Epidemiological studies in idiopathic pulmonary fibrosis: pitfalls in methodologies and data interpretation

Antonella Caminati[1,3], Fabiana Madotto[2,3], Giancarlo Cesana[2], Sara Conti[2] and Sergio Harari[1]

Affiliations: [1]U.O. di Pneumologia e Terapia Semi-Intensiva Respiratoria – Servizio di Fisiopatologia Respiratoria ed Emodinamica Polmonare, Ospedale San Giuseppe – Multimedica IRCCS, Milan, Italy. [2]Research Center on Public Health, Dept of Health Science, University of Milano-Bicocca, Monza, Italy. [3]Both authors contributed equally.

Correspondence: Antonella Caminati, 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, Italy. E-mail: lafitta@libero.it

ABSTRACT Data on incidence, prevalence and mortality of idiopathic pulmonary fibrosis (IPF) are sparse and vary across studies. The true incidence and prevalence of the disease are unknown. In general, the overall prevalence and incidence reported in European and Asian countries are lower than those reported in American studies. In recent years, the epidemiological approach to IPF has been difficult for many reasons. First, the diagnostic criteria of the disease have changed over time. Secondly, the coding system used for IPF in administrative databases, the most common data source used to study this aspect of the disease, has been modified in the past few years. Finally, the study design, the methodology and the population selected in each of the studies are very different. All these aspects make comparisons among studies very difficult or impossible. In this review, we list the main issues that might arise when comparing different studies and that should be taken into consideration when describing the state of epidemiological knowledge concerning this pathology.

@ERSpublications
An accurate understanding of IPF epidemiology is important; comparisons of studies must consider their heterogeneity http://ow.ly/P5sZF

Introduction

Idiopathic pulmonary fibrosis (IPF) is a specific and progressive form of fibrotic interstitial pneumonia of unknown cause with a severe prognosis. Median survival is estimated to be 3–5 years from the time of diagnosis [1, 2], and the disease prognosis is unpredictable [2, 3]. IPF was considered a rare disease with no possible therapy until a few years ago. After several negative studies, two drugs have recently been approved for the treatment of IPF: pirfenidone and nintedanib. This has started a new era for the management of this devastating disease [4–7]. Despite its clinical relevance, and the extensive research into its pathophysiology and possible therapeutic approaches, IPF remains an uncommon disorder and only a small number of studies have focused on its epidemiology. The epidemiology of IPF has not yet been completely investigated. This is due to the relative rarity of the disease, which makes it difficult to enrol adequately sized study samples [8, 9]. Most information on the epidemiology of IPF comes from the USA [10–14] and Europe [15–25], where most studies have been conducted, although we also have some information from Asian countries [26, 27].

As highlighted in two recent systematic reviews on IPF epidemiology [28, 29], estimated mortality and morbidity rates are highly variable. KAUNISTO et al. [28] suggested that the observed differences are probably due to various and heterogeneous methods of investigation and not to true geographical differences, as some authors suggest. KAUNISTO et al. [28] also give some indications of possible reasons for the observed variability in epidemiological parameters, such as different diagnostic criteria or disease definition. In the most recent review, published in 2015 [29], that systematically searched for all population-based studies on incidence and mortality of IPF, the authors listed the strengths and weaknesses of different study designs and stressed that such differences in design make it difficult to produce global incidence and mortality estimates through meta-analysis.

Since these aspects are crucial for correctly approaching the IPF literature, in this review we list the main issues that might arise when comparing different studies and the issues that should be taken into consideration when describing the state of epidemiological knowledge concerning this pathology.

Methods

We integrated the systematic review of literature on IPF epidemiology carried out by KAUNISTO et al. [28] adding studies published before January 01, 2001, and between January 09, 2012 and April 23, 2015. We searched for "idiopathic pulmonary fibrosis" AND "epidemiology" in the MEDLINE (PubMed) and Science Citation Index (Web of Science) databases and selected all publications reporting measures of incidence, prevalence and mortality of IPF. We identified 22 publications fulfilling the aforementioned criteria, 13 of which corresponded to those included in the review by KAUNISTO et al. [28]. Finally, we added a further recent study that did not fulfil the inclusion criteria but reported the results of a German IPF registry [30].

Results

Three commonly used methods for the evaluation of IPF epidemiology in a population have been used in the different published studies: national registries of interstitial lung diseases (ILDs), questionnaires send to pulmonologists and the most common method used in the IPF epidemiological studies, the use of pre-existing databases [28]. The results published from these studies are very different (table 1).

