





# The prognostic value of cardiopulmonary exercise testing in interstitial lung disease: a systematic review

Shaney L. Barratt <sup>1,2</sup>, Richard Davis<sup>1</sup>, Charles Sharp <sup>3</sup> and John D. Pauling<sup>4,5</sup>

**Affiliations:** <sup>1</sup>Academic Respiratory Unit, School of Clinical Sciences, University of Bristol, Bristol, UK. <sup>2</sup>Bristol Interstitial Lung Disease Service, North Bristol NHS Trust, Westbury-on-Trym, Bristol, UK. <sup>3</sup>Gloucestershire Hospitals NHS Foundation Trust, Gloucester, UK. <sup>4</sup>Department of Pharmacy and Pharmacology, University of Bath, Bath, UK. <sup>5</sup>Royal National Hospital for Rheumatic Diseases, Royal United Hospitals NHS Foundation Trust, Bath, UK.

**Correspondence:** Shaney L. Barratt, Consultant Respiratory Physician and Honorary Associate Researcher, North Bristol NHS Trust, Southmead Hospital, Westbury-on-Trym, Bristol BS10 5NB, UK. E-mail: shaney.barratt@nbt.nhs.uk

**ABSTRACT** The heterogeneity of interstitial lung disease (ILD) results in prognostic uncertainty concerning end-of-life discussions and optimal timing for transplantation. Effective prognostic markers and prediction models are needed. Cardiopulmonary exercise testing (CPET) provides a comprehensive assessment of the physiological changes in the respiratory, cardiovascular and musculoskeletal systems in a controlled laboratory environment. It has shown promise as a prognostic factor for other chronic respiratory conditions. We sought to evaluate the prognostic value of CPET in predicting outcomes in longitudinal studies of ILD.

MEDLINE, Embase and the Cochrane Database of Systematic Reviews were used to identify studies reporting the prognostic value of CPET in predicting outcomes in longitudinal studies of ILD. Study quality was assessed using the Quality in Prognosis Study risk of bias tool.

Thirteen studies were included that reported the prognostic value of CPET in ILD. All studies reported at least one CPET parameter predicting clinical outcomes in ILD, with survival being the principal outcome assessed. Maximum oxygen consumption, reduced ventilatory efficiency and exercise-induced hypoxaemia were all reported to have prognostic value in ILD. Issues with study design (primarily due to inherent problems of retrospective studies, patient selection and presentation of numerous CPET parameters), insufficient adjustment for important confounders and inadequate statistical analyses limit the strength of the conclusions that can be drawn at this stage.

There is insufficient evidence to confirm the value of CPET in facilitating “real-world” clinical decisions in ILD. Additional prospective studies are required to validate the putative prognostic associations reported in previous studies in carefully phenotyped patient populations.

 @ERSpublications

**There is presently insufficient evidence to confirm the value of CPET in facilitating “real-world” clinical decisions in ILD. Additional prospective studies are required to validate the putative prognostic associations reported in previous studies.** <https://bit.ly/3dfp5kq>

**Cite this article as:** Barratt SL, Davis R, Sharp C, *et al.* The prognostic value of cardiopulmonary exercise testing in interstitial lung disease: a systematic review. *ERJ Open Res* 2020; 6: 00027-2020 [<https://doi.org/10.1183/23120541.00027-2020>].



## Introduction

The heterogeneity of interstitial lung disease (ILD) [1, 2] presents challenges for patients and clinicians in terms of treatment choices, optimal timing of end-of-life discussions [3], or referral for transplantation [4] and clinical trial design [5, 6].

Cardiopulmonary exercise testing (CPET) provides a comprehensive assessment of the physiological changes that occur in the respiratory, cardiovascular and musculoskeletal systems during exercise, in a controlled laboratory environment [7, 8], and is considered the gold standard for evaluating maximal/symptom-limited exercise tolerance in patients with pulmonary and cardiac disease [9]. Although CPET has been available for decades, recent evidence is emerging to support its use in the prognostication of chronic cardiopulmonary disease [10, 11], with increasing interest in its application in ILD [12].

Maximum/peak oxygen consumption ( $\text{VO}_{2\text{max}}$  or peak  $\text{VO}_2$ ) is a measurement of the capacity for aerobic exercise and is determined by variables that define oxygen delivery by the Fick equation [13]. In patients with ILD, limitations on exercise may be the consequence of either ventilatory mechanical limitation (by reaching their ventilatory ceiling, typically thought to be 80% of maximal voluntary ventilation (MVV)), abnormal gas exchange (or reduction in ventilatory efficiency; indicated by variables such as the increment in minute ventilation (VE) relative to carbon dioxide production ( $\text{VE}/\text{VCO}_2$ )) and/or diffusion limitation (indicated by variables such as reduction in oxygenation >5% or hypoxia at anaerobic threshold (AT)/peak exercise) [13].

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 and support cohort enrichment for ILD clinical trials.

## Materials and methods

The study protocol was prepared in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [14] and registered in the International Prospective Register of Systematic Reviews (PROSPERO 110198/2018) (study commencement date 1 November 2018, completion date 30 September 2019). In brief, eligible studies included cohort (retrospective or prospective) studies reporting the prognostic value of CPET results in adult populations of ILD.

The primary objective was to evaluate the prognostic value of CPET in predicting disease course and outcomes in longitudinal studies of ILD. We explored the relationship between CPET and a broad range of relevant clinical outcomes including, but not limited to, relevant disease outcomes (*e.g.* death, hospitalisation), potential surrogates of disease severity (*e.g.* worsening lung physiology, *etc.*), and future deterioration in health-related quality of life (HRQoL) and/or functional status. Where possible, a comparison of the prognostic value of CPET was made across different ILD subtypes.

Studies were excluded if an ILD cohort was not described and reported separately. Non-original research publications and abbreviated reports were excluded. Randomised controlled trials (RCTs) were excluded as we did not expect this to be an appropriate methodological design for assessing the prognostic value of CPET. An amendment to our originally registered protocol (English language articles only) enabled the inclusion of a relevant non-English (French) publication.

The search criteria were developed in accordance with search recommendations for systematic reviews of evaluations of prognostic variables [15]. Electronic searches were performed in MEDLINE, Embase and the Cochrane Database of Systematic Reviews (CDSR), with no publication date or language restrictions. Full details of the specific search criteria applied are presented as supplementary material 1. All titles and abstracts were screened independently by two review authors (RD and CS), and agreement was assessed using Cohen's kappa statistics [16]. Any discrepancies/disagreements were resolved by discussion between reviewers and included a third party (SLB) if necessary.

A formal systematic review management platform was not used for this study. EndNote (Alfasoft Limited) was used to facilitate the combination of multiple database results and deduplication. A standardised data extraction form was used (independently by RS and CS, with subsequent verification by SLB) to extract relevant study details from selected studies.

A meta-analysis was planned if appropriate and feasible. A narrative, qualitative data synthesis was planned if wide heterogeneity in study design and CPET analysis precluded quantitative analysis. Study quality was assessed using the QUIPS (Quality in Prognosis Study) risk of bias tool by two reviewers (RD and CS) [17], with agreement measured using Cohen's kappa (supplementary material 2).

## Results

Simultaneous searches of Embase (n=573) and MEDLINE (n=373), performed on 13 April 2019, identified 946 articles. A search of the CDSR did not identify any additional studies. As anticipated, we did not identify any relevant RCTs in our study selection process and no studies were excluded on the basis of an RCT design. After removal of duplicates, 658 titles and abstracts were screened for eligibility. There was moderate initial agreement between the two reviewers (Cohen's kappa 0.462, supplementary material 3). Discordance for 20 studies was due to a single non-clinically trained reviewer choosing to include questionable studies for consideration (all of which were easily resolved through discussion and subsequently excluded). Due to the nature of discordance, it was not felt that retraining reviewers and formally repeating the title and abstract selection process would benefit the review process.

Following full text review, 13 studies were eligible for full data extraction (figure 1).

### Study design

Table 1 summarises the study design and reported findings of the final 13 studies.

The majority were retrospective cohort analyses (11/13, 85%), with variable follow-up (the majority <4 years [7, 18–25], one study with 5 year follow-up [26]; range 23 days–20 years follow-up [19, 27]).

