Author:
Davis, Richard J

Title:
Evaluation of Cardiopulmonary Exercise Testing (CPET) as a Prognostic Tool in Idiopathic Pulmonary Fibrosis (IPF)

Author:
Davis, Richard J

Title:
Evaluation of Cardiopulmonary Exercise Testing (CPET) as a Prognostic Tool in Idiopathic Pulmonary Fibrosis (IPF)

Evaluation of cardio-pulmonary exercise testing (CPET) as a prognostic tool in Idiopathic Pulmonary Fibrosis (IPF)

Richard Jon Davis BSc

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Masters of Science by Research in the Faculty of Health Sciences, Bristol Medical School

Words (excluding references): 22,802

Abstract

Personalised medicine is a medical approach that emphasises the customisation of healthcare, with all decisions and practices being tailored to individual patients.

Idiopathic pulmonary fibrosis (IPF) is a progressive fibrosing condition of the lungs with a median survival of 2-3 years from diagnosis. There is however vast heterogeneity in terms of presenting features, severity and disease course. Individual survival varies greatly as a result, leading to difficulties for patients and clinicians in terms of end-of-life discussions, treatment choices and conduct of clinical trials.

Clinicians would benefit from tools that would help to better predict clinical progression or track response to therapy. Several prognostic tools have been used in IPF with variable success. CardioPulmonary Exercise Testing (CPET) has been proposed as a potentially effective tool for the early detection of gas exchange abnormalities in lung diseases but its prognostic value remains uncertain. There are limited data available on the use of CPET as a predictive tool for disease progression in the setting of IPF, with a weak correlation between CPET and mortality reported in small cohorts. The predictive value of CPET in determining future disease progression and its relationship with Quality of Life (QoL) measurements and lung physiology is not known.

This thesis aims to test the hypothesis that CPET would be feasible in a population of mild to moderate IPF patients and more sensitive to change in patient's health status than 6 Minute Walk Test (6MWT), Forced Vital Capacity (FVC) or Transfer factor for carbon monoxide (TL_{CO}), the routine clinical tests used globally today.

Table of Contents

List of figures and tables

Figures:

Tables:

Dedication

For the patients around the world living with IPF

Acknowledgements

Many thanks to my supervisors Professor Nick Maskell and Professor Ann Millar for your time given and experience passed on.

Thanks to Dr Shaney Barratt for all your help, guidance and huge support throughout this research period. I hope the contents of this thesis will lead to a larger clinical trial for patients with IPF to better understand their clinical course.

Thanks to Dr Huzaifa Adamali for your clinical support and leadership of the Bristol ILD service and to Dr Charlie Sharpe for your thoughts and ideas when setting this challenge

I would also like to thank Anna Morley, Naomi Rippon and all the respiratory research team for your support and kindness in addition to Sarah Mulholland and Michelle Morales for their help in guidance during recruitment.

Finally, I would like to thank Jason Viner, Catherine Dixon and the Bristol respiratory physiology team for all the effort and support during this study.

Published output

Published abstracts/conference proceedings

The use of cardiopulmonary exercise testing in Idiopathic Pulmonary Fibrosis: Feasibility and correlation with quality of life measures.

British Thoracic Society, London, UK 2019

Davis R, Viner J, Dixon C, Morley A, Adamali H, Maskell N, Barratt SL

The use of cardiopulmonary exercise testing in Idiopathic Pulmonary Fibrosis: correlation with baseline quality of life measures.

American Thoracic Society, Philadelphia, USA 2020

Davis R, Viner J, Dixon C, Morley A, Adamali H, Maskell N, Barratt SL

****The prognostic value of cardiopulmonary exercise testing in interstitial lung disease: A systematic literature review***

European Respiratory Journal Open Research, 2020

Shaney L. Barratt, **Richard Davis,** Charles Sharp, John D. Pauling

A role for cardiopulmonary exercise testing in detecting physiological changes underlying health status in Idiopathic pulmonary fibrosis: a feasibility study

BMC Pulmonary Medicine – revised submission under review

Davis R, Dixon C, Millar AB, Maskell N, Barratt SL ***Author**

declaration:

The final draft of the systematic review was primarily written by Dr S Barratt (SB) & Dr J Pauling (JP). Full literature search and review was undertaken by RD & CS. All duplication removal was made by RD. Bias review was independently made and agreed between RD & SB.

Signed: [redacted]

1st Author

Signed: [redacted]

Final Author

Author’s declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: [redacted] DATE: 02-JUL-2021

Abbreviations

ACCP	American College of Chest Physicians
AEC	Alveolar Epithelial Cells
AE-IPF	Acute Exacerbation of IPF
ALAT	Latin American Thoracic Society
ALI	Acute Lung Injury
AT	Anaerobic threshold
ATS	American Thoracic Society
COPD	Chronic Obstructive Pulmonary Disease
CPET	Cardio-Pulmonary Exercise Testing
CPI	Composite physiologic index
CRP	Clinical–Radiographic–Physiologic
CTD	Connective Tissue Disease
CTIMP	Clinical Trials of Investigational Medicinal Products
EBV	Epstein - Barr virus
EIB	Exercise Induced Bronchospasm
EMA	European Medicines Agency
EMT	Epithelial-Mesenchymal Transition
FDA	US Food and Drug Administration
FEV_1	Forced Expiratory Volume in 1 second
f-IPF	Familial IPF
FVC	Forced Vital Capacity
GAP index	Gender Age Physiology index
GORD	Gastro-oesophageal reflux disease

HCV	Hepatitis C Virus
HRCT	High-resolution computed tomography
HRQoL	Health Related Quality of Life
IIP	Idiopathic Interstitial Pneumonias
ILD	Interstitial Lung Disease
IPF	Idiopathic Pulmonary Fibrosis
IPF-PROM	Idiopathic Pulmonary Fibrosis – Patient Reported Outcome Measure
IRAS	Integrated Research Application System
JRS	Japanese Respiratory Society
K-BILD	Kings-Brief Interstitial Lung Disease
LVRS	Lung Volume Reduction Surgery
MDT	Multi-disciplinary team
MRC	Medical Research Council
NICE	National Institute for Health and Care Excellence
NLR	Neutrophil: Lymphocyte ratio
OUES	Oxygen Uptake Efficiency Slope
PR	Pulmonary Rehabilitation
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QoL	Quality of Life
REC	Regional Ethics Committee
RER	Respiratory Exchange Ratio
SD	Standard Deviation
SGRQ	St. Georges Respiratory Questionnaire
SpO_2	Oxygen saturation
TBLC	Transbronchial Lung Cryobiopsy

TL_{CO}	Transfer Factor for Carbon Monoxide
VATS	Video-assisted thoracoscopic
VAS	Visual Analogue Scale
VE	Minute ventilation
V_E/VCO_2	Ventilatory equivalent for carbon dioxide
V_E/VO_2	Ventilatory equivalent for oxygen consumption
VO_2 max/kg	maximum rate of oxygen consumption per kilogram
4MGS	4 Metre Gait Speed
6MWT	6-Minute Walk Test

Chapter 1. Introduction

The Interstitial Lung Diseases (ILDs) also known as the Diffuse Parenchymal Lung Diseases (DPLD) are a group of over 200 heterogeneous acute and chronic lung disorders. Whilst some of these lung disorders may be triggered by an underlying autoimmune condition, or occur secondary to a hazardous chemical such as asbestos, among these disorders are the Idiopathic Interstitial Pneumonias (IIPs), all with an unknown aetiology and include Idiopathic Pulmonary Fibrosis (IPF) (Figure 1.1) (Demedts and Costabel 2002).

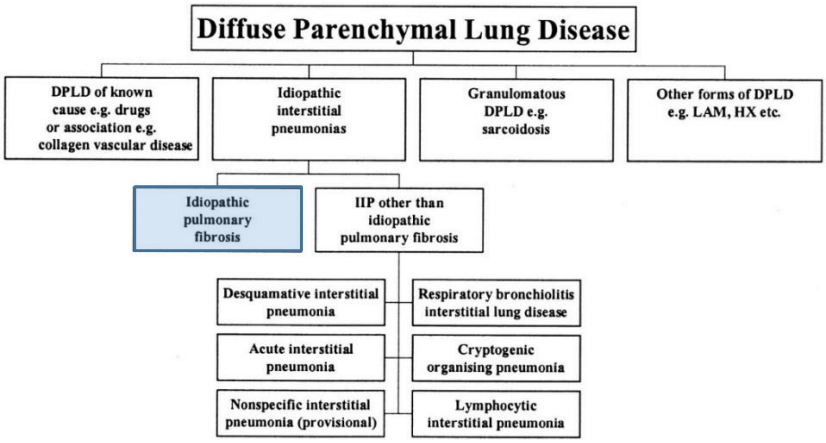


Figure 1.1. Classification of ILD. Adapted from ATS/ERS international multidisciplinary consensus classification of the idiopathic interstitial pneumonias. (Demedts and Costabel 2002)

1.1. Idiopathic Pulmonary Fibrosis

IPF is a progressive and in most cases fatal pulmonary disease of unknown cause, affecting gaseous exchange at the lung parenchyma. Although the causal pathology for disease onset is uncertain and discussed later in this chapter, it is apparent the body's repair cascade malfunctions leading to aberrant wound healing, irreversible epithelial remodelling and the formation of scar tissue (Figure 1.2) (Chambers 2008).

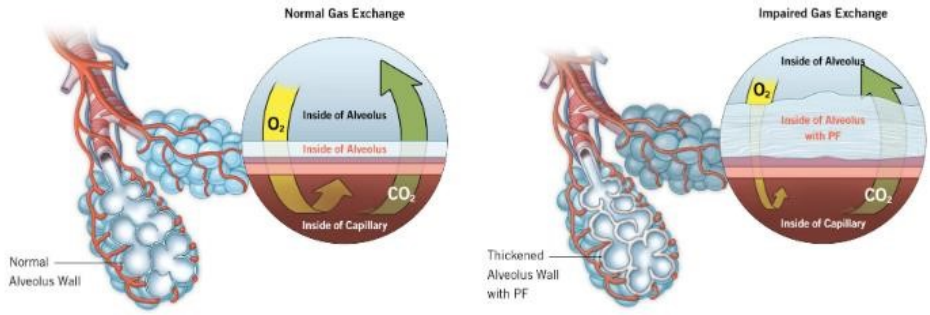


Figure 1.2. Graphical description of the thickening of alveoli walls in pathogenesis of IPF. Copied with permission from PulmonaryFibrosis.org (Foundation 2020).

1.1.2. Epidemiology

Whilst classified as a rare disease (this is defined by the European Medicines Agency (EMA) as a disease affecting fewer than 5 in 10,000 across the European Union) there are an estimated 32,500 patients living in the UK with IPF, with approximately 6000 new cases per year. IPF largely affects patients over the age of 60 and is significantly more prevalent in males than females (Snell, Strachan and Hubbard 2016), although there is no evidence to suggest a difference of IPF incidence based on race, ethnic group or social environment (Meltzer and Noble 2008). In the US, due to a rising number & aging population, it is thought there will be a doubling of IPF cases between 2005 and 2030 (Fernández Pérez et al. 2010).

1.1.3. Proposed Risk Factors for disease development

The origin of the initial epithelial injury, thought to be the triggering factor of this disease is largely unknown (idiopathic), however a number of exposures have been suggested as risk factors for development of IPF. The most common of these being a history of smoking and older age (Bellou et al. 2017) although no direct causal association has been established for either of these and estimates of effect vary between studies depending on study design and case definition (Ryu et al. 2001).

Occupational exposures, such as farm workers, vets, gardeners and metal or steel industry workers have an independent association with the development of the high-resolution computed tomography (HRCT) pattern associated with but not definitive for IPF, usual interstitial pneumonia (UIP). Further work from Baumgartner and colleagues suggested raising birds, hairdressing and exposure to vegetable or animal dust were significantly associated with a diagnosis of IPF when compared to agematched controls (odds ratio for all, in excess of 4.0) (Baumgartner et al. 2000).

Furthermore, one study suggested up to 87% of IPF patients report the symptom of gastrooesophageal reflux disease (GORD), leading to the hypothesis that micro-aspiration of acid into the lungs is a significant initiating factor for this disease (Raghu et al. 2006a). Support for this hypothesis has grown by evidence linking GORD and the presence of hiatus hernia (Fein et al. 1999), suggesting that hiatus hernia may be a co-factor in not only the pathogenesis of IPF, but also the subsequent disease progression (Mackintosh et al. 2019).

Other potential risk factors for development of IPF include diabetes (Gribbin, Hubbard and Smith 2009) and viral infection. Whilst some evidence exists pointing towards a role of viruses in acute exacerbations of IPF (AE-IPF), including Epstein Barr virus (EBV) & Hepatitis C (HCV) (ATS/ERS Statement 2000) their exact role remains uncertain. The propensity of both virus' and bacterial infection to cause alveolar-epithelial cell injury has been well documented, however their role in pathogenesis of the

disease is little understood (Yin et al. 2020). A recent paper from the UK has also examined a causal effect of telomere shortening as a mechanism for pathogenesis that is unique to IPF and is not seen in other chronic lung diseases, such as COPD (Duckworth et al. 2020). Finally, the role of genetics has been long explored as a cause of IPF, with one estimation that up to 20% of IPF cases could have a genetic foundation (García-Sancho et al. 2011). Whilst numerous studies have investigated the role of individual genes in IPF pathogenesis, a gain of function MUC5B gene polymorphism appears to be the strongest genetic risk factor and effect size for IPF (Evans et al. 2016) (Figure 1.3) and has been implicated in familial forms (Yang et al. 2015). Animal models observing the overexpression of this gene have suggested impaired mucus clearance, leading to the initiation of fibrosis as a plausible scientific pathway, although further research on this continues (Hancock et al.
2018).

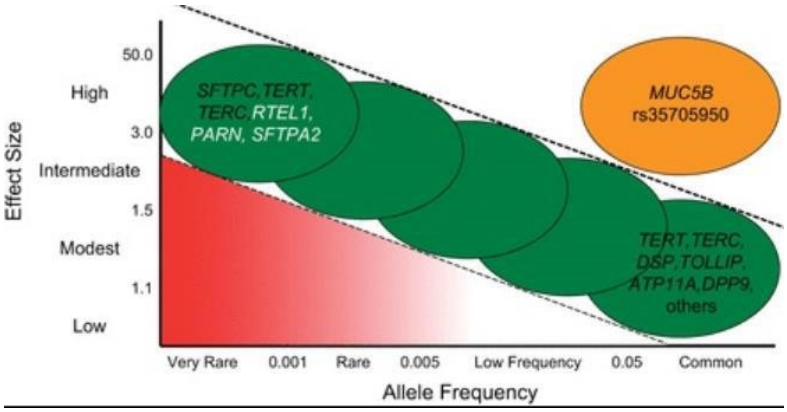


Fig 1.3. Relationship between allele frequency and penetrance of the risk allele. Adapted from Figure 1 (Antonarakis et al. 2010)

Genetic profiling as a potential prognostic tool has to date been little explored. An earlier age of onset of familial IPF (f-IPF) is associated with a more aggressive disease course (Krauss et al. 2019) and an ERS taskforce is due to publish a consensus paper on genetic involvement in progressive fibrosis within the coming twelve months (Borie and Van Moorsel 2021).

1.2. Pathogenesis of IPF

There remains no clear consensus on the pathogenesis of IPF. Generalised inflammation progressing to widespread parenchymal fibrosis was an historical paradigm (Kim, Collard and King 2006) that has become less popular over the last decade and increasingly it is considered, that repetitive epithelial injury and activation of fibroblasts are critical early events. This in turn triggers a cascade eventually leading to reorganisation of pulmonary tissue where fibrosis is believed to predominate over inflammation (Selman and Pardo 2014). The alveolar wall, the site of gaseous exchange between the

lungs and the bloodstream, is made largely of two types of Alveolar Epithelial cells (AECs), type I and type II, with the former covering up to 95% of the alveolar epithelial surface (McElroy and Kasper 2004). Type I cells increase the surface area of the alveolar wall and facilitate gaseous exchange. Type II AEC's secrete surfactant in order to lower surface tension and prevent alveolar atelectasis. Upon injury, type II AECs are stimulated to proliferate and then differentiate, to facilitate replacement of type I AECs. This process is known as re-epithelialisation and such a wound healing process is essential to restore the barrier integrity. An aberrant response to repeated injury of this surface is now thought to be the primary mechanism of IPF (Herzog et al. 2008).

Damage to the alveolar capillary membrane is a triggering event that mediates a down-stream cascade resulting in epithelial-mesenchymal transition (EMT) whereby epithelial cells undertake molecular changes gaining a mesenchymal phenotype. A subsequent increase in fibroblast differentiation, activation of the coagulation cascade, fibroblast/myofibroblast differentiation and recruitment to sites of injury with deposition of extracellular matrix, and disruption of normal lung architecture results in scar tissue formation. This can be seen as the characteristic usual interstitial pneumonia (UIP) scarring pattern observed on high-resolution computed tomography (HRCT) in IPF patients. Such a radiological finding may not be conclusive of an IPF diagnosis but is highly suggestive given an absence of other clinical or serological history (Figure 1.4) (Paolocci et al. 2018).

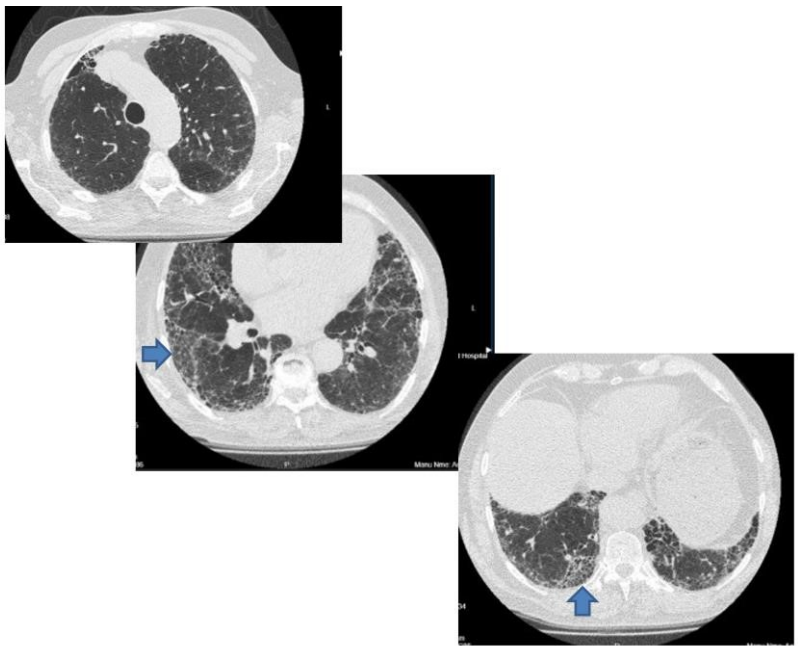


Figure 1.4. Patient with diagnosed IPF with HRCT pattern consistent with UIP. Reticulation with honeycombing. Picture shown with permission from Bristol ILD Service MDT.

1.3. Clinical Course of IPF

The most frequent symptom patients will report to their healthcare professional is breathlessness, with which many patients will have experienced a progressive worsening over >6-month period (ATS/ERS Statement 2000) .A non-productive cough is another common symptom at presentation, whilst on examination, lower lobe, bibasilar 'velcro-like' crackles are present in approximately 80% of IPF patients. Although less common, finger clubbing is reported in 30-50% of patients at initial presentation (Nakamura and Suda 2015).

The clinical course varies greatly between patients and is very difficult to predict (Ley, Collard and King 2011, Raghu et al. 2011), (Kim et al. 2006) as seen graphically in figure 1.5 (Kim, Perlman and Tomic 2015). Whilst some patients rapidly deteriorate within months, others may have a much slower disease progression. At any point in the disease trajectory, acute deteriorations may be experienced. Traditionally, this acute worsening (including new radiographic abnormalities on HRCT), accompanied by the absence of specific cause for this disease change, is known as an Acute Exacerbation of IPF (AEIPF). More recent thinking from an international expert panel following a literature review in to these often catastrophic events, has suggested AE-IPF share many clinical features with patients experiencing Acute Lung Injury (ALI) from a known cause (e.g. Infection) and such events should be labelled as 'triggered' or 'idiopathic', with an awareness that outcome can often be similar (Collard et al. 2016). After an AE-IPF, in hospital mortality rates are extremely high with a mortality rate up to 85% and mean survival rate of 3-13 days (Collard et al. 2016). This disease heterogeneity makes it difficult for clinicians to give an accurate prognosis at presentation and poses challenges with regards to timing of lung transplantation and palliative care. Understanding an individual's risk of rapid progression or susceptibility to AE-IPF, may allow clinicians to increase observation and provide an improved personal care plan.

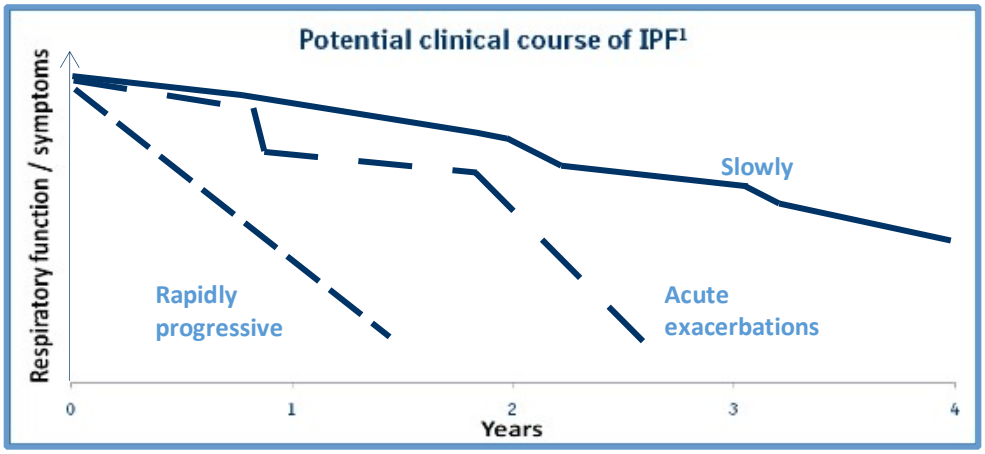


Figure 1.5. Natural history of idiopathic pulmonary fibrosis. Adapted from (Kim et al. 2015).

1.4. Diagnosis

Within the UK, specialist tertiary referral centres have been introduced alongside a National Institute for Health and Care Excellence (NICE) clinical guideline to support the diagnosis and management of IPF (2017). As outlined in the recently updated International consensus diagnosis guidelines agreed by the American Thoracic Society (ATS), European Thoracic Society (ERS), Japanese Respiratory Society (JRS) and the Latin American Thoracic Society (ALAT), the gold standard of IPF diagnosis is through multi-disciplinary team (MDT) discussion, with the clinical team, radiologist and when applicable, pathologist (Raghu et al. 2018). In the context of a classical radiological pattern associated with IPF and no clear alternative cause, a diagnosis of IPF can be made by MDT consensus without the need for lung biopsy.

These International diagnostic guidelines suggest biopsies should be performed where diagnosis remains uncertain from HRCT. However, this is a cause of debate internationally given the high 30-day post-operative complication rate and 3-4% mortality (Kaarteenaho 2013) following video-assisted thoracoscopic (VATS) biopsy. There is an increasing use of transbronchial lung cryobiopsy (TBLC), with a reduced complication and mortality rate. Producing smaller sample yields, its diagnostic use across all ILDs remains uncertain, however in UIP specifically, results are suggestive of being similar to those gained via VATS (Zaizen et al. 2019). The first set of guidelines for the use of cryobiopsy in ILD have recently been published by expert panel agreement (Maldonado et al. 2020).

Average life expectancy for an IPF patient is 2-3 years (Ley et al. 2011) and whilst respiratory failure, often due to an acute worsening of the disease with an exacerbation of the disease (AE-IPF) is the leading course of death, up to 35% of deaths may be unrelated to respiratory failure (cardiac or noncardiopulmonary) (Daniels, Yi and Ryu 2008). This prognosis for IPF patients is significantly worse than many cancer outcomes with a review in America suggesting a 5-year life expectancy around 20% (Kim et al. 2006) (see Figure 1.6).

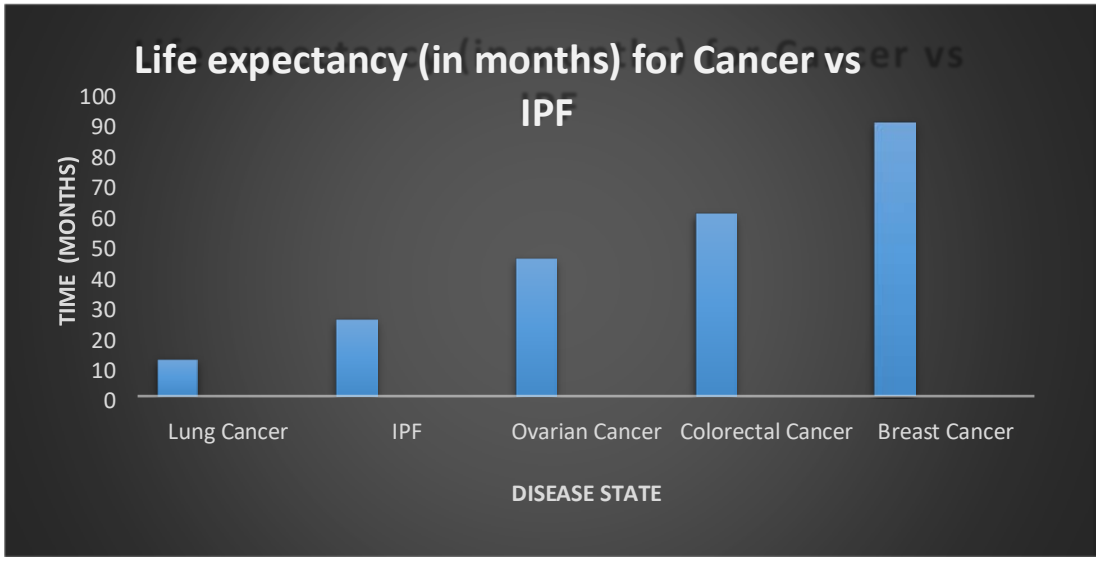


Fig 1.6. Life expectancy of patients with IPF vs different cancers. Adapted from (Kim et al. 2015).

To date, there remains no cure for this condition; two anti-fibrotic medications have been approved for the treatment of IPF (nintedanib and pirfenidone) and have both been shown to slow down lung function decline in large scale clinical trials and real-world data (Richeldi et al. 2014), (King et al. 2014, Rodríguez-Portal 2018). Care should focus on holistic supportive management with involvement of Pulmonary Rehabilitation (PR), Oxygen services, symptom control, early referral to lung transplant where appropriate and palliative care (Shaw et al. 2017).

As well as the extensive mortality and morbidity an IPF diagnosis presents, the healthcare costs are disparate to the disease prevalence. A Medicare analysis of an over 65-year-old population in the USA showed the healthcare costs both pre and post IPF diagnosis to be almost double that of an age matched non IPF cohort, with increased outpatient and emergency department visits (Morrow 2019).

1.5. CURRENT PROGNOSTIC TOOLS IN IPF

1.5.1 Health related quality of life (HRQoL) assessments in IPF

Until recently, few quality of life measurement questionnaires or disease specific instruments have been available to a treating physician to gain an accurate understanding of an individual's perception of their disease severity, how this can impact a patient's psychological wellbeing and also their ability to exercise or perform usual activities of daily living. The relationship between such measures and the standard functional testing used for disease severity (FVC, DL_{CO}) is also largely unexplored. In 2004, a systematic review by Swigris et al. (Swigris et al. 2005) concluded the variability between study

subjects was not fully explained by the patient's breathlessness or lung function results and an individual's QoL scores may provide unique information aside from clinical measurements.

The most widely used assessment of QoL in IPF patients to date is the St. Georges Respiratory Questionnaire (SGRQ). Developed in 1991 and containing a total of 50 items, the questions are specific to the measurement of impact of obstructive airways disease, especially Chronic Obstructive Pulmonary Disease (COPD). The utility of SGRQ with IPF patients remains questionable and a review of its use in this population concluded further research was needed with regards its validity in IPF, (Swigris et al. 2014) although a small retrospective study undertaken in Japan has suggested a score >30 has an independent mortality prognosis in patients with IPF (Furukawa et al. 2017).

To overcome the inherent problems of the lack of disease specificity of SGRQ, 20 years after its first iteration, the original authors created an IPF specific SGRQ-I. The item number was reduced to 34 with different scales and weightings to the original and the removal of airways particular effects (such as wheeze). However, the prospective validation of this questionnaire remains unclear and its use globally within clinical trials or in daily clinic use is limited (Yorke, Jones and Swigris 2010).

More recently, Patel et al. developed an ILD specific questionnaire named King's Brief Interstitial Lung Disease Health Status Questionnaire (K-BILD) (Patel et al. 2012). With a much-reduced item burden, comprising 15 questions across 3 domains (breathlessness and activity, chest symptoms and psychological), this questionnaire has been validated across several ILD cohorts (Wapenaar et al. 2017), including large numbers of IPF patients. The condensed question load and time consumption allows its use across multiple clinical settings for quick assessment of a patients' own perspective on their disease. In 2019, K-BILD was used for the first time in a large phase III clinical study for a wider ILD cohort (Flaherty et al. 2019) and appears to now be the gold standard across fibrosing lung conditions with validated minimal clinically important differences (MCIDs) now established across different ILD cohorts (Sinha et al. 2019) for total and individual domain scores.

Finally, and as yet un-validated questionnaire across larger cohorts, is the Idiopathic Pulmonary Fibrosis - Patient Reported Outcome Measure (IPF-PROM). This set of questions has been further refined to just 13 questions across 4 domains (physical experience of breathlessness, psychological experience of breathlessness, emotional well-being and energy) (Russell 2017). This questionnaire remains the only IPF specific tool available for patient reported outcomes and assessment of health status and may add to our clinical knowledge of an individual's disease perception and potential course. Both questionnaires can be seen within the appendices (see appendix A and B).

1.5.2 Static measurements

As has been highlighted, predicting the prognosis to individual patients is extremely challenging. As such, a number of prediction tools have been designed and are proposed. Forced Vital Capacity (FVC), the amount of air a patient can exhale in a single breath (measured in ml) and specifically a decline in a patient's expected percent predicted FVC is now widely accepted in larger phase II and phase III clinical trials as a surrogate marker for mortality (Karimi-Shah and Chowdhury 2015). The magnitude of change measured in FVC over a period of time, to be suggestive of a poorer outcome, however, still causes debate today. Richeldi and colleagues have concluded a ≥10% decline in FVC over 12 months, not only signals a clinically meaningful change to an individual's lung capacity but also maintains prognostic accuracy (Richeldi et al. 2012). A further physiological test measuring the ability of the lungs to transfer gases across the alveolar:blood stream barrier, routinely performed by IPF patients is that of diffusion capacity (or Transfer factor) of the lungs for carbon monoxide (TL_{CO}). A number of studies have alluded to the idea that this measurement is more accurate than FVC at a given point in the patient's disease, and can predict with greater accuracy, the severity of IPF. Sharpe et al suggested TL_{CO} outperformed FVC % predicted when considering 12- and 24-month survival (Sharp, Adamali and Millar 2017) however, data from a patient registry in the US (Snyder et al. 2019) has proposed that a low FVC % predicted has similar prognostic predictive accuracy as a low TL_{CO} % predicted. Furthermore, the difficulty in the reproducibility of this test and the inability of some patients to perform the breath hold needed to undertake the measurement has meant TL_{CO} alone, is rarely used to predict survival.

A number of composite mortality prediction tools have been designed with the most widely used being the gender, age, physiology model (GAP). With a simple scoring system, this combines a patient's gender, age and certain physiological scores and places them in to three stages (I – III) to predict a survival probability (in months) (Ley et al. 2012) (Figure 1.7). Accurate on a population level, its accuracy for individual predication is naturally variable and has greatly limited its routine use with patients.

Predictors	1-Year Mortality			2-Year Mortality			All Follow-Up		
	HR	p-value	Points	HR	p-value	Points	HR	p-value	Points
Gender									
Male	1.06	0.810	1	1.11	0.610	1	1.13	0.528	1
Female	-	-	0	-	-	0	-	-	0
Age, years									
>65	1.52	0.123	4	1.49	0.078	4	1.54	0.056	4
61-65	1.07	0.839	1	1.18	0.576	2	1.19	0.558	2
≤60	-	-	0	-	-	0	-	-	0
Baseline FVC, % Predicted									
<50	4.27	0.002	15	3.29	0.001	12	3.14	0.002	11
50-75	3.32	0.001	12	2.35	0.001	9	2.37	0.001	9
>75	-	-	0	-	-	0	-	-	0
Baseline DL_{CO}, % Predicted									
Unable to perform	9.5	<0.001	23	7.37	<0.001	20	8.10	<0.001	21
≤35	2.89	0.054	11	2.77	0.018	10	3.05	0.009	11
36-55	1.77	0.282	6	1.74	0.170	6	1.83	0.135	6
>55	-	-	0	-	-	0	-	-	0
ΔFVC, % Predicted									
≤-10	3.46	<0.001	12	2.68	<0.001	10	2.78	<0.001	10
-10 to -5	1.65	0.074	5	1.57	0.050	4	1.63	0.031	5
>-5	-	-	0	-	-	0	-	-	0
History of Respiratory Hospitalization									
Yes	3.94	<0.001	14	3.84	<0.001	13	3.83	<0.001	13
No	-	-	0	-	-	0	-	-	0

Figure 1.7. Longitudinal GAP staging score and probability of survival by stage. Adapted from (Ley et al. 2012).

One recent addition to the traditional GAP scoring, was suggested by Mikolasch et al (Mikolasch, Sahota and Garthwaite 2018). The tallying of the standard GAP score, in addition to the Neutrophil: Lymphocyte ratio (NLR), previously shown to be prognostic in certain cancers (Ren et al. 2019), may provide improved projective accuracy across a broad IPF patient spectrum with a high NLR predictive of earlier mortality (Figure 1.8).

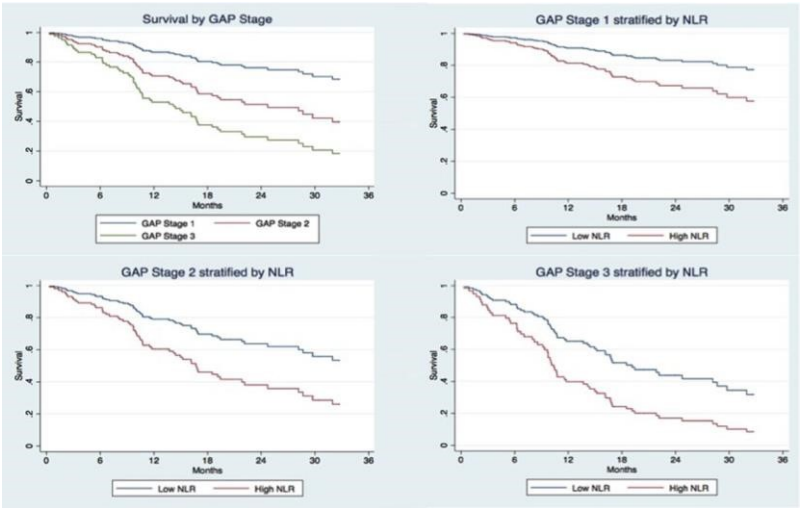
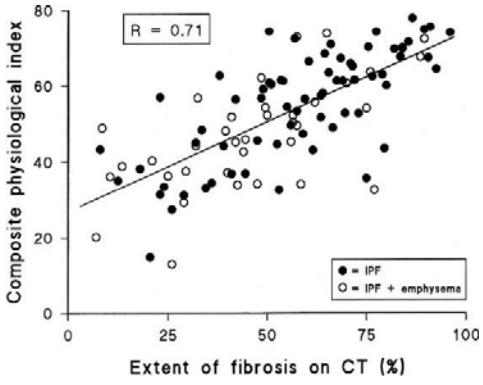


Figure 1.8. GAP stage NLR stratified Cox proportional hazard survival curves. Shown with permission (Mikolasch et al. 2018).

In this small study (n=75), the authors concluded a high NLR in IPF patients doubled the possibility of death in the follow up period. This was independent of GAP score, but it is suggested this biomarker ratio, in conjunction with the GAP scoring can refine and add accuracy to prediction of an individual's prognosis.

Largely now superseded by the arrival of CPI (described below), the clinical–radiographic–physiologic (CRP) scoring system uses seven variables (dyspnoea, chest HRCT, spirometry, lung volume, TL_{CO}, resting alveolar-arterial PO2, and exercise O_2 saturation). This score was initially devised to assess the severity of disease and suitability for transplant whereby each variable was graded in to between 2 – 9 levels of severity (dependant on the range of data available). The variables were subsequently weighted to ensure equivalence across each (Watters et al. 1986). The CRP has more recently been adapted and evaluated to predict patient survival (King et al. 2001) although remains little used in clinical or research practice.

