 0.9840 to 0.9962) and 0.983 (95% CI 0.9686 to 0.9972) respectively. Scores for each patient are shown in Appendix Table 1a.

6.4.2 Clinical segmentation scores

6.4.2.1 Sheffield internal validation

There were no failures (RadSeg score 0) and all cases were successfully segmented by the algorithm. 146 (82%) cases had a full lung segmentation without any clinically significant error (RadSeg score 3) and 31 (18%) had a minor clinically insignificant error (RadSeg score

2). No cases were segmented with a clinically significant error (RadSeg score 1). The most common reasons for segmentation errors were consolidation (9, 29%), atelectasis (7, 23%) and pleural effusion (6, 19%). These findings are shown in Table 2.

6.4.2.2 Stanford external validation

There were no failures (RadSeg score 0) and all cases were successfully segmented by the algorithm. 20 (71%) cases had a full lung segmentation without any clinically significant error (RadSeg score 3) and 7 (25%) had a minor clinically insignificant error (RadSeg score 2). The reasons for segmentation errors were atelectasis, pleural effusion and basal severe fibrosis.

The one case with a clinically significant error had a patulous dilated oesophagus secondary to scleroderma, in which the gas filled oesophagus was included within the lung segmentation.

6.5 Discussion

This study presents a state-of-the-art 3D lung segmentation algorithm with clinical validation across two centres in distinct well phenotyped clinical cohorts. The model achieved high technical precision, indicated by high DICE and NSD scores, and clinical utility, indicated by high scores on radiological review. To our knowledge, this is the first model developed specifically in CTPA imaging and in a cohort of PH patients.

The study differs significantly in study design, patient cohort and model evaluation compared to most segmentation studies currently in the literature. The methods were specifically designed to address common limitations that make clinical use and translation challenging. These are a lack of clinical evaluation of the segmentation output from expert radiologists, lack of diverse, heterogenous real world clinical cohorts, and external validation. This study addresses each limitation by clinically evaluating model outputs, and by developing and testing the model on two different patient cohorts from two centres.

Importance of radiological review of segmentation outputs

The most common parameter used to report and compare segmentation performance is the Dice Similarity Coefficient (DSC). However, technical parameters alone are insensitive in assessing clinical utility. Almost all studies show DSC scores >0.97, but failure analysis of cases with suboptimal performance is rarely reported (14). DSC itself is known to be highly reliant on structure size and providing artificially high scores by ignoring missing values [205]. The structure size limitation is particularly appropriate in the clinical use of lung segmentation, where due to the large lung volume, small volume but highly clinically significant segmentation errors - such as segmenting the oesophagus or failing to segment basal ground glass change in interstitial lung disease - have minimal impact on the reported DSC score.

We therefore propose and utilise a clinical segmentation scoring system (RadSeg score), which accounts for the small volume and clinical significance of errors. Our model had no major clinically significant errors when validated against scans from the reference centre (Sheffield) and only one after deployment in an external centre (Stanford). As radiologists are the target user in mind for such AI models, their involvement in the development and clinical evaluation of such models, we believe will help create trust in the model's output and help increase uptake in routine clinical practice [60,206].

Analysis of suboptimal outputs

Transparency of the algorithm limitations and causes of suboptimal performance is important to clinical trust. The training and test cohorts in our study included cases with severe parenchymal abnormalities (Figure 2, panel C), and the majority of these accurately segmented. The most common reasons for suboptimal segmentation (RadSeg Score 1 or 2) were consolidation and atelectasis. Accurate segmentation of these high-density parenchymal abnormalities is challenging, and is a limitation common to all lung segmentation models and studies [196,200]. The causes for suboptimal performance are hard to directly compare, as they are often not reported or characterised.

Importance of external deployment and validation.

Our model had no failures either in the internal or external test sets, with only one clinically significant error in the external set. External validation of any AI model is vital to test model performance and generalisability. However only 6% of published deep learning studies in the field of diagnostics performed external validation [207]. The majority of studies with external

validation demonstrate diminished algorithm performance [208]. Within the domain of lung segmentation, Yoo et al found a significant drop off (DSC 99.4 in internal test to DSC 95.3 in external test) in the performance on their 3D U-Net based model on external validation [201]. This study performs external validation not only after deployment in a different centre, but on a different clinical cohort of patients, and demonstrates good results on clinical evaluation. The external cohort is heterogeneous both in containing patients with and without PH, but also a random mix of different volumetric chest CT protocols, with and without contrast opacification.

Need for heterogeneous 'real world' clinical cohorts

Our technical results of mean DSC of 0.990 and NSD of 0.983 compare favourably against other published 3D and 2D segmentation approaches on non-contrast scans [196,199,201,209,210]. This result likely can be attributed to the use of a state-of-the-art nnU-net architecture combined with a heterogeneous 'real world' clinical training cohort with a mix of pathology. Hofmanninger et al. showed that the accuracy and reliability of lung segmentation algorithms on difficult cases primarily relies on the strength of the training data, more so than the underlying model architecture [196]. This effect is particularly pronounced on external validation against a variety of test sets, with algorithms trained on more diverse data being more generalisable. Previous studies have been limited by training on relatively homogenous cohorts, which is a by-product of utilising public datasets, which only contain normal scans or variations of a specific disease class [203,211].

Implications for practice, and intended clinical use

Accurate lung segmentation is a necessity for any quantitative chest CT analysis. We envisage clinical use of this model as the first pre-processing step in quantitative CT analysis models such as lung parenchymal disease severity quantification in PH. There is a clinical need and great interest in better characterising the severity and extent of lung parenchymal disease in PH [212]. Qualitative routine radiological report descriptions of parenchymal disease have shown to be prognostic biomarkers [178]. A new IPAH phenotype has been identified with distinct radiological and clinical characteristics [213]. These clinical scenarios will benefit greatly from automated end-to-end lung segmentation and parenchymal disease quantification models. The implications of this work also exist beyond PH, given the good performance on the external test cohort. CTPA is routinely performed for the acute diagnosis of PE. Automated lung segmentation models can enable further quantitative vessel analysis

and clot burden estimation [214,215]. With appropriate transfer learning to further improve performance, this model may be used as a tool for both contrast and non-contrast imaging.

Limitations

The training set contains only patients with a known diagnosis of pulmonary hypertension, and thin-slice volumetric CTPA protocol scans, from a single tertiary centre with a predominantly white European population. The range of lung disease in this cohort has been previously reported, and whilst this represents a realistic heterogeneous clinical cohort of patients, there is a relative skew and bias to pathological cases due to a lack of truly 'normal' scans [178]. A limitation to using real world clinical data is a lack of 'rare' cases in the training data which can limit performance. Future work will seek to address the limitations of this study by developing and testing the DL model in a large cohort of multi-ethnicity patients. Despite good performance in the heterogeneous external test cohort, the intended clinical use of the developed model at this stage is limited to patients with suspected diagnosis of PH.

In conclusion, we developed a 3D nnU-net based model for lung segmentation in CTPA imaging that is highly accurate, clinically evaluated, and externally tested in well phenotyped patient cohorts with a spread of lung disease. The model is suited to important clinical scenarios of lung disease quantification in pulmonary hypertension.

6.6 Figures

6.6.1 Fig 1: STROBE flow diagram

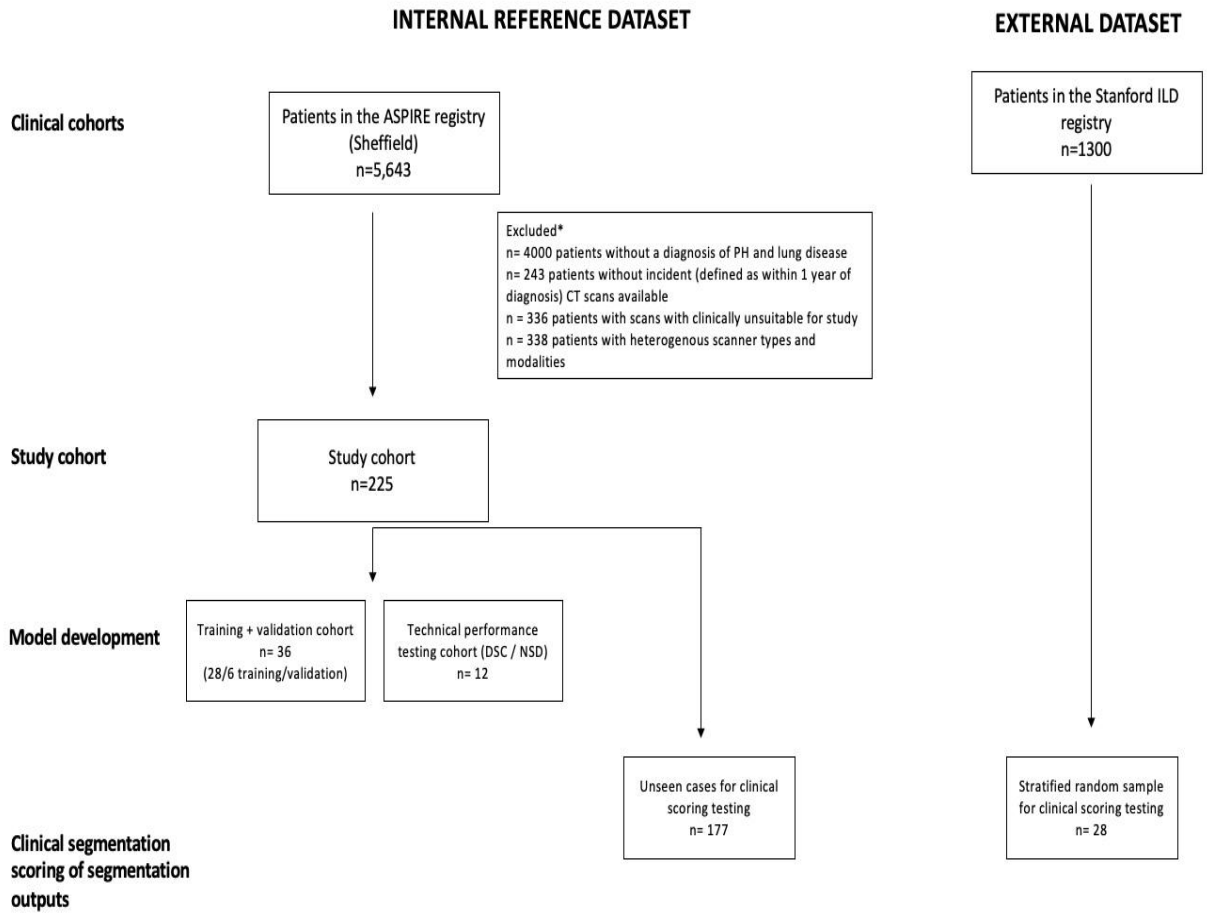


Figure 6.1 STROBE diagram showing patient selection in both internal reference datasets (Sheffield, ASPIRE cohort) and external datasets (Stanford ILD cohort).

Patient groups per study stage (model development and clinical segmentation scoring) are also shown.

6.6.2 Fig 2: Example of RadSeg scores

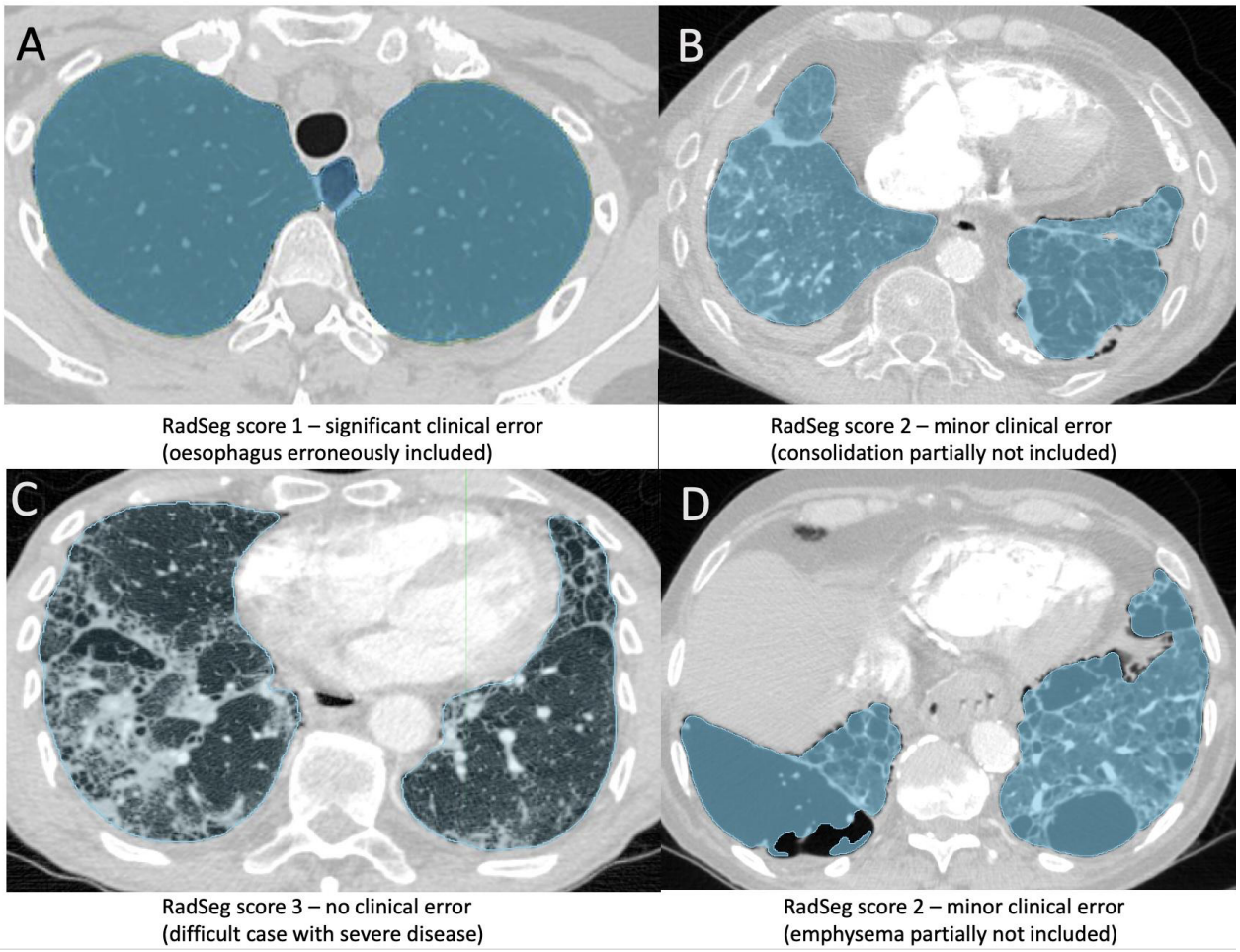


Figure 6.2 Examples of cases with corresponding RadSeg scores.

A (score 1) - example of significant clinical error (oesophagus erroneously included). B (score 2) - example of minor clinical error (consolidation at left lung base partially not included). C (score 3) - example of no clinical error in a difficult case with severe lung disease. D (score 2) - another example of minor clinical error (bullous emphysema partially not included)

6.7 Tables

6.7.1 Table 1: Patient characteristics

Characteristic	**Sheffield Reference Cohort N = 225[1]**	**Stanford External Cohort N = 28**
Age at scan	68 (60, 74)	62 (51, 74)
Female gender	106 (47%)	22 (79%)
Body Mass Index	27.2 (23.9, 31.8)	6 (21%)
WHO Functional Class		
2	30 (13%)	N/A
3	126 (56%)	N/A
4	67 (30%)	N/A
Baseline Intermittent Shuttle Walk Test distance (m)	90 (30, 182)	N/A
PFT - FVC % predicted	83 (67, 105)	64 (52, 76)
PFT - FEV1 % predicted	70 (49, 89)	68 (52, 81)
PFT - FEV1 / FVC Ratio	68 (52, 76)	79 (75, 84)
PFT - DLCO % predicted	29 (22, 47)	63 (58, 68)
Smoker (ever)		
Never	26 (34%)	15 (54%)
Ever	50 (66%)	13 (46%)
Pack years	24 (20, 40)	N/A
RHC - Mean Right Atrial Pressure (mmHg)	9.0 (6.0, 14.0)	N/A
RHC - Mean Pulmonary Arterial Pressure (mPAP)	45 (35, 53)	N/A
RHC - Pulmonary Arterial Wedge Pressure (mmHg)	11.0 (9.0, 14.0)	N/A

RHC - Cardiac index (L/min x m^-2)	2.58 (2.05, 3.29)	N/A
RHC - Pulmonary Vascular Resistance (Wood Units)	7.1 (4.1, 10.7)	N/A
RHC - Mixed venous oxygen saturation SvO2 %	65 (59, 70)	N/A
[1] Median (IQR); n (%)		

Table 6.1 Patient characteristics for both internal reference and external cohorts

Abbreviations used – WHO – World Health Organisation, PFT – Pulmonary Function Test, FEV1 – Forced Expiratory Volume in 1 second, FVC – Forced Vital Capacity, DLco – Diffusing capacity of carbon monoxide, RHC – right heart catheterisation.

6.7.2 Table 2: Clinical evaluation of radiological segmentation (RedSeg) results

	Sheffield Reference Cohort N = 177[1]	**Stanford External Cohort N = 28[1]**
RadSeg score		
0 (failed to output a segmentation)	0	0
1 (segmented with clinically significant error)	0	1 (3.6%)
2 (segmented with minor clinically insignificant error)	31 (18%)	7 (25%)
3 (segmented without any clinical significant errors)	146 (82%)	20 (71%)
Failure analysis - reasons for RadSeg score 1 or 2		
Consolidation	9 (29%)	0
Atelectasis	7 (23%)	3 (38%)

Pleural effusion	6 (19%)	2 (25%)
Bullous emphysema	3 (9.7%)	0
Apical scarring	2 (6.5%)	0
Lung mass/nodule	2 (6.5%)	0
Fibrosis	0	2 (25%)
Azygous fissure (thickened)	1 (3.2%)	0
Collapsed lobe	1 (3.2%)	0
Patulous dilated oesophagus		1 (12%)
[1] n (%)		

Table 6.2 Clinical evaluation of radiological segmentation (RedSeg) results and failure analysis for suboptimal performance in each cohort.

