**Supplementary**

**Appendix 1**

Supplemental methods

Table S1 Missing data in the analysis

Table S2 Specific diagnosis of CTD-ILD patients

Table S3 Baseline demographics and clinical characteristics in HCs and CTD-ILD

Table S4 Baseline serum tumor markers of HCs and CTD-ILD patients

Table S5 Correlation coefficient values of the tumor markers and lung functions

Table S6 Baseline demographics and clinical characteristics in specific diagnosis

Table S7 The association of CEA and CA125 with PPF in specific diagnosis

Table S8 Immunosuppressants treatment during follow-up

Table S9 Baseline treatment from different groups

Table S10 The associations of CEA and CA125 with PPF excluded acute worsening

Table S11 The associations of CEA and CA125 with PPF (emphysema excluded)

Figure S1 Kaplan-Meier curves for all-cause mortality and cumulative hazard for AE.

Figure S2 Subgroup analysis of the proportional effect of elevated CEA concentrations on PPF.

Figure S3 Subgroup analysis of the proportional effect of elevated CA125 concentrations in PPF.

Figure S4 Patterns of pulmonary fibrosis progression.

Figure S5 Changes in of CEA and CA125 over time (n=90) stratified by PPF.

Figure S6 Graphical abstract.

**Appendix 1**

**#Supplemental methods**

##Sample size evaluation

When determining the sample size for this study, two key factors were considered. Firstly, as a cohort study, the power for the primary endpoint (the occurrence of progressive pulmonary fibrosis [PPF]) is calculated using PASS software based on a two-sided *t-*test with a significance level of 5%. With a sample size of 92 patients in low exposure group and 92 patients in high exposure group, the trial will have more than 90% power to detect a difference between liraglutide and liraglutide placebo in the proportion of patients with PPF, given that the assumed probabilities to experience progression is 20% for low exposure group and 40% for high exposure group. As for the regression test for the primary outcome of PPF, the sample size followed the EPV (events per variable) principle. This principle suggests a minimum of 5 events per predictor variable, ideally 10, to ensure an adequate sample size for the regression analysis.

##Missing data

Multiple imputation by chained equations was used to impute follow-up data for those patients who missed their follow-up visits (there were no missing values in the covariates measured at baseline). Variables that were related to the outcome, and/or related to loss-to-follow-up were included in the imputation model, including age, sex, smoking history, baseline FVC% pred., baseline DLCO% pred., pulmonary hypertension, emphysema, honeycombing, bronchiectasis and antifibrotic treatment. The final data set was based on the median (for continuous variables) or the mode (for categorical variables) of the 5 imputed data sets. Table S1 included the level of missing data in each of the covariates included in the analysis. The missing covariates all had very low levels of missingness.

**Schedule of assessments**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variables | Study period | | | | | | |
| Enrolment | Study period for the PPF patients | | | | | Follow-up of survival and AE/end of studya |
| Time point (month) | -1 | T0 (baseline) | T1 (3 months) | T2 (6 months) | T3 (9 months) | T4 (12 months) | Every 6 months |
| Visit window (week) |  |  | ±2 | ±2 | ±2 | ±2 | ±2 |
| Enrolment: |  |  |  |  |  |  |  |
| Eligibility screen | X |  |  |  |  |  |  |
| Informed consent | X |  |  |  |  |  |  |
| Demographics | X |  |  |  |  |  |  |
| Medical history | X |  |  |  |  |  |  |
| Assessment: |  |  |  |  |  |  |  |
| Sociodemographic characteristics |  | X |  |  |  |  |  |
| Physical examination |  | X | X | X | X | X |  |
| ANA |  | X |  |  |  |  |  |
| Autoantibodies |  | X |  |  |  |  |  |
| Tumor markers |  | X |  |  |  |  |  |
| Self-reported symptoms |  | X | X | X | X | X |  |
| HRCT test |  | Xb | Xc | X | Xc | X |  |
| Lung function tests |  | X | Xc | X | Xc | X |  |
| Primary outcome |  |  |  |  |  |  |  |
| Pulmonary fibrosis progression |  | X | X | X | X | X |  |
| Secondary outcomes |  |  |  |  |  |  |  |
| Acute exacerbation |  | X | X | X | X | X | X |
| All-cause mortality |  | X | X | X | X | X | X |

a, after the first year of follow-up, survival and acute exacerbation information will be collected by telephone, medical records, and/or clinical visits approximately every 6 months (±2 weeks) until death, loss of follow-up, or end of the study. b, the results of a chest HRCT performed within three months prior to the enrollment visit (if available) are acceptable, provided there is no reason to suspect any clinical changes, per clinician discretion. HRCT must be repeated if the patient had CT in over 3 months prior to enrollment. c, the assessment of lung function and HRCT at this visit is not obligatory unless there is suspect deterioration in clinical manifestations (determined by the clinician). AE, acute exacerbation; ANA, antinuclear antibodies; HRCT, high resolution computed tomography; PPF, progressive pulmonary fibrosis.

