Tracking patterns of disease incidence and mortality is fundamental to disease control. The story of lung cancer is exemplary; the 20th century epidemic of lung cancer was first detected in western countries as mortality rates rose; initially it affected males more than females; and rates were not uniform across regions and between countries [1]. The epidemiological studies, motivated by the changing patterns of lung cancer occurrence, identified smoking, and occupational and environmental agents as causes of this highly fatal malignancy. Lung cancer mortality rates, which are very close to incidence rates because of the high case-fatality rate of lung cancer, and incidence rates, tracked through cancer registries, have now fallen in the USA and much of Europe in response to successful tobacco control and reduction of exposures to occupational and environmental carcinogens. In the example of lung cancer, tracking of occurrence is facilitated by the close correspondence of mortality with incidence, such that mortality is a reasonable index of occurrence, by the long standing surveillance of cancer through population-based cancer registries in many countries of Europe and much of the USA [2], and by improved diagnosis of lung cancer by increasingly accurate diagnostic methods.

By contrast, for idiopathic pulmonary fibrosis (IPF), there has long been uncertainty about incidence and mortality rates and the associated disease burden [3, 4]. The sources of uncertainty include the complexities of establishing a diagnosis, lack of validity of the reported and/or noted diagnosis in medical records and physicians variation in diagnostic approaches over time and among clinicians, differences in utilisation of ICD codes and lack of specific ICD codes for true diagnosis of IPF, and variations across geographic regions and countries [5]. Unlike lung cancer, mortality rates are not a highly valid indicator of incidence [5] and sustained registries for IPF (and other lung diseases) are not in place.

In this issue of the European Respiratory Journal, Hutchinson et al. [6] report the findings of a systematic review on the published evidence available on IPF incidence and mortality. They used an appropriate approach to cull through the literature, finding 34 publications from 21 countries meeting selection criteria. The reports have inherent limitations associated with epidemiological studies and are mixed in their sources that include: 1) administrative databases, e.g. unvalidated insurance claim data; 2) local medical records; 3) questionnaires to survey medical providers; and 4) routine mortality data. Administrative datasets and mortality statistics have the advantage of availability, while inevitably limited by diagnostic misclassification and lack of standardisation. Questionnaire surveys are an imperfect tool for capturing new and prevalent cases. Studies based on uniform case identification and review of medical records have the potential to provide accurate information for the covered population, but such studies have not been maintained for substantial periods of time or in multiple populations. For example, Coultas
et al. [7] carefully ascertained all IPF cases in one county in the USA and subjected medical information to a standardised review; however, the project spanned only 3 years.

While the data captured by Hutchinson et al. [6] have these inherent and inevitable limitations and comparability among datasets cannot be assured, some informative patterns were identified. One useful finding was the description of the range of incidence rates, estimated at 3–9 cases per 100,000 per year in Europe and North America, and some evidence that incidence is rising. For comparison, the incidence of lung cancer among males in the USA is presently 70.1 per 100,000 (age-standardised to the US population) [8]. For some cancers, such as laryngeal cancer and cervical cancer, incidence rates are comparable to those of IPF at 3.2 and 7.7 per 100,000, respectively, reinforcing the proposition that IPF is an overlooked disease [8]. For Europe and North America, mortality rates for IPF overlapped with the range of incidence rates emphasised by Hutchinson et al. [6], providing further support for their “best guess”.

The evidence indicating a possible rise in occurrence across the four decades covered (approximately 1970–2010) is intriguing and could reflect increasing rates of detection and also a shift in the distribution of aetiological factors. Undoubtedly, there has been an increasing likelihood of diagnosis over this period, given increasing awareness on the part of physicians, the growth of pulmonary medicine, and the implementation of increasingly sensitive imaging techniques. In addition, an improved standard of care of patients with IPF may have contributed to lower mortality rates observed in recent clinical trials and thus, the observed increase in prevalence may in part be due to the apparent increased survival [9, 10]. By definition, the causes of IPF are unknown, but epidemiological studies have identified a number of potential environmental risk factors with the strongest evidence for cigarette smoking besides genetic risks and risks associated with microaspiration from gastro-oesophageal reflux [1, 4, 5].

Above all, the systematic review by Hutchinson et al. [6] highlights the need for systematic tracking of the occurrence of IPF. Among the approaches for doing so, administrative databases and electronic health records that are maintained for large populations, e.g. a national healthcare system or other delivery systems, combined with mortality data may represent the most feasible alternatives, particularly as electronic health records gain richness and specificity and tools for using them become more sophisticated in search capabilities. Already, such databases have been used, including those for Medicare beneficiaries and other health insurers in the USA [10–12], residents of Olmsted County in the USA which is covered by the Mayo Clinic [13], and from a sample of general practitioners in the UK [14], for example. If such databases are used for tracking the occurrence of IPF, nested studies could be implemented for validation of diagnoses and for making adjustments for errors in classification.

Basic epidemiological information has long been lacking for IPF, as documented by Hutchinson et al. [6]. Aetiological, genetic, mechanistic and clinical trial research has provided some insights into the course of the disease, but insufficient attention has been given to assuring that we understand population patterns of IPF occurrence and the associated burden of disease [15]. It will only be through ongoing large-scale epidemiological investigations that determination of the effectiveness of recently approved efficacious interventions for IPF or development of other potential strategies for disease control through primary or secondary prevention will be possible [16, 17]. The findings of Hutchinson et al. [6] document that IPF should not be dismissed as merely a rare disease and that better ways of reporting/documenting the ascertained diagnosis of IPF utilising 2011 diagnostic criteria and new approaches for monitoring are needed, and this should be established based around large databases and emerging tools for clinical informatics [5].

References