After reviewing the 23 selected studies, we decided to distinguish between those using administrative databases [9, 11–14, 18, 19, 24, 25, 31, 32, 34, 35] and those using clinical data [11, 13, 15, 16, 18, 20–24, 26, 27, 30, 32, 33], as they have completely different features. We then structured our review into three parts: we will first discuss comparison issues within the group of studies based on administrative databases, similarly we will then address the group of studies based on clinical data, and finally, we will highlight the major challenges in comparing the two groups.

Comparability among studies based on administrative databases

Studies which used administrative databases looked at databases recording hospital admissions [11–14, 18, 19, 25, 32, 34], outpatient visits [11–14, 32, 34] and death certificates [9, 11, 24, 25, 31, 35]. While the database on death certificates generally covers the whole population of a country or of a defined area, information about hospitalisations and outpatient claims is usually stored to trace the healthcare expenditures borne by insurers or national healthcare systems and consequently it could cover the whole population or only a portion of it. Therefore, selection bias might occur in those studies concerning countries where the national healthcare system does not provide a universal coverage or where insurance data concerns only a portion of the population with specific characteristics (e.g. high or low socioeconomic status). Such bias would affect the comparability of two studies based on administrative databases. For example, one of the largest studies carried out in the USA [12] used the claim database of a large health plan to define the incidence and the prevalence of IPF. The overall rates of prevalence were estimated to range between 14 and 43 per 100 000 person-years, while the incidence ranged between 7 and 16 per 100 000 person-years, depending on the case definition. It should be noted that patients covered by the health plan might not be representative for the whole population, because they might have a higher socioeconomic status given that they can afford a health plan.

Furthermore, in these databases patients with IPF are identified through algorithms based on the diagnoses reported in such archives and usually coded using the International Classification of Disease (ICD) ninth or 10th revision (ICD-9CM or ICD-10). IPF was not given a diagnostic code in the ICD until the ninth revision (ICD-9) at the end of the 1970s. However, selection algorithms are not homogenous; therefore, when comparing studies of this kind we should ask ourselves if the differences in the selection criteria can lead to the identification of samples of patients with different characteristics (e.g. different disease severity). If we compare the study by VON PLESSEN et al. [19] with that of RAGHU et al. [14],

TABLE 1 Studies on idiopathic pulmonary fibrosis (IPF) epidemiology

First author [ref.]	Country and study period	Study methods and case definition	Study population	Age limitations	Rates per 100 000 person-years
MANNINO [31]	USA, 1979–1991	Source: death certificate reports compiled by the National Center for Health Statistics Cases identified based on: ICD-8CM codes and ICD-9CM codes (515, 516.3), used to identify PF	US population (N/A)	None	Mortality for PF: 330 (95% CI N/A)
COULTAS [11]	USA, 1988–1990	Source: population-based ILD registry, clinical and administrative databases Cases identified based on: physician diagnosis, following a common guideline	480 577 (year: 1990)	≥18 years (stratified data for ≥35 years)	Prevalence of IPF: Males 20.2 (95% CI N/A); Females 13.2 (95% CI N/A) Incidence of IPF: Males: 10.7 (95% CI N/A); Females: 7.4 (95% CI N/A)
OLSON [9]	USA, 1992–2003	Source: death certificate reports compiled by the National Center for Health Statistics Cases identified based on: ICD-9CM codes (515, 516.3) and ICD-10CM codes (J84.1), used to identify PF	US population (N/A)	None (stratified data for ≥45 years)	Mortality for PF: 508 (95% CI N/A)
RAGHU [12]	USA, 1996–2000	Source: healthcare claims databases of a health plan Cases identified based on: ICD9-CM codes (516.3), with two algorithms (broad case definition# and narrow case definition¶)	2 149 054	≥18 years	Broad case definition: Prevalence of IPF 42.7 (95% CI N/A); Incidence of IPF 16.3 (95% CI N/A) Narrow case definition: Prevalence of IPF 14.0 (95% CI N/A); Incidence of IPF 6.8 (95% CI N/A)
FERNÁNDEZ PÉREZ [13]	USA, 1997–2005	Source: clinical and administrative database of medical care Cases identified based on: ICD9-CM codes (516.3), hospital international classification of diseases-adapted codes (517, 519), physician diagnosis, ATS/ERS consensus 2002 [37] Two algorithms (narrow case definition+ and broad case definition§)	~128 000 (all ages)	≥50 years	Broad case definition: Prevalence of IPF 63.0 (95% CI 36.4–89.6); Incidence of IPF 17.4 (95% CI 12.4–22.4) Narrow case definition: Prevalence of IPF 27.9 (95% CI 10.4–45.4); Incidence of IPF 8.8 (95% CI 5.3–12.4)
RAGHU [14]	USA, 2000–2011	Source: administrative claims from a 5% random sample of Medicare beneficiaries Cases identified based on: ICD9-CM codes (516.3), with two algorithms (broad case definition# and narrow case definition¶)	2 871 436	≥65 years	Broad case definition: Prevalence of IPF 82.6 (95% CI N/A) in 2001; 233.3 (95% CI N/A) in 2011 Incidence of IPF 31.1 (95% CI N/A) in 2001; 43.0 (95% CI N/A) in year 2011 Narrow case definition: Prevalence of IPF 30.5 (95% CI N/A) in 2001; 145.7 (95% CI N/A) in 2011 Incidence of IPF 15.9 (95% CI N/A) in 2001; 31.1 (95% CI N/A) in 2011
NAVARATNAM [24]	UK, 1968–2009	Source: the Health Improvement Network database (primary care), death certificates from the Office for National Statistics Cases identified based on: ICD-8CM codes (517), ICD-9CM codes (516.3, 515), ICD-10CM codes (J84.1), the Read codes (H563.00, H563.11, H563.12, H563100, H563z00)	Population of the study area (N/A)	≥40 years	Incidence of IPF 2000–2009: 7.44 (95% CI 7.12–7.77) Mortality for IPF 1968–2008: 2.54 (95% CI 2.52–2.56)