There were two prospective studies [28, 29]. One investigated the relationship between CPET and survival characteristics in IPF with variable follow-up between 9 and 64 months [28]. The other used CPET as part of a wider investigation into the role of exercise testing in the prognostication of ILD and followed patients for a fixed period of 40 months [29].

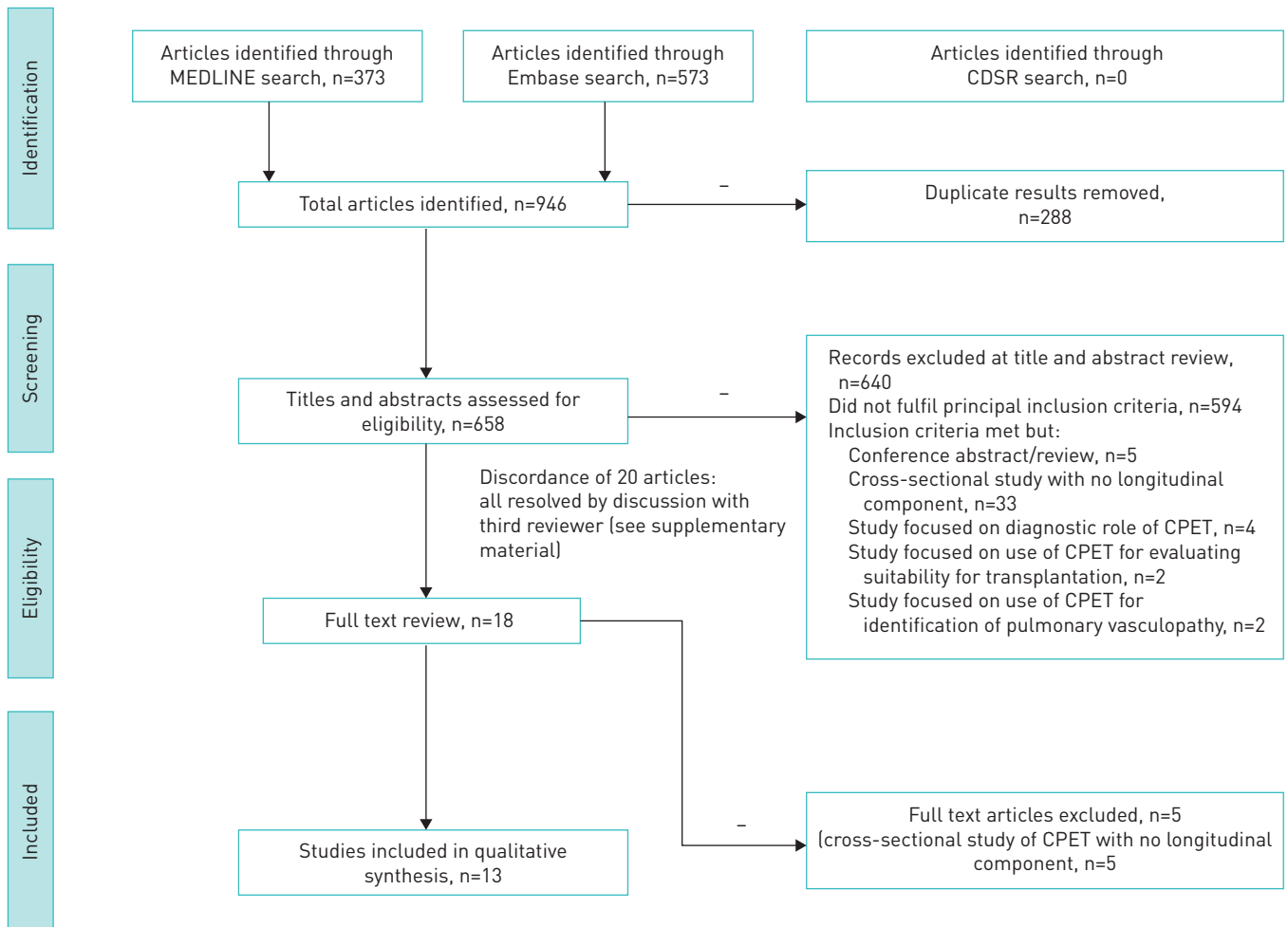


FIGURE 1 Study selection flow diagram presented according to PRISMA statement. CDSR: Cochrane Database of Systematic Reviews; CPET: cardiopulmonary exercise testing.

TABLE 1 Study characteristics of papers selected for full data extraction

First author, date [ref], origin	Description	Study population and attrition	CPET method and CPET parameters	Exclusion	Disease outcomes	Statistical methods to investigate CPET and outcome	Summary of key reported outcomes	Comments
<b>TRIANTELLIDOU 2013 [28], Greece</b>	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/VCO <sub>2</sub> slope, VO <sub>2</sub> peak/kg, VE/VCO <sub>2</sub> ratio at AT.	Significant PH (PASP >45 mmHg 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/VCO <sub>2</sub> slope, VO <sub>2</sub> peak·kg <sup>-1</sup> and VE/VCO <sub>2</sub> at AT were significant survival predictors. Optimal model for mortality risk estimation combined VO <sub>2</sub> peak·kg <sup>-1</sup> with DLCO (p<0.0001). Per 1 unit increase in VO <sub>2</sub> peak·kg <sup>-1</sup> (1 mL·kg <sup>-1</sup> ·min <sup>-1</sup> ) and DLCO% (1%), mortality rate was reduced by 32% and 13%, respectively. VO <sub>2</sub> peak threshold of 14.2 mL·min <sup>-1</sup> ·kg <sup>-1</sup> 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.
<b>VAINSELBOIM 2016 [29], Israel</b>	Prospective observational study evaluating role of 12 week exercise training programme 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 VO <sub>2</sub> ·kg <sup>-1</sup> , peak work rate, VE/VO <sub>2</sub> nadir, VE/VCO <sub>2</sub> ratio at AT, tidal volume reserve.	Non-IPF ILD. Clinically unstable in preceding 3–6 months, severe comorbid 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 increased mortality associated with cut-off points for: peak work rate <62 watts [AUC 0.854, 0.73–0.98 CI, p=0.005], peak VO <sub>2</sub> <13.8 mL·kg <sup>-1</sup> ·min <sup>-1</sup> [AUC 0.731, 0.56–0.90 CI, p=0.031], tidal volume reserve <0.48 L·breath <sup>-1</sup> [AUC 0.810, 0.66–0.96 CI, p=0.01], VE/VCO <sub>2</sub> at AT >34 [AUC 0.783, 0.6–0.96 CI, p=0.02] and VE/VO <sub>2</sub> nadir >34 [AUC 0.736, 0.56–0.90, 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 VO <sub>2</sub> 4.4 (0.94–20.3), tidal volume reserve 7.6 (1.6–35.2), VE/VO <sub>2</sub> nadir 8.3 (2.2–31.6), VE/VO <sub>2</sub> at AT 4.6 (1.2–17.3). Non-survivors were characterised by higher dyspnoea levels, the presence of PH (assessed by ECHO sPAP>35 mmHg), and CPET markers of reduced ventilatory efficiency (VE/VO <sub>2</sub> nadir p=0.039, VE/VCO <sub>2</sub> at AT p=0.008) and reduced exercise capacity (peak work rate p=0.01, peak VO <sub>2</sub> p=0.02).	Prospective observational study analysis as part of a wider single-centre RCT. Underpowered to detect survival differences between groups. Small sample size. Higher prevalence of PH in non-survivors.