6.8 Supplemental Material / Appendix

6.8.1 DICE and accuracy scores for each patient in performance testing.

Case	**Accuracy**	**Dice Similarity Coefficient (DSC)**	**Normalised Surface Distance (NSD)**
1	0.999	0.995	0.988
2	0.995	0.967	0.941
3	0.999	0.996	0.994
4	0.999	0.996	0.994
5	0.998	0.993	0.990
6	0.998	0.987	0.991
7	0.999	0.995	0.991
8	0.997	0.974	0.930
9	0.998	0.992	0.992
10	0.999	0.995	0.989
11	0.999	0.995	0.999
12	0.999	0.996	0.996
Mean (95% Confidence Interval)	**0.998 (95% CI 0.9976 to 0.9989)**	**0.990 (95% CI 0.9840 to 0.9962)**	**0.983 (95% CI 0.9686 to 0.9972)**

Table 6.3 DICE and accuracy scores against validation cases.

6.8.2 DICOM information for each cohort

Characteristic	**Sheffield Reference Cohort N = 225[1]**	**Stanford External Cohort N = 28**
Scanner manufacturer		
GE MEDICAL SYSTEMS	225 (100%)	6 (21%)
SIEMENS		21 (75%)
TOSHIBA		1 (4%)
Scanner model name		
LightSpeed Pro 32	64 (26%)	0 (0%)
LightSpeed VCT	161 (72%)	4 (14%)
SOMATOM Definition AS+		7 (25%)
SOMATOM Definition Edge		4 (14%)
SOMATOM Definition Flash		2 (7%)
SOMATOM Force		3 (11%)
Sensation 64		5 (18%)
Aquilion		1 (4%)
Scan slice thickness		
0.625	225 (100%)	1 (4%)
1.0		20 (71%)
1.25		5 (18%)
2.0		2 (7%)
kvp		
80	0 (0%)	1 (4%)

100	39 (17%)	4 (14%)
120	186 (83%)	23 (82%)
[1] n (%)		

Table 6.4 Scan DICOM (Digital Imaging and Communications in Medicine) information for each cohort.

7 Improved quantification and prognostication of lung disease on CT in pulmonary hypertension by combining the strengths of AI and radiologists: a retrospective multicentre validation study

Authors: **Krit Dwivedi** BMBS BMedSci (Hons) FRCR, Michael Sharkey MEng MSc, Liam Delaney MBChB (Hons) BMedSci (Hons), Samer Alabed MD FRCR, Smitha Rajaram FRCR, Catherine Hill Mb ChB MRCP FRCR, Christopher Johns MBBS PhD FRCR, Alex Rothman MA BM BCh MRes PhD MRCP EEGC, A A Roger Thompson MRCP PhD, Jim Wild PhD, Robin Condliffe BSc (Hons) MBChB MRCP MD, David G Kiely BSc (Hons) MD FRCP, Andrew J Swift MD, PhD, FRCR

Submitted to Radiology

Contribution: First author. Contributor Roles Taxonomy (CRediT) – conceptualisation, data curation, formal analyses, funding analysis, methodology, project administration, resources, software, validation, visualisation, writing – original draft, review, and editing).

I lead this study. I performed all the data analysis and independently wrote the original draft for this study. This was reviewed by the collaborators and supervisors DGK and KD. MS helped and led with writing the code for making the lung segmentation and classification model. All other authors had a small involvement with data gathering, support, project administration and resources.

7.1 Abstract

Background Gold standard for Computed Tomography (CT) lung disease quantification is visual scoring by specialist radiologists. Due to overlapping characteristics, distinguishing between Idiopathic Pulmonary Arterial Hypertension (IPAH) and PH secondary to Chronic Lung Disease (PH-CLD) is challenging in patients with 'mild' lung disease. This retrospective multicentre study deploys an AI model to quantify and establish the prognostic value of common lung parenchymal patterns.

Methods

521 consecutive patients with incident CT imaging between 2001-19 were included from the ASPIRE registry. AI quantified the percentage of normal lung, ground glass, ground glass with reticulation (GGR), emphysema, honeycombing and fibrosis. Fibrosis severity was scored by sub-specialist radiologists. Multivariate cox regression adjusting for age, sex, WHO function class, and diffusing capacity of carbon monoxide (DLCO) was performed. Findings were externally validated in 246 patients from 33 centres and 37 scanners.

Findings

All patterns were prognostic univariate predictors. GGR% (HR 1.02, p=0.015) and fibrosis% (HR 1.01, p=0.05) were prognostic multivariate predictors (Hazard ratio reflects 1% increase in lung involvement)[I]. 2% GGR and 4% fibrosis corresponded to 20% 1-year mortality. In the external cohort, these thresholds were multivariate prognostic predictors (2% GGR HR 1.74, p=0.011 and 4% fibrosis HR 1.85, p=0.004). In 300 patients scored by radiologists as having 'no' fibrosis, AI identified minor disease (1.2% GGR) which was prognostic (HR 1.03, p=0.006). Adding GGR to a predictive model of radiologically scored disease significantly improved the model (c-index 0.763 vs 0.742, p=0.038).

Conclusion

AI quantified GGR and fibrosis are prognostic markers for survival, independent of age, sex, WHO function class and DLCO. AI is sensitive to minor lung disease, and when used in

[I] This has been slightly modified from the original published manuscript to improve clarity.

combination with radiological reporting, provides additional predictive value. These findings have implications for patient phenotyping, radiological reporting, and patient management.

7.2 Research in context panel

7.2.1 Evidence before this study

Patients with evidence of significant precapillary pulmonary hypertension (PH) but with a 'mild' degree of parenchymal lung disease on Computational Tomography (CT) pose a diagnostic dilemma between two phenotypes - Idiopathic Pulmonary Arterial Hypertension (IPAH) and PH secondary to Chronic Lung Disease (PH-CLD). Current guidance assigns these patients as IPAH, but recent evidence has shown they have poorer outcomes compared to those with 'no' lung disease. There is a need for better characterising lung disease in these patients. Recent Artificial Intelligence (AI) advances have enabled truly quantitative approaches to CT which can directly quantify individual imaging features.

PubMed was searched on 13 December 2022 using the terms "pulmonary hypertension" AND "artificial intelligence" AND "lung disease", with no date or language restrictions. Variations of those search themes were combined with the OR operator. The search retrieved 38 results. These included papers that investigated the use of machine learning in PH in other diagnostic modalities, such as ECG, cardiac MR, or blood profiling. There were two reviews on the current use of machine learning in PH, and a further review of the imaging modalities used in PH. One study used a composite of imaging, clinical, lab and haemodynamic variables to identify additional disease phenotypes in PH.

Of the three directly related studies, two used machine learning methods on CT angiography to assess coronary and vascular features respectively, in patients with PH. The final paper used a CT based mathematical model to distinguish Chronic Thromboembolic Pulmonary Hypertension (CTEPH) from pulmonary embolism. No studies applied quantitative chest CT models that assessed parenchymal lung disease.

7.2.2 Added value of this study

Current lung disease severity assessment on CT is limited to a subjective visual score for an overall disease (eg, 'none'/'mild'/'moderate'/'severe' for fibrosis). This study better characterises lung disease on CT as a percentage of involvement of individual radiological parenchymal patterns, and establishes their respective prognostic impact. Ground glass change with reticulation is an important prognostic marker that is currently overlooked. The study establishes the added value of this combined quantitative AI approach when used in

adjunct to routine reporting by a specialist radiologist. Importantly, we externally validate our findings with imaging from 37 centres, performed on 33 scanners.

7.2.3 Implications of all the available evidence

Better characterisation of lung disease as a percentage of lung involvement of individual radiological patterns will help in improving patient phenotyping. Lung disease severity is important to clinical outcomes, referral decisions and management. This automated AI approach identifies prognostic imaging biomarkers, which can be combined with clinical, genomic and multimodality approaches to identify or discover new clusters and sub-phenotypes of disease.

7.3 List of abbreviations

CT	Computed Tomography
CTPA	Computed Tomography Pulmonary Angiogram
PH	Pulmonary Hypertension
PAH	Pulmonary Arterial Hypertension
IPAH	Idiopathic Pulmonary Arterial Hypertension
LOESS	LOcally Estimated Scatterplot Smoothing
ILD	Interstitial Lung Disease
DSC	Dice Score Coefficient
NSD	Normalised Surface Distance
HRCT	High Resolution Computed Tomography
AI	Artificial Intelligence
GDPR	General Data Protection Regulation
ASPIRE	Assessing the Severity of Pulmonary Hypertension In a Pulmonary Hypertension REferral Centre
PH-CLD	Pulmonary Hypertension associated with Chronic Lung Disease
HU	Hounsfield units
WHO	World Health Organisation
PFT	Pulmonary Function Test
FEV1	Forced Expiratory Volume in 1 second
FVC	Forced Vital Capacity
DL_{CO}	Diffusing capacity of carbon monoxide
RHC	Right Heart Catheterisation

7.4 Introduction

Lung disease severity in Pulmonary Hypertension (PH) is an important factor to accurately phenotype patients between Group 1 Pulmonary Arterial Hypertension (PAH) and Group 3 Pulmonary Hypertension associated with Chronic Lung Disease (PH-CLD). In practice, distinguishing between these two groups is challenging due to overlapping clinical characteristics, particularly in the most common form of PAH, Idiopathic Pulmonary Arterial Hypertension (IPAH)[42,47,213]. Patients presenting with 'mild' to 'moderate' lung disease pose a clinical diagnostic dilemma of substantial diagnostic and prognostic importance, as only patients diagnosed as PAH are eligible for novel targeted therapies which improve survival [3]. Current guidance recommends patients with mild lung disease be classified as IPAH [3]. However, recent studies have demonstrated that these patients have poorer outcomes, and likely represent a new distinct phenotype [36,41,213]. Chest Computed Tomography (CT) imaging is routinely performed for diagnosis but is not currently used for prognostication [30].

The gold-standard for lung parenchymal disease assessment is a visual assessment by a specialist radiologist. However, this is not reproducible and there exists significant inter and intra observer viability even amongst specialist radiologists[84,86]. Quantification is also limited to at best a subjective semi-quantitative score, such as none/mild/moderate/severe. Advances in artificial intelligence (AI) deep-learning approaches have made possible truly quantitative analysis of imaging features [49]. AI models can automatically identify the parenchyma, classify regions of disease, and provide metrics such as volume and percentage of lung involvement of common radiological parenchymal patterns [49,80,88].

This retrospective study deploys an automated end-to-end CT AI model that quantifies the extent of lung disease on a large consecutive clinical cohort of patients with a diagnosis of IPAH and PH-CLD. The principal hypothesis is that AI quantified percentage of individual radiological parenchymal patterns is prognostic, independent from and of added value beyond current established methods of lung disease severity assessment - pulmonary function testing (DLCO) and semi-quantitative radiological assessment. The primary aim of the study is to test this hypothesis with imaging performed across multiple centres and scanner types. Secondary aims are to derive thresholds for any significant lung disease features and externally validate their added prognostic value in predicting patient survival.

7.5 Methods

7.5.1 Study population

All incident patients with a primary assigned diagnosis of IPAH or PH-CLD were identified from the ASPIRE (Assessing the Severity of Pulmonary Hypertension In a Pulmonary Hypertension REferral Centre) registry. The combined cohort, irrespective of original assigned diagnosis, was chosen to minimise misclassification, selection or misdiagnosis bias. The registry contains comprehensive multimodality clinical and imaging data from patients across the United Kingdom, with a referral population 15-20 million patients[213]. Inclusion criteria was CT imaging performed at time of diagnosis, which was appropriate for AI quantitative analysis – a thin slice chest CT pulmonary angiogram (CTPA) without severe respiratory motion artefact. Exclusion criteria were patients with PH-CLD associated with conditions other than COPD and/or emphysema or ILD, and those with two or more radiological features of possible pulmonary veno-occlusive disease[167]. These eligibility criteria were chosen to represent a consecutive patient cohort with a spectrum of lung disease and PH aetiology. The cohort was divided into an internal derivation and external validation cohort based on location of their imaging. Mortality data were obtained from systems linked to the National Health Service Personal Demographics Service (PDS), which is updated when a death is registered in the UK. Patients who emigrated were excluded, as were patients without a record on the PDS. Patients undergoing lung transplantation were censored at the time of surgery, and mortality data were collected using a census date of May 31, 2019.

7.5.2 CT image acquisition, analysis and quantitative AI model

CT scans were performed in multiple centres on scanners from all major manufacturers with varying technical parameters. All scans were reported by experienced sub-speciality chest radiologists, who scored fibrosis semi-quantitatively as none/mild/moderate/severe as previously described [178,213]. The five parenchymal patterns, classified in accordance with the Fleishner Society glossary of terms, are ‘normal lung’, ‘ground glass’(GG), ‘ground glass with reticulation’ (GGR), ‘emphysema’ and ‘honeycombing [216]. ‘Fibrosis’ is a composite variable formed by summating regions of GG, GGR and honeycombing to reflect routine radiological reporting practice in these patients.

A supervised quantitative CT AI classification model was developed, as previously described [217]. In brief, the model takes a CTPA scan as an input and provides the respective percentage of lung volume involvement of each pattern as an output. Expert labelling of these patterns was performed by two qualified radiologists, with 12 and 2 years of specialist experience in cardiothoracic imaging. Lung segmentation is performed using a trained CTPA nnU-net based model[202]. Parenchyma classification is performed using a patch-based sliding window approach with a trained DenseNet-121. The parenchyma is divided into multiple overlapping patches, which are classified, then reconstructed to form a classified mask. Parameters hypertuned and tested for optimal DenseNet-121 performance were patch size, dimensionality, selection radius, minimum fill factor, and number of patches. The optimum hyperparameters were found to be 2.5D patches, pixel size of 64, selection radius of 3mm, minimum fill factor of 0.625 and the maximum number of patches. Post-processing was performed on the 'Emphysema' class using a patch threshold of 925 Hounsfield Units (HU) and fill factor of >5%.

7.5.3 Statistics

All analysis is performed at the patient level with one single corresponding incident CTPA. This is a retrospective post-hoc analysis of routinely collected prospective data. Analysis was performed with R version 4.1.2. Categorical data are presented as number and percentage, continuous data as median and IQR. To compare the derivation and validation cohorts, two-sample Welch t-tests or Wilcoxon rank sum tests were used for continuous data. Categorical data were compared by Pearson's $\chi 2$ test or by Fisher's exact test. Correlations with pulmonary function tests and association with radiologically reported disease severity was performed on a full cohort level.

Cox proportional hazard's regression was used to determine association between different CT parenchymal features and survival. Hazard ratios are presented with 95% confidence intervals and p value of <0.05 was considered significant. Unadjusted univariate and three separate adjusted multivariate models incorporating known prognostic variables were created for each AI CT feature. Model 1 adjusted for patient demographics – age, sex and WHO functional class; model 2 adjusted for all demographics and additionally haemodynamic disease severity by including pulmonary vascular resistance (PVR); and model 3 adjusted for all demographics and DLCO. In the internal derivation cohort, thresholds for 20% 1-year

mortality were derived for features that were independent predictors after adjusting for age, sex, WHO FC and DLCO (model 3), from locally estimated scatterplot smoothing (LOESS) curves. Patients in the external validation cohort were stratified by the derived thresholds. All three multivariate models were again analysed for the external validation cohort. Age, PVR and DLCO was stratified by thresholds of 50 years, 5 Wood Units and 45% DLCO, as per prior studies. Survival estimates from the time of enrolment were done by Kaplan-Meier analyses and log-rank test.

In a subgroup of patients scored as 'no disease' for fibrosis by radiologists, cox proportional hazard's regression was used to determine association between the percentage of lung involvement of disease features and survival. To assess whether CT AI features provide additional prognostic value, three different multivariate models were tested in the external derivation cohort. Model A included demographics alone – age, sex, and WHO function class. Model B included demographics and additionally the radiological scoring of none/mild/moderate/severe. Model C included demographics, radiological scoring and additionally prognostic AI metrics. The model c-index and Akaike information criterion (AIC) were measured to assess the relative prognostic strength of the model. A better model fit is indicated by a higher c-index, which represents the rate of correct survival predictions, and a lower AIC, which represents rate of prediction errors. In addition, the likelihood ratio test was performed to assess if there is a statistically significant difference between models.

7.5.4 Ethics

Ethical approval was granted by the Institutional Review Board and approved by the National Research Ethics Service (16/YH/0352). All research was conducted in agreement with the Declaration of Helsinki and the European General Data Protection Regulation.

7.5.5 Role of the funding sources

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

7.6 Results

7.6.1 Patient and cohort characteristics

From 5643 patients in the ASPIRE registry, 521 patients between February 2001 and January 2019 met the eligibility criteria and were included for analysis. 275 patients had CT imaging in our centre on four different scanners and formed the internal derivation cohort. 246 with CT imaging performed in 37 centres on 33 different scanners formed the external validation cohort. Patient characteristics are described in Table 1 and scanner type information in supplementary table s1a. Data availability for each variable is provided in appendix table s4a.

The combined cohort had a representative wide range of lung disease on pulmonary function testing (Interquartile range, IQR, 22-49 for DLCO, 72-107 for FVC percent predicted and 55-90 for FEV1 percent predicted) and haemodynamic disease severity on right heart catheterisation (PVR IQR 5.0-11.6). There was a 50% even split between IPAH and PH-CLD. Patients in the external validation cohort were younger, more female, and had more severe haemodynamic disease diagnosed as IPAH. They had less severe lung disease quantified by FEV1 and FVC, but there was no significant difference in DLCO between groups. There was no significant difference in the extent of AI quantified normal lung, GG, GGR, honeycombing or fibrosis. The internal cohort had slightly higher extent of emphysema (8% IQR 1-27% vs 3% IQR 0-20%) compared to the external cohort.

7.6.2 Correlations with existing methods of quantifying lung disease severity

Percentage of normal had moderate positive correlation with DLCO (R = 0.56). Percentage of fibrosis and GG had moderate negative correlation with DLCO (R = +0.41 and +0.46 respectively). Percentage of GG, emphysema and honeycombing had a weak negative correlation with DLCO (R = -0.32, -0.35, -0.34 respectively). These and further correlations for FEV1 and FVC percent predicted are presented in Appendix table S2a.

AI quantified emphysema and fibrosis corresponded well with radiological assessment. There were significant between-group difference in percentage of emphysema and fibrosis when stratified by their respective radiological severity (1%, 4%, 23% and 57% for none, mild,

moderate, and severe emphysema, p<0.001 and 4%, 9%, 31% and 54% for none, mild, moderate and severe fibrosis, p<0.001). Appendix table S3a describes this and further significant between group differences (p<0.001 for all) for GG, GGR and honeycombing when grouped by radiologically reported fibrosis severity.