**#Results**

Table S1 Missing data in the analysis

|  |  |  |
| --- | --- | --- |
| Variables | N=224 | Percentage |
| Measures |  |  |
| FVC pred% at 6 months | 10 | 4.46% |
| DLCO pred% at 6 months | 10 | 4.46% |
| FVC pred% at 12 months | 14 | 6.25% |
| DLCO pred% at 12 months | 14 | 6.25% |
| Outcomes |  |  |
| Progressive pulmonary fibrosis | 5 | 2.23% |
| Acute exacerbation | 11 | 4.91% |

FVC, forced vital capacity; DLCO, diffusion capacity of the lung for carbon monoxide.

Table S2 Specific diagnosis of CTD-ILD patients

|  |  |  |  |
| --- | --- | --- | --- |
| Entities | All, N=224 | Non-PPF, N=161 | PPF, N=63 |
| RA | 19 (8.5) | 15 (9.3) | 4 (6.3) |
| IIM | 82 (36.6) | 67 (41.6) | 15 (23.8) |
| pSS | 46 (20.5) | 28 (17.4) | 18 (28.6) |
| SLE | 2 (0.9) | 1 (0.6) | 1 (1.6) |
| SSc | 13 (5.8) | 10 (6.2) | 3 (4.8) |
| OCTD | 24 (10.7) | 12 (7.5) | 12 (19.0) |
| UCTD | 38 (17.0) | 27 (16.8) | 11 (17.5) |

Values were given as the n (%). CTD, connective tissue disease; IIM, idiopathic inflammatory myopathy; ILD, interstitial lung disease; MCTD, mixed connective tissue disease; OCTD, overlap connective tissue disorders; PPF, progressive pulmonary fibrosis; SLE, systemic lupus erythematosus; pSS, primary Sjögren syndrome; RA, rheumatoid arthritis; SSc, systemic sclerosis; UCTD, undifferentiated connective tissue disease.

Table S3 Baseline demographics and clinical characteristics in HCs and CTD-ILD

|  |  |  |  |
| --- | --- | --- | --- |
| Variables | HCs | CTD-ILD | P value |
| N | 63 | 224 |  |
| Male, n% | 28(44.4) | 80 (35.7) | 0.21 |
| Age, years | 62.6 (13.7) | 60.1(11.6) | 0.41 |
| Smoking status |  |  |  |
| Ever-smokers, n (%) | 10 (15.9) | 58 (25.9) | 0.10 |
| Pack-years | 15 (10-60) | 32.29 (14.8-40) | 0.50 |

Data are presented as mean (standard deviation), median (first and third quartiles), number (%), or number. HCs, healthy controls; CTD, connective tissue disease; ILD, interstitial lung disease.

Table S4 Baseline serum tumor markers of HCs and CTD-ILD patients

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variables | HCs | Non-PPF | PPF | Pvalue |
| N | 63 | 161 | 63 |  |
| SCC ng/mL | 0.7 (0.6-1.03) | 0.8 (0.5-1) | 0.9 (0.5-1.1) | 0.06 |
| >1.5 ng/mL, n (%) | 4 (6.3) | 15 (9.3) | 8 (12.7) | 0.47 |
| CEA, ng/mL | 0.94 (0.41-1.52) | 1.48 (0.86-2.47)# | 2.09 (1.36-3.9)#\* | <0.001 |
| >5 ng/mL, n (%) | 0 | 11 (6.8) | 9 (14.2) | 0.007 |
| CA19-9 U/mL | 9.29 (6.83-14.56) | 14.12 (8.98-30.74)# | 13.64 (8.7-42.3)# | <0.001 |
| >37 U/mL, n (%) | 0 | 34 (21.1) | 17 (26.9) | <0.001 |
| SF μg/L | 155.3 (94.38-218.65) | 369.6 (117.95-440.32)# | 227.8 (126.8-440.3)# | <0.001 |
| >291 μg/L, n (%) | 7 (11.1) | 90 (55.9) | 25 (39.7) | <0.001 |
| AFP ng/mL | 2.60 (1.78-3.50) | 2.48 (1.15-3) | 2.45 (1.2-3.5) | 0.04 |
| >8.1 ng/mL, n (%) | 1 (1.6) | 2 (1.2) | 3 (4.7) | 0.24 |
| CYFRA21-1 ng/mL | 1.76 (1.28-2.30) | 4.13 (2.53-5.74)# | 4.26 (2.5-5.7)# | <0.001 |
| >2.08 ng/mL, n (%) | 22 (34.9) | 133 (82.6) | 49 (77.8) | <0.001 |
| CA125, U/mL | 7.05 (5.18-10.48) | 13.5 (8.55-22.80)# | 17.4 (10-26.5)#\* | <0.001 |
| >30.2 U/mL, n (%) | 1 (1.59) | 17 (10.56) | 12 (19.05) | 0.006 |
| NSE ng/mL | 14.74 (11.46-17.21) | 16.44 (13.32-18.83)# | 15.66 (12.9-18.7) | 0.006 |
| >16.3 ng/mL, n (%) | 19 (30.2) | 82 (50.9) | 28 (44.4) | 0.02 |
| CA724 U/mL | 1.60 (0.97-2.69) | 3.63 (1.44-7.11)# | 2.95 (1.75-7.1) # | 0.003 |
| >8.2 U/mL, n (%) | 5 (7.9) | 21 (13) | 7 (11.1) | 0.56 |

