



LETTERS

Towards individualised medicine for airways disease: identifying clinical phenotype groups

To the Editors:

Asthma and chronic obstructive pulmonary disease (COPD) are heterogeneous diseases [1, 2]. In diagnosing a patient with asthma or COPD, an individual is thereby lumped together with other patients whose disease phenotype and response to treatment may be quite different. Doing so simplifies treatment guidelines and facilitates the application of evidence-based medicine but there is a risk that interventions that could provide benefit to certain disease subgroups are overlooked. At present, it is not known to what degree splitting of airways disease groups can lead to improved health outcomes and this question remains a focus of intense research. The approach in recent years has been to re-examine the classification of airways disease to identify disease subgroups that may respond to treatments in different ways.

There have been several examples that illustrate the potential for this. It has been shown in both asthma and COPD that the response to corticosteroids can be predicted by sputum eosinophilia [3, 4] and that lung volume reduction surgery is of more benefit to those with predominantly upper lobe emphysema [5].

More recently, there have been attempts to explore phenotypes with methods that are less reliant on *a priori* assumptions about the best way to split disease categories into clinically meaningful groups. Cluster analysis is a tool that can identify subsets of patients with airways disease who have similar characteristics. We have previously used cluster analysis to identify five phenotypes of airways disease based on nine key disease variables in a sample of adults selected at random from the New Zealand community [6]. The identification of these groups describes associations within the data but does not demonstrate that the groups are relevant clinically. To do this, potential phenotypes require prospective validation with clinical interventional trials [7].

Another limitation of cluster analysis is that one cannot prospectively determine which cluster an individual belongs to. It is therefore desirable to generate an allocation rule that allows an individual's phenotype to be identified at the point of study entry and in the clinic [8]. Classification and regression trees (CART) are exploratory techniques that can produce simple allocation rules for groups based on variables that describe individuals without any particular assumptions about statistical distributions. In the current study, we apply CART to our identified phenotype clusters as a proof of concept of the utility this technique.

Methods of the Wellington Respiratory Survey have been reported in detail elsewhere [6]. In brief, participants were

randomly selected from the New Zealand community and attended the respiratory physiology laboratory for an assessment that included detailed respiratory questionnaires, full pulmonary function testing, bronchodilator reversibility testing, exhaled nitric oxide analysis and blood tests. Participants with evidence of airways disease and complete data ($n=175$) were included in a cluster analysis that was carried out using the following variables: pre-bronchodilator forced expiratory volume in 1 s (FEV₁), post-bronchodilator change in FEV₁, pre-bronchodilator FEV₁/forced vital capacity ratio from baseline, functional residual capacity, diffusing capacity of the lung for carbon monoxide, serum immunoglobulin E concentration, exhaled nitric oxide fraction (FeNO), presence of chronic sputum production and lifetime tobacco cigarette consumption. Five clusters emerged using an agglomerative hierarchical clustering method and four clusters using a divisive method.

These clusters can be loosely described as follows. Cluster 1, $n=15$: severe overlap group with atopic asthma, chronic bronchitis and emphysema in smokers. Cluster 2, $n=14$: pure emphysema in smokers. Cluster 3, $n=30$: atopic asthma with eosinophilic airways inflammation. Cluster 4, $n=78$: mild airflow obstruction without eosinophilic airways inflammation. Cluster 5, $n=38$: simple chronic bronchitis in nonsmokers.

The allocation rules were developed using regression trees with the R statistical software (University of Vienna, Vienna, Austria) using the "tree" package with default settings and the cluster allocation that had previously been determined. Potential predictor variables were chosen from the original nine key variables for clinical relevance. The variables used were pack-years of tobacco smoke exposure, FEV₁, post-bronchodilator change in FEV₁, reversibility and FeNO as continuous variables and sputum production as a dichotomous variable. Cross-validation suggested that our choice of five splits in the four variables for the agglomerative clusters produced a reasonable compromise between reduction in deviance and misclassification rate.

The following allocation rules predicted cluster membership. Cluster 1: sputum producers with ≥ 16.9 pack-yr of tobacco smoke exposure. Cluster 2: non-sputum producers with FeNO < 47 ppb and either FEV₁ $< 51.8\%$ predicted or FEV₁ $\geq 51.8\%$ predicted and ≥ 25.6 pack-yr of tobacco smoke exposure. Cluster 3: non-sputum producers with FeNO ≥ 47.0 ppb. Cluster 4: non-sputum producers with FeNO < 47.0 ppb, FEV₁ $\geq 51.8\%$ predicted and < 25.6 pack-yr of tobacco smoke exposure. Cluster 5: sputum producers with < 16.9 pack-yr of tobacco smoke exposure.