7.6.3 Survival analysis

7.6.3.1 Internal derivation cohort

Univariate and multivariate associations between each parenchymal pattern and mortality in the internal cohort are presented in table 2. Median follow-up duration was 1.96 (0.91, 4.35) years. The outcome of mortality occurred in 183 (67%) patients. Two (0.7%) underwent transplantation. All lung disease features were significant (p<0.001) on univariate analysis, after adjusting for demographics (multivariate model 1 including age, sex and WHO FC) and additionally for PVR (p<0.001, model 2 including age, sex, WHO FC and PVR). In model 3, after adjusting for demographics and DLCO, percentage of normal (HR 0.99, 95% CI 0.98-1.00, p=0.003), GGR (HR 1.02, 95% CI 1.00-1.04, p=0.015) and fibrosis (HR 1.01, 95% CI 1.00-1.01, p=0.050) remained significant prognostic predictors for survival. Percentage of normal, GGR and fibrosis were established as significant independent predictors for mortality, and LOESS curves for each variable are presented in figure 3. Derived thresholds for 20% 1-year mortality were 61% for normal lung (39% abnormal lung), 3% for GGR and 8% for fibrosis.

7.6.3.2 External validation cohort

Patients were stratified by the above identified thresholds for percent normal, GGR and fibrosis in the external validation cohort. Univariate and multivariate associations between each parenchymal pattern and mortality in the internal cohort are presented in table 3. Median follow-up duration was 1.95 (0.80, 4.07) years. The outcome of mortality occurred in 118 (48%) patients. Two (0.8%) underwent transplantation. All three patterns were significant (p<0.001) on univariate analysis and after adjusting for demographics (p<0.001, multivariate model 1 including age, sex and WHO FC) and additionally for PVR (p<0.001, model 2 including age, sex, WHO FC and PVR). In model 3, after adjusting for demographics and DLCO, GGR (HR 1.74, 95% CI 1.12-2.68, p=0.011) and fibrosis (HR 1.85, 95% CI 1.22-2.80, p=0.004) remained significant prognostic predictors for survival.

Kaplan-Meier survival analysis demonstrated that patients with lung disease above the identified thresholds had significantly (log rank p<0.0001, figure 4) poorer survival. When grouped by the threshold of 2% GGR, the estimated survival rates were 71% vs 89% for 1 year, 38% vs 68% for 3-year and 24% vs 60% for 5-year survival.

7.6.4 Additional prognostic value of CT AI features

Of 444 patients with radiological reports available, 300 (67.6%) were reported as having no fibrosis. On quantitative analysis, these patients had a minor degree of ground glass (2% (0,5)), GGR (1.2% (0.3,4.6)), honeycombing (0.47% (0.24,0.96)) and fibrosis (4% (1,11)). Despite the small volumes of disease, GGR (HR 1.03, 95% CI 1.01-1.06, p=0.006), honeycombing (HR 1.13, 95% CI 1.07-1.19) and fibrosis (HR 1.01, 95% CI 1.00-1.02, p=0.019) were significant univariate predictors for mortality in this sub-group (table 4).

In the external validation cohort, Model A incorporating demographics alone (age, sex, and WHO FC) had a c-index of 0.721 and AIC of 1090 (table 5). Model B which additionally incorporated radiological semi-quantitative fibrosis severity scoring had a significantly (p=0.036) better predictive value than Model A with a c-index of 0.742 and AIC of 604. Model C which additionally incorporated percentage of GGR on top of radiological severity scoring and demographics had significant (p=0.038) additional predictive value compared to Model B, with a c-index of 0.763 and AIC of 601.

7.7 Discussion

The key finding of this study is that AI quantified lung parenchymal patterns on CT are prognostic for survival in patients with PH and lung disease. This is independent of demographics, PVR, DLCO, and externally validated across imaging from 37 centres. AI is sensitive to minor lung changes and when used in combination with radiological reporting, it provides additional predictive value. Of 38 prior studies in the domain of PH, AI and lung disease, this is the first study to deploy or clinically evaluate an AI quantitative CT model for lung disease assessment in PH.

The motivation of this work was to use AI on a routine investigation (CT) to help solve an important clinical problem. Patients with significant precapillary PH, unremarkable spirometry, no thromboembolic disease, and absence of other risk factors but with a 'mild' degree of parenchymal lung disease are a significant population that cannot currently confidently be phenotyped. Accurate phenotyping is vital as it informs treatment and management. Current guidelines recommend these patients be classified as IPAH [26]. However, these patients have been shown to have poorer survival and treatment response compared to those with truly 'classical' IPAH with no lung disease [41]. A recent seminal study by Hoeper et al characterises this problem and identifies a new phenotype of 'IPAH with a lung phenotype' which lies in between the spectrum of Group 1 IPAH and Group 3 PH-CLD [213]. There is a clear clinical need to better characterise and quantify lung disease in these patients, to inform their phenotyping. All existing work uses either PFTs (DLCO) or semi-quantitative radiological scoring for lung disease assessment. PFTs show significant variability, low reproducibility, and can be insensitive to change in disease progression [56–58,113],[J]. Radiological scoring is limited at subjective scoring, and has significant inter- and intra-observer variability even amongst specialist radiologists.[84,86]. This is typically provided as a categorical score (none/mild/moderate/severe). The AI approach in this study provides a continuous percentage of disease at an individual lung parenchymal pattern (ground glass with reticulation) level. There exist a few continuous visual scoring scales in which radiologists quantify to the nearest 5% lobar disease, but these have been limited to the

[J] Please refer to the background section for more detail on the strengths and limitations of pulmonary function tests.

research domain [113]. Visually quantifying each lung parenchymal pattern for each lobe is time consuming and impractical in routine clinical practice.

Independence from PVR and DLCO is important as it demonstrates the strong prognostic predictive strength of our findings. The PVR threshold of 5 WU was chosen based on recent studies that demonstrate its significance in predicting mortality across both IPAH and PH-CLD[51,163,164]. The DLCO threshold of 45% was derived from previous studies that determined the strong prognostic value of this threshold [41,53,166,168 168,169]. DLCO is often the only biomarker used for lung disease severity across multiple studies and is part of the REVEAL 2.0 risk score for PAH risk-stratification [8,36,41,213]. CT AI quantified lung disease metrics correlate well with DLCO and radiological scoring across multiple scanners and centres, which adds clinical confidence to the AI outputs. In the internal cohort, percentage of normal, GGR and fibrosis were significant predictors after adjusting for DLCO. In the external cohort, GGR and fibrosis remained significant after adjusting for DLCO, and there were significant between group differences in survival between patients stratified by our derived thresholds (figure 4). We postulate this may be because DLCO is a functional marker of gas exchange, whereas CT features are a more gross anatomical marker of architectural distortion. As both CT and DLCO are recommended by guidelines and performed in routine practice, combining the respective strengths of both modalities could aid in prognostication.

The predictive strength of our model for mortality was significantly improved when we added AI derived GGR percentage on top of radiological lung disease severity reporting. In patients who were radiologically scored as having ‘no’ fibrosis, the AI picked up minor levels of lung disease, which were significant predictors for mortality. This may be because such minor disease is currently overlooked or not felt to be significant enough by radiologists to be classified as ‘mild disease’. In idiopathic pulmonary fibrosis, automated quantitative CT was found superior to visual scoring [88], and provided additional prognostic value over visual scoring [113]. Radiologists and AI systems each have respective strengths and limitations, and combining both together has shown improved overall performance in other settings such as breast cancer diagnosis [218]. Radiologists use their clinical expertise to assess multiple features simultaneously and assess their relative clinical significance. This transparent, explainable assessment is vital and central to image analysis. However, accurate volume level quantification of parenchymal features on 500+ slices is an impractical, unrealistic task for a radiologist in routine reporting. Similar to automated segmentation studies, which aid

radiological reporting, this study proposes the use of such AI CT approaches as an adjunct radiological reporting, and demonstrates the additional prognostic value of this combined approach.

The AI model and findings were externally validated with imaging across 33 scanners and 37 centres. Our study is substantially larger and more representative (521 patients, 246 with external validation) than all three prior studies which have deployed AI on chest CT imaging in PH. They had 125[219] and 42 patients [220], and 82 scans (patient numbers unreported)[221]. None performed external validation. External deployment of our model and clinical validity of our findings in a different PH centre or registry with differing population demographics is warranted future work which would help better evaluate generalisability.

We acknowledge limitations inherent in the retrospective setting, within chest CT imaging, and imaging AI studies. The study was designed to minimise misclassification and selection bias and be representative of a 'real-world' consecutive clinical cohort with a spectrum of lung disease in PH. Whilst all patients were assigned diagnosis in a specialist centre, we cannot fully exclude the possibility of some misclassification bias and missed cases, including those never identified nor referred in non-specialist settings. Variations in breath holding by patients during CT scan acquisition can cause respiratory motion artefacts. Severe artefact limits traditional visual assessment for radiologists and can lead to misinterpretation. This study excluded patients with severe artefact, and it is possible for the model to overcall severely artifactual regions as pathology. Objective 'gold-standard' level classification of lung parenchyma, agreeable by all readers, is fundamentally challenging due to variability between highly experienced radiologists. Certain patterns such as severe subpleural emphysema with interstitial thickening look visually similar to honeycombing and there is some overlap in these cases. The intended clinical use and deployment of the model is as an adjunct to reporting by a radiologist, who will be aware of such inherent limitations. All AI classification models have inherent, difficult to quantify, biases skewed by their development training set. To best account for these, we preferentially chose a highly heterogeneous external validation cohort to validate our findings. Our thresholds, although validated externally, might not be generalisable to all different sites, and may need to shift to adjust for different prevalence rates.

In conclusion, this study better characterises lung disease on CT using AI by both increasing the precision of assessment as a percentage of total lung involvement, and specificity, by quantifying individual radiological parenchymal features. These features provide additional prognostic value beyond current established methods of lung disease assessment in PH and can be used in combination with radiological reporting to improve prognostication. These findings have implications for patient phenotyping, which affect treatment and management decisions.

7.8 Figures

7.8.1 Fig 1 - STROBE diagram

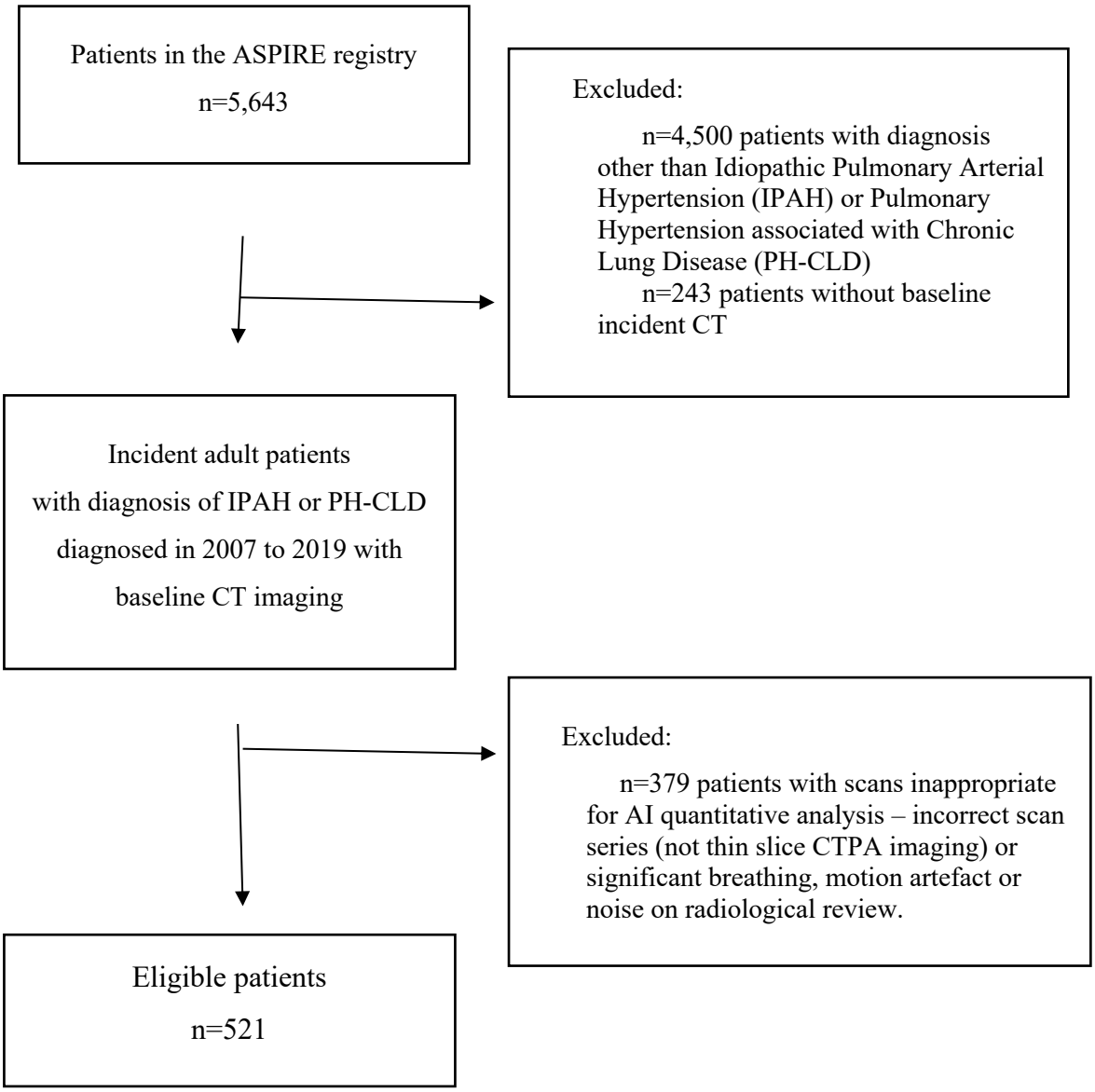


Figure 7.1 STROBE diagram showing patient selection.

Figure 1: STROBE diagram showing patient selection. Abbreviations used: ASPIRE (Assessing the Severity of Pulmonary Hypertension In a Pulmonary Hypertension REferral Centre) registry, IPAH (Idiopathic Pulmonary Arterial Hypertension), PH-CLD (Pulmonary Hypertension associated with Chronic Lung Disease), CT (Computed Tomography), CTPA (Computed Tomography Pulmonary Angiogram).

7.8.2 Fig 2 - AI output

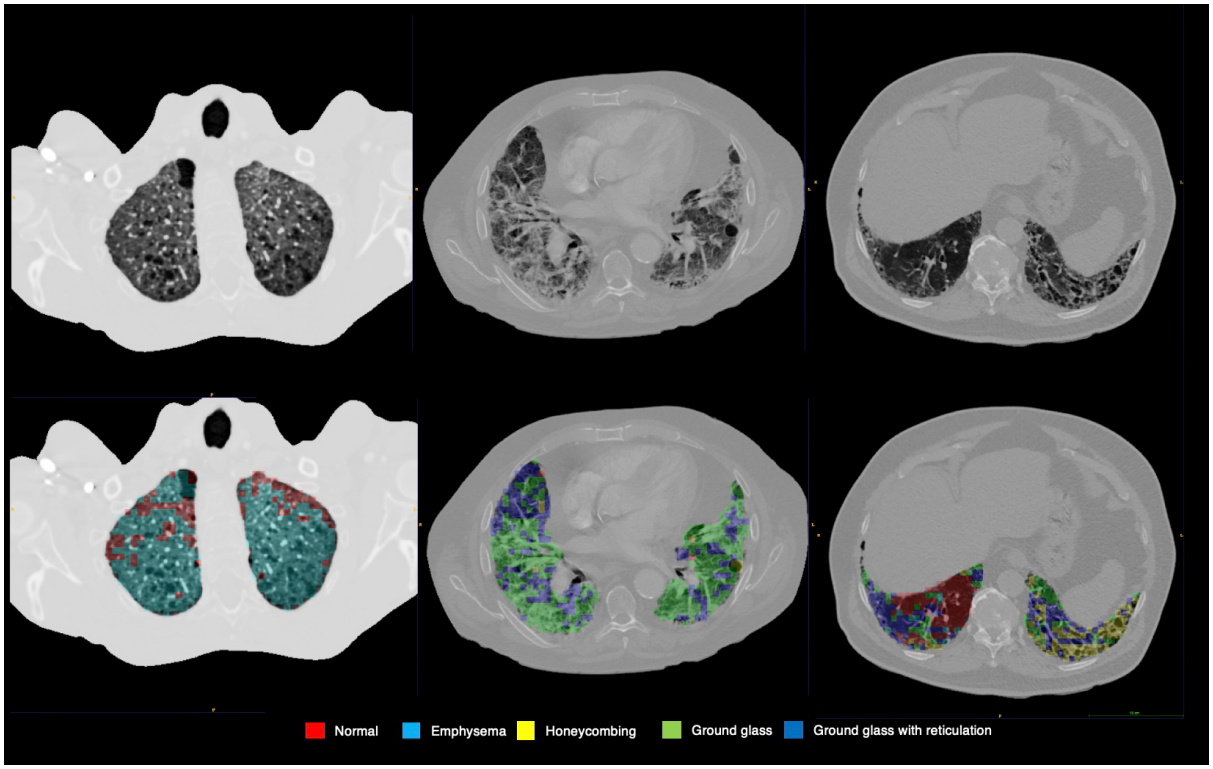


Figure 7.2: Example outputs from the AI model.

Figure 2: Example outputs from the AI model. Lung parenchyma is automatically segmented then classified into five lung parenchymal patterns - normal (red), emphysema (teal), honeycombing (yellow), ground glass (green) and ground glass with added reticulation (blue). This is done for every slice of the CT pulmonary angiogram.

7.8.3 Fig 3 - Kaplan Meier curves stratified by significant lung disease thresholds in external cohort

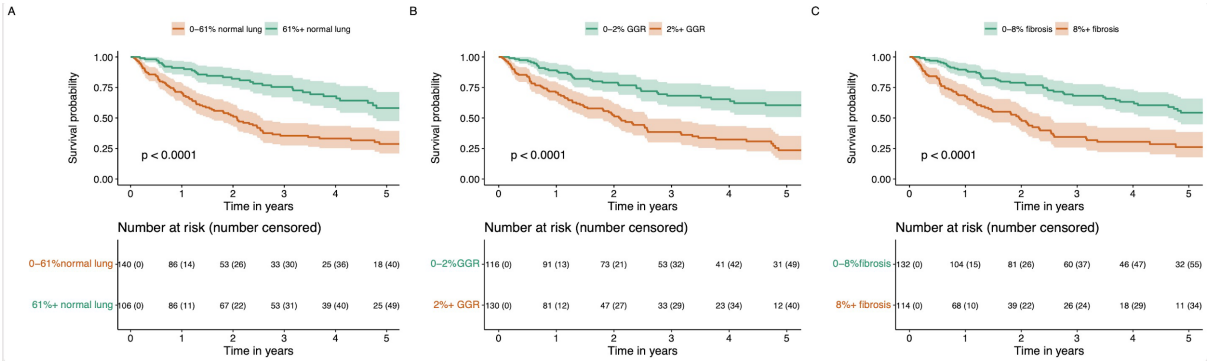


Figure 7.3 Kaplan-Meier survival estimates for patients in the external validation cohort (n=246) stratified into groups by thresholds for significant disease derived in the internal derivation cohort (n=275).

