IM - ORIGINAL

The natural history of idiopathic pulmonary fibrosis in a large European population: the role of age, sex and comorbidities

Antonella Caminati[1] · Fabiana Madotto[2,3] · Sara Conti[3] · Giancarlo Cesana[3] · Lorenzo Mantovani[2,3] · Sergio Harari[1,4,5]

Abstract

Placebo arms of clinical trials provide an opportunity to investigate the natural history of idiopathic pulmonary fibrosis (IPF) but these patients are not representative of the real life IPF population. Objective of this article is to evaluate patients' characteristics of incident IPF cases and their impact on mortality and hospitalizations risk. We conducted a retrospective cohort study using data from administrative databases from 2000 to 2010. Based on different algorithms reported in literature, incident IPF cases were identified. We applied Cox proportional hazards models to assess relationship between patients' characteristics, mortality and hospitalization. According to three case definitions, we identified 2338, 460 and 1704 incident IPF cases. Mean age at diagnosis was about 72 years, the proportion of male varied between 59 and 62% and patients with at least one chronic disease were between 70 and 74%. Age, male sex and comorbidities were associated to worse outcomes. Congestive heart failure (CHF), diabetes and cancer were conditions associated to mortality, while those associated to hospitalization were CHF and chronic obstructive pulmonary disease. Our data source provided one of the largest samples of unselected patients with a long follow-up period. Using different algorithms proposed and validated in literature, we observed that mortality and hospitalization rate are high in patients with IPF and age, sex and comorbidities significantly affect clinical outcomes. Females show a significant survival advantage over males, even after adjusting for age and comorbidities. Patients with pre-existing diseases, especially those with pulmonary and cardiovascular diseases are at higher risk.

Keywords Epidemiology · Administrative databases · Mortality · Survival · Hospitalization

A. Caminati and F. Madotto equal contribution—share first authorship

✉ Antonella Caminati
lafitta@libero.it

1 U.O. di Pneumologia e Terapia Semi-Intensiva Respiratoria, Servizio di Fisiopatologia Respiratoria ed Emodinamica Polmonare. Ospedale San Giuseppe, MultiMedica IRCCS, Via San Vittore 12, 20123 Milan, MI, Italy

2 Value-based Healthcare Unit, IRCCS MultiMedica, Via Milanese 300, Sesto San Giovanni, 20099 Milan, Italy

3 Research Centre on Public Health, Department of Medicine and Surgery, University of Milano-Bicocca, Via Cadore 48, 20900 Monza, Monza e Brianza, Italy

4 Clinica Medica, Ospedale San Giuseppe MultiMedica IRCCS, Via San Vittore 12, 20123 Milan, Italy

5 Dipartimento di Scienze Cliniche e di Comunità, Università di Milano, Milan, Italy

Introduction

Idiopathic pulmonary fibrosis (IPF) is a disease with a severe prognosis and a mean survival from the time of diagnosis estimated from retrospective studies of 3–5 years [1–5]. However, the natural history of IPF is highly variable and it is very difficult to predict the disease course in individual patient [6–9]. Placebo arms of recent phase II and III clinical trials provide an opportunity to investigate the natural history of lung function decline in patients with IPF [10–20]. However, these results should be regarded with caution, due to relatively short follow-up times (1–2 years); moreover, patients enrolled in clinical trials do not represent a random sample of the whole IPF population, many patients are in fact excluded for age, disease severity and comorbidities. For this reason, the probability of death estimated among patients enrolled in placebo arms of clinical trials has proved to be substantially lower as compared to real practice, while estimated survival appeared to be longer [21, 22].

In this context, healthcare administrative databases can represent a useful tool to conduct historical cohort studies on unselected populations to describe the real progression of a rare disease and to investigate the role of several health determinants on main disease outcomes. In literature, many researchers used this approach to estimate incidence and prevalence rate of IPF in specific geographic areas and their time trends [6, 23–30]. In these studies, the diagnostic code-based algorithms used to identify IPF cases are different according to the specific information gathered and available in the databases. Although some algorithms are clinically validated and their performances well-known, few studies have implemented this approach to analyse the main outcomes of the disease [31].

In this study, we used healthcare administrative databases of Lombardy Region (10 million inhabitants) to identify a retrospective cohort of unselect patients with IPF and analyse the progression of disease, in terms of mortality and hospitalizations following its onset. Secondary objective was to evaluate the impact of demographic characteristics and clinical complexity of patients on these outcomes.

Materials and methods

Study design, data collection and study population

We conducted a retrospective observational study using healthcare administrative data of the Lombardy Region that provides universal coverage within the Italian National Health Service (NHS).

At regional level, data regarding the main provided health services are stored in several large databases for administrative and epidemiological purposes. A data warehouse (called DENALI) has been created that collect and arrange health information of all Lombardy Region population [28].

A peculiarity of DENALI is the implementation of probabilistic record linkage to match data belonging to the same individual, but stored in different datasets. This linkage technique uses properties of variables common to different databases to determine the probability that two records refer to the same person [32, 33], and it therefore provides accurate matching when datasets do not share a single common identifier or when the identifier contains errors or omissions [34].

To perform our study, we used data collected in DENALI from 2000 to 2010, related to demographic aspects (sex, date of birth, date of death, date of emigration) and to the main characteristics of inpatient and outpatient claims, occurred in public or private hospitals and funded by the public healthcare system.

