CORRESPONDENCE

Longitudinal lung function in idiopathic NSIP: what are we missing?

To the Editors:

In a recent article, PARK *et al.* [1] described the clinical course and changes in lung function of 83 patients classified as idiopathic nonspecific interstitial pneumonia (NSIP). PARK *et al.* [1] found that 81% of patients with fibrotic NSIP had improved or stable lung function over time after initial treatment. Specifically, the authors showed an increase in the percentage of patients with an improvement in forced vital capacity and diffusing capacity of the lung for carbon monoxide over time, as well as an increase in the magnitude of change in observed lung function. However, due to the lack of complete longitudinal data for all patients, such results must be interpreted with caution.

Missing data is a problem that plagues all statistical analyses, particularly those involving longitudinal data from an observational study. When ''missingness'' of a variable depends on the value of that variable, this can lead to systematic bias [2]. In the case of the study by PARK *et al.* [1], those patients with worse lung function were more likely to die, thereby precluding any subsequent assessment of their lung function. This may explain, in part, the continued ''improvement'' in lung function observed over time, as these sicker patients drop out of the cohort. This effect has been widely described, both in medicine [3] and in other fields of study [4], and is sometimes referred to as ''survivor bias'' or ''survivorship bias''.

An alternative data analysis approach to that presented by PARK *et al.* [1] would be to perform a ''complete case'' analysis, excluding those patients with incomplete lung function data. Although complete case analysis is not immune to bias [2], it may be helpful in testing the robustness of the reported findings. Another alternative would be to create a separate category for those patients with missing data, and then report the percentage of patients that were improved, stable, worsened and missing at each time-point with respect to the total number of patients at baseline. This would give a more conservative assessment of the clinical course of disease and illustrate the degree to which missing lung function data could influence the results. Assuming that deaths accounted for the majority of missing lung function data, it might also be reasonable to simply categorise the missing group together with those whose lung function worsened, resulting in a much lower percentage of subjects with improved lung function. Regardless of the analytical approach used, it is always valuable to describe the missing data population as fully as possible, so that differences in baseline features and survival may be appreciated.

We commend PARK *et al.* [1] for characterising the clinical course of patients with idiopathic NSIP, a heterogeneous condition for which long-term data is relatively lacking. Future prospective studies are clearly needed to further define the natural history of this disorder.

H. Chen, A. Gomez, A. Shum and H.R. Collard

Division of Pulmonary and Critical Care Medicine, University of California, San Francisco, CA, USA.

Correspondence: H. Chen, 350 Parnassus Avenue, Suite 609, San Francisco, CA 94143-0924, USA. E-mail: hubert.chen@ucsf.edu

REFERENCES

1 Park IN, Jegal Y, Kim DS, *et al.* Clinical course and lung function change of idiopathic nonspecific interstitial pneumonia. *Eur Respir J* 2009; 33: 68–76.
2 Little RJA, Rubin DB. Statistical Analysis with Missing Data. 2nd Edn. New York, John Wiley & Sons, 2002.
3 Glymour MM, Greenland S. Study design and conduct: Causal diagrams. *In*: Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd Edn. Philadelphia, Lippincott Williams & Wilkins, 2008; p. 198.
4 Elton EJ, Gruber MJ, Blake CR. Survivor bias and mutual fund performance. *Rev Fin* 1996; 9: 1097–1120.

From the authors:

We thank H. Chen and co-workers for the interest in our recent paper in the *European Respiratory Journal* [1], and for raising an important point with good suggestions.

We agree with their opinion, missing data is one of the biggest problems in retrospective studies and we should have discussed this in more detail. However, we still think that the missing data of our study did not markedly influence our results and conclusion [1].

Although there were many missing data in the pulmonary function tests (PFT), the clinical course of most patients were well documented. As stated in our study, 81% of patients had improved or were stable after the initial treatment [1]. Although follow-up PFT was variable, as the treatment duration was different for each case, PFT was performed at the time of completion of the initial treatment in all patients,