Figure 3: Kaplan-Meier survival estimates for patients in the external validation cohort (n=246) stratified into groups by thresholds for significant disease derived in the internal derivation cohort (n=275). Figure 4A represents percentage of normal lung (grouped by threshold of 61%), 4B percentage of ground glass with additional reticulation (GGR) (grouped by threshold of 2%), and 4C percentage of fibrosis (grouped by threshold of 8%). Log-rank p<0.0001 for all three between group comparisons.

7.9 Tables

7.9.1 Table 1 - demographics and clinical characteristics

	Full Cohort	**Cohorts**		
Characteristic	**N = 521[1]**	**Internal derivation, N = 275[1]**	**External validation, N = 246[1]**	**p-value[2]**
Age at diagnosis, years	67 (57, 74)	68 (60, 75)	65 (51, 72)	**<0.001**
Sex, female	270 (52%)	128 (47%)	142 (58%)	**0.011**
Body Mass Index	27.6 (24.0, 32.0)	27.5 (24.1, 32.4)	27.9 (23.9, 31.7)	0.8
WHO Function Class				0.4
II	67 (13%)	37 (14%)	30 (12%)	
III	309 (60%)	155 (57%)	154 (63%)	
IV	143 (28%)	81 (30%)	62 (25%)	
Pulmonary Function Tests				
FVC, percent predicted	90 (72, 107)	85 (67, 105)	93 (76, 107)	**0.007**
FEV1, percent predicted	75 (55, 90)	70 (51, 89)	78 (59, 91)	**0.014**
FEV1 / FVC ratio	68 (57, 77)	68 (55, 77)	69 (60, 78)	0.4
DLCO, percent predicted	31 (22, 49)	30 (22, 47)	32 (23, 53)	0.2
Haemodynamics				
RAP, mm Hg	9.0 (6.0, 13.0)	8.5 (6.0, 13.0)	10.0 (7.0, 13.0)	0.060
mPAP, mm Hg	46 (38, 54)	44 (35, 52)	48 (41, 56)	**<0.001**
PAWP, mm Hg	11.0 (9.0, 14.0)	11.0 (9.0, 14.0)	11.0 (8.0, 13.0)	**0.038**
Cardiac index, L/min/m2	2.41 (1.88, 3.02)	2.56 (1.95, 3.17)	2.29 (1.81, 2.78)	**<0.001**
PVR, wood units	8.0 (5.0, 11.6)	7.2 (4.2, 10.8)	9.4 (6.4, 13.2)	**<0.001**
SvO2, %	64 (58, 70)	65 (59, 70)	64 (56, 69)	0.061
PH phenotype				**<0.001**
Group 1 IPAH	259 (50%)	110 (40%)	149 (61%)	
Group 3 PH-CLD	262 (50%)	165 (60%)	97 (39%)	
AI Quantified CT parenchymal pattern (percentage of)				
Normal lung	59 (37, 74)	59 (35, 76)	58 (38, 71)	0.2
Ground glass	3 (1, 10)	3 (1, 9)	3 (1, 11)	0.6
Ground glass with additional reticulation (GGR)	3 (1, 11)	3 (1, 11)	2 (1, 10)	0.4
Honeycombing	0.72 (0.32, 1.57)	0.64 (0.28, 1.84)	0.77 (0.39, 1.36)	0.4
Fibrosis	7 (2, 24)	7 (2, 24)	6 (2, 23)	0.6

Emphysema	6 (0, 24)	8 (1, 27)	3 (0, 20)	**0.002**
[2]Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test				

Table 7.1 Patient characteristics for the full, international derivation and external validation cohorts, with group comparison p-values for between cohort comparisons.

Table 1: Patient characteristics for the full, international derivation and external validation cohorts, with group comparison p-values for between cohort comparisons. Categorical data are shown as n and (%) of the respective population. Continuous data are depicted as median [Q1, Q3]. Bold text represents significant between group difference.

Definition of abbreviations: BMI, body mass index; IPAH, idiopathic pulmonary arterial hypertension; PH, pulmonary hypertension; PH-CLD, Pulmonary Hypertension associated with Chronic Lung Disease; WHO FC, World Health Organization Functional Class; FVC, forced vital capacity; FEV1, forced expiratory volume in 1 s; DLCO, diffusion capacity of the lung for carbon monoxide; RA, right atrial pressure; mPAP, mean pulmonary arterial pressure; PAWP, pulmonary arterial wedge pressure; CI, cardiac index; PVR, pulmonary vascular resistance; SvO2, mixed-venous oxygen saturation; AI, artificial intelligence; GGR, ground glass with reticulation

7.9.2 Table 2 – Univariate and multivariate regression analysis for survival in internal cohort

	Univariate				Multivariate 1 (adjusting for demographics - age, gender, WHO FC)			Multivariate 2 (adjusting for demographics and PVR)			Multivariate 3 (adjusting for demographics and DLCO)		
AI Quantified CT feature (percentage of)	**N**	**HR[1]**	**95% CI[1]**	**p-value**	**HR[1]**	**95% CI[1]**	**p-value**	**HR[1]**	**95% CI[1]**	**p-value**	**HR[1]**	**95% CI[1]**	**p-value**
Normal lung	275	0.98	0.97, 0.98	**<0.001**	0.98	0.98, 0.99	**<0.001**	0.98	0.98, 0.99	**<0.001**	0.99	0.98, 1.00	**0.003**
Ground glass	275	1.02	1.01, 1.03	**<0.001**	1.02	1.00, 1.03	**0.017**	1.02	1.00, 1.03	**0.026**	1.01	0.99, 1.02	0.3
Ground glass with additional reticulation (GGR)	275	1.04	1.03, 1.06	**<0.001**	1.04	1.02, 1.05	**<0.001**	1.03	1.02, 1.05	**<0.001**	1.02	1.00, 1.04	**0.015**
Honeycombing	275	1.06	1.04, 1.09	**<0.001**	1.05	1.02, 1.07	**0.002**	1.04	1.02, 1.07	**0.003**	1.03	1.00, 1.06	0.082
Fibrosis	275	1.02	1.01, 1.02	**<0.001**	1.01	1.01, 1.02	**<0.001**	1.01	1.01, 1.02	**<0.001**	1.01	1.00, 1.01	**0.050**
Emphysema	275	1.01	1.01, 1.02	**<0.001**	1.01	1.00, 1.02	**0.012**	1.01	1.00, 1.02	**0.016**	1.01	1.00, 1.02	0.057

Table 7.2 Unadjusted univariate and adjusted multivariate cox proportional hazard's regression for each AI quantified CT feature in the internal derivation cohort (n=275 patients).

Table 2: Unadjusted univariate and adjusted multivariate cox proportional hazard's regression for each AI quantified CT feature in the internal derivation cohort (n=275 patients). Multivariate model 1 adjusted for patient demographics – age (grouped at 50 years), sex and WHO functional class; model 2 adjusted for all demographics and additionally haemodynamic disease severity by including pulmonary vascular

resistance (PVR, grouped at 5 WU); and model 3 adjusted for all demographics and DLCO (grouped at 45%). Bold text represents significant prediction.

7.9.3 Table 3 - Univariate and multivariate regression analysis for survival in external cohort

	Univariate				Multivariate 1 (adjusting for demographics - age, gender, WHO FC)			Multivariate 2 (adjusting for demographics and PVR)			Multivariate 3 (adjusting for demographics and DLCO)		
AI Quantified CT feature (grouped by significant derived threshold)	**N**	**HR[1]**	**95% CI[1]**	**p-value**	**HR[1]**	**95% CI[1]**	**p-value**	**HR[1]**	**95% CI[1]**	**p-value**	**HR[1]**	**95% CI[1]**	**p-value**
Normal lung	246	0.37	0.25, 0.55	**<0.001**	0.45	0.30, 0.67	**<0.001**	0.49	0.32, 0.76	**<0.001**	0.67	0.42, 1.06	0.082
Ground glass with additional reticulation (GGR)	246	2.78	1.88, 4.11	**<0.001**	2.00	1.34, 2.96	**<0.001**	2.13	1.40, 3.22	**<0.001**	1.74	1.12, 2.68	**0.011**
Fibrosis	246	2.71	1.86, 3.93	**<0.001**	2.07	1.41, 3.03	**<0.001**	2.03	1.36, 3.04	**<0.001**	1.85	1.22, 2.80	**0.004**

Table 7.3 Unadjusted univariate and adjusted multivariate cox proportional hazard's regression performed in the external validation cohort (n=246 patients) for each AI quantified CT feature grouped by thresholds for significant disease derived in the internal cohort

Table 3: Unadjusted univariate and adjusted multivariate cox proportional hazard's regression performed in the external validation cohort (n=246 patients) for each AI quantified CT feature grouped by thresholds for significant disease derived in the internal cohort. Multivariate model 1 adjusted for patient demographics – age (grouped at 50 years), sex and WHO functional class; model 2 adjusted for all demographics and additionally haemodynamic disease severity by including pulmonary vascular resistance (PVR, grouped at 5 WU); and model 3 adjusted for all demographics and DLCO (grouped at 45%). Bold text represents significant prediction.

7.9.4 Table 4 – Significance and extent of AI quantified disease in those patients radiologically scored as 'no fibrosis'

Characteristic	N	Median (IQR)	HR[1]	95% CI[1]	p-value
Percentage ground glass	300	2 (0,5)	1.01	0.99, 1.03	0.3
Percentage ground glass with added reticulation	300	1.2 (0.3,4.6)	1.03	1.01, 1.06	**0.006**
Percentage honeycombing	300	0.47 (0.24, 0.96)	1.13	1.07, 1.19	**<0.001**
Percentage fibrosis	300	4 (1,11)	1.01	1.00, 1.02	**0.019**
	[1] HR = Hazard Ratio, CI = Confidence Interval				

Table 7.4 For sub-group of patients scored by radiologists as having no fibrosis (n=300 patients), extent of and univariate cox proportional hazard's regression analysis for each significant AI quantified CT feature.

Table 4: For sub-group of patients scored by radiologists as having no fibrosis (n=300 patients), extent of and univariate cox proportional hazard's regression analysis for each significant AI quantified CT feature. Bold text represents significant prediction.

7.9.5 Table 5 - Additive value of AI metrics on radiological fibrosis severity scoring

Multivariate Cox proportional hazard's regression model for survival	**C-index**	**AIC**	**P value**
Model A – Demographics alone	0.721	1090	Reference standard
Model B -Demographics + Radiological severity scoring (none/mild/mod/severe)	0.742	604	**P=0.036** (compared to Model A - demographics alone)
Model C - Demographics + Radiological fibrosis severity scoring (none/mild/mod/severe) + AI radiological GGR severity quantification (continuous variable)	0.763	601	**P=0.038** (compared to Model B - demographics and radiological severity scoring)

Table 7.5 Comparison of predictive strength of three multivariate cox proportional hazard's regression models in the external validation cohort (n=246 patients).

Table 5: Comparison of predictive strength of three multivariate cox proportional hazard's regression models in the external validation cohort (n=246 patients). A higher c-index indicates a better model fit and a lower Akaike information criterion (AIC) indicates a relative lower prediction error. Bold text represents significant log-rank difference between models by likelihood ratio test. Adding AI quantified ground glass reticular severity to the multivariate cox proportional hazard's regression model for survival significantly improves the predictive model (c-index 0.763 vs 0.742, AIC 601 vs 604, log-rank p= 0.038)`

7.10 Supplementary material

7.10.1 Table S1a - Scanner types between cohorts

Characteristic	Internal, N = 275[1]	External, N = 246[1]
manufacturer		
GE MEDICAL SYSTEMS	223 (81%)	63 (26%)
SIEMENS	0 (0%)	113 (46%)
TOSHIBA	52 (19%)	58 (24%)
Philips	0 (0%)	12 (4.9%)
manufacturer_model_name		
LightSpeed VCT	160 (58%)	22 (8.9%)
LightSpeed Pro 32	63 (23%)	3 (1.2%)
Aquilion PRIME	37 (13%)	12 (4.9%)
Aquilion	0 (0%)	40 (16%)
Sensation 16	0 (0%)	29 (12%)
SOMATOM Definition AS+	0 (0%)	23 (9.3%)
Aquilion ONE	15 (5.5%)	5 (2.0%)
Optima CT660	0 (0%)	15 (6.1%)
SOMATOM Definition AS	0 (0%)	13 (5.3%)
SOMATOM Definition Flash	0 (0%)	11 (4.5%)
Sensation 64	0 (0%)	8 (3.3%)
LightSpeed16	0 (0%)	7 (2.8%)
Brilliance 64	0 (0%)	6 (2.4%)
Sensation 40	0 (0%)	5 (2.0%)
SOMATOM Definition	0 (0%)	5 (2.0%)
Definition AS	0 (0%)	4 (1.6%)
Definition AS+	0 (0%)	4 (1.6%)
Revolution EVO	0 (0%)	4 (1.6%)
Definition	0 (0%)	3 (1.2%)
Ingenuity Core 128	0 (0%)	3 (1.2%)
LightSpeed Plus	0 (0%)	3 (1.2%)
LightSpeed Ultra	0 (0%)	3 (1.2%)
SOMATOM Definition Edge	0 (0%)	3 (1.2%)
Discovery CT750 HD	0 (0%)	2 (0.8%)
Ingenuity CT	0 (0%)	2 (0.8%)
LightSpeed Pro 16	0 (0%)	2 (0.8%)
Sensation Cardiac	0 (0%)	2 (0.8%)
Aquilion Prime SP	0 (0%)	1 (0.4%)
BrightSpeed	0 (0%)	1 (0.4%)
iCT 256	0 (0%)	1 (0.4%)
Optima CT540	0 (0%)	1 (0.4%)

Perspective	0 (0%)	1 (0.4%)
Sensation 16 with Akron Q tube	0 (0%)	1 (0.4%)
Volume Zoom	0 (0%)	1 (0.4%)
[1]n (%)		

Table 7.6 Scanner types between internal derivation and external validation cohorts.

Appendix Table S1a: Scanner types between internal derivation and external validation cohorts.

7.10.2 Association between AI CT disease and radiological reporting

AI quantified CT feature / Reported fibrosis severity	**None, N = 300[1]**	**Mild, N = 51[1]**	**Moderate, N = 36[1]**	**Severe, N = 39[1]**	**p-value[2]**
Ground glass	2 (0, 5)	4 (1, 10)	10 (5, 22)	16 (6, 30)	**<0.001**
Ground glass with additional reticulation	1 (0, 5)	4 (1, 11)	15 (9, 25)	26 (12, 34)	**<0.001**
Fibrosis	4 (1, 11)	9 (3, 26)	31 (19, 43)	54 (26, 69)	**<0.001**
Honeycombing	0.47 (0.24, 0.96)	0.69 (0.30, 1.25)	2.35 (1.06, 4.98)	4.20 (1.75, 12.27)	**<0.001**
Reported emphysema severity	**None, N = 237[1]**	**Mild, N = [1]**	**Moderate, N = 36[1]**	**Severe, N = 39[1]**	**p-value[2]**
Emphysema	1 (0, 7)	4 (0, 13)	23 (11, 37)	57 (33, 71)	**<0.001**

Table 7.7 Association between AI quantified CT features and radiologically reported semi-quantitative disease severity

Appendix Table S2a: Association between AI quantified CT features and radiologically reported semi-quantitative disease severity. Data presented as median (IQR). Kruskal-Wallis rank sum test used to compare groups.

7.10.3 Available data for each variable

Characteristic	**Internal, N = 275**	**External, N = 246**
Age at diagnosis, years	275 (100%)	246 (100%)

Sex, female	275 (100%)	246 (100%)
Body Mass Index	245 (89%)	226 (92%)
WHO Function Class	273 (99%)	246 (100%)
FVC, percent predicted	255 (93%)	238 (97%)
FEV1, percent predicted	256 (93%)	239 (97%)
FEV1 / FVC ratio	253 (92%)	239 (97%)
DLCO, percent predicted	234 (85%)	216 (88%)
RAP, mm Hg	270 (98%)	226 (92%)
mPAP, mm Hg	273 (99%)	241 (98%)
PAWP, mm Hg	265 (96%)	223 (91%)
Cardiac index, L/min/m2	262 (95%)	215 (87%)
PVR, wood units	265 (96%)	221 (90%)
SvO2, %	248 (90%)	209 (85%)
PH phenotype	275 (100%)	246 (100%)
Treatment	274 (100%)	231 (94%)

Table 7.8 Available data for each variable.

Table S3a: Available data for each variable. Data are shown as n and (%) of the respective population.

7.10.4 Fig S1a - LOESS curves

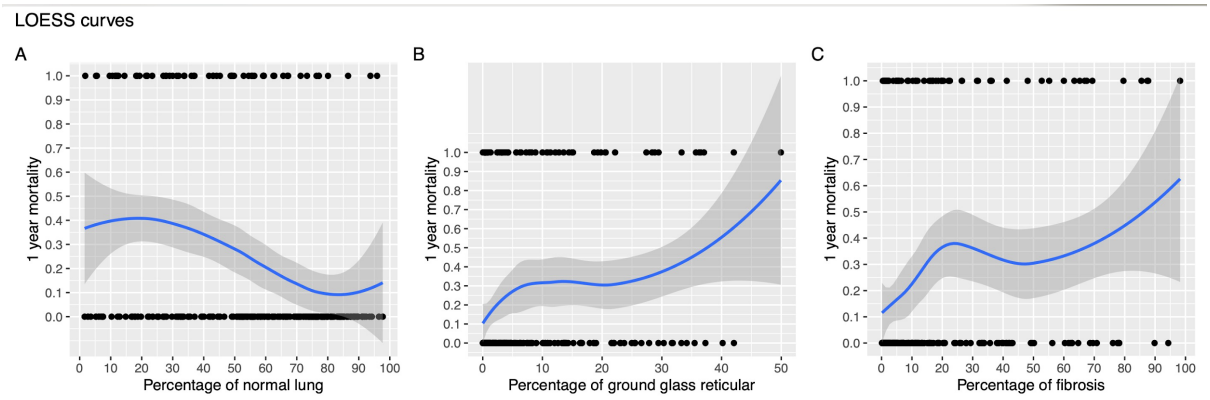


Figure 7.4 LOESS (LOcally Estimated Scatterplot Smoothing) curves

Figure s1a: LOESS (LOcally Estimated Scatterplot Smoothing) curves with percentage of 1 year mortality on the y-axis and percentage of AI quantified disease feature on the x-axis. Analysis performed in the internal derivation cohort (n= 275 patients). Figure 3A represents percentage of normal lung, 3B percentage of ground glass with additional reticulation (ground glass reticular) and 3C percentage of fibrosis.