Since there is no univocal IPF case definition based on diagnosis codes reported in health claims, we used multiple analyses identifying IPF cases through three algorithms similar to those reported in recent studies, that have a clinical validation in literature and that are adaptable to Italian available data [23, 35, 36] (Table 1). The first definition of IPF case was based on the algorithm proposed by Raghu et al. [23, 35] and clinically validated by Esposito et al. [31], on patients at least 50 years old; the second one was based on algorithm proposed and validated by Ley [36]. The third case definition was also based on Ley's algorithm but we decided to remove the inclusion criteria "at least 2 claims for either ICD-9 code 516.3 at least 1 month apart", to include patients who did not have a second claim for IPF disease because: (1) the mild expression of disease did not require further access to the NHS, or (2) the severe clinical conditions led to death in a short time.

We excluded patients who had been living in Lombardy for less than 5 years at the time of the first medical claim with a diagnosis code for IPF (index event), to identify newly diagnosed IPF cases (incident cases). We chose a washout period of 5 years because the median survival of IPF patients is 2–5 years from diagnosis [1, 5]. Thus, for each case definition, we finally obtained the cohort of Lombardy dwellers with a new diagnosis of IPF between January 1st 2005 and December 31st 2010 (study period).

After the index event, we assessed vital status, time to death and hospitalizations occurred in the study period for each enrolled patient. Among hospitalizations, we also identified the subgroup of admissions due to an acute respiratory event, as those recorded with a first diagnosis of acute respiratory failure (ICD-9-CM code 518.81), acute lung oedema (518.84), pneumonia (486) or congestive heart failure (CHF) (428.0).

From all inpatients claims occurred during the five years preceding the index event, we identified coexisting chronic conditions, using ICD-9-CM diagnostic codes according to the algorithms proposed by Quan [37]. The coexisting pathologies were also aggregated into a comorbidity score, known as the Charlson comorbidity index (CCI) [38].

Statistical analysis

At the index event, we evaluated demographic characteristics (age at onset, sex) and coexisting chronic conditions. We compared data of patients selected based on the three IPF case definitions: differences were evaluated with Pearson χ^2 test for nominal and discrete variables and one-way ANOVA for continuous variables. Non-parametric tests were used if the variable distribution was not normal [39]. Bonferroni's corrections was used to assess differences between cohorts.

For each cohort identified through the three IPF case definitions, we estimated mean annual mortality rate (per 1,000 person-years), mean annual incidence rate for hospitalization and for acute respiratory hospitalization (per person-year), presenting them with exact 95% confidence intervals (95%

Table 1 Algorithms used to identify IPF cases in Italian healthcare administrative databases

	Case definition
Definition 1 based on Raghu et al. (Broad definition) [23, 35]	Inclusion criteria
	Age ≥ 50yrs at time of 1st claim forICD-9-CM code 516.3
	At least one claim withICD-9-CM diagnosis code of 516.3
	Exclusion criteria
	Any other diagnosis code for ILD‡ on or after the date of the last recorded diagnosis of IPF
Definition 2 based on Ley et al. [36]	Inclusion criteria
	Age ≥ 50yrs at time of 1st claim for either ICD-9-CM code 516.3
	At least 2 claims for either ICD-9 code 516.3 at least 1 month apart
	Any chest CT procedure claim* prior to the 1st diagnostic claim for IPF
	Exclusion criteria
	Any claims for alternative ILD codes† occurring on or after the first claim for IPF
Definition 3 based on Ley et al.[36]	Inclusion criteria
	Age ≥ 50yrs at time of 1st claim for either ICD-9-CM code 516.3
	Any chest CT procedure claim* prior to the 1st diagnostic claim for IPF
	Exclusion criteria
	Any claims for alternative ILD codes† occurring on or after the first claim for IPF

CT computed tomography, *ICD-9* international classification of diseases, 9th revision, *ICD-9-CM* international classification of diseases, 9th revision, clinical modification, *ILD* interstitial lung disease, *IPF* idiopathic pulmonary fibrosis

*Chest CT procedure code: ICD-9-CM 87.41

†ILD diagnoses, ICD-9 codes: 135, 237.7, 272.7, 277.3, 277.8, 446.21, 446.4, 495, 500–505, 506.4, 508.1, 508.8, 516.0, 516.1, 516.32–516.37, 516.2, 516.8, 516.9, 517.0, 517.2, 517.8, 518.3, 555, 710.0, 710.0–710.4, 714.0, 714.81, 720, and 759.5

‡ILD diagnoses, ICD-9 codes: 135, 237.7, 272.7, 277.3, 277.8, 446.21, 446.4, 495, 500–505, 506.4, 508.1, 508.8, 516.0, 516.1, 516.2, 516.8, 516.9, 517.0, 517.2, 517.8, 518.3, 555, 710.0, 710.0–710.4, 714.81, 720, and 759.5

CI). We carried out overall and sex-stratified survival analyses using the Kaplan–Meier approach to estimate mean survival time, mean time to the first hospitalization and mean time to first admission for acute respiratory event after the index event. Log-rank test was used to assess the significance of the difference between male and female patients. Cox proportional hazard models were applied to investigate the relationship between potential risk factors (age at onset, sex, CCI and comorbidities) and outcomes (death and hospital admission). The test proposed by Harrel and Lee based on Schoenfeld residuals was used to check the proportional hazards assumption.

For all statistical tests, a pre-specified two-sided α of 0.05 was regarded as statistically significant.

All analyses were performed using SAS software, version 9.4 (SAS Institute, Cary, NC, USA) and R, version 3.5.2 (R Project for Statistical Computing, http://www.R-project.org).