Continued

TABLE 1 Continued

First author, date [ref], origin	Description	Study population and attrition	CPET method and CPET parameters	Exclusion	Disease outcomes	Statistical methods to investigate CPET and outcome	Summary of key reported outcomes	Comments
<b>KING 2001 [23], USA</b>	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)O <sub>2</sub> corrected for FiO <sub>2</sub> , VD/VT, VO <sub>2</sub> , maximal workload.	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, 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. Multivariable 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 (included age, smoking history, clubbing, extent of profusion of interstitial opacities, presence/absence of PH on CXR, % predicted TLC and PaO <sub>2</sub> at the end of maximal exercise). Exercise PaO <sub>2</sub> only exercise variable included in the model, accounting for 10.5% of score [PaO <sub>2</sub> 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 used CXR rather than HRCT in 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.
<b>Miki 2003 [21], Japan</b>	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). PaO <sub>2</sub> , PaCO <sub>2</sub> , HR, respiratory frequency (f), Vt, VE, peak VO <sub>2</sub> , VE/VO <sub>2</sub> , VE/VCO <sub>2</sub> , VO <sub>2</sub> /HR, AaDO <sub>2</sub> and PaO <sub>2</sub> slope.	CTD, sarcoid, OP, EP, HP, cardiac disease, anaemia, primary cardiac disease, 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 follow-up.	Respiratory death	Exercise parameters (between groups split by PaO <sub>2</sub> 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 PaO <sub>2</sub> 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, VO <sub>2</sub> max (HR 0.997, 0.995–0.999 CI, p=0.012), VO <sub>2</sub> /HRR max (HR 0.69, 0.51–0.93 CI, p=0.014), PaO <sub>2</sub> slope (HR 0.68, 0.51–0.89 CI, p=0.006), VE/VCO <sub>2</sub> (HR 1.04, 1.006–1.07 CI, p=0.020) and age (HR 1.1, 1.02–1.18 CI, p=0.014) associated with survival in IPF. On multiple regression, PaO <sub>2</sub> slope (HR 0.84, 0.73–0.97 CI, p=0.015) and age (HR 1.096, 1.01–1.19 CI, p=0.025) independently related to survival. When PaO <sub>2</sub> slope was divided into steep (<–60 mmHg·L <sup>-1</sup> ·min <sup>-1</sup> ) and gentle (>–60 mmHg·L <sup>-1</sup> ·min <sup>-1</sup> ), median survival time after CPET significantly shorter in steep group (1.6 versus 4.5 years).	Retrospective, single-centre cohort. Large number of exclusion criteria. Outcomes limited to respiratory deaths.

Continued

TABLE 1 Continued

First author, date [ref], origin	Description	Study population and attrition	CPET method and CPET parameters	Exclusion	Disease outcomes	Statistical methods to investigate CPET and outcome	Summary of key reported outcomes	Comments
<b>FELL 2009 [24], USA</b>	Retrospective study evaluating prognostic value of CPET in IPF. Mean follow-up not reported.	117 IPF pts. 10 pts excluded from survival analysis as VO <sub>2</sub> max changed between baseline and 6 months.	Cycle ergometer. Blood gas analysis. Peak VO <sub>2</sub> ·kg <sup>-1</sup>	Patients with CTD, occupational or environmental exposure, histological pattern other than UIP.	Survival	Multivariate Cox proportional hazard models studied the predictive value of peak VO <sub>2</sub> adjusting for age, gender, smoking status, baseline FVC% and baseline DLCO%. Resulting HR were plotted against peak VO <sub>2</sub> to determine thresholds. Survival thresholds examined with Kaplan–Meier survival curves, log-rank tests and multivariate Cox proportional hazard models.	Peak VO <sub>2</sub> ·kg <sup>-1</sup> examined as a continuous variable did not predict survival HR 0.969 (p=0.55). Baseline threshold peak VO <sub>2</sub> <8.3 mL·kg <sup>-1</sup> ·min <sup>-1</sup> 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 <i>priori</i> plan. Small number of pts below VO <sub>2</sub> 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.
<b>WALLAERT 2011 [22], France</b>	Retrospective multicentre study evaluating prognostic role of CPET in determining 3-year survival in IPF.	63 IPF patients	Cycle ergometer. Blood gas analysis. Peak VO <sub>2</sub> ·kg <sup>-1</sup> , VE/VO <sub>2</sub> at ventilatory threshold, VE/CO <sub>2</sub> , (VO <sub>2</sub> /HRR), P(A-a)O <sub>2</sub> , 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 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/VO <sub>2</sub> at ventilatory threshold, FVC (% pred) and P(A-a)O <sub>2</sub> . The most appropriate logistic regression model incorporated two variables, with the lowest 3 year survival when TLC (<65%) and VE/VO <sub>2</sub> 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.

Continued

TABLE 1 Continued

First author, date [ref], origin	Description	Study population and attrition	CPET method and CPET parameters	Exclusion	Disease outcomes	Statistical methods to investigate CPET and outcome	Summary of key reported outcomes	Comments
<b>GLÄSER 2013 [18], Germany</b>	Retrospective study evaluating predictive value of CPET measures for the presence of PH in IPF. 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 $\text{VO}_2$ , $\text{VO}_2$ at AT ( $\text{mL}\cdot\text{min}^{-1}$ ), VE/MVV, VE versus $\text{VCO}_2$ slope, VE max, Vt max, Vt max/IC, VE/MVV.	Pts with left heart disease (ECHO $\pm$ PWP > 14 mmHg 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 chi-squared test used for comparison of IPF pts with/without PH. Cox proportional hazards analysis used for pulmonary variables and endpoint. 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. Presence of PH best predicted by gas exchange efficiency during exercise and peak oxygen uptake (VE versus $\text{VO}_2$ slope pred ( $\geq 152.4$ , AUC 0.938, 0.892–0.984 CI) and $\text{VO}_2$ peak ( $\leq 56.3$ , AUC 0.832, 0.753–0.911 CI)). By univariate analysis, the presence of PH (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 $\text{VO}_2$ pred (HR 0.96, $p=0.001$ ) and $\text{VO}_2$ at AT pred (HR 0.97, $p=0.017$ ) also being statistically significant. In multivariate analysis, invasively measured pH and peak $\text{VO}_2$ pred were independent predictors for survival.	Retrospective multicentre study. Potential recruitment bias due to selected cohort (specialist centres, excluded left heart disease).
<b>VAN DER PLAS 2014 [20], Netherlands</b>	Retrospective study exploring predictive value of CPET and ECHO parameters for survival in IPF. Mean follow-up 42.3 $\pm$ 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), $\text{VO}_2$ peak (% pred), VE peak (% pred), breathing reserve (%), HRR peak (% pred), VE/ $\text{VCO}_2$ ratio at AT, $\text{VO}_2$ /HRR (% pred), $\text{ETCO}_2$ 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 and CPET parameters. Patients were grouped into those with/without sPAP $\geq 40$ mmHg and differences in exercise parameters analysed with unpaired <i>t</i> -test or chi-squared test. ROC curve analysis was used to determine variables that predict sPAP $\geq 40$ mmHg. 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 (76%) had a reduced $\text{VO}_2$ peak (i.e. < 84% predicted). VE/ $\text{VCO}_2$ at AT was significantly higher in patients with sPAP $\geq 40$ mmHg ( $n=11$ ) compared to those with sPAP $\leq 40$ mmHg ( $n=27$ ), (54.0 $\pm$ 21.9 versus 37.9 $\pm$ 7.5, $p=0.021$ ). VE/ $\text{VCO}_2$ at AT was shown to be a good predictor of sPAP $\geq 40$ mmHg by ROC curve analysis but only VE/ $\text{VCO}_2$ at AT and not sPAP $\geq 40$ mmHg was shown to predict survival. Pts with VE/ $\text{VCO}_2$ at AT $\leq 45$ ( $n=24$ ) had a significantly better prognosis than those with VE/ $\text{VCO}_2$ $\geq 45$ ( $n=14$ ), 81.3 $\pm$ 14.1 versus 21.0 $\pm$ 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/ $\text{VCO}_2$ at AT for survival.	Retrospective analysis of prospective database. Single centre.