The Composite Physiological Index (CPI) (Wells et al. 2003) is a second relatively simple prognostic scoring system, which may have greater accuracy than GAP staging when comparing mortality data to three years (Lee et al. 2018). The CPI not only includes physiology results but also those from HRCT, importantly taking in to account the co-existence of emphysema and fits disease extent against pulmonary function testing, providing a more accurate prognostic assessment that pulmonary function tests alone (Figure 1.9).



The formula for the CPI was as follows: extent of disease on CT = 91.0 - (0.65 x percent predicted diffusing capacity for carbon monoxide [DLCO]) - (0.53 x percent predicted FVC) + (0.34 x percent predicted FEV1).

Figure 1.9. Composite Physiologic Index relationship to Disease Extent Observed by Computed Tomography. Shown with permission (Wells et al. 2003).

The final and most recently published tool for IPF prognosis is a staging instrument developed by Torrisi et al.; the Torvan Index (Torrisi et al. 2019). The calculation of a scoring system assigned to pre-existing comorbidities in addition to the standard GAP scores, as seen in figure 1.11 below, has

provided significantly improved accuracy of performance across two independent cohorts although its validation & use in large scale clinical trials remains uncertain.

Predictors	Points (sparse model)		Points (full model)	
Age years				
⩽55	0		0	
56–70	6		6	
>70	9		8	
FVC % pred				
>80	0		0	
61–80	1		1	
⩽60	6		6	
***D*LCO % pred**				
>60	0		0	
31–60	6		6	
⩽30	8		7	
Unable to perform	9		9	
Diabetes mellitus	1		2	
Systemic hypertension	/		1	
GORD	1 (absence)		2 (absence)	
Pulmonary hypertension	2		2	
Major depressive disorder	1		3	
Lung cancer	6		6	
Valvular heart disease	5		6	
Atrial arrhythmias	6		6	
Points (both for sparse and full models)	<14	14–16	17–22	⩾23
TORVAN stage	I	II	III	IV

Points are assigned to each variable in order to obtain a final score. Based on that, patients can be grouped into four different staging categories. FVC: forced vital capacity; *D*LCO: diffusing capacity of the lung for carbon monoxide; GORD: gastro-oesophageal reflux disease.

Figure 1.10. Torvan scoring system and staging. With permission (Torrisi et al. 2019).

Although relatively quick and easy to perform, the true prognostic value given to these static measurements remains unclear and it is likely an individual's dynamic longitudinal changes, for example dyspnoea, HRCT changes and physiological changes over a given time period, will allow a much greater accuracy of a likely survival estimate (Kolb and Collard 2014). Furthermore, whilst there remains no consensus globally as to the preferred predictive tool of choice, clinicians will be making their individual choice for preference and communication of such information to patients will remain variable.

1.5.3 Dynamic measurements

A small number of dynamic exercise tests have been utilised to assess IPF outcomes including Cardiopulmonary Exercise testing (CPET) (Fell et al. 2009), 4 metre gait speed (4MGS) (Nolan et al. 2019) and the Six Minute Walk Test (6MWT) (du Bois et al. 2011). Routinely used around the world, patients undertaking the 6MWT are asked to shuttle walk as far as they can within a six-minute period. Total distance covered (in metres) and also change in distance over time are two outcomes shown to have good prediction of mortality. Du Bois et al showed a reduction of 50m over a 24-week period to have a 4-fold increase in death within a 12-month period (du Bois et al. 2011), whereas Caminati and

colleagues showed patients walking less than 212m in 6 minutes had a poorer outcome (Caminati et al. 2009). Whilst this test has a number of potential benefits in that it has widespread availability, is easily reproducible and inexpensive with limited experience needed from the physiologist perspective, there are a number of potential pit falls. Patient variability in terms of test effort will have a significant effect on the outcome results; fragility, age, gender and a patient's cognitive function can all provide limitations to its usefulness (Heresi and Dweik 2011).

A more recently explained dynamic test is the 4 MGS. Awaiting further validation, early data whereby the time of an individual to walk a 4-meter distance is recorded has suggested 4MGS correlated significantly with 6MWT and total KBILD although its prognostic use is unexplored.

Cardiopulmonary exercise testing has been extensively utilised within secondary care settings but to date, there is limited knowledge of its use in the IPF setting. The test itself allows assessment and causes of limitations to exercise to be identified via numerous gaseous exchange measurements during exercise stress to the ventilatory, circulatory and muscular systems. In addition to conventional exercise testing, CPET allows for the measurements of breathing (ventilation) and gas volumes, both inhaled (oxygen) and exhaled (carbon dioxide) to be measured. This provides the clinician with a multi organ view of the transport and delivery of oxygen to the muscles (mitochondria) and its use during activity (Figure 1.12).

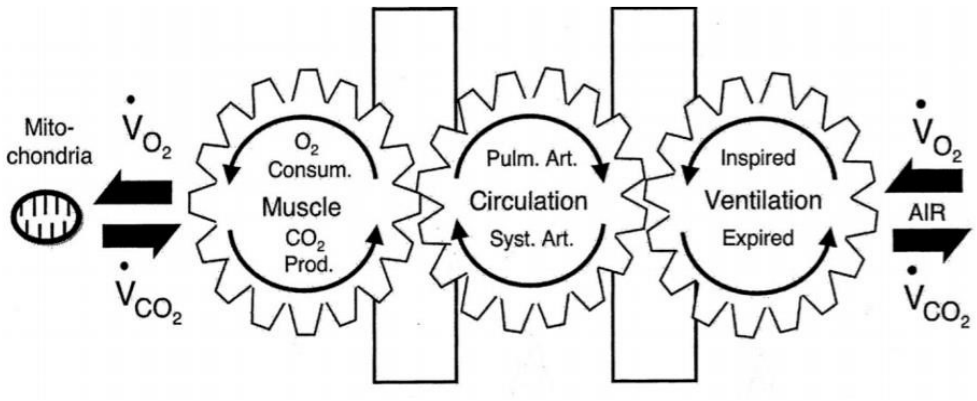


Figure 1.11. Gas exchange and O_2 utilisation during CPET testing (Wasserman 2012).

CPET has been performed for over 50 years (Schraufnagel and Agostoni 2017) although for much of this time, only in adults with single disease to aid the evaluation of dyspnoea. Today it is routinely used to assess a much more heterogeneous population, from young to old and fit to frail across numerous disease states and those with multiple comorbidities. Its use within the field of cardiology is perhaps the most widely explored for evaluation of prognosis, diagnosis and assessment of risk in

patient known to have cardiac disease (Albouaini et al. 2007, Akinpelu 2018). However, its reach extends well beyond assessment of the heart and it is now widely used in the pre-operative assessment to aid peri-operative planning and assess a patient's suitability and potential survival of surgery for organ transplant, especially of the heart (Ong et al. 2000) and lung (Dudley and El-Chemaly 2012). The standardisation of procedures, including COPD, cystic fibrosis, Exercise Induced Bronchospasm (EIB) and Lung Volume Reduction Surgery (LVRS) among others were outlined in International guidelines published jointly by the American Thoracic Society (ATS) and American College of Chest Physicians as far back as 2002 (Society and Society 2002). More recently in 2019, the European Respiratory Society published an updated standardisation of CPET with a greater focus on the procedural aspect across all chronic lung diseases (Radtke et al. 2019).

Over the course of a CPET, usually lasting between 8-12 minutes, thousands of measurements will be taken from each breath and heartbeat which will be averaged out over a short period of time (often 30 seconds) and plotted and presented on a 9-panel plot (Figure 1.13) which can essentially be split in to four categories:

1) Work rate (Watts)

2) Gas exchange (including O_2 consumption VO_2, CO_2 production VCO_2 and the output of VCO_2 / VO_2 known as the Respiratory Exchange Ratio or RER)

3) Ventilatory Assessment (including ventilation rate V_E, ventilatory equivalents for O_2 and CO_2 - V_E/VO_2 and V_E/VCO_2 and oxygen saturation)

4) Cardiac Assessment (including Heart rate, blood pressure and ECG changes)

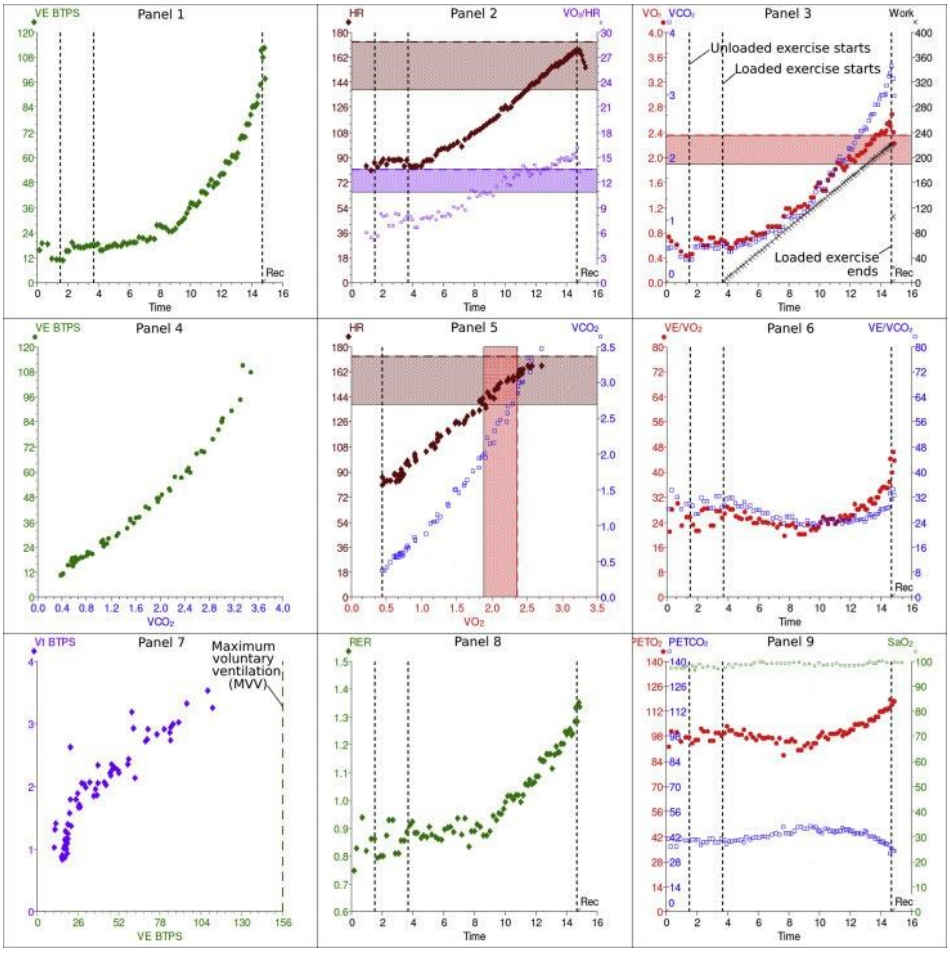


Figure 1.12. Example of 9-panel plot in a healthy individual. This presentation of outputs was described by Wasserman and colleagues in 2012 (Wasserman 2012) with 30-second averaging of data. Panels 2 and 3 present circulatory parameters, panels 5 and 9 represent ventilatory parameters whilst 4, 6 And 7 represent ventilatory efficiency. Panel 8 (RER) shows metabolic changes. AT is most easily calculated in panel 3 with the divergence of VCO_2 against VO_2.

Experience of using such data presentation will lead a physiologist or clinician to have focus on these different plots dependent upon the measurement they have an emphasis on. Generally, panels 2,3 and 5 represent the cardiac outputs, the ventilatory system is shown in plots 1, 4 and 7 whilst the remaining graphs (6, 8 and 9), represent the ventilation-perfusion relationship.

As the work requirements increase with load, so do the muscles energy needs, initiating the use of anaerobic metabolism as the oxygen supply cannot meet the muscles metabolic demands. This causes a rise in blood lactate concentrations alongside metabolic acidosis. The resulting hydrogen ion generation in the cells (from lactic acid dissociation) are buffered by bicarbonate resulting in a sharp rise in VCO_2 when compared to VO_2 (which until this point will rise at a similar rate). This point, which is most clearly seen in panel 3 in figure 1.12 utilising the VCO_2 versus VO_2 is known as the Anaerobic Threshold (AT) and is often observed at 50-60% of VO_2 peak. This method of calculation is the V Slope method and is utilised in our study. This point is important as it can be representative of an individual's

cardiovascular fitness and its observation (or lack of) can often distinguish between the cause of exercise limitation being cardiac or ventilatory associated (Albouaini et al. 2007).

Essential to the interpretation of CPET results is the patient's attainment of a maximal effort test and the parameters to assess this are outlined later in the thesis. The test itself can take place on a treadmill with the patient walking/running or on a cycle ergometer. Often the use of a treadmill can provide greater VO_2 peak values, simply by the greater muscle volume being worked, but it may not be suitable for all. The choice of each is often made by the physiologist and may be dependent on patient preference or indeed ability/fragility.

Despite the abundance of data derived from CPET it does come with potential issues when compared to more widely used functional tests. Costs in an ever increasing financially restricted health service, including both equipment and staffing per test are well in excess of 6MWT and standard LFTs. Although serious adverse events during CPETs are rare, this must still be considered on an individual basis. In a large Korean retrospective cohort of nearly 1500 elderly patients, serious cardiovascular complications occurred in 0.2% (Kim et al. 2019). Finally, and more pertinent to the current day, infection control and associated time and costs is a consideration.

Previous studies with the use of CPET across all ILD subtypes have identified declining functional capacity and muscle weakness as strongly predictive of disease progression and increased mortality (Panagiotou, Polychronopoulos and Strange 2016), whilst measures of gas exchange may be more valuable predictors of outcome than measures of lung mechanics (Lederer et al. 2006, Flaherty et al. 2006, Ley et al. 2011). Nevertheless, further efforts to develop definitive prediction models are required for clinical practice (Kolb and Collard 2014, Ley et al. 2011). Our current knowledge on the prognostic use of CPET in IPF patients is further explored in Chapter 2.

Chapter 2: CPET in IPF, our current understanding

2.1 Introduction

As discussed in section 1.5.3, cardio-pulmonary exercise testing (CPET) provides a comprehensive assessment of the physiological changes in the respiratory, cardiovascular, and musculoskeletal systems in a controlled laboratory environment (Layton et al. 2017), that has shown promise in terms of prognostic value in a number of chronic respiratory conditions (Ferrazza et al. 2009, Arena and Sietsema 2011). My hypothesis that CPET variables could be more sensitive to change in patient's health status than the more traditional lung function parameters (FVC, TL_{CO} or 6MWT) arose from the volume of data suggesting other potentially useful markers for example, progression of fibrosis on CT bare no correlation with these traditional (and globally accepted) lung function measurements (Clukers et al. 2018), (Hayton et al. 2019). Furthermore, patient reported outcomes within larger IPF cohort studies has shown no correlation to lung physiology measurements (Richeldi et al. 2014, King et al. 2014, Noble et al. 2011). Whilst CPET has been widely used across other disease states to predict clinical outcomes and prognosis, especially among cardiology, oncology and suitability for transplantation (Patel et al. 2019, Kleber and Köln 2018, Ney et al. 2016), relatively little is known about its role in IPF. I was keen to confirm a belief that not only such a test was safely achievable in an IPF patient population but also, that relationships between CPET outcomes, patient's prognosis and self-perceived QoL existed.

To provide a detailed understanding of the currently available data on the longitudinal use of CPET as a prognostic tool in IPF patients and in order to gain insight in to difficulties previously encountered by research teams around the world in such study cohorts, a formal systematic review was undertaken to better inform the study protocols. The primary objective of this search was two-fold. Firstly, to better comprehend the safety of this test in our likely cohort, enabling the formulation of the inclusion and exclusion criteria and understand the feasibility of recruitment in a mild to moderate population. Secondly, to gain an insight in the current knowledge gaps in the prognostic use of CPET in both IPF and the wider ILD umbrella diseases and evaluate the current understanding of CPET in predicting disease-specific outcomes in long term follow up of ILD populations. The outcomes of this review were utilised to formulate a clear understanding of published data, inform our hypothesis and ultimately finalise the study design. Should the hypothesis be correct and a prognostic role for CPET be confirmed, it could be used to guide earlier intervention for at-risk patients, support cohort enrichment for ILD clinical trials and allay anxiety and unnecessary monitoring amongst patients with stable ILD.

A brief online non-systematic search of the literature suggested small numbers of studies in IPF and thus the decision was made to include all ILDs and not just IPF to broaden the potential reader interest of the review.

To explore the data available, we undertook a full literature review of all studies where CPET variables had been used to estimate prognosis across all ILDs. The study selection (according to PRISMA statement) can be seen in appendix C, however, of the 946 articles identified by the search criteria and subsequent reduction to 658 after duplication extraction, only 18 papers went through to full review. By far the dominant reason for rejection was the inclusion criteria not being met, but other reasons included the lack of longitudinal data, congress abstracts, and the primary disease state being studied was not ILD, see figure 2.1.

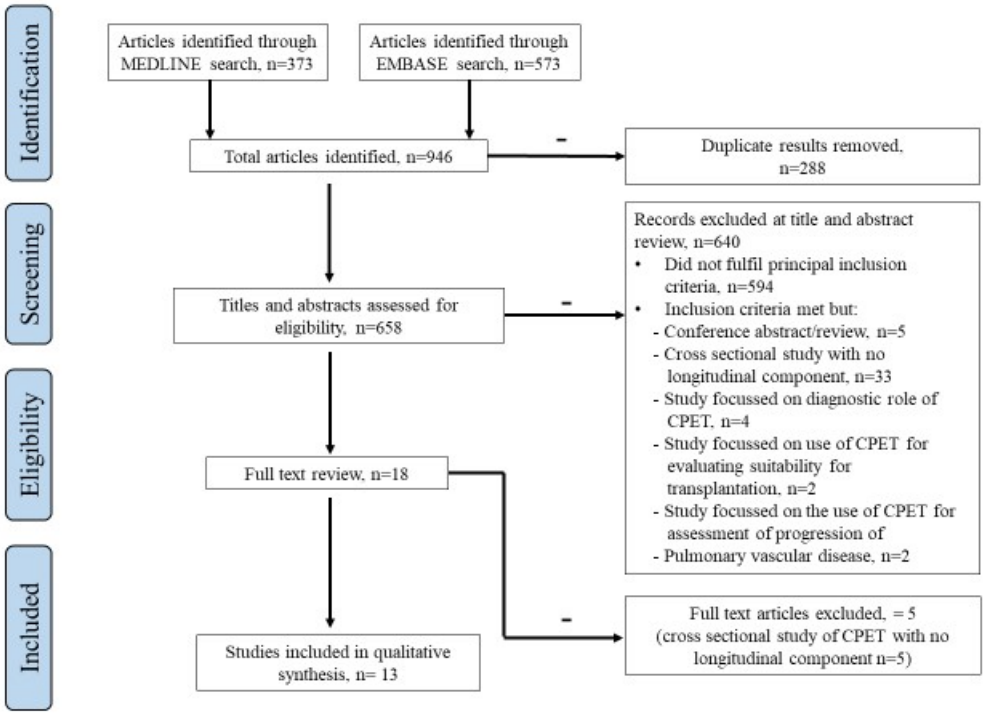


Figure 2.1. Study selection flow diagram presented according to PRISMA statement.

After full review of 18 papers and exclusion of a further 5 studies due again, to the lack of a longitudinal component to the research, only 13 studies remained. Of the studies included, 8 were specifically focussed on IPF patients although only 2 provided prospective analysis. The full published article of this review, including data from non-IPF cohorts can be seen in appendix D.

2.2. Materials and methods

The protocol for this review was prepared in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Shamseer et al. 2015) and registered in the

International Prospective Register of Systematic Reviews (PROSPERO 110198/2018). The benefits of utilising PRISMA provides a standardised structure to the writing, allowing critical review of the strengths and weakness', allowing direct replication of the methods undertaken at a future point in time. Furthermore, is indicative of the quality of the work undertaken (Moher et al. 2009). Prospective registration of the review on PROSPERO gives a published record of the planned work to be undertaken, potentially reducing the risk of bias and importantly, avoiding duplication of work (Stewart, Moher and Shekelle 2012). In line with PROSPERO registration and the need with our initial findings on the limited data available for IPF populations alone, the review team made an amendment to the protocol to include the term 'ILD' rather than 'IPF', furthermore, the allowance of non-English language publications was accepted to incorporate all available data.

2.2.1 Eligibility criteria

Studies that reported the relationship between CPET assessment and disease progression, prognosis or the presence/emergence of specific clinical outcomes of ILD were included.

Using the PICO framework outlined below, we evaluated publications that fulfilled the following criteria:

***P**opulation*

Adults (18 years or older) with a diagnosis of ILD (including but not limited to idiopathic pulmonary fibrosis, CTD-related ILD and sarcoid-related ILD).

***I**ntervention*

Studies reporting the outcome of CPET assessment as a prognostic factor. All available methods of 1) performing formal CPET and 2) reporting CPET results were included.

***C**omparison*

Patients with/who developed relevant outcomes (see below) were compared with those who did not, using CPET testing at baseline in both groups.

***O**utcome measures*

The primary objective was to evaluate the prognostic value of CPET in predicting disease course and outcomes in longitudinal (retrospective or prospective) studies of ILD. The relationship between CPET results and a number of clinically relevant outcomes including, but not limited to, relevant clinical

phenotype and disease demographics (e.g. disease duration, gender, age, lung physiology), disease outcomes (e.g. death, hospitalisation), surrogates of disease severity (including, but not limited to lung physiology, circulating biomarkers etc.), health-related quality of life (HRQoL) and functional status, were examined.

2.2.2. Study design

Eligible studies included cohort (retrospective or prospective) and observational longitudinal studies, that reported outcomes at a time point distinct from the baseline CPET (i.e. were of an appropriate design to evaluate prognostic value). The following types of studies were excluded: 1) animal studies 2) studies including patients with lung disease where an ILD cohort was not described and reported separately 3) studies designed to develop or validate health measurement scales 4) randomized controlled trials 5) case reports 6) qualitative research 7) non-original research publications (i.e., editorials, reviews) 8) abbreviated reports (e.g. letters to editors) and conference proceedings.

2.2.3. Search strategy

The search criteria were developed in accordance with search recommendations for systematic reviews of evaluations of prognostic variables (Altman et al. 2000). Electronic searches were performed in Medline and EMBASE, with no publication date or language restrictions. Full details of the specific search criteria can be seen in figure 2.2.

((Cardiopulmonary exercise test*) OR (cardiopulmonary exercise) OR (exercise test*)) **AND** ((idiopathic pulmonary fibrosis) OR (pulmonary fibrosis) OR (interstitial lung disease) OR (idiopathic interstitial pneumonia) OR (Cryptogenic fibrosing alveolitis) OR (fibrosing alveolitis) OR (Connective tissue disease-related interstitial lung disease) OR (Connective tissue disease-associated interstitial lung disease) OR (rheumatoid lung) OR (systemic sclerosis) OR (scleroderma) OR (polymyositis) OR (myositis)) **AND** ((cohort studies) OR (longitudinal studies) OR (case-control studies) OR (follow-up studies) OR (retrospective studies) OR (prospective studies) OR (incidence) OR (mortality) OR (follow-up studies) OR (prognos*) OR (predict*) OR (course) OR (prognostic) OR (prognosis) OR (progression) OR (future) OR (development) OR (outcome) OR (treatment outcome) OR (disease-free survival) OR (treatment failure) OR (morbidity) OR (mortality) OR (survival rate) OR (survival) OR (cause of death) OR (survival analysis)).

Figure 2.2. Specific search criteria performed in Medline and EMBASE.

All titles and abstracts generated by the search criteria were screened independently by myself and a second independent reviewer, identifying those studies relevant and eligible for full text review. Agreement between reviewers in the study selection process was assessed using Cohen's Kappa

statistics (Cohen 1968). This methodology used to assess reliability between those rating variables within data is now more widely used than assigning an arbitrary percentage score previously utilised, as it accounts for the possibility of chance agreement. The kappa score has a range from -1 to +1 although there is some disagreement on its agreement values in health studies, given a value of 0.41 is suggestive of assessor agreement (McHugh 2012).

Any discrepancies/disagreements within our review were resolved by discussion between reviewers and included a third party where necessary. Discussions between reviewers resolved any discrepancies at each stage of the study selection process. Review articles or editorials focussing on the "prognostic aspects of cardiopulmonary exercise testing in Interstitial Lung Disease" were also reviewed, to facilitate a grey search of cited manuscripts within these reviews.

2.2.4. Data extraction

A standardised form was used to independently extract relevant study details from each of the selected studies that included: date of publication, journal or publication source, study design, initial population of the study, study inclusion criteria, study exclusion criteria, CPET method, CPET analysis endpoints, disease outcomes assessed and a summary of key findings. Study corresponding authors were contacted when clarification was required. See Appendix C.

2.2.5. Risk of bias assessment

The QUIPS (Quality in Prognosis Study) risk of bias tool was used to assess the risk of bias within every included study (Huguet et al. 2013), see figure 2.3. Assessment in this way, scored across 6 domains places a high (red), medium (amber) or low (green) risk against the various categories to interpret the interrater agreement across the reviewed papers.

Authors	The Quality in Prognostic studies (QUIPS) tool assessment of bias					
	Study participation	Study Attrition	Prognostic factor measurement	Outcome measurement	Study confounding	Statistical analysis and reporting
King et al. 2001						
Miki et al. 2003						
Kawut et al. 2005						
Swigris et al. 2009						
Fell et al. 2009						
Wallaert et al. 2011						
Kollert et al. 2011						
Lopes et al. 2012						
Triantafillidou et al. 2013						
Gläser et al. 2013						
Van der Plas et al. 2014						
Vainshelboim et al. 2016						
Layton et al. 2017						

Figure 2.3. QUIPS tool assessment of bias.

2.3. Results

Study selection

Initial simultaneous searches in EMBASE (n=573) and Medline (n=373), performed in early 2019, identified a total of 946 articles. After removal of duplicates (n=288), 658 articles generated by the search were screened for eligibility and exclusion criteria based on titles and abstract review. There was moderate initial agreement between the two reviewers (Cohen's kappa 0.462 – see appendix E), with discordance in 20 abstracts, that was easily resolved through discussion. Eighteen articles proceeded to full text review and this led exclusion of a further 5 studies. A total of 13 studies were deemed eligible for inclusion. The full study selection process is detailed in figure 2.1.

2.3.1. Geographical participation and date of publication

Five studies were undertaken in Europe (5/13, 38%)(Triantafillidou et al. 2013, van der Plas et al. 2014, Gläser et al. 2013, Wallaert et al. 2011, Kollert et al. 2011), five in USA (5/13, 38%)(Fell et al. 2009, Kawut et al. 2005, Layton et al. 2017, Swigris et al. 2009, King et al. 2001) and the remainder in Israel (Vainshelboim et al. 2016), Japan (Miki et al. 2003) and Brazil (Lopes et al. 2012). The majority of studies were published in the last 10 years (10/13, 77%)(Layton et al. 2017, Triantafillidou et al. 2013, van der Plas et al. 2014, Gläser et al. 2013, Vainshelboim et al. 2016, Lopes et al. 2012, Wallaert et al. 2011, Swigris et al. 2009, King et al. 2001, Kollert et al. 2011) and only three studies published in the years preceding 2009 (Miki et al. 2003, Fell et al. 2009, Kawut et al. 2005).

2.3.2. Study characteristics

Most studies were retrospective cohort analyses (11/13, 85%), with variable follow-up periods (range 23 days (Kawut et al. 2005) - 20 years (Swigris et al. 2009)). The majority of retrospective studies evaluated independent risk factors for survival or mortality outcomes in ILD (9/11, 82%) and had an average follow up time of between 1-4 years (Gläser et al. 2013, Kawut et al. 2005, Layton et al. 2017, van der Plas et al. 2014, Miki et al. 2003, Vainshelboim et al. 2016, Triantafillidou et al. 2013, Wallaert et al. 2011, King et al. 2001). The longest planned follow up was in a study examining a non- IPF cohort (of systemic sclerosis ILD patients) which was truncated at 20 years (Swigris et al. 2009).

There were two prospective studies (Triantafillidou et al. 2013, Vainshelboim et al. 2016). One investigating the relationship between CPET and survival characteristics in IPF had a variable duration of follow up between 9-64 months (Triantafillidou et al. 2013). The other prospective study used CPET

as part of a wider investigation into the role of exercise testing in the prognostication of ILD and followed patients up for a fixed period of 40 months (Vainshelboim et al. 2016).

2.3.3. Study populations

Of the studies identified, 8/13 (62%) exclusively recruited patients with IPF, two recruited only sarcoidosis patients (Lopes et al. 2012, Kollert et al. 2011), and one study systemic-sclerosis associated ILD (Swigris et al. 2009). The remaining studies (2/13, 15%) evaluated more heterogeneous cohorts of ILD patients referred for lung transplantation assessment (Kawut et al. 2005, Layton et al. 2017).

The prognostic value of CPET has been retrospectively reported in a total of 703 patients with IPF, and prospectively in a further 59 patients in 2 small, single centre studies (n=25 (Triantafillidou et al. 2013) and n=34 (Vainshelboim et al. 2016)). Patients were recruited to studies according to consensus statements on the diagnosis of IPF available at the time of enrolment; the 2000 American Thoracic Society (ATS) international consensus statement for the diagnosis of IPF (Miki et al. 2003, Triantafillidou et al. 2013, van der Plas et al. 2014, Fell et al. 2009, King et al. 2001, ATS/ERS Statement 2000) and the later 2002 ATS/ERS International consensus classification of the idiopathic interstitial pneumonias (including IPF) (Wallaert et al. 2011, Kawut et al. 2005, Society and Society 2002). The updated 2011 ATS/ERS/JRS/ALAT evidence based guidelines for the diagnosis of IPF (Raghu et al. 2011) were applied in all (Triantafillidou et al. 2013, Gläser et al. 2013, Layton et al. 2017, Vainshelboim et al. 2016) but one of the studies (van der Plas et al. 2014) published after 2011 (the latter was a retrospective study that may have recruited patients prior to the publication of the 2011 guidelines).

We identified two retrospective studies that examined the role of CPET in predicting outcomes in mixed populations of ILD patients (Layton et al. 2017, Kawut et al. 2005). Cumulative patient numbers were small (a heterogeneous group of connective tissue disorders n=28, HP n=8, unclassifiable ILD n=7, sarcoid n=15, IIP n=21 (NSIP n=18, COP, DIP, COP). Whilst the cohorts could be considered to be representative of mixed ILD cohorts, patient numbers for each subtype were too small to consider each subgroup separately.

With regards to the study participant populations, the QUIPS risk of bias was considered to be low for only 3/13 (23%) studies (Triantafillidou et al. 2013, Wallaert et al. 2011, Vainshelboim et al. 2016), with the majority regarded as having a moderate (6/13, 46%) or high (4/13, 31%) (Miki et al. 2003, Layton et al. 2017, Kawut et al. 2005, Kollert et al. 2011) risk of bias. The generalisability of one study was potentially limited by the reported high diagnostic lung biopsy rate for IPF patients (64% (75/117) (Fell et al. 2009), as previously alluded to, a condition that can often be confidently diagnosed without biopsy in the presence of typical radiological findings and by consensus agreement in the multidisciplinary team setting (Walsh et al. 2016) and thus raising concerns as to whether this cohort

was representative of IPF populations in the 'real world'. Two studies examined disease outcomes that necessitated a particular baseline clinical phenotype e.g. recruitment from source populations referred for lung transplant evaluation and thus by definition only analysed selected cohorts of advanced ILD patients (Layton et al. 2017, Kawut et al. 2005). Others incorporated a *priori* patient grouping, for example the presence of pulmonary hypertension (Gläser et al. 2013), to enrich populations with patients at high risk of developing outcomes of interest, or required the active exclusion of patients with a relevant phenotype e.g. those that died from a cause other than respiratory failure (Miki et al. 2003).

Study attrition was generally reported to be low, which may reflect the retrospective nature of the majority of the studies identified. The QUIPS risk of bias for study attrition was reported to be high in two studies, increasing the potential for selection bias; >25% patients identified were excluded from the analysis by Lopes et al.(Lopes et al. 2012), whilst in the study by King et al. (King et al. 2001), 34% (80/238) of the originally identified population were excluded from inclusion in the final analysis because of incomplete data sets.

2.3.4. Prognostic factor measurement

CPET was the sole prognostic factor for the majority of studies 8/13 (62%), with a minority using CPET as part of a broader repertoire of exploratory physiological tests including 6MWT (Kawut et al. 2005, Triantafillidou et al. 2013, Layton et al. 2017) or lung function parameters (Gläser et al. 2013). One study used CPET in conjunction with clinical, radiological and resting physiological tests to devise a scoring system to predict survival in newly diagnosed cases of IPF (the CRP score: Clinical Radiological Physiological score) (King et al. 2001).

In two studies, CPET was used as the principal method to achieve a standardised form of maximal exercise (Kollert et al. 2011, Swigris et al. 2009) where upon arterial blood gas sampling or peripheral oxygenation measurements were taken to determine the effect of exercise on gas exchange. In both of these studies, typical CPET measures, such as maximal oxygen consumption (VO_{2max}) were not recorded.

Across all studies, the bias rating for prognostic factor measurement using the QUIPS tool was considered low-to-moderate (figure 2.3), with the majority of studies reporting a standardised approach to CPET and analysis that would be easily reproducible and less amenable to bias. Most studies provided a sufficient description of the CPET protocol used, adhering to the 2003 American Thoracic Society statement on cardiopulmonary exercise testing (Society and Physicians 2003) (6/10, 60%)(Kawut et al. 2005, Wallaert et al. 2011, van der Plas et al. 2014, Layton et al. 2017, Triantafillidou et al. 2013, Gläser et al. 2013). Others used the European Respiratory Society 1997 (Miki et al. 2003)

and updated 2007 (Vainshelboim et al. 2016, Palange et al. 2007) recommendations. In others important details were missing e.g. if oxygenation was measured during CPET (van der Plas et al. 2014). Variation in the methodological approach to CPET was also observed. For example, in one study, oxygen usage during CPET was an inclusion criteria (Layton et al. 2017), whilst in another, supplemental oxygen during exercise was supplied variably to participants depending on a pre-study requirement for home oxygen or saturation on room air <90% (Kawut et al. 2005). In 7/13 (54%) studies, blood gas analysis was used to assess the adequacy of gas exchange during exercise (Fell et al. 2009, Miki et al. 2003, Lopes et al. 2012, Wallaert et al. 2011, Kollert et al. 2011, Swigris et al. 2009, King et al. 2001), whilst the remainder used pulse oximetry, considered by some experts to be a suboptimal substitute (Society and Physicians 2003). A broad range of quantitative CPET parameters were presented/analysed raising the possibility of reporting bias (see later).

All but one study used cycle ergometry. Treadmill exercise testing was used as the method of CPET in the remaining study; in which exercise increments were chosen for participants based on patient's daily activities and parameters of resting pulmonary function, raising concerns whether a standardised approach had been adopted (Miki et al. 2003). Additionally, non-uniform speed increases, often inherent to treadmill testing, results in nonlinear metabolic rate increases and fundamental difficulties in calculating an accurate external work rate and an estimation of peak VO2. Thus, direct comparisons of peak VO2 obtained during treadmill testing studies cannot be compared with those obtained from cycle ergometry studies.

2.3.5. Outcome measurement

The most commonly reported outcome was mortality/survival 11/13 (85%). The majority of these studies that used survival/mortality as an outcome measurement (10/11, 91%) examined all-cause mortality, considering death or lung transplantation as composite endpoint. One study used an outcome measurement that was restricted to respiratory deaths only (Miki et al. 2003) and another study assessed the discriminatory ability of CPET to identify patients who would die on the lung transplant list before receiving transplantation (Kawut et al. 2005). Other outcomes included interceding pulmonary hypertension (PH) (Gläser et al. 2013) and decline in pulmonary function (FVC and DL_{CO}) or duration of immunosuppressive therapy in two non-IPF cohorts with sarcoidosis (Lopes et al. 2012, Kollert et al. 2011).

Using the QUIPS tool, the risk of bias in the approach to outcome measure assessment was considered low-to-moderate, in all studies.

2.4 Reported prognostic associations of CPET in the IPF cohorts

All studies reported at least 1 positive association between CPET and clinical outcomes, raising the possibility of positive reporting bias. Significant heterogeneity in study design, study populations (and classification criteria adopted), CPET protocols, CPET endpoints and defined endpoints precluded any useful attempt at meta-analysis.