Results

According to the aforementioned definitions of IPF cases, we identified 2,338 (definition 1), 460 (definition 2), and 1,704 (definition 3) incident IPF cases with a mean age at diagnosis of about 72 years and a proportion of male patients varying between 59 and 62% (Table 2). At the index event, patients without a chronic comorbidity were around 26% and 30%, according to IPF case definition. The most prevalent coexisting diseases among IPF patients were: other pulmonary diseases (41–42%), CHF (19–23%), diabetes (17–19%), cerebrovascular diseases (13–15%), tumour (12–15%), and myocardial infarction (11%-12%). Overall, the complexity of female patients at diagnosis was significantly lower than that of male patients, regardless of the IPF case definition: 22% of males had not comorbidities while the percentage in females varied between 33 and 41% (Tables e1–e3). Based on definition 1, 43.4% of patients died during follow-up period with a mean annual mortality rate of 213.6 (per 1,000 person-years) (95% CI 213.2–214.0) (Table 3). This rate increased to 234.8 (95% CI 234.2–235.3) and to 267.5 (95% CI 266.5–268.6) applying definition 3 and definition 2, respectively.

In IPF patients, factors associated with mortality were age, male sex and CCI.

(Table e4, Fig. 1a–c, Figure e2A–e2C). When definition 2 was used, this last factor was not a statistically significant predictor (Table e4, Figure e1A–e1C).

Using definition 1, about 70% of study population had a hospital admission for any cause and the mean annual rate (per person-year) was 1.67 (95% CI 1.63–1.71) (Table 3).

Table 2 Characteristics of the study population at index event, according to IPF case definition

	Definition 1	Definition 2	Definition 3	p value
N	2,338	460	1,704	
Age (years), mean ± SD	72.63 ± 9.52	71.76 ± 9.19	72.63 ± 9.34	*0.4690*
Males, n (%)	1,373 (58.73)	286 (62.17)	1,023 (60.04)	*0.3429*
Charlson comorbidity index				
Median [IQR]	1 [0–3]	1 [0–2.5]	1 [0–3]†	*0.0396*
Classes				*0.1110*
0	632 (27.03)	136 (29.57)	442 (25.94)	
1–2	1,024 (43.80)	209 (45.43)	729 (42.78)	
>2	682 (29.17)	115 (25.00)	533 (31.28)	
Chronic comorbidities				
Myocardial infarction	260 (11.12)	51 (11.09)	208 (12.21)	*0.5386*
Congestive heart failure	516 (22.07)	87 (18.91)	387 (22.71)	*0.2161*
Peripheral vascular disease	188 (8.04)	40 (8.70)	151 (8.86)	*0.6339*
Cerebrovascular diseases	347 (14.84)	58 (12.61)	255 (14.96)	*0.4199*
Cerebrovascular disease	337 (14.41)	54 (11.74)	248 (14.55)	*0.2794*
Paraplegia and hemiplegia	37 (1.58)	6 (1.30)	25 (1.47)	*0.8893*
Dementia	36 (1.54)	4 (0.87)	20 (1.17)	*0.3985*
Chronic pulmonary diseases	952 (40.72)	190 (41.30)	737 (43.25)	*0.2672*
Peptic ulcer disease	42 (1.80)	6 (1.30)	28 (1.64)	*0.7430*
Liver disease	207 (8.85)	36 (7.83)	152 (8.92)	*0.7481*
Mild liver disease	206 (8.81)	36 (7.83)	151 (8.86)	*0.7680*
Moderate or severe liver disease	16 (0.68)	2 (0.43)	12 (0.70)	*0.8102*
Diabetes	394 (16.85)	88 (19.13)	308 (18.08)	*0.3856*
Diabetes without complications	381 (16.30)	86 (18.70)	299 (17.55)	*0.3468*
Diabetes with complications	71 (3.04)	15 (3.26)	51 (2.99)	*0.9566*
Renal disease	191 (8.17)	29 (6.30)	145 (8.51)	*0.3029*
Tumour	304 (13.00)	54 (11.74)	256 (15.02)	*0.0826*
Cancer	299 (12.79)	52 (11.30)	252 (14.79)	*0.0697*
Metastatic carcinoma	55 (2.35)	13 (2.83)	47 (2.76)	*0.6690*
AIDS/HIV	2 (0.09)	1 (0.22)	2 (0.12)	*0.5154*

IQR interquartile range (median, first quartile, third quartile), *SD* standard deviation

*Comparison versus "Definition 1", p value < 0.05 (Bonferroni's correction)

†Comparison versus "Definition 2", p value < 0.05 (Bonferroni's correction)

Concerning the hospital admission due to an acute respiratory event, 24% were hospitalized and the mean annual rate (per person-year) was 0.18 (95% CI 0.17–0.19). Based on definition 2, hospitalization rates and proportions of patients with at least one hospital admission were statistically higher compared to those obtained using definition 1. Using definition 3, the results related to the hospitalizations during the follow-up were similar to those obtained by applying definition 1 and they were statistically lower than those obtained by definition 2. Based on definition 1 and 3, the predictors associated to hospitalization for any cause and an acute respiratory event were age, male sex and CCI (Table e4, Figs. 1d–I, e2D–e2I).Using definition 2, age and male sex were statically significant predictors for acute respiratory hospitalization (Table e4, Figure e1D–e1I).

According to definition 1 and definition 3, chronic conditions associated to mortality in IPF patients were CHF, tumour and diabetes (Table 4). Concerning predictors for hospitalizations after IPF onset, CHF and COPD were associated to this outcome.