TABLE 1 Continued

First author [ref.]	Country and study period	Study methods and case definition	Study population	Age limitations	Rates per 100 000 person-years
Kolek [15]	Czech Republic, 1981–1990	Source: retrospective observational study in 24 centres Cases identified based on: physician diagnosis of CFA	4 033 010 (year 1990)	None	Prevalence of CFA 6.5 (95% CI N/A) in 1981; 12.1 (95% CI N/A) in 1990 Incidence of CFA 0.74 (95% CI N/A) in 1981; 1.28 (95% CI N/A) in 1990
von Plessen [19]	Norway, 1984–1998	Source: hospital records Cases identified based on: ICD-8CM codes (517, 517.00, 517.01, 517.09) and ICD-9CM codes (515, 516.3, 516.8, 516.9), used to identify CFA	254 999 (year 1999)	≥16 years	All CFA cases: Prevalence 23.4 (95% CI 14.9–33.0) Incidence 4.3 (95% CI N/A) CFA cases after clinical validation: Prevalence 11.4 (95% CI N/A) Incidence 2.7 (95% CI N/A)
Gribbin [23]	UK, 1991–2003	Source: the Health Improvement Network database (primary care) Cases identified based on: physician diagnosis	1 555 152 (all ages)	≥40 years	Incidence of IPF 4.6 (95% CI 4.3–4.9)
Thomeer [16]	Belgium, 1992–1996	Source: population-based ILDs registry in 20 centres Cases identified based on: physician diagnosis, ATS/ERS consensus 2000 [36]	5 768 925	None	Prevalence of IPF 1.25 (95% CI N/A) Incidence of IPF 0.22 (95% CI N/A)
Hodgson [18]	Finland, 1997–1998	Source: pulmonary clinic databases nationwide (29 clinics) Cases identified based on: ICD10-CM codes (J84.1), physician diagnosis, ATS/ERS guidelines 2000 [36]	5 170 000	None	Prevalence of IPF 16–18 (95% CI N/A)
Tinelli [21]	Italy, 1998–2005	Source: national register for ILDs, voluntary participation Cases identified based on: physician diagnosis	N/A	None	Number of IPF cases: 864, rates N/A
Xaubet [20]	Spain, 2000–2001	Source: population-based ILDs registry in 23 respiratory centres Cases identified based on: physician diagnosis, ATS/ERS consensus 2000 [36] and 2002 [37]	6 700 000	None	Incidence of ILDs 7.6 (95% CI N/A) IPF constituted 38.6% of cases
Hyldgaard [32]	Denmark, 2003–2009	Source: clinical and administrative databases, patients who were diagnosed with ILDs Cases identified based on: ICD10-CM codes, validation with ATS/ERS consensus 2011 [1]	Population of the study area (N/A)	None	Incidence of ILDs 4.1 (95% CI N/A) IPF constituted 28.0% of cases
Karakatsani [22]	Greece, 2004	Source: ad hoc questionnaires for investigating incidence and prevalence of IPF Cases identified based on: physician diagnosis, ATS/ERS consensus 2002 [37]	~3 360 000	≥16 years	Prevalence of IPF 3.38 (95% CI N/A) Incidence of IPF 0.93 (95% CI N/A)
Agabiti [25]	Italy, 2005–2009	Source: hospital admissions database, mortality database of the regional public health service in Lazio Cases identified based on: ICD9-CM codes (516.3)	4 727 710	≥18 years	All IPF cases: Prevalence 25.6 (95% CI 25.1–26.2) Incidence 7.5 (95% CI 7.3–7.7) IPF "confident" definition after hospital chart audit: Prevalence 31.6 (95% CI 30.9–32.2) Incidence 9.3 (95% CI 9.2–9.4)
Musellim [33]	Turkey, 2007–2009	Source: ad hoc questionnaires, prospective study (31 centres in 19 cities) Cases identified based on: physician diagnosis, ATS/ERS guidelines 2002 [37], to identify ILDs	N/A	None	Incidence of ILDs 25.8 (95% CI N/A) IPF constituted 19.9% of cases