8 Discussion

** This is a brief, overarching discussion of the work. Please also refer to individual chapters for additional dedicated discussion for each aspect of the thesis.*

This thesis establishes CT lung parenchymal patterns as prognostic biomarkers in patients with precapillary pulmonary hypertension (IPAH and PH-CLD). It further develops and clinically validates a deep-learning automated AI model which quantifies the extent of these parenchymal patterns, and their respective prognostic strength. Both are novel contributions to the fields of medical imaging and respiratory medicine.

Chapter two is a state-of-the-art literature review (published in *Diagnostics*) performed at the outset of the PhD, which establishes the context in which this thesis and its work is performed. AI has recently enabled a quantitative, data-first approach to medical imaging and made automated end-to-end approaches to image analysis possible. Within Chest CT imaging, this thesis is the first body of work using AI for lung parenchymal analysis in PH. The review outlined the promises and challenges associated with this approach.

Two significant contributions of the thesis, beyond the published chapters and scientific discoveries, is the creation of the large clinico-radiological database (detailed in chapter three), and workflows for AI development within the research group. The experience of training and building an in-house end-to-end automated AI model from routine NHS imaging data I feel was particularly instrumental in my development as a clinical academic radiologist. This required collaboration between multiple domain experts – radiologists, clinicians, clinical scientists, physicists, and computer scientists. All members of Sheffield Cardiothoracic Imaging Research group can use the database and AI pipeline on other research questions going forward.

Chapter four (published in *European Respiratory Journal Open Access*) achieves the primary aim of the thesis and establishes the clinical prognostic significance of CT lung parenchymal patterns in patients with PH and lung disease. I wrote code within R to automatically search CT reports using a regex string search algorithm, but I manually validated and checked each report. The approach here was semi-quantitative, as CT reports and subjective visual scores were used

to analyse the scans within the database. This is limited compared to an automated AI approach but has the advantage of being widely and more immediately clinically applicable. The work won the prestigious Radiological Society of North America (RSNA) Trainee Prize in Chest imaging. The take home message for radiologists was the importance of quantifying emphysema and fibrosis severity. Fibrosis is a predictor of outcome independent of age, sex, WHO function class, PH severity and DLCO. Routine clinical reports of CT lung parenchymal disease identify groups of patients IPAH and CLD-PH with significantly different prognoses. After this work, I performed a service evaluation to assess if patients who undergo CT scans for PH diagnosis have their emphysema or fibrosis quantified as none/mild/moderate/severe. Over 18 years and 660 patients, emphysema severity was quantified in 91.1% of patients and fibrosis is 86.0%. The work won the second prize at the Royal College of Radiologists RADIANT Audit Day 2022.

Chapter five (published in *The Lancet Respiratory Medicine*) is a large comprehensive international multi-centre registry analysis which defines a new phenotype of Pulmonary Hypertension. The proposed IPAH-LD phenotype is characterised by a smoking history and DLCO <45%, but also has distinct CT features. The work has been cited 30 times in the last 9 months since publication and was an entire session of discussion at the European Respiratory Conference 2022. It is probable this work will lead to a reclassification of the Group 1 PAH and establishes that better characterisation of lung disease in PH patients is warranted[213]. The work has wide ranging implications for patient diagnosis, risk-stratification, management, and therapy decisions.

Chapter 6 (submitted to *European Radiology*) describes the development and clinical evaluation of a novel CTPA lung segmentation algorithm. The model was developed in-house using the state-of-the-art nnU-net and has a high technical accuracy (0.998) with no failures. An important aspect of the work was collaboration with the Stanford Center for Artificial Intelligence in Medicine & Imaging (AIMI). The algorithm was installed and deployed at AIMI prior to external validation on a different disease cohort (ILD). Within the medical imaging AI domain, few studies take this step of deployment at an external centre. An important message of the paper was the need for clinical and radiological review of AI outputs and limitations of traditional technical metrics of assessment. The work was presented at the 2022 UK Imaging and Oncology conference as an oral presentation and won the 2022 Grainger Prize.

Chapter 7 (submitted to *Lancet Digital Health*) describes the clinical utility of the automated lung parenchymal disease quantification model developed for this thesis. The model automatically classifies the lung parenchyma into five classes – normal, emphysema, ground glass, ground glass with reticulation, and honeycombing. The percentage of lung involvement of each parenchymal pattern is provided as an output. The prognostic strength of this continuous variable as a metric for lung disease assessment is established in the study. Combining radiological (semi-qualitative, none/mild/moderate/severe) and AI assessment (continuous variable as 0-100% of lung involvement) together improves the prognostic predictive strength of CTPA imaging as an imaging biomarker. In patients that had their fibrosis quantified has 'none' by radiologists, the AI identified minor lung disease, which was of prognostic significance. A particular strength of the study is external validation of the developed AI model across 33 centres and 37 scanners.

8.1 Limitations

* *Please also refer to individual chapters for additional dedicated discussion on limitations and solutions for each aspect of the thesis. In particular, section 2.6 is a detailed discussion on limitations, challenges and solutions to quantitative chest CT and AI.*

The thesis was designed to address and rectify common limitations in AI medical imaging studies – lack of comprehensive clinical data, lack of expert labelling for training data and lack of external validation in a diverse 'real world' clinical cohort. The clinico-radiological dataset created from the ASPIRE registry for this thesis is likely the largest in this patient group. It is 'tidy', contextual, anonymised, and quality controlled, therefore reaches the highest level, Level A, on the MIDaR scale[150]. 'Tidy' data refers to a structured database suited for programmatic statistical analysis[162]. All labelling has been performed by two FRCR radiologists with subspeciality experience in cardiothoracic imaging. The findings in this thesis are validated in a realistic clinical cohort with multiple scanners from multiple centres.

In addition to the specific limitations detailed in each chapter, there are some overarching limitations. All analysis is performed on retrospective data. All efforts have been made to account for the multiple inherent biases of this approach, as detailed in the individual chapters, but these are impossible to entirely account for. Whilst all patients were assigned diagnosis in a specialist centre, we cannot fully exclude the possibility of some misclassification bias and missed cases, including those never identified nor referred in non-specialist settings. An

unavoidable limitation of a clinical cohort spanning 18 years is the variance in diagnostic criteria, treatment options, patient management guidelines, and inherent quality of CT scanners. Prospective validation of these findings is required. Although the database incorporates data from multiple centres and scanners, the current overall automated end-to-end AI approach is deployed in a single centre. External deployment of the model and clinical validity of our findings in a different PH centre or registry with different population demographics would help better demonstrate generalisability.

Chapter 5 validates the 'IPAH lung phenotype' group in the COMPERA cohort, but a limitation of this work is differences between the study cohorts. COMPERA is a prospective registry which collects data of patients confirmed as suffering from pulmonary hypertension and treated with targeted therapies[176,213]. ASPIRE includes all patients suspected of having pulmonary hypertension, not just those that receive therapies. This may introduce some selection bias for COMPERA, as patients who do not receive treatment are not included. The difference is most pronounced in Group 3 PH, where all COMPERA patients received treatment, compared to 52% in ASPIRE. However, the impact of treatment in Group 3 PH is not well understood and it is uncertain if it improves outcomes. Group 3 also is a highly heterogenous group, and there may be within group differences in patients that are not recognised in the summary statistics. This is an avenue of research I will be investigating in detail, as described later in the 'Future Works' section. Other differences between cohorts is the median follow-up time (3.9 years for COMPERA compared to 4.5 years for ASPIRE) and omission of patients with risk factors for cardiovascular disease in COMPERA. These limitations are important to acknowledge, but they do not significantly detract from the strong overall message of the paper. The 'IPAH lung phenotype' has strikingly similar clinical characteristics and survival compared to 'Classical IPAH' and 'Group 3 PH' across two large independent multi-centre patient registries.

Radiological assessment is ultimately a subjective visual process. Reports between radiologists differ significantly in their style and content, and there exists significant inter-observer even between highly experienced radiologists[84–86]. Adjacent to this thesis, I am supervising a trainee in conducting a systematic review and meta-analysis of interobserver variability among radiologists in the reporting of chest CT studies, which investigates this in more detail[222].

It is important to acknowledge that truly 'ground truth' level objective classifications of the lung parenchyma, agreeable by all readers, is fundamentally challenging. This is a limitation acknowledged across other medical imaging domains[223,224]. All AI classification models have inherently, difficult to quantify, biases skewed by their development training set. The AI model will learn from the preferences, biases, and opinions of the two radiologists that labelled the training data. In the future, we aim to involve more radiologists of differing levels of expertise to create a more representative model. Certain patterns such as severe subpleural emphysema with interstitial thickening look visually identical to honeycombing and there is some overlap or overcall in these cases. There is also some overlap between severe traction bronchiectasis and honeycombing at the lung bases. CT features of 'fibrosis' are context dependant. In the context of pulmonary hypertension and lung disease, where scleroderma and interstitial lung disease are highly prevalent, it was decided to include extent of 'ground glass', 'ground glass with reticulation' and 'honeycombing' together in 'fibrosis'. A limitation of this approach is that this is not the case for all chest CT and lung disease. 'Ground glass' change is a non-specific feature and can also be associated with infective changes or malignancy. Therefore, this specific AI model would likely call infective changes as 'fibrosis'. The intended clinical use and deployment of the model is as an adjunct to reporting by a radiologist, who will be aware of such inherent limitations in image analysis. Ground glass change can be further characterised as centrilobular or diffuse ground glass change, but this was not done in this thesis due to the high degree of overlap and difficulty distinguishing between both patterns. It may be possible in the future to create a downstream two-class classification algorithm which sub-classifies regions classified as 'ground glass' in this approach.

The lung airways and vasculature were not individually segmented in this model. These approaches have corresponding limitations, particularly in more distal airways and vessels. Even on CTPA imaging, it is highly challenging to segment pulmonary arteries from veins[97,225]. A limitation of including airways and vasculature in the parenchymal classification is potential noise from these imaging features, but this was felt to be a better trade-off. Errors within a lung airway or vascular segmentation tool would propagate down to the classification model and reliance on multiple segmentation algorithm would also make deployment in external centres more challenging. Using percentage of lung involvement as a means of quantifying disease also has limitations. There are visual radiological differences between mild and severe parenchymal features, but the AI was not trained to stratify and distinguish these. Radiologically severe and highly localised disease would have a low percentage of lung involvement, but pathologically

be more significant. Fibrosis and interstitial lung disease is associated with lung volume loss, which influences the percentage of lung involvement for parenchymal features in these patients[88], However, the impact of this is difficult to accurately quantify, as non-fibrosed regions of lung can expand to partially compensate for this effect. There was no spirometric gating of CT scans, which could mean a variable state of inflation of the lungs. Spirometric gating however is not widely used in clinical practise and AI tools developed using such approaches would likely not be translatable to routine clinical practice.

Each AI model is trained to answer a specific question, in the case of this thesis, prognostic impact. The algorithm is therefore appropriately trained on a biased dataset of patients with a known assigned diagnosis of IPAH or PH-CLD. The AI model is trained and tested only on contrast enhanced CTPA imaging, which is the standard acquisition protocol used in these patients. However, non-contrast CT imaging is overall more commonly in other lung diseases such as COPD, and ILD. There are significant differences in the appearances of parenchymal between non-contrast and CTPA imaging due to the uptake of contrast from the lung parenchyma. It therefore would not be appropriate as a screening or diagnostic tool in other diseases or non-contrast CT without further training, adaptive methods, or transfer learning. The algorithm has been developed and tested in a predominantly white European population, limited by the demographics of the prevalent population. Future work will seek to address the limitations of this study by testing the model in larger cohorts of multi-ethnicity patients.

9 Future Work

Several avenues of potential future research have been identified from the multiple studies within this thesis. I aim to work on these future projects at a post-doctoral level as a clinical lecturer.

In depth phenotyping and characterisation of Group 3 PH to investigate treatment response. This thesis studies both Group 1 and Group 3 PH, and defined a new phenotype of 'IPAH lung phenotype' which appears to lie in between. Group 3 PH is a highly heterogenous group including diseases with different pathological and physiological processes. The two major subgroups are Group 3.1 – 'Obstructive disease or emphysema' and 3.2 'Restrictive lung disease.' Currently, all Group 3 patients are managed similarly, with the recommendation to maximise the treatment of their underlying lung disease. PAH specific vasoactive therapies, which have significantly improved survival in Group 1, are not indicated for use in Group 3.

Several randomised control trails have been performed in both Subgroup 3.1 and 3.2 with mixed results. In ILD (Group 3.2), Endothelin receptor antagonists (bosentan, ambrisentan and macitentan) were found to have no significant effect or were associated with worse outcomes in patients [63,226–228]. Sildenafil was also shown to stabilise 6-min walk distance and improve quality of life in one study, but two other subsequent studies failed to reach their primary endpoints [229,230]. In COPD (Group 3.1), studies on bosentan, sildenafil and tadalafil in patients with moderate PH on echocardiography failed to reach their primary endpoint and show a benefit [231–233]. In contrast, the SPHERIC-1 trial showed a significant improvement in PVR and quality of life in a small number of patients with more severe RHC proven PH [234]. Most promisingly, the INCREASE study demonstrated an improvement in exercise capacity and reducing in clinical worsening events in ILD [65]. Subgroup analysis also interestingly revealed that the benefits were seen only seen in patients with significant heamodynamic disease (defined as PVR > 4 WU)[51].

A significant limitation of the literature on the efficacy of PAH therapies on Group 3 is the large variability of the selection criteria between trails. Trials included or excluded patients based on extremes of either mild or severe heamodynamic disease or lung disease as quantified by pulmonary function tests and qualitative scoring of chest CT. The studies that showed benefit tended to be in those with severe haemodynamic PH [65,234]. There is a great opportunity

and great clinical need for further studies in this area that can identify a sub-phenotype that may benefit from PH therapy. A hypothesis is this phenotype may be characterised with severe haemodynamic disease, but only mild or moderate lung disease on pulmonary function testing and CT imaging. Discovery of such a sub-phenotype would open a further avenue for mechanistic research and drug development.

Improve understanding of the overall impact of lung disease in pulmonary hypertension. There is active interest within the PH community to better characterise lung disease and its prognostic impact. Chapter 6, the *Lancet Res Med* paper on the new IPAH-lung disease phenotype, was one of the main topics of discussion at the 2022 European Respiratory Society conference. The work described above for Group 3 can also be applied to Group 1 and the newly defined 'IPAH lung phenotype'. Should these patients be treated more or less aggressively compared to the 'Classical IPAH' group? Could aggressive treatment improve their prognosis, and if so, by how much compared to 'Classical IPAH' and 'Group 3 PH'? There is conflicting evidence surrounding treatment response, partially due to differing definitions of what truly constitutes a new phenotype. This thesis has investigated survival as the endpoint outcome for prognosis, but treatment response is an important endpoint to investigate. There is industry and community interest in identifying sub-groups within phenotypes that could be treatment responders. A limitation to this analysis is that treatment in PH is complex, and patients undergo combination therapy with multiple overlapping agents. The duration of treatment with specific agents, prior treatments, and severity of disease on commencement are important but hard to control confounders. This work would very much require prospective, randomised control trials with careful but broad selection criteria.

Investigate impact of DLCO, KCO and derive discriminating cut-offs. Current work uses a DLCO threshold of 45% percent predicted as a threshold for binary 'low' or 'high' DLCO. This was taken from multiple studies in this domain, but this choice of threshold has not formally been investigated. Future work could investigate the prognostic impact of DLCO as a continuous variable (0-100% percent predicted) and use different thresholds that account for its bimodal distribution. This analysis would help better identify an appropriate threshold for when DLCO is used in addition to the quantitative CT metrics, as it may be that they both have supplementary effects on prognosis. This would be important work as all PH patients routinely undergo DLCO testing and CT imaging, and a more discriminating cut-off could increase their prognostic predictive value. Going further, DLCO is derived from multiplying the Va (alveolar

volume) and Kco (CO transfer coefficient)[235]. The impact of these separately can be assessed, as it may be that KCO is more discriminating than DLCO when used with CT features. There exists some work in the literature investigating this in PH secondary to systemic sclerosis, where KCO changes predicted mortality, but this was less than DLCO[236]. Our work will investigate the relative impact of this in the spectrum of IPAH, IPAH lung phenotype and Group 3 PH (including sub-phenotypes to adjust for heterogeneity).

Prospective validation, external and clinical deployment. As stated in the limitations and discussion, the work described is retrospective and the AI model has been deployed in a single PH centre. Clinical deployment of the AI model, and subsequent prospective validation of its additive value on radiological reporting can help better understand the limitations of the model. Clinical deployment would involve working with senior clinical imaging scientists to incorporate the model as a part of an existing workflow and pipeline. It also requires gaining appropriate regulatory approvals. The cardiothoracic research group for this project has succeed in clinical translation of research models. An automated cardiac MRI segmentation tool was successfully clinically deployed and won the 2022 Medipex NHS Innovation Award[237]. Challenges of deploying any research model into clinical use are both technical and regulatory. From a technical perspective, the clinical service uses an existing set of workflows and pipelines, and the research model needs to be able to interact with the existing imaging pipeline. For example, in the clinical domain, the DICOM image format is regularly used, and tools (scanners, PACS system and any existing additional software used) need to work in that format. In the research domain, the DICOM format is often converted to the NIFTI (Neuroimaging Informatics Technology Initiative) format as this is more amenable to data analysis and machine learning. The pipeline for the AI model itself has multiple intermediary files and steps that generates other formats. For clinical deployment, all these steps and files need to be streamlined and made interactive with each other. The end output of the AI model needs to be a DICOM file, that can then be integrated back to the clinical imaging PACS system. Each step needs to pass clinical information governance and data regulatory steps. Integrating any AI model into clinical use requires regulatory approval, discussion and evaluation of the potential clinical impact, and careful consideration of the patient population in which it is used. For the AI model developed in this thesis, we feel the approach of generating an additional scan ‘series’ would work well, and would only be deployed in the PH population on which it was trained. This additional scan ‘series’ would contain a labelled DICOM image with the AI output mask which visualises what each region of the CT scan is classified as by

the AI model. Explain ability of the model outputs is an important step towards deployment of AI in clinical practice. The reporting radiologist can then have access to this additional information to aid in making their clinical opinion. The AI model itself will not 'report' any findings and will only be used as an adjunct additional tool.