Continued

TABLE 1 Continued

First author, date [ref], origin	Description	Study population and attrition	CPET method and CPET parameters	Exclusion	Disease outcomes	Statistical methods to investigate CPET and outcome	Summary of key reported outcomes	Comments
<b>KOLLERT 2011 [25], Germany</b>	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)O <sub>2</sub>	Patients who could not complete CPET >6 min, in the absence of extra-cardiopulmonary 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)O <sub>2</sub> 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 FVC (OR 0.954, 0.917–0.992 CI, p=0.009) and P(A-a)O <sub>2</sub> (OR 1.098, 1.039–1.160 CI, p<0.0001) during exercise were independently associated with a need for prolonged immunosuppressive treatment. No other CPET variables reported.	No other CPET variables described in analysis and thus potential for reporting bias. Unable to determine exact clinical characteristics of this longitudinal cohort from the data presented.
<b>LOPES 2012 [26], Brazil</b>	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 VO <sub>2</sub> (% pred), % peak VO <sub>2</sub> at lactate threshold, VCO <sub>2</sub> /VO <sub>2</sub> , VO <sub>2</sub> /HRR, maximum respiratory rate, breathing reserve, HRR, P(A-a)O <sub>2</sub> , ΔSpO <sub>2</sub> , 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 5 years 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. MLR used to identify factors independently related to decreased lung function.	Significant reductions in FVC (relative variation –5.1% [–23.1 to 0%]) and DLCO (relative variation –2.5% [–44.4 to 0.93%]) at 5 years follow-up. Peak VO <sub>2</sub> (% pred), breathing reserve, maximum RR, P(A-a)O <sub>2</sub> and ΔSpO <sub>2</sub> correlated with FVC and DLCO values that had declined >10% from baseline (p<0.0001 for all parameters). P(A-a)O <sub>2</sub> >22 mmHg (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.

Continued



TABLE 1 Continued

First author, date [ref], origin	Description	Study population and attrition	CPET method and CPET parameters	Exclusion	Disease outcomes	Statistical methods to investigate CPET and outcome	Summary of key reported outcomes	Comments
LAYTON 2017 [7], USA	Retrospective study evaluating predictive value of CPET for 1-year transplant-free survival in a population of ILD patients undergoing lung transplant evaluation.	192 pts had CPET performed on oxygen. Four tests terminated due to oxygen desaturation (nadir SpO <sub>2</sub> <80% despite 30% FiO <sub>2</sub> ). Three tests terminated early due to low ETCO <sub>2</sub> (<18 mmHg) or elevated ETCO <sub>2</sub> (>60 mmHg), reducing cohort to 185 pts.	Cycle ergometer, pulse oximetry. Peak VO <sub>2</sub> (mL·kg <sup>-1</sup> ·min <sup>-1</sup> , % predicted), workload (watts, % predicted), VE/VCO <sub>2</sub> slope (% predicted), ETCO <sub>2</sub> mmHg and O <sub>2</sub> 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 1 year).	Comparison of variables between those who died / transplanted (D/LTx) and those who survived transplant-free were compared using two-sample independent <i>t</i> -test. Survival was calculated by Kaplan–Meier method, with univariable Cox regression analysis to identify predictors of 1 year 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.	79 D/LTx during follow-up period. 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 than 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, AUC=0.740) nadir CPET SpO <sub>2</sub> <86% despite 30% FiO <sub>2</sub> (HR 2.27, 1.41–3.68 CI, AUC=0.645) FVC% predicted <45% (HR 1.82, 1.15–2.87 CI, AUC 0.624) as discriminatory parameters predicting 1-year mortality or need for transplant. 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 covariate 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 programmes.
KAWUT 2005 [19], USA	Retrospective study of CPET and 6MWT 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. SaO <sub>2</sub> (unloaded, peak, recovery), peak VO <sub>2</sub> ·kg <sup>-1</sup> , VO <sub>2</sub> /HR peak, VCO <sub>2</sub> unloaded, VE unloaded.	Pts evaluated at another lung transplantation centre. Other forms of DPLD, e.g. LAM, pulmonary Langerhans cell histiocytosis/histiocytosis X, EP and granulomatous DPLD, e.g. sarcoidosis.	All-cause mortality. Death on the lung transplantation waiting list.	Cox proportional hazards regression to identify predictors of time-to-death. Individual models were constructed using LTx as a time-dependent 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 post-transplantation). 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 6MWT <350 m (HR 4.6, 1.5–14.2 CI, p=0.009), peak VO <sub>2</sub> ·kg <sup>-1</sup> (HR 0.88, 0.79–0.99 CI, p=0.039) (no threshold determined) and VE/VCO <sub>2</sub> >46 (p=0.05) (were each associated with increased risk of death). SpO <sub>2</sub> <95% during unloaded exercise had 75% chance of dying on transplantation list (sensitivity 86%, specificity 89%). 67% chance of death if 6MWT <350 m.	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 programmes.

Continued

TABLE 1 Continued

First author, date [ref], origin	Description	Study population and attrition	CPET method and CPET parameters	Exclusion	Disease outcomes	Statistical methods to investigate CPET and outcome	Summary of key reported outcomes	Comments
SWIGRIS 2009 [27], USA	Retrospective study exploring prognostic role of SpO <sub>2</sub> and SaO <sub>2</sub> 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. SpO <sub>2</sub> and SaO <sub>2</sub> at rest and during maximal exercise (SpO <sub>2</sub> max). VO <sub>2</sub> max measured but not reported.	Pulmonary hypertension, overlap syndromes.	Mortality	Cox proportional hazard models were used to examine the prognostic capabilities of SpO <sub>2</sub> , 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, SpO <sub>2</sub> predicted mortality; SpO <sub>2</sub> max <89% (HR 2.4, 95% CI 1.2 to 4.9, p=0.02), SpO <sub>2</sub> max fall >4% from baseline (HR 2.4, 95% CI 1.1 to 5.0, p=0.02), alongside ΔSpO <sub>2</sub> (HR 1.08, 95% CI 1.03 to 1.14, p=0.002). Controlling for FVC%, the ΔSpO <sub>2</sub> 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.

Abbreviations: ΔSpO<sub>2</sub>: difference between peak and resting oxygen saturation; 6MWT: 6-minute walk test distance; AaDO<sub>2</sub>: alveolar–arterial oxygen pressure difference; AT: anaerobic threshold; AUC: area under the curve; BR: breathing reserve [1 – (VE during exercise/MV)] × 100; CI: confidence interval; COP: cryptogenic organising pneumonia; CPET: cardiopulmonary exercise testing; CPI: composite physiologic index; CTD: connective tissue disease; CXR: chest X-ray; D: died/deaths; DLCO: diffusion capacity of lungs for carbon dioxide; DPLD: diffuse parenchymal lung disease; ECHO: echocardiogram; EP: eosinophilic pneumonia; ETCO<sub>2</sub>: end tidal carbon dioxide; FiO<sub>2</sub>: fraction of inspired oxygen; FVC: forced vital capacity; HP: hypersensitivity pneumonitis; HR: hazard ratio; HRCT: high-resolution computed tomography; HRR: heart rate; IC: inspiratory capacity; ILD: interstitial lung disease; IPF: idiopathic pulmonary fibrosis; LAM: lymphangioliomyomatosis; LTx: lung transplantation; max: maximal; MLR: multiple logistic regression; MVV: maximum voluntary ventilation (can be measured or estimated as FEV1 × 41); OR: odds ratio; PaCO<sub>2</sub>: partial pressure of carbon dioxide; PaO<sub>2</sub>: partial pressure of oxygen; pts: patients; NSIP: non-specific interstitial pneumonia; P(A-a)O<sub>2</sub>: alveolar–arterial oxygen pressure gradient at peak exercise; PH: pulmonary hypertension; pred: predicted; PVD: peripheral vascular disease; PWP: pulmonary capillary wedge pressure; RCT: randomised controlled trial; RHC: right heart catheter; ROC: receiver operating characteristic curve; RR: respiratory rate; SaO<sub>2</sub>: oxygen saturation of arterial blood; sPAP: systolic pulmonary artery pressure; SpO<sub>2</sub>: oxygen saturation measured by pulse oximetry; SSc: systemic sclerosis; TLC: total lung capacity; UIP: usual interstitial pneumonia; VCO<sub>2</sub>: carbon dioxide production; VD/VT: physiological dead space/tidal volume ratio; VE: minute ventilation; VE/VCO<sub>2</sub>: ventilatory equivalent for carbon dioxide; VE/VO<sub>2</sub>: ventilatory equivalent for oxygen; VO<sub>2</sub>: oxygen uptake; VO<sub>2</sub> slope: PaO<sub>2</sub> plotted against VO<sub>2</sub>; VO<sub>2</sub>/HRR max or oxygen pulse: oxygen delivery per heartbeat; VT: ventilatory threshold (highest VO<sub>2</sub> sustained without lactic acidosis); Vt: tidal volume; tidal volume reserve: Vt max-Vt resting.