TABLE 1   Continued

First author [ref.]	Country and study period	Study methods and case definition	Study population	Age limitations	Rates per 100 000 person-years
Behr [30]	Germany, 2012–2014	Source: national register for IPF (19 centres) Cases identified based on: physician diagnosis	N/A	⩾18 years	Number of IPF cases: 502 Number of prevalent IPF cases: 331 Number of incident IPF cases: 171 Rates N/A
Lai [34]	Taiwan, 1997–2007	Source: administrative databases from Bureau of National Health Insurance and Ministry of the Interior, regarding health services and deaths Cases identified based on: ICD9-CM codes (516.3), with two algorithms (broad case definition[#] and narrow case definition[¶])	2 619 534 (all ages)	⩾18 years	Broad case definition: Prevalence of IPF 0.7 (95% CI N/A) in 1998; 6.4 (95% CI N/A) in 2007 Incidence of IPF 0.6 (95% CI N/A) in 1998; 1.4 (95% CI N/A) in 2007 Narrow case definition: Prevalence of IPF 0.5 (95% CI N/A) in 1998; 4.9 (95% CI N/A) in 2007 Incidence of IPF 0.5 (95% CI N/A) in 1998; 1.2 (95% CI N/A) in 2007
Natsuizaka [27]	Japan, 2003–2007	Source: clinical database of patients with IPF and the Certificate of Medical Benefit Cases identified based on: physician diagnosis, ATS/ERS guidelines 2000 [36]	5 572 770	None	Prevalence of IPF 10 (95% CI N/A) Incidence of IPF 2.23 (95% CI N/A)
Ohno [26]	Japan, 2005	Source: clinical database of patients with IIP and the Certificate of Medical Benefit Cases identified based on: physician diagnosis	127 756 815	None	Prevalence of IIP 3.44 (95% CI N/A) Number of IPF prevalent cases: 1322 Number of IPF incident cases: 545
Hutchinson [35]	Worldwide, 1999–2012	Source: mortality data from national statistics agencies Cases identified based on: ICD10-CM codes (J84 and J84.1)	N/A	None	Mortality for IPF 2010 J84: between 4.60 (Sweden) and 9.93 (Japan) (95% CI N/A) J84.1: between 4.63 (Spain) and 7.19 (England and Wales) (95% CI N/A)

ICD: International Classification of Diseases; PF: pulmonary fibrosis; N/A: not available; ILD: interstitial lung disease; ATS: American Thoracic Society; ERS: European Respiratory Society; CFA: cryptogenic fibrosing alveolitis; IIP: idiopathic interstitial pneumonia. [#]: age ⩾18 years, at least one medical claim with a diagnosis code for IPF (ICD-9-CM 516.3), no medical claims with a diagnosis code for any other ILDs on or after the date of the last medical claim with a diagnosis code for IPF; [¶]: met broad case definition (see [#]) plus had at least one medical claim with a procedure code for surgical lung biopsy, transbronchial lung biopsy or high-resolution computed tomography (HRCT) on or before the date of the last medical claim with a diagnosis code for IPF; [+]: usual interstitial pneumonia (UIP) on surgical lung biopsy specimen or a definite UIP pattern on a HRCT image; [§]: UIP on surgical lung biopsy specimen or a definite or possible UIP pattern on a HRCT image.

we should keep in mind that the first study selected all diagnoses coded as 515, 516.3, 516.8 and 516.9 (ICD9-CM) as potential IPF cases, while the second study focused on 516.3.