External deployment of the model in other PH centres through collaborations would further strengthen the evidence base. This would require further significant technical and regulatory work. The model would first need to be packaged in a way that it can be shared and then steps would need to be taken by the other institution to correctly integrate it with their respective PACS and radiology systems. The lung segmentation model in the thesis has been successfully deployed at the Stanford Center for Artificial Intelligence in Medicine and Imaging in a research domain, and the aim would be for similar collaborations at first for the entire parenchymal classification model.

Quantifying other imaging features to create ensemble prognostic imaging models.
The AI methodologies and approaches used in this thesis to quantify lung parenchymal patterns can be used to automate the quantification of other imaging features. Within chest CT, potential targets include pulmonary embolism clot burden quantification, and automated measurement of cardiac and vascular structures. Some of the initial development work for these approaches has been carried out in parallel. A medium-term goal would be creating an ensemble overall CT model which incorporates multiple AI models that quantify and automate these different imaging features together. Each incident CTPA could then have its lung parenchyma quantified, cardiac and mediastinal structures measured, and clot burden quantified. This would aid significantly in radiological reporting, but also provide quantitative endpoints which can be used for assessing prognostic impact. The data can be put together in a machine learning or linear regression model, which could potentially be of higher prognostic strength.

The longer-term goal is to create a multi-modality imaging model which incorporates the relative strength of each imaging modality. For example, echocardiography is the first line recommended tool for screening for pulmonary hypertension, chest CT is the gold standard for lung disease assessment and clot burden estimation, and cardiac MRI is the gold standard for cardiac structural and functional assessment. The research department is working on AI models for each modality to automate the assessment and quantification of imaging features. Combining all these together, to replicate the current clinical workflow for the radiologist,

could greatly aid in patient phenotyping. An even more ambitious target would be combine all imaging features with clinical data such as pulmonary function tests, exercise tests and patients demographics for a 'big data' machine learning approach to see if such a model could accurately aid with screening and diagnosis of PH.

Direct comparison of AI metrics against expert radiologists. The current approach validates the AI approach against radiologists using two approaches. Firstly, it directly compares the AI classified outputs as a patch against the labelled patches by the radiologists. Secondly, it correlates the overall AI reported extent of disease against reports from multiple expert radiologists in normal clinical practice. Further prospective direct validation can be performed in which the same CT scan is reported concurrently by both an expert radiologist and the AI. Lobar segmentation can be performed, and each lobe divided into uniform equal regions, for which a radiologist estimates the extent of the lung parenchymal features to the nearest 5%. This approach has been previously used in Idiopathic Pulmonary Fibrosis when establishing the value of quantitative CT approaches[27,113]. A benefit of this approach would be a more systematic and detailed understanding of the limitations of the current AI model. For example, it would help establish whether the current model classifies traction bronchiectasis as 'honeycombing' and investigate what cardiac pulsation and breathing artefacts are quantified as. A limitation of this approach is the time intensive and laborious nature of the work, but this would help further clinically validate the AI outputs. Feedback from other radiologists, who will be the end users of the AI model, can also help shape future development to improve model usability and performance.

Investigating impact in other PH phenotypes and diseases and application to non-contrast CT. The AI package has been developed and trained specifically using patients with IPAH or PH-CLD. The CT texture patterns identified by the algorithm – such as ground glass, emphysema and fibrosis - are the basic foundational features of CT Chest pathology. With appropriate adaptive methods and transfer learning, the model could therefore be developed to be more widely applicable in other diseases. Within PH, the connective tissue disease phenotype (PH Group 1.4.1 as per latest classification) is an ideal target to first investigate the generalisability of the model as it is most adjacent to IPAH and is commonly associated with fibrotic lung disease. Outside of PH, pulmonary embolism and COVID-19 are diseases where CTPA imaging is common. Longer term and with substantially more development, the model could be adapted potentially to work in both non-contrast and contrast-enhanced CT imaging.

This would allow for uses in diseases which would benefit from severity quantification such as asthma, COPD and ILD.

10 Conclusion

This thesis establishes lung parenchymal patterns on routine CTPA imaging as prognostic imaging biomarkers for survival in patients with IPAH and PH-CLD. It further develops and establishes the clinical significance of an automated AI model to quantify these lung parenchymal patterns. Combining the strengths of AI and radiologists together improves the prognostic predictive strength of lung disease assessment. The 'in-house' development and external clinical evaluation of these AI models are significant milestones within the cardiothoracic medical imaging research group, and within the domain of medical imaging artificial intelligence more broadly. This work has advanced the understanding of the impact of lung disease in this cohort of patients, and it will contribute to a likely reclassification and sub-phenotyping of 'IPAH' into 'Classical IPAH' and 'IPAH with lung disease' in the next PH guidelines[213]. This will have implications on patient diagnosis, risk-stratification, treatment decisions and management.

Finally, the thesis has been a journey of personal and professional development beyond the described traditional 'research outputs'. Learning coding and statistical programming are personal highlights for skills I have developed during my PhD. The opportunity to collaborate and work with world experts in the fields of AI in medicine (Prof Curtis Langlotz, Director of Stanford Center for Artificial Intelligence in Medicine and Chair of Board for Radiological Society of North America, co-author of chapter 6) and Pulmonary Hypertension (Prof Marius Hoeper, PH Lead for PH at Hannover Medical School and section editor of 2022 European PH guidelines, co-author of chapter 5) has been instrumental to my development as a clinical academic and my understanding of the domain.

I finish this thesis fully committed to a substantive clinical academic career, and eternally grateful to the Wellcome Trust, my supervisors, my family and all for supporting my development.

Publications, Presentations and Awards

Publications directly related to thesis

1. Hoeper, M, **Dwivedi, K** [...] Kiely D, Halank M. Phenotyping of idiopathic pulmonary arterial hypertension: a registry analysis. *The Lancet Respiratory Medicine,* Jul 2022. (Impact Factor: 102.6)
2. Mamalakis, M, **Dwivedi, K**, [...], Swift, A. A transparent artificial intelligence framework to assess lung disease in pulmonary hypertension. *Sci Rep* 13, 3812 (2023). (IF: 5.0)
3. Sharkey, M, [...] **Dwivedi, K** [...] Swift AJ. Fully automatic cardiac four chamber and great vessel segmentation on CT pulmonary angiography using deep learning. *Frontiers in Cardiovascular Medicine*. Sept 2022. (IF: 5.05)
4. Alkhanfar D [...] **Dwivedi, K** [...] Swift A. *European Respiratory Journal Open Research,* 2021.… Swift, A. Severe pulmonary hypertension associated with lung disease is characterised by a loss of small pulmonary vessels on quantitative computed tomography. *European Respiratory Journal Open Research,* 2021. (IF: *awaited*)
5. **Dwivedi K,** Condliffe C, et al. Computed tomography lung parenchymal descriptions in routine radiological reporting have diagnostic and prognostic utility in patients with idiopathic pulmonary arterial hypertension and pulmonary hypertension associated with lung disease. *European Respiratory Journal Open Research,* 2022. (IF: 4.2)
6. **Dwivedi K**, Sharkey, M et al. Pulmonary Hypertension in Association with Lung Disease: Quantitative CT and Artificial Intelligence to the Rescue? State-of-the-Art Review. *Diagnostics,* (2021), 11 (4). (IF: 3.70)

International presentations

1. Quantitative Computed Tomography Lung Parenchymal Features Can Aid in Phenotyping of Group 1 and Group 3 Pulmonary Hypertension. American Thoracic Society Conference 2023 (accepted to be presented), Washington, USA.
2. Computed Tomography (CT) Features Are of Diagnostic Utility in Pre Diagnosis Idiopathic Pulmonary Arterial Hypertension (IPAH): A Case Controlled Study. Poster presentation. American Thoracic Society Conference 2022, San Francisco, USA.
3. Fully automatic cardiac and great vessel segmentation on CT pulmonary angiography (CTPA) using deep learning. Poster presentation. American Thoracic Society Conference 2022, San Francisco, USA.
4. Deep learning approaches to classify lung parenchymal disease on CT scans. Poster presentation. American Thoracic Society Conference 2022, San Francisco, USA.

5. Prognostic importance of centrilobular ground glass changes in patients with Idiopathic Pulmonary Arterial Hypertension (IPAH). Poster presentation. European Respiratory Society Congress 2022, Barcelona, Spain.
6. Reporting of prognostically important lung disease severity on CT in Pulmonary Hypertension with lung disease at a specialist referral centre. Royal College of Radiologists RADIANT Audit Day. Poster presentation. London, UK.
7. Pulmonary vascular resistance thresholds of 3, 5 and 10 WU are associated with significantly different survival in pulmonary hypertension with chronic lung disease. Poster presentation. Pulmonary Vascular Research Institute Conference 2022, Athens, Greece.
8. Fibrosis on Computed Tomography (CT) is an independent predictor of mortality in PH-CLD (Pulmonary Hypertension with Chronic Lung Disease). Poster presentation. Pulmonary Vascular Research Institute Conference 2022, Athens, Greece.
9. Deep learning approach for lung segmentation in computed tomography pulmonary angiography (CTPA). UK Imaging and Oncology Congress (UKIO), Liverpool, UK.
10. CT lung parenchymal patterns as significant prognostic imaging biomarkers in pulmonary hypertension with lung disease. Poster presentation, RSNA 'Chest' Trainee Prize winner. Radiology Society of North America, Chicago, 2021.

Leadership and management roles related to thesis

1. Committee member, Royal College Radiologists Artificial Intelligence Working Group.
2. Artificial Intelligence lead, Royal College of Radiologists RADIANT (Radiology Academic Network for Trainees)
3. Supervisor for NIHR Academic Foundation Programme, Sheffield
4. Supervisor for Special Study Module, University of Sheffield Medical School.
5. Peer reviewer for Clinical Radiology

Invited talks

1. Academia and research as a trainee and Academic Clinical Fellow, Royal College of Radiologists Research Day 2022.
2. The challenge of phenotyping IPAH: an emerging role for CT. National Pulmonary Hypertension Forum 2022.
3. AI in Chest CT and Pulmonary Hypertension, Stanford Centre for Artificial Intelligence in Medicine (AIMI), University of Stanford, 2022.
4. AI for chest diagnosis, Regional Chest and Cardiac Day, Yorkshire Denary, 2021.

Awards

1. Radiological Society of North America (RSNA) Trainee Research Prize, RSNA Congress 2021.
2. First Prize, Grainger Price Meeting, 2022.
3. Second Prize, Royal College of Radiologists RADIANT Audit Day, 2022.

Bibliography

1. Hatano, S., Strasser, T. & World Health Organization. *Primary pulmonary hypertension : report on a WHO meeting, Geneva, 15-17 October 1973 / edited by Shuichi Hatano and Toma Strasser*. (World Health Organization, 1975).
2. Simonneau, G. *et al.* Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *Eur. Respir. J.* **53**, (2019).
3. Humbert, M. *et al.* 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur. Heart J.* **43**, 3618–3731 (2022).
4. Pulmonary hypertension specialist centres. https://www.phauk.org/treatment-for-pulmonary-hypertension/pulmonary-hypertension-specialist-centres/.
5. National Pulmonary Hypertension Audit 11th Annual Report - NHS Digital. https://digital.nhs.uk/data-and-information/publications/statistical/national-pulmonary-hypertension-audit/11th-annual-report.
6. Galiè, N. *et al.* 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *Eur. Heart J.* **37**, 67–119 (2016).
7. Iannuzzi, G. L., D'Alto, M., Formisano, R. & Maniscalco, M. Biomarkers in clinical management of pulmonary hypertension: has the emperor no clothes? A call for action. *Biomark. Med.* **13**, 235–238 (2019).
8. Benza, R. L. *et al.* Predicting Survival in Patients With Pulmonary Arterial Hypertension: The REVEAL Risk Score Calculator 2.0 and Comparison With ESC/ERS-Based Risk Assessment Strategies. *Chest* **156**, 323–337 (2019).
9. Anderson, J. J. *et al.* Retrospective validation of the REVEAL 2.0 risk score with the Australian and New Zealand pulmonary hypertension Registry cohort. *Chest* **157**, 162–172 (2020).
10. Sitbon, O. *et al.* Validation of two predictive models for survival in pulmonary arterial hypertension. *Eur. Respir. J.* **46**, 152–164 (2015).
11. Benza, R. L. *et al.* REVEAL risk scores applied to riociguat-treated patients in PATENT-2: Impact of changes in risk score on survival. *J. Heart Lung Transplant.* **37**, 513–519 (2018).

12. Benza, R. L. *et al.* REVEAL risk score in patients with chronic thromboembolic pulmonary hypertension receiving riociguat. *J. Heart Lung Transplant.* **37**, 836–843 (2018).
13. Kane, G. C. *et al.* Integration of clinical and hemodynamic parameters in the prediction of long-term survival in patients with pulmonary arterial hypertension. *Chest* **139**, 1285–1293 (2011).
14. Lewis, R. A. *et al.* Identification of Cardiac Magnetic Resonance Imaging Thresholds for Risk Stratification in Pulmonary Arterial Hypertension. *Am. J. Respir. Crit. Care Med.* **201**, 458–468 (2020).
15. Kylhammar, D. *et al.* A comprehensive risk stratification at early follow-up determines prognosis in pulmonary arterial hypertension. *Eur. Heart J.* **39**, 4175–4181 (2018).
16. Hoeper, M. M. *et al.* COMPERA 2.0: a refined four-stratum risk assessment model for pulmonary arterial hypertension. *Eur. Respir. J.* **60**, (2022).
17. Hoeper, M. M. *et al.* Mortality in pulmonary arterial hypertension: prediction by the 2015 European pulmonary hypertension guidelines risk stratification model. *Eur. Respir. J.* **50**, (2017).
18. Boucly, A. *et al.* Risk assessment, prognosis and guideline implementation in pulmonary arterial hypertension. *Eur. Respir. J.* **50**, (2017).
19. Cooper, B. G. *et al.* The Global Lung Function Initiative (GLI) Network: bringing the world's respiratory reference values together. *Breathe (Sheff)* **13**, e56–e64 (2017).
20. Low, A. T., Medford, A. R. L., Millar, A. B. & Tulloh, R. M. R. Lung function in pulmonary hypertension. *Respir. Med.* **109**, 1244–1249 (2015).
21. Sun, X.-G., Hansen, J. E., Oudiz, R. J. & Wasserman, K. Pulmonary function in primary pulmonary hypertension. *J. Am. Coll. Cardiol.* **41**, 1028–1035 (2003).
22. Meyer, F. J. *et al.* Peripheral airway obstruction in primary pulmonary hypertension. *Thorax* **57**, 473–476 (2002).
23. Hurdman, J. *et al.* Pulmonary hypertension in COPD: results from the ASPIRE registry. *Eur. Respir. J.* **41**, 1292–1301 (2013).
24. Nathan, S. D., Shlobin, O. A., Ahmad, S., Urbanek, S. & Barnett, S. D. Pulmonary hypertension and pulmonary function testing in idiopathic pulmonary fibrosis. *Chest* **131**, 657–663 (2007).
25. Andersen, K. H. *et al.* Prevalence, predictors, and survival in pulmonary hypertension related to end-stage chronic obstructive pulmonary disease. *J. Heart Lung Transplant.* **31**, 373–380 (2012).

26. Nathan, S. D. *et al.* Pulmonary hypertension in chronic lung disease and hypoxia. *Eur. Respir. J.* **53**, (2019).
27. Jacob, J. *et al.* Serial CT analysis in idiopathic pulmonary fibrosis: comparison of visual features that determine patient outcome. *Thorax* **75**, 648–654 (2020).
28. Rajaram, S. *et al.* CT features of pulmonary arterial hypertension and its major subtypes: a systematic CT evaluation of 292 patients from the ASPIRE Registry. *Thorax* **70**, 382–387 (2015).
29. Swift, A. J. *et al.* Diagnostic accuracy of CT pulmonary angiography in suspected pulmonary hypertension. *Eur. Radiol.* **30**, 4918–4929 (2020).
30. Kiely, D. G. *et al.* EXPRESS: Statement on imaging and pulmonary hypertension from the Pulmonary Vascular Research Institute (PVRI). *Pulm. Circ.* 2045894019841990 (2019).
31. Tan, R. T. *et al.* Utility of CT scan evaluation for predicting pulmonary hypertension in patients with parenchymal lung disease. Medical College of Wisconsin Lung Transplant Group. *Chest* **113**, 1250–1256 (1998).
32. Palmer, S. M. *et al.* Massive pulmonary edema and death after prostacyclin infusion in a patient with pulmonary veno-occlusive disease. *Chest* **113**, 237–240 (1998).
33. Foley, R. W. *et al.* Computed tomography appearances of the lung parenchyma in pulmonary hypertension. *Br. J. Radiol.* 20200830 (2020).
34. Seeger, W. *et al.* Pulmonary hypertension in chronic lung diseases. *J. Am. Coll. Cardiol.* **62**, D109-16 (2013).
35. Elia, D. *et al.* Pulmonary hypertension and chronic lung disease: where are we headed? *Eur. Respir. Rev.* **28**, (2019).
36. Kovacs, G. *et al.* Pulmonary Vascular Involvement in Chronic Obstructive Pulmonary Disease. Is There a Pulmonary Vascular Phenotype? *Am. J. Respir. Crit. Care Med.* **198**, 1000–1011 (2018).
37. Chaouat, A. *et al.* Severe pulmonary hypertension and chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **172**, 189–194 (2005).
38. Kiely, D. G., Elliot, C. A., Sabroe, I. & Condliffe, R. Pulmonary hypertension: diagnosis and management. *BMJ* **346**, f2028 (2013).
39. Swietlik, E. M., Gräf, S. & Morrell, N. W. The role of genomics and genetics in pulmonary arterial hypertension. *Glob Cardiol Sci Pract* **2020**, e202013 (2020).
40. Morrell, N. W. *et al.* Genetics and genomics of pulmonary arterial hypertension. *Eur. Respir. J.* **53**, (2019).