The prognostic role of peak VO_2 has been examined across several studies of IPF. Fell et al. (Fell et al. 2009) retrospectively suggested a baseline threshold of peak VO_2 8.3ml/kg/min predicted survival in 117 patients with IPF (peak VO_2 <8.3ml/kg/min HR 3.24, CI 1.10-9.56, p=0.03). Patient numbers in the subgroup with peak VO_2 < 8.3ml/kg/min were small however (n=8, 7%), compared to the 46% patients that actually died, suggesting that the threshold sensitivity was not high. In another study, Triantafillidou et al. (Triantafillidou et al. 2013) prospectively identified a threshold of 14.2ml/kg/min for survival in 25 patients with moderate IPF (mean FVC 77.5 ±21.8), whilst Vainshelboim et al. (Vainshelboim et al. 2016) suggested VO_2 <13.8 ml/kg/min as one of 5 CPET parameter thresholds (peak work rate, tidal volume reserve, V_E/VO_2 nadir and V_E/VCO_2 at AT) predicting survival in a prospective cohort study of 34 patients with IPF. Finally, Gläser et al. (Gläser et al. 2013) identified that the presence of PH (invasively assessed by right heart catheter) and peak VO_2 % predicted were the only variables independently predictive of survival in a retrospective cohort of 133 patients, and application of % predicted values showed statically significant superiority to absolute data values. These results contrast with the findings of other studies where no independent association between survival in IPF and peak VO_2 has been demonstrated (Wallaert et al. 2011, Miki et al. 2003). Heterogeneity in terms of disease severity, follow-up periods and accompanying disease co-morbidity may have impacted on results of these studies and larger prospective studies are required to ascertain the prognostic role of peak VO_2 in predicting IPF survival.

Gläser et al. found that the development of interceding PH in IPF was best predicted by reductions in ventilatory efficiency, the V_E/VCO_2 $slope_{pred}$ (cut off of ≥152.4, AUC 0.938; CI 0.892-0.984), with a sensitivity of 87.2% and specificity of 88.4%, but analysis of PH subgroup alone did not identify any CPET parameters that provided independent prognostic information. V_E/VCO_2 at AT has also been shown to be a discriminating factor to determine the presence of PH across a cohort of IPF patients (adjusted OR 1.182; CI 1.029-1.384, p=0.021, n=81), but once again the prognostic value of this parameter has not been determined (Boutou et al. 2011).

The prognostic value of an alternative measure of ventilatory efficiency, the ventilatory equivalent for carbon dioxide at AT (V_E/VCO_2 at AT), in predicting survival in IPF has also been examined (van der Plas

et al. 2014). In a retrospective study of 38 IPF patients, those with V_E/VCO_2 at AT >45 had a significantly worse survival compared to patients with V_E/VCO_2 at AT ≤ 45 (HR 4.58, p=0.001), and this parameter remained a strong predictor even after correcting for functional severity of ILD, highlighting its possible use in the early detection of vascular impairment. Furthermore, the ventilator equivalent for oxygen at AT (V_E/VO_2 at AT) >45 was reported to be an independent poor predictor of 3 year-survival in a cohort of 63 IPF patients (Wallaert et al. 2011), findings consistent with the univariate analysis of Miki et al. (Miki et al. 2003). Results suggest that the magnitude of hyperventilation at ventilatory threshold may be determining prognostic value, but further prospective studies are required to confirm the value of these parameters of ventilatory efficiency in the prognostication of IPF.

Exercise induce hypoxaemia was also considered as a potential prognostic factor in IPF. Miki et al (Miki et al. 2003) found that only two factors, age and PaO_2 slope (defined as change in arterial oxygen pressure in mmHg / change in VO_2 uptake during exercise ($\Delta PaO_2/\Delta VO_2$)), provided independent prognostic information in a cohort of 41 IPF patients (HR 1.096, CI 1.012-1.187, p=0.025 and HR 0.841, CI 0.731-0.967, p=0.015 respectively) and stratification of patients according to this slope (≤60mmHg/l/min or >-60mmHg/l/min) identified significant differences in median survival (1.6 years vs 4.5 years respectively). Measurement of this parameter does however, require invasive arterial blood gas analysis during exercise testing, that is unavailable in the many clinical exercise laboratories. In the study by King et al. (King et al. 2001), PaO_2 at the end of maximal exercise was the only CPET derived parameter included in their comprehensive clinical-radiologic-physiologic scoring model to predict survival in IPF, and when weighted, accounted for as much as 10.5% of the maximum score in the complete model. Nevertheless, there were methodological limitations in this latter study; only 158/238 patients performed exercise testing and patients received supplemental oxygen when significant hypoxaemia ensued.

As a consequence of the utilisation of numerous different CPET parameters, CPET cut-off values, and timing of mortality evaluation, it was not possible to determine definitive thresholds for mortality or the development of pulmonary hypertension based on the analysed data.

2.5. Study confounders, statistical analysis and reporting across all studies

The majority of studies were considered to be at 'high' risk of bias due to inadequate account of potential confounding factors or methods of statistical analysis/reporting (figure 2.3).

The data used in the majority of studies was obtained from existing databases and/or case note review (n=11, 85%). As the data was not collected as part of a designed study, several potential confounders variables were not recorded, for example the presence of co-morbid disease (Wallaert et al. 2011, Fell

et al. 2009, Miki et al. 2003, Gläser et al. 2013, Lopes et al. 2012, Swigris et al. 2009), body mass index (Triantafillidou et al. 2013, Fell et al. 2009, Miki et al. 2003, Kawut et al. 2005, van der Plas et al. 2014, Lopes et al. 2012) and smoking status (Wallaert et al. 2011, Kawut et al. 2005, Gläser et al. 2013, Lopes et al. 2012).

The most important potential confounder was baseline 'disease severity' which was only specifically addressed as a confounder in one study; through the inclusion of lung function parameters and a composite physiological index (as markers of disease severity) into the Cox regression model used for analysis (van der Plas et al. 2014). This same study also stratified patients in an attempt to control for other potential confounders. Patients were sub-grouped into those with a systolic pulmonary artery pressure greater than or less than 40mmHg, in an attempt to control for interceding pulmonary hypertension, but this reduced subgroup sample sizes and thus may have reduced the statistical power to detect an effect.

Eligibility criteria were used to increase uniformity of study participants and reduce potential confounders. For example, two studies used participants referred for transplantation and thus by definition analysed distinct cohorts of more advanced patients but this selection bias reduced the generalisability of results (van der Plas et al. 2014, Layton et al. 2017). Other studies focused on healthier populations of ILD patients who did not need supplemental oxygen during CPET testing, but this, unsurprisingly, resulted in low mortality rates (n<10) leading to reporting bias (Vainshelboim et al. 2016, Fell et al. 2009, Triantafillidou et al. 2013).

Multiple regression analysis was the dominant statistical methodology used to determine the relationship between CPET parameters and clinical outcomes in ILD. Whilst this approach is generally considered to be one of the better statistical approaches to minimise unknown confounders, many of the studies reported on sample sizes much smaller than the minimum requirement for multiple logistic regression analysis as determined by Bujang et al. (Bujang, Sa'at and Sidik 2017). Furthermore, of all of the studies examined, only one detailed an *a priori* power calculation (Vainshelboim et al. 2016) and important consideration taken forward to the design of this study to better prove my hypothesis. Many studies were likely to be underpowered to detect the outcomes proposed. This research remains a feasibility study but may go some way to predict expected outcomes in future trials and aid such power calculations.

Stepwise multiple regression was used by some studies to determine the optimal model parameters to predict increased mortality (Triantafillidou et al. 2013, King et al. 2001). One criticism of this statistical approach is that model selection is conducted through parameter inference, which may lead to over-fitting of some parameters or exclusion of confounders that are not statistically significant

(Whittingham et al. 2006). Furthermore, the order of parameter entry (or deletion) and the number of parameters, can also affect the selected model (Derksen and Keselman 1992), whilst the multiple hypotheses tests, performed as part of this analysis, increases the probability of Type I error (Whittingham et al. 2006). The authors of one study did however attempt to overcome some of these limitations by checking for consistency between forward selection and backward elimination algorithms (Triantafillidou et al. 2013).

2.6. Discussion

Maximum oxygen consumption (VO_{2max}) is a measurement of the capacity for aerobic exercise and is determined by variables that define oxygen delivery by the Fick equation (Society and Physicians 2003); thus gas exchange across the lung, oxygen content of blood, oxygen delivery to tissues and oxygen uptake in the tissues can all affect the VO_{2max}. In healthy individuals, constraints of the cardiovascular system are most responsible for limiting VO_{2max} (Wagner 1996, Stickland et al. 2012). In patients with ILD, limitation to exercise may generally occur as a consequence of one of more of: 1) ventilatory mechanical limitation (unable to increase tidal volume (V_T) sufficiently and may reach their maximal predicted minute ventilation (% pred VE_{max})), 2) abnormal gas exchange (or reduction in ventilatory efficiency, indicated by variables such as the increment in minute ventilation (V_E) relative to carbon dioxide production (CO_2; V_E/VCO_2) 3) and/or diffusion limitation (indicated by variables such as reduction in oxygenation ≥ 4% or hypoxia at anaerobic threshold (AT)/peak exercise).

To my knowledge, this is the first study to systematically review and critically appraise studies that have reported the prognostic value of CPET in ILD. This field has gained recent attention with the majority of studies published within the last 8 years. Thirteen studies were identified that examined the prognostic value of CPET in ILD, all of which reported a prognostic role for CPET parameters in predicting clinical outcomes in ILD, with survival being the principle clinical outcome measured. Issues with study quality (relating primarily to the inherent problems of retrospective studies, patient selection and presentation of numerous CPET parameters), limits the strength of conclusions that can be drawn from the studies reviewed and thus whilst the associations presented shed important light to the potential role of CPET in disease prognostication in ILD, there is insufficient evidence at the moment to support its use in facilitating 'real-world' clinical decisions.

The exclusion of unpublished studies (e.g. conference abstracts) and abbreviated reports from this review may also increase the potential for publication bias, although this *priori* decision was taken to

ensure sufficient information was available to enable detailed data extraction from each study and aid this study design.

One published article was identified that described the prognostic value of CPET in IPF that was not originally eligible for inclusion in our study analysis due to the full text being published in French (Wallaert et al. 2011). As touched on, the decision was taken to amend our published protocol to include this study as the subject of the study was deemed to be important by independent reviewers.

This work has identified several considerations for future prognostic studies of CPET in ILD, including my research. Common to many human diseases, the disease progression in ILD is likely influenced by a complex interplay of patient, genetic, environmental and treatment factors. As such, a multivariable approach to the design and analysis of any future prognostic studies of ILD is essential if we are to confirm a specific role for CPET in routine monitoring. In contrast to randomised controlled trials, there are no robust standards defining the need to register or publish protocols for prognostic research and as such it is not always transparent whether statistical analysis were part of a *priori* plan (Hemingway, Riley and Altman 2009). Almost all studies in this review examined multiple prognostic CPET variables and as such there is potential for selective reporting bias that I intend to overcome by more stringent protocol registration with pre-specified outcomes of interest. A confounder overcome within my study by pre-identified CPET variables for evaluation.

2.7. Conclusion

The quality of existing studies on the role of CPET in the prognostication of ILD limits the conclusions that can be drawn from such work. Larger prospective studies are needed to establish the role of CPET in the longitudinal assessment of ILD in the future.

The review has however guided my study development and provided additional knowledge to the protocol. From the data available, CPET has been performed in a similar IPF cohort to the planned patient group from the Bristol ILD service. The team gained confidence in its safety in a moderate patient group (as defined by an FVC % predicted between 50-80%) and as a result, extended the inclusion criteria to include such patients. Unsure of the likely decline in exercise capacity of the more moderate group, the decision was made to only repeat CPET in the mild cohort and to enrich for this patient group (30 mild: 20 moderate) to reduce potential loss to follow up.

Whilst the majority of the studies are retrospective, and therefore open to suggestion of bias, the enrolment will be prospective in order to reduce this chance. Furthermore, this study will have predefined CPET parameters to avoid the previously mentioned selection bias of CPET outcomes. The use of patient reported QoL measures is limited throughout the literature review and to make this

study more focused on patient outcomes, the decision was made to include two different IPF specific questionnaires and assess the relationships of how the patient perceives their disease against clinical outcomes, making this study unique to those previously reported. Whilst many of the studies primary endpoints include mortality or time to transplantation, my initial discussions focused on correlations with the commonly used lung function outcomes of FVC and DL_{CO} % predicted. The former being widely accepted by international licensing authorities as an acceptable marker for mortality. As outlined in chapter 1, a great volume of data exists on the prognostic use of declines in these PFT measurements and initial thoughts were that a 12 month follow up may provide sufficient evidence for a confident prediction on the prognostic markers of CPET. However, the follow up times throughout this evidence base suggested this may not be long enough due to the heterogeneity of IPF. As a result of this, the decision was to increase to the total follow up to a maximum of five years to gain a better understanding of prognostic use of these outcomes.

Chapter 3: Feasibility of CPET in a mild to moderate IPF population

3.1. Introduction:

Hypothesis: CPET is feasible in a population of mild to moderate IPF patients and more sensitive to change in patient's health status than 6 Minute Walk Test (6MWT), Forced Vital Capacity (FVC) or Transfer factor for carbon monoxide (TL_{CO}), the routine clinical tests used globally today.

Hypothesis Generation: Whilst the use of exercise for the purposes of prognostication has been well explored in various IPF cohorts, this has very much focused on the use of a 6MWT, utilising either the patient's total distance or distance as a % predicted, as discussed in chapter 1. After initial discussion with the NBT physiology team, it was clear that, at the time, CPET was rarely used in the wider respiratory department and few, if any IPF patients were assessed in this way. At this point I initiated a literature review of all the evidence for the use of CPET as a prognostic tool in IPF and as seen in the previous chapter, *'The quality of existing studies on the role of CPET in the prognostication of ILD limits the conclusions that can be drawn from such work'.* Having been guided by the review and with questions outstanding from the initial hypothesis, the decision was made to proceed with a prospective feasibility study to gain a greater understanding of both the safety of a maximal exercise

test in this population as well as explore pre-determined parameters CPET can provide as to their discrete prognostic value for a disease inherently difficult to predict individual outcomes.

Secondary to the exercise testing was the addition of health status questionnaires. Numerous questionnaires have been used across the globe within IPF studies although to date, these have not been designed specifically for ILD or IPF patients. Permission was obtained from the authors of both K-BILD (Kings College Hospital NHS Trust) and IPF-PROM (Imperial College Healthcare NHS Trust) to better understand the relationship between exercise outcomes and an individual's perception of their quality of life and health status. The validity of K-BILD has been documented and indeed, the questionnaire is now widely used across numerous late phase ILD studies, however the same cannot be said of IPF-PROM. This study has been listed by the author (A.M Russell) as part of its validation and the hope is, with its increasing use in different IPF cohorts that longitudinal MCIDs can be attained for total and domain scores to better understand its meaning.

Research question: In order to test this hypothesis, the study set out to investigate the feasibility and safety of CPET in a population of mild to moderate IPF patients, a test previously not routinely considered for such patients and one for which a relatively small amount of data exists for its safety and outcomes. Secondly, with longitudinal follow up, the outcomes will provide a greater understanding and answer questions on the relationship of CPET parameters to more sensitive changes in patient's own perceived health status and allow comparisons to 6 Minute Walk Test (6MWT), Forced Vital Capacity (FVC) or Transfer factor for carbon monoxide (TL_{CO}), the routine clinical tests used globally today.

3.2 Methods

This prospective observational follow-up study was conducted at the North Bristol ILD Service located at Southmead Hospital, Bristol. Written informed consent was obtained from each of the study participants prior to enrolment and study participation. Subjects were recruited between June 2019 and May 2020 from the outpatient clinic environment and all have a multidisciplinary team meeting consensus diagnosis of IPF based upon the ATS/ERS/JRC/ALAT 2011 guidelines (Raghu et al. 2011). The reasons for this enrolment timeline were two-fold. Firstly, the deadlines given by the sponsor for study and MSc data collection completion, but also, from a retrospective search of IPF patient numbers attending clinics at Southmead Hospital and the author's personal time allowance for this study, it was felt 11 months would be necessary for successful enrolment of the planned numbers.

Due to the previously discussed data including IPF patients undertaking CPET, seen in Chapter 2, a cautious approach to patient enrolment to this study was undertaken to maximise patient safety. A high threshold of >50% TL_{CO} was included for entry to the study and resulted in a mean TL_{CO} % of 61%

at baseline. This was with the knowledge that non-exercise, interventional studies of IPF patients entering large phase III global trials have a mean TL_{CO} % between 40-50% (Richeldi et al. 2014, King et al. 2014, Noble et al. 2011). Results derived from this patient cohort may provide evidence of safety to allow more inclusive criteria when considering lung function parameters, especially TL_{CO} and could allow for the study of a more moderate to severe IPF population potentially providing results across a wider heterogeneity of patient types.

Two major amendments were submitted to the REC and subsequently approved for changes to the study protocol during the recruitment and subsequent follow up stages. Evidence became available (Thomas et al. 2019) that the aging lung can differ in speed in its functional decline. Given the average age of this cohort being greater than 70, this influenced the study recruitment. Considering this, the strict use of the FEV_1/FVC ratio cut off being above 0.7 (as a determining measurement of obstructive vs restrictive disease) was changed. The decision was made to alter the exclusion criteria to read '<70% unless within normal range for age (pre bronchodilator)'. Although two individuals had been screen failed prior to this change, it did allow inclusion beyond this accepted change (Table 3.1).

The second, and increasingly important adjustment to the protocol, was the study follow up time period. This major study amendment permitted follow up time for lung function and vital status to be observed for a period of up to 5 years. This decision was made given the heterogeneity of our study subjects disease course as previously discussed. Although always progressive, the disease path over 12 months, in this milder cohort may not provide the evidence of progression needed to make any firm conclusions on the many CPET variables available. The aim of an extended observation period will allow assessment of continued lung function testing (as part of standard clinical care) against the parameters of CPET tested at baseline and 12-month follow-up (where applicable). Given these results over an extended period, the aim is for this feasibility study to aid power calculations and application for a larger future study.

3.2.1 Study Population

Dependant up on an individual's FVC% predicted being above or below 80%, patients were sub divided in to a 'mild' or 'moderate' category. Although evidence for such staging based solely on % predicted FVC is limited, it is likely for accuracy, additional variables should be utilised (Kolb and Collard 2014), this provided an additional safety measure for follow up. Those patients with a milder disease would undertake both a baseline and repeat CPET at the 12 months follow up. Due to the uncertainty of the ability of those with reduced lung function to perform a maximal exercise test, the decision was made to take only baseline measurements from CPET at the start of the study. Key exclusion criteria were also primarily in place for reasons of safety and in line with ATS/ACCP guidelines (Society and Physicians 2003): Patients requiring oxygen treatment (due to the deficiency of safety evidence of undertaking CPET in this patient cohort), history of infarction within 6 months or unstable angina within 1 month, unstable cardiac disease & mobility issues that would impair exercise performance

resulting in a non-maximal test. The indications for exercise termination included acute myocardial infarction (MI) or suspicion of MI, onset of moderate-severe angina, serious dysrhythmias & at the request of the subject.

Key Inclusion Criteria	*Key Exclusion criteria*
Male or female aged ≥ 40yrs	**Cognitive behaviour/Inability to perform CPET**
Multidisciplinary team meeting consensus diagnosis of IPF based upon the ATS/ERS/JRC/ALAT 2011 guidelines	**Mobility issues**
Chest high resolution computed tomography (HRCT) and, if available, surgical lung biopsy pattern consistent with diagnosis of IPF	• **History of myocardial infarction (MI) within 6 months or unstable angina within 1 month.** • **Severe or untreated arterial hypertension (>200mmHg systolic at rest, >120mmHg diastolic**
FVC ≥ 50% and <80% predicted (Moderate) or ≥80% predicted (Mild)	**FEV1/FVC < 0.7 unless within normal range for age (prebronchodilator)**
DLCO ≥ 50% predicted	
	Patients using O2 treatment

Table 3.1. Key study inclusion/exclusion criteria.

Patients were followed up at 12 months ± 3 weeks from the initial exercise test, with a further 4 years of scheduled follow up to planned. This 12 month initial period of follow-up was chosen given the acceptance by global licensing authorities including the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) among others that the key lung function comparator (FVC) is a suitable marker of decline in IPF patients (Collard et al. 2003, Behr et al. 2015) and can be prognostic of mortality in this population over such a period of time (Zappala et al. 2010). Figure 3.1 below highlights the patient journey through the study protocol.

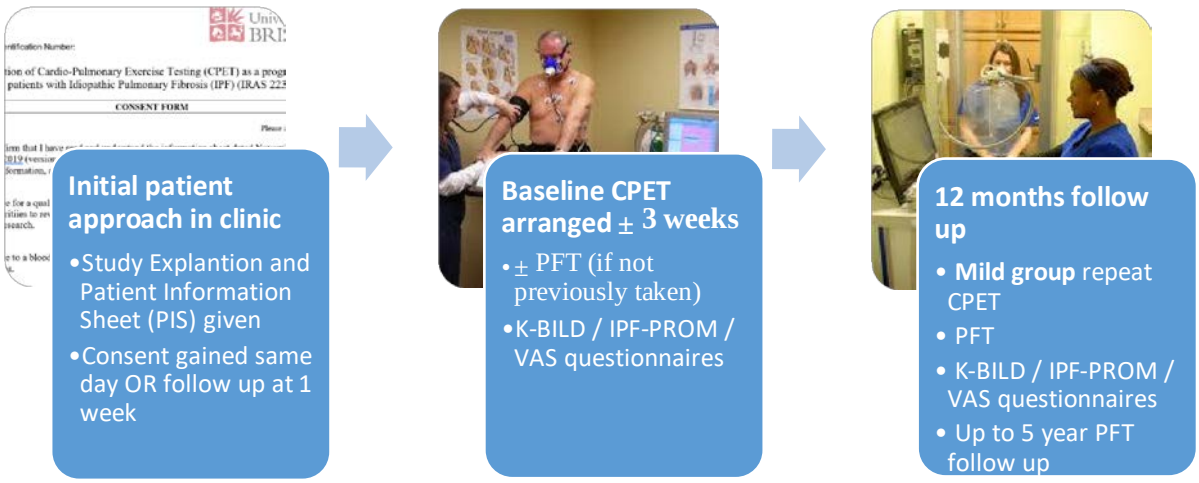


Figure 3.1. Patient flow for study involvement from initial approach in clinic visit to extended follow up to 5 years.

3.3. Ethics

The study was sponsored by the University of Bristol and was reviewed by the Bristol Interstitial Lung Disease clinical team at Southmead hospital, the research and Innovation department at North Bristol NHS Trust, The Health Research Authority (IRAS 223450), an independent research ethics committee and the Academic Respiratory Unit, University of Bristol. Following a major amendment to the study protocol to allow ≤ 5-year follow-up of data from routine clinical lung function appointments, all patients were asked for further written consent.

Potential participants were initially approached by members of their clinical care team, to brief them on the purpose of this research. At this time, all patients were reassured that there was no requirement to participate and decline for participation would not change their individual care plan. Following this and after an initial interest in study participation, a PIS was provided to each individual and a carer should it be requested, explaining the purposes of the research and the details of their study involvement. Each participant was provided the opportunity to discuss the study with a member of the clinical or research team should they wish. Should any of the patients at this point feel sufficiently informed and motivated to sign the consent form, they could do so. Otherwise, all were given opportunity to consider the study further. Those indicating interest in participating were given the opportunity to go away and consider their options further and contact the research team at a later stage should they be willing to participate (given maximum of 2 weeks). A member of the research team undertook a follow up phone call at one week after the initial approach. Consent was taken at the clinical trials unit or outpatients department on the Southmead Hospital site, by a member of the research team, for those that did not consent at the initial screening visit.

No patients were enrolled into the study that were (or had been in the previous 4 months) involved in an interventional study, including Clinical Trials of Investigational Medicinal Products (CTIMPs) of any description. Patients could partake if they were involved in questionnaire-based studies.

All data will be stored securely for a period of 5 years. Participants consented for their information to be stored in its anonymised form for this length of time. The cross-referencing list and data will be stored electronically on a Bristol University protected computer for a period of 5 years also. Beyond this date a new ethics application would be sought should it be required.

3.4. Lung Function Tests

3.4.1. Pulmonary Function Tests

Pulmonary function tests were undertaken by all patients ± 3 weeks from the date of CPET. All tests were performed in accordance with ATS/ERS guidelines (Graham et al. 2019) and repeat measurements taken after 12-months.

Included in the measurements recorded during pulmonary function testing were forced expiratory volume during first second of expiration (FEV1), forced vital capacity (FVC), total lung capacity (TLC) and single breath transfer factor for carbon monoxide (TL_{CO}) measured by an nSpire HDpft (nSpire Health GmbH, Germany). These chosen measurements make up the standard data set for routine clinical lung function tests within the hospital and can reliably be replicated over the full follow-up period.

Measured parameters were presented as actual values (litres) as well as percent predicted values of the European Community for Coal and Steel reference values.

All patients were asked to score their self-perceived breathlessness during their activities of daily living according to the Medical Research Council (MRC) dyspnoea scale (see Figure 3.2).

Grade	Degree of breathlessness related to activities
1	Not troubled by breathlessness except on strenuous exercise
2	Short of breath when hurrying on the level or walking up a slight hill
3	Walks slower than most people on the level, stops after a mile or so, or stops after 15 minutes walking at own pace
4	Stops for breath after walking about 100 yds or after a few minutes on level ground
5	Too breathless to leave the house, or breathless when undressing

Fig 3.2. The MRC breathlessness scale (adapted from (Fletcher 1952).

3.4.2. 6-Minute Walk Test

Study participants undertook a baseline and follow-up 6 MWT with a member of the Trust physiology team conducted in accordance with the American Thoracic Society guidelines (Laboratories 2002) in a 15m corridor at the hospital's respiratory outpatient unit. Measurements utilised for the purposes of this study were total distance achieved (m), % of theoretical distance walked alongside oxygen saturation at initiation of test and minimum saturation level during.

3.4.3. Cardio-Pulmonary Exercise Test:

All patients underwent a symptom limiting CPET that was performed and assessed by the respiratory physiology team at North Bristol Trust in line with the guidelines of the ATS/ACCP (Society and Physicians 2003), (ergoselect 100, ergoline GmbH, Germany).

Workload was determined primarily by reviewing the patient's normal intensities of exercise and previous levels of activity (with increases between 5W & 15W) but also additional factors including gender and body size in order to ensure a maximal effort was achieved within 8-12 minutes. Despite a recent ERS CPET standardisation statement suggesting a respiratory exchange ratio (RER) >1.05 (Radtke et al. 2019) is indicative of a maximal effort test, this has been contradicted by Thomas et al (Thomas et al. 2019, Thomas and Sylvester 2020). To be certain of a successful test, the team's decision was an RER ≥1.1 was required, although this also needed to be accompanied by at least one of the following markers (as determined by the physiologist) to confirm a maximum effort; maximum heart rate (HR max) > 80% of maximum predicted HR which was calculated by the sum of 220 – patient age, maximum minute ventilation during exercise >85% predicted based on maximum voluntary ventilation (MVV) at rest, and finally a plateau in VO_2 with an increased workload.

A decision for this study was made by the clinical and physiology team to use a cycle ergometer. This choice was based upon the perceived higher safety in this vulnerable patient group over a treadmill, although as suggested by Alessandro Mezzani (Mezzani 2017) in a recent published seminar, a number of other benefits may also be present (Figure 3.3).

	Treadmill	Cycle Ergometer
Higher peak oxygen uptake	X	
Easier implementation of ramp protocols		X
Possibility to quantify external work		X
Higher ECG quality		X
Possibility to obtain blood specimens during exercise		X
Higher safety		X
Possible use in supine position		X
Smaller size		X
Less noisy		X
Lower cost		X
Ease of movement		X
Greater experience in Europe		X

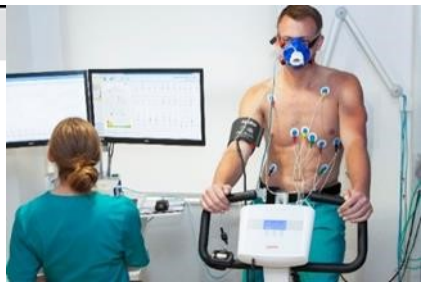


Figure 3.3. Advantages/ disadvantages of treadmill versus cycle ergometer for CPET.

Data collection and analysis were made by nSpire Zan 600 USB system (nSpire Health GmbH, Germany). The test could be discontinued at the discretion of the attending physiologist although no minimum $Sp0_2$ was pre-defined.

The protocol included a seated rest period for 3 minutes to allow the patient familiarisation with the equipment and apparatus (facemask and ECG probes), followed by unweighted peddling for the same time, allowing baseline VO_2 measurements to be attained. Subjects were asked to maintain a rate of 60 revolutions per minute throughout the exercise period. Anaerobic Threshold (AT) was calculated by the v slope method (VCO_2 vs VO_2).

Following cessation of the test, the subject remained seated with unweighted peddling and reduced cadence for a recovery period for a further 2 minutes (see Figure 3.4.)

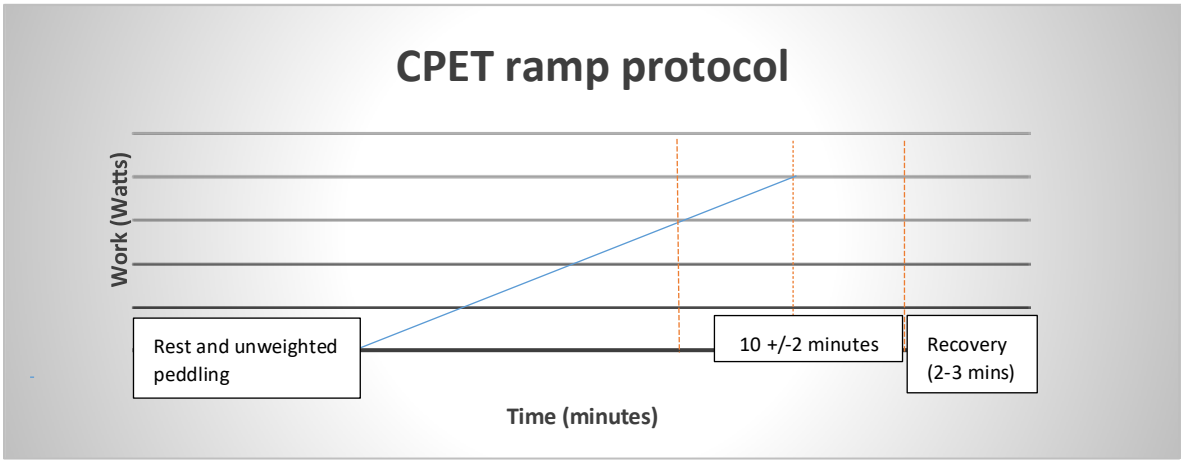


Fig 3.4. CPET ramp protocol with a target of 10 minutes to patient exhaustion.

Continuous measurements for the following pre-defined variables were made during the exercise period:

· Peak oxygen consumption (VO_2 peak, ml/kg/min),

· Peak oxygen consumption at anaerobic threshold (AT)

· Carbon dioxide production (VCO_2)

· Peak minute ventilation (VE peak) - (marker of ventilatory function during exercise),

· VE/VCO_2 slope as derived from the above values - (reflects changes in ventilatory drive)

· Peripheral capillary oxygen saturation SpO_2 - (marker of hypoxaemia indicating possible ventilatory limitation to exercise)

· Peak power output (W) - (marker of musculoskeletal function)

· Heart rate (HR) - (marker of cardiac function during exercise), ·

Breathing reserve (BR) - (potential indicator of ventilatory defect)

3.4.4. Questionnaires:

All patients completed two Visual Analogue Scores (VAS) namely a standardised day/night cough score and the Bristol VAS (measuring breathlessness and fatigue) (Yates et al. 2018a) as well as two ILD specific Quality of Life questionnaires (QoL), Kings Brief ILD (K-BILD) (Patel et al. 2012) & IPF-Patient Reported Outcome Measures (IPF-PROM) (Russell 2017). These were all repeated at the 12 months follow up, with the patients blinded to their baseline answers. Due to the lockdown of hospitals due

to COVID-19, those patients unable to attend study appointments for follow up, were asked to complete the questionnaires at home and return by post.

3.5. Primary & Secondary outcomes

Given the limited data of use of CPET in such an IPF cohort, this feasibility study was aimed to gain insights on the practicality of undertaking CPET in this mild to moderate group of patients, understand the likely attrition rates of study participants as well as the safety of such a strenuous exercise test in this study population. Secondary endpoints, gained from insights of previous analysis on the use of CPET prognostication in IPF, included the change from baseline at 12 months for CPET variables, correlation between CPET parameters and lung function changes as well as changes to health status (using questionnaires) over a 1-year period.

3.6. Statistical Analysis

Categorical variables are reported as absolute numbers and percentages. Normality of continuous data was initially verified using D'Agostino and Pearson normality test. Mean and standard deviation (SD) were used to describe parametric data; median and interquartile range (IQR, in brackets) for nonparametric data. Differences between two groups were verified by t-test with Welch's correction (continuous data), χ2-tests (categorical data) and paired t-tests for comparison in variables from baseline to 12 months. Pearson's correlation was used to determine correlations between parametric variables. Data were analysed using GraphPad Prism version 8.0. A P value of <0.05 was considered statistically significant.

3.7. Discussion

Ethics approval: 'Always allow longer than you expect' would be a take home message having been through the process. Although the IRAS application itself has been relatively straight forward, ensuring timely approval from both the sponsor and NHS centre was at times difficult and ultimately led to an approximate 6-month delay in study initiation. The author's presentation to the REC, including numerous questions from both lay members and healthcare professionals on the committee, surrounding employment within a pharmaceutical company was of particular interest, highlighting the high degree of cynicism of the industry as a whole.

One final delay in study initiation came as a result of the required Research Passport needed by both the sponsor and North Bristol NHS Trust to allow the investigator to undertake such research within

Southmead Hospital. The process for this remains unclear and is ill-defined. Thanks must go to individuals across many university sites who pointed me in a direction to someone who 'may know'.

Personal key learnings from the methodology have however been the initial hypothesis generation, functionality of the integrated research application system (IRAS) and local Regional Ethics Committee (REC) presentation. All of which were unknown at the start of this research.

3.7.1. Challenges of study design:

Study Funding: Although student study time was funded for the period of data collection, this was limited, and the hope was to gain approval for a National Institute for Health Research (NIHR) Clinical Research Network (CRN) funding to provide some additional nursing support from the academic research team. Whilst the precedent for such an exercise study in a similar patient cohort although lacking a longitudinal component had previously been set, our application was declined. A subsequent review was requested and despite approval for an amount of funds being awarded by the national respiratory CRN lead, this was later overturned. Although not essential to the setup of the study, it did place further pressure on timelines for the trawl of outpatient clinic appointments, patient interaction and introduction of the study and gaining of consent, resulting ultimately in an additional delay in patient recruitment.

Study Administration: Taking a large proportion of workload throughout the study period, the administration needed has provided great learning simply in the time allowance. Ranging from initial study documents (Patient Information Sheet (PIS), consent forms), ethics approval, Clinical Research Folder (CRF) through to the search of clinic appointments & naturally exercise appointments with the physiology team. The teamwork needed for larger studies has been very evident and whilst single handed management has been sustainable for such a feasibility study, it would not be recommended for anything much larger.

COVID-19 pandemic: Like every study ongoing around the globe, this research has been hugely disrupted by the virus. Due to study timelines and follow up, the last patient out was due in May 2020, with almost a third of repeat CPET test due between February and May 2020. Whilst the investigator was able to remotely contact all outstanding patients at the initiation of lockdown, enabling completion of questionnaires by post, there remains an amount of missing data. As can be seen from figure 3.1 and as will be discussed further in the next chapter, five patients in total (4 mild and 1 moderate) were lost to the study due to COVID-19. As a result of the temporary closure of all lung physiology testing and the subsequent capacity reduction upon reinstatement (due to the need to ventilate rooms post patient and requirement for PPE), routine lung function data was able to be collected in subsequent months with the median time in months (IQR) quoted. It was not possible to

undertake further CPET testing after the initial lockdown in early 2020. Study continuation was sought by the sponsor to enable the 5-year follow up amendment to take place (beyond its planned closure in May 2020). This delay in approval, due to the sponsors workload with COVID-19 and allowance of reconsent via email or post, led sadly to the loss of three patients to this extended follow up, who deceased in this interim period. One further deceased patient had given previous reconsent at 1 year follow up.

Chapter 4: Results

A total of 74 consecutive IPF patients were prospectively assessed for eligibility and subsequently approached for entry to the study. Screen failure and decline to partake excluded 32 patients, leading to enrolment and consent gained for 42 patients. Prior to study commencement and baseline testing, a further 4 patients withdrew consent, 5 patients developed exclusions (2 newly planned surgery, 2 undiagnosed cardiac arrhythmia and 1 frailty leading to inability to perform CPET). A further patient died prior to study initiation, see figure 4.1.