Discussion

Our study considered incident IPF cases in the Lombardy region (about 10 million inhabitants), the most populous region in Italy, in the period between 2005 and 2010. The study describes survival, hospitalizations and frequency of co-morbidities in an unselected IPF cohort, with typical demographic and clinical characteristics, followed up for a long period (five years of time), before the introduction

Table 3 Outcomes of interest occurred during follow-up period, according to IPF case definition

| Outcome of interest	Definition 1	Definition 2	Definition 3	p value
Mortality				
Deaths, n (%)	1,015 (43.41)	263 (56.96)*	759 (44.54)†	<.0001
Mean annual rate (per 1,000 person-years) (95% CI)	213.60 (213.18–214.01)	267.53 (266.51–268.55)*	234.76 (234.24–235.29)*†	<.0001
Time (months), mean ± SE	41.40 ± 0.68	36.05 ± 1.35*	39.80 ± 0.81	0.0112
Hospital admission				
Mean annual rate (per person-year) (95% CI)	1.67 (1.63–1.71)	2.60 (2.49–2.70)*	1.70 (1.66–1.75)†	<.0001
Patients with at least 1 hospitalization, n (%)	1,628 (69.63)	445 (96.74)*	1,177 (69.07)†	<.0001
Time (months) to first hospitalization, mean ± se	17.08 ± 0.56	7.25 ± 0.44*	17.02 ± 0.66†	<.0001
Acute respiratory hospitalization				
Mean annual rate (per person-year) (95% CI)	0.18 (0.17–0.19)	0.33 (0.29–0.36)*	0.19 (0.17–0.20)†	<.0001
Patients with at least 1 hospitalization, n (%)	563 (24.08)	205 (44.57)*	415 (24.35)†	<.0001
Time (months) to first hospitalization, mean ± se	50.99 ± 0.73	38.52 ± 1.56*	49.94 ± 0.88†	<.0001

CI confidence interval; SE standard error

*Comparison versus "Definition 1", p value < 0.05 (Bonferroni's correction)

†Comparison versus "Definition 2", p value < 0.05 (Bonferroni's correction)

Table 4 Multivariable Cox proportional hazards models

	Definition 1	Definition 2	Definition 3
	Hazard ratio (95%CI)	Hazard ratio (95%CI)	Hazard ratio (95%CI)
Mortality			
Age (years)	1.058 (1.050–1.066)	1.035 (1.020–1.050)	1.051 (1.042–1.060)
Sex (ref. Females)	1.457 (1.279–1.659)	1.574 (1.213–2.043)	1.473 (1.266–1.713)
Congestive heart failure (ref. No)	1.365 (1.185–1.572)	–	1.357 (1.154–1.595)
Tumour (ref. No)	1.528 (1.293–1.806)	–	1.476 (1.228–1.775)
Liver disease (ref. No)	1.314 (1.076–1.604)	–	–
Cerebrovascular Disease or paraplegia-hemiplegia (ref. No)	1.217 (1.036–1.429)	–	–
Diabetes (ref. No)	1.210 (1.034–1.415)	–	1.235 (1.035–1.475)
1st hospitalization			
Age (years)	1.012 (1.006–1.018)	–	1.012 (1.005–1.019)
Sex (ref. Females)	1.225 (1.104–1.358)	–	1.286 (1.138–1.453)
Congestive heart failure (ref. No)	1.383 (1.220–1.568)	–	1.355 (1.173–1.564)
COPD (ref. No)	1.288 (1.162–1.427)	1.373 (1.133–1.664)	1.313 (1.163–1.481)
Diabetes (ref. No)	1.178 (1.031–1.346)	–	–
Cerebrovascular Disease or paraplegia-hemiplegia (ref. No)	1.155 (1.002–1.331)	–	–
Tumour (ref. No)	—	–	1.210 (1.027–1.426)
1st acute respiratory hospitalization			
Age (years)	1.026 (1.017–1.036)	1.016 (1.000–1.032)	1.020 (1.009–1.032)
Sex (ref. Females)	1.464 (1.228–1.745)	1.523 (1.132–2.050)	1.398 (1.139–1.716)
Congestive heart failure (ref. No)	1.754 (1.453–2.118)	1.425 (1.023–1.986)	1.809 (1.457–2.246)
COPD (ref. No)	1.663 (1.404–1.969)	1.445 (1.088–1.918)	1.586 (1.300–1.936)
Diabetes (ref. No)	1.235 (1.002–1.523)	–	–

