Logistic regression analysis

These exploratory analyses were performed with either inclusion of all variables or stepwise forward selection. All study questions except dyspnea upon strong exertion and cough appeared as statistically significant predictors in at least one of the comparisons, in addition to age and sex. The functional measures identified as relevant comprised expiratory flow rates (MEF_{25}, MEF_{50}, MEF_{75}, MMEF), depending on the comparison. Due to collinearity, the results of the logistic regression analyses were not very consistent and robust against inclusion or exclusion of variables. We therefore focused on the other approaches that are better capable of describing complex relationships without the introduction of difficult-to-interpret interaction terms.

Further decision trees

Again using the CHAID algorithm, we established separate trees for asthma or COPD versus control (*Figures S1* and *S2*). All eight questions (wheezing in the last 12 months, self-diagnosed allergic rhinitis, dyspnea at strong or mild exertion, cough, phlegm, current smoker, ex-smoker), moreover all spirometric parameters (FEV_1, FVC, FEV_1/FVC, MEF_{25}, MEF_{50}, MEF_{75}, MMEF), as well as age, sex and BMI were offered to the algorithm which selected the optimal criteria at each node. To account for the dependence of lung function parameters on anthropometric characteristics, z-scores were used for the spirometric parameters. The diagnostic accuracy was 70.5% and 88.1%, respectively. We also used the whole group of subjects without asthma or COPD instead of the control group for comparison, i.e., the reference group included subjects with other respiratory diseases. Under this condition, the overall diagnostic accuracy for the tree was slightly lower (62.6%) but its structure and the variables involved remained the same.

The trees shown in *Figures S1* and *S2* were then used in an attempt to identify subjects with the diagnosis of both asthma and COPD, requiring that both trees indicated the respective diagnosis. Among the 34 patients, 32 (94%) were correctly identified as having asthma and COPD, however among the 659 patients having either asthma or COPD but not both, only 36% were correctly identified. This indicates that with the tools used in the present study that were limited to a minimal set a proper diagnosis of the combination of asthma and COPD cannot be achieved. For this purpose, more advanced procedures such as bronchial provocation tests or bronchodilator tests or the assessment of the concentration of exhaled nitric oxide (FeNO) is necessary.