### Patient populations

The majority (8/13, 62%) exclusively recruited patients with IPF, with retrospective assessment of 703 IPF patients and prospective assessment of 59 IPF patients. Classification of IPF was based on accepted criteria used at the time of enrolment: the 2000 American Thoracic Society (ATS) international consensus (IC) statement [1, 20, 21, 23, 24, 28] and the later 2002 ATS/ERS (European Respiratory Society) IC classification of the idiopathic interstitial pneumonias (including IPF) [19, 22, 30]. The updated 2011 ATS/ERS/JRS/ALAT guidelines for the diagnosis of IPF [31] were applied in all [7, 18, 28, 29] but one of the studies [20] published after 2011 (the latter was a retrospective study that may have recruited patients prior to the 2011 guidelines).

Two retrospective studies explored the prognostic role of CPET in 144 histologically confirmed sarcoidosis patients [25, 26], representing Scadding disease stages 1–4 [32]. Only one retrospective study had examined the prognostic role of CPET in systemic sclerosis ILD (SSc-ILD) (n=83) [27]. Patients with SSc met classification criteria adopted by the 1980 American Rheumatology Association [33] and those with SSc sine scleroderma met criteria proposed by Poormoghim and colleagues [34]. A diagnosis of ILD was based on chest radiography in 60/83 patients [27].

The prognostic role of CPET in other secondary causes of ILD (such as myositis, occupational causes of ILD and hypersensitivity pneumonitis (HP)) and/or other forms of idiopathic interstitial pneumonias (IIP) has not been well studied. Two retrospective studies have reported the prognostic value of CPET in mixed ILD populations referred for lung transplantation [7, 19], but low patient numbers precluded useful subgroup analyses.

The majority of studies had a moderate (6/13, 46%) or high (4/13, 31%) [7, 19, 21, 25] risk of bias for participant selection. For example, generalisability in one study was limited by the lack of clearly defined clinical characteristics (e.g. Scadding disease stage) in patients followed longitudinally (102/149) [30]. Studies enrolling from populations referred for lung transplantation resulted in selected cohorts of advanced ILD patients [7, 19]. Others incorporated *a priori* patient grouping, for example the presence of pulmonary hypertension (PH) [18], to enrich populations with those at higher risk of outcomes of interest, or actively excluded relevant patients, e.g. those that died from a cause other than respiratory failure [21].

Study attrition was generally low, consistent with the retrospective nature of the majority of studies. The QUIPS risk of bias for study attrition was high in two studies. Over 25% of patients were excluded from the analyses by Lopes *et al.* [26] (due to smoking history, concomitant respiratory disease, cardiac disease and neuromuscular disease). In another study, 34% (80/238) of the original study population were excluded from the analysis because of incomplete data sets [23].

### 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 [7, 19, 28] or lung function parameters [18]. One study incorporated CPET with other clinical, radiological and resting physiological assessments to devise a scoring system to predict survival in newly diagnosed cases of IPF (the clinical–radiological–physiological score) [23].

In two studies, CPET was the principal method of achieving maximal exercise [25, 27], with arterial blood gas sampling or peripheral oxygenation measurements used to determine the effect of exercise on gas exchange. In both studies, typical CPET measures, such as peak  $\text{VO}_2$ , were not reported.

The bias rating for prognostic factor measurement using the QUIPS tool was generally low to moderate (figure 2), with the majority of studies reporting a standardised approach to CPET (albeit individualised for each study) and analysis that would be easily reproducible and not amenable to bias. Most studies provided a sufficient description of the CPET protocol used (6/10, 60%), adhering to the 2003 ATS statement on CPET testing [7, 18–20, 22, 28]. Variance in the use of supplemental oxygen during CPET was observed; oxygen usage was an inclusion criteria in one study [7], whilst in others, supplemental oxygen was applied variably, depending on a pre-study requirement for home oxygen or saturation on room air <90% [19]. In 7 of 13 (54%) studies, blood gas analysis was used to assess the adequacy of gas exchange during exercise [21–27], whilst the remainder used pulse oximetry, considered by some experts to be a suboptimal substitute [13]. 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, with exercise increments based on a patient's daily activities and parameters of resting pulmonary function; this raises concerns regarding variation amongst subjects [21]. Furthermore, inherent differences in physiological responses recorded by the two ergometers during incremental exercise have been well defined, and make direct comparison of the two methods problematic [35, 36].

First author, year	Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Study confounding	Statistical analysis and reporting
KING 2001 [23]	Amber	Red	Amber	Green	Amber	Amber
MIKI 2003 [21]	Red	Green	Amber	Amber	Red	Red
KAWUT 2005 [19]	Red	Green	Amber	Amber	Red	Red
SWIGRIS 2009 [27]	Amber	Green	Green	Green	Amber	Red
FELL 2009 [24]	Amber	Green	Amber	Green	Red	Red
WALLAERT 2011 [22]	Green	Green	Amber	Green	Red	Red
KOLLERT 2011 [25]	Red	Green	Amber	Amber	Red	Red
LOPES 2012 [26]	Amber	Red	Amber	Green	Red	Red
TRIAANTAFILLIDOU 2013 [28]	Green	Green	Green	Green	Amber	Amber
GLÄSER 2013 [18]	Amber	Green	Amber	Green	Red	Amber
VAINSELBOIM 2016 [29]	Green	Green	Green	Green	Red	Red
LAYTON 2017 [7]	Red	Green	Amber	Green	Red	Amber

**FIGURE 2** The Quality in Prognosis Study (QUIPS) risk of bias tool assessment of included studies. Green indicates low risk of bias, amber indicates a moderate risk and red indicates a high risk of bias.

#### **Outcome measurement**

Eleven of 13 studies (85%) evaluated mortality. The majority of these (10/11, 91%) examined all-cause mortality, considering death or lung transplantation as composite endpoint. The remaining study used an outcome measurement that was restricted to respiratory deaths only [21]. One study assessed the discriminatory ability of CPET to identify patients who would die on the lung transplant list before receiving transplantation [19]. Other outcomes included interceding PH [18], decline in pulmonary function (FVC, forced vital capacity), decline in DLCO (diffusion capacity for carbon monoxide) and/or duration of immunosuppressive therapy in sarcoidosis [25, 26].

The risk of bias for outcome measure assessment was considered low to moderate across all studies (fig 2).

#### **Reported prognostic associations of CPET in ILD**

All studies reported at least one positive association between CPET and clinical outcomes, raising the possibility of positive reporting bias. A summary of the main findings is presented in table 2. Significant heterogeneity in study design, study populations (and classification criteria adopted), CPET protocols, CPET endpoints and defined endpoints precluded a meta-analysis.

#### *Maximal oxygen consumption*

The prognostic value of measures of maximal oxygen consumption during CPET on ILD outcomes have been reported in 10/13 (77%) studies (table 2).

Peak  $\text{VO}_2 \cdot \text{kg}^{-1}$  inversely correlated with increased 1-year mortality in two cohorts of patients with severe ILD referred for lung transplantation [7, 19], whilst peak  $\text{VO}_2$  thresholds ranging from  $<8.3$  to  $<14.2 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  [24, 28, 29] were reported to predict mortality in IPF. These results contrasted with the findings of other studies that failed to identify any significant association [20–22].

#### *Ventilatory efficiency*

The prognostic value of the ventilatory equivalent for  $\text{CO}_2$  at AT ( $\text{VE}/\text{VCO}_2$  at AT) at levels ranging between  $>34$  and  $>46$  was reported to predict survival in IPF [7, 19, 20], even after correcting for functional severity of ILD [20] (table 2).

The ventilatory equivalent for oxygen at AT ( $\text{VE}/\text{VO}_2$  at AT) was also reported to be a poor predictor of survival in IPF patients [21, 22] and whilst  $\text{VE}/\text{VO}_2$  was associated with worse IPF survival in the derivation cohort of the clinical–radiological–physiological multimodal score, even after adjustment for age and smoking status, it was not included as a parameter in the final model [21].