CI confidence interval; COPD: chronic obstructive pulmonary disease
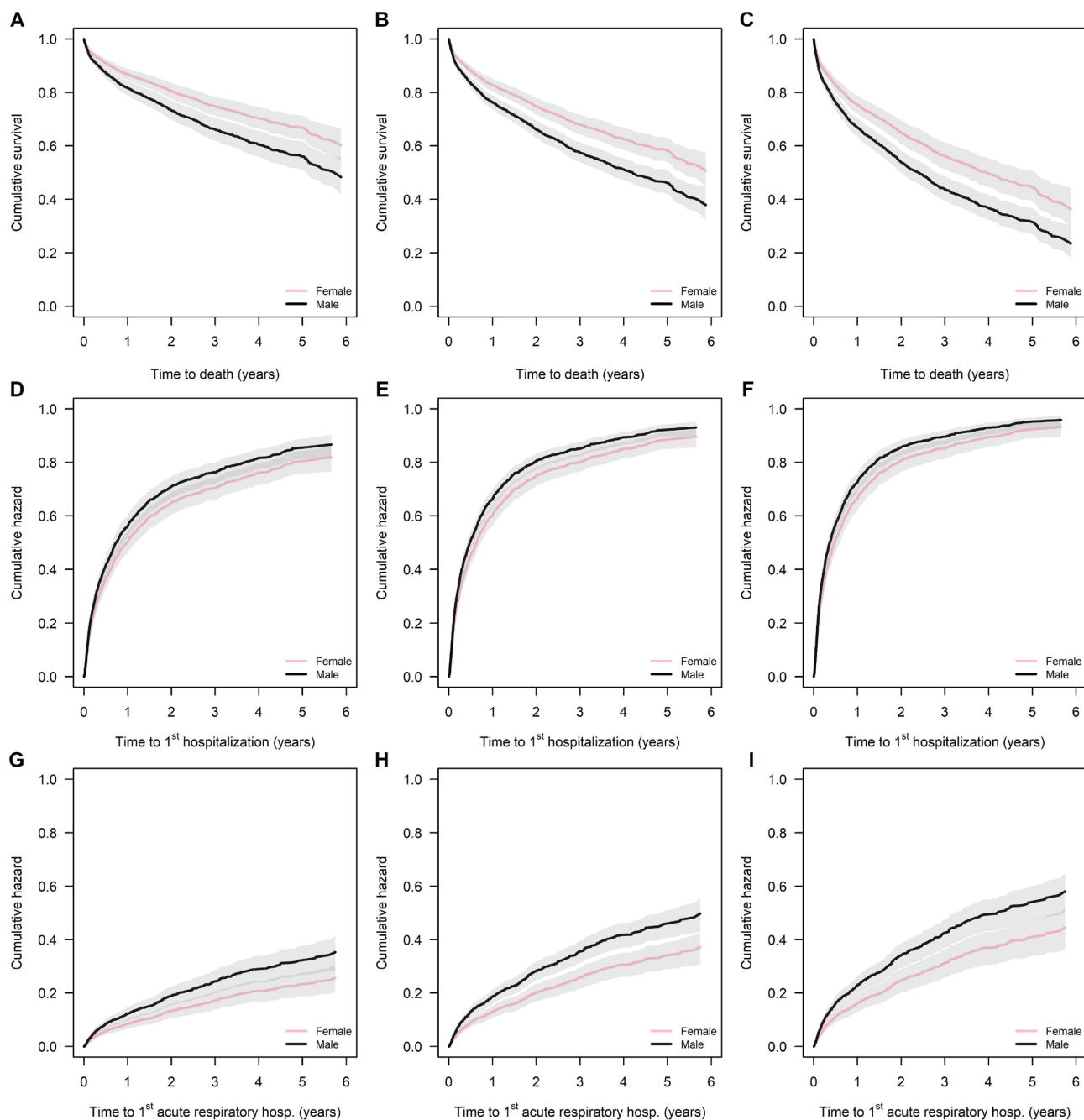
Fig. 1 Results of Cox proportional hazards models in incident IPF cases identified by definition 1.Survival function and probability of hospitalization during follow-up time. **a–c** Survival functions estimated by sex, for 70-year-old patient (mean age at IPF onset) and no chronic comorbidities (CCI equal to 0) (panel **a**), CCI equal to 1 or 2 (panel **b**) and CCI more than 2 (panel **c**). **d–f** Probability of first hospitalization(ordinary or day hospital) estimated by sex, for 70-year-old patient (mean age at IPF onset) and no chronic comorbidities (CCI equal to 0) (panel **d**), CCI equal to 1 or 2 (panel **e**) and CCI more than 2 (panel **f**). **g–l** Probability of first acute respiratory hospitalization estimated by sex, for 70-year-old patient (mean age at IPF onset) and no chronic comorbidities (CCI equal to 0) (panel **J**), CCI equal to 1 or 2 (panel **k**) and CCI more than 2 (panel **l**)

in clinical practice of antifibrotic therapies, pirfenidone and nintedanib. We used data from the healthcare administrative databases of the regional Health System; this approach allowed us to analyse a large unselected cohort of IPF patients and to follow them for a long period after

the disease onset. Using different algorithms proposed and validated in literature [23, 35, 36], we observed that mortality and hospitalization rate are high in patients with IPF and age, sex and comorbidities significantly affect clinical outcomes occurred after the onset. The median survival of IPF

has been established to be between 3.5 and 4.4 years in the US and 3.1 years in the UK and the cause of death was often the disease itself [7, 26]. Analyses of subgroups are consistent in affirming that the elderly are the most affected demographic subgroup [6, 7, 23, 26, 40, 41, 42, 43] and the ones with the poorest prognosis [44]; regarding sex, men have been shown to be more susceptible to IPF [7, 23, 24, 26], even if women seem to have a higher rate of increase in IPF mortality [6]. Our results confirm the severe prognosis of the disease with a mean survival from diagnosis of 3–3.4 years. Comparisons are difficult due to the use of different methods to identify the study populations and different study designs. In this unselected IPF population, males are more frequently represented and show a greater prevalence of comorbidity than females. As in a previous study [45], in our population, females show a significant survival advantage over males, even after adjusting for age and chronic comorbidities.

Hospitalizations are common events in patients with IPF [46, 47]. Studies on IPF patients have analysed the significance of hospitalization [48–50] and non-elective hospitalization is usually of clinical significance to the patient and it is associated with a high risk of subsequent mortality [51]. For this reason, hospitalization has been used as an endpoint in IPF clinical trials [15–17]. Hospitalization is common in our patients with IPF during the course of the disease with a mean annual rate (per person-year) varied between 1.67 (95% CI 1.63–1.71) and 2.60 (95% CI 2.49–2.70). Hospitalization is an event also reported in clinical trials, but occurs with a much lower frequency [52]. This may be due to better case selection, enrolment of patients with a generally less severe disease, fewer comorbidities and not elderly, and/or a shorter follow-up period.