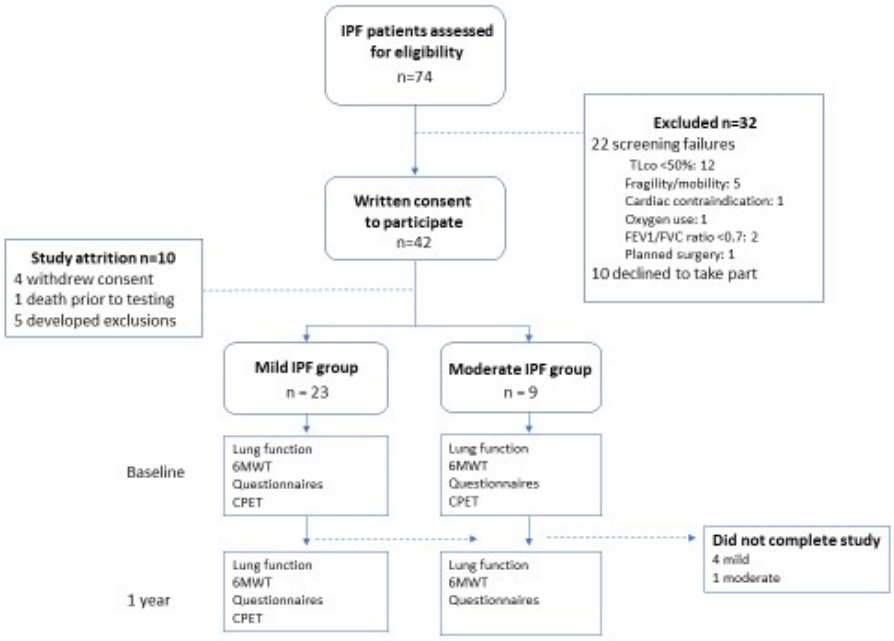


Figure 4.1. Study enrolment.

A total of 31 patients (23 moderate and 9 mild) underwent baseline testing, however due to the impact of COVID-19 previously discussed, a further 5 patients (4 mild and 1 moderate) were lost to follow-up, leaving a final population cohort of 27 patients (19 mild and 8 moderate). This patient group was

similar to those observed across other IPF study populations, with a male predominance (n=22, 82%) and a mean age of 75 years (SD ± 1.5 years). Although symptomatic at baseline, with a median Medical Research Council (MRC) breathlessness score of 2 (IQR 2-3), indicating that individuals feel 'short of breath when hurrying on a level or when walking up a slight hill', lung function was relatively well preserved with an FVC % predicted of 91% and gas transfer (TL_{CO}) of 61%. Approximately one third (33%, n=9) of the final study population (5 mild and 4 moderate) received antifibrotics during the 12month observational period. The most common comorbidities included gastro-oesophageal reflux disease (GORD), hypertension and coronary artery disease, with over half the patients (52%) having >1 comorbidity. At completion of 1 year follow up, all patients remained alive. See table 4.1.

Characteristic	**Overall n=27**	**Mild n=19**	**Moderate n=8**	**P value**
Age (yrs) mean (SD)	75 (+/- 1.5)	75 (+/- 1.5)	74 (+/- 1.5)	0.440
Gender (male n, %)	22, 82%	14, 74%	8, 100%	0.280
Smoking history (n, %) *Current* *Never* *Ex-smoker*	 0, 0% 10, 37% 17, 63%	 0, 0% 9, 47% 10, 53%	 0, 0% 1, 12.5% 7, 87.5%	0.552
BMI (kg/m)	28.5 (+/- 0.86)	28.9 (+/-1.2)	27.7 (+/- 1.0)	0.585
MRC (median, IQR)	2 (2-3)	2 (2-3)	2 (2-3)	0.964
Antifibrotics (n,%)	9, 33%	5, 26%	4, 50%	0.375
Co-morbidities				n/s
Gastro-oesophageal reflux *Hypertension* *Coronary artery disease* *Diabetes*	11 10 10 3	6 7 4 1	5 3 6 2	
Lung function parameters				
FVC (L)	2.96 (+/- 0.14)	3.11 (+/- 0.18)	2.61 (+/- 0.16)	0.050
FVC (% predicted)	91 (+/- 3.1)	99 (+/- 2.3)	70 (+/- 1.8)	**<0.0001**
FEV1/FVC ratio	78 (+/-1.5)	76 (+/- 1.8)	82 (+/- 2.6)	0.080
TL_{CO} % predicted	61 (+/- 1.7)	63 (+/- 2.2)	57 (+/- 2.2)	0.055
6MWT	**n=25**	**n=18**	**n=7**	
Distance achieved (m) % theoretical distance (m)	350 (+/- 12.7) 78 (+/- 3.3)	349 (+/- 15.7) 78 (+/- 3.9)	354 (+/- 22.1) 76 (+/- 6.4)	0.827 0.821
CPET	**n=27**	**n=19**	**n=8**	
VO_2 peak/kg (ml/kg/min)	20.9 (+/- 0.75)	20.6 (+/- 0.97)	21.7 (+/- 1.11)	0.489
VO_2/kg at AT (ml/kg/min)	13.6 (+/- 0.67)	13.8 (+/- 0.83)	13.0 (1.13)	0.585
VE peak (L/min)	69.9 (+/- 4.07)	69.0 (+/- 5.00)	72.1 (+/- 7.38)	0.731
VE peak (% predicted)	74.5 (+/- 2.8)	71.0 (+/- 3.2)	82.9 (+/- 4.4)	**0.045**
VE/VCO_2 at AT	28.2 (+/- 0.59)	28.7 (+/- 0.76)	27.2 (+/- 0.83)	0.201
Minimum O_2 saturation during CPET (%)	91 (+/- 0.9)	91 (+/- 1.2)	92 (+/- 1.0)	0.723
Peak work rate (W)	104.8 (+/- 5.08)	103.6 (+/- 6.74)	107.6 (+/- 6.62)	0.673
Peak work (% predicted)	44 (+/- 1.7)	44 (+/- 2.2)	43 +/- 2.2)	0.565
HR (bpm)	141 (+/- 4.2)	141 (+/- 4.9)	139 (+/- 8.7)	0.793
HR (% predicted)	97 (+/- 2.9)	98 (+/- 3.4)	95 (+/- 5.7)	0.711
BR max (L/min); median (IQR)	26.5 (18.4-32.5)	28.8 (18.8-33.3)		0.083
K-BILD questionnaire	**n=27**	**n=19**	20.4 (12-24.9), n=7 **n=8** 60 (+/- 2)	

Total *Psychological domain* *Breathlessness and activity domain* *Chest symptoms domain*	65 (+/- 2) 69 (+/- 3) 56 (+/- 2) 78 (+/- 3)	64 (+/- 2) 69 (+/- 4) 52 (+/- 3) 76 (+/- 4)	 61 (+/- 3) 52 (+/- 3) 71 (+/- 5)	0.058 0.128 0.267 0.133
IPF-PROM questionnaire	**n=27**	**n=19**	**n=8**	
Total *Physical breathlessness* *Psychological breathlessness* *Well-being* *Energy*	20 (+/- 5) 5 (+/- 0.3) 5 (+/- 0.3) 5 (+/- 0.4) 5 (+/- 0.3)	20 (+/- 1.05) 5 (+/- 0.3) 5 (+/- 0.3) 5 (+/- 0.5) 5 (+/- 0.4)	21 (+/- 1) 5 (+/- 0.5) 6 (+/- 0.5) 5 (+/- 0.4) 5 (+/- 0.3)	0.337 0.264 0.121 0.662 0.565
VAS Cough (cm) (median, (IQR))	1.7 (0.8-2.8)	1.5 (0.2-2.6)	2.3 (1.4-3.0)	0.135
Bristol VAS breathlessness (cm)	1.9 (0.8-3.3)	1.8 (0.8-3.4)	2.4 (1.1-4.7)	0.630
Bristol VAS fatigue (cm)	3.7 (1.1-5.1)	3.7 (1.1-5.1)	3.8 (1.4-6.3)	0.457

Table 4.1. Baseline characteristics of IPF participants All data shown as mean with standard deviation (SD) unless otherwise stated. All questionnaire results and visual analogue scores presented as median with interquartile range (IQR). A $p<0.05$ was considered statistically significant.

Reasons for exercise cessation across the mild and moderate cohort fell into two categories, namely breathlessness and muscle (leg) fatigue (n=10, 37% and n=17, 63% respectively). Within the mild patient group who were undertaking repeat CPET at 12 months, this was more evenly split, with breathlessness being the cause of discontinuation in 9 patients (47%) and leg fatigue in 10 (53%).

Across the whole cohort, the percentage of patients stating they undertake regular exercise was 56% (n=15) although no reasons were given as to why the remaining patients chose not to exercise routinely and no difference between the mild and moderate groups for those undertaking exercise routinely was observed (58% vs 50% respectively) ($p>0.05$).

One patient described dizziness related to his breathlessness during CPET, but no other adverse events were recorded. There were no serious adverse events

4.1. Baseline measurement comparisons between mild and moderate groups:

From table 1, it can be seen baseline demographics between the mild and moderate IPF groups were statistically comparable. As per the study inclusion and subgroup definitions, participants in the moderate IPF group had a statistically lower baseline FVC% predicted when compared to those in the milder group (moderate 70% ± 1.8 vs mild 99% ± 2.3, $p<0.0001$). Furthermore, although not reaching statistical significance, there was a trend towards a reduced TL_{CO} in those patients within the more advanced disease moderate subgroup (moderate 57% ± 2.2 vs mild 63% ± 2.2, $p= 0.055$). No difference was seen in either of the 6MWT parameters of distance achieved or % of theoretical distance.

Baseline QoL questionnaires showed lower scores (indicating a reduced patient perceived QoL) for the patients with moderate disease for total K-BILD score (moderate 60 ±2 vs mild 64 ±2, p= 0.058) as well as a similar trend in the individual domains of chest symptoms and psychological when compared to those with mild disease, although these values were not statistically significant. When comparing the results within the IPF-PROM total and domain scores, where a higher score is indicative of a reduced quality of life, no significant differences were seen. Study patients showed an increase in psychological experience of breathlessness (moderate 6± 0.5 vs mild 5± 0.3, p= 0.121) although again, this was not statistically meaningful.

No differences were observed in the Visual Analogue Scores (VAS) for cough, breathlessness or fatigue between the two groups.

Participants across both groups achieved the anaerobic threshold (AT) during baseline testing and a respiratory exchange ratio (RER) of >1.10 allowing an expectation of a maximal effort test. No tests were stopped prematurely. Baseline CPET values were all within expected published ranges for age (Society and Physicians 2003) being relatively well preserved, not unexpected given the milder disease nature of this patient population. Of note, was a significant difference between groups in % predicated peak minute ventilation (moderate 82.9% ± 4.4 vs mild 71.0% ± 3.2, p= 0.045) likely indicating the increased necessity for oxygen uptake in this patient group although no evidence is published to suggest any prognostic use for this difference.

Our initial baseline data analysis correlating CPET variables with QoL scores resulted in acceptance of two posters at national and international congress (Winter British Thoracic Society 2019 and American Thoracic Society 2020), see appendix F and G. From this initial data of CPET variables measured, some significant correlations to both questionnaires were observed. At anaerobic threshold (AT), VO_2 peak/kg positively correlated with K-BILD total scores (r=0.42, p=0.03), breathlessness/activity (r=0.47, P=0.014) and chest domains (r=0.44, p=0.002) (Pearson's correlation). Similarly, IPF-PROM total score and wellbeing domains significantly correlated with VO_2 peak (r=-0.43, p=0.02 and r=-0.44, p=0.02) with a trend towards statistical significance for total IPF-PROM and VO_2 peak at AT (p=0.06). The ventilatory equivalents for oxygen ($VE/VO_{2)}$ at AT also strongly correlated with total K-BILD score (r=0.39, p=0.001) although there were no significant correlations with the individual domains of either questionnaire.

Baseline Variable	**Correlation coefficient**	**P value (Spearman's**
Peak VO_2/kg/min at AT		
K-BILD		

vs Breathlessness and Activity domain	0.47	**0.014**
vs Chest domain	0.44	**0.002**
vs Total	0.42	**0.03**
IPF-PROM		
vs Total score	-0.36	0.06
Peak VO_2		
K-BILD		
vs Chest domain	0.42	**0.03**
IPF-PROM		
vs Total score	-0.43	**0.02**
vs wellbeing domain	-0.44	**0.02**
FVC % predicted		
vs Total K-BILD		0.14
vs Total IPF-PROM		0.50
TL_{CO} % predicted		
vs Total K-BILD		0.16
vs Total IPF-PROM		0.32

Table 4.2. Baseline CPET parameter correlation to QOL scores and pulmonary function testing.

With regards the more widely performed exercise test of 6MWT and total questionnaire scores, total scores for both K-BILD and IPF-PROM significantly correlated with 6MWT distance (K-BILD r=0.44, p=0.03 and IPF-PROM r=-0.43, p=0.03). The expected difference between groups in distance walked was not observed, possibly due to the inherent difficulties of this test, with regards individual motivation, baseline fitness and mobility issues affecting movement. Baseline and minimum SpO_2 results from 6MWT did not show significant correlations (Total K-BILD p=0.25 and p=0.32 respectively, Total IPF-PROM p=0.53 and p=0.55 respectively). Considering the results from the lung functions parameters, again no significant correlations were observed (FVC % predicted: Total K-BILD, p=0.14; Total IPF-PROM p=0.50 and TL_{CO} % predicted: Total K-BILD, p=0.16; Total IPF-PROM, p=0.32).

No significant correlation between baseline CPET parameters and VAS scores were seen (p>0.05).

4.2. Measurements at 1 year follow up

Total study cohort

Previous work undertaken by Collard et al. (Collard et al. 2003) have described the utility of % FVC decline as a surrogate marker of survival in IPF patients. Their work has shown, over a similar 12 month follow up period, a drop of 10% FVC % predicted will significantly reduce expected mortality versus a stable or improving IPF population (with regards FVC % predicted). In our study cohort, at one year

follow up, the mean reduction in FVC % predicted for the whole study group (n=27) was -3.56% (± 1.37, p=0.04). Whilst statistics dictate this to be significant, it is likely to be below those deemed clinically meaningful. Similarly, with TL_{CO} % predicted, data from recent Czech registry data (Doubková et al. 2018), confirmed a 12 month drop of 15% TL_{CO} predicted conferred a worse overall survival. Our study group had a mean decline of 3.23% (± 1.47, p=0.04), again statistically significant but well below the data evidence for clinical significance.

For the 6MWT parameters, there was no statistically significant reduction in either distance achieved (-4.8m ± 6.9, p=0.50) or in the % theoretical distance 0.15% (±1.6, p=0.92).

Within the K-BILD questionnaire, there was a statistically significant reduction in the breathlessness and activity domain from baseline to follow up of 4.81 points (±1.96, p=0.02). This drop in score is suggestive of a worsening health status, although it is difficult to predict the clinical significance. Published data (Sinha et al. 2019) has suggested a minimum clinically important difference (MCID) of a 7-point reduction for this domain, although this was assessed from an ILD cohort of only 57 patients, of which 17 were IPF. As a reduction in activity and inability to perform the usual daily activities is often an early effect mentioned by IPF patients, such a 5-point drop could potentially hold some merit in disease development and QoL.

Parameter	**Change at one year follow up (n=27)**	**P value**
Lung function /walk test	**Mean % change (+/- SD)**	
FVC % predicted	-3.6 (+/- 1.4)	**0.015**
TL_{CO} % predicted	-3.2 (+/- 1.5)	**0.037**
6MWT distance % theoretical distance	-0.16 (+/- 1.6)	0.920
K-BILD questionnaire	**Mean unit change (+/- SD)**	
Total *Psychological domain* *Breathlessness and activity domain* *Chest symptoms domain*	-2.30 (+/- 1.73) -1.71 (+/- 3.13) -4.81 (+/-1.96) -2.15 (+/- 2.77)	0.194 0.590 **0.021** 0.447
IPF-PROM		
Total *Physical breathlessness* *Psychological breathlessness* *Well-being* *Energy*	0.52 (+/- 0.75) 0.11 (+/- 0.25) 0.33 (+/- 0.27) -0.33 (+/- 0.23) 0.37 (+/- 0.21)	0.408 0.663 0.232 0.164 0.086
VAS Cough (cm) median	-0.21	0.601
Bristol VAS breathlessness (cm) median	0.0	0.876
Bristol VAS fatigue (cm) median	-0.10	0.925

Table 4.3. Total cohort change in Lung function, K-BILD, IPF-PROM and Visual analogue scores of patients at 1 year follow up. Paired t-test or Wilcoxon paired signed rank test.

There were no statistically significant differences in the VAS scores for cough, breathlessness and fatigue or in the IPF-PROM from baseline to 12 months (p>0.05).

4.2.1. Mild group repeat CPET follow up

Aside from the 4 mild patients lost to follow up, a further 6 were unable to attend their planned test date due to the impact of the coronavirus COVID-19 pandemic and prohibited the return of these patients to the hospital for non-emergency care. This left a total of thirteen patients who returned for repeat CPET at 1 year.

As can be seen in table 4.4, several CPET parameters saw a statistically significant decline over the 12month period. Perhaps the most studied variable within CPET testing, that of VO_2 peak declined by an average of 2.5 ml/kg/min across this small subset of patients (21.58 ml/kg/min ± 0.8 vs 19.08 ± 0.8, p=0.017). According to a German registry of over 10,000 health volunteers undertaking CPET, this peak figure would place our study patients in the bottom 10% with age matched healthy individuals (Rapp et al. 2018). Although as alluded to in earlier chapters, this follow up measurement is well above the suggested threshold by Fell et al. of a peak VO_2 <8.3ml/kg/min (HR 3.24, CI 1.10-9.56, p=0.03) being prognostic for early mortality (Fell et al. 2009). In the absence of minimally clinically important differences (MCID) being established for CPET variables in any IPF population, interpretation of such a decline in this mild group is difficult to quantify. Mezzani (Mezzani 2017) stated an average individual, after the age of 30 years will decrease their VO_2 peak by approximately 10% per decade due to numerous factors including stroke volume and the muscular ability to utilise oxygen with increasing age. This being the case, our cohort exhibit an accelerated decline in this output. One patient did not reach AT at follow up and this data was removed from the CPET follow up comparisons seen in table 4.4.

Several other physiological parameters resulting from CPET provided statistically significant differences from baseline to follow up (see table 4.4), including;

- VO_2 peak at AT (14.12 ml/kg/min ± 0.92 vs 11.82 ± 0.5, p=0.044
- Minute ventilation (VE) peak (75.31 L/min ± 5.8 vs 66.08 ± 6.0, p=0.007)
- Peak work (106W ± 7.3 VS 90.77 ± 7.2, p=0.022)
- Heart rate (HR) (142.3 bpm ± 6.7 vs 133.0 ± 6.2, p=0.040)
- Breathing reserve (BR max) at AT (21.8 (12.4-34.2) vs 33.8 (20.2-55.7), p=0.0002)

CPET parameters	Baseline (n=13)	Follow up (n=12)	P value

VO_2 peak (ml/kg/min)	21.58 ± 0.8	19.08 ± 0.8	**0.017**
VO_2 peak at AT (ml/kg/min)	14.19 ± 0.9	11.83 ± 0.5, n=12	**0.044**
VE peak (L/min)	75.31 ± 5.8	66.08 ± 6.0	**0.007**
VE peak % pred	75.54 ± 3.7	65.88 ± 3.4	**0.007**
VE/VCO_2 at AT	29.68 ± 0.9	31 ± 1.3, n=12	0.353
Minimum O_2 saturation during CPET (%)	91.46 ± 1.5	87.92 ± 1.9, n=12	0.182
Peak Work (W)	106.9 ± 7.3	90.77 ± 7.2	**0.022**
Peak Work (% predicted)	44.31 ± 1.1	37.69 ± 2.4	**0.002**
HR (bpm)	142.3 ± 6.7	133 ± 6.2	**0.040**
HR (% predicted)	98.69 ± 4.7	91.77 ± 4.7	**0.022**
BR max (median, (IQR))	21.8 (12.4-34.2)	33.8 (20.2-55.7)	**0.0002**
6MWT			
Distance achieved (m)	346.9 ± 20.5	340.8 ± 20.1	0.563
% theoretical distance (m)	76.38 ± 5.1	76 ± 4.7	0.872
Lung function			
FVC % predicted	98.77 ± 2.4	93.38 ± 2.9	**0.010**
TL_{CO} % predicted	62.31 ± 2.7	59.31 ± 3.3	0.161

Table 4.4. Baseline and 1 year follow up data for patients within mild group (those with matched tests).

As within the overall study population, there was a significant reduction in FVC% predicted at 12 months although again, the mean percent predicted decline was <10% (baseline FVC 98.77% ± 2.44 vs follow up FVC 93.88% ± 2.9, P=0.01). Whilst Zappala et.al (Zappala et al. 2010), described an association of a more marginal FVC % decline (5–10%) and overall mortality, this was considered significant over a 6 month follow up, so interpretation of such a decline over a year is more difficult.

In this same mild cohort of patients, statistically significant reductions were observed in both the domains of breathlessness & activity (-7.21 ± -2.98, p=0.033) and chest (-9.59 ± -4.15, p=0.040) with a trend towards significance in the total score (-5.57 ± -2.88, p=0.077) at follow up. Notably, these mean unit changes of total K-BILD score (-5.57) and breathlessness and activity domain (-7.21) exceed the MCID reported by the questionnaire authors (5 and 7 unit change respectively) (Sinha et al. 2019). When relating this significance to actual patient numbers, a total of 5/13 reached the MCID for total score and 8/13 for the breathlessness domain. Furthermore, the reduction of >-9 units is approaching the estimated MCID of -11 unit decline for chest symptoms.

The second questionnaire, IPF-PROM, showed a statistically significant worsening in the psychological experience of breathlessness (0.76 ± 0.34, p=0.044). To date, no unit MCID for changes in domain score have been published and therefore the clinical significance of such a change is unknown.

K-BILD questionnaire	**Unit change (mean, SD) n=13**	**P value**
Total	-5.57 (+/-2.88)	0.077
Psychological domain	-5.17 (+/- 5.2)	0.304
Breathlessness and activity domain	-7.21 (+/- 2.98)	**0.033**
Chest symptoms domain	-9.59 (+/- 4.15)	**0.040**
IPF-PROM		
Total	1.54 (+/- 0.89)	0.109
Physical breathlessness	0.54 (+/- 0.39)	0.189
Psychological breathlessness	0.77 (+/- 0.34)	**0.044**
Well-being	-0.15 (+/- 0.34)	0.656
Energy	0.39 (+/- 0.39)	0.337
VAS Cough (cm) median	-1.6	0.391
Bristol VAS breathlessness (cm) median	0.0	0.716
Bristol VAS fatigue (cm) median	-0.1	0.956

Table 4.5. Change in K-BILD, IPF-PROM and Visual analogue scores of mild IPF patients with repeat CPET at 1 year follow up. Results shown as mean change in questionnaire score with standard deviation (SD), unless otherwise stated. Paired t-test used for parametric data and Wilcoxon matched pairs signed rank test for non-parametric data.

No statistical significance was seen for the VAS scores for cough, breathlessness or fatigue from baseline to one year (p>0.05).

Finally, we explored the correlation between the changes of K-BILD at 12 months and the CPET variables at baseline to better understand how exercise testing may reflect the patients experience of their disease. Of interest, a lower VO_2 peak/ml/min at anaerobic threshold correlated with greater declines in total K-BILD score (r=-0.62, p=0.024) and the psychological domain at follow up (r=-0.63, p=0.022). No other CPET parameters significantly correlated to changes in K-BILD score.

As seen previously, the results of the baseline FVC% predicted and TL_{CO}% predicted showed no significant correlation to the changes in K-BILD over the 12 months follow up (p=0.70 and p=0.62 respectively).

When considering the reasons for CPET discontinuation in the follow up tests, seven patients cited breathlessness as the cause to stop (54%) whilst 5 patients complained of muscle fatigue (38%). A

single patient suggested a dry mouth was their reason for ending the test. No serious adverse events were reported.

4.3 Discussion

Whilst CPET remains the gold standard for pulmonary exercise testing (Ferrazza et al. 2009), given its unique assessment of any ventilatory, cardiac and metabolic limitations to exercise and the safety of such a test is well documented across multiple morbidities (Patel et al. 2019, Kleber and Köln 2018, Ney et al. 2016), its utility in the IPF clinical setting is minimal, perhaps due to the lack of evidence to support its use either as a marker of current disease state or indeed as a prognostic tool. This data deficiency was highlighted in our systematic review, seen in chapter 2. The heterogeneity of IPF disease course allied to the retrospective nature of the majority of published studies has greatly limited conclusions that could be drawn. Furthermore, whilst estimates of MCID have been given for the QoL questionnaire, K-BILD, suggestive that a patient's perception of their own illness may be indicative of the disease state, this is not the case for CPET parameters in our patient cohort, where MCID's are yet to be established.

This small feasibility study has provided strong evidence that CPET can be undertaken in a mild to moderate IPF population successfully. Whilst side effects of the test were minimal and transient (slight dizziness and dry mouth both n=1), study attrition was high with just under two thirds (64%) of those enrolled completing the protocol. This can somewhat be explained in the latter end of the study by the effects of COVID-19 and patient isolation, however a significant number were unable to initiate the study after giving consent (10 patients, 24%). This is largely due to the patient demographics and comorbidities (frailty & undiagnosed cardiac disease) and will go some way to aiding power calculation for larger similar studies.

Interpretation of the CPET data suggests peak VO_2 is associated with a clinically meaningful patientperceived reduction in health status despite only a limited change and relative stability of lung function parameters (<10% decline in FVC and <15% decline in TL_{CO}). Whilst a similar correlation between VO_2 peak and QoL scores has been seen before, notably El Naggar (El Naggar 2017), who concluded a negative correlation (r=-0.35) between VO_2 peak performance and a non-IPF specific health questionnaire (Saint Georges Respiratory Questionnaire, SGRQ), no longitudinal measurements were explored. Interestingly, the same study noted a correlation between VO_2 and TL_{CO} (r=-0.53), a result not replicated in our study. In our cohort, VO_2 peak correlated with patient reported outcome measures at baseline. This correlation remained over the 12 months of follow up with significant declines in both measurements. The extent to which VO_2 peak is a useful predictive marker in IPF patients is unclear with conflicting data from different study populations. Values of peak VO_2 ranging

from <8.3 to <14.2 ml/kg/min (Fell et al. 2009, Triantafillidou et al. 2013, Vainshelboim et al. 2016), have reported to predict mortality, whilst others have failed to identify a significant association (Wallaert et al. 2011, Miki et al. 2003). Extended follow up of our prospective cohort may shed further evidence as to the true impact of this measurement on the long-term outcomes in IPF patients.

When considering the milder population and specifically their follow up results, the decline in exercise performance from baseline to one year is marked. Whilst a limited reduction in FVC% predicted is observed and expected, this is accompanied with a reduced minute ventilation (VE) and an increased breathing reserve (BR). A BR >20% is suggestive that this cohort is not adversely affected by ventilatory limitation. Assessment of baseline and follow up activity levels of participants does not suggest a significant change of habits, with 42% and 46% taking regular exercise at baseline and follow up respectively. Patients did not report cardiac or pulmonary vascular dysfunction and test instructions were identical on each occasion. One possible explanation for such a decline may lie in the patient's perception of disease, specifically breathlessness. We have seen the MCID for the K-BILD breathlessness and activity domain was surpassed (>7-unit reduction) over the follow up period. The hypothesis for the reduced exercise performance is the self-imposed deconditioning of participants who fear exercise will lead to a breathless episode. As with any population, healthy or not, reduced activity levels will lead to a reduction in exercise tolerance. Vainshelboim (Vainshelboim 2016) has previously described a similar phenomenon, studying the benefits of exercise training in IPF patients (figure 5.1) and whether such deconditioning can be reversed. No specific exercise routine was dictated within this study and therefore it is not possible to speculate if this could have affected our results.

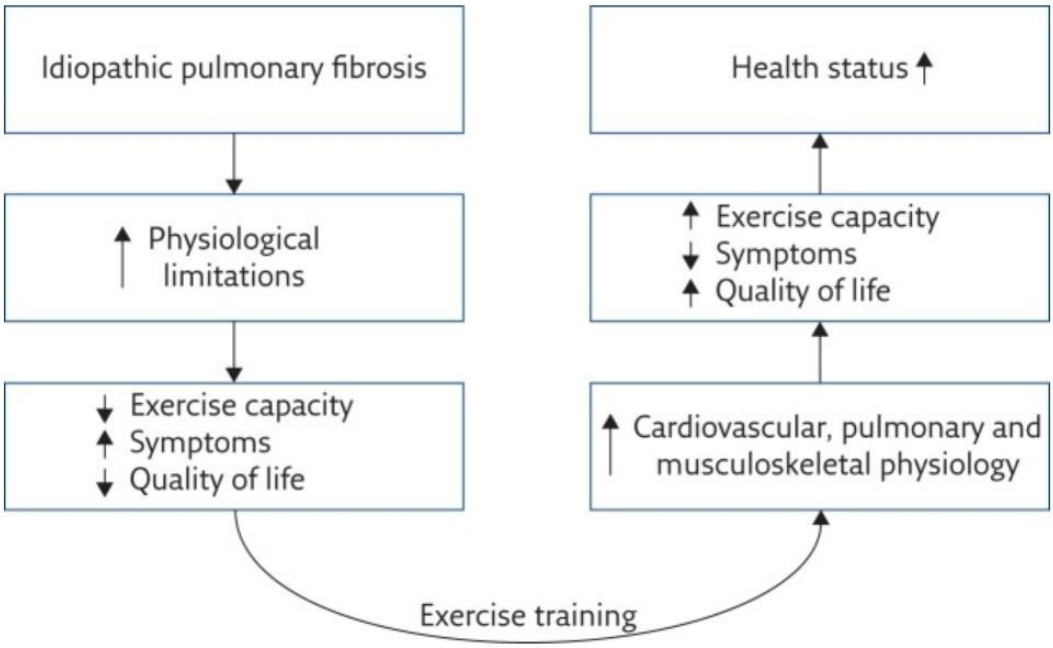


Figure 5.1. Suspected mechanism for patient deconditioning and reversal with exercise training.

Examining other outputs of our mild group CPET follow up results, a large number of variables showed a statistically significant decline over 1 year although interpretation of their clinical meaning is difficult given the lack of data for MCID in such a patient population. Layton et al (Layton et al. 2017) explored the prognostic use of CPET outputs for a more severe ILD cohort awaiting transplantation. Of note, was the workload % predicted cut off of 35% (HR = 4.71, 95% CI = 2.64–8.38) being an independent predictor of mortality or the need to consider transplantation. Whilst Leyton's study did not look at longitudinal changes, the decline in our mild group mean did approach this value (37.69% ± 2.4) and further follow up over time may again provide greater knowledge to the importance of this measurement.

Consideration of the 6MWT results for this same population suggests this to be a less specific test for measurement of exercise decline. Du Bois et al (du Bois et al. 2011) have previously described an MCID of >50m over a 24-week period being predictive of a fourfold increase in risk of death at 1 year. This parameter, alongside the % distance predicted remained stable in our cohort with no significant difference between baseline and follow up, contradictory to the exercise reduction observed on CPET. One possible explanation for this could be the reduced baseline achievements of our group (mean= 350m vs 392m) potentially meaning maximum distanced was not attained, or simply that the outputs of CPET provide greater insight across cardiovascular, pulmonary and skeletal muscle systems and may be more explicit in determining functional changes.

The use, for the first time in an exercise study, of two IPF specific patient reported health questionnaires has provide an interesting insight into an individual's perception of their own disease and exercise capabilities. As touched upon already, it appears that a number of baseline parameters across both K-BILD and IPF-PROM correlate with the outputs of CPET, including those more closely related to an exercise challenge test, namely the chest and breathlessness domains of K-BILD and the energy domain of IPF-PROM as well as the total scores across both. The period of extended follow up of our patient group may again shed more light on the importance of these scores from a prognostic perspective. From our 12-month data in the mild cohort, it is of interest to see where the statistically significant declines are seen, similar to above, the chest and breathlessness domains of K-BILD and the psychological experience of breathlessness in IPF-PROM. Whilst K-BILD has been more widely used and is validated across multiple study populations, the increase in this IPF-PROM domain (0.77 (+/- 0.34), p=0.044) remains unclear given the lack of MCID values. However, this finding could support the psychological aspect and an individual's fear of hypoxia leading to patient deconditioning, given the physical experience of breathlessness remained stable at 1 year follow up.

Finally, despite previous studies undertaken within the Bristol ILD service (Yates et al. 2018a) confirming the validity of visual analogue scales (VAS) across dyspnoea and fatigue and its correlation to changes in total K-BILD score, this was not seen in this study group, with little or no change in scores from baseline to 12 months. Results for the cough VAS were however in line with this same study, finding no significant changes over the year.

Questions remain over the utility of FVC% and TL_{CO}% predicated as either a marker of current disease or indeed future progression. We found no correlations either at baseline or follow up between any of the CPET outputs or questionnaires. Whilst some reports outlined in chapter 2 suggest a correlation of VO_2 and DL_{CO}, it is possible the milder disease state of this cohort did not allow sufficient deterioration over this initial 12 months follow up. This can potentially be confirmed over subsequent years and repeat lung function testing. Furthermore, despite our total cohort FVC% decline over the 12 months follow up being statistically significant with a loss of -3.6% (± 1.4), this was only marginally greater than that seen in a general population study of over 60 year olds, giving an annual expected FVC% decline of −2.46% (−3.07–−1.85) (Luoto et al. 2019). It is probable these standard PFT measurements will provide greater prognostic utility when used in conjunction with other dynamic measurements of an individual's health status.

4.3.1. Limitations

Several limitations of this study exist making interpretation difficult to assess across a wider IPF population. Firstly, and perhaps most importantly, the study participants were a relatively small and homogenous sample of patients. Whilst there is no standardised definition for the severity of IPF disease, according to the pulmonary function parameters currently used in clinical trials, patients with a FVC >50–55% of predicted and a TL_{CO} >35–40% of predicted are typically diagnosed as having mildto-moderate disease, while patients with severe or advanced disease present with FVC and TL_{CO} values lower than the abovementioned thresholds (Caminati et al. 2017). With our inclusion criteria meaning all patients needed both FVC% predicted and TL_{CO} % predicted >50%, suggestive of a milder cohort, patients in our moderate cohort did seem to be more impaired in terms of exercise limitations and desaturation. Such inclusion may limit the overall utility of results, particularly in terms of feasibility of CPET across IPF phenotypes; for example, those with exercise induced pulmonary hypertension versus those with relatively normal pulmonary vascular response to exercise, and the risk of Type II error may be relatively high. Whilst our preference would be to have had a broader range of symptomatic patients, safety evidence to support CPET in more severe patient was very limited and

recruiting patients with a higher MRC score could have led to early completion of testing, before the limit of the pulmonary and cardiovascular systems had been reached, adversely influencing the results. It may be possible to utilise other outcomes of CPET not requiring a maximal effort, for example an Oxygen Uptake Efficiency Slope (OUES), although this measurement was not available at the time of initiation of this study on the analysis dataset provided. Our knowledge gained of the performance and safety of the test of those patients classified as moderate in this study (FVC % predicted <80%) would give more confidence to allow a follow up CPET test at the 12-month review period and potentially expand the inclusion criteria below the 50% thresholds.

Almost a quarter of patients who gave consent for entry in the study developed exclusions or were lost to follow up. Part of this can be explained by the unforeseen COVID-19 pandemic, during which this group of patients were classified as 'extremely vulnerable' and asked to isolate making a follow up CPET test impossible. Whilst this may play an important role to inform power calculations for future studies involving CPET as an outcome measure by utilising the expected mean changes across CPET variables and lung function parameters from this study population, it again limits the generalisability of our outcomes due to resulting smaller sample size.

This study adds confidence and supports the feasibility of IPF patients undertaking repeated CPET in a mild to moderate population. Some ad-hoc patient feedback has suggested a positive outlook towards their individual exercise tolerance as most have self-imposed 'restrictions' on exercise due to the fear of breathlessness and none of the study population had undertaken a maximal exercise effort since a diagnosis of IPF.

4.3.2. Conclusion

The study outcomes provide some evidence that CPET could be a useful tool to assess the change in an individual's health status over time and may add to the armoury of clinicians when faced with the difficult patient discussions around disease progression and prognosis.

Future work should look to confirm the results of this study in a larger, more heterogenous IPF population. This study has provided strong evidence of the safety of CPET in a mild to moderate cohort and furthermore, patients are willing to partake in such a test in a longitudinal study. Inclusion of a more severe patient group likely measured by FVC and/or DL_{CO} % predicted can be combined with the patient's own assessment of their functional ability (e.g. K-BILD). Establishing MCID for longitudinal measurements derived from CPET will provide a significant move forward in our knowledge and in turn, utility of the test. Furthermore, exploration of relationships between CPET outputs and other

assessment tools, potentially quantitative fibrosis CT- derived measures, individual comorbidities, drug treatments and QoL scores provide a more accurate estimation of a patient's expected disease course.