While it is likely that studies using different ICD codes to detect IPF will lead to the selection of nonhomogeneous patients, even if the ICD codes are the same the accuracy of the reported diagnosis or cause of death might vary across different geographical areas or even within the same area. The accuracy of the reported diagnosis depends on the background of the physician who registered the diagnosis: using this data source, it is impossible to ascertain whether this physician was a pneumologist or not. The accuracy also depends on the study period, as diagnostic procedures have greatly improved in recent decades making diagnosis more and more precise. In particular, the introduction and spread of high-resolution computed tomography (HRCT) has completely changed IPF diagnosis. During the 1990s, HRCT scans became routine in the evaluation of patients with ILD. As the technology advanced, HRCT proved to be a simple, noninvasive, relatively inexpensive method (compared with surgical lung biopsy) to make a confident diagnosis of IPF in many patients. The role of bronchoalveolar lavage and transbronchial biopsy in IPF diagnosis has also changed over time and the relevance of these two methodologies has been revisited in the latest statement from the American Thoracic Society (ATS)/European Respiratory Society (ERS)/Japanese Respiratory Society/Latin American Thoracic Association [1].

Another factor that can cause differences over time and affect the comparability among studies is the introduction and evolution of guidelines to identify ILD in general, and more specifically IPF. A first consensus statement for the diagnosis and the management of IPF was published in 2000 [36] and was then revised in 2011 [1]. These statements have determined a more precise definition of the radiological and histological diagnosis of IPF than before. Changes in classification of idiopathic interstitial pneumonias (IIP) and diagnostics make comparison between studies conducted in different decades difficult [36, 37]. For example, it is hard to compare estimated incidence and prevalence rates from a study based on the adult population in Bernalillo County, NM, USA in 1988–1990 [11], *i.e.* 10 years before the first consensus statement, with those of a more recent study carried out after the first consensus statement in the USA from 2001 to 2011 [14].

In addition to this, one should also look at when the guidelines were actually introduced into clinical practice: the gap between first publication and the actual adoption of these guidelines can be different among countries, but also within the same country. Therefore, even studies carried out in the same period might gather cases diagnosed following different guidelines. This might also affect estimates of the temporal trends within the same study, for example in Olmsted County, MN, USA [13] during the period 1997–2005, using a broad case definition, incidence ranged between 24.6 per 100 000 person-years during the period 1997–1999 and 11.0 per 100 000 person-years during the period 2003–2005. We should keep in mind that the observed decrease might be due to the introduction of the first ATS/ERS consensus statement [36]. A similar observation applies to a study conducted in Taiwan during the period 1997–2007 [34]. Another important feature that should also be considered is the presence of a clinical validation of the diagnosis reported in the administrative archives: not all of the studies ascertain the diagnosis through a clinical revision of each patient's medical record. The proportion of patients wrongly classified as IPF cases is likely to be higher in studies lacking a validation. Studies that determined the cases with a clinical validation usually integrated administrative and clinical databases [11, 13, 18, 19, 32].

Even with clinical validation, in the case of administrative databases it is difficult to go back to the true disease onset. We can assume that the onset corresponds to the first access of healthcare services, but different cultural and economic reasons might lead the IPF patient to seek medical care at different stages during the clinical course of the disease.

Moreover, studies which only use hospital admission databases cannot detect those patients who are followed solely through outpatient visits and who are never hospitalised (*e.g.* those who die before their first hospitalisation). This affects the comparison between studies which considered both outpatient and inpatient claims [11–14, 32, 34] and those which focused only on inpatient claims [18, 19, 25].

If we again focus on the identification of disease onset, it should be noted that some patients might have accessed healthcare services with an IPF diagnosis before the study period; therefore, studies that use a washout period probably produce less biased estimates of incidence and prevalence than those that don't consider this issue. For example, RAGHU *et al.* [14] required a washout period of 1 year before the first available diagnosis of IPF, unlike VON PLESSEN *et al.* [19] who did not make any restrictions. Also the length of the washout period might affect the estimates.

Last but not least, there are some methodological aspects regarding rate calculations that are crucial when wanting to compare estimated rates. First, since it has been shown that age and gender are risk factors for IPF and that the demographic structures of the populations involved in the studies might be different, a

proper comparison should aim to standardise the results [38], although this is possible only if the authors provide age- and gender-specific rates, using the same age classes. For example, we should be careful when comparing the estimates from Bernalillo County [11] and Olmsted County [13]. Although they both provide age- and gender-specific rates, they use different age-classes. Both studies use 10-year-wide classes, but the first study starts at 35 years of age, while the second starts at 50 years of age. In addition, it is advisable to provide a variability measure (standard error or confidence interval) of the estimated rates, because this allows reliable statistical comparisons to be built. Finally, if we are comparing overall rates we should always verify if the studies applied different age limits to the study population: the overall rates estimated in the US Medicare population [14] are not directly comparable to those estimated within a US population covered by a health plan [12], because the first study involves only patients aged 65 years and older, while the second also involves younger patients (⩾18 years of age).