In literature, prevalence data on comorbidities in IPF are sparse. The reported prevalence of comorbidities is variable; this depends on the type of studies and how comorbidities were studied. Comorbidities may affect quality of life of IPF patients and potentially influence survival; it is debatable whether treating comorbidities influences the clinical outcome [5, 42], even if it is reasonable to assume that improved treatment of comorbidities may have a favourable impact on the clinical course of these patients [42, 53]. Our results confirm that comorbidities are frequently observed in patients with IPF. The high prevalence of COPD patients comes as no surprise, since both diseases are related to smoking. In our cohort, we found a high prevalence of neoplasms (12–15%). Previous single center retrospective studies reported a high prevalence of these comorbidities and showed as some of these comorbidities influenced survival [42, 54]. We confirmed previous observations and outlined that the number of comorbidities is important for survival estimation [41, 42].

Population-based studies of IPF, as our study is, allow to analyse large population for long periods of time, but

have been limited by reliance on diagnostic code-based algorithms that lack clinical validation. The poor positive predictive value (PPV) of the IPF algorithm is likely due to a combination of misdiagnosis at the clinical level and miscoding at the administrative level [31]. A modified IPF algorithm was derived and validated to optimize the PPV by Ley's study [36]. Applying the Ley et al. [36] criteria to our population, the results were identical both regarding the prevalence of comorbidities and the effect of age and sex on prognosis (death and hospitalization). With the application of Ley et al. [36] criteria, the number of comorbidities does not influence the patient's prognosis: this could be due to the fact that with these criteria, the number of the population has decreased a lot and this does not allow to reach statistical. The Ley et al. criteria required at least two distinct IPF diagnosis codes at least one month apart; we have also introduced a modified Ley criteria believing that the need for at least two IPF diagnosis codes can exclude more serious patients and patients with very mild disease. Even with this change, the results have been substantially identical.

The consistency of the results obtained with the application of the different criteria supports the robustness and confidence of the results themselves.

According to previous results, in our study, chronic conditions associated to mortality in IPF patients were CHF, tumour and diabetes [41, 42]. We also observed that CHF and COPD were predictors for hospitalization.

There are some limitations in our study linked to the use of administrative healthcare data, as discussed in previous studies [28, 55]. First, the case definition used to identify IPF patients in this study is not validated with respect to medical record review and if diagnoses were based on multidisciplinary discussions is unknown. Second, pertinent clinical data such as smoking and occupational status, as well as environmental exposures and pulmonary function tests are not captured in our database and their effects on our findings are therefore unknown. Finally, in our analysis, we have not been able to establish the cause of death (whether due to the disease itself or to comorbidities) due to the absence of death certificate data.

Despite these limitations, our study analysed a very large population of unselected incident IPF cases over a long period of 5 years and our results were confirmed using different algorithms of IPF case definitions.

Our findings represent a retrospective observation of real life data in a period prior to the era of antifibrotic therapy and which therefore reflects the natural history of the disease. The mortality data, hospitalizations and comorbidities observed in our study probably come closer to the real situation than the same data obtained from clinical trials in which the patient population is definitely more homogeneous and well studied but also more selected and less severe. Differences between the data obtained from clinical trials

and real life also emerge from registries such as the German INSIGHT-IPF registry, where the yearly mortality, for example, turned out to be higher than the rate observed in clinical trials (observed mortality rate 14.2% per year vs. 7–8% in the placebo group of various clinical trials) [55, 56, 57]. The data collection process is different from ours, but the German study and our experience stress the marked difference in the clinical history of IPF that emerges with respect to the data provided by large randomized controlled clinical trials.

It should be emphasized that the data in our study relate to a time when the currently approved antifibrotic drugs for IPF therapy were not yet available. It will be interesting to see whether the introduction of these drugs, which may slow the progression of the disease, will in the coming years lead to a change in survival that can be detected in studies on real life patients like ours.

Lastly, one important finding from our study is the significant effect of sex on the clinical history of the disease. If these observations will be confirmed by further studies, we believe that sex should be taken into account in patient stratification when designing future clinical trials to avoid possible confounding factors in interpreting the results.

Conclusions

Since the Italian healthcare system is universal, our data source provided us with one of the largest samples of patients ever considered, without age limitations, and with a long follow-up period. Our results suggest that the burden of IPF could be considerable, potentially requiring substantial health care resources, as IPF patients present many comorbidities and are likely to be hospitalized.

These real life data confirm the poor prognosis for IPF, the high mortality and risk of hospitalization and the frequent presence of comorbidities related in part to the age of the affected population and partly to smoking. Our data provide evidence that the disease prognosis is significantly worse in men both in terms of higher and earlier mortality and earlier hospitalization. From the methodological point of view, our data are confirmed and validated after application of Ley's criteria and definition and this support the confidence on our results.