 http://dx.doi.org/10.21037/jtd-20-3539

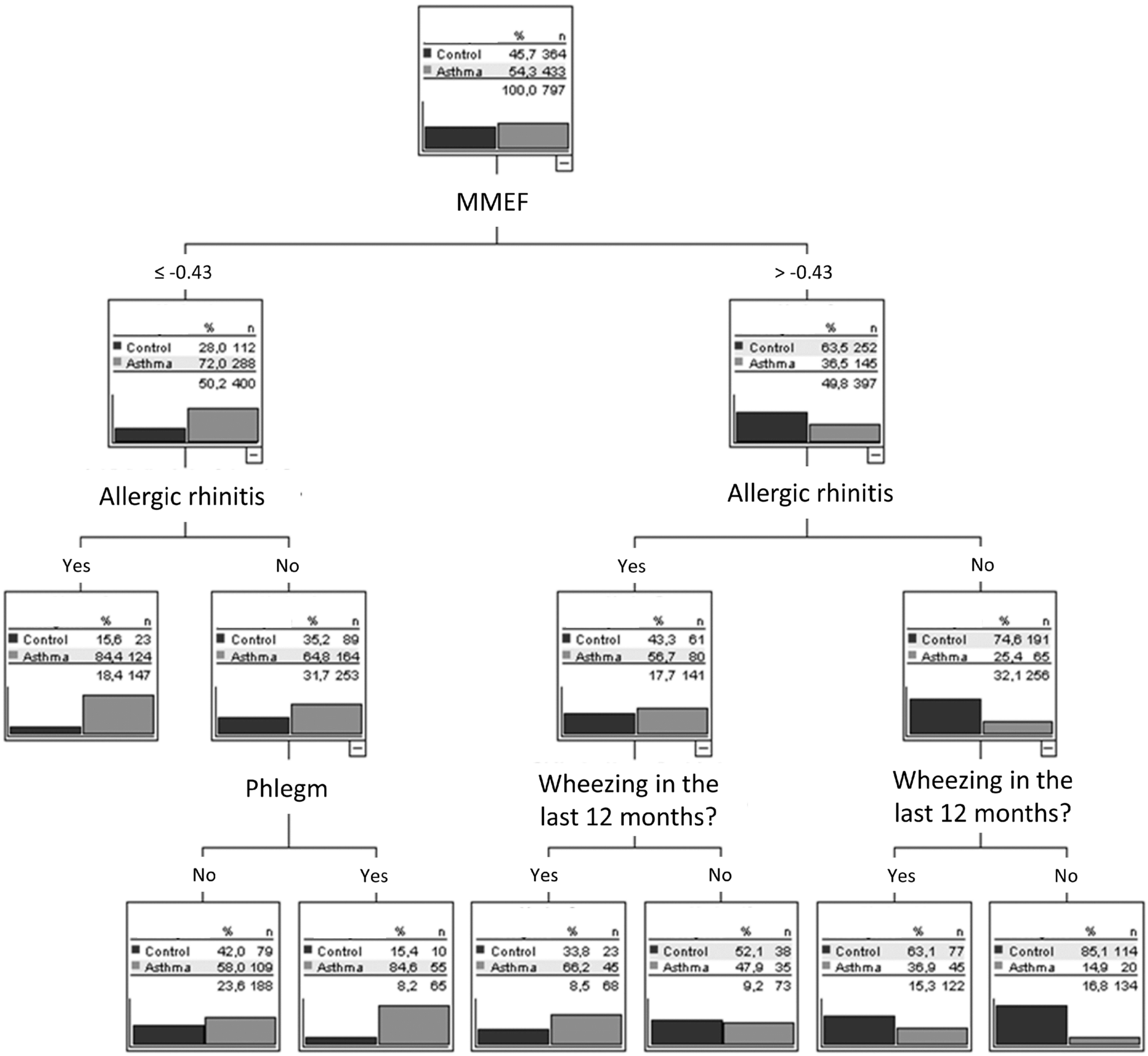

Figure S1 Decision tree for the comparison of asthma and control subjects obtained by the CHAID algorithm. The figure shows the average result of a 10-fold cross-validation. To account for the dependence of lung function parameters on anthropometric characteristics, z-scores were used for the spirometric parameters. FEV_1, forced expiratory volume in one second; FVC, forced vital capacity; MEF_{25}, MEF_{50}, MEF_{75}, mean expiratory flow rates at 25%, 50% and 75% of vital capacity; MMEF, maximal mid-expiratory flow rate.

 http://dx.doi.org/10.21037/jtd-20-3539

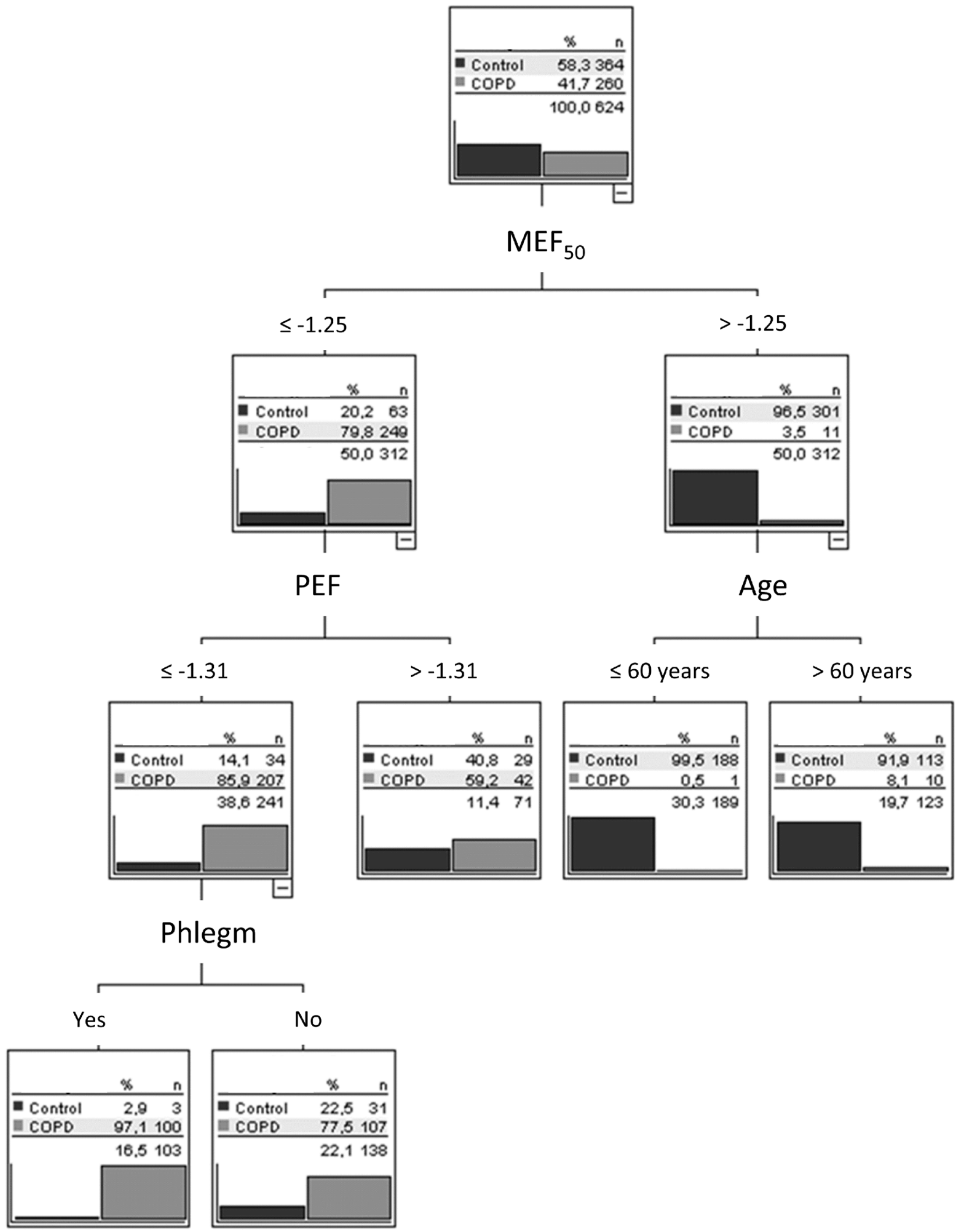

Figure S2 Decision tree for the comparison of COPD and control subjects obtained by the CHAID algorithm. The figure shows the average result of a 10-fold cross-validation. To account for the dependence of lung function parameters on anthropometric characteristics, z-scores were used for the spirometric parameters. FEV_1, forced expiratory volume in one second; FVC, forced vital capacity; MEF_{25}, MEF_{50}, MEF_{75}, mean expiratory flow rates at 25%, 50% and 75% of vital capacity; MMEF, maximal mid-expiratory flow rate; PEF, peak expiratory flow.

 http://dx.doi.org/10.21037/jtd-20-3539