#, Pvalue <0.05 compared with HCs; \*, P value <0.05 compared with non-PPF. Values were given as the median (first and third quartiles) or n (%). The recommended normal ranges used were: CEA ≤5 ng/mL; SCC ≤1.5 ng/mL; NSE ≤16.3 ng/mL; CYFRA21-1 ≤2.08 ng/mL; SF ≤291 μg/L; AFP ≤8.1 ng/mL; CA125 ≤30.2 U/mL; CA19-9 ≤37 U/mL and CA724 ≤8.2 U/mL. AFP, alpha fetoprotein; CA125, carbohydrate antigen 125; CA199, carbohydrate antigen 199; CA724, carbohydrate antigen 724; CEA, carcinoembryonic antigen; CTD, connective tissue disease; CYFRA21-1, cytokeratin fraction 21-1; HCs, healthy controls; ILD, interstitial lung disease; PPF, progressive pulmonary fibrosis; NSE, neuron-specific enolase; SCC, squamous cell carcinoma antigen; SF, serum ferritin.

Table S5 Correlation coefficient values of the tumor markers and lung functions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tumor markers | FVC% pred | P value | DLCO% pred | P value |
| SCC | 0.058 | 0.39 | –0.095 | 0.16 |
| CEA | –0.044 | 0.52 | –0.213 | 0.001 |
| CA19-9 | 0.002 | 0.98 | –0.139 | 0.04 |
| SF | 0.051 | 0.45 | –0.023 | 0.74 |
| AFP | 0.055 | 0.41 | –0.119 | 0.08 |
| CYFRA21-1 | –0.069 | 0.30 | –0.144 | 0.03 |
| CA125 | –0.159 | 0.02 | –0.282 | <0.001 |
| NSE | –0.107 | 0.11 | –0.187 | 0.005 |
| CA724 | 0.057 | 0.40 | –0.097 | 0.15 |

AFP, alpha fetoprotein; CA125, carbohydrate antigen 125; CA199, carbohydrate antigen 199; CA724, carbohydrate antigen 724; CEA, carcinoembryonic antigen; CYFRA21-1, cytokeratin fraction 21-1; DLCO, diffusion capacity of the lung for carbon monoxide; FVC, forced vital capacity; NSE, neuron-specific enolase; SCC, squamous cell carcinoma antigen; SF, serum ferritin.

Table S6 Baseline demographics and clinical characteristics in specific diagnosis