41. Lewis, R. A. *et al.* Mild parenchymal lung disease and/or low diffusion capacity impacts survival and treatment response in patients diagnosed with idiopathic pulmonary arterial hypertension. *Eur. Respir. J.* **55**, (2020).
42. Condliffe, R., Kiely, D. G. & Lewis, R. A. Mild parenchymal lung disease is still lung disease. *The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology* vol. 56 (2020).
43. Sommer, N. *et al.* Current and future treatments of pulmonary arterial hypertension. *Br. J. Pharmacol.* **178**, 6–30 (2021).
44. Farber, H. W. *et al.* Five-Year outcomes of patients enrolled in the REVEAL Registry. *Chest* **148**, 1043–1054 (2015).
45. Rose, L. *et al.* Survival in pulmonary hypertension due to chronic lung disease: Influence of low diffusion capacity of the lungs for carbon monoxide. *J. Heart Lung Transplant.* **38**, 145–155 (2019).
46. Galié, N. *et al.* Ambrisentan therapy for pulmonary arterial hypertension. *J. Am. Coll. Cardiol.* **46**, 529–535 (2005).
47. Olschewski, H. The Challenge to Decide between Pulmonary Hypertension Due to Chronic Lung Disease and PAH with Chronic Lung Disease. *Diagnostics* **11**, (2021).
48. Olschewski, H. *et al.* Inhaled iloprost for severe pulmonary hypertension. *N. Engl. J. Med.* **347**, 322–329 (2002).
49. Dwivedi, K. *et al.* Pulmonary Hypertension in Association with Lung Disease: Quantitative CT and Artificial Intelligence to the Rescue? State-of-the-Art Review. *Diagnostics* **11**, 679 (2021).
50. Hurdman, J. *et al.* ASPIRE registry: assessing the Spectrum of Pulmonary hypertension Identified at a REferral centre. *Eur. Respir. J.* **39**, 945–955 (2012).
51. Kiely, D. G. & Condliffe, R. Assessing pulmonary hypertension severity in lung disease is a key step to improving outcomes: embrace resistance and don't be pressurised to go with the flow. *The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology* vol. 58 (2021).
52. Lettieri, C. J., Nathan, S. D., Barnett, S. D., Ahmad, S. & Shorr, A. F. Prevalence and outcomes of pulmonary arterial hypertension in advanced idiopathic pulmonary fibrosis. *Chest* **129**, 746–752 (2006).
53. Hoeper, M. M. *et al.* Diffusion Capacity and Mortality in Patients With Pulmonary Hypertension Due to Heart Failure With Preserved Ejection Fraction. *JACC Heart Fail* **4**, 441–449 (2016).

54. Sivova, N. *et al.* Relevance of partitioning DLCO to detect pulmonary hypertension in systemic sclerosis. *PLoS One* **8**, e78001 (2013).
55. Chandra, S. *et al.* Carbon monoxide diffusing capacity and mortality in pulmonary arterial hypertension. *J. Heart Lung Transplant.* **29**, 181–187 (2010).
56. Hruby, J. & Butler, J. Variability of routine pulmonary function tests. *Thorax* **30**, 548–553 (1975).
57. Magnussen, H., Vaz Fragoso, C. A., Miller, M. R. & Brusasco, V. Spirometry Variability Must Be Critically Interpreted before Negating a Clinical Diagnosis of Chronic Obstructive Pulmonary Disease. *American journal of respiratory and critical care medicine* vol. 197 835–836 (2018).
58. Khalil, N. *et al.* Phase 2 clinical trial of PBI-4050 in patients with idiopathic pulmonary fibrosis. *Eur. Respir. J.* **53**, (2019).
59. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H. & Aerts, H. J. W. L. Artificial intelligence in radiology. *Nat. Rev. Cancer* **18**, 500–510 (2018).
60. RCR position statement on artificial intelligence. https://www.rcr.ac.uk/posts/rcr-position-statement-artificial-intelligence.
61. Khemasuwan, D., Sorensen, J. S. & Colt, H. G. Artificial intelligence in pulmonary medicine: computer vision, predictive model and COVID-19. *Eur. Respir. Rev.* **29**, (2020).
62. Weitzenblum, E., Chaouat, A., Canuet, M. & Kessler, R. Pulmonary hypertension in chronic obstructive pulmonary disease and interstitial lung diseases. *Semin. Respir. Crit. Care Med.* **30**, 458–470 (2009).
63. Raghu, G. *et al.* Treatment of idiopathic pulmonary fibrosis with ambrisentan: a parallel, randomized trial. *Ann. Intern. Med.* **158**, 641–649 (2013).
64. Nathan, S. D. *et al.* Riociguat for idiopathic interstitial pneumonia-associated pulmonary hypertension (RISE-IIP): a randomised, placebo-controlled phase 2b study. *Lancet Respir Med* **7**, 780–790 (2019).
65. Waxman, A. *et al.* Inhaled Treprostinil in Pulmonary Hypertension Due to Interstitial Lung Disease. *N. Engl. J. Med.* **384**, 325–334 (2021).
66. McKinney, S. M. *et al.* International evaluation of an AI system for breast cancer screening. *Nature* **577**, 89–94 (2020).
67. Grace, K., Salvatier, J., Dafoe, A., Zhang, B. & Evans, O. When Will AI Exceed Human Performance? Evidence from AI Experts. *arXiv [cs.AI]* (2017).
68. Moravčík, M. *et al.* DeepStack: Expert-level artificial intelligence in heads-up no-limit poker. *Science* **356**, 508–513 (2017).

69. Silver, D. *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
70. Liu, X. *et al.* A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health* **1**, e271–e297 (2019).
71. Augusto, J. B. *et al.* Diagnosis and risk stratification in hypertrophic cardiomyopathy using machine learning wall thickness measurement: a comparison with human test-retest performance. *The Lancet Digital Health* **3**, e20–e28 (2021).
72. Yang, J. *et al.* Unsupervised Domain Adaption With Adversarial Learning (UDAA) for Emphysema Subtyping on Cardiac CT Scans: The Mesa Study. in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* 289–293 (2019).
73. Harvey, H. Why AI will not replace radiologists - Towards Data Science. *Towards Data Science* https://towardsdatascience.com/why-ai-will-not-replace-radiologists-c7736f2c7d80 (2018).
74. Sathyakumar, K., Munoz, M., Singh, J., Hussain, N. & Babu, B. A. Automated Lung Cancer Detection Using Artificial Intelligence (AI) Deep Convolutional Neural Networks: A Narrative Literature Review. *Cureus* **12**, e10017 (2020).
75. Cui, S. *et al.* Development and clinical application of deep learning model for lung nodules screening on CT images. *Sci. Rep.* **10**, 13657 (2020).
76. Svoboda, E. Artificial intelligence is improving the detection of lung cancer. *Nature* **587**, S20–S22 (2020).
77. Armato, S. G., 3rd *et al.* The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med. Phys.* **38**, 915–931 (2011).
78. van Timmeren, J. E., Cester, D., Tanadini-Lang, S., Alkadhi, H. & Baessler, B. Radiomics in medical imaging-"how-to" guide and critical reflection. *Insights Imaging* **11**, 91 (2020).
79. Sollini, M., Antunovic, L., Chiti, A. & Kirienko, M. Towards clinical application of image mining: a systematic review on artificial intelligence and radiomics. *Eur. J. Nucl. Med. Mol. Imaging* **46**, 2656–2672 (2019).
80. Chen, A., Karwoski, R. A., Gierada, D. S., Bartholmai, B. J. & Koo, C. W. Quantitative CT Analysis of Diffuse Lung Disease. *Radiographics* **40**, 28–43 (2020).
81. Silva, M., Milanese, G., Seletti, V., Ariani, A. & Sverzellati, N. Pulmonary quantitative CT imaging in focal and diffuse disease: current research and clinical applications. *Br. J. Radiol.* **91**, 20170644 (2018).

82. Lessmann, N. *et al.* Automated Assessment of CO-RADS and Chest CT Severity Scores in Patients with Suspected COVID-19 Using Artificial Intelligence. *Radiology* 202439 (2020).
83. Salisbury, M. L. *et al.* Idiopathic Pulmonary Fibrosis: The Association between the Adaptive Multiple Features Method and Fibrosis Outcomes. *Am. J. Respir. Crit. Care Med.* **195**, 921–929 (2017).
84. Watadani, T. *et al.* Interobserver variability in the CT assessment of honeycombing in the lungs. *Radiology* **266**, 936–944 (2013).
85. Bankier, A. A., De Maertelaer, V., Keyzer, C. & Gevenois, P. A. Pulmonary emphysema: subjective visual grading versus objective quantification with macroscopic morphometry and thin-section CT densitometry. *Radiology* **211**, 851–858 (1999).
86. Widell, J. & Lidén, M. Interobserver variability in high-resolution CT of the lungs. *Eur J Radiol Open* **7**, 100228 (2020).
87. Flaherty, K. R. *et al.* Idiopathic interstitial pneumonia: do community and academic physicians agree on diagnosis? *Am. J. Respir. Crit. Care Med.* **175**, 1054–1060 (2007).
88. Jacob, J. *et al.* Automated Quantitative Computed Tomography Versus Visual Computed Tomography Scoring in Idiopathic Pulmonary Fibrosis. *J. Thorac. Imaging* **31**, 304–311 (2016).
89. Recht, M. & Bryan, R. N. Artificial Intelligence: Threat or Boon to Radiologists? *J. Am. Coll. Radiol.* **14**, 1476–1480 (2017).
90. Clinical radiology UK workforce census report 2018. https://www.rcr.ac.uk/publication/clinical-radiology-uk-workforce-census-report-2018.
91. McDonald, R. J. *et al.* The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. *Acad. Radiol.* **22**, 1191–1198 (2015).
92. Shiraishi, J., Li, Q., Appelbaum, D. & Doi, K. Computer-aided diagnosis and artificial intelligence in clinical imaging. *Semin. Nucl. Med.* **41**, 449–462 (2011).
93. Fazal, M. I., Patel, M. E., Tye, J. & Gupta, Y. The past, present and future role of artificial intelligence in imaging. *Eur. J. Radiol.* **105**, 246–250 (2018).
94. Shiraishi, J. *et al.* Usefulness of temporal subtraction images for identification of interval changes in successive whole-body bone scans: JAFROC analysis of radiologists' performance. *Acad. Radiol.* **14**, 959–966 (2007).
95. Annarumma, M. *et al.* Automated Triaging of Adult Chest Radiographs with Deep Artificial Neural Networks. *Radiology* **291**, 196–202 (2019).

96. Kulik, T. J. *et al.* Pulmonary arterial hypertension: what the large pulmonary arteries tell us. *Pediatr. Cardiol.* **32**, 759–765 (2011).
97. Lesage, D., Angelini, E. D., Bloch, I. & Funka-Lea, G. A review of 3D vessel lumen segmentation techniques: models, features and extraction schemes. *Med. Image Anal.* **13**, 819–845 (2009).
98. Nardelli, P. *et al.* Pulmonary Artery-Vein Classification in CT Images Using Deep Learning. *IEEE Trans. Med. Imaging* **37**, 2428–2440 (2018).
99. Armstrong, I. *et al.* The patient experience of pulmonary hypertension: a large cross-sectional study of UK patients. *BMC Pulm. Med.* **19**, 67 (2019).
100. Jairam, P. M. *et al.* Incidental findings on chest CT imaging are associated with increased COPD exacerbations and mortality. *Thorax* **70**, 725–731 (2015).
101. Binder, P., Batmanghelich, N. K., Estepar, R. S. J. & Golland, P. Unsupervised Discovery of Emphysema Subtypes in a Large Clinical Cohort. *Mach Learn Med Imaging* **10019**, 180–187 (2016).
102. Yang, J. *et al.* Unsupervised Discovery of Spatially-Informed Lung Texture Patterns for Pulmonary Emphysema: The MESA COPD Study. *Med. Image Comput. Comput. Assist. Interv.* **10433**, 116–124 (2017).
103. Shin, K. E., Chung, M. J., Jung, M. P., Choe, B. K. & Lee, K. S. Quantitative computed tomographic indexes in diffuse interstitial lung disease: correlation with physiologic tests and computed tomography visual scores. *J. Comput. Assist. Tomogr.* **35**, 266–271 (2011).
104. Park, H. J. *et al.* Texture-Based Automated Quantitative Assessment of Regional Patterns on Initial CT in Patients With Idiopathic Pulmonary Fibrosis: Relationship to Decline in Forced Vital Capacity. *AJR Am. J. Roentgenol.* **207**, 976–983 (2016).
105. Condon, D. F., Nickel, N. P., Anderson, R., Mirza, S. & de Jesus Perez, V. A. The 6th World Symposium on Pulmonary Hypertension: what's old is new. *F1000Res.* **8**, (2019).
106. Hemnes, A. & PVDOMICS Study Group. PVDOMICS: Early Clinical Findings Across the Spectrum of Pulmonary Hypertension. in *A105. GLORY DAYS: THE LATEST CLINICAL RESEARCH IN PAH* A2517–A2517 (American Thoracic Society, 2019).
107. Zhu, N. *et al.* Novel risk genes and mechanisms implicated by exome sequencing of 2572 individuals with pulmonary arterial hypertension. *Genome Med.* **11**, 69 (2019).
108. National Cohort Study of Idiopathic and Heritable PAH. https://www.ipahcohort.com/.
109. Evans, J. D. W. *et al.* BMPR2 mutations and survival in pulmonary arterial hypertension: an individual participant data meta-analysis. *Lancet Respir Med* **4**, 129–137 (2016).

110. Elliott, C. G. *et al.* Relationship of BMPR2 mutations to vasoreactivity in pulmonary arterial hypertension. *Circulation* **113**, 2509–2515 (2006).
111. Dunmore, B. J., Jones, R. J., Toshner, M. R., Upton, P. D. & Morrell, N. W. Approaches to treat pulmonary arterial hypertension by targeting bmpr2 - from cell membrane to nucleus. *Cardiovasc. Res.* (2021) doi:10.1093/cvr/cvaa350.
112. van der Bruggen, C. E. *et al.* Bone Morphogenetic Protein Receptor Type 2 Mutation in Pulmonary Arterial Hypertension: A View on the Right Ventricle. *Circulation* **133**, 1747–1760 (2016).
113. Jacob, J. *et al.* Mortality prediction in idiopathic pulmonary fibrosis: evaluation of computer-based CT analysis with conventional severity measures. *Eur. Respir. J.* **49**, (2017).
114. Maldonado, F. *et al.* Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis. *Eur. Respir. J.* **43**, 204–212 (2014).
115. Feng, D.-Y. *et al.* Selection of glucocorticoid-sensitive patients in interstitial lung disease secondary to connective tissue diseases population by radiomics. *Ther. Clin. Risk Manag.* **14**, 1975–1986 (2018).
116. Mooney, J. J. *et al.* Radiographic fibrosis score predicts survival in hypersensitivity pneumonitis. *Chest* **144**, 586–592 (2013).
117. Walsh, S. L. F., Sverzellati, N., Devaraj, A., Wells, A. U. & Hansell, D. M. Chronic hypersensitivity pneumonitis: high resolution computed tomography patterns and pulmonary function indices as prognostic determinants. *Eur. Radiol.* **22**, 1672–1679 (2012).
118. Regan, E. A. *et al.* Genetic epidemiology of COPD (COPDGene) study design. *COPD* **7**, 32–43 (2010).
119. Sieren, J. P. *et al.* SPIROMICS Protocol for Multicenter Quantitative Computed Tomography to Phenotype the Lungs. *Am. J. Respir. Crit. Care Med.* **194**, 794–806 (2016).
120. Han, M. K. *et al.* Frequency of exacerbations in patients with chronic obstructive pulmonary disease: an analysis of the SPIROMICS cohort. *Lancet Respir Med* **5**, 619–626 (2017).
121. Han, M. K. *et al.* Chronic obstructive pulmonary disease exacerbations in the COPDGene study: associated radiologic phenotypes. *Radiology* **261**, 274–282 (2011).
122. Milanese, G., Silva, M. & Sverzellati, N. Lung volume reduction of pulmonary emphysema: the radiologist task. *Curr. Opin. Pulm. Med.* **22**, 179–186 (2016).

123. Sverzellati, N. *et al.* Reliability of quantitative computed tomography to predict postoperative lung function in patients with chronic obstructive pulmonary disease having a lobectomy. *J. Comput. Assist. Tomogr.* **29**, 819–824 (2005).
124. Röhrich, S. *et al.* Prospects and Challenges of Radiomics by Using Nononcologic Routine Chest CT. *Radiology: Cardiothoracic Imaging* **2**, e190190 (2020).
125. Ng, C. S., Wells, A. U. & Padley, S. P. A CT sign of chronic pulmonary arterial hypertension: the ratio of main pulmonary artery to aortic diameter. *J. Thorac. Imaging* **14**, 270–278 (1999).
126. Hansell, D. M. *et al.* CT staging and monitoring of fibrotic interstitial lung diseases in clinical practice and treatment trials: a position paper from the Fleischner Society. *Lancet Respir Med* **3**, 483–496 (2015).
127. Labaki, W. W. *et al.* The Role of Chest Computed Tomography in the Evaluation and Management of the Patient with Chronic Obstructive Pulmonary Disease. *Am. J. Respir. Crit. Care Med.* **196**, 1372–1379 (2017).
128. Acharya, U. R., Hagiwara, Y., Sudarshan, V. K., Chan, W. Y. & Ng, K. H. Towards precision medicine: from quantitative imaging to radiomics. *J. Zhejiang Univ. Sci. B* **19**, 6–24 (2018).
129. Herold, C. J. *et al.* Imaging in the Age of Precision Medicine: Summary of the Proceedings of the 10th Biannual Symposium of the International Society for Strategic Studies in Radiology. *Radiology* **279**, 226–238 (2016).
130. Abramson, R. G. *et al.* Methods and challenges in quantitative imaging biomarker development. *Acad. Radiol.* **22**, 25–32 (2015).
131. Quantitative Imaging Biomarkers Alliance. https://www.rsna.org/research/quantitative-imaging-biomarkers-alliance.
132. van Griethuysen, J. J. M. *et al.* Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **77**, e104–e107 (2017).
133. Mackin, D. *et al.* Measuring Computed Tomography Scanner Variability of Radiomics Features. *Invest. Radiol.* **50**, 757–765 (2015).
134. Andrearczyk, V., Depeursinge, A. & Müller, H. Learning cross-protocol radiomics and deep feature standardization from CT images of texture phantoms. in *Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications* vol. 10954 109540I (International Society for Optics and Photonics, 2019).
135. Shafiq-Ul-Hassan, M. *et al.* Voxel size and gray level normalization of CT radiomic features in lung cancer. *Sci. Rep.* **8**, 10545 (2018).