Sadly, since the study closure, four patients have deceased. Next steps will be to continue observations of the twenty-three patients through routine clinical physiology appointments and a hope to add to our knowledge of the usefulness of CPET longitudinal data and prognostic accuracy and its potential for wider use in the clinical setting, akin to the 6MWT. With data from this increased timeframe, the team anticipate presenting the most comprehensive prospective follow up of IPF patients undertaking CPET to date. With ongoing surveillance of participant lung function, and improved knowledge of morbidity and mortality in this group of patients, the aim is to accurately predict study numbers for a larger, multi-site study for the use of CPET in IPF patients. To utilise our knowledge gained to inform inclusion and exclusion criteria and importantly, secure funding to fully answer the questions only partially answered on the prognostic use of CPET in IPF.

Chapter 5: Personal review and learnings

Study reflections

Despite the numerous challenges posted over the last 3 ½ years, since the inception of the idea of undertaking a research MSc, it is with a significantly greater understanding of clinical research that I am writing now. Whilst much of the research process can be controlled, from hypothesis to endpoints, recruitment to follow up, the 'known unknowns', for me, ethics approval provided a steep learning curve. Resulting in a very early extension to the study timelines (and University fees), undertaking the completion of the IRAS form, presentation to the Regional Ethics Committee (REC) and subsequent major amendments were an eye opener. Accompanied by this and touched on earlier is the sheer volume of administration needed prior and during the study, both in electronic and paper form and a necessity for the sponsor and NHS research alike. If there was an ability to turn back time, or pass learnings to future students, a few changes could save many months of process:

Research Passport – a requirement to undertake research on the NHS premises, this littleknown approval process remains confused. There is a necessity to streamline this process, agree the timelines for application and better understand where such approval comes from.

Short courses – Offered across the Medical School, attendance of several courses over the study period has been essential. From statistics through to paper and thesis writing, these days have added greatly to undertaking research and such opportunities should be maximised.

Funding – Whilst not always possible or even available, I would be keen to secure funding to aid additional nurse support for such studies. The part time nature of this MSc has at times led to difficulties in my person clinic attendance, primarily during recruitment but also follow up. It is important to recognise the additional work the clinical research team undertook on numerous occasions, to gain consent, undertake questionnaire and ensure the clinical research folder was up to date. Furthermore, the amazing work of the physiology team over the study period to ensure patient testing was on time according to study protocol and minimise travel of patients to the hospital for multiple appointments.

With the benefit of hindsight and a better understanding of the safety of CPET in IPF patients, this study would undoubtably benefit from a wider inclusion criteria. Our knowledge gained from this and other ongoing research in this field would likely allow us to revise downwards the lower limits of lung function parameters and permit a repeat of exercise testing of all participants. Given the acceptance of this study group to repeat CPET after 12 months, it could also add value to continue annual exercise testing alongside their routine clinical lung function to better understand the changes within individual CPET parameters and the longitudinal importance of each in predictions of prognosis.

The near future presents opportunities for publication of this data and presentation across physiology and respiratory conferences. Longer term data will aid our understanding of the importance of individual outcomes across all the tested parameters (CPET, PFTs, questionnaires) and armed with this information, allied to our improved understanding of safety, the hope would be to initiate a large scale study to validate such findings and the prognostic use of CPET itself.

One of the greatest rewards from the last couple of years has been the direct interaction with patients. Having worked in the therapy area for more than 10 years from early phase drug discovery through to large global phase III studies from within a pharmaceutical setting, whilst patients remain front of mind, they are often of sight. It has been clear throughout how enthusiastic this group of patients are to support and add to our understanding of the disease they are living with on a daily basis. In the knowledge results of such studies may not benefit them directly, it was important to all, anything they could do to benefit others was a ‘must’. It was a privilege to meet each of them, often with carers, who share a common goal in one day finding a cure for IPF.

The arrival of an unprecedented global pandemic undoubtably adversely affected our desired outcomes, reducing follow up numbers and the availability of clinical data (PFTs) due to the advice for

all study participants to shield for many months due to their vulnerability to the virus. Teamwork and support of the sponsor meant we were able to maximise our data return via postal questionnaires and e mail reconsent. Whilst only emergency visits were allowed into hospital and clinic appointments moving online, routine lung function testing was halted for a period towards the close of the study. Any patients with an outstanding test (for study purposes) were contacted by the physiology team to arrange the earliest possible appointment on the re-opening of the department.

On a personal note, the undertaking of this study has given me insight into set up, delivery and interpretation of clinical research. From hypothesis generation, study design, presenting to a REC and patient enrolment through to the outputs of data presentation and publication. A revised paper of this study has been submitted to BMC Pulmonary Medicine which can be seen in Appendix H. A first for me was to present my own generated data at a national respiratory congress alongside experts in the field of ILD. The work undertaken in this study and within the literature review has been cited at subsequent pulmonary congresses and stands Bristol ILD service in good shape for future collaborations within the field of exercise and ILD.

Most of all, it has provided me an unparalleled opportunity to engage with such an enthusiastic and willing group of patients.

My hope is, this data will add something to the multitude of ongoing work globally in IPF prognosis, to allow clinicians to better answer their most important question of what this diagnosis means to each individual.

References:

Akinpelu, D. 2018. Treadmill stress testing periprocedural care retrieved from https://emedicine.medscape.com/article/1827089-periprocedure.

Albera, C. (2011) Challenges in idiopathic pulmonary fibrosis trials: the point on end-points. *Eur Respir Rev,* 20**,** 195-200.

Albouaini, K., M. Egred, A. Alahmar & D. J. Wright (2007) Cardiopulmonary exercise testing and its application. *Postgrad Med J,* 83**,** 675-82.

Altman, D. G., J. J. Decks, M. Clarke & C. Cates (2000) The quality of systematic reviews. High quality reporting of both randomised trials and systematic reviews should be priority. *BMJ,* 321**,** 297; author reply 298-9.

Antonarakis, S. E., A. Chakravarti, J. C. Cohen & J. Hardy. 2010. Mendelian disorders and multifactorial traits: the big divide or one for all? In *Nat Rev Genet*, 380-4. England.

Arena, R. & K. E. Sietsema (2011) Cardiopulmonary exercise testing in the clinical evaluation of patients with heart and lung disease. *Circulation,* 123**,** 668-80.

ATS/ERS Statement. 2000. American Thoracic Society. Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. American Thoracic Society (ATS), and the European Respiratory Society (ERS).

Barratt, S. L., R. Davis, C. Sharp & J. D. Pauling (2020) The prognostic value of cardiopulmonary exercise testing in interstitial lung disease: a systematic review. *ERJ Open Res,* 6.

Baumgartner, K. B., J. M. Samet, D. B. Coultas, C. A. Stidley, W. C. Hunt, T. V. Colby & J. A. Waldron (2000) Occupational and environmental risk factors for idiopathic pulmonary fibrosis: a multicenter case-control study. Collaborating Centers. *Am J Epidemiol,* 152**,** 307-15.

Behr, J., F. Bonella, R. Bonnet, S. Gläser, C. Grohé, A. Günther, D. Koschel, M. Kreuter, D. Kirsten, C. Krögel, P. Markart, J. Müller-Quernheim, C. Neurohr, M. Pfeifer, A. Prasse, N. Schönfeld, J. Schreiber, H. Wirtz, C. Witt & U. Costabel (2015) [Position Paper: Significance of the Forced Vital Capacity in Idiopathic Pulmonary Fibrosis]. *Pneumologie,* 69**,** 455-8.

Bellaye, P. S. & M. Kolb (2015) Why do patients get idiopathic pulmonary fibrosis? Current concepts in the pathogenesis of pulmonary fibrosis. *BMC Med,* 13**,** 176.

Bellou, V., L. Belbasis, A. Konstantinidis & E. Evangelou (2017) Tobacco smoking and risk for idiopathic pulmonary fibrosis: a prospective cohort study in UK Biobank. *European Respiratory Journal,* 50**,** PA4887.

Borie, R. & C. Van Moorsel. 2021. A statement on the impact of genetics in Pulmonary Fibrosis. Genetics of ILD; Task Force ERS; 2020-2021.

Boutou, A. K., G. G. Pitsiou, I. Trigonis, D. Papakosta, P. K. Kontou, N. Chavouzis, C. Nakou, P. Argyropoulou, K. Wasserman & I. Stanopoulos (2011) Exercise capacity in idiopathic pulmonary fibrosis: the effect of pulmonary hypertension. *Respirology,* 16**,** 451-8.

Bujang, M. A., N. Sa'at & T. M. I. T. A. B. Sidik (2017) Determination of Minimum Sample Size Requirement for Multiple Linear Regression and Analysis of Covariance Based on Experimental and Non-experimental Studies. *Epidemiology Biostatistics and Public Health,* 14.

Caminati, A., A. Bianchi, R. Cassandro, M. R. Mirenda & S. Harari (2009) Walking distance on 6-MWT is a prognostic factor in idiopathic pulmonary fibrosis. *Respir Med,* 103**,** 117-23.

Caminati, A., R. Cassandro, O. Torre & S. Harari (2017) Severe idiopathic pulmonary fibrosis: what can be done? *Eur Respir Rev,* 26.

Chambers, R. C. (2008) Abnormal wound healing responses in pulmonary fibrosis: focus on coagulation signalling. *European Respiratory Review,* 17**,** 130.

Clukers, J., M. Lanclus, B. Mignot, C. Van Holsbeke, J. Roseman, S. Porter, E. Gorina, E. Kouchakji, K. E. Lipson, W. De Backer & J. De Backer (2018) Quantitative CT analysis using functional imaging is superior in describing disease progression in idiopathic pulmonary fibrosis compared to forced vital capacity. *Respir Res,* 19**,** 213.

Cohen, J. (1968) Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull,* 70**,** 213-20.

Collard, H. R., T. E. King, B. B. Bartelson, J. S. Vourlekis, M. I. Schwarz & K. K. Brown (2003) Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med,* 168**,** 538-42.

Collard, H. R., C. J. Ryerson, T. J. Corte, G. Jenkins, Y. Kondoh, D. J. Lederer, J. S. Lee, T. M. Maher, A. U. Wells, K. M. Antoniou, J. Behr, K. K. Brown, V. Cottin, K. R. Flaherty, J. Fukuoka, D. M. Hansell, T. Johkoh, N. Kaminski, D. S. Kim, M. Kolb, D. A. Lynch, J. L. Myers, G. Raghu, L. Richeldi, H. Taniguchi & F. J. Martinez (2016) Acute Exacerbation of Idiopathic Pulmonary Fibrosis. An International Working Group Report. *Am J Respir Crit Care Med,* 194**,** 265-75.

Daniels, C. E., E. S. Yi & J. H. Ryu (2008) Autopsy findings in 42 consecutive patients with idiopathic pulmonary fibrosis. *Eur Respir J,* 32**,** 170-4.

Demedts, M. & U. Costabel (2002) ATS/ERS international multidisciplinary consensus classification of the idiopathic interstitial pneumonias. *European Respiratory Journal,* 19**,** 794.

Derksen, S. & H. J. Keselman (1992) Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. *British Journal of Mathematical & Statistical Psychology,* 45**,** 265-282.

Doubková, M., J. Švancara, M. Svoboda, M. Šterclová, V. Bartoš, M. Plačková, L. Lacina, M. Žurková, I. Binková, R. Bittenglová, V. Lošťáková, Z. Merta, L. Šišková, R. Tyl, P. Lisá, H. Šuldová, F. Petřík, J. Pšikalová, V. Řihák, T. Snížek, P. Reiterer, J. Homolka, P. Musilová, J. Lněnička, P. Palúch, R. Hrdina, R. Králová, H. Hortvíková, J. Strenková & M. Vašáková (2018) EMPIRE Registry, Czech Part: Impact of demographics, pulmonary function and HRCT on survival and clinical course in idiopathic pulmonary fibrosis. *Clin Respir J,* 12**,** 1526-1535.

du Bois, R. M., D. Weycker, C. Albera, W. Z. Bradford, U. Costabel, A. Kartashov, L. Lancaster, P. W. Noble, S. A. Sahn, J. Szwarcberg, M. Thomeer, D. Valeyre & T. E. King (2011) Six-minute-walk test in idiopathic pulmonary fibrosis: test validation and minimal clinically important difference. *Am J Respir Crit Care Med,* 183**,** 1231-7.

Duckworth, A., M. A. Gibbons, R. J. Allen, H. Almond, R. N. Beaumont, A. R. Wood, K. Lunnon, M. A. Lindsay, L. V. Wain, J. Tyrrell & C. J. Scotton (2020) Telomere length and risk of idiopathic pulmonary fibrosis and chronic obstructive pulmonary disease: a mendelian randomisation study. *Lancet Respir Med*.

Dudley, K. A. & S. El-Chemaly (2012) Cardiopulmonary exercise testing in lung transplantation: a review. *Pulm Med,* 2012**,** 237852.

El Naggar, T. 2017. Evaluation of the questionnaires' validity in assessing the severity of idiopathic pulmonary fibrosis with high-resolution computed tomography, lung diffusion and cardiopulmonary exercise test*.,* eds. A. Said & S. Sharkawy. Egypt J Bronchol **11**, 141-148.

Enright, P. L. & D. L. Sherrill (1998) Reference equations for the six-minute walk in healthy adults. *Am J Respir Crit Care Med,* 158**,** 1384-7.

Evans, C. M., T. E. Fingerlin, M. I. Schwarz, D. Lynch, J. Kurche, L. Warg, I. V. Yang & D. A. Schwartz (2016) Idiopathic Pulmonary Fibrosis: A Genetic Disease That Involves Mucociliary Dysfunction of the Peripheral Airways. *Physiol Rev,* 96**,** 1567-91.

Fein, M., M. P. Ritter, T. R. DeMeester, S. Oberg, J. H. Peters, J. A. Hagen & C. G. Bremner (1999) Role of the lower esophageal sphincter and hiatal hernia in the pathogenesis of gastroesophageal reflux disease. *J Gastrointest Surg,* 3**,** 405-10.

Fell, C. D., L. X. Liu, C. Motika, E. A. Kazerooni, B. H. Gross, W. D. Travis, T. V. Colby, S. Murray, G. B. Toews, F. J. Martinez & K. R. Flaherty (2009) The prognostic value of cardiopulmonary exercise testing in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med,* 179**,** 402-7.

Fernández Pérez, E. R., C. E. Daniels, D. R. Schroeder, J. St Sauver, T. E. Hartman, B. J. Bartholmai, E. S. Yi & J. H. Ryu (2010) Incidence, prevalence, and clinical course of idiopathic pulmonary fibrosis: a population-based study. *Chest,* 137**,** 129-37.

Ferrazza, A. M., D. Martolini, G. Valli & P. Palange (2009) Cardiopulmonary exercise testing in the functional and prognostic evaluation of patients with pulmonary diseases. *Respiration,* 77**,** 317.

Flaherty, K. R., A. C. Andrei, S. Murray, C. Fraley, T. V. Colby, W. D. Travis, V. Lama, E. A. Kazerooni, B. H. Gross, G. B. Toews & F. J. Martinez (2006) Idiopathic pulmonary fibrosis: prognostic value of changes in physiology and six-minute-walk test. *Am J Respir Crit Care Med,* 174**,** 803-9.

Flaherty, K. R., A. U. Wells, V. Cottin, A. Devaraj, S. L. F. Walsh, Y. Inoue, L. Richeldi, M. Kolb, K. Tetzlaff, S. Stowasser, C. Coeck, E. Clerisme-Beaty, B. Rosenstock, M. Quaresma, T. Haeufel, R. G. Goeldner, R. Schlenker-Herceg, K. K. Brown & I. T. Investigators (2019) Nintedanib in Progressive Fibrosing Interstitial Lung Diseases. *N Engl J Med,* 381**,** 1718-1727.

Fletcher, C. 1952. The clinical diagnosis of pulmonary emphysema - and experimental study. Proc R Soc Med. 1952 Se;45(9):577-84. PMID: 13003946.

Foundation, P. F. 2020. PF Overview. retrieved from https://pulmonaryfibrosis.org/life-withpf/about-pf.

Furukawa, T., H. Taniguchi, M. Ando, Y. Kondoh, K. Kataoka, O. Nishiyama, T. Johkoh, J. Fukuoka, K. Sakamoto & Y. Hasegawa (2017) The St. George's Respiratory Questionnaire as a prognostic factor in IPF. *Respir Res,* 18**,** 18.

García-Sancho, C., I. Buendía-Roldán, M. R. Fernández-Plata, C. Navarro, R. Pérez-Padilla, M. H. Vargas, J. E. Loyd & M. Selman (2011) Familial pulmonary fibrosis is the strongest risk factor for idiopathic pulmonary fibrosis. *Respir Med,* 105**,** 1902-7.

Gläser, S., A. Obst, B. Koch, B. Henkel, A. Grieger, S. B. Felix, M. Halank, L. Bruch, T. Bollmann, C. Warnke, C. Schäper & R. Ewert (2013) Pulmonary hypertension in patients with idiopathic pulmonary fibrosis - the predictive value of exercise capacity and gas exchange efficiency. *PLoS One,* 8**,** e65643.

Gordon, J. K. & R. T. Domsic (2016) Clinical Trial Design Issues in Systemic Sclerosis: an Update. *Curr Rheumatol Rep,* 18**,** 38.

Graham, B. L., I. Steenbruggen, M. R. Miller, I. Z. Barjaktarevic, B. G. Cooper, G. L. Hall, T. S. Hallstrand, D. A. Kaminsky, K. McCarthy, M. C. McCormack, C. E. Oropez, M. Rosenfeld, S.

Stanojevic, M. P. Swanney & B. R. Thompson (2019) Standardization of Spirometry 2019 Update. An Official American Thoracic Society and European Respiratory Society Technical Statement. *Am J Respir Crit Care Med,* 200**,** e70-e88.

Gribbin, J., R. Hubbard & C. Smith (2009) Role of diabetes mellitus and gastro-oesophageal reflux in the aetiology of idiopathic pulmonary fibrosis. *Respir Med,* 103**,** 927-31.

Hancock, L. A., C. E. Hennessy, G. M. Solomon, E. Dobrinskikh, A. Estrella, N. Hara, D. B. Hill, W. J. Kissner, M. R. Markovetz, D. E. Grove Villalon, M. E. Voss, G. J. Tearney, K. S. Carroll, Y. Shi, M. I. Schwarz, W. R. Thelin, S. M. Rowe, I. V. Yang, C. M. Evans & D. A. Schwartz (2018) Muc5b overexpression causes mucociliary dysfunction and enhances lung fibrosis in mice. *Nat Commun,* 9**,** 5363.

Hayton, C., D. Terrington, A. M. Wilson, N. Chaudhuri, C. Leonard & S. J. Fowler (2019) Breath biomarkers in idiopathic pulmonary fibrosis: a systematic review. *Respir Res,* 20**,** 7.

Hemingway, H., R. D. Riley & D. G. Altman (2009) Ten steps towards improving prognosis research. *BMJ,* 339**,** b4184.

Heresi, G. A. & R. A. Dweik (2011) Strengths and limitations of the six-minute-walk test: a model biomarker study in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med,* 183**,** 1122-4.

Herzog, E. L., A. R. Brody, T. V. Colby, R. Mason & M. C. Williams (2008) Knowns and unknowns of the alveolus. *Proc Am Thorac Soc,* 5**,** 778-82.

Huguet, A., J. A. Hayden, J. Stinson, P. J. McGrath, C. T. Chambers, M. E. Tougas & L. Wozney (2013) Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. *Syst Rev,* 2**,** 71.

Johnson, C., I. Pinal-Fernandez, R. Parikh, J. Paik, J. Albayda, A. L. Mammen, L. Christopher-Stine & S. Danoff (2016) Assessment of Mortality in Autoimmune Myositis With and Without Associated Interstitial Lung Disease. *Lung,* 194**,** 733-7.

Jones, N. L., L. Makrides, C. Hitchcock, T. Chypchar & N. McCartney (1985) Normal standards for an incremental progressive cycle ergometer test. *Am Rev Respir Dis,* 131**,** 700-8.

Kaarteenaho, R. (2013) The current position of surgical lung biopsy in the diagnosis of idiopathic pulmonary fibrosis. *Respir Res,* 14**,** 43.

Karimi-Shah, B. A. & B. A. Chowdhury (2015) Forced vital capacity in idiopathic pulmonary fibrosis-FDA review of pirfenidone and nintedanib. *N Engl J Med,* 372**,** 1189-91.

Kawut, S. M., M. K. O'Shea, M. N. Bartels, J. S. Wilt, J. R. Sonett & S. M. Arcasoy (2005) Exercise testing determines survival in patients with diffuse parenchymal lung disease evaluated for lung transplantation. *Respir Med,* 99**,** 1431-9.

Key, A. L., K. Holt, A. Hamilton, J. A. Smith & J. E. Earis (2010) Objective cough frequency in Idiopathic Pulmonary Fibrosis. *Cough,* 6**,** 4.

Kim, B. J., Y. Kim, J. Oh, J. Jang & S. M. Kang (2019) Characteristics and Safety of Cardiopulmonary Exercise Testing in Elderly Patients with Cardiovascular Diseases in Korea. *Yonsei Med J,* 60**,** 547-553.

Kim, D. S., H. R. Collard & T. E. King (2006) Classification and natural history of the idiopathic interstitial pneumonias. *Proc Am Thorac Soc,* 3**,** 285-92.

Kim, H. J., D. Perlman & R. Tomic (2015) Natural history of idiopathic pulmonary fibrosis. *Respir Med,* 109**,** 661-70.

King, T. E., W. Z. Bradford, S. Castro-Bernardini, E. A. Fagan, I. Glaspole, M. K. Glassberg, E. Gorina, P. M. Hopkins, D. Kardatzke, L. Lancaster, D. J. Lederer, S. D. Nathan, C. A. Pereira, S. A. Sahn, R. Sussman, J. J. Swigris, P. W. Noble & A. S. Group (2014) A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med,* 370**,** 2083-92.

King, T. E., S. Safrin, K. M. Starko, K. K. Brown, P. W. Noble, G. Raghu & D. A. Schwartz (2005) Analyses of efficacy end points in a controlled trial of interferon-gamma1b for idiopathic pulmonary fibrosis. *Chest,* 127**,** 171-7.

King, T. E., J. A. Tooze, M. I. Schwarz, K. R. Brown & R. M. Cherniack (2001) Predicting survival in idiopathic pulmonary fibrosis: scoring system and survival model. *Am J Respir Crit Care Med,* 164**,** 1171-81.

Kleber, F. X. & P. J. Köln (2018) Oxygen consumption trajectory flattening-yet another cardiopulmonary exercise testing parameter in chronic heart failure. *Eur J Heart Fail,* 20**,** 1125-1127.

Kolb, M. & H. R. Collard (2014) Staging of idiopathic pulmonary fibrosis: past, present and future. *Eur Respir Rev,* 23**,** 220-4.

Kollert, F., B. Geck, R. Suchy, R. A. Jörres, M. Arzt, D. Heidinger, O. W. Hamer, A. Prasse, J. MüllerQuernheim, M. Pfeifer & S. Budweiser (2011) The impact of gas exchange measurement during exercise in pulmonary sarcoidosis. *Respir Med,* 105**,** 122-9.

Krauss, E., G. Gehrken, F. Drakopanagiotakis, S. Tello, R. C. Dartsch, O. Maurer, A. Windhorst, D. von der Beck, M. Griese, W. Seeger & A. Guenther (2019) Clinical characteristics of patients with familial idiopathic pulmonary fibrosis (f-IPF). *BMC Pulm Med,* 19**,** 130.

Laboratories, A. C. o. P. S. f. C. P. F. (2002) ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med,* 166**,** 111-7.

Layton, A. M., H. F. Armstrong, H. P. Kim, K. S. Meza, F. D'Ovidio & S. M. Arcasoy (2017) Cardiopulmonary exercise factors predict survival in patients with advanced interstitial lung disease referred for lung transplantation. *Respir Med,* 126**,** 59-67.

Lederer, D. J., S. M. Arcasoy, J. S. Wilt, F. D'Ovidio, J. R. Sonett & S. M. Kawut (2006) Six-minute-walk distance predicts waiting list survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med,* 174**,** 659-64.

Lee, S. H., J. S. Park, S. Y. Kim, D. S. Kim, Y. W. Kim, M. P. Chung, S. T. Uh, C. S. Park, S. W. Park, S. H. Jeong, Y. B. Park, H. L. Lee, J. W. Shin, E. J. Lee, J. H. Lee, Y. Jegal, H. K. Lee, Y. H. Kim, J. W. Song & M. S. Park (2018) Comparison of CPI and GAP models in patients with idiopathic pulmonary fibrosis: a nationwide cohort study. *Sci Rep,* 8**,** 4784.

Ley, B., H. R. Collard & T. E. King (2011) Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med,* 183**,** 431-40.

Ley, B., C. J. Ryerson, E. Vittinghoff, J. H. Ryu, S. Tomassetti, J. S. Lee, V. Poletti, M. Buccioli, B. M. Elicker, K. D. Jones, T. E. King & H. R. Collard (2012) A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Ann Intern Med,* 156**,** 684-91.

Lopes, A. J., S. L. Menezes, C. M. Dias, J. F. Oliveira, M. R. Mainenti & F. S. Guimarães (2012) Cardiopulmonary exercise testing variables as predictors of long-term outcome in thoracic sarcoidosis. *Braz J Med Biol Res,* 45**,** 256-63.

Luoto, J., M. Pihlsgård, P. Wollmer & S. Elmståhl (2019) Relative and absolute lung function change in a general population aged 60-102 years. *Eur Respir J,* 53.

Mackintosh, J. A., S. R. Desai, H. Adamali, K. Patel, F. Chua, A. Devaraj, V. Kouranos, M. Kokosi, G. Margaritopoulos, E. A. Renzoni, A. U. Wells, P. L. Molyneaux, S. Kumar, T. M. Maher & P. M. George (2019) In patients with idiopathic pulmonary fibrosis the presence of hiatus hernia is associated with disease progression and mortality. *Eur Respir J,* 53.

Maldonado, F., S. K. Danoff, A. U. Wells, T. V. Colby, J. H. Ryu, M. Liberman, M. M. Wahidi, L. Frazer, J. Hetzel, O. B. Rickman, F. J. F. Herth, V. Poletti & L. B. Yarmus (2020) Transbronchial Cryobiopsy for the Diagnosis of Interstitial Lung Diseases: CHEST Guideline and Expert Panel Report. *Chest,* 157**,** 1030-1042.

McElroy, M. C. & M. Kasper (2004) The use of alveolar epithelial type I cell-selective markers to investigate lung injury and repair. *Eur Respir J,* 24**,** 664-73.

McHugh, M. L. (2012) Interrater reliability: the kappa statistic. *Biochem Med (Zagreb),* 22**,** 276-82.

Meltzer, E. B. & P. W. Noble (2008) Idiopathic pulmonary fibrosis. *Orphanet J Rare Dis,* 3**,** 8.

Mezzani, A. (2017) Cardiopulmonary Exercise Testing: Basics of Methodology and Measurements. *Ann Am Thorac Soc,* 14**,** S3-S11.

Miki, K., R. Maekura, T. Hiraga, Y. Okuda, T. Okamoto, A. Hirotani & T. Ogura (2003) Impairments and prognostic factors for survival in patients with idiopathic pulmonary fibrosis. *Respir Med,* 97**,** 482-90.

Mikolasch, T., J. Sahota & H. Garthwaite. 2018. Neutrophil lymphocyte ratio (NLR) as a predictive biomarker in idiopathic pulmonary fibrosis (IPF). Thorax 2018; 73: A88-A89.

Moher, D., A. Liberati, J. Tetzlaff, D. G. Altman & P. Group (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol,* 62**,** 1006-12.

Molgat-Seon, Y., M. R. Schaeffer, C. J. Ryerson & J. A. Guenette (2020) Cardiopulmonary Exercise Testing in Patients With Interstitial Lung Disease. *Front Physiol,* 11**,** 832.

Morrow, T. (2019) Improving outcomes and managing costs in idiopathic pulmonary fibrosis. *Am J Manag Care,* 25**,** S204-S209.

Mura, M., M. A. Porretta, E. Bargagli, G. Sergiacomi, M. Zompatori, N. Sverzellati, A. Taglieri, F. Mezzasalma, P. Rottoli, C. Saltini & P. Rogliani (2012) Predicting survival in newly diagnosed idiopathic pulmonary fibrosis: a 3-year prospective study. *Eur Respir J,* 40**,** 101-9.

Nakamura, Y. & T. Suda (2015) Idiopathic Pulmonary Fibrosis: Diagnosis and Clinical Manifestations. *Clin Med Insights Circ Respir Pulm Med,* 9**,** 163-71.

Navaratnam, V., K. M. Fleming, J. West, C. J. Smith, R. G. Jenkins, A. Fogarty & R. B. Hubbard (2011) The rising incidence of idiopathic pulmonary fibrosis in the U.K. *Thorax,* 66**,** 462-7.

Ney, M., M. J. Haykowsky, B. Vandermeer, A. Shah, M. Ow & P. Tandon (2016) Systematic review: pre- and post-operative prognostic value of cardiopulmonary exercise testing in liver transplant candidates. *Aliment Pharmacol Ther,* 44**,** 796-806.

Noble, P. W., C. Albera, W. Z. Bradford, U. Costabel, M. K. Glassberg, D. Kardatzke, T. E. King, L. Lancaster, S. A. Sahn, J. Szwarcberg, D. Valeyre, R. M. du Bois & C. S. Group (2011) Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet,* 377**,** 1760-9.

Nolan, C. M., M. Maddocks, T. M. Maher, W. Banya, S. Patel, R. E. Barker, S. E. Jones, P. M. George, P. Cullinan & W. D. C. Man (2019) Gait speed and prognosis in patients with idiopathic pulmonary fibrosis: a prospective cohort study. *European Respiratory Journal,* 53**,** 1801186.

O'Donnell, D. E., J. Ora, K. A. Webb, P. Laveneziana & D. Jensen (2009) Mechanisms of activityrelated dyspnea in pulmonary diseases. *Respir Physiol Neurobiol,* 167**,** 116-32.

Ong, K. C., J. P. Benedicto, A. H. Chan, Y. S. Tan & Y. Y. Ong (2000) Cardiopulmonary exercise testing in heart transplant candidates. *Ann Acad Med Singap,* 29**,** 442-6.

Palange, P., S. A. Ward, K. H. Carlsen, R. Casaburi, C. G. Gallagher, R. Gosselink, D. E. O'Donnell, L. Puente-Maestu, A. M. Schols, S. Singh, B. J. Whipp & E. T. Force (2007) Recommendations on the use of exercise testing in clinical practice. *Eur Respir J,* 29**,** 185-209.

Panagiotou, M., V. Polychronopoulos & C. Strange (2016) Respiratory and lower limb muscle function in interstitial lung disease. *Chron Respir Dis,* 13**,** 162-72.

Paolocci, G., I. Folletti, K. Torén, M. Ekström, M. Dell'Omo, G. Muzi & N. Murgia (2018) Occupational risk factors for idiopathic pulmonary fibrosis in Southern Europe: a case-control study. *BMC Pulm Med,* 18**,** 75.

Patel, A. S., R. J. Siegert, K. Brignall, P. Gordon, S. Steer, S. R. Desai, T. M. Maher, E. A. Renzoni, A. U. Wells, I. J. Higginson & S. S. Birring (2012) The development and validation of the King's Brief Interstitial Lung Disease (K-BILD) health status questionnaire. *Thorax,* 67**,** 804-10.

Patel, A. S., R. J. Siegert, D. Creamer, G. Larkin, T. M. Maher, E. A. Renzoni, A. U. Wells, I. J. Higginson & S. S. Birring (2013) The development and validation of the King's Sarcoidosis Questionnaire for the assessment of health status. *Thorax,* 68**,** 57-65.

Patel, N., A. G. Powell, J. R. Wheat, C. Brown, I. R. Appadurai, R. G. Davies, D. M. Bailey & W. G. Lewis (2019) Cardiopulmonary fitness predicts postoperative major morbidity after esophagectomy for patients with cancer. *Physiol Rep,* 7**,** e14174.

Poormoghim, H., M. Lucas, N. Fertig & T. A. Medsger (2000) Systemic sclerosis sine scleroderma: demographic, clinical, and serologic features and survival in forty-eight patients. *Arthritis Rheum,* 43**,** 444-51.

Quanjer, P. H., G. J. Tammeling, J. E. Cotes, L. M. Fabbri, H. Matthys, O. F. Pedersen, R. Peslin, J. Roca, P. J. Sterk & W. T. Ulmer (1993) Symbols, abbreviations and units. Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. *Eur Respir J Suppl,* 16**,** 85-100.

Radtke, T., S. Crook, G. Kaltsakas, Z. Louvaris, D. Berton, D. S. Urquhart, A. Kampouras, R. A. Rabinovich, S. Verges, D. Kontopidis, J. Boyd, T. Tonia, D. Langer, J. De Brandt, Y. M. J. Goërtz, C. Burtin, M. A. Spruit, D. C. W. Braeken, S. Dacha, F. M. E. Franssen, P. Laveneziana, E. Eber, T. Troosters, J. A. Neder, M. A. Puhan, R. Casaburi, I. Vogiatzis & H. Hebestreit (2019) ERS statement on standardisation of cardiopulmonary exercise testing in chronic lung diseases. *Eur Respir Rev,* 28.

Raghu, G., H. R. Collard, J. J. Egan, F. J. Martinez, J. Behr, K. K. Brown, T. V. Colby, J. F. Cordier, K. R. Flaherty, J. A. Lasky, D. A. Lynch, J. H. Ryu, J. J. Swigris, A. U. Wells, J. Ancochea, D. Bouros, C. Carvalho, U. Costabel, M. Ebina, D. M. Hansell, T. Johkoh, D. S. Kim, T. E. King, Y. Kondoh, J. Myers, N. L. Müller, A. G. Nicholson, L. Richeldi, M. Selman, R. F. Dudden, B. S. Griss, S. L. Protzko, H. J. Schünemann & A. E. J. A. C. o. I. P. Fibrosis (2011) An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med,* 183**,** 788-824.

Raghu, G., T. D. Freudenberger, S. Yang, J. R. Curtis, C. Spada, J. Hayes, J. K. Sillery, C. E. Pope & C. A. Pellegrini (2006a) High prevalence of abnormal acid gastro-oesophageal reflux in idiopathic pulmonary fibrosis. *Eur Respir J,* 27**,** 136-42.

Raghu, G., M. Remy-Jardin, J. L. Myers, L. Richeldi, C. J. Ryerson, D. J. Lederer, J. Behr, V. Cottin, S. K. Danoff, F. Morell, K. R. Flaherty, A. Wells, F. J. Martinez, A. Azuma, T. J. Bice, D. Bouros, K. K. Brown, H. R. Collard, A. Duggal, L. Galvin, Y. Inoue, R. G. Jenkins, T. Johkoh, E. A. Kazerooni, M. Kitaichi, S. L. Knight, G. Mansour, A. G. Nicholson, S. N. J. Pipavath, I. Buendía-Roldán, M. Selman, W. D. Travis, S. Walsh, K. C. Wilson & E. R. S. American Thoracic Society, J.panese Respiratory Society, and Latin American Thoracic Society (2018) Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med,* 198**,** e44-e68.

Raghu, G., D. Weycker, J. Edelsberg, W. Z. Bradford & G. Oster (2006b) Incidence and prevalence of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med,* 174**,** 810-6.

Rapp, D., J. Scharhag, S. Wagenpfeil & J. Scholl (2018) Reference values for peak oxygen uptake: cross-sectional analysis of cycle ergometry-based cardiopulmonary exercise tests of 10 090 adult German volunteers from the Prevention First Registry. *BMJ Open,* 8**,** e018697.

Ren, F., T. Zhao, B. Liu & L. Pan (2019) Neutrophil-lymphocyte ratio (NLR) predicted prognosis for advanced non-small-cell lung cancer (NSCLC) patients who received immune checkpoint blockade (ICB). *Onco Targets Ther,* 12**,** 4235-4244.

Richeldi, L., R. M. du Bois, G. Raghu, A. Azuma, K. K. Brown, U. Costabel, V. Cottin, K. R. Flaherty, D. M. Hansell, Y. Inoue, D. S. Kim, M. Kolb, A. G. Nicholson, P. W. Noble, M. Selman, H. Taniguchi, M. Brun, F. Le Maulf, M. Girard, S. Stowasser, R. Schlenker-Herceg, B. Disse, H. R. Collard & I. T. Investigators (2014) Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med,* 370**,** 2071-82.

Richeldi, L., C. J. Ryerson, J. S. Lee, P. J. Wolters, L. L. Koth, B. Ley, B. M. Elicker, K. D. Jones, T. E. King, J. H. Ryu & H. R. Collard (2012) Relative versus absolute change in forced vital capacity in idiopathic pulmonary fibrosis. *Thorax,* 67**,** 407-11.