Compliance with ethical standards

References

1. Bjoraker JA, Ryu JH, Edwin MK, Myers JL, Tazelaar HD, Schroeder DR et al (1998) Prognostic significance of histopathologic subsets in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 157:199–203
2. Nicholson AG, Colby TV, du Bois RM, Hansell DM, Wells AU (2000) The prognostic significance of the histologic pattern of interstitial pneumonia in patients presenting with the clinical entity of cryptogenic fibrosing alveolitis. Am J Respir Crit Care Med 162:2213–2217
3. Flaherty KR, Toews GB, Travis WD, Colby TV, Kazerooni EA, Gross BH et al (2002) Clinical significance of histological classification of idiopathic interstitial pneumonia. Eur Respir J 19:275–283
4. Harari S, Caminati A (2015) Idiopathic pulmonary fibrosis: from clinical trials to real life experience. Eur Respir Rev 24:420–427
5. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al (2011) ATS/ERS/JRS/ALAT committee on idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med 183(6):788–824
6. Olson AL, Swigris JJ, Lezotte DC, Norris JM, Wilson CG, Brown KK (2007) Mortality from pulmonary fibrosis increased in the United States from 1992 to 2003. Am J Respir Crit Care Med 176(3):277–284
7. Fernández Pérez ER, Daniels CE, Schroeder DR, St SJ, Hartman TE, Bartholmai BJ et al (2010) Incidence, prevalence, and clinical course of idiopathic pulmonary fibrosis: a population-based study. Chest 137(1):129–137
8. Nathan SD, Shlobin OA, Weir N, Ahmad S, Kaldjob JM, Battle E et al (2011) Long-term course and prognosis of idiopathic pulmonary fibrosis in the new millennium. Chest 140(1):221–229
9. Brown AW, Shlobin OA, Weir N, Albano MC, Ahmad S, Smith M et al (2012) Dynamic patient counseling: a novel concept in idiopathic pulmonary fibrosis. Chest 142(4):1005–1010
10. Raghu G, Brown KK, Bradford WZ, Starko K, Noble PW, Schwartz DA, et al (2004) Idiopathic pulmonary fibrosis study group. A placebo-controlled trial of interferon gamma-1b in patients with idiopathic pulmonary fibrosis. N Engl J Med 350(2):125–133
11. King TE Jr, Albera C, Bradford WZ, Costabel U, Hormel P, Lancaster L, et al (2009) INSPIRE Study Group. Effect of interferon gamma-1b on survival inpatients with idiopathic pulmonary fibrosis (INSPIRE): a multicentre, randomised, placebo-controlled trial. Lancet 374 (9685):222–228
12. King TE Jr, Behr J, Brown KK, du Bois RM, Lancaster L, de Andrade JA et al (2008) BUILD-1: a randomized placebo controlled trial of bosentan in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 177(1):75–81
13. King TE Jr, Brown KK, Raghu G, du Bois RM, Lynch DA, Martinez F et al (2011) BUILD-3: a randomized, controlled trial of bosentan in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 184(1):92–99

14. Noble PW, Albera C, Bradford WZ, Costabel U, Glassberg MK, Kardatzke D, et al (2011) CAPACITY Study Group. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomized trials. Lancet 377 (9779):1760–1769

15. Raghu G, Behr J, Brown KK, Egan JJ, Kawut SM, Flaherty KR, et al (2013) ARTEMIS-IPF investigators. Treatment of idiopathic pulmonary fibrosis with ambrisentan: a parallel, randomized trial. Ann Intern Med 158 (9):641–649

16. Raghu G, Collard HR, Anstrom KJ, Flaherty KR, Fleming TR, King TE Jr et al (2012) Idiopathic pulmonary fibrosis: clinically meaningful primary endpoints in phase 3 clinical trials. Am J Respir Crit Care Med 185(10):1044–1048

17. Noth I, Anstrom KJ, Calvert SB, de Andrade J, Flaherty KR, Glazer C et al (2012) Idiopathic pulmonary fibrosis clinical research network (IPFnet) A placebo controlled randomized trial of warfarin in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 186(1):88–95

18. Raghu G, Anstrom KJ, King TE Jr, Lasky JA, Martinez FJ (2012) Idiopathic Pulmonary Fibrosis Clinical Research Network. Prednisone, azathioprine, and *N*-acetylcysteine for pulmonary fibrosis. N Engl J Med 366 (21): 1968–1977

19. Richeldi L, Costabel U, Moises S, Kim DS, Hansell DM, Nicholson AG et al (2011) Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis. N Engl J Med 365:1079–1087

20. Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U et al (2014) Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. N Engl J Med 370:2071–2082

21. Collard HR, Brown KK, Martinez JJ, Raghu G, Roberts RS, Anstrom KJ (2014) Study design implications of death and hospitalization as end points in idiopathic pulmonary fibrosis. Chest 146(5):1256–1262

22. Harari S, Caminati A (2015) Idiopathic pulmonary fibrosis: from clinical trials to real-life experiences. Eur Respir Rev 24(137):420–427

23. Raghu G, Weycher D, Edelsberg J, Bradford WZ, Oster G (2006) Incidence and prevalence of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 174:810–816

24. Gribbin J, Hubbard RB, Le Jeune I, Smith CJP, West J, Tata LJ (2006) Incidence and mortality of idiopathic pulmonary fibrosis and sarcoidosis in the UK. Thorax 61:980–985

25. Ohno S, Nakaya T, Bando M, Sugyma Y (2008) Idiopathic pulmonary fibrosis—results from a Japanese nationwide epidemiological survey using individual clinical records. Respirology 13:926–928

26. Navaratnam V, Fleming KM, West J, Smith CJP, Jenkins RG, Fogarty A, Hubbard RB (2011) The rising incidence of idiopathic pulmonary fibrosis in the UK. Thorax 66:462–467

27. Lai CC, Wang C-Y, Lu H-M, Chen L, Teng N-C, Yan Y-H et al (2012) Idiopathic pulmonary fibrosis in Taiwan—a population-based study. Respir Med 106:1566–1574

28. Harari S, Madotto F, Caminati A, Conti S, Cesana GC (2016) Epidemiology of idiopathic pulmonary fibrosis in Northern Italy. PLoS ONE 11:e0147072

29. Raghu G, Chen SY, Hou Q, Yeh WS, Collard HR (2016) Incidence and prevalence of idiopathic pulmonary fibrosis in US adult 18–64 years old. Eur Respir J 48:179–186

30. Tarride JE, Hopkins RB, Burke N, Guertin JR, O'Reilly D, Fell CD et al (2018) Clinical and economic burden of idiopathic pulmonary fibrosis in Quebec, Canada. Clinicoecon Outcomes Res 10:127–137