Comparability among studies based on clinical databases

In the second group of studies, based on clinical databases, we included all national registries of ILDs, studies with questionnaires where data collection was based on the activity of physicians [15, 16, 20–22, 27, 30, 33] and studies based on existing clinical databases, such as The Health Improvement Network in the UK [24, 26, 32].

In these studies, the quality of the collected data relies on the background information received from the different physicians, on the ATS/ERS guidelines used at the time of the study and on the available diagnostic procedures. Given that these features can vary over time, we should be careful when comparing studies carried out in different years, such as the Greek study [22] that identified IPF cases in 2004 according to the ATS/ERS guidelines of 2000 [36] and the registry built in Flanders [16], that did not follow a guideline as it identified cases during the period 1992–1996.

Furthermore, physicians who take part in registries and questionnaire-based studies are volunteers. This could generate a selection bias: differences in the activity of the physicians in reporting patient numbers and data might lead to the selection of a sample of patients that is not representative of the whole population and, consequently, to the estimation of nongeneralisable epidemiological parameters. As reported by Kaunisto et al. [28], it is essential to know the extent and demographic structure of the population covered by registries or assisted by the physicians, in order to properly estimate the denominator for the epidemiological calculations. As mentioned previously, demographic structures also affect comparability among estimates, and therefore, authors should provide age- and gender-specific rates, in order to standardise the results.

From a statistical point of view, once again, comparisons would require a variability measure for the estimated rates. Finally, age limitations should be carefully checked.

Studies based on clinical databases versus studies based on administrative databases

To conclude this review, we would like to describe further issues that might complicate the comparison between studies which use administrative and clinical data.

First of all, the disease onset estimated in studies based on administrative databases is likely to be further from the real one, as opposed to that estimated in the other group of studies. Therefore, the age distribution of incident cases might differ in the two groups of studies due to the study design. For example, age at onset might be systematically lower in clinical studies, because IPF patients access inpatient or outpatient services after the first physician diagnosis. In addition to this, clinical studies identify patients at the first diagnosis made by the physician [15, 16, 21–23, 27, 33], while studies based on administrative data identify the patient at his or her first outpatient or inpatient claim [12, 14, 19, 25, 34]. If a subject has no such claims, this results in an underestimation of the number of incident cases in studies based on administrative data.

Finally, the ATS/ERS guidelines used to identify the patient are usually stated in clinical studies [20, 22, 27, 33], while they are unknown in studies based on administrative databases, because, as previously mentioned, it is impossible to know when each guideline was actually implemented. This inevitably leads to the selection of nonhomogenous patients.

Discussion

In this review we have outlined the differences in the various studies on IPF epidemiology. We tried to understand if the differences in results between studies are true differences or if they actually come from the different methodologies used or are due to the difficulty in comparing results in this type of study. The relative rarity of IPF has constituted a challenge for investigators with an interest in its epidemiology and, before 1990, it had discouraged the performance of large scale epidemiological studies [8]. The study of

epidemiology in IPF has been difficult for many other reasons. First, the definition of the disease has changed over time. Secondly, the diagnosis codes used to identify IPF in the administrative databases, the most common approach used in IPF epidemiology studies, has also changed over time. Moreover, the ICD-9-CM code 516.3 is not specific for IPF and may include other IIPs, such as nonspecific interstitial pneumonia or acute interstitial pneumonia. For studies based on administrative databases, it should be stressed that such databases are not universally available or easy to access for epidemiological analyses. For example, in some countries legal issues might impede the use of health administration data for research purposes.

The differences that exist between studies, regarding the methodology employed for the determination of incidence and prevalence of IPF, make comparison very difficult. The result of this situation is that the true incidence and prevalence of IPF cannot be properly established, as they vary across studies [39]. In recent years, more studies have been published which focus on epidemiology of IPF in Europe, USA and Asian countries and, despite the differences between studies, some important information has been acquired.

IPF is a disease of older age and is prevalent in males. All the epidemiological studies agree on this aspect of the disease [11–14, 21, 23–25, 27, 33, 34] and confirm the results from clinical studies [4–7]. The only exception was the study conducted in Norway, where 55% of the incident cases were women. Norwegian women were amongst the heaviest smokers in the world and smoking is known to be risk factor for the development of IPF, so this may explain the predominance of females observed in this study [19].

The mean survival of an IPF patient is 3–5 years from different studies [13, 14].