Rodríguez-Portal, J. A. (2018) Efficacy and Safety of Nintedanib for the Treatment of Idiopathic Pulmonary Fibrosis: An Update. *Drugs R D,* 18**,** 19-25.

Russell A, Wickremasinghe M, Renzoni E, Adamali H & F. S. Boreal Z, Maher TM, Ng Man Kwong G, Saketkoo LA, Fleming S, Cullinan P. 2018. The idiopathic pulmonary fibrosis patients reported outcome measure (IPF-PROM) is reliable and valid for use in populations with IPF. In *British Thoracic Society*, A47. Thorax.

Russell, A.-M. 2017. Development and preliminary Testing of the Idiopathic Pulmonary Fibrosis Patient Reported Outcome Measure (IPF-PROM): UK and Ireland multi-centre Study. American Journal od Respiratory and Critical Care Medicine 2017;195:A1031

Ryu, J. H., T. V. Colby, T. E. Hartman & R. Vassallo (2001) Smoking-related interstitial lung diseases: a concise review. *Eur Respir J,* 17**,** 122-32.

S, C., H. A & P. B. 2000. *Regression Analysis by Example*.

SCADDING, J. G. (1961) Prognosis of intrathoracic sarcoidosis in England. A review of 136 cases after five years' observation. *Br Med J,* 2**,** 1165-72.

Schraufnagel, D. E. & P. Agostoni (2017) Cardiopulmonary Exercise Testing. *Ann Am Thorac Soc,* 14**,** S1-S2.

Schroedl, C. J., S. E. Yount, E. Szmuilowicz, P. J. Hutchison, S. R. Rosenberg & R. Kalhan (2014) A qualitative study of unmet healthcare needs in chronic obstructive pulmonary disease. A potential role for specialist palliative care? *Ann Am Thorac Soc,* 11**,** 1433-8.

Selman, M. & A. Pardo (2014) Revealing the pathogenic and aging-related mechanisms of the enigmatic idiopathic pulmonary fibrosis. an integral model. *Am J Respir Crit Care Med,* 189**,** 1161-72.

Shamseer, L., D. Moher, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew, P. Shekelle, L. A. Stewart & P.P. Group (2015) Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ,* 350**,** g7647.

Sharp, C., H. I. Adamali & A. B. Millar (2017) A comparison of published multidimensional indices to predict outcome in idiopathic pulmonary fibrosis. *ERJ Open Res,* 3.

Shaw, J., T. Marshall, H. Morris, C. Hayton & N. Chaudhuri (2017) Idiopathic pulmonary fibrosis: a holistic approach to disease management in the antifibrotic age. *J Thorac Dis,* 9**,** 4700-4707.

Sinha, A., A. S. Patel, R. J. Siegert, S. Bajwah, T. M. Maher, E. A. Renzoni, A. U. Wells, I. J. Higginson & S. S. Birring (2019) The King's Brief Interstitial Lung Disease (KBILD) questionnaire: an updated minimal clinically important difference. *BMJ Open Respir Res,* 6**,** e000363.

Snell, N., D. Strachan & R. Hubbard. 2016. Epidemiology of idiopathic pulmonary fibrosis in the UK: findings from the British lung foundation's 'respiratory health of the nation' project. Thorax. Snyder, L., M. L. Neely, A. S. Hellkamp, E. O'Brien, J. de Andrade, C. S. Conoscenti, T. Leonard, S. Bender, M. Gulati, D. A. Culver, R. J. Kaner, S. Palmer, H. J. Kim & I.-P. R. investigators (2019) Predictors of death or lung transplant after a diagnosis of idiopathic pulmonary fibrosis: insights from the IPF-PRO Registry. *Respir Res,* 20**,** 105.

Society, A. T. & A. C. o. C. Physicians (2003) ATS/ACCP Statement on cardiopulmonary exercise testing. *Am J Respir Crit Care Med,* 167**,** 211-77.

Society, A. T. & E. R. Society (2002) American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. *Am J Respir Crit Care Med,* 165**,** 277-304.

Steen, V. D. & T. A. Medsger (2007) Changes in causes of death in systemic sclerosis, 1972-2002. *Ann Rheum Dis,* 66**,** 940-4.

Stewart, L., D. Moher & P. Shekelle (2012) Why prospective registration of systematic reviews makes sense. *Syst Rev,* 1**,** 7.

Stickland, M. K., S. J. Butcher, D. D. Marciniuk & M. Bhutani (2012) Assessing exercise limitation using cardiopulmonary exercise testing. *Pulm Med,* 2012**,** 824091.

Sue, D. Y. & K. Wasserman (1991) Impact of integrative cardiopulmonary exercise testing on clinical decision making. *Chest,* 99**,** 981-92.

Swigris, J. J., D. Esser, C. S. Conoscenti & K. K. Brown (2014) The psychometric properties of the St George's Respiratory Questionnaire (SGRQ) in patients with idiopathic pulmonary fibrosis: a literature review. *Health Qual Life Outcomes,* 12**,** 124.

Swigris, J. J., W. G. Kuschner, S. S. Jacobs, S. R. Wilson & M. K. Gould (2005) Health-related quality of life in patients with idiopathic pulmonary fibrosis: a systematic review. *Thorax,* 60**,** 588-94.

Swigris, J. J., X. Zhou, F. S. Wamboldt, R. du Bois, R. Keith, A. Fischer, G. P. Cosgrove, S. K. Frankel, D. Curran-Everett & K. K. Brown (2009) Exercise peripheral oxygen saturation (SpO2) accurately reflects arterial oxygen saturation (SaO2) and predicts mortality in systemic sclerosis. *Thorax,* 64**,** 626-30.

Thomas, E. T., M. Guppy, S. E. Straus, K. J. L. Bell & P. Glasziou (2019) Rate of normal lung function decline in ageing adults: a systematic review of prospective cohort studies. *BMJ Open,* 9**,** e028150.

Thomas, M. & K. Sylvester (2020) An RER of 1.05 should not be used to determine maximal effort during CPET. *European Respiratory Journal,* 56**,** 3787.

Torrisi, S. E., B. Ley, M. Kreuter, M. Wijsenbeek, E. Vittinghoff, H. R. Collard & C. Vancheri (2019) The added value of comorbidities in predicting survival in idiopathic pulmonary fibrosis: a multicentre observational study. *Eur Respir J,* 53.

Triantafillidou, C., E. Manali, P. Lyberopoulos, L. Kolilekas, K. Kagouridis, S. Gyftopoulos, K. Vougas, A. Kotanidou, M. Alchanatis, A. Karakatsani & S. A. Papiris (2013) The Role of Cardiopulmonary Exercise Test in IPF Prognosis. *Pulm Med,* 2013**,** 514817.

Vainshelboim, B. (2016) Exercise training in idiopathic pulmonary fibrosis: is it of benefit? *Breathe (Sheff),* 12**,** 130-8.

Vainshelboim, B., J. Oliveira, B. D. Fox & M. R. Kramer (2016) The Prognostic Role of Ventilatory Inefficiency and Exercise Capacity in Idiopathic Pulmonary Fibrosis. *Respir Care,* 61**,** 1100-9.

van der Plas, M. N., C. van Kan, J. Blumenthal, H. M. Jansen, A. U. Wells & P. Bresser (2014) Pulmonary vascular limitation to exercise and survival in idiopathic pulmonary fibrosis. *Respirology,* 19**,** 269-275.

Wagner, P. D. (1996) Determinants of maximal oxygen transport and utilization. *Annu Rev Physiol,* 58**,** 21-50.

Wallaert, B., A. Guetta, L. Wemeau-Stervinou, G. Tercé, M. Valette, R. Nevière & B. Aguilaniu (2011) [Prognostic value of clinical exercise testing in idiopathic pulmonary fibrosis]. *Rev Mal Respir,* 28**,** 290-6.

Walsh, S. L. F., A. U. Wells, S. R. Desai, V. Poletti, S. Piciucchi, A. Dubini, H. Nunes, D. Valeyre, P. Y. Brillet, M. Kambouchner, A. Morais, J. M. Pereira, C. S. Moura, J. C. Grutters, D. A. van den Heuvel, H. W. van Es, M. F. van Oosterhout, C. A. Seldenrijk, E. Bendstrup, F. Rasmussen, L. B. Madsen, B. Gooptu, S. Pomplun, H. Taniguchi, J. Fukuoka, T. Johkoh, A. G. Nicholson, C. Sayer, L. Edmunds, J. Jacob, M. A. Kokosi, J. L. Myers, K. R. Flaherty & D. M. Hansell (2016) Multicentre evaluation of multidisciplinary team meeting agreement on diagnosis in diffuse parenchymal lung disease: a case-cohort study. *Lancet Respir Med,* 4**,** 557-565.

Wapenaar, M., A. S. Patel, S. S. Birring, R. T. V. Domburg, E. W. Bakker, V. Vindigni, C. M. Sköld, V. Cottin, C. Vancheri & M. S. Wijsenbeek (2017) Translation and validation of the King's Brief Interstitial Lung Disease (K-BILD) questionnaire in French, Italian, Swedish, and Dutch. *Chron Respir Dis,* 14**,** 140-150.

Wasserman, K. 2012. *Principles of exercise testing and interpretation : including pathophysiology and clinical applications*. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Wasserman, K., W. W. Stringer, R. Casaburi, A. Koike & C. B. Cooper (1994) Determination of the anaerobic threshold by gas exchange: biochemical considerations, methodology and physiological effects. *Z Kardiol,* 83 Suppl 3**,** 1-12.

Watters, L. C., T. E. King, M. I. Schwarz, J. A. Waldron, R. E. Stanford & R. M. Cherniack (1986) A clinical, radiographic, and physiologic scoring system for the longitudinal assessment of patients with idiopathic pulmonary fibrosis. *Am Rev Respir Dis,* 133**,** 97-103.

Wells, A. U., J. Behr, U. Costabel, V. Cottin, V. Poletti, L. Richeldi & E. I. C. Group (2012) Hot of the breath: mortality as a primary end-point in IPF treatment trials: the best is the enemy of the good. *Thorax,* 67**,** 938-40.

Wells, A. U., S. R. Desai, M. B. Rubens, N. S. Goh, D. Cramer, A. G. Nicholson, T. V. Colby, R. M. du Bois & D. M. Hansell (2003) Idiopathic pulmonary fibrosis: a composite physiologic index derived from disease extent observed by computed tomography. *Am J Respir Crit Care Med,* 167**,** 962-9.

Whittingham, M. J., P. A. Stephens, R. B. Bradbury & R. P. Freckleton (2006) Why do we still use stepwise modelling in ecology and behaviour? *J Anim Ecol,* 75**,** 1182-9.

Wu, W., S. Jordan, M. O. Becker, R. Dobrota, B. Maurer, H. Fretheim, S. Ye, E. Siegert, Y. Allanore, A. M. Hoffmann-Vold & O. Distler (2018) Prediction of progression of interstitial lung disease in patients with systemic sclerosis: the SPAR model. *Ann Rheum Dis,* 77**,** 1326-1332.

Yang, I. V., T. E. Fingerlin, C. M. Evans, M. I. Schwarz & D. A. Schwartz (2015) MUC5B and Idiopathic Pulmonary Fibrosis. *Ann Am Thorac Soc,* 12 Suppl 2**,** S193-9.

Yates, H., H. I. Adamali, N. Maskell, S. Barratt & C. Sharp (2018a) Visual analogue scales for interstitial lung disease: a prospective validation study. *QJM,* 111**,** 531-539.

--- (2018b) Visual analogue scales for interstitial lung disease: a prospective validation study. *QJM*.

Yin, Q., M. J. Strong, Y. Zhuang, E. K. Flemington, N. Kaminski, J. A. de Andrade & J. A. Lasky (2020) Assessment of viral RNA in idiopathic pulmonary fibrosis using RNA-seq. *BMC Pulm Med,* 20**,** 81.

Yorke, J., P. W. Jones & J. J. Swigris (2010) Development and validity testing of an IPF-specific version of the St George's Respiratory Questionnaire. *Thorax,* 65**,** 921-6.

Zaizen, Y., Y. Kohashi, K. Kuroda, K. Tabata, Y. Kitamura, A. Hebisawa, Y. Saito & J. Fukuoka (2019) Concordance between sequential transbronchial lung cryobiopsy and surgical lung biopsy in patients with diffuse interstitial lung disease. *Diagn Pathol,* 14**,** 131.

Zappala, C. J., P. I. Latsi, A. G. Nicholson, T. V. Colby, D. Cramer, E. A. Renzoni, D. M. Hansell, R. M. du Bois & A. U. Wells (2010) Marginal decline in forced vital capacity is associated with a poor outcome in idiopathic pulmonary fibrosis. *Eur Respir J,* 35**,** 830-6.

Appendices :

Appendix A: K-BILD Questionnaire

King's Brief ILD Questionnaire (K-BILD)

This questionnaire is designed to assess the impact of your lung disease on various aspects of your everyday life. Read each question carefully and answer by SELECTING the response that best applies to you. Please answer ALL questions, as honestly as you can.

PATIENT INFORMATION:

Patient Identifier: .....

Date: .....

1. In the last 2 weeks, I have been breathless climbing stairs or walking up an incline or hill.

1. Every time
2. Most times
3. Several Times
4. Sometimes
5. Occasionally
6. Rarely
7. Never

2. In the last 2 weeks, because of my lung condition, my chest has felt tight.

1. All of the time
2. Most of the time
3. A good bit of the time
4. Some of the time
5. A little of the time
6. Hardly any of the time
7. None of the time

3. In the last 2 weeks have you worried about the seriousness of your lung complaint?

1. All of the time
2. Most of the time
3. A good bit of the time
4. Some of the time
5. A little of the time
6. Hardly any of the time
7. None of the time

4. In the last 2 weeks have you avoided doing things that make you breathless?

1. All of the time
2. Most of the time
3. A good bit of the time
4. Some of the time
5. A little of the time
6. Hardly any of the time
7. None of the time

5. In the last 2 weeks have you felt in control of your lung condition?

1. None of the time
2. Hardly any of the time
3. A little of the time
4. Some of the time
5. A good bit of the time
6. Most of the time
7. All of the time

6. In the last 2 weeks, has your lung complaint made you feel fed up or down in the dumps?

1. All of the time
2. Most of the time
3. A good bit of the time
4. Some of the time
5. A little of the time
6. Hardly any of the time
7. None of the time

7. In the last 2 weeks, I have felt the urge to breathe, also known as 'air hunger'.

1. All of the time
2. Most of the time
3. A good bit of the time
4. Some of the time
5. A little of the time
6. Hardly any of the time
7. None of the time

8. In the last 2 weeks, my lung condition has made me feel anxious.

1. All of the time
2. Most of the time
3. A good bit of the time
4. Some of the time
5. A little of the time
6. Hardly any of the time
7. None of the time

9. In the last 2 weeks, how often have you experienced 'wheeze' or whistling sounds from your chest?

1. All of the time
2. Most of the time
3. A good bit of the time
4. Some of the time
5. A little of the time
6. Hardly any of the time
7. None of the time

10. In the last two weeks how much of the time have you felt your lung disease is getting worse?

1. All of the time
2. Most of the time
3. A good bit of the time
4. Some of the time
5. A little of the time
6. Hardly any of the time
7. None of the time

11. In the last 2 weeks has your lung condition interfered with your job or other daily tasks?

1. All of the time
2. Most of the time
3. A good bit of the time
4. Some of the time
5. A little of the time
6. Hardly any of the time
7. None of the time

12. In the last 2 weeks have you expected your lung complaint to get worse?

1. All of the time
2. Most of the time
3. A good bit of the time
4. Some of the time
5. A little of the time
6. Hardly any of the time
7. None of the time

13. In the last 2 weeks, how much has your lung condition limited you carrying things, for example, groceries?

1. All of the time
2. Most of the time
3. A good bit of the time
4. Some of the time
5. A little of the time
6. Hardly any of the time
7. None of the time

14. In the last 2 weeks, has your lung condition made you think more about the end of your life?

1. All of the time
2. Most of the time
3. A good bit of the time
4. Some of the time
5. A little of the time
6. Hardly any of the time
7. None of the time

15. Are you financially worse off because of your lung condition?

1. A significant amount
2. A large amount
3. A considerable amount
4. A reasonable amount
5. A small amount
6. Hardly at all
7. Not at all

Thank you for completing this questionnaire

Appendix B: IPF-PROM Questionnaire

Patient Identifier :

Date:

The Idiopathic Pulmonary Fibrosis Patient Reported Outcome Measure

This questionnaire is designed to help us learn more about how Idiopathic Pulmonary Fibrosis affects your life

The information and the answers you give will be treated with the utmost confidentiality

There are no right or wrong answers

Please read each item and place an **'X'** in the box that best matches your experience over the last two weeks

If you do not experience an item put an **'X'** in the 'none' box.

Please respond to all items.

We would like to thank you very much for taking the time to answer these questions and help us with our research

This research was supported by a research fellowship from the National Institute of Health Research, UK

During the last **two weeks** how would you rate your overall quality of life?

Excellent	Good	Fair	Poor	Very Poor
☐	☐	☐	☐	☐

During the last **two weeks**, because of IPF, how much of the time have you ……..

1. Felt that your breathing difficulties have affected your quality of life?

None of the time	A little of the time	Most of the time	All of the time
☐	☐	☐	☐

During the last **two weeks**, because of IPF, how much of the time have you ……..

2. Felt that your fear of getting out of breath has limited your daily life?

None of the time	A little of the time	Most of the time	All of the time
☐	☐	☐	☐

During the last **two weeks**, because of IPF, how much of the time have you ……..

3. Stopped you doing any of the things you like to do?

None of the time	A little of the time	Most of the time	All of the time
☐	☐	☐	☐

Please make sure you have put an 'X' in one box for each question before moving on to the next page…….

During the last **two weeks**

All of the time

, because of IPF, how much of the time have you……..

4. Felt breathless with gentle physical exercise?
   None of the time A little of the time Most of the time

During the last **two weeks**, because of IPF, how much of the time have you ……..

5. Stopped for breath when walking at your own pace on the flat level (e.g. along the pavement; at home)
   None of the time A little of the time Most of the time All of the time

During the last **two weeks**, because of IPF, how much of the time have you ……..

6. Felt breathless with any of your everyday activities?
   None of the time A little of the time Most of the time All of the time

Please make sure you have put an 'X' in one box for each question before moving on to the next page…….

During the last **two weeks**

All of the time

, how much of the time have you ……..

7. Felt that IPF has brought worry to your life?

None of the time A little of the time Most of the time

During the last **two weeks**, because of IPF, how much of the time have you ……..

8. Felt frightened about the future?

None of the time A little of the time Most of the time All of the time

During the last **two weeks**, how much of the time ……..

9. Has it been difficult to manage the uncertainty of living with IPF?

None of the time A little of the time Most of the time All of the time

Please make sure you have put an 'X' in one box for each question before moving on to the next page…….

During the last **two weeks**

All of the time

, because of IPF, how much of the time have you ……..

10. Felt lethargic?
    None of the time A little of the time Most of the time

During the last **two weeks**, because of IPF, how much of the time have you ……..

11. Felt too tired to do your usual everyday activities?
    None of the time A little of the time Most of the time All of the time

During the last **two weeks**, because of IPF, how much of the time have you ……..

12. Felt frustrated by being too tired to do the things you like to do?
    None of the time A little of the time Most of the time All of the time

Please make sure you have put an 'X' in one box for each question

Thank you for completing this questionnaire

During the last **two weeks**

All of the time

Appendix C: Literature review study selection

Author, Date & Origin	Description	Study population & attrition	CPET method and CPET parameters	Exclusion	Disease outcomes	Statistical methods to investigate CPET & outcome	Summary of key reported outcomes	Comments
King et al. 2001 USA	Retrospective analysis of clinical, radiological and physiological parameters predicting survival in IPF. Median follow up 20 months (maximum 14.8 years).	238 IPF pts with histological UIP. 80 pts excluded from the final model derivation.	Cycle ergometer, blood gas analysis. P (A-a)O2 corrected for FiO2, VD/VT, VO2, maximal work load	CTD, left ventricular failure, occupational and environmental exposure, or history of drug exposure known to cause pulmonary fibrosis. Incomplete case records.	Survival (defined as death or time of censoring: Censored if still alive at last contact n=79, received single LTx n=11, double LTx n=1, or heart and LTx n=1 or e) died from other cause than IPF n=12).	Kaplan-Meier survival curves developed for group as whole and stratified by sex, age and smoking status. Univariate Cox proportional hazards regression analysis (adjusted for age and smoking) for each variable. Variables with p<0.25 included in multivariate analysis. Pearson's correlation to avoid multicollinearity. Forward elimination process used to develop preliminary model. mulitvariable influential points removed. Composite scoring system developed, weighting categories according to p values and HR, and using Akaike's Information criteria.	155 D (125 IPF, 19 other causes, 11 unknown and attributed to IPF). 105 patients censored (n=79 alive at time of analysis, n=13 LTx, n=12 non-IPF deaths, n=1 lost to follow up). Composite scoring model developed to predict survival in IPF which included age, smoking history, clubbing, extent of profusion of interstitial opacities, presence/absence of PH on CXR, % predicted TLC and PaO2 at the end of maximal exercise. Exercise PaO2 only exercise variable included in the model, accounting for 10.5% of score (PaO2 maximal exercise HR 0.74, CI 0.67-0.82, p<0.0001).	CPET performed in study as part of wider analysis of predictive factors in IPF. Histological UIP increased potential selection bias of a less severe IPF population. The radiological component of the scoring system used CXR rather than HRCT (HRCT not obtained during early years of the study). Only 158/238 (66%) of the original cohort were used to derive the complete model and thus possibility for selection bias.
Miki et al. 2003 Japan	Retrospective study: Evaluation of the predictive value of CPET for IPF respiratory deaths. Mean follow up 2.7 years (7.2 months - 9.0 years).	41 IPF pts.	Exercise treadmill (Sheffield protocol). *Pa*O2, *Pa*CO2, HR, respiratory frequency (f), *Vt*, VE, peak VO2, VE/VO2, VE/VCO2, VO2/HR, *Aa DO2* & *Pa*O2-slope.	CTD, Sarcoid, OP, EP, HP, Cardiac disease, anaemia, primary cardiac disease, anaemia, PVD, cancer, pleural/chest wall disorders including respiratory muscle weakness. Steroid or immunosuppressive treatment prior to study entry. Death from a non-respiratory cause during followup.	Respiratory death	Exercise parameters (between groups split by PaO2-slope) compared using Mann-Whitney. Univariate Cox proportional hazards model to compare initial parameters then entered into multiple regression analysis using stepwise evaluation. Relationship between Pao2 slope and other variables were analysed by linear regression with stepwise technique. Survival times compared using Kaplan Meier curves and statistical significance determined by log rank test.	23 respiratory deaths. Median survival 2.9 years. In univariate analysis, VO2 max (HR 0.997, 0.995-0.999 CI, p=0.012), VO2/HRR max) (HR 0.69, 0.51-0.93 CI, p= 0.014), *Pa*O2 slope (HR 0.68, 0.51-0.89 CI, p=0.006), VE/VCO2 (HR 1.04, 1.006-1.07 CI, p=0.020) & age (HR 1.1, 1.02-1.18 CI, p=0.014) associated with survival in IPF. On multiple regression, PaO2 slope (HR 0.84, 0.730.97 CI, p=0.015) and age (HR 1.096, 1.01-1.19 CI, p=0.025) independently related to survival. PaO2 at rest and during maximum exercise did not influence survival. When PaO2 slope was divided into steep (<- 60mmHg/l/min) and gentle (>-60mmHg/l/min), median survival time after CPET significantly shorter in steep group (1.6 vs 4.5 yrs).	Retrospective, single centre cohort. Large number of exclusion criteria. Outcomes limited to respiratory deaths. PaO2 slope (as an indicator of exercise induced hypoxaemia) had the greatest correlation with survival.
Kawut et al. 2005, USA	Retrospective study of CPET and 6MWTD variables associated with survival in pts referred for lung transplant. Median follow up 271 days (23-983).	51 pts with IIP or DPLD of known cause (e.g. drugs, occupational or environmental exposures, CTD) referred for lung transplant.	Cycle ergometer. Pulse oximetry. SaO2 (unloaded, peak, recovery), Peak VO2/kg, VO2/HR peak, VCO2 unloaded, VE unloaded.	Pts evaluated at another lung transplantation centre. Other forms of DPLD e.g. LAM, pulmonary Langerhans' cell histiocytosis/histiocyt osis X, EP and granulomatous DPLD e.g. sarcoidosis.	All-cause mortality. Death on the lung transplantat ion waiting list.	Cox proportional hazards regression to identify predictors of time-to-death. Individual models were constructed using LTx as a timedependant covariate to 'control' for receiving a LTx. ROC curve analysis was used to define cut-off for variables associated with dying on the transplantation list.	7 lung transplantations and 17 deaths (1 posttransplantation). 28/51 (55%) UIP/IPF, CTD-UIP (n=4), NSIP (n=6), HP (n=2), DIP (n=1), COP (n=1), LIP (n=1) and unclassifiable ILD (n=7). A 6MWTD <350m (HR 4.6, 1.5-14.2 CI, p=0.009), peak VO2/kg (HR 0.88, 0.790.99 CI, p=0.039) (no threshold determined) and VE/VCO2 >46 (p=0.05) (non-proportional and increased over time so could not be estimated with a single HR) were each associated with increased risk of death. Sp02 <95% during unloaded exercise had 75% chance of dying on transplantation list (sensitivity 86%, specificity 89%). 67% chance of death if 6MWTD <350m.	Retrospective single centre cohort. Only half pts reached AT which limited analysis (low number of endpoints). Additional oxygen use during CPET was variable. Generalisability questionable As highly selected cohort of severe ILD. Source population, patterns of referral to transplant centre, waiting times and cohort characteristics may differ from other transplant programs.

	Retrospective							
Swigris et al. 2009, USA	Retrospective study exploring prognostic role of SpO2 and SaO2 at rest and during maximal exercise in SSc-ILD exercise. Median follow up 7.1 years.	83 patients with SSc-ILD	Cycle ergometer. Blood gas analysis and pulse oximetry. SpO2 and SaO2 at rest and during maximal exercise (SpO2 max). VO2 max measured but not reported.	Pulmonary hypertension, overlap syndromes.	Mortality	Cox proportional hazard models were used to examine the prognostic capabilities of SpO2, dichotomised by <89% or ≥ 89% and also as continuous variables. Kaplan Meier survival curves were generated.	39 deaths (number of transplantations not recorded). In Cox proportional hazards models, SpO2 predicted mortality; SpO2max < 89% (HR 2.4, 95% CI 1.2 to 4.9, p =0.02), SpO2max fall >4% from baseline (HR 2.4, 95% CI 1.1 to 5.0, p = 0.02), alongside ΔSp02 (HR 1.08, 95% CI 1.03 to 1.14, p = 0.002). Controlling for FVC%, the ΔSp02 remained a significant predictor of mortality (HR 1.07, 95% CI 1.01 to 1.14, p = 0.02). No other CPET variables reported.	No other CPET variables described in analysis and thus potential for reporting bias.
Fell et al. 2009, USA	study evaluating prognostic value of CPET in IPF. Mean follow up not reported.	117 IPF pts. 10 pts excluded from survival analysis as VO2 max changed between baseline and 6 months.	Cycle ergometer. Blood gas analysis. Peak VO2/kg	Patients with CTD, occupational or environmental exposure, histological pattern other than UIP	Survival	Multivariate Cox proportional hazard models studied the predictive value of peak VO2 adjusting for age, gender, smoking status, baseline FVC % and baseline DLCO%. Resulting HR were plotted against peak VO2 to determine thresholds. Survival thresholds examined with Kaplan Meier survival curves, log-rank tests and Multivariate Cox proportional hazard models.	Peak VO2/kg examined as a continuous variable did not predict survival HR 0.969 (p=0.55). However, a baseline threshold peak VO2 <8.3ml/kg/min was associated with an increased risk of death (n=8; HR 3.24 1.10-9.56 CI, p=0.03). No other CPET variables reported.	Retrospective, single centre study. Number of deaths in each group not reported. Analysis was not by a priori plan. Small number pts below VO2 max threshold in analysis. Caution in interpreting generalisability to IPF population as 64% (75/117) required a surgical lung biopsy for diagnosis. No other CPET outcomes reported.
Wallaert et al. 2011, France	Retrospective multicentre study evaluating prognostic role of CPET in determining 3-year survival in IPF.	63 IPF patients	Cycle ergometer. Blood gas analysis. Peak VO2/kg, VE/VO2 at ventilatory threshold, VE/CO2, (VO2/HRR), P(A-a)O2, ventilatory reserve and lactate.	Non-IPF associated ILD. Pts in which blood gas analysis had not been performed.	3-year survival (absence of D or LTx).	Demographic data, resting pulmonary function and CPET parameters in the survivors were compared to the those who died/received lung transplantation by univariate survival analysis. Multivariate logistic regression analysis explored prognosis at 3 years. Kaplan Meier curve and log-rank test was performed, with model validation by ROC curve analysis.	19 patients: D (n=14) or LTx (n=5) at 3 years. Multivariate logistic regression analysis highlighted four parameters to be independently correlated with mortality: TLC (% pred), VE/VO2 at ventilatory threshold, FVC (% pred) and P(A-a)O2. The most appropriate logistic regression model incorporated two variables, with the lowest 3 year survival when TLC (<65%) and VE/VO2 at ventilatory threshold (>45) (AUC 0.811, sensitivity was 98%, specificity 50%, positive predictive value 80% and the negative predictive value 64%).	Retrospective study. Presence of PH not studied. Inadequate description of exclusion criteria.
Kollert et al. 2011, Germany	Retrospective study evaluating whether gas exchange during CPET reflects disease activity and clinical course in sarcoidosis. 2 year follow up	149 histologically confirmed sarcoidosis. Analysis of 102 patients (47 incomplete notes)	Cycle ergometer, capillary blood gas analysis. P (A-a)O2	Patients who could not complete CPET > 6 minutes, in the absence of extracardiopulmonary limitations. Patients with clinical signs of acute infection. For the longitudinal subgroup analysis: Patients with incomplete records	Longitudinal component: Duration of immunosuppressive therapy (no treatment, treatment ≤ 1 year, treatment > 1 year)	Associations between sarcoidosis clinical parameters (including the need for prolonged immunosuppressive therapy > 1 year) and P(A-a)O2 during exercise were assessed by analysis of variance statistical methodology. Univariate then multivariate backward binary logistic regression analysis used to assess clinical variables independently associated with need for prolonged immunosuppression.	Multivariate regression analysis suggested that FVC (OR 0.954, 0.917-0.992 CI, p =0.009) and P(A-a)O2 (OR 1.098, 1.039-1.160 CI, p<0.0001) during exercise were independently associated with a need for prolonged immunosuppressive treatment. Unable to determine exact clinical characteristics of this longitudinal cohort from the data presented. No other CPET variables reported.	No other CPET variables described in analysis and thus potential for reporting bias.
Lopes et al. 2012, Brazil	Retrospective study to identify CPET measures that predict FVC and DLco progression over 5 years in patients with thoracic sarcoidosis.	42 pts with histologically confirmed sarcoidosis.	Cycle ergometer, blood gas analysis. Peak VO2 (% pred), % peak VO2 at lactate threshold,VCO2/VO2, VO2/HRR, maximum respiratory rate, breathing reserve, HRR, P(A-a)O2, ΔSp02, Δlactate.	History of smoking. Mycobacterial infection, exposure to aero-contaminants or medications known to cause granulomatous disorders. Those with known medical history or laboratory diagnosis of concomitant respiratory, cardiac or neuromuscular disease.	Decline FVC% and DLco%	FVC/DLCO variation over study period evaluated by Wilcoxon signed rank test. Correlations between CPET measures and FVC/DLCO variation over 5years used Spearman's rank correlation (except breathing reserve and relative variations of FVC). ROC curve analysis used to determine cut offs for CPET measurements are predictors for lung function decline. Multiple logistic regression used to identify factors independently related to decreased lung function.	Statistically significant reductions in FVC (relative variation -5.1% (-23.1% - 0%) and DLCO (relative variation -2.5% (-44.4% - 0.93%) at 5 years follow up. Peak VO2 (% pred), breathing reserve, maximum RR, P(A-a)O2 and ΔSp02 correlated with FVC and DLCO values that had declined >10% from the initial values measured (p<0.0001 for all parameters). P(A-a)O2 >22mmHg (RR 70.0 p=0.001) and breathing reserve <40% (RR 20.8, p=0.014) independently predicted lung function decline (FVC % pred and DLCO% pred).	Retrospective, single centre study. Potential for selection bias (tertiary centre for sarcoid – more likely to have severe patients). Small number of patients resulting in high RR values. Cardiac circulatory status not determined.

	Retrospective							
Triantafillidou et al. 2013, Greece	Prospective, study evaluating prognostic role of 6MWT and CPET in IPF. Follow up 9-64 months.	25 pts with IPF	Cycle ergometer, pulse oximetry. VE/VCO2 slope, VO2 peak/kg, VE/VCO2 ratio at AT.	Significant PH (PASP>45mmHg on ECHO), pts taking beta blockers. Pulmonary fibrosis due to environmental and occupational exposure, drug toxicity or autoimmune rheumatological disease	Survival	Parameters of study were evaluated by Wald test, likelihood ratio test and the score (log rank) tests with Bonferroni correction. Parameters achieving statistical significance were then evaluated in a multiple regression Cox proportional hazard model with a stepwise model selection.	8 D by end of the observation period. 21 patients reached the AT. VE/VCO2 slope, VO2 peak/kg & VE/VCO2 at AT were significant survival predictors. Optimal model for mortality risk estimation combined VO2 peak/kg with DLCO (P<0.0001). Per 1 unit increase in VO2 peak/kg (1mL/kg min) and DLCO% (1%), mortality rate is reduced by 32% and 13% respectively. VO2 peak threshold of 14.2mL/min/kg was associated with an increased mortality risk.	Prospective study with low mortality rate in small numbers of pts. Data generated from sub-analysis of RCT.
Gläser et al. 2013, Germany	study evaluating predictive value of CPET measures for the presence of PH in IPF. Investigate diagnostic and prognostic use of gas exchange during CPET in pts with/out PH. Follow up 2 years.	135 pts (73 with PH) IPF. No follow up data for 2 pts, reducing cohort to 133.	Cycle ergometer, pulse oximetry. Peak VO2, VO2 at AT (ml/min), VE/MVV, VE vs VCO2 slope, VE max, Vt max, Vt max/ IC, VE/MVV.	Pts with left heart disease (ECHO ± PWP>14mmHg by RHC), non-IPF pulmonary fibrosis and/or PH resulting in a life expectancy <24 months, inability to perform CPET due to orthopaedic or neurological impairment.	Interceding pulmonary hypertension. Survival (death and lung transplantation combined endpoint)	Mann-Whitney or X-test used for comparison of IPF pts with/without PH. Cox proportional hazards analysis used for pulmonary variables and end point. Kaplan Meier survival plots constructed with differences in survival analysed by log-rank test. Cut off values for best discrimination determined using ROC curve analysis.	37 D and 6 LTx during follow up. The presence of PH is best predicted by gas exchange efficiency during exercise and peak oxygen uptake (VE vsVO2 slope pred (≥ 152.4, AUC 0.938, 0.892-0.984 CI) and VO2 peak pred (≤ 56.3, AUC 0.832, 0.753-0.911 CI)). By univariate analysis, the presence of PH as determined by RHC was the most powerful prognosticator in IPF (whole group) (mPAP HR 1.07, 1.04-1.11 CI), with CPET outcomes of peak VO2 pred (HR 0.96 p=0.001) and VO2 at AT pred (HR 0.97 P=0.017) also being statistically significant. In multivariate analysis, invasively measured PH and peak VO2 pred were independent predictors for survival.	Retrospective multicentre study. Potential recruitment bias due to selected cohort (specialist centres, excluded left heart disease).
Van der Plas et al. 2014, Netherlands	Retrospective study exploring predictive value of CPET and ECHO parameters for survival in IPF. Mean follow-up 42.3+/- 42.2 months.	38 pts with IPF. Follow up for 3 pts who received transplantation was censored at date of transplantation.	Cycle ergometer. Peak workload (% predicted), VO2 peak (% pred), VE peak (% pred), breathing reserve (%), HRR peak (% pred), VE/ VCO2 ratio at AT, VO2/HRR (% pred), ETCO2 at max (kPa)	Non -IPF ILD. Pts where CPET and ECHO were performed more than 2 weeks apart.	Survival	Pearson's correlation coefficients were calculated for sPAP & CPET parameters. Patients were grouped into those with/without sPAP≥ 40mmHg and differences in exercise parameters analysed with unpaired *t*-test or chisquare test. ROC curve analysis was used to determine variables that predict sPAP≥ 40mmHg. Kaplan-Meier survival curves then evaluated the prognostic value of these parameters on survival. HRs were calculated using multivariate Cox proportional hazard models (with FVC and CPI included in the model to correct for functional severity of IPF) to determine predictive value of parameters on survival.	24 D and 3 LTx during follow up. 29/38 (765) had a reduced VO2 peak (ie. <84% predicted). Mean peak VO2 5.5ml/min/kg; 66.6% predicted). VE/VCO2 at AT was significantly higher in patients with sPAP ≥ 40mmHg (n=11) compared to those with sPAP≤ 40mmHg (n=27), (54.0±21.9 vs 37.9±7.5, p=0.021). VE/VCO2 at AT was shown to be a good predictor of sPAP ≥ 40mmHg by ROC curve analysis but only VE/VCO2 at AT and not sPAP ≥ 40mmHg was shown to predict survival. Pts with VE/VCO2 at AT ≤ 45 (n=24) had a significantly better prognosis that those with VE/VCO2 ≥ 45 (n=14), 81.3±14.1 vs 21.0±4.9 months respectively; HR 4.58, p=0.001. Parameters reflecting functional severity of IPF did not add to the predictive value of VE/VCO2 at AT for survival.	Retrospective analysis of prospective database. Single centre.
Vainshelboim et al. 2016, Israel	Prospective, observational study evaluating role of 12 week exercise training program on survival at 40 months follow up. Evaluation of the role of CPET variables in the prognostication of IPF.	34 pts with IPF	Cycle ergometer, pulse oximetry. Peak VO2/kg, peak work rate, VE/VO2 nadir , VE/VCO2 ratio at AT, tidal volume reserve.	Non-IPF ILD. Clinically unstable in preceding 3-6 months, severe co-morbid illness, unstable cardiac disease and any orthopaedic or neurological contraindications to CPET.	Mortality or transplantation	ROC curve analysis was used to determine cut off points of CPET variables for mortality. Cox regression analysis for survival analysis and comparison between significant cut-off points (log rank test). HR for death or LTx (Wald test).	9 deaths and 2 LTx (considered fatalities in statistical analysis).Poorer survival and significant increased risk of mortality associated with cut off points for: peak work rate<62 watts (AUC 0.854, 0.73-0.98 CI, p=0.005), peak VO2 <13.8mL/kg/min (AUC 0.731, 0.56-0.9, p=0.031), tidal volume reserve <0.48 L/breath (AUC 0.810, 0.660.96, p=0.01), VE/VCO2 at AT >34 (AUC 0.783, 0.60.96, p=0.02) & VE/VO2 nadir >34 (AUC 0.736, 0.560.9, p=0.002). Bivariate analysis of these cut offs (above and below the threshold) revealed HRs as follows: Peak work rate 9.2 (1.9-42.6), Peak VO2 4.4 (0.94-20.3), Tidal volume reserve 7.6 (1.6-35.2), VE/VO2 nadir 8.3 (2.231.6), VE/VO2 at AT 4.6(1.2-17.3). Non survivors were characterised by higher dyspnoea levels, the presence of PH (assessed by ECHO sPAP>35mmHg), and CPET markers of reduced ventilatory efficiency (VE/VO2 nadir p=0.039, VE/VCO2 at AT p=0.008) and reduced exercise capacity (Peak work rate p=0.01, Peak VO2 p=0.02). Exercise training intervention had no survival benefit over standard care. Higher prevalence of PH in non survivors.	Prospective observational study analysis as part of a wider single centre RCT. Underpowered to detect survival differences between groups. Small sample size.