TABLE 2 Reported associations between CPET parameters and outcomes in studies of ILD

CPET measurement	Studies	Threshold	Outcome	
<b>Maximal oxygen consumption (peak VO<sub>2</sub>)</b>	FELL 2009	No association as continuous variable (HR 0.969, p=0.55). Peak VO <sub>2</sub> <8.3 mL·kg <sup>-1</sup> ·min <sup>-1</sup> associated with worse outcome (n=8; HR 3.24, 1.10–9.56 CI, p=0.03).	Survival in IPF.	
	GLÄSER 2013	Peak VO <sub>2%</sub> pred <56.3% Peak VO <sub>2%</sub> pred (multivariate analysis) – no threshold determined.	Presence of PH. Survival in IPF.	
	KAWUT 2005	Peak VO <sub>2</sub> ·kg <sup>-1</sup> (no threshold determined), associated with worse outcome.	Mortality at 1 year of mixed ILD patients referred for transplantation.	
	KING 2001	Methodology suggested peak VO <sub>2</sub> was recorded but result not reported in results section.	Survival in IPF.	
	KÖLLERT 2011	Not measured.	Prolonged immunosuppressive therapy (>1 year) in sarcoidosis.	
	LAYTON 2017	Peak VO <sub>2</sub> ·kg <sup>-1</sup> and peak VO <sub>2%</sub> pred (association with univariate analysis but not multivariate). A 1 mL·kg <sup>-1</sup> ·min <sup>-1</sup> greater VO <sub>2</sub> reduced the risk of mortality/transplantation by 9%.	1-year mortality or transplantation in mixed population of ILD.	
	LOPES 2012	Peak VO <sub>2</sub> <50% pred (association on univariate but not multivariate analysis).	Decline of >10% FVC% pred and DLCO % pred at 5 years follow-up from baseline, in thoracic sarcoidosis.	
	MIKI 2003	Peak VO <sub>2</sub> (associated with worse outcome using univariate analysis, but not on multivariate).	Respiratory deaths in IPF.	
	SWIGRIS 2009	Although measured, not part of planned statistical analysis.	Mortality in SSc-ILD.	
	TRIANTAFILLIDOU 2013	Peak VO <sub>2</sub> <14.2 mL·kg <sup>-1</sup> ·min <sup>-1</sup> associated with worse outcome and further enforced when the model combines DLCO.	Survival in IPF.	
	VAINSELBOIM 2016	Peak VO <sub>2</sub> <13.8 mL·kg <sup>-1</sup> ·min <sup>-1</sup> associated with worse outcome (AUC 0.731, 0.56–0.9, p=0.031).	Mortality or transplantation in IPF.	
	VAN DER PLAS 2014	Peak VO <sub>2</sub> ·kg <sup>-1</sup> – no association.	Survival in IPF.	
	WALLAERT 2011	Peak VO <sub>2</sub> and peak VO <sub>2</sub> ·kg <sup>-1</sup> – no association as continuous variable.	3-year survival in IPF.	
	<b>Ventilatory efficiency (VE/VO<sub>2</sub>, VE/VCO<sub>2</sub>)</b>	FELL 2009	Not measured	Survival in IPF.
		GLÄSER 2013	VE/VCO <sub>2</sub> slope <sub>pred</sub> ≥152.4 predicted outcome (sensitivity 87.2%, specificity 88.4%).	Development of interceding PH in IPF.
KAWUT 2005		VE/VCO <sub>2</sub> >46 associated with worse outcome. The risk was non-proportional so could not be estimated with a single hazard ratio.	All-cause mortality at 1 year of mixed ILD patients referred for transplantation.	
KING 2005		VE/VO <sub>2</sub> associated with worse outcome when results adjusted for age and smoking status (HR 1.06). Not included in multivariable model.	Survival in IPF.	
KÖLLERT 2011		Not measured	Prolonged immunosuppressive therapy (>1 year) in sarcoidosis.	
LAYTON 2017		VE/VCO <sub>2</sub> slope (association with univariate analysis but not multivariate)	1-year mortality or transplantation in mixed population of ILD.	
LOPES 2012		Not measured	Decline in FVC and DLCO at 5 years in sarcoidosis.	
MIKI 2003		VE/VO <sub>2</sub> at max VE/VCO <sub>2</sub> at max (associated with worse outcome using univariate analysis, but not on multivariate).	Respiratory deaths in IPF.	
SWIGRIS 2009		Not measured	Survival in SSc-ILD.	
TRIANTAFILLIDOU 2013		VE/VCO <sub>2</sub> slope and higher VE/VCO <sub>2</sub> at AT predicted worse outcome.	Survival in IPF.	
VAINSELBOIM 2016		VE/VCO <sub>2</sub> at AT >34 and nadir VE/VO <sub>2</sub> >34 predicted worse outcome in univariate and bivariate analysis	Mortality in IPF.	

Continued

TABLE 2 Continued

CPET measurement	Studies	Threshold	Outcome
Diffusion limitation or exercise-induced hypoxaemia	VAN DER PLAS 2014	VE/VCO <sub>2</sub> at AT >45 associated with poorer survival (HR 4.58, p=0.001), even after correcting for lung function severity.	Survival in IPF.
	WALLAERT 2011	VE/VO <sub>2</sub> at AT >45 associated with worse outcome (multivariate analysis).	3-year survival in IPF.
	FELL 2009	Resting PaO <sub>2</sub> was associated with worse outcome (HR 0.934) when adjusted for age, sex, baseline physiology and smoking status. No threshold could be determined.	Survival in IPF.
	GLÄSER	Although SpO <sub>2</sub> monitored during CPET, not included in analysis.	Survival in IPF or development of interceding PH.
	KAWUT 2013	SaO <sub>2</sub> <95% during unloaded exercise (one of several variables) predicting worse outcome (p=0.0025).	All-cause mortality at 1 year of mixed ILD patients referred for transplantation.
		SaO <sub>2</sub> <95% during unloaded exercise (one of several variables) predicting worse outcome (sens. of 86%, spec. 89%).	Death on waiting list for lung transplantation.
	KING 2001	PaO <sub>2</sub> at maximal exercise associated with worse outcome and included in multivariable model (accounted for as much as 10.5% of the maximum score in the model).	Survival in IPF.
	KOLLERT 2011	P(A-a)O <sub>2</sub> associated with worse outcome (multivariate analysis, OR 1.098, p<0.001).	Prolonged immunosuppressive therapy (>1 year) in sarcoidosis.
	LAYTON 2017	Nadir CPET SpO <sub>2</sub> <86% independently associated with worse outcome (HR 2.27, p=0.001). Risk of death/lung transplantation increased two-fold when SpO <sub>2</sub> <86%.	1-year mortality or transplantation in mixed population of ILD.
	LOPES 2012	P(A-a)O <sub>2</sub> >22 mmHg associated with worse outcome (multivariate analysis, RR 70.0, p<0.001).	Decline of >10% FVC% pred and DLCO % pred at 5 years follow-up from baseline, in thoracic sarcoidosis.
	MIKI 2003	PaO <sub>2</sub> slope ( $\Delta$ PaO <sub>2</sub> / $\Delta$ VO <sub>2</sub> ) predicted worse outcome (multivariate analysis, HR 0.841, p=0.015). Those stratified $\leq -60$ mmHg·L <sup>-1</sup> ·min <sup>-1</sup> associated with worse survival (1.6 years <i>versus</i> 4.5 years).	Respiratory deaths in IPF.
	SWIGRIS 2009	SpO <sub>2</sub> at maximum exercise <89% (HR 2.4) or SpO <sub>2</sub> at maximum exercise fall >4 points from baseline (HR 2.4) associated with worse outcome.	Survival in SSc-ILD.
	TRIANTAFILLIDOU 2013	SpO <sub>2</sub> at peak exercise – no association	Survival in IPF.
	VAINSELBOIM 2016	Although SpO <sub>2</sub> monitored during CPET, not included in analysis.	Mortality in IPF.
	VAN DER PLAS 2014	Not specifically reported on.	Survival in IPF.
WALLAERT 2011	Higher P(A-a)O <sub>2</sub> associated with worse outcome using multivariate analysis, but was not included in the final logistic regression model.	3-year survival in IPF.	