31. Esposito DB, Lanes D, Donneyong M, Holick CN, Lasky JA, Lederer D, et al (2015) Idiopathic pulmonary fibrosis in United States Automated Claims. Incidence, prevalence, and algorithm validation. Am J Respir Crit Care Med 192(10):1200–1207

32. Fellegi IP, Sunter AB (1969) A theory for record linkage. J Am Stat Assoc 64:1183–1210

33. Newcombe HB (1988) Handbook of record linkage: methods for health and statistical studies, administration, and business. Oxford University Press, Incorporate

34. Fornari C, Madotto F, Demaria M, Romanelli A, Pepe P, Raciti M et al (2008) Record-linkage procedures in epidemiology: an Italian multicentre study. Epidemiol Prev 32:79–88

35. Raghu G, Chen SY, Yeh WS, Maroni B, Li Q, Lee Y-C, Collard HR (2014) Idiopathic pulmonary fibrosis in US Medicare beneficiaries aged 65 years and older: incidence prevalence and survival 2001–11. Lancet Respir Med 2(7):566–572

36. Ley B, Urbania T, Husson G, Vittinghoff E, Brush DR, Eisner MD et al (2017) Code-based diagnostic algorithms for idiopathic pulmonary fibrosis. Case validation and improvement. Ann Am Thorac Soc 14:880–887

37. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J-C et al (2005) Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 43:1130–1139

38. Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 40:373–383

39. David S (2007) Handbook of parametric and nonparametric statistical procedures. Ed. Chapman & Hall/CRC (4th edition).

40. Ley B, Ryerson CJ, Vittinghoff E, Ryu JH, Tomassetti S, Lee JS et al (2012) A multidimensional index and staging system for idiopathic pulmonary fibrosis. Ann Intern Med 156:684–691

41. Torrisi SE, Ley B, Kreuter M, Wijsenbeek M, Vittinghoff E, Collard HR, Vancheri C (2019) The added value of comorbidities in predicting survival in idiopathic pulmonary fibrosis: a multicenter observational study. Eur Respir J 53(3):1801587

42. Kreuter M, Ehlers-Tenenbaum S, Palmowski K, Bruhwyler J, Oltmanns U, Muley T et al (2016) Impact of comorbidities on mortality inpatients with idiopathic pulmonary fibrosis. PLoS ONE 11(3):e0151425

43. Mannino DM, Etzel RA, Parrish RG (1996) Pulmonary fibrosis deaths in the United States, 1979–1991: an analysis of multiple-cause mortality data. Am J Respir Crit Care Med 153:1548–1552

44. Ley B, Collard HR, King TE Jr (2011) Clinical course and prediction of survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 183:431–440

45. Han MK, Murray S, Fell CD, Flaherty KR, Toews GB, Myers J et al (2008) Sex differences in physiological progression of idiopathic pulmonary fibrosis. Eur Respir J 31:1183–1188

46. Collard HR, Chen SY, Yeh WS, Li Q, Lee Y-C, Wang A, Raghu G (2015) Health care utilization and costs of idiopathic pulmonary fibrosis in U.S. Medicare beneficiaries aged 65 years and older. Ann Am Thorac Soc 12(7): 981–987

47. Raimundo K, Chang E, Broder MS, Alexander K, Zazzali J, Swigris JJ (2016) Clinical and economic burden of idiopathic pulmonary fibrosis: a retrospective cohort study. PBM Pulm Med 16:2

48. Song JW, Hong SB, Lim CM, Koh Y, Kim DS (2011) Acute exacerbation of idiopathicpulmonary fibrosis: incidence, risk factors and outcome. Eur Respir J 37(2):356–363

49. Martinez FJ, Safrin S, Weycker D, Starko KM, Bradford WZ, King TE Jr, et al (2005) IPF Study Group. The clinical course of patients with idiopathic pulmonary fibrosis. Ann Intern Med 142 (12pt 1):963–967

50. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A et al (2011) Ascertainment of individual risk of mortality for patients with idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 184(4):459–466

51. Brown AW, Fischer CP, Shlobin OA, Buhr RG, Shahzad Ahmad S, Weir NA, Nathan SD (2015) Outcomes after hospitalization in idiopathic pulmonary fibrosis. A cohort study. Chest 147(1):173–179

52. Ley B, Bradford WZ, Vittinghoff E, Weycker D, du Bois RM, Collard HR (2016) Predictors of mortality poorly predict common

measures of disease progression in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 194:711–718

53. Caminati A, Lonati C, Cassandro R, Elia D, Pelosi G, Torre O et al (2019) Comorbidities on IPF: an underestimated issue. Eur Respir Rev 28(153):190044

54. Hyldgaard C, Hilberg O, Bendstrup E (2014) How does comorbidity influence survival in idiopathic pulmonary fibrosis? Respir Med 108(4):647–653

55. Caminati A, Madotto F, Cesana GC, Conti S, Harari S (2015) Epidemiological studies in idiopathic pulmonary fibrosis: pitfalls in methodologies and data interpretation. Eur Respir Rev 24:436–444

56. Behr J, Wirtz H, Pittrow D et al (2014) Clinical outcomes of patients with idiopathic pulmonary fibrosis in clinical practice: follow-up results of the INSIGHTS-IPF Registry. Eur Respir J 17(7):A600

57. Behr J, Prasse A, Wirtz H, Koschel D, Pittrow D, Held M et al (2020) Survival and course of lung function in the presence or absence of antifibrotic treatment in patients with idiopathic pulmonary fibrosis: long-term results of the INSIGHTS-IPF registry. Eur Respir J. https://doi.org/10.1183/13993003.02279-2019