	Retrospective							
Layton et al. 2017, USA	study evaluating predictive value of CPET for one-year transplant free survival in a population of ILD patients undergoing lung transplant evaluation.	192 pts had CPET was performed on oxygen. Four tests terminated due to oxygen desaturation (nadir SpO2 < 80% despite 30% FiO2). 3 tests terminated early due to low ETCO2 (<18mmHg) or elevated ETCO2 (>60mmHg), reducing cohort to 185 pts.	Cycle ergometer, pulse oximetry. Peak VO2 (ml/kg/min, % predicted), workload (watts, % predicted), VE/VCO2 slope (% predicted), ETCO2 mmHg & O2 pulse.	Pts not being evaluated for lung transplant, those that did not require oxygen with exercise, no follow up data available at 1 year post CPET.	Survival without the need for transplantation (at one year).	79 deaths/transplantations during follow up period. Comparison of variables between those who died /transplanted (D/LTx) and those who survived transplant free were compared using 2-sample independent t test. Survival was calculated by Kaplan-Meier method, with univariable Cox regression analysis to identify predictors of 1yr transplant free survival. Multivariable cox model with forward stepwise elimination method to identify prediction of transplant free survival (and to predict survival excluding those transplanted. ROC used to test thresholds of these predictors.	Mixed cohort of ILD patients analysed: IPF n=135 (70%), sarcoidosis n=15 (8%), HP n=6 (3%), NSIP n=12 (6%), ILD with mixed connective tissue disorder n=24 (13%). 113/192 (59%) survived transplant free. More patients with sarcoidosis in the survival transplant free group then the D/LTx group and more patients with NSIP in the D/LTx group (p=0.028). Multivariable cox regression identified CPET variables of peak workload <35% predicted (HR 4.71, 2.64-8.38 CI and AUC =0.740) and nadir CPET SpO2 <86% despite 30% FiO2 (HR 2.27, 1.41-3.68 CI, AUC=0.645) as discriminatory parameters predicting one-year mortality or need for transplant, alongside FVC% predicted <45% (HR 1.82, 1.15-2.87 CI AUC 0.624). Notably the presence of PH (present in 50% pts determined by combination of RHC or ECHO) was not an independent predictor of prognosis in this study.	Retrospective, single centre cohort. Potential for selection bias, unidentified confounding and missing co-variate data. Generalisability to general ILD patients questionable as highly selected cohort of advanced ILD patients. Source population, patterns of referral transplant, waiting times and cohort characteristics may differ from other transplant programs.

Table 1: Study characteristics of papers selected for full data extraction.

Abbreviations: Pts, patients; IPF, idiopathic pulmonary fibrosis; ILD, interstitial lung disease; NSIP, non-specific interstitial pneumonia; HP, hypersensitivity pneumonitis; CTD, connective tissue disease; SSc, systemic sclerosis; COP, cryptogenic organising pneumonia; UIP, usual interstitial pneumonia; DPLD, diffuse parenchymal lung disease; LAM, lymphangioleiomyomatosis; EP, eosinophilic pneumonia; 6MWTD, 6-minute walk test distance; CPET, cardiopulmonary exercise testing; AaDO2, alveolar-arterial oxygen pressure difference; FiO2, fraction of inspired oxygen; VD/VT, physiological dead space/tidal volume ratio; VO2, oxygen uptake, VCO2, carbon dioxide production; PaO2, partial pressure of oxygen; PaCO2, partial pressure of carbon dioxide; VT, ventilatory threshold (highest VO2 sustained without lactic acidosis); AT, anaerobic threshold; Vt, tidal volume; tidal volume reserve, Vt max-Vt resting; IC, inspiratory capacity; VE, minute ventilation; breathing reserve, VE/MVV during exercise; VE/VO2, peak oxygen uptake; VE/VCO2, ventilatory efficiency; kg, kilograms; OR, odds ratio; HR, hazards ratio; HRR, heart rate; SaO2, oxygen saturation of arterial blood; SpO2, oxygen saturations measured by pulse oximetry; max, maximal; MVV, maximum voluntary ventilation (can be measured or estimated as FEV1 X 41); Δ, change in; VO2 slope, PaO2 plotted against VO2; VO2/HRR max or oxygen pulse, oxygen delivery per heartbeat; ETCO2, end tidal carbon dioxide; D, died/deaths; LTx, lung transplantation, sPAP, systolic pulmonary artery pressure, ROC, receiver operating characteristic curve; FVC, forced vital capacity; CPI, composite physiologic index; TLC, total lung capacity; DLCO, diffusion capacity of lungs for carbon dioxide; RHC; right heart catheter; ECHO, echocardiogram; PH, pulmonary hypertension; PWP, pulmonary capillary wedge pressure; AUC, area under the curve; RCT, randomised controlled trial; CI, confidence interval; pred, predicted; %, percentage; P(A-a)O2, alveolar-arterial oxygen pressure gradient at peak exercise; ΔSp02, difference between peak and resting oxygen saturation, CXR, chest X-ray; PVD; peripheral vascular disease, HRCT, high resolution computed tomography; RR, respiratory rate.

Appendix D: Published literature review- ERJ Open Research: August 2020

The prognostic value of cardiopulmonary exercise testing in interstitial lung disease: A systematic literature review

Shaney L. Barratt [1,2], Richard Davis[1], Charles Sharp[3], John D. Pauling [4,5]

[1]Academic Respiratory Unit, School of Clinical Sciences, University of Bristol, Bristol, UK [2]Bristol Interstitial Lung Disease Service, North Bristol NHS Trust, Southmead, Bristol, UK

[3]Gloucestershire Hospitals NHS Foundation Trust, Gloucestershire, UK.

[4]Department of Pharmacy and Pharmacology, University of Bath, Bath, UK

[5]Royal National Hospital for Rheumatic Diseases, Royal United Hospitals NHS Foundation Trust, Bath, UK.

Abstract

Background

Interstitial lung disease (ILD) heterogeneity poses challenges in terms of prognostication; including end of life discussions and optimal timing for transplantation. Efforts are required to develop definitive prediction models for use in clinical practice. Cardio-Pulmonary Exercise Testing (CPET) provides a comprehensive assessment of the physiological changes in the respiratory, cardiovascular, and musculoskeletal systems in a controlled laboratory environment, that has shown promise in terms of prognostic value in a number of chronic respiratory conditions.

Methods

We conducted a systematic review to identify CPET variables that predicted longitudinal outcomes in ILD. Two databases were searched to identify all studies reporting prognostic value of CPET in predicting disease-specific outcomes in longitudinal studies of ILD. Two authors independently reviewed and extracted data from acceptable studies.

Results

The initial search identified 658 unique citations. Thirteen studies were identified that examined the prognostic value of CPET in ILD, all of which reported a prognostic role for CPET parameters in predicting clinical outcomes in ILD, with survival being the principle clinical outcome measured. Issues with study quality (relating primarily to the inherent problems of retrospective studies, patient selection and presentation of numerous CPET parameters), limits the strength of conclusions that can be drawn from the studies reviewed.

Conclusions

There is insufficient evidence to support its use in facilitating 'real-world' clinical decisions. Additional prospective studies are required to validate these preliminary findings.

Introduction

The Interstitial Lung Diseases (ILD) are a group of heterogeneous diseases characterised by alveolar and interstitial damage, varying degrees of inflammation and/or fibrosis, architectural distortion and impaired gas exchange. ILD may be attributed to a known cause (e.g. drugs, connective tissue disease or inhalation of dusts or organic antigens) or unknown cause, such as Idiopathic pulmonary fibrosis (IPF) (2000). The prognosis is often poor, particularly for IPF which is typically progressive with a median survival of 2-5 years from diagnosis (Ley et al. 2011). ILD is also the leading cause of disease related mortality in connective tissue diseases (CTD) such as systemic sclerosis (SSc) (Steen and Medsger 2007) and myositis (Johnson et al. 2016). There is however vast heterogeneity in terms of presenting features, severity, disease course, treatment response and individual survival (Bellaye and Kolb 2015). This leads to challenges for patients and clinicians in terms of end of life discussions (Schroedl et al. 2014), treatment choices, optimal timing for transplantation (Mura et al. 2012) and conduct of clinical trials (Albera 2011, Gordon and Domsic 2016).

Previous studies of ILD have identified declining functional capacity and muscle weakness as strongly predictive of disease progression and increased mortality(Panagiotou et al. 2016), whilst measures of gas exchange may be more valuable predictors of outcome than measures of lung mechanics (Lederer et al. 2006, Flaherty et al. 2006, Ley et al. 2011). Hypoxia after 6-minute walk test (6MWT) and a history of arthritis appear to predict progression in SSc-ILD (Wu et al. 2018). Nevertheless, further efforts to develop definitive prediction models are required for clinical practice (Kolb and Collard 2014, Ley et al. 2011).

Cardio-Pulmonary Exercise Testing (CPET) provides a comprehensive assessment of the physiological changes in the respiratory, cardiovascular, and musculoskeletal systems in a controlled laboratory environment (Layton et al. 2017), that has shown promise in terms of prognostic value in a number of chronic respiratory conditions (Ferrazza et al. 2009, Arena and Sietsema 2011).

The primary objective of this systematic literature review was to evaluate the prognostic value of CPET in predicting disease-specific outcomes in longitudinal studies of ILD. If a prognostic role for CPET were confirmed, it could be used to guide earlier intervention for at-risk patients, support cohort enrichment for ILD clinical trials and allay anxiety and unnecessary monitoring amongst patients with stable ILD.

Materials and methods

Reporting of protocol and review registration

The study protocol was prepared in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Shamseer et al. 2015) and registered in the International Prospective Register of Systematic Reviews (PROSPERO 110198/2018).

Eligibility criteria

Studies that reported the relationship between CPET assessment and disease progression, prognosis or the presence/emergence of specific clinical outcomes of ILD were included.

Using the PICO framework, we evaluated publications that fulfilled the following criteria:

Population

Adults (18 years or older) with a diagnosis of ILD (including but not limited to idiopathic pulmonary fibrosis, CTD-related ILD and sarcoid-related ILD).

Intervention

Studies reporting the outcome of CPET assessment as a prognostic factor. All available methods of 1) performing formal CPET and 2) reporting CPET results were included.

Comparison

Patients with/who developed relevant outcomes (see below) were compared with those who did not, using CPET testing at baseline in both groups.

Outcome measures

The primary objective was to evaluate the prognostic value of CPET in predicting disease course and outcomes in longitudinal (retrospective or prospective) studies of ILD. The relationship between CPET results and a number of clinically relevant outcomes including, but not limited to, relevant clinical phenotype and disease demographics (e.g. disease duration, gender, age, lung physiology), disease outcomes (e.g. death, hospitalisation), surrogates of disease severity (including, but not limited to lung physiology, circulating biomarkers etc.), health-related quality of life (HRQOL) and functional status, were examined.

Study design

Eligible studies included cohort (retrospective or prospective) and observational longitudinal studies, that reported outcomes at a time point distinct from the baseline CPET (i.e. were of an appropriate design to evaluate prognostic value). The following types of studies were excluded: 1) animal studies 2) studies including patients with lung disease where an ILD cohort was not described and reported separately 3) studies designed to develop or validate health measurement scales 4) randomized controlled trials 5) case reports 6) qualitative research 7) non-original research publications (i.e., editorials, reviews) 8) abbreviated reports (e.g. letters to editors) and conference proceedings. An amendment to our originally registered protocol (English language articles only) was made to enable the inclusion of a relevant non-English (French) publication.

Search strategy

The search criteria were developed in accordance with search recommendations for systematic reviews of evaluations of prognostic variables (Altman et al. 2000). Electronic searches were performed in Medline and EMBASE, with no publication date or language restrictions. Full details of the specific search criteria applied are detailed in the supplementary material (Supplementary material 1). All titles and abstracts generated by the search criteria were screened independently by two review authors (R.D. and C.S.) identifying those studies relevant and eligible for full text review. Agreement between reviewers in the study selection process was assessed using Cohen's Kappa statistics (Cohen 1968). Any discrepancies/disagreements were resolved by discussion between reviewers and included a third party (SLB) if necessary. Discussions between reviewers resolved any discrepancies at each stage of the study selection process. Review articles or editorials focussing on the "prognostic aspects of cardiopulmonary exercise testing in Interstitial Lung Disease" were also reviewed, to facilitate a grey search of cited manuscripts within these reviews.

Data extraction

A standardised form was used (initially by RS and CS, with subsequent verification by SB) to independently extract relevant study details from each of the selected studies that included: date of publication, journal or publication source, study design, initial population of the study, study inclusion criteria, study exclusion criteria, CPET method, CPET analysis endpoints, disease outcomes assessed and a summary of key findings. Study corresponding authors were contacted when clarification was required.

Data synthesis

Formal meta-analysis was planned if appropriate and feasible. In anticipation of wide heterogeneity in design, CPET methods and CPET analysis, a narrative, qualitative synthesis of results was planned if quantitative analysis was not possible.

Risk of bias assessment

The QUIPS (Quality in Prognosis Study) risk of bias tool was used to assess the risk of bias within every included study (Huguet et al. 2013). Kappa statistics were applied to assess preliminary agreement between reviewers for bias assessment. Discussions between reviewers were undertaken to achieve consensus where discordance existed in the preliminary bias assessment for any domain (Supplementary material 2).

Results Study selection

Initial simultaneous searches in EMBASE (n=573) and Medline (n=373), performed on 13th April 2019, identified a total of 946 articles. After removal of duplicates (n=288), 658 articles generated by the search were screened for eligibility and exclusion criteria based on titles and abstract review. There was moderate initial agreement between the two reviewers (Cohen's kappa 0.462 – see Supplementary material 3), with discordance in 20 abstracts, that was easily resolved through discussion. Eighteen articles proceeded to full text review and this led exclusion of a further 5 studies. A total of 13 studies were deemed eligible for inclusion. The full study selection process is detailed in Figure 1. Table 1 summarises the study design and reported findings of the thirteen papers that proceeded to full data extraction.

Geographical participation and date of publication

Three studies were undertaken in Europe (5/13, 38%)(Triantafillidou et al. 2013, van der Plas et al. 2014, Gläser et al. 2013, Wallaert et al. 2011, Kollert et al. 2011), four in USA (5/13, 38%)(Fell et al. 2009, Kawut et al. 2005, Layton et al. 2017, Swigris et al. 2009, King et al. 2001) and the remainder in Israel (Vainshelboim et al. 2016), Japan (Miki et al. 2003) and Brazil (Lopes et al. 2012). The majority of studies were published in the last 10 years (10/13, 70%)(Layton et al. 2017, Triantafillidou et al. 2013, van der Plas et al. 2014, Gläser et al. 2013, Vainshelboim et al. 2016, Lopes et al. 2012, Wallaert et al. 2011, Swigris et al. 2009, King et al. 2001, Kollert et al. 2011) and only three studies published in the years preceding 2009 (Miki et al. 2003, Fell et al. 2009, Kawut et al. 2005).

Study characteristics

Most studies were retrospective cohort analyses (11/13, 85%), with variable follow-up periods (range 23 days(Kawut et al. 2005) - 20 years(Swigris et al. 2009)). The majority of retrospective studies evaluated independent risk factors for survival or mortality outcomes in ILD (9/11, 82%) and had an average follow up time of between 1-4 years (Gläser et al. 2013, Kawut et al. 2005, Layton et al. 2017, van der Plas et al. 2014, Miki et al. 2003, Vainshelboim et al. 2016, Triantafillidou et al. 2013, Wallaert et al. 2011, King et al. 2001). The longest planned follow up was in a study examining a cohort of systemic sclerosis ILD patients (truncated at 20 years) (Swigris et al. 2009).

There were two prospective studies (Triantafillidou et al. 2013, Vainshelboim et al. 2016). One investigating the relationship between CPET and survival characteristics in IPF had a variable duration of follow up between 9-64 months (Triantafillidou et al. 2013). The other prospective study used CPET as part of a wider investigation into the role of exercise testing in the prognostication of ILD and followed patients up for a fixed period of 40 months (Vainshelboim et al. 2016).

Study populations

Of the studies identified, 8/13 (62%) exclusively recruited patients with IPF, two recruited only sarcoidosis patients (Lopes et al. 2012, Kollert et al. 2011), and one study systemic-sclerosis associated ILD (Swigris et al. 2009). The remaining studies (2/13, 15%) evaluated more heterogeneous cohorts of ILD patients referred for lung transplantation assessment (Kawut et al. 2005, Layton et al. 2017).

The prognostic value of CPET has been retrospectively reported in a total of 703 patients with IPF, and prospectively in a further 59 patients in 2 small, single centre studies (n=25 (Triantafillidou et al. 2013) and n=34 (Vainshelboim et al. 2016) respectively). Patients were recruited to studies according to consensus statements on the diagnosis of IPF available at the time of enrolment; the 2000 American

Thoracic Society (ATS) international consensus statement for the diagnosis of IPF(2000, Miki et al. 2003, Triantafillidou et al. 2013, van der Plas et al. 2014, Fell et al. 2009, King et al. 2001) and the later 2002 ATS/ERS (European Respiratory Society) international consensus classification of the idiopathic interstitial pneumonias (including IPF) (Wallaert et al. 2011, Kawut et al. 2005, Society and Society 2002). The updated 2011 ATS/ERS/JRS/ALAT evidence based guidelines for the diagnosis of IPF (Raghu et al. 2011) were applied in all (Triantafillidou et al. 2013, Gläser et al. 2013, Layton et al. 2017, Vainshelboim et al. 2016) but one of the studies (van der Plas et al. 2014) published after 2011 (the latter was a retrospective study that may have recruited patients prior to the publication of the 2011 guidelines).

The prognostic role of CPET in outcomes of secondary causes of ILD (such as connective tissue disease (CTD), drug-induced ILD, occupational causes of ILD and hypersensitivity pneumonitis (HP)) in granulomatous disease or in other forms of idiopathic interstitial pneumonias (IIP) has not been extensively reported. No prospective studies were identified. Only one retrospective study was identified that examined the prognostic role of measures obtained during CPET in a cohort of SSc associated ILD patients (n=83) (Swigris et al. 2009). Patients with SSc met diagnostic criteria adopted by the 1980 American Rheumatology Association and those with SSc sine scleroderma met criteria proposed by Poormoghim and colleagues (Poormoghim et al. 2000). A diagnosis of ILD was based on chest radiography in n=60 patients (Swigris et al. 2009).

Two additional retrospective studies have explored the role of CPET in predicting longitudinal outcomes in a total of 144 histologically confirmed sarcoidosis patients (Lopes et al. 2012, Kollert et al. 2011), representing Scadding disease stages 1-4 (SCADDING 1961).

We identified two retrospective studies that examined the role of CPET in predicting outcomes in mixed populations of ILD patients (Layton et al. 2017, Kawut et al. 2005). Cumulative patient numbers were small (a heterogeneous group of connective tissue disorders n=28, HP n=8, unclassifiable ILD n=7, sarcoid n=15, IIP n=21 (NSIP n=18, COP, DIP, COP). Whilst the cohorts could be considered to be representative of mixed ILD cohorts, patient numbers for each subtype were too small to consider each subgroup separately.

With regards to the study participant populations, the QUIPS risk of bias was considered to be low for only 3/13 (23%) studies (Triantafillidou et al. 2013, Wallaert et al. 2011, Vainshelboim et al. 2016), with the majority regarded as having a moderate (6/13, 46%) or high (4/13, 31%) (Miki et al. 2003, Layton et al. 2017, Kawut et al. 2005, Kollert et al. 2011) risk of bias. The generalisability of one study was potentially limited by the reported high diagnostic lung biopsy rate for IPF patients (64% (75/117)

(Fell et al. 2009), a condition that can often be confidently diagnosed without biopsy in the presence of typical radiological findings and by consensus agreement in the multidisciplinary team setting (Walsh et al. 2016) and thus raising concerns as to whether this cohort was representative of IPF populations in the 'real world'. The generalisability of a further study that assessed the extent to which gas exchange measurements could predict the need for prolonged immunosuppressive therapy in sarcoidosis, was limited by the lack of clearly defined clinical characteristics e.g Scadding disease stage, in the subset of patients followed longitudinally (102/149)(Kollert et al. 2011). Two studies examined disease outcomes that necessitated a particular baseline clinical phenotype e.g. recruitment from source populations referred for lung transplant evaluation and thus by definition only analysed selected cohorts of advanced ILD patients (Layton et al. 2017, Kawut et al. 2005). Others incorporated a *priori* patient grouping, for example the presence of pulmonary hypertension (Gläser et al. 2013), to enrich populations with patients at high risk of developing outcomes of interest, or required the active exclusion of patients with a relevant phenotype e.g. those that died from a cause other than respiratory failure (Miki et al. 2003).

Study attrition was generally reported to be low, which may reflect the retrospective nature of the majority of the studies identified. The QUIPS risk of bias for study attrition was reported to be high in two studies, increasing the potential for selection bias; >25% patients identified were excluded from the analysis by Lopes et al.(Lopes et al. 2012) (15 pts excluded: smoking history (n=10), concomitant respiratory disease (n=2), cardiac disease (n=2), neuromuscular disease (n=1), reducing final cohort to 42 pts), whilst in the study by King et al. (King et al. 2001), 34% (80/238) of the originally identified population were excluded from inclusion in the final analysis because of incomplete data sets.

Prognostic factor measurement

CPET was the sole prognostic factor for the majority of studies 8/13 (62%), with a minority using CPET as part of a broader repertoire of exploratory physiological tests including 6MWT (Kawut et al. 2005, Triantafillidou et al. 2013, Layton et al. 2017) or lung function parameters (Gläser et al. 2013). One study used CPET in conjunction with clinical, radiological and resting physiological tests to devise a scoring system to predict survival in newly diagnosed cases of IPF (the CRP score: Clinical Radiological Physiological score) (King et al. 2001).

In two studies, CPET was used as the principle method to achieve a standardised form of maximal exercise (Kollert et al. 2011, Swigris et al. 2009) where upon arterial blood gas sampling or peripheral

oxygenation measurements were taken to determine the effect of exercise on gas exchange. In both of these studies, typical CPET measures, such as maximal oxygen consumption (VO_{2max}) were not recorded.

Across all studies, the bias rating for prognostic factor measurement using the QUIPS tool was considered low-to-moderate (Table 2), with the majority of studies reporting a standardised approach to CPET and analysis that would be easily reproducible and less amenable to bias. Most studies provided a sufficient description of the CPET protocol used, adhering to the 2003 American Thoracic Society statement on cardiopulmonary exercise testing (Society and Physicians 2003) (6/10, 60%)(Kawut et al. 2005, Wallaert et al. 2011, van der Plas et al. 2014, Layton et al. 2017, Triantafillidou et al. 2013, Gläser et al. 2013). Others used the European Respiratory Society 1997 (Miki et al. 2003) and updated 2007 (Vainshelboim et al. 2016, Palange et al. 2007) recommendations. In others important details were missing e.g. if oxygenation was measured during CPET (van der Plas et al. 2014). Variation in the methodological approach to CPET was also observed. For example, in one study, oxygen usage during CPET was an inclusion criteria (Layton et al. 2017), whilst in another, supplemental oxygen during exercise was supplied variably to participants depending on a pre-study requirement for home oxygen or saturation on room air <90% (Kawut et al. 2005). In 7/13 (54%) studies, blood gas analysis was used to assess the adequacy of gas exchange during exercise (Fell et al. 2009, Miki et al. 2003, Lopes et al. 2012, Wallaert et al. 2011, Kollert et al. 2011, Swigris et al. 2009, King et al. 2001), whilst the remainder used pulse oximetry, considered by some experts to be a suboptimal substitute (Society and Physicians 2003). A broad range of quantitative CPET parameters were presented/analysed (summarised in Table 1), raising the possibility of reporting bias (see later).

All but one study used cycle ergometry. Treadmill exercise testing was used as the method of CPET in the remaining study; in which exercise increments were chosen for participants based on patient's daily activities and parameters of resting pulmonary function, raising concerns whether a standardised approach had been adopted (Miki et al. 2003). Additionally, non-uniform speed increases, often inherent to treadmill testing, results in nonlinear metabolic rate increases and fundamental difficulties in calculating an accurate external work rate and an estimation of peak VO2. Thus direct comparisons of peak VO2 obtained during treadmill testing studies cannot be compared with those obtained from cycle ergometry studies.

Outcome measurement

The most commonly reported outcome was mortality/survival 11/13 (85%). The majority of these studies that used survival/mortality as an outcome measurement (10/11, 91%) examined all-cause

mortality, considering death or lung transplantation as composite end-point. One study used an outcome measurement that was restricted to respiratory deaths only (Miki et al. 2003) and another study assessed the discriminatory ability of CPET to identify patients who would die on the lung transplant list before receiving transplantation (Kawut et al. 2005). Other outcomes included interceding pulmonary hypertension (PH) (Gläser et al. 2013) and decline in pulmonary function (FVC and DLCO) or duration of immunosuppressive therapy in sarcoidosis (Lopes et al. 2012, Kollert et al. 2011).

Using the QUIPS tool, the risk of bias in the approach to outcome measure assessment was considered low-to-moderate, in all studies.

Reported prognostic associations of CPET in ILD

All studies reported at least 1 positive association between CPET and clinical outcomes, raising the possibility of positive reporting bias (Table 1). Significant heterogeneity in study design, study populations (and classification criteria adopted), CPET protocols, CPET endpoints and defined endpoints precluded any useful attempt at meta-analysis.

Idiopathic pulmonary fibrosis

The prognostic role of peak VO_2 has been examined across several studies of IPF. Fell et al. (Fell et al. 2009) retrospectively suggested a baseline threshold of peak VO_2 8.3ml/kg/min predicted survival in 117 patients with IPF (peak VO_2 <8.3ml/kg/min HR 3.24, CI 1.10-9.56, p=0.03). Patient numbers in the subgroup with peak VO_2 < 8.3ml/kg/min were small however (n=8, 7%), compared to the 46% patients that actually died, suggesting that the threshold sensitivity was not high. In another study, Triantafillidou et al. (Triantafillidou et al. 2013) prospectively identified a threshold of 14.2ml/kg/min for survival in 25 patients with moderate IPF (mean FVC 77.5 ± 21.8), whilst Vainshelboim et al. (Vainshelboim et al. 2016) suggested VO_2 <13.8 ml/kg/min as one of 5 CPET parameter thresholds (peak work rate, tidal volume reserve, V_E/VO_2 nadir and V_E/VCO_2 at AT) predicting survival in a prospective cohort study of 34 patients with IPF. Finally, Gläser et al. identified that the presence of PH (invasively assessed by right heart catheter) and peak VO_2 % predicted were the only variables independently predictive of survival in a retrospective cohort of 133 patients, and application of % predicted values showed statically significant superiority to absolute data values. These results contrast with the findings of other studies where no independent association between survival in IPF and peak VO_2 has been demonstrated (Wallaert et al. 2011, Miki et al. 2003). Heterogeneity in terms

of disease severity, follow-up periods and accompanying disease co-morbidity may have impacted on results of these studies and larger prospective studies are required to ascertain the prognostic role of peak VO_2 in predicting IPF survival.

Gläser et al. found that the development of interceding PH in IPF was best predicted by reductions in ventilatory efficiency, the V_E/VCO_2 $slope_{pred}$ (cut off of ≥152.4, AUC 0.938; CI 0.892-0.984), with a sensitivity of 87.2% and specificity of 88.4%, but analysis of PH subgroup alone did not identify any CPET parameters that provided independent prognostic information. V_E/VCO_2 at AT has also been shown to be discriminating factor to determine the presence of PH across a cohort of IPF patients (adjusted OR 1.182; CI 1.029-1.384, p=0.021, n=81), but once again the prognostic value of this parameter has not been determined (Boutou et al. 2011).

The prognostic value of an alternative measures of ventilatory efficiency, the ventilatory equivalent for carbon dioxide at AT ($V_E/VCO_{2\ at}$ AT), in predicting survival in IPF has also been examined (van der Plas et al. 2014). In a retrospective study of 38 IPF patients, those with $V_E/VCO_{2\ at}$ AT >45 had a significantly worse survival compared to patients with $V_E/VCO_{2\ at}$ AT ≤ 45 (HR 4.58, p=0.001), and this parameter remained a strong predictor even after correcting for functional severity of ILD, highlighting its possible use in the early detection of vascular impairment. Furthermore, the ventilator equivalent for oxygen at AT (V_E/VO_2 at AT) >45 was reported to be an independent poor predictor of 3 yearsurvival in a cohort of 63 IPF patients (Wallaert et al. 2011), findings consistent with the univariate analysis of Miki et al. (Miki et al. 2003). Results suggest that the magnitude of hyperventilation at ventilatory threshold may be determining prognostic value, but further prospective studies are required to confirm the value of these parameters of ventilatory efficiency in the prognostication of IPF.

Exercise induce hypoxaemia was also considered as a potential prognostic factor in IPF. Miki et al (Miki et al. 2003) found that only two factors, age and PaO_2 slope (defined as change in arterial oxygen pressure in mmHg / change in VO_2 uptake during exercise ($\Delta PaO_2/\Delta VO_2$)), provided independent prognostic information in a cohort of 41 IPF patients (HR 1.096, CI 1.012-1.187, p=0.025 and HR 0.841, CI 0.731-0.967, p=0.015 respectively) and stratification of patients according to this slope (≤60mmHg/l/min or >-60mmHg/l/min) identified significant differences in median survival (1.6 years vs 4.5 years respectively). Measurement of this parameter does however, require invasive arterial blood gas analysis during exercise testing, that is unavailable in the many clinical exercise laboratories. In the study by King et al. (King et al. 2001), PaO_2 at the end of maximal exercise was the only CPET

derived parameter included in their comprehensive clinical-radiologic-physiologic scoring model to predict survival in IPF, and when weighted, accounted for as much as 10.5% of the maximum score in the complete model. Nevertheless, there were methodological limitations in this latter study; only 158/238 patients performed exercise testing and patients received supplemental oxygen when significant hypoxaemia ensued.

As a consequence of the utilisation of numerous different CPET parameters, CPET cut-off values, and timing of mortality evaluation, it was not possible to determine definitive thresholds for mortality or the development of pulmonary hypertension based on the analysed data.

Non-IPF Interstitial Lung Disease

In mixed populations of ILD patients with advanced disease and referred for lung transplantation (Layton et al. 2017, Kawut et al. 2005), oxygen saturations during CPET exercise were amongst the variables that were predictive of lung transplantation or death. Layton et al. (Layton et al. 2017) demonstrated a workload threshold during CPET <35%, nadir CPET SpO_2 <86% and FVC <45% were predictive of lung transplantation or death within 1 year of their CPET, with HR of 4.71 (CI 2.64-8.38, p<0.001) and HR 2.27 (1.41-3.68, p=0.001) and HR 1.82 (1.15-2.87, p=0.01), respectively. Kawut et al. (Kawut et al. 2005) identified more than 15 different variables, including at least 6 CPET variables (SaO_2 unloaded, peak and recovery, VO_2/kg peak, VO_2/Heart rate, carbon dioxide production, minute ventilation, and V_E/VCO_2) that predicted all-cause mortality in this patient population, with worse survival for patients with SaO_2 <95% during unloaded exercise (Kaplan Meier log rank test p=0.0025) or low six minute walk test distance (6MWTD) <350m (p=0.001). The authors went on to study factors that might predict death on the transplantation list and identified that desaturation during exercise or 6MWTD <350m, were again associated with poor outcomes; for a patient with SaO_2 <95% during unloaded exercise, there was a 75% chance of dying on the list.

Across the two studies that examined longitudinal outcomes in sarcoidosis, the alveolar-arterial oxygen pressure gradient during exercise P(A-a)O2, a measure of arterial desaturation during exercise, was independently associated with both the need for prolonged immunosuppressive therapy (>1 year) (odds ratio (OR) 1.098 (CI 1.039-1.160, p<0.001)(Kollert et al. 2011) and decline in pulmonary function at 5 years (>10% decline in FVC or DLCO from baseline)(with $P(A\text{-}a)O_2$ >22mmHg Relative Risk (RR) 70.0 (CI 3.03-161.3, p<0.001)(Lopes et al. 2012).

Only one study specifically examined the role of CPET in predicting survival in SSc-ILD (Swigris et al. 2009). Whilst typical CPET measurements e.g. VO_{2max} were recorded, attempts to correlate these with survival were not described. In similarity to the studies in sarcoidosis and IPF, diffusion limitation, measured in this study as the change in peripheral pulse oximetry during CPET exercise (SpO_2), correlated with survival. The risk of death was 2.4 times greater in SSc-ILD individuals whose SpO_2max fell <89% (Hazards Ratio (HR) 2.4, CI 1.2-4.9, p=0.02), and 2.4 times greater for subjects whose SpO_2 max fell >4% from baseline (HR 2.4, CI 1.1-5.0, p=0.02).

Further interpretation of the prognostic value of CPET parameters in ILD is limited by moderate-tohigh risk of bias across domains the QUIPS tool, with a 'high' risk of bias rating present in at least one QUIP domain in 9/10 (90%) studies. The main issues with study quality were related to confounding and statistical domains (see Table 2) and are discussed in more detail in following sections.

Study confounders, statistical analysis and reporting

The majority of studies were considered to be at 'high' risk of bias due to inadequate account of potential confounding factors or methods of statistical analysis/reporting (Table 2).

The data used in the majority of studies was obtained from existing databases and/or case note review (85%, n=11). As the data was not collected as part of a designed study, several potential confounders variables were not recorded, for example the presence of co-morbid disease (Wallaert et al. 2011, Fell et al. 2009, Miki et al. 2003, Gläser et al. 2013, Lopes et al. 2012, Swigris et al. 2009), body mass index (Triantafillidou et al. 2013, Fell et al. 2009, Miki et al. 2003, Kawut et al. 2005, van der Plas et al. 2014, Lopes et al. 2012) and smoking status (Wallaert et al. 2011, Kawut et al. 2005, Gläser et al. 2013, Lopes et al. 2012).