136. Shafiq-Ul-Hassan, M. *et al.* Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med. Phys.* **44**, 1050–1062 (2017).
137. Ligero, M. *et al.* Minimizing acquisition-related radiomics variability by image resampling and batch effect correction to allow for large-scale data analysis. *Eur. Radiol.* (2020) doi:10.1007/s00330-020-07174-0.
138. Mühlberg, A. *et al.* The Technome - A Predictive Internal Calibration Approach for Quantitative Imaging Biomarker Research. *Sci. Rep.* **10**, 1103 (2020).
139. Parmar, C. *et al.* Robust Radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One* **9**, e102107 (2014).
140. Kamnitsas, K. *et al.* Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. *arXiv [cs.CV]* (2016).
141. Zhang, S. *et al.* Computer-Aided Diagnosis (CAD) of Pulmonary Nodule of Thoracic CT Image Using Transfer Learning. *J. Digit. Imaging* **32**, 995–1007 (2019).
142. Park, C. M. Can Artificial Intelligence Fix the Reproducibility Problem of Radiomics? *Radiology* vol. 292 374–375 (2019).
143. Madani, A., Van Muylem, A. & Gevenois, P. A. Pulmonary Emphysema: Effect of Lung Volume on Objective Quantification at Thin-Section CT. *Radiology* **257**, 260–268 (2010).
144. Galbán, C. J. *et al.* Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nat. Med.* **18**, 1711–1715 (2012).
145. Sverzellati, N. *et al.* CT-based weight assessment of lung lobes: comparison with ex vivo measurements. *Diagn. Interv. Radiol.* **19**, 355–359 (2013).
146. He, L. *et al.* Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. *Sci. Rep.* **6**, 34921 (2016).
147. Duman, I. E., Cimsit, C., Yildizeli, S. O. & Cimsit, N. C. Parenchymal density changes in acute pulmonary embolism: Can quantitative CT be a diagnostic tool? A preliminary study. *Clin. Imaging* **41**, 157–163 (2017).
148. Sun, C., Shrivastava, A., Singh, S. & Gupta, A. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. in *2017 IEEE International Conference on Computer Vision (ICCV)* 843–852 (2017).
149. Halevy, A., Norvig, P. & Pereira, F. The Unreasonable Effectiveness of Data. *IEEE Intell. Syst.* **24**, 8–12 (2009).
150. Harvey, H. & Glocker, B. A standardised approach for preparing imaging data for machine learning tasks in radiology. *Artificial Intelligence in Medical Imaging* (2019).

151. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016).
152. Kohli, M. D., Summers, R. M. & Geis, J. R. Medical Image Data and Datasets in the Era of Machine Learning-Whitepaper from the 2016 C-MIMI Meeting Dataset Session. *J. Digit. Imaging* **30**, 392–399 (2017).
153. Mutasa, S., Sun, S. & Ha, R. Understanding artificial intelligence based radiology studies: What is overfitting? *Clin. Imaging* **65**, 96–99 (2020).
154. Nagendran, M. *et al.* Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* **368**, m689 (2020).
155. Zech, J. R. *et al.* Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Med.* **15**, e1002683 (2018).
156. Balki, I. *et al.* Sample-Size Determination Methodologies for Machine Learning in Medical Imaging Research: A Systematic Review. *Can. Assoc. Radiol. J.* **70**, 344–353 (2019).
157. Cruz Rivera, S. *et al.* Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Lancet Digit Health* **2**, e549–e560 (2020).
158. Liu, X. *et al.* Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health* **2**, e537–e548 (2020).
159. Reyes, M. *et al.* On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities. *Radiol Artif Intell* **2**, e190043 (2020).
160. The Lancet Respiratory Medicine. Opening the black box of machine learning. *Lancet Respir Med* **6**, 801 (2018).
161. Hedgire, S. S., Baliyan, V., Ghoshhajra, B. B. & Kalra, M. K. Recent advances in cardiac computed tomography dose reduction strategies: a review of scientific evidence and technical developments. *J Med Imaging (Bellingham)* **4**, 031211 (2017).
162. Wickham, H. Tidy Data. *J. Stat. Softw.* **59**, 1–23 (2014).
163. Zeder, K. *et al.* Elevated pulmonary vascular resistance predicts mortality in COPD patients. *Eur. Respir. J.* (2021) doi:10.1183/13993003.00944-2021.
164. Olsson, K. M. *et al.* Pulmonary vascular resistance predicts mortality in patients with pulmonary hypertension associated with interstitial lung disease: results from the COMPERA registry. *Eur. Respir. J.* (2021) doi:10.1183/13993003.01483-2021.

165. Godinas, L., Harari, S., Barberà, J. A. & Montani, D. Mild parenchymal lung disease is still lung disease. *The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology* vol. 56 (2020).
166. Hoeper, M. M. *et al.* Idiopathic pulmonary arterial hypertension phenotypes determined by cluster analysis from the COMPERA registry. *J. Heart Lung Transplant.* (2020) doi:10.1016/j.healun.2020.09.011.
167. Resten, A. *et al.* Pulmonary hypertension: CT of the chest in pulmonary venoocclusive disease. *AJR Am. J. Roentgenol.* **183**, 65–70 (2004).
168. Trip, P. *et al.* Severely reduced diffusion capacity in idiopathic pulmonary arterial hypertension: patient characteristics and treatment responses. *Eur. Respir. J.* **42**, 1575–1585 (2013).
169. Olsson, K. M., Fuge, J., Meyer, K., Welte, T. & Hoeper, M. M. More on idiopathic pulmonary arterial hypertension with a low diffusing capacity. *Eur. Respir. J.* **50**, (2017).
170. Nolan, R. L. *et al.* Pulmonary cholesterol granulomas in patients with pulmonary artery hypertension: chest radiographic and CT findings. *AJR Am. J. Roentgenol.* **172**, 1317–1319 (1999).
171. Horton, M. R. & Tuder, R. M. Primary pulmonary arterial hypertension presenting as diffuse micronodules on CT. *Crit. Rev. Comput. Tomogr.* **45**, 335–341 (2004).
172. Sheehan, R., Perloff, J. K., Fishbein, M. C., Gjertson, D. & Aberle, D. R. Pulmonary neovascularity: a distinctive radiographic finding in Eisenmenger syndrome. *Circulation* **112**, 2778–2785 (2005).
173. Griffin, N., Allen, D., Wort, J., Rubens, M. & Padley, S. Eisenmenger syndrome and idiopathic pulmonary arterial hypertension: do parenchymal lung changes reflect aetiology? *Clin. Radiol.* **62**, 587–595 (2007).
174. Rich, S. *et al.* Primary pulmonary hypertension. A national prospective study. *Ann. Intern. Med.* **107**, 216–223 (1987).
175. Ling, Y. *et al.* Changing demographics, epidemiology, and survival of incident pulmonary arterial hypertension: results from the pulmonary hypertension registry of the United Kingdom and Ireland. *Am. J. Respir. Crit. Care Med.* **186**, 790–796 (2012).
176. Hoeper, M. M. *et al.* Elderly patients diagnosed with idiopathic pulmonary arterial hypertension: results from the COMPERA registry. *Int. J. Cardiol.* **168**, 871–880 (2013).
177. Schiess, R. *et al.* Tobacco smoke: a risk factor for pulmonary arterial hypertension? A case-control study. *Chest* **138**, 1086–1092 (2010).

178. Dwivedi, K. *et al.* Computed tomography lung parenchymal descriptions in routine radiological reporting have diagnostic and prognostic utility in patients with idiopathic pulmonary arterial hypertension and pulmonary hypertension associated with lung disease. *ERJ Open Res* **8**, (2022).
179. D'Alonzo, G. E. *et al.* Survival in patients with primary pulmonary hypertension. Results from a national prospective registry. *Ann. Intern. Med.* **115**, 343–349 (1991).
180. Hoeper, M. M. & Vonk-Noordegraaf, A. Is there a vanishing pulmonary capillary syndrome? *Lancet Respir Med* **5**, 676–678 (2017).
181. Seimetz, M. *et al.* Inducible NOS inhibition reverses tobacco-smoke-induced emphysema and pulmonary hypertension in mice. *Cell* **147**, 293–305 (2011).
182. Barst, R. J. *et al.* A comparison of continuous intravenous epoprostenol (prostacyclin) with conventional therapy for primary pulmonary hypertension. *N. Engl. J. Med.* **334**, 296–301 (1996).
183. Rubin, L. J. *et al.* Bosentan therapy for pulmonary arterial hypertension. *N. Engl. J. Med.* **346**, 896–903 (2002).
184. Galiè, N. *et al.* Sildenafil citrate therapy for pulmonary arterial hypertension. *N. Engl. J. Med.* **353**, 2148–2157 (2005).
185. Galiè, N. *et al.* Tadalafil therapy for pulmonary arterial hypertension. *Circulation* **119**, 2894–2903 (2009).
186. Galiè, N. *et al.* Ambrisentan for the treatment of pulmonary arterial hypertension: results of the ambrisentan in pulmonary arterial hypertension, randomized, double-blind, placebo-controlled, multicenter, efficacy (ARIES) study 1 and 2. *Circulation* **117**, 3010–3019 (2008).
187. Galiè, N. *et al.* Initial use of ambrisentan plus tadalafil in pulmonary arterial hypertension. *N. Engl. J. Med.* **373**, 834–844 (2015).
188. Ghofrani, H.-A. *et al.* Riociguat for the treatment of pulmonary arterial hypertension. *N. Engl. J. Med.* **369**, 330–340 (2013).
189. Pulido, T. *et al.* Macitentan and morbidity and mortality in pulmonary arterial hypertension. *N. Engl. J. Med.* **369**, 809–818 (2013).
190. Sitbon, O. *et al.* Selexipag for the Treatment of Pulmonary Arterial Hypertension. *N. Engl. J. Med.* **373**, 2522–2533 (2015).
191. Valentin, S. *et al.* Outcomes of patients with decreased arterial oxyhaemoglobin saturation on pulmonary arterial hypertension drugs. *Eur. Respir. J.* **58**, (2021).

192. Montani, D. *et al.* Clinical phenotypes and outcomes of heritable and sporadic pulmonary veno-occlusive disease: a population-based study. *Lancet Respir Med* **5**, 125–134 (2017).
193. Eyries, M. *et al.* Familial pulmonary arterial hypertension by KDR heterozygous loss of function. *Eur. Respir. J.* **55**, (2020).
194. Hesamian, M. H., Jia, W., He, X. & Kennedy, P. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *J. Digit. Imaging* **32**, 582–596 (2019).
195. Lenchik, L. *et al.* Automated Segmentation of Tissues Using CT and MRI: A Systematic Review. *Acad. Radiol.* **26**, 1695–1706 (2019).
196. Hofmanninger, J. *et al.* Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. *Eur Radiol Exp* **4**, 50 (2020).
197. Cao, W., Wu, R., Cao, G. & He, Z. A Comprehensive Review of Computer-Aided Diagnosis of Pulmonary Nodules Based on Computed Tomography Scans. *IEEE Access* **8**, 154007–154023 (2020).
198. Carvalho, L. E., Sobieranski, A. C. & von Wangenheim, A. 3D Segmentation Algorithms for Computerized Tomographic Imaging: a Systematic Literature Review. *J. Digit. Imaging* **31**, 799–850 (2018).
199. Gerard, S. E. *et al.* CT Image Segmentation for Inflamed and Fibrotic Lungs Using a Multi-Resolution Convolutional Neural Network. *arXiv [eess.IV]* (2020).
200. Sousa, J. *et al.* Lung Segmentation in CT Images: A Residual U-Net Approach on a Cross-Cohort Dataset. *NATO Adv. Sci. Inst. Ser. E Appl. Sci.* **12**, 1959 (2022).
201. Yoo, S. J. *et al.* Automated Lung Segmentation on Chest Computed Tomography Images with Extensive Lung Parenchymal Abnormalities Using a Deep Neural Network. *Korean J. Radiol.* **22**, 476–488 (2021).
202. Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211 (2021).
203. Roth, H. *et al.* Rapid Artificial Intelligence Solutions in a Pandemic - The COVID-19-20 Lung CT Lesion Segmentation Challenge. *Res Sq* (2021) doi:10.21203/rs.3.rs-571332/v1.
204. Guo, F. *et al.* Improving cardiac MRI convolutional neural network segmentation on small training datasets and dataset shift: A continuous kernel cut approach. *Med. Image Anal.* **61**, 101636 (2020).
205. Reinke, A. *et al.* Common limitations of performance metrics in biomedical image analysis. (2022).

206. Hasani, N. *et al.* Trustworthy Artificial Intelligence in Medical Imaging. *PET Clin.* **17**, 1–12 (2022).
207. Kim, D. W., Jang, H. Y., Kim, K. W., Shin, Y. & Park, S. H. Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers. *Korean J. Radiol.* **20**, 405–410 (2019).
208. Yu, A. C., Mohajer, B. & Eng, J. External Validation of Deep Learning Algorithms for Radiologic Diagnosis: A Systematic Review. *Radiol Artif Intell* **4**, e210064 (2022).
209. Sforazzini, F. *et al.* Deep Learning-based Automatic Lung Segmentation on Multiresolution CT Scans from Healthy and Fibrotic Lungs in Mice. *Radiol Artif Intell* **4**, e210095 (2022).
210. Park, J. *et al.* Fully Automated Lung Lobe Segmentation in Volumetric Chest CT with 3D U-Net: Validation with Intra- and Extra-Datasets. *J. Digit. Imaging* **33**, 221–230 (2020).
211. Yang, J. *et al.* Autosegmentation for thoracic radiation treatment planning: A grand challenge at AAPM 2017. *Med. Phys.* **45**, 4568–4581 (2018).
212. Nathan, S. D. *et al.* Impact of the new definition for pulmonary hypertension in patients with lung disease: an analysis of the United Network for Organ Sharing database. *Pulm. Circ.* **11**, 2045894021999960 (2021).
213. Hoeper, M. M. *et al.* Phenotyping of idiopathic pulmonary arterial hypertension: a registry analysis. *Lancet Respir Med* (2022) doi:10.1016/S2213-2600(22)00097-2.
214. Kay, F. U. *et al.* Translation of Quantitative Imaging Biomarkers into Clinical Chest CT. *Radiographics* **39**, 957–976 (2019).
215. Gopalan, D. & Gibbs, J. S. R. From Early Morphometrics to Machine Learning-What Future for Cardiovascular Imaging of the Pulmonary Circulation? *Diagnostics (Basel)* **10**, (2020).
216. Hansell, D. M. *et al.* Fleischner Society: glossary of terms for thoracic imaging. *Radiology* **246**, 697–722 (2008).
217. Sharkey, M. J., Dwivedi, K., Alabed, S. & Swift, A. J. Deep learning automated quantification of lung disease in pulmonary hypertension on CT pulmonary angiography: A preliminary clinical study with external validation. *arXiv [eess.IV]* (2023).
218. Leibig, C. *et al.* Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. *Lancet Digit Health* **4**, e507–e519 (2022).

219. Shikhare, S. *et al.* Right-to-left ventricle ratio determined by machine learning algorithms on CT pulmonary angiography images predicts prolonged ICU length of stay in operated chronic thromboembolic pulmonary hypertension. *Br. J. Radiol.* **95**, 20210722 (2022).
220. Rahaghi, F. N. *et al.* Quantification of Arterial and Venous Morphologic Markers in Pulmonary Arterial Hypertension Using CT Imaging. *Chest* **160**, 2220–2231 (2021).
221. Jimenez-Del-Toro, O. *et al.* A lung graph model for the radiological assessment of chronic thromboembolic pulmonary hypertension in CT. *Comput. Biol. Med.* **125**, 103962 (2020).
222. PROSPERO: Systematic review and meta-analysis of interobserver variability among radiologists in the reporting of chest CT studies. *NIHR PROSPERO* https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=361803.
223. Wilson, M. *et al.* Validation and Clinical Applicability of Whole-Volume Automated Segmentation of Optical Coherence Tomography in Retinal Disease Using Deep Learning. *JAMA Ophthalmol.* **139**, 964–973 (2021).
224. Webb, J. M. *et al.* Comparing deep learning-based automatic segmentation of breast masses to expert interobserver variability in ultrasound imaging. *Comput. Biol. Med.* **139**, 104966 (2021).
225. Mühlfeld, C. *et al.* Recent developments in 3-D reconstruction and stereology to study the pulmonary vasculature. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **315**, L173–L183 (2018).
226. King, T. E., Jr *et al.* BUILD-3: a randomized, controlled trial of bosentan in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **184**, 92–99 (2011).
227. Raghu, G. *et al.* Macitentan for the treatment of idiopathic pulmonary fibrosis: the randomised controlled MUSIC trial. *Eur. Respir. J.* **42**, 1622–1632 (2013).
228. Corte, T. J. *et al.* Bosentan in pulmonary hypertension associated with fibrotic idiopathic interstitial pneumonia. *Am. J. Respir. Crit. Care Med.* **190**, 208–217 (2014).
229. Han, M. K. *et al.* Sildenafil preserves exercise capacity in patients with idiopathic pulmonary fibrosis and right-sided ventricular dysfunction. *Chest* **143**, 1699–1708 (2013).
230. Kolb, M. *et al.* Nintedanib plus sildenafil in patients with idiopathic pulmonary fibrosis. *N. Engl. J. Med.* **379**, 1722–1731 (2018).
231. Stolz, D. *et al.* A randomised, controlled trial of bosentan in severe COPD. *Eur. Respir. J.* **32**, 619–628 (2008).
232. Blanco, I. *et al.* Sildenafil to improve respiratory rehabilitation outcomes in COPD: a controlled trial. *Eur. Respir. J.* **42**, 982–992 (2013).

233. Goudie, A. R., Lipworth, B. J., Hopkinson, P. J., Wei, L. & Struthers, A. D. Tadalafil in patients with chronic obstructive pulmonary disease: a randomised, double-blind, parallel-group, placebo-controlled trial. *Lancet Respir. Med.* **2**, 293–300 (2014).
234. Vitulo, P. *et al.* Sildenafil in severe pulmonary hypertension associated with chronic obstructive pulmonary disease: A randomized controlled multicenter clinical trial. *J. Heart Lung Transplant.* **36**, 166–174 (2017).
235. Hughes, J. M. B. & Pride, N. B. Examination of the carbon monoxide diffusing capacity (DL(CO)) in relation to its KCO and VA components. *Am. J. Respir. Crit. Care Med.* **186**, 132–139 (2012).
236. Nihtyanova, S. I. *et al.* Dynamic Prediction of Pulmonary Hypertension in Systemic Sclerosis Using Landmark Analysis. *Arthritis Rheumatol* **75**, 449–458 (2023).
237. Medipex NHS Innovation Awards 2022a – NIHR CYP MedTech. https://cypmedtech.nihr.ac.uk/2022/10/13/medipex-nhs-innovation-awards-2022a/.