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variables | RA | pSS | IIM | OCTD | UCTD | Others\* |
| N | 19 | 46 | 82 | 24 | 38 | 15 |
| Male, n (%) | 7 (36.8) | 12 (26.1) | 38 (46.3) | 3 (12.5) | 14 (36.8) | 6 (40.0) |
| Age, years | 66.3±7.9 | 64.8±10.9 | 55.6±10.1 | 60.5±12.5 | 61.8±12.9 | 57.5±12.1 |
| BMI, kg/m2 | 26.4±4.4 | 25.4±3.34 | 26.1±3.4 | 24.4±3.7 | 25.4±3.6 | 23.7±2.3 |
| Smoking status |  |  |  |  |  |  |
| Ever-smokers, n (%) | 5 (26.3) | 13 (28.3) | 25 (30.5) | 4 (16.7) | 8 (21.1) | 3 (20.0) |
| Pack-years | 30.0 (20.0, 40.0) | 40.0 (20.0, 50.0) | 23.88 (10.0, 40.0) | 25.0 (18.5, 32.5) | 27.5 (13.1, 41.3) | 52.5 (41.3, 63.8) |
| Clubbing of fingers | 2 (10.5) | 14 (30.4) | 17 (20.7) | 7 (29.2) | 16 (42.1) | 5 (33.3) |
| ANA, n (%) | 9 (47.4) | 32 (69.6) | 39 (47.6) | 19 (79.2) | 21 (55.3) | 11 (73.3) |
| Lung function |  |  |  |  |  |  |
| FVC% pred. | 95.9 (87.0, 115.3) | 94.1 (80.7, 107.1) | 82.0 (68.4, 91.6) | 85.1 (74.1, 106.6) | 86.7 (67.2, 101.0) | 78.1 (60.3, 96.5) |
| DLCO% pred. | 63.5 (48.2, 74.4) | 57.0 (46.8, 68.8) | 58.1 (49.2, 70.8) | 56.6 (45.0, 60.6) | 53.1 (42.0, 64.4) | 54.4 (46.3, 68.8) |
| HRCT, n (%) |  |  |  |  |  |  |
| Honeycombing | 3 (15.8) | 16 (34.8) | 15 (18.3) | 6 (25.0) | 9 (23.7) | 2 (13.3) |
| Emphysema | 1 (5.3) | 5 (10.9) | 11 (13.4) | 1 (4.2) | 4 (10.5) | 0 (0.0) |
| Bronchiectasis | 6 (31.6) | 28 (60.9) | 41 (50.0) | 11 (45.8) | 14 (36.8) | 8 (53.3) |
| Comorbidities, n (%) |  |  |  |  |  |  |
| Gastroesophageal reflux | 5 (26.3) | 11 (23.9) | 17 (20.7) | 4 (16.7) | 10 (26.3) | 6 (40.0) |
| Pulmonary hypertension | 3 (15.8) | 1 (2.2) | 9 (11.0) | 8 (33.3) | 5 (13.2) | 4 (26.7) |
| Diabetes mellitus | 8 (42.1) | 11 (23.9) | 14 (17.1) | 8 (33.3) | 12 (31.6) | 2 (13.3) |
| Coronary heart disease | 1 (5.3) | 8 (17.4) | 5 (6.1) | 2 (8.3) | 8 (21.1) | 3 (20.0) |
| Osteoporosis | 9 (47.4) | 14 (30.4) | 31 (37.8) | 11 (45.8) | 7 (18.4) | 5 (33.3) |
| Treatment, n (%) |  |  |  |  |  |  |
| Glucocorticoids | 17 (89.5) | 39 (84.8) | 78 (95.1) | 22 (91.7) | 33 (86.8) | 14 (93.3) |
| Immunosuppressants | 17 (89.5) | 36 (78.3) | 66 (80.5) | 20 (83.3) | 27 (71.1) | 12 (80.0) |
| Anti-fibrotic treatment | 2 (10.5) | 11 (23.9) | 13 (15.9) | 0 (0.0) | 12 (31.6) | 3 (20.0) |
| Fibrosis progression, n (%) | 4 (21.1) | 18 (39.1) | 15 (18.3) | 12 (50.0) | 11 (29.0) | 3 (20.0) |
| Death, n (%) | 2 (28.6) | 6 (17.1) | 13 (17.3) | 4 (21.1) | 7 (22.6) | 2 (20.0) |
| Acute exacerbation, n (%) | 4 (21.1) | 7 (16.3) | 21 (26.6) | 11 (45.8) | 12 (31.6) | 3 (20.0) |
| Follow-up duration, month | 52.0 (36.5, 56.0) | 48.5 (33.0, 52.0) | 37.0 (23.5, 48.8) | 38.0 (29.5, 55.3) | 36.5 (20.5, 44.8) | 42.0 (31.0, 55.5) |

\*, others: SLE and SSc. Data are presented as mean ± standard deviation, median (first and third quartiles), number (%), or number. ANA, antinuclear antibodies; BMI, body mass index, DLCO, diffusion capacity of the lung for carbon monoxide; FVC, forced vital capacity; HRCT, high resolution computed tomography; IIM, idiopathic inflammatory myopathy; MCTD, mixed connective tissue disease; OCTD, overlap connective tissue disorders; PPF, progressive pulmonary fibrosis; SLE, systemic lupus erythematosus; pSS, primary Sjögren syndrome; SSc, systemic sclerosis; RA, rheumatoid arthritis; UCTD, undifferentiated connective tissue disease.