Comparison between predicted cluster membership using these allocation rules and actual cluster membership from

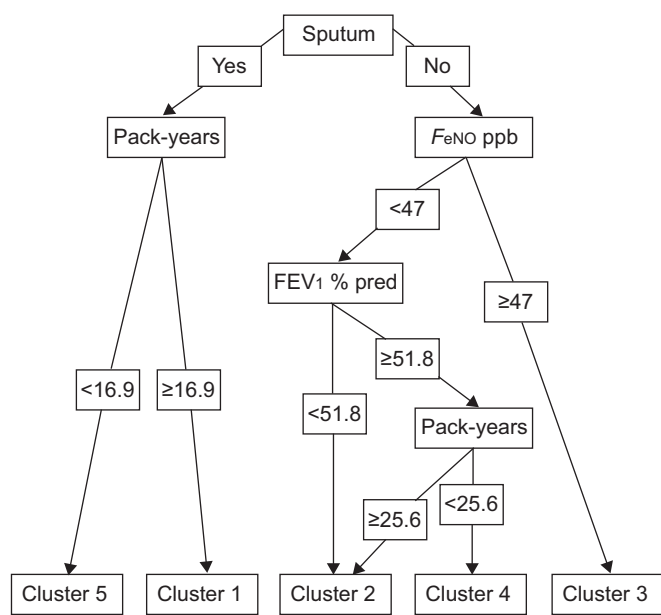


FIGURE 1. Allocation rules predicting cluster membership. Summary of previous study [6]. FeNO: exhaled nitric oxide fraction; FEV1: forced expiratory volume in 1 s; % pred: % predicted.

the original cluster analysis resulted in a misclassification rate of 19 (10.9%) out of 175. Allocation rules may be graphically represented as a decision tree, which is intuitive and easily applied in a clinical setting (fig. 1).

A separate CART analysis of four clusters identified in the same subjects using a divisive clustering algorithm [6] resulted in allocation rules based only on FEV1 and sputum production and a misclassification rate of only 4.6%.

This use of CART analysis illustrates the potential of this technique to help translate groups defined by cluster analysis into practically identifiable phenotype groups. This is a critical step if the insights into disease heterogeneity derived from techniques such as cluster analysis are to be translated into meaningful health benefits. An ideal allocation rule would be simple to administer, using only variables that could be collected in routine clinical care and yet accurately predict the phenotype for a particular patient [8].

There are several limitations inherent in deriving allocation rules using CART analysis, which the present report also demonstrates. CART analysis tends to over-fit the particular data set on which it is used and statistical tests of the value of the splitting variables are not available [9] so prospective validation is required. Putative groups identified by these allocation rules will never coincide exactly with the clusters identified from the original study but 89.1% of the subjects were correctly assigned. This compares favourably with an allocation rule described in a severe asthma cohort in the USA [10], where 80% of subjects were correctly allocated using a different CART analysis package based on three variables: pre-bronchodilator FEV1, post-bronchodilator FEV1 and age of onset of asthma.

The cluster analysis on which our allocation rules are based used several variables that are not often part of routine clinical practice. The resulting allocation rules rest on the measurement of FeNO,

which is not widely available in the clinical setting. However, FeNO was not required for allocating subjects into the alternative four cluster classification. The process of generating these rules does not take these practical considerations into account and ideally allocation rules would use only routinely available clinical information, but a simple rule risks missing important physiological information and an increase in misallocation.

The allocation rules presented here represent a proof of concept only. The groupings presented in figure 1 have not been prospectively evaluated for clinical relevance and should not be applied to clinical practice in any way.

CART analysis can be used to derive allocation rules that allow disease groups identified through cluster analysis to be prospectively identified in the real world. This will enable trials to test interventions in putative phenotypes, a necessary step towards personalised medicine for airways disease.

J. Travers*, **M. Weatherall*#**, **J. Fingleton*†** and **R. Beasley*+**

*Medical Research Institute of New Zealand, #University of Otago Wellington, †Victoria University of Wellington, and +Wellington Hospital, Capital and Coast District Health Board, Wellington, New Zealand.

Correspondence: R. Beasley, Medical Research Institute of New Zealand, Private Bag 7902, Wellington 6242, New Zealand. E-mail: Richard.Beasley@mrinz.ac.nz

Statement of Interest: None declared.

REFERENCES

- 1 Anderson GP. Endotyping asthma: new insights into key pathogenic mechanisms in a complex, heterogeneous disease. *Lancet* 2008; 372: 1107–1119.
- 2 Mannino DM. COPD: epidemiology, prevalence, morbidity and mortality, and disease heterogeneity. *Chest* 2002; 121: Suppl. 5, 121S–126S.
- 3 Berry M, Morgan A, Shaw DE, *et al.* Pathological features and inhaled corticosteroid response of eosinophilic and non-eosinophilic asthma. *Thorax* 2007; 62: 1043–1049.
- 4 Brightling CE, Monteiro W, Ward R, *et al.* Sputum eosinophilia and short-term response to prednisolone in chronic obstructive pulmonary disease: a randomised controlled trial. *Lancet* 2000; 356: 1480–1485.
- 5 Fishman A, Martinez F, Naunheim K, *et al.* A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med* 2003; 348: 2059–2073.
- 6 Weatherall M, Travers J, Shirtcliffe PM, *et al.* Distinct clinical phenotypes of airways disease defined by cluster analysis. *Eur Respir J* 2009; 34: 812–818.
- 7 Han MK, Agusti A, Calverley PM, *et al.* Chronic obstructive pulmonary disease phenotypes: the future of COPD. *Am J Respir Crit Care Med* 2010; 182: 598–604.
- 8 Fingleton J, Weatherall M, Beasley R. Towards individualised treatment in COPD. *Thorax* 2011; 66: 363–364.
- 9 Crawley MJ. The R Book. Chichester, John Wiley and Sons, 2007.
- 10 Moore WC, Meyers DA, Wenzel SE, *et al.* Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med* 2010; 181: 315–323.

DOI: 10.1183/09031936.00122811