Generally, the prevalence and incidence of IPF are higher in American studies [11–14] than in European [15, 16, 18–20, 22–25, 32] and Asian studies [26, 27, 34]. Although it is unclear whether this is due to true differences in geographic and ethnic aspects or simply due to methodological differences in the way the studies were conducted (as discussed earlier).

Generally, studies based on administrative databases involve a larger population but data re-evaluation and validation is more difficult, when compared with studies based on clinical data that involve a smaller group of patients but with more controlled data.

There seems to be geographic variability in IPF epidemiology because some differences have been observed across regions using similar methodologies [18, 23, 24]. These differences may be due to environmental exposure or ethnic background [40–42]. For example, important differences have been observed in the causes of death between a recent Japanese study and American studies. In the Japanese study, a total of 328 patients (59.3% of patients with IPF) died from various causes. Focusing on the sample of patients with IPF, the most common cause of death was an acute exacerbation of the disease (40% of the events), followed by lung cancer (11%), pneumonia (7%) and cardiovascular diseases (3%) [27]. The percentage of acute exacerbations was higher than in American studies [9, 13, 31, 43] and the authors suggested an ethnic difference, particularly between East Asian and Western populations.

In conclusion, epidemiological estimates from different studies cannot be directly compared without taking their heterogeneity into consideration and keeping in mind the numerous intervening elements. However, it is clear that an accurate understanding of the epidemiology of IPF is important in order to estimate the size of the problem. The existence of uniform, international diagnostic guidelines for a definitive IPF definition and of a clear systematic classification code system enables the collection of more accurate and comparable data in the future.

References

1   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.

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

3   King TE Jr., Pardo A, Selman M. Idiopathic pulmonary fibrosis. *Lancet* 2011; 378: 1949–1961.

4   Noble PW, Albera C, Bradford WZ, *et al*. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet* 2011; 377: 1760–1769.

5   King TE Jr., Bradford WZ, Castro-Bernardini S, *et al*. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2014; 370: 2083–2092.

6   Richeldi L, Costabel U, Selman M, *et al*. Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis. *N Engl J Med* 2011; 365: 1079–1087.

7   Richeldi L, du Bois RM, Raghu G, *et al*. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014; 370: 2071–2082.

8   Coultas DB, Hubbard R. Epidemiology of idiopathic pulmonary fibrosis. *In*: Lynch JP. Idiopathic Pulmonary Fibrosis. New York, Taylor & Francis, 2003; pp. 1–30.

9 Olson AL, Swigris JJ, Lezotte DC, *et al.* Mortality from pulmonary fibrosis increased in the United States from 1992 to 2003. *Am J Respir Crit Care Med* 2007; 176: 277–284.

10 Cherniack RM, Colby TV, Flint A, *et al.* Quantitative assessment of lung pathology in idiopathic pulmonary fibrosis. The BAL Cooperative Group Steering Committee. *Am Rev Respir Dis* 1991; 144: 892–900.

11 Coultas DB, Zumwalt RE, Black WC, *et al.* The epidemiology of interstitial lung diseases. *Am J Respir Crit Care Med* 1994; 150: 967–972.

12 Raghu G, Weycker D, Edelsberg J, *et al.* Incidence and prevalence of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2006; 174: 810–816.

13 Fernández Pérez ER, Daniels CE, Schroeder DR, *et al.* Incidence, prevalence, and clinical course of idiopathic pulmonary fibrosis: a population-based study. *Chest* 2010; 137: 129–137.

14 Raghu G, Chen SY, Yeh WS, *et al.* Idiopathic pulmonary fibrosis in US Medicare beneficiaries aged 65 years and older: incidence, prevalence, and survival, 2001–11. *Lancet Respir Med* 2014; 2: 566–572.

15 Kolek V. Epidemiology of cryptogenic fibrosing alveolitis in Moravia and Silesia, in the period 1981–1990. *Internista* 1995; 3: 105–108.

16 Thomeer M, Demedts M, Vandeurzen K. Registration of interstitial lung diseases by 20 centres of respiratory medicine in Flanders. *Acta Clin Belg* 2001; 56: 163–172.

17 Thomeer MJ, Costabel U, Rizzato G, *et al.* Comparison of registries of interstitial lung diseases in three European countries. *Eur Respir J* 2001; 18: Suppl. 32, 114s–118s.

18 Hodgson U, Laitinen T, Tukiainen P. Nationwide prevalence of sporadic and familial idiopathic pulmonary fibrosis: evidence of founder effect among multiplex families in Finland. *Thorax* 2002; 57: 338–342.

19 von Plessen C, Grinde O, Gulsvik A. Incidence and prevalence of cryptogenic fibrosing alveolitis in a Norwegian community. *Respir Med* 2003; 97: 428–435.