Table S7 The association of CEA and CA125 with PPF in specific diagnosis

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variables | CEA | | |  | CA125 | | |
| OR | 95%CI | P value |  | OR | 95%CI | P value |
| RA | 10.59 | 0.41–274.46 | 0.16 |  | 1.21 | 0.43–3.41 | 0.72 |
| pSS | 2.10 | 0.84–5.23 | 0.11 |  | 1.09 | 0.51–2.34 | 0.83 |
| IIM | 1.32 | 1.04–2.74 | 0.045 |  | 1.07 | 0.60–1.92 | 0.83 |
| OCTD | 0.82 | 0.33–2.05 | 0.63 |  | 1.15 | 0.43–3.05 | 0.79 |
| UCTD | 4.74 | 1.45–15.52 | 0.01 |  | 2.08 | 0.76–5.66 | 0.13 |
| Others | 1.525 | 0.13–20.79 | 0.75 |  | 1.18 | 0.65–2.14 | 0.59 |

Model: crude continuous model. CA125, carbohydrate antigen 125; CEA, carcinoembryonic antigen; CI, confidence interval; OR, odds ratio; IIM, idiopathic inflammatory myopathy; MCTD, mixed connective tissue disease; OCTD, overlap connective tissue disorders; PPF, progressive pulmonary fibrosis; pSS, primary Sjögren syndrome; RA, rheumatoid arthritis; UCTD, undifferentiated connective tissue disease.

Table S8 Immunosuppressants treatment during follow-up

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Immunosuppressant | CEA | | | CA125 | | |
| ≤1.70 ng/mL | >1.70 ng/mL | P value | ≤14.95 U/mL | >14.95 U/mL | P value |
| Cyclophosphamide | 68 (60.7) | 69 (61.6) | 0.89 | 69 (61.6) | 68 (60.7) | 0.89 |
| Mycophenolate mofetil | 13 (11.6) | 20 (17.9) | 0.19 | 15 (13.4) | 18 (16.1) | 0.57 |
| Azathioprine | 5 (4.5) | 2 (1.8) | 0.25 | 3 (2.7) | 4 (3.6) | 0.70 |
| Tacrolimus | 17 (15.2) | 10 (8.9) | 0.15 | 14 (12.5) | 13 (11.6) | 0.84 |
| Exposure, years | 1 (0.75-1.67) | 1 (0.75-1.5) | 0.70 | 1 (1-2) | 1 (1-1.5) | 0.22 |

Data was presented as n (%) or median (first and third quartiles). CA125, carbohydrate antigen 125; CEA, carcinoembryonic antigen.

Table S9 Baseline treatment from different groups

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Baseline treatment | CEA | | | CA125 | | |
| ≤1.70 ng/mL | >1.70 ng/mL | Pvalue | ≤14.95 U/mL | >14.95 U/mL | P value |
| Glucocorticoids | 97 (86.6) | 93 (83) | 0.46 | 90 (80.4) | 100 (89.3) | 0.06 |
| Immunosuppressant | 90 (80.4) | 88 (78.6) | 0.74 | 88 (78.6) | 90 (80.4) | 0.74 |
| Cyclophosphamide | 49 (43.8) | 52 (46.4) | 0.68 | 49 (43.8) | 52 (46.4) | 0.69 |
| Mycophenolate mofetil | 6 (5.4) | 11 (9.8) | 0.21 | 9 (8.0) | 8 (7.1) | 0.80 |
| Azathioprine | 7 (6.3) | 3 (2.7) | 0.19 | 5 (4.5) | 5 (4.5) | >0.99 |
| Tacrolimus | 11 (9.8) | 7 (6.3) | 0.33 | 10 (8.9) | 8 (7.1) | 0.62 |

Data was presented as n (%). CA125, carbohydrate antigen 125; CEA, carcinoembryonic antigen.

Table S10 The associations of CEA and CA125 with PPF excluded acute worsening

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variables | Categorical | | | |  | Continuous | | | |
| Model 1 | | Model 2 | |  | Model 1 | | Model 2 | |
| OR (95% CI) | P value | OR (95% CI) | P value |  | OR (95% CI) | P value | OR (95% CI) | P value |
| CEA | 2.62 (1.35-5.09) | 0.005 | 2.46 (1.22-4.97) | 0.01 |  | 1.67 (1.09-2.55) | 0.02 | 1.67 (1.05-2.66) | 0.03 |
| CA125 | 1.87 (0.98-3.57) | 0.056 | 2.06 (1.01-4.19) | 0.047 |  | 1.32 (0.93-1.88) | 0.13 | 1.39 (0.94-2.06) | 0.10 |



Model 1: crude; Model 2: adjusted for DLCO% pred, age, pulmonary hypertension, honeycombing, emphysema. CA125, carbohydrate antigen 125; CEA, carcinoembryonic antigen; CI, confidence interval; OR, odds ratio; PPF, progressive pulmonary fibrosis.