The most important potential confounder was baseline 'disease severity' which was only specifically addressed as a confounder in one study; through the inclusion of lung function parameters and a composite physiological index (as markers of disease severity) into the Cox regression model used for analysis (van der Plas et al. 2014). This same study also stratified patients in an attempt to control for other potential confounders. Patients were sub-grouped into those with a systolic pulmonary artery pressure greater than or less than 40mmHg, in an attempt to control for interceding pulmonary hypertension, but this reduced subgroup sample sizes and thus may have reduced the statistical power to detect an effect.

Eligibility criteria were used to increase uniformity of study participants and reduce potential confounders. For example two studies used participants referred for transplantation and thus by definition analysed distinct cohorts of more advanced patients but this selection bias reduced the generalisability of results (van der Plas et al. 2014, Layton et al. 2017). Other studies focused on healthier populations of ILD patients who did not need supplemental oxygen during CPET testing, but this, unsurprisingly, resulted in low mortality rates (n<10) leading to reporting bias (Vainshelboim et al. 2016, Fell et al. 2009, Triantafillidou et al. 2013).

Multiple regression analysis was the dominant statistical methodology used to determine the relationship between CPET parameters and clinical outcomes in ILD. Whilst this approach is generally considered to be one of the better statistical approaches to minimise unknown confounders, many of the studies reported on sample sizes much smaller than the minimum requirement for multiple logistic regression analysis as determined by Bujang et al. (Bujang et al. 2017). Furthermore, of all of the studies examined, only one detailed an *a priori* power calculation (Vainshelboim et al. 2016). Many studies were likely to be underpowered to detect the outcomes proposed.

Stepwise multiple regression was used by some studies to determine the optimal model parameters to predict increased mortality (Triantafillidou et al. 2013, King et al. 2001). One criticism of this statistical approach is that model selection is conducted through parameter inference, which may lead to over-fitting of some parameters or exclusion of confounders that are not statistically significant (Whittingham et al. 2006). Furthermore, the order of parameter entry (or deletion) and the number of parameters, can also affect the selected model (Derksen and Keselman 1992), whilst the multiple hypotheses tests, performed as part of this analysis, increases the probability of Type I error (Whittingham et al. 2006). The authors of one study did however attempt to overcome some of these limitations by checking for consistency between forward selection and backward elimination algorithms (Triantafillidou et al. 2013).

Only one study specifically attempted to reduce multicollinearity between parameters considered for inclusion into the multiple regression analysis (King et al. 2001). Multicollinearity is more common in observational studies and if ignored may lead to unreliable estimates of regression coefficients, inclusion of redundant variables and increased type II error (S, A and B 2000). King et al. (King et al. 2001) used Pearson's correlation coefficient to detect variables that were highly correlated; for example PaO2 at maximal exercise and resting PaO2, entering only the most statistically significant variables into the multivariable model.

Discussion

Maximum oxygen consumption (VO_{2max}) is a measurement of the capacity for aerobic exercise and is determined by variables that define oxygen delivery by the Fick equation (Society and Physicians 2003); thus gas exchange across the lung, oxygen content of blood, oxygen delivery to tissues and oxygen uptake in the tissues can all affect the VO_{2max}. In healthy individuals, constraints of the cardiovascular system are most responsible for limiting VO_{2max} (Wagner 1996, Stickland et al. 2012). In patients with ILD, limitation to exercise may generally occur as a consequence of one of more of: 1) ventilatory mechanical limitation (unable to increase tidal volume (V_T) sufficiently and may reach their maximal predicted minute ventilation (% pred V_{Emax})), 2) abnormal gas exchange (or reduction in ventilatory efficiency, indicated by variables such as the increment in minute ventilation (V_E) relative to carbon dioxide production (CO_2; V_E/VCO_2) 3) and/or diffusion limitation (indicated by variables such as reduction in oxygenation ≥ 4% or hypoxia at anaerobic threshold (AT)/peak exercise).

To our knowledge, this is the first study to systematically review and critically appraise studies that have reported the prognostic value of CPET in ILD. This field has gained recent attention with the majority of studies published within the last 8 years. Thirteen studies were identified that examined the prognostic value of CPET in ILD, all of which reported a prognostic role for CPET parameters in predicting clinical outcomes in ILD, with survival being the principle clinical outcome measured. Issues with study quality (relating primarily to the inherent problems of retrospective studies, patient selection and presentation of numerous CPET parameters), limits the strength of conclusions that can be drawn from the studies reviewed and thus whilst the associations presented shed important light to the potential role of CPET in disease prognostication in ILD, there is insufficient evidence at the moment to support its use in facilitating 'real-world' clinical decisions.

The exclusion of unpublished studies (e.g. conference abstracts) and abbreviated reports from this review may also increase the potential for publication bias, although this *priori* decision was taken to ensure sufficient information was available to enable detailed data extraction from each study.

We identified one study that described the prognostic value of CPET in IPF that was not originally eligible for inclusion in our study analysis due to the full text being published in French (Wallaert et al. 2011). The decision was taken to amend our published protocol to include this study as the subject of the study was deemed to be important by independent reviewers.

This work has identified a number of considerations for future prognostic studies of CPET in ILD. Common to many human diseases, the disease progression in ILD is likely influenced by a complex

interplay of patient, genetic, environmental and treatment factors. As such, a multivariable approach to the design and analysis of future prognostic studies of ILD is essential if we are to confirm a specific role for CPET in routine monitoring. In contrast to randomised controlled trials, there are no robust standards defining the need to register or publish protocols for prognostic research and as such it is not always transparent whether statistical analysis were part of a *priori* plan (Hemingway et al. 2009). Almost all studies in this review examined multiple prognostic CPET variables and as such there is potential for selective reporting bias that could be largely overcome by more stringent protocol registration with pre-specified outcomes of interest.

Conclusion

The quality of existing studies on the role of CPET in the prognostication of ILD limits the conclusions that can be drawn from such work. Larger prospective studies are needed to establish the role of CPET in the longitudinal assessment of ILD in the future.

Appendix E: Details of inter-reviewer agreement during initial title and abstract review for eligible articles

Details of agreement between reviewers for the title and abstract review

		Reviewer 2		
Reviewer 1		Include	Exclude	**Totals**
	Include	9	20	**29**
	Exclude	0	629	**629**
Totals		**9**	**649**	**658**

Number of observed agreements: 638 (97.0% of the observations)
Number of agreements expected by chance: 620.8 (94.35% of the observations)

Kappa = 0.462 (95% confidence interval 0.267-0.658)
SE of kappa= 0.100
The strength of agreement is considered to be 'moderate'.

Of the 20 citations for which there was disagreement, 9 papers were included in full text analysis. This resulted in a total of 18 papers proceeding to full text review.

Appendix F: Poster accepted for Winter British Thoracic Society 2019

The use of cardiopulmonary exercise testing in Idiopathic Pulmonary Fibrosis: Feasibility and correlation with quality of life measures.

Davis R[1], Viner J[2], Dixon C[2], Morley A[1], Adamali H[3], Maskell N[1], Barratt SL[1,3]

1. Academic Respiratory Unit, University of Bristol, Bristol, UK
2. Respiratory Physiology, North Bristol NHS Trust, Bristol, UK
3. Interstitial Lung Disease Service, North Bristol NHS Trust, Bristol, UK

Introduction:

The heterogeneity of idiopathic pulmonary fibrosis (IPF) in terms of disease course and treatment response leads to challenges for patients and clinicians in terms of optimal timing for transplantation(Mura et al. 2012) and/or end of life discussions (Schroedl et al. 2014). The use of cardiopulmonary exercise testing (CPET) in IPF prognostication remains largely unexplored.

Objectives:

1) To explore the feasibility of undertaking CPET in this population
2) To explore the correlation between baseline CPET variables, physiological variables and quality of life (QOL) scores.

Methods:

Consecutive IPF patients (n=74) were approached, with prospective recruitment of 42 participants to the study. Patients with FVC <50% and/or DLCO <50% were excluded. King's Brief ILD (K-BILD) questionnaire assessed QOL. Patients undertook incremental exercise testing to maximal exertion using a cycle ergometer, with contemporaneous physiological testing (FVC, DLCO).

Results

32 patients were excluded from the study (22 screening failures, 10 declined), with study attrition of an additional 10 patients (n=4 withdrew consent, n=1 death prior to testing, n=5 developed exclusions). Thirty-two patients (23 mild IPF with FVC>80%, 9 moderate IPF with FVC 50-80%), 26M:6F and median age (IQR) 75 years (71-79), underwent CPET. One patient failed to reach anaerobic threshold (AT) and was excluded from the analysis. Median (IQR) pulmonary and exercise results were: FVC 92% (75-102), DLCO 62% (54-69), minimum SpO_2 93%

(88-95), VO_2 peak/kg 21 (17.4-23.8) $mL.kg^{-1}.min^{-1}$ and V_E/VCO_2 27.2 (25.4-30.5). Median (IQR) QOL scores for each domain were: total K-BILD 64.4 (58.1-68.7), psychological 68.3 (56.9-80.9), breathlessness/activity (B/A) 50.2 (48-

62.7) and chest symptoms 85.2 (85.2-100).

VO_2 peak/kg correlated with chest (r=0.36, p=0.049) and B/A (r=0.43, p=0.016) domains of the K-BILD questionnaire. VO_2 peak/kg at AT also correlated with total K-BILD scores r=0.37, p=0.039 and chest domains (r=0.535, p=0.002). Total KBILD scores did not correlate with %FVC (r=0.26, p=0.15), %DLCO predicted (r=0.11, p=0.544) or SpO_2 (r=0.01, p=0.959) (Spearman's).

Conclusions

Initial results suggest CPET is a feasible method of testing in mild-moderate IPF. Whilst QOL did not correlate with baseline FVC and DLCO, the relationship between oxygen consumption and QOL measures, requires further

exploration. Longitudinal data will hopefully provide further information on the usefulness of CPET as a prognostic marker.

Appendix G: Poster accepted for American Thoracic Society 2020

The use of cardiopulmonary exercise testing in Idiopathic Pulmonary Fibrosis: correlation with baseline quality of life measures.

Davis R[1], Viner J[2], Dixon C[2], Morley A[1], Adamali H[3], Maskell N[1], Barratt SL[1,3]

1. Academic Respiratory Unit, University of Bristol, Bristol, UK

2. Respiratory Physiology, North Bristol NHS Trust, Bristol, UK
3. Interstitial Lung Disease Service, North Bristol NHS Trust, Bristol, UK

Rationale:

The heterogeneity of idiopathic pulmonary fibrosis (IPF) in terms of disease course and treatment response leads to challenges for patients and clinicians in terms of optimal timing for transplantation and/or end of life discussions. The use of cardiopulmonary exercise testing (CPET) in IPF prognostication remains largely unexplored. We report on the preliminary baseline data obtained from a study evaluating the role of CPET as a prognostic tool in patients with IPF.

Objectives:

To explore the correlation between baseline CPET variables, physiological variables and quality of life (QOL) scores.

Methods:

Consecutive IPF patients (n=74) were approached, with prospective recruitment of 42 participants to the study. Patients with FVC <50% and/or DLCO <50% were excluded. King's Brief ILD (K-BILD) and IPF-Patient Reported Outcome Measure (IPF-PROM) questionnaires assessed QOL. Patients undertook incremental exercise testing to maximal exhaustion using a cycle ergometer, with contemporaneous physiological testing (FVC, DLCO).

Results

32 patients were excluded from the study (22 screening failures, 10 declined), with study attrition of an additional 10 patients (n=4 withdrew consent, n=1 death prior to testing, n=5 developed exclusions). Thirty-two patients (23 mild IPF with FVC>80%, 9 moderate IPF with FVC 50-80%), 26M:6F and median age (IQR) 75 years (71-79), underwent CPET. One patient failed to reach anaerobic threshold (AT) and was excluded from the analysis. Median (IQR) pulmonary and exercise results were: FVC 92% (75-102), DLCO 62% (54-69), minimum SpO_2 93% (88-95), VO_2 peak/kg 21 (17.4-23.8) $mL.kg^{-1}.min^{-1}$ and V_E/VCO_2 27.2 (25.4-30.5).

VO_2 peak/kg correlated with chest (r=0.36, p=0.049) and B/A (r=0.43, p=0.016) domains of the K-BILD questionnaire, in addition to total (r = -0.46, p=0.009) and energy (r= -0.45, p=0.012) domains of the IPF-PROM. VO_2 peak/kg at AT also correlated with total K-BILD scores r=0.37, p=0.039 and chest domains (r=0.535, p=0.002), but only the energy domain (r= -0.38, p=0.036) of IPF-PROM. Total K-BILD and IPF-PROM scores did not correlate with baseline FVC % predicted, DLCO % predicted or minimum SpO_2 (Spearman's rank).

Conclusions

In this cohort of mild-moderate IPF patients, baseline physiological testing did not correlate with patient QOL measures. The relationship between oxygen consumption during CPET and QOL measures, requires further exploration. Longitudinal data will hopefully provide further information on the usefulness of CPET as a prognostic marker.

Appendix H: Revised paper submitted to BMC Pulmonary Medicine

Title: A role for cardiopulmonary exercise testing in detecting physiological changes underlying health status in Idiopathic pulmonary fibrosis: a feasibility study

Authors: Davis R[1], Dixon C[2], Millar AB[1,2], Maskell NA[2], Barratt SL[1,2]

Affiliations:

[1]Academic Respiratory Unit, School of Clinical Sciences, University of Bristol, Bristol, UK

[2]Bristol Interstitial Lung Disease Service, North Bristol NHS Trust, Bristol, UK

Abstract

Introduction

There is limited data available on the use of CPET as a predictive tool for disease outcomes in the setting of IPF. We investigated the feasibility of undertaking CPET and the relationship between CPET and quality of life measurements in a well-defined population of mild and moderate IPF patients.

Methods

A prospective, single-centre observational study.

Results

Thirty-two IPF patients (mild n=23, moderate n=9) participated in the study, n=13 mild patients attended for repeat CPET testing at 12 months. At baseline, total K-BILD scores and total IPF-PROM scores significantly correlated with 6MWT distance, but not with baseline FVC % predicted, TL_{CO} % predicted, baseline or minimum SpO_2. VO_2 peak/kg at AT positively correlated with total scores, breathlessness/activity and chest domains of the K-BILD questionnaire ($p<0.05$). VO_2 peak significantly correlated with total IPF PROM scores and wellbeing domains ($p<0.05$), with a trend towards statistical significance for total IPF-PROM and VO_2 peak/kg at anaerobic threshold ($p=0.06$).

There was a statistically significant reduction in FVC% predicted at 12 months follow up, although the mean absolute decline was <10% ($p<0.05$). During this period VO_2 peak significantly reduced (21.6 ml/kg/min±2.9 vs 19.1±2.8; $p=0.017$), with corresponding reductions in total K-BILD and breathlessness/activity domains that exceeded the MCID for responsiveness. Lower baseline VO_2 peak/kg at anaerobic threshold correlated with greater declines in total K-BILD scores ($r = -0.62$, 0.024) at 12 months. Whilst baseline FVC% predicted or TL_{CO} % predicted did not predict change in health status,

Conclusion

We have shown that it is feasible to undertake CPET in patients with mild to moderate IPF.

CPET measures of VO_2 peak correlated with both baseline and change in K-BILD measurements at 1 year, despite relatively stable standard lung function (declines of <10% in FVC), suggesting its potential sensitivity to detect physiological changes underlying health status.

Introduction

Idiopathic Pulmonary Fibrosis (IPF) is a progressive fibrosing lung disease of increasing prevalence(Navaratnam et al. 2011), associated with median survival of only 3-5 years from diagnosis (Raghu et al. 2006b, Ley et al. 2011). Disease heterogeneity continues to present challenges for clinicians with regards to prognostication and optimal timings for lung transplantation and/or advanced care planning (Mura et al. 2012, Schroedl et al. 2014). In the setting of large-scale clinical trials, a decline in forced vital capacity (FVC) has been used as a primary outcome measure(Richeldi et al. 2014, King et al. 2014) and as a surrogate for mortality, although this has not been universally endorsed(Wells et al. 2012, King et al. 2005).

Cardiopulmonary exercise testing (CPET) is the considered the gold standard for evaluating maximal/symptom-limited exercise tolerance, encompassing respiratory, cardiovascular and musculoskeletal assessments, in a controlled laboratory environment(Layton et al. 2017, Sue and Wasserman 1991, Palange et al. 2007).

However, there is limited data available on the use of CPET as a predictive tool for disease outcomes in the setting of IPF. A recent systematic review identified only two small-scale prospective studies that investigated the role of CPET in the prognostication in IPF (Triantafillidou et al. 2013, Vainshelboim et al. 2016) and concluded that there was insufficient evidence to support its use in facilitating 'real world' clinical decisions at the current time.

We have investigated the feasibility of undertaking CPET in a population of mild and moderate IPF patients in terms of the attrition of participants, information on safety data, and willingness to engage with the study protocol. Secondary end-points included: the change in CPET parameters over a 1 year period and the correlation between baseline CPET parameters and change in lung function, 6MWT and health status at 1 year.

We hypothesised that CPET would be feasible in population of mild to moderate IPF patients and more sensitive to change in patient's health status than 6MWT, FVC or TL_{CO}.

Methods

Study design

This was a prospective, single-centre observational study undertaken at a large secondary care institution in the UK, providing secondary and tertiary care to patients with Interstitial Lung disease (ILD) within the South-West of England. The study was approved by the Health Research Authority and Research Ethics Committees (IRAS 223450).

Study Subjects

Patients with a multidisciplinary team (MDT) diagnosis of IPF, based upon the American Thoracic Society/European Respiratory Society 2018 guidelines(Raghu et al. 2018), were prospectively recruited to the study between June 2018 and May 2019. Written informed consent was obtained from each patient.

Patients were divided into a 'mild' or 'moderate' category dependent on their baseline Forced Vital Capacity (FVC ≥ 50% - <80%: moderate; FVC ≥ 80% mild). Those patients in the 'mild' disease group would undertake both a baseline and repeat CPET at 12 months. It was decided by the study committee, due to the uncertainty of the ability of those with reduced lung function to perform a maximal exercise test, that those in the 'moderate' disease group would undertake only a baseline CPET test.

Inclusion and exclusion criteria

Inclusion criteria were an MDT consensus diagnosis of IPF, male or female aged ≥ 40yrs, TLCO ≥ 50% predicted and FVC ≥50% with written informed consent for study participation.

Key exclusion criteria were: FEV1/FVC ratio < the lower limit of normal, mobility issues preventing the participant to undertake cycle ergometry, history of myocardial infarction (MI) within 6 months or unstable angina within 1 month, uncontrolled arrhythmias causing symptoms or haemodynamic compromise, history of recent syncope (within last 6 months), acute thrombosis within previous 6 months, cognitive impairment/ inability to perform CPET, severe or untreated arterial hypertension (>200mmHg systolic at rest, >120mmHg diastolic) and patients using oxygen treatment.

Participant testing

Pulmonary function testing

Pulmonary function tests were performed in accordance with ATS/ERS guidelines(Graham et al. 2019), using the European Community of Coal and Steel (ECCS) reference equations(Quanjer et al. 1993). Forced expiratory volume during first second of expiration (FEV_1), forced vital capacity (FVC), and transfer factor for carbon monoxide (TLCO) were undertaken at baseline (within 4 weeks of CPET) and at 12 months (± 4 weeks) (nSpire HDpft, nSpire Health GmbH, Germany). The MRC score, age (years), height (meters), and body weight of the patients (kilograms) were also recorded.

6-minute walking test (6MWT)

A 6MWT was performed at baseline (and within 3 months of CPET) according to ATS guidelines(Laboratories 2002), using the Enright reference equation(Enright and Sherrill 1998). The following data were collected and analysed: distance achieved (metres), oxygen saturation at the initiation of the test, the minimum saturation level, percentage of theoretical distance achieved and at the end of the test.

Cardiopulmonary Exercise Testing (CPET)

CPET was performed using a standardized protocol in accordance with the American Thoracic Society/American College of Chest Physicians (ATS/ACCP) statement(Society and Physicians 2003), using Wasserman(Wasserman et al. 1994) and Jones(Jones et al. 1985) reference equations. All patients underwent a symptom-limited CPET to exhaustion or intolerability with an electromagnetically braked cycle ergometer (Ergoselect 100, ergoline GmbH, Germany) using a ramp protocol over 8-12 minutes. The protocol included 3 min of rest, 2 min of unloaded cycling (at 60 revolutions per min), followed by a progressively increasing work rate in a ramp fashion, and a

recovery period (patient dependent). The work rate increment for each ramped exercise test was selected depending on the patient's level of daily activity (either 5 or 10 watt/min ramp).

Subjects were asked to maintain a rate of 60 revolutions per minute throughout the exercise period. Several markers were used to determine if a maximal effort test was performed; a respiratory exchange ratio (RER; VCO_2/VO_2) ≥1.1, maximum heart rate (HR max) > 80% of maximum predicted HR (220-age), maximum minute ventilation during exercise >85% predicted based on MVV at rest (maximum voluntary ventilation) and a plateau in VO_2 with an increased workload. CPET could be discontinued at the discretion of the supervising attendant if clinically indicated.

Cardiopulmonary data were collected and analysed with nSpire Zan 600 USB system (nSpire Health GmbH, Germany).

The following parameters were recorded:

- Peak oxygen consumption (VO_2 peak, ml/kg/min),
- Oxygen consumption at anaerobic threshold
- Carbon dioxide production (VCO_2)
- Peak minute ventilation (VE peak)(marker of ventilatory function during exercise),
- VE/VCO_2 slope as derived from the above values (reflects changes in ventilatory drive)
- Peripheral capillary oxygen saturation SpO_2 (marker of hypoxaemia indicating possible ventilatory limitation to exercise)
- Peak power output (W)(marker of musculoskeletal function)
- Heart rate (HR) (marker of cardiac function during exercise), □Breathing reserve (BR)

Anaerobic threshold was determined noninvasively through the plot of VCO_2 versus VO_2 (V-slope method). Predicted minute ventilation was automatically calculated by the software as a function of maximal voluntary ventilation (MVV), where MVV = FEV1 L x 40. The BR was automatically calculated from the software as the difference between the maximum voluntary ventilation at rest and the peak ventilation. The % predicted VO_2 peak and % theoretical VO_2 peak at AT were not determined in this study as it was felt that populations deriving existing reference equations and normal standards were

not representative of the male predominant and elderly population, characteristic of IPF patients/populations.

Health status questionnaires and patient-reported outcomes

Patients were asked to complete the King's Brief ILD questionnaire (K-BILD)(Patel et al. 2013) and IPF-Patient reported outcome measure (IPF-PROM) (Russell A et al. 2018), in addition to the Leicester Cough visual analogue scale (VAS) for cough (Key et al. 2010) and Bristol VAS for breathlessness and fatigue (Yates et al. 2018b), at baseline and at 12 months.

The K-BILD is a self-completed health status questionnaire that comprises 15 items in three domains of psychological, breathlessness and activities and chest symptoms. The K-BILD scoring system implements logit transformation of raw item response scores to provide total score ranges of 0–100, where 100 represents best health status. The minimally important clinical difference (MCID) for the logit version of the K-BILD questionnaire is 5 for total K-BILD, 6 for Psychological, 7 for Breathlessness and activities and 11 for Chest symptoms (Sinha et al. 2019).

The IPF-PROM(Russell A et al. 2018) is a self-completed 12 item health status questionnaire that measures the physical and psychological experience of breathlessness; emotional well-being and fatigue. The questionnaire has been validated in terms of face and content validity. The scores range from 12-48, where 48 indicates worst health status.

Outcomes

We wished to study the feasibility of undertaking CPET in a population of mild and moderate IPF patients: the attrition of participants, information on safety data, and willingness to engage with the study protocol. Secondary end-points included: the change in CPET parameters over a 1 year period and the correlation between baseline CPET parameters and change in lung function, 6MWT and health status at 1 year.

Statistical analysis

Categorical variables are reported as absolute numbers and percentages. Normality of continuous data was initially verified using D'Agostino and Pearson normality test. Mean and standard deviations (SD) were used to describe parametric data; median and interquartile range (IQR, in brackets) for nonparametric data. Differences among two groups were verified by t-test with Welch's correction (continuous data), χ^2-tests (categorical data) and paired t-tests for comparison in variables from baseline to 12 months. Pearson's correlation was used to determine correlations between parametric variables. Data were analysed using GraphPad Prism version 8.0. A p value of <0.05 was considered statistically significant.

Results

Study population

Forty-two consecutive IPF patients were prospectively enrolled to the study. Four patients subsequently withdrew consent, 1 patient died and 5 patients developed exclusion criteria prior to commencement of the study. A further 5 patients did not complete the study and were lost to follow up (4 mild, 1 moderate). The final population studied thus consisted of 27 patients (mild n=19 and moderate n=8) (Figure 1). Patients were predominantly male (n=22, 82%) with a mean age of 75 years (±6.0 years) and were symptomatic at baseline with a median MRC breathlessness score of 2 (IQR 23). Approximately one third (33%, n=9) of patients (mild n=5, moderate n=4) received antifibrotics during the observational period. At completion of 1 year follow up, all patients remained alive.

Feasibility

There was excellent willingness to engage with the study protocol.

All patients achieved a RER > 1.1 and the vast majority of patients also achieved >80% of their maximal predicted heart rate (25/27, 93%) and /or had limited breathing reserve, providing corroboration that patients performed at maximal effort.

At baseline, all participants achieved the anaerobic threshold during testing and at 1 year follow up only one patient failed to achieve the anaerobic threshold.

Breathlessness and fatigue were the most commonly cited reasons for terminating CPET. Of patients completing the study (n=27; 19 mild and 9 moderate IPF), baseline CPET was terminated due to breathlessness in 37% (10/27), the majority of which had mild IPF (90%, 9/10). Leg/muscle fatigue was cited as a reason for terminating CPET in 63% patients (17/27), of which 59% (10/17) had mild disease. There were no significant differences in the reasons for terminating CPET between those that completed and did not complete follow-up.

At 1 year repeat CPET, 54% (7/13) described breathlessness as the reason for stopping and 38% (5/13) muscle fatigue. A dry mouth was cited as the main contributing reason for stopping in one patient.

One patient described dizziness related to his breathlessness during CPET but no other adverse events were recorded. There were no serious adverse events.

Baseline measurements between mild and moderate IPF groups

Baseline demographics between mild and moderate IPF groups were statistically comparable (Table 1). As per a priori subgroup definitions, participants in the moderate IPF group had a statistically lower baseline FVC % predicted compared to those in the mild IPF group (mild 99% predicted ±10.0, range 85-125% predicted) vs moderate 70% predicted ±5.1, range 62-75% predicted, $p<0.0001$). In keeping with these findings there was a trend towards a lower TLco in those within the moderate IPF group (mild 63% predicted ±9.5), range 50-83 % predicted vs moderate 57% predicted ±6.2, range 50-65% predicted, $p=0.055$). No significant difference in the 6MWT distance measured between mild and moderate groups was observed.

Patients with moderate disease had numerically lower total K-BILD, chest symptom and psychological domain scores compared to those with mild disease, although values were not statistically different

(total K-BILD mild disease 67 (±10.3) vs moderate disease 60 (±6.6), p=0.058). There were no significant differences in VAS scores of cough, breathlessness or fatigue between mild and moderate IPF groups or IPF-PROM measurements.

Whilst baseline CPET values were all within 'normal' published ranges (Society and Physicians 2003), peak minute ventilation (% predicted) was significantly higher for those with moderate IPF compared to those with mild disease (mild 71.0% (±13.9) vs moderate 82.9% (±12.4, p=0.045) (Table 1).

Of the baseline CPET parameters measured, VO_2 peak/kg at anaerobic threshold positively correlated with total scores (r=0.42, p=0.03), breathlessness/activity (r=0.47, p=0.014) and chest domains (r=0.44, p=0.02) of the K-BILD questionnaire (Pearson's correlation). Similarly, total IPF PROM scores and wellbeing domains significantly correlated with VO_2 peak (r=-0.43, p=0.02 and r=-0.44, p=0.02), with a trend towards statistical significance for total IPF-PROM and VO_2 peak/kg at anaerobic threshold (p=0.06). VE/VCO_2 at anaerobic threshold also correlated with total K-BILD score (r=0.39; p=0.001) at baseline, although there were no significant correlations with the individual domains of the questionnaire or IPF-PROM.

Total K-BILD scores (r=0.44, p=0.03) and total IPF-PROM scores (r=-0.43, p=0.03) both significantly correlated with 6MWT distance, but not with baseline FVC % predicted (Total K-BILD, p=0.14; Total IPF-PROM p=0.50), TL_{CO} % predicted (Total K-BILD p=0.16; Total IPF-PROM p=0.32), baseline or minimum SpO_2 (Total K-BILD p=0.25 and p=0.32, respectively, Total IPF PROM p=0.53 and p=0.55, respectively). There were no significant correlations between baseline CPET parameters and VAS scores (p>0.05).

Measurements at 1 year follow up

Total IPF cohort

At one year of follow up, the mean reduction in FVC and TL_{CO} % predicted for the whole IPF cohort (n=27) was -3.6% (±7.1, p=0.02) and -3.2% (±7.5, p=0.04) respectively (Supplementary Table 1). Whilst statistically significant, values were below those deemed clinically significant (Collard et al. 2003, Flaherty et al. 2006). There was no significant reduction in 6MWT distance achieved (mean reduction 4.8m±34.7, p=0.50; mean reduction in 6MWT distance as % theoretical distance 0.2% (±7.9, p=0.92). There was a statistically significant reduction in the breathlessness and activity domain scores of the

K-BILD questionnaire from baseline to 1 year, suggesting worse health status (-4.8±10.2, p=0.02), although this did not reach the published MCID for responsiveness for this domain (Sinha et al. 2019). There were no statistically significant differences in the VAS scores for cough, breathlessness and fatigue or in the IPF-PROM from baseline to one year (p>0.05).

Follow up of mild IPF group with repeat CPET

Thirteen patients from the mild IPF group returned for repeat CPET at 1 year; the coronavirus COVID19 pandemic prohibited the return of the remaining six patients at the one year follow-up time point as planned. All but one patient achieved anaerobic threshold.

Upon repeat CPET testing there were statistically significant declines in the VO_2 peak (21.6 ml/kg/min±2.9 vs 19.1±2.8; p=0.017), VO_2 peak at AT (14.2 ml/kg/min±3.2 vs 11.8±1.6, p=0.044) VE peak (75.3 L/min ±20.9 vs 66.1±21.6; p=0.007), peak work (106.9 W±26.3 vs 90.8±25.9; p=0.022) heart rate response (142.3 bpm±24.0 vs 133±22.3; p=0.040) and increased breathing reserve at anaerobic threshold (BRmax) (21.8 L/min (12.4-34.2) vs 33.8 (20.2-55.7); p=0.0002), compared to baseline values (Table 2).

There was a statistically significant reduction in FVC% predicted at 12 months, although the mean absolute decline was <10% (baseline FVC 98.8% predicted±8.5 vs follow up FVC 93.4% predicted±10.3, p=0.01). In these same patients, statistically significant reductions in breathlessness/activity (-7.2±10.8; p=0.033) and chest (-9.6±-15.0; p=0.040) domain scores of the K-BILD questionnaire were observed, with a trend towards statistical significance for reduction in the total B-ILD score (-5.6±-10.4; p=0.077) at follow up. Notably, the mean unit change of total K-BILD and breathlessness/activity domain scores exceeded the minimally clinically important difference previously reported (5 and 7 unit change respectively)(Sinha et al. 2019), with 5/13 patients achieving the MCID for total K-BILD score and 8/13 for the breathlessness/activity domains (Table 3).

There were no statistically significant differences in the VAS scores for cough, breathlessness or fatigue VAS score from baseline to one year (p>0.05). There was statistically significant worsening in the psychological experience of breathlessness as reported by the IPF-PROM (0.8±1.2, p=0.044), although the clinical significance of this small statistical change is not clear.

Reductions in K-BILD scores observed at 12 months were correlated with baseline CPET measurements. Lower baseline VO_2 peak/kg at anaerobic threshold correlated with greater declines in total K-BILD scores (r = -0.62, 0.024) and psychological domains of the K-BILD at follow up (r= -0.63, p=0.022). No other baseline CPET parameters significantly correlated with change in K-BILD score in this small cohort, including peak work rate. Furthermore, there was no significant correlation with the baseline FVC% predicted (p=0.70) or TL_{CO}% predicted (p=0.62) and change in K-BILD score (Pearson's correlation).

Discussion

CPET is considered the gold standard for evaluating exertional dyspnoea and exercise intolerance in patients with cardiorespiratory conditions (Molgat-Seon et al. 2020), yet currently lacks a defined role in the management of ILD (Raghu et al. 2011, Molgat-Seon et al. 2020). A recent systematic review by our group highlighted the insufficient available evidence to support the use of CPET in disease prognostication in ILD, emphasising that heterogeneity in terms of the ILD populations studied and the retrospective nature of the majority of published studies limited the conclusions that could be drawn (Barratt et al. 2020). Furthermore, the minimally clinically important differences for CPET parameters in ILD have not been established.

Our study has shown that CPET can be undertaken in both mild and moderate populations of IPF patients, without any significant adverse events, although study attrition was high and complicated by COVID-19 restrictions, such that only 64% patients completing the protocol.

Our prospective data suggests that baseline CPET VO_2 peak is associated with clinically meaningful patient-perceived reduction in health status at 1 year, in spite of comparatively stable lung function parameters (<10% decline in FVC and <15% decline in TLco). VO_2 peak is an integrated measure of respiratory, cardiovascular and neuromuscular function (Society and Physicians 2003). In a progressive disease such as IPF, the finding of reduced exercise performance at one year was not a surprising one. However, results suggest that this reduction was not as a consequence of ventilatory limitation. There was no change in the CPET ventilatory mode and the development of cardiac +/- pulmonary vascular dysfunction was not apparent. One possible explanation might be that patients became more deconditioned with reduced activity levels in response to their perceived worsening of breathlessness. El Naggar et al (El Naggar. 2017) have previously shown that VO_2 peak correlated with health status of IPF patients at baseline as determined by the St Georges questionnaire but longitudinal changes in

CPET parameters and associated health status were not explored. In our cohort, VO_2 peak during exercise correlated with patient reported outcome measures at baseline, it significantly declined at 12 months and also correlated with the change in patient reported health status at 12 months.

Existing literature conflicts as to whether VO_2 peak might predict disease outcomes in IPF; peak VO_2 thresholds ranging from <8.3 to <14.2ml/kg/min (Triantafillidou et al. 2013, Fell et al. 2009, Vainshelboim et al. 2016) have been reported to predict mortality in IPF, whilst others studies have failed to identify any significant association (Wallaert et al. 2011, Miki et al. 2003, van der Plas et al. 2014). Ongoing follow up of our prospective cohort will be used to further study the use of baseline CPET parameters to predict longer-term outcomes in these patients.

It is recognised that this study has limitations. Firstly, and perhaps most importantly, the study was conducted on a relatively small and homogenous sample of patients. This limits the overall generalisability of results, particularly in terms of feasibility of CPET across IPF phenotypes; for example those with exercise induced pulmonary hypertension versus those with relatively normal pulmonary vascular response to exercise, and the risk of Type II error may be relatively high.

The vast majority of patients had mild IPF (72%) with a median MRC score of 2; again limiting the generalisability of results. Whilst it would have been preferable to have a broader range of symptomatic patients, exercise in patients with high MRC scores would be very restricted leading to early completion of tests before the limit of pulmonary and cardiovascular systems had been reached (O'Donnell et al. 2009) and thus negatively influencing the results. A further limitation of the study was that almost a quarter of patients enrolled in the study developed exclusions to CPET or were lost to follow up; a factor that will be helpful to inform power calculations for future studies involving CPET as an outcome measure. It was decided by the study committee in the planning of the protocol that due to the uncertainty of the ability and the safety of those with reduced lung function to perform a maximal exercise test, those in the 'moderate' disease group would undertake only a baseline CPET test. In retrospect, it would have been more valuable to undertake repeat CPET on all enrolled participants. With the experience gained from this study, this is something that could explored in the future. Finally, the COVID-19 pandemic adversely affected the ability to perform follow-up CPET testing, particularly in this highly vulnerable group of individuals; consequently the resulting sample size was small.

Conclusion

In conclusion, our study provides the initial data to support the feasibility of CPET in at least mildmoderate populations of IPF and the ability of repeated CPET to assess the change in health status over time. This may be clinically applied in the future to assess the response to pharmacological or non-pharmacological interventions from the patient's perspective. Future work should concentrate on examining the relationship between CPET parameters, lung function and CT-derived measures of disease, establishing the MCID for longitudinal change in CPET in ILD.