Abbreviations: AT: anaerobic threshold; DLCO: diffusion capacity of lungs for carbon dioxide; FVC: forced vital capacity; HR: hazard ratio; ILD: interstitial lung disease; IPF: idiopathic pulmonary fibrosis; max: maximum; OR: odds ratio; P(A-a)O<sub>2</sub>: alveolar-arterial oxygen pressure gradient at peak exercise; pred: predicted; RR: relative risk; SaO<sub>2</sub>: oxygen saturation of arterial blood; sens.: sensitivity; spec.: specificity; SpO<sub>2</sub>: oxygen saturation measured by pulse oximetry; SSc: systemic sclerosis; VE/VCO<sub>2</sub>: ventilatory equivalent for carbon dioxide; VE/VO<sub>2</sub>: ventilatory equivalent for oxygen; VO<sub>2</sub>: oxygen uptake.

#### *Diffusion limitation or exercise-induced hypoxaemia*

Exercise-induced hypoxaemia was reported as a potential prognostic factor for survival in IPF [21, 23]. PaO<sub>2</sub> at the end of maximal exercise was the only CPET-derived parameter included in the comprehensive clinical–radiological–physiological multimodal score predicting survival in IPF, and when weighted, accounted for as much as 10.5% of the maximum score in the final model [23].

In mixed populations of ILD patients with advanced disease and referred for lung transplantation [7, 19], desaturation during CPET was reported to be predictive of lung transplantation or death.

In two studies examining longitudinal outcomes in sarcoidosis, the alveolar–arterial oxygen pressure gradient during exercise  $P(A-a)O_2$  (a measure of arterial desaturation during exercise) was independently associated with both the need for prolonged (>1 year) immunosuppressive therapy [25] and decline in pulmonary function at 5 years [26].

Finally, in the single study of SSc-ILD [27], akin to studies of sarcoidosis and IPF, diffusion limitation, measured in this study as the change in peripheral oxygenation ( $SpO_2$ ) during CPET, correlated with survival.

### Study confounders

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 (fig 2). The data used in the majority of studies was obtained from existing databases and/or case note review (85%,  $n=11$ ) and as such, the contribution of potential important confounders such as comorbid disease [18, 21, 22, 24, 26, 27], body mass index [19–21, 24, 26, 28] and smoking status [18, 19, 22, 26] was not recorded. Baseline “disease severity” was only specifically addressed as a potential confounder by one study [20]. The use of variable levels of supplemental oxygen (or uncertain inspired oxygen concentrations) in some of the reviewed studies [7, 19] is also a major limitation that potentially impacts on the accuracy of peak  $VO_2$ .

As discussed previously, studies reporting outcomes in subjects referred for transplantation reduces the generalisability of the study findings [7, 20], selecting cohorts of more advanced ILD patients. Other studies focused on healthier populations of ILD patients (*e.g.* not requiring supplemental oxygen during CPET), and this unsurprisingly resulted in lower mortality rates ( $n<10$ ) [24, 28, 29].

Multiple logistic regression (MLR) was the dominant statistical methodology used to determine the relationship between CPET parameters and clinical outcomes in ILD. Whilst this approach adjusts for the effects of known confounders, most of the study sample sizes were smaller than the proposed minimum requirement for MLR analysis [37]. Only one study reported an *a priori* power calculation to influence sample size [29]; others were underpowered to detect the outcomes proposed.

Stepwise multiple regression was used by some studies to determine the optimal model parameters to predict increased mortality [23, 28]. This approach uses parameter inference, which may lead to over-fitting of some parameters or exclusion of confounders that do not reach statistical significance [38]. Furthermore, the number of parameters or order entry (or deletion) can also affect the selected model [39] and affect the likelihood of type I error [38]. Only one study specifically attempted to reduce multicollinearity [23], which if overlooked can increase the risk of type II error [40].

### Discussion

Clinicians would benefit from reliable prognostic markers for patients with ILD to enable timelier referral for transplantation, improved monitoring of existing therapies, and to determine the efficacy of novel treatments in clinical trials [12, 41].

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. Thirteen studies were identified; survival was the principal clinical outcome measured. The utilisation of numerous methodologies, CPET parameters and timing of mortality evaluation prevented the determination of definitive CPET thresholds for predicting outcomes in ILD. Due to the clinical diversity of the studies and moderate risk of bias in all studies in at least one domain of the QUIPS tool, meta-analysis was not possible. It was also felt that meta-analysis might overstate the findings of these small-scale, poorly matched studies.

There were conflicting results with regards to the prognostic role of maximal oxygen consumption in predicting survival, which may in part be attributable to the heterogeneity of the studies concerned. Whilst reductions in ventilatory efficiency have been reported to predict both the presence of PH [42] and development of interceding PH in IPF cohorts [18], an independent prognostic value in IPF patients was not determined. The magnitude of hyperventilation at ventilatory threshold does, however, warrant further exploration as a prognostic factor in ILD, particularly as a marker of concurrent cardiopulmonary vascular impairment. Exercise-induced hypoxaemia was another potential prognostic outcome reported in several studies. A study directly comparing the longitudinal prognostic value of CPET with alternative forms of exercise testing, such as 6-minute walk testing, could therefore be justified.

Issues around study design (relating primarily to the inherent problems of retrospective studies, patient selection and presentation of numerous CPET parameters), insufficient adjustment for confounding

variables and inadequate statistical analyses limits the strength of conclusions that can be drawn from the studies undertaken to date. Whilst the associations presented shed important light on the potential role of CPET in disease prognostication in ILD, there is currently insufficient evidence to support its use in facilitating “real-world” clinical decisions and larger prospective studies are required. In planning future clinical studies, rigorously phenotyped patient cohorts, characterised using standardised definitions and with external validation or multicentre cohorts, will be imperative to try to overcome some of the challenges encountered by studying heterogeneous ILD populations.

Several practical challenges of CPET, including lack of measurement standardisation, non-uniform parameter availability from different instrument manufacturers, provision of adequate training of personnel, availability of equipment in secondary care, establishment of optimal exercise duration and ramping protocol, alongside individual patient safety considerations, such as desaturation to prohibitive levels in advanced ILD, will all need to be addressed prior to its consideration in clinical practice in ILD patient populations. The absence of sufficient longitudinal data to identify a minimally clinically important change in CPET values in ILDs is a further obstacle that will also need to be overcome [12, 43].

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 probably 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 RCTs 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 was part of the *a priori* plan [44]. 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. It is important that relevant study confounders are taken into consideration in future studies examining the prognostic value of CPET in ILD to establish whether CPET provides additional prognostic value beyond more easily obtainable clinical and physiological outcomes.

### Conclusions: take-home message

CPET may have a role as a prognostic factor in ILD but the quality of existing studies and lack of MCID values in ILDs limits the conclusions that can be drawn at present. Large carefully designed prospective studies are needed to establish the role of CPET in the longitudinal assessment of ILD in the future.

Acknowledgements: Independent statistical advice was sought from Paul White (statistician at the University of the West of England, Bristol, UK) to confirm that a meta-analysis was not possible.

Author contributions: S.L. Barratt is the guarantor of the content of the manuscript, including data and analysis. C. Sharp and R. Davis undertook the initial literature review and data extraction. This was then verified by S.L. Barrett and J.D. Pauling who prepared the manuscript. R. Davis and C. Sharp verified the manuscript.

Conflict of interest: S.L. Barratt reports personal fees for an advisory board and financial support to attend an educational conference from Boehringer Ingelheim outside the submitted work. R. Davis is an employee of Boehringer Ingelheim. Boehringer Ingelheim did not have any involvement in the research or the preparation of this manuscript. C. Sharp has nothing to disclose. J.D. Pauling reports personal fees from Boehringer Ingelheim; grants, personal fees and non-financial support for attendance at educational meetings from Actelion Pharmaceuticals; and personal fees from Sojournix Pharma, all outside the submitted work.