Table S11 The associations of CEA and CA125 with PPF (emphysema excluded)

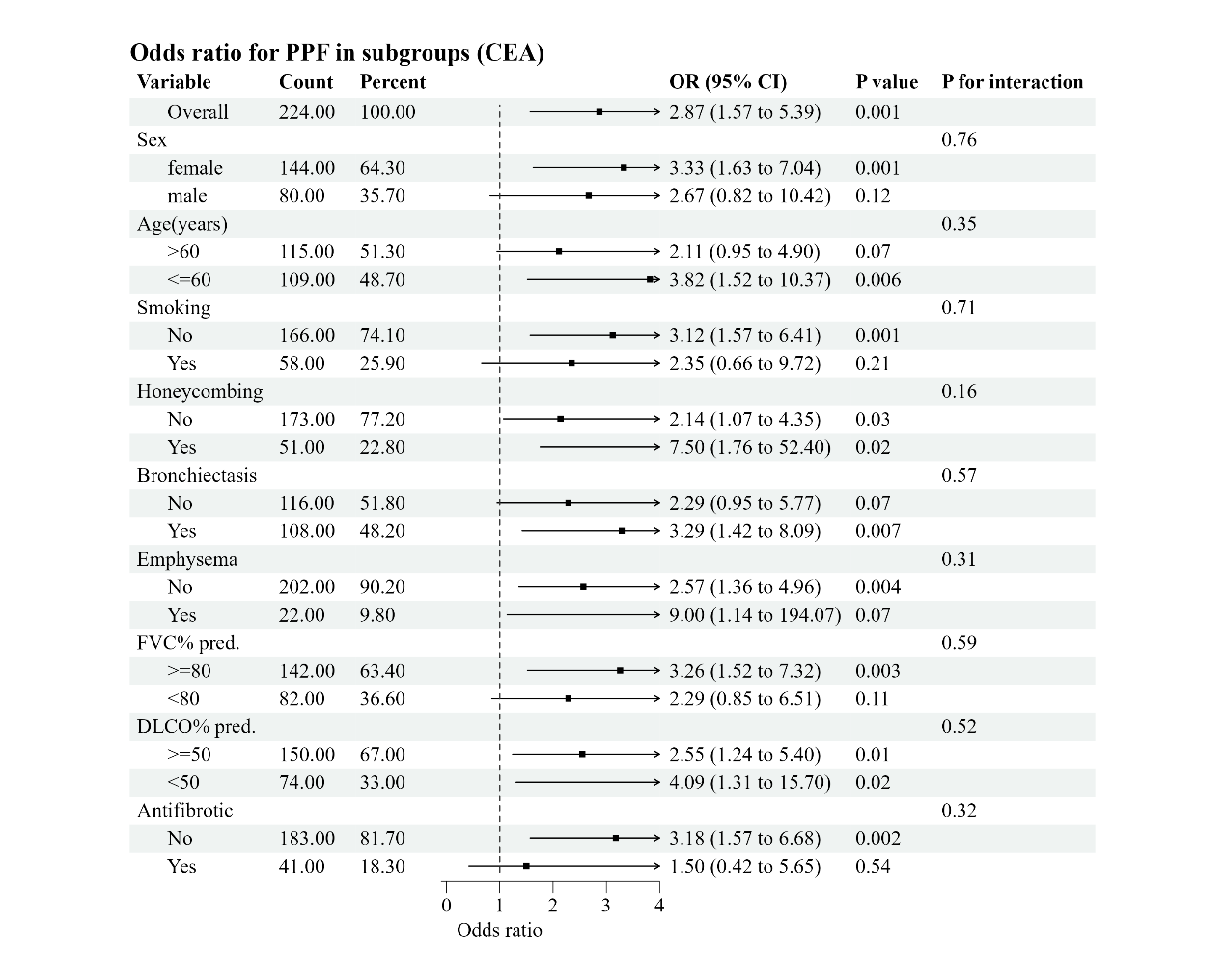


|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variables | Categorical | | | |  | Continuous | | | |
| Model 1 | | Model 2 | |  | Model 1 | | Model 2 | |
| OR (95% CI) | P value | OR (95% CI) | P value |  | OR (95% CI) | P value | OR (95% CI) | P value |
| CEA | 2.57 (1.34-4.88) | 0.004 | 2.51 (1.29-4.90) | 0.007 |  | 1.75 (1.14-2.69) | 0.01 | 1.76 (1.12-2.76) | 0.02 |
| CA125 | 1.89 (1.02-3.56) | 0.046 | 1.95 (0.99-3.83) | 0.053 |  | 1.40 (0.98-2.00) | 0.07 | 1.44 (0.98-2.12) | 0.06 |