20 Xaubet A, Ancochea J, Morell F, *et al.* Report on the incidence of interstitial lung diseases in Spain. *Sarcoidosis Vasc Diffuse Lung Dis* 2004; 21: 64–70.

21 Tinelli C, De Silvestri A, Richeldi L, *et al.* The Italian register for diffuse infiltrative lung disorders (RIPID): a four-year report. *Sarcoidosis Vasc Diffuse Lung Dis* 2005; 22: Suppl. 1, S4–S8.

22 Karakatsani A, Papakosta D, Rapti A, *et al.* Epidemiology of interstitial lung diseases in Greece. *Respir Med* 2009; 103: 1122–1129.

23 Gribbin J, Hubbard RB, Le Jeune I, *et al.* Incidence and mortality of idiopathic pulmonary fibrosis and sarcoidosis in the UK. *Thorax* 2006; 61: 980–985.

24 Navaratnam V, Fleming KM, West J, *et al.* The rising incidence of idiopathic pulmonary fibrosis in the U.K. *Thorax* 2011; 66: 462–467.

25 Agabiti N, Porretta MA, Bauleo L, *et al.* Idiopathic pulmonary fibrosis (IPF) incidence and prevalence in Italy. *Sarcoidosis Vasc Diffuse Lung Dis* 2014; 31: 191–197.

26 Ohno S, Nakaya T, Bando M, *et al.* Idiopathic pulmonary fibrosis – results from a Japanese nationwide epidemiological survey using individual clinical records. *Respirology* 2008; 13: 926–928.

27 Natsuizaka M, Chiba H, Kuronuma K, *et al.* Epidemiologic survey of Japanese patients with idiopathic pulmonary fibrosis and investigation of ethnic differences. *Am J Respir Crit Care Med* 2014; 190: 773–779.

28 Kaunisto J, Salomaa ER, Hodgson U, *et al.* Idiopathic pulmonary fibrosis – a systematic review on methodology for the collection of epidemiological data. *BMC Pulm Med* 2013; 13: 53.

29 Hutchinson J, Fogarty A, Hubbard R, *et al.* Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. *Eur Respir J* 2015; 46: 795–806.

30 Behr J, Kreuter M, Hoeper MM, *et al.* Management of patients with idiopathic pulmonary fibrosis in clinical practice: the INSIGHTS-IPF registry. *Eur Respir J* 2015; 46: 186–196.

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

32 Hyldgaard C, Hilberg O, Muller A, *et al.* A cohort study of interstitial lung diseases in central Denmark. *Respir Med* 2014; 108: 793–799.

33 Musellim B, Okumus G, Uzaslan E, *et al.* Epidemiology and distribution of interstitial lung diseases in Turkey. *Clin Respir J* 2014; 8: 55–62.

34 Lai CC, Wang CY, Lu HM, *et al.* Idiopathic pulmonary fibrosis in Taiwan – a population-based study. *Respir Med* 2012; 106: 1566–1574.

35 Hutchinson JP, McKeever TM, Fogarty AW, *et al.* Increasing global mortality from idiopathic pulmonary fibrosis in the twenty-first century. *Ann Am Thorac Soc* 2014; 11: 1176–1185.

36 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.

37 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.

38 Boyle P, Parkin DM. Statistical methods for registries. *In*: Breslow NE, Day NE. Statistical Methods in Cancer Research. Lyon, International Agency for Research on Cancer, 1980.

39 Nalysnyk L, Cid-Ruzafa J, Rotella P, *et al.* Incidence and prevalence of idiopathic pulmonary fibrosis: review of the literature. *Eur Respir Rev* 2012; 21: 355–361.

40 Baumgartner KB, Samet JM, Coultas DB, *et al.* Occupational and environmental risk factors for idiopathic pulmonary fibrosis: a multicenter case-control study. Collaborating Centers. *Am J Epidemiol* 2000; 152: 307–315.

41 Miyake Y, Sasaki S, Yokoyama T, *et al.* Occupational and environmental factors and idiopathic pulmonary fibrosis in Japan. *Ann Occup Hyg* 2005; 49: 259–265.

42 Lederer DJ, Arcasoy SM, Barr RG, *et al.* Racial and ethnic disparities in idiopathic pulmonary fibrosis: a UNOS/OPTN database analysis. *Am J Transplant* 2006; 6: 2436–2442.

43 Panos RJ, Mortenson RL, Niccoli SA, *et al.* Clinical deterioration in patients with idiopathic pulmonary fibrosis: causes and assessment. *Am J Med* 1990; 88: 396–404.