### References

- 1 American Thoracic Society. Idiopathic Pulmonary Fibrosis: Diagnosis and Treatment. International Consensus Statement. American Thoracic Society (ATS), and the European Respiratory Society (ERS). *Am J Respir Crit Care Med* 2000; 161: 646–664.
- 2 Bellaye PS, Kolb M. Why do patients get idiopathic pulmonary fibrosis? Current concepts in the pathogenesis of pulmonary fibrosis. *BMC Med* 2015; 13: 176.
- 3 Schroedel CJ, Yount SE, Szmuiłowicz E, *et al.* A qualitative study of unmet healthcare needs in chronic obstructive pulmonary disease. A potential role for specialist palliative care? *Ann Am Thorac Soc* 2014; 11: 1433–1438.
- 4 Mura M, Porretta MA, Bargagli E, *et al.* Predicting survival in newly diagnosed idiopathic pulmonary fibrosis: a 3-year prospective study. *Eur Respir J* 2012; 40: 101–109.
- 5 Albera C. Challenges in idiopathic pulmonary fibrosis trials: the point on end-points. *Eur Respir Rev* 2011; 20: 195–200.
- 6 Gordon JK, Domsic RT. Clinical trial design issues in systemic sclerosis: an update. *Curr Rheumatol Rep* 2016; 18: 38.
- 7 Layton AM, Armstrong HF, Kim HP, *et al.* Cardiopulmonary exercise factors predict survival in patients with advanced interstitial lung disease referred for lung transplantation. *Respir Med* 2017; 126: 59–67.
- 8 Sue DY, Wasserman K. Impact of integrative cardiopulmonary exercise testing on clinical decision making. *Chest* 1991; 99: 981–992.
- 9 Palange P, Ward SA, Carlsen KH, *et al.* Recommendations on the use of exercise testing in clinical practice. *Eur Respir J* 2007; 29: 185–209.



- 10 Ferrazza AM, Martolini D, Valli G, *et al.* Cardiopulmonary exercise testing in the functional and prognostic evaluation of patients with pulmonary diseases. *Respiration* 2009; 77: 3–17.
- 11 Arena R, Sietsema KE. Cardiopulmonary exercise testing in the clinical evaluation of patients with heart and lung disease. *Circulation* 2011; 123: 668–680.
- 12 Bonini M, Fiorenzano G. Exertional dyspnoea in interstitial lung diseases: the clinical utility of cardiopulmonary exercise testing. *Eur Respir Rev* 2017; 26: 160099.
- 13 American Thoracic Society, American College of Chest Physicians. ATS/ACCP Statement on Cardiopulmonary Exercise Testing. *Am J Respir Crit Care Med* 2003; 167: 211–277.
- 14 Shamseer L, Moher D, Clarke M, *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015; 350: g7647.
- 15 Altman DG, Deeks JJ, Clarke M, *et al.* The quality of systematic reviews. High quality reporting of both randomised trials and systematic reviews should be priority. *BMJ* 2000; 321: 297–299.
- 16 Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968; 70: 213–220.
- 17 Huguet A, Hayden JA, Stinson J, *et al.* Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. *Syst Rev* 2013; 2: 71.
- 18 Gläser S, Obst A, Koch B, *et al.* Pulmonary hypertension in patients with idiopathic pulmonary fibrosis - the predictive value of exercise capacity and gas exchange efficiency. *PLoS One* 2013; 8: e65643.
- 19 Kawut SM, O'Shea MK, Bartels MN, *et al.* Exercise testing determines survival in patients with diffuse parenchymal lung disease evaluated for lung transplantation. *Respir Med* 2005; 99: 1431–1439.
- 20 van der Plas MN, van Kan C, Blumenthal J, *et al.* Pulmonary vascular limitation to exercise and survival in idiopathic pulmonary fibrosis. *Respirology* 2014; 19: 269–275.
- 21 Miki K, Maekura R, Hiraga T, *et al.* Impairments and prognostic factors for survival in patients with idiopathic pulmonary fibrosis. *Respir Med* 2003; 97: 482–490.
- 22 Wallaert B, Guetta A, Wemeau-Stervinou L, *et al.* [Prognostic value of clinical exercise testing in idiopathic pulmonary fibrosis]. *Rev Mal Respir* 2011; 28: 290–296.
- 23 King TE, Tooze JA, Schwarz MI, *et al.* Predicting survival in idiopathic pulmonary fibrosis: scoring system and survival model. *Am J Respir Crit Care Med* 2001; 164: 1171–1181.
- 24 Fell CD, Liu LX, Motika C, *et al.* The prognostic value of cardiopulmonary exercise testing in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2009; 179: 402–407.
- 25 Kollert F, Geck B, Suchy R, *et al.* The impact of gas exchange measurement during exercise in pulmonary sarcoidosis. *Respir Med* 2011; 105: 122–129.
- 26 Lopes AJ, Menezes SL, Dias CM, *et al.* Cardiopulmonary exercise testing variables as predictors of long-term outcome in thoracic sarcoidosis. *Braz J Med Biol Res* 2012; 45: 256–263.
- 27 Swigris JJ, Zhou X, Wamboldt FS, *et al.* Exercise peripheral oxygen saturation (SpO<sub>2</sub>) accurately reflects arterial oxygen saturation (SaO<sub>2</sub>) and predicts mortality in systemic sclerosis. *Thorax* 2009; 64: 626–630.
- 28 Triantafyllidou C, Manali E, Lyberopoulos P, *et al.* The role of cardiopulmonary exercise test in IPF prognosis. *Pulm Med* 2013; 2013: 514817.
- 29 Vainshelboim B, Oliveira J, Fox BD, *et al.* The prognostic role of ventilatory inefficiency and exercise capacity in idiopathic pulmonary fibrosis. *Respir Care* 2016; 61: 1100–1109.
- 30 American Thoracic Society, European Respiratory Society. American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. *Am J Respir Crit Care Med* 2002; 165: 277–304.
- 31 Raghu G, Collard HR, Egan JJ, *et al.* An Official ATS/ERS/JRS/ALAT Statement: Idiopathic Pulmonary Fibrosis: Evidence-based Guidelines for Diagnosis and Management. *Am J Respir Crit Care Med* 2011; 183: 788–824.
- 32 Scadding JG. Prognosis of intrathoracic sarcoidosis in England. A review of 136 cases after five years' observation. *Br Med J* 1961; 2: 1165–1172.
- 33 Massi AT. Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). *Arthritis Rheum* 1980; 23: 581–590.
- 34 Poormoghim H, Lucas M, Fertig N, *et al.* Systemic sclerosis sine scleroderma: demographic, clinical, and serologic features and survival in forty-eight patients. *Arthritis Rheum* 2000; 43: 444–451.
- 35 Gibbons RJ, Balady GJ, Beasley JW, *et al.* ACC/AHA Guidelines for Exercise Testing. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Exercise Testing). *J Am Coll Cardiol* 1997; 30: 260–311.
- 36 McArdle WD, Katch FI, Pechar GS. Comparison of continuous and discontinuous treadmill and bicycle tests for max Vo<sub>2</sub>. *Med Sci Sports* 1973; 5: 156–160.
- 37 Bujang MA, Sa'at N, Sidik TMITAB. Determination of minimum sample size requirement for multiple linear regression and analysis of covariance based on experimental and non-experimental studies. *Epidemiol Biostat Public Health* 2017; 14: e12117-1-9.
- 38 Whittingham MJ, Stephens PA, Bradbury RB, *et al.* Why do we still use stepwise modelling in ecology and behaviour? *J Anim Ecol* 2006; 75: 1182–1189.
- 39 Derksen S, Keselman HJ. Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. *Br J Math Stat Psychol* 1992; 45: 265–282.
- 40 Chatterjee S, Hadi AS, Bertram P. Regression Analysis by Example (Series in Probability and Statistics). 3rd Edn Hoboken, Wiley, 2000.
- 41 Nathan SD, Meyer KC. IPF clinical trial design and endpoints. *Curr Opin Pulm Med* 2014; 20: 463–471.
- 42 Boutou AK, Pitsiou GG, Trigonis I, *et al.* Exercise capacity in idiopathic pulmonary fibrosis: the effect of pulmonary hypertension. *Respirology* 2011; 16: 451–458.
- 43 Collard HR, King TE, Bartelson BB, *et al.* Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2003; 168: 538–542.
- 44 Hemingway H, Riley RD, Altman DG. Ten steps towards improving prognosis research. *BMJ* 2009; 339: b4184.