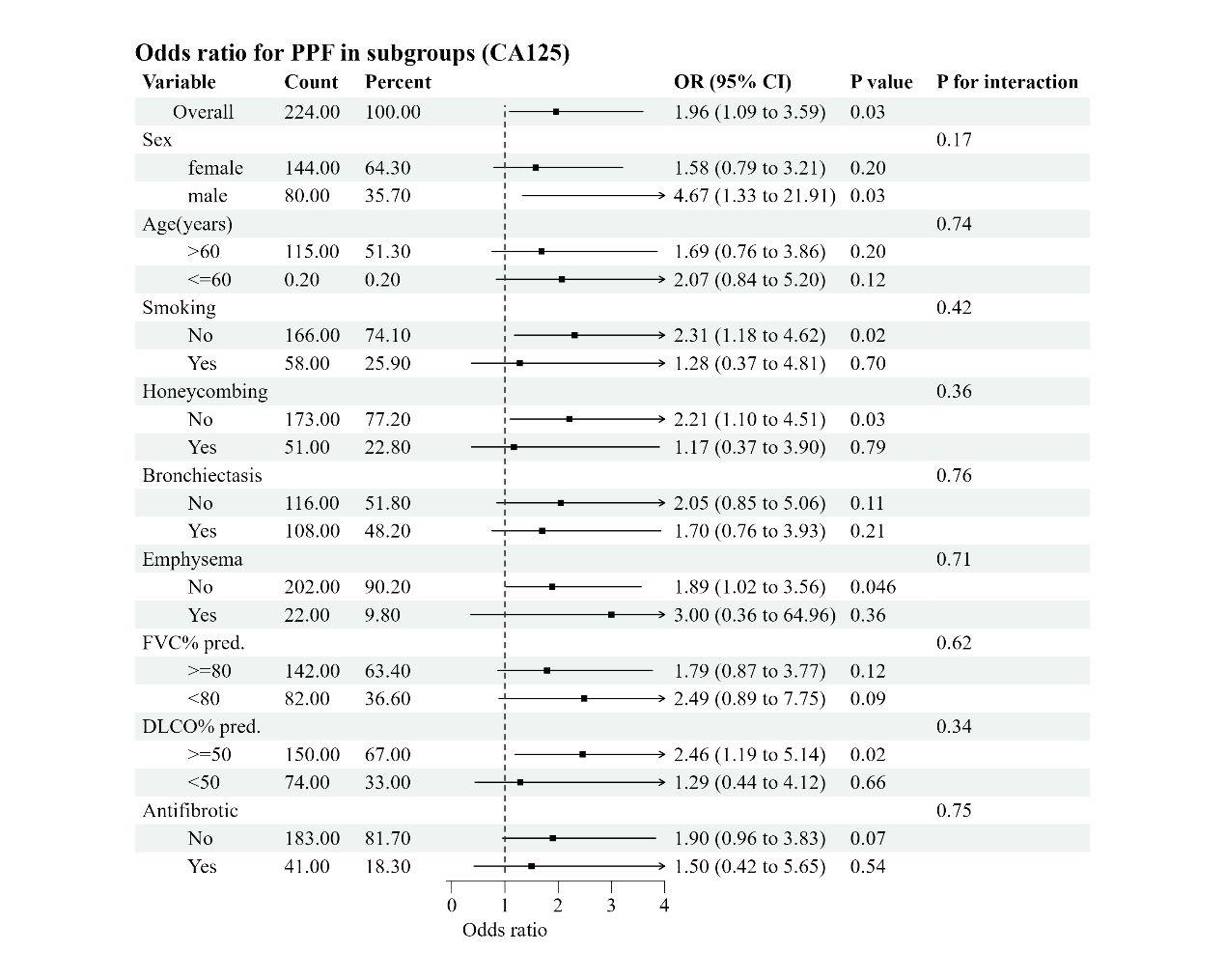
Model1: crude; Model2: adjusted for DLCO% pred, age, pulmonary hypertension, honeycombing. CA125, carbohydrate antigen 125; CEA, carcinoembryonic antigen; CI, confidence interval; OR, odds ratio; PPF, progressive pulmonary fibrosis.

****

**Figure S1** Kaplan-Meier curves for all-cause mortality and cumulative hazard for AE. (A) Kaplan-Meier curves showing patients in high CEA group had worse OS compared with those in low CEA group (P<0.001). (B) Kaplan-Meier curves showing OS did not differed in patients in high CA125 group and low CA125 group (P=0.31). (C) Cumulative hazard for AE differed significantly in patients in high CEA group and low CEA group (P<0.001). (D) Cumulative hazard for AE differed significantly in patients in high CA125 group and low CA125 group (P=0.005). AE, acute exacerbation; CA125, carbohydrate antigen 125; CEA, carcinoembryonic antigen; OS, overall survival.

****

**Figure S2** Subgroup analysis of the proportional effect of elevated CEA concentrations on PPF. CEA, carcinoembryonic antigen; CI, confidence interval; DLCO, diffusion capacity of the lung for carbon monoxide; FVC, forced vital capacity; OR, odds ratio; PPF, progressive pulmonary fibrosis.

****

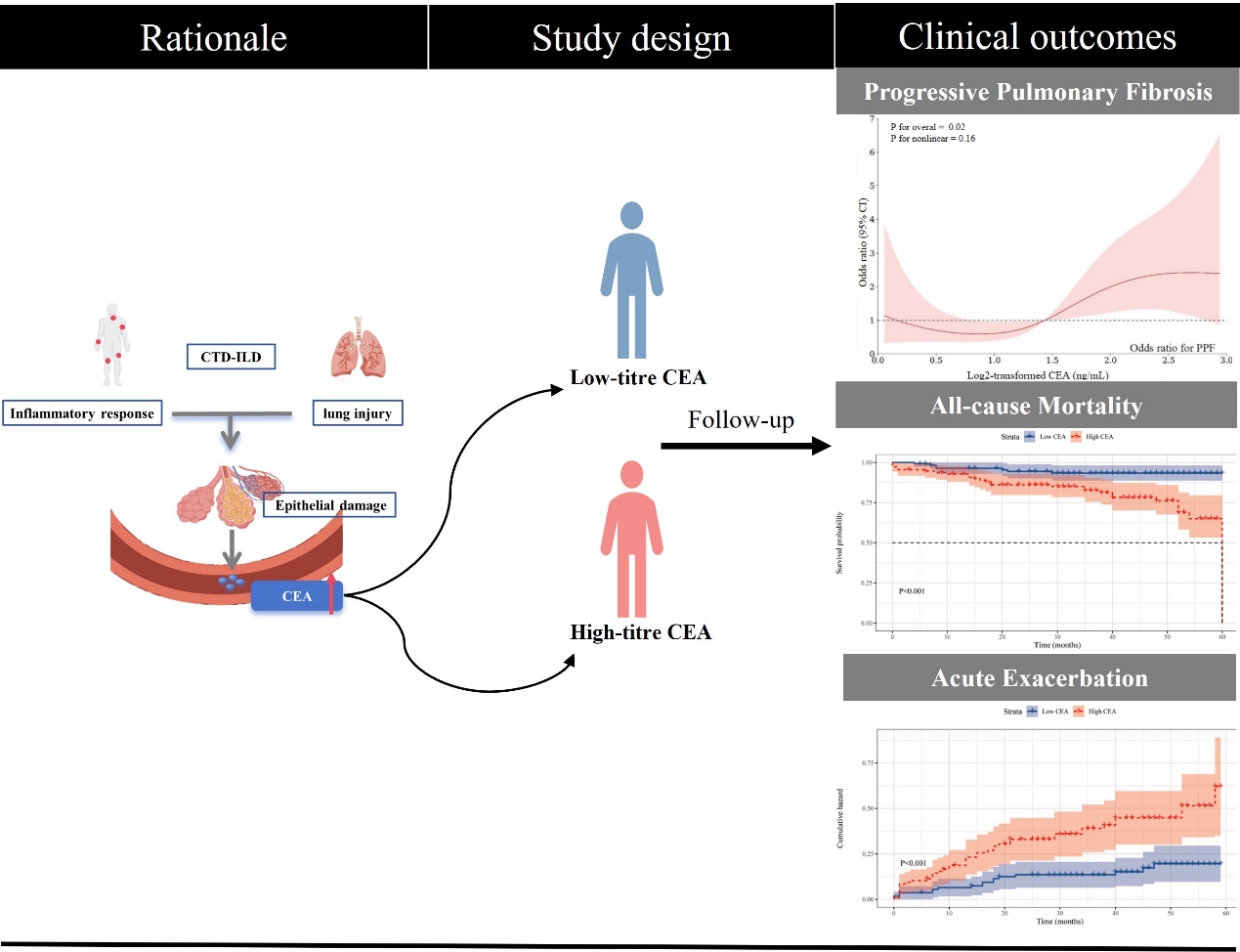
**Figure S3** Subgroup analysis of the proportional effect of elevated CA125 concentrations in PPF patients. CA125, carbohydrate antigen 125; CI, confidence interval; DLCO, diffusion capacity of the lung for carbon monoxide; FVC, forced vital capacity; OR, odds ratio; PPF, progressive pulmonary fibrosis.



**Figure S4** Patterns of pulmonary fibrosis progression. In 13 (20.63%) of these patients, progressive pulmonary fibrosis occurred after one acute exacerbation.



**Figure S5** Changes in of CEA and CA125 over time (n=90) stratified by PPF. The median time for the follow-up duration was 12 (9-12) months. No difference was observed for CEA and CA125 changes over time in both non-PPF (A/C, n=54) and PPF (B/D, n=36) group. CEA, carcinoembryonic antigen; CA125, carbohydrate antigen 125; PPF, progressive pulmonary fibrosis.



**Figure S6** Graphical abstract. CEA, carcinoembryonic antigen; CTD, connective tissue disease; ILD, interstitial lung disease.