



Published in final edited form as:

*Thorax*. 2011 October ; 66(10): 903–909. doi:10.1136/thx.2010.146118.

## The Impact of Nonlinear Smoking Effects on the Identification of Gene-by-Smoking Interactions in COPD Genetics Studies

P.J. Castaldi<sup>1,2</sup>, D.L. Demeo<sup>2</sup>, C.P. Hersh<sup>2</sup>, D.A. Lomas<sup>3</sup>, I.C. Soerheim<sup>4</sup>, A. Gulsvik<sup>4</sup>, P. Bakke<sup>4</sup>, Stephen Rennard<sup>5</sup>, Peter Pare<sup>6</sup>, Jørgen Vestbo<sup>7,8</sup>, AATGM Investigators, ICGN Investigators, and E.K. Silverman<sup>2</sup>

<sup>1</sup>Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, USA

<sup>2</sup>Channing Laboratory and Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA, USA

<sup>3</sup>Department of Medicine, University of Cambridge, Cambridge Institute for Medical Research, Cambridge, UK

<sup>4</sup>Haukeland University Hospital, Bergen, Norway

<sup>5</sup>University of Nebraska, Omaha, Nebraska

<sup>6</sup>Division of Respiratory Medicine and the James Hogg iCAPTURE Centre for Cardiovascular and Pulmonary Research, University of British Columbia, St. Paul's Hospital, Vancouver, Canada

<sup>7</sup>Department of Cardiology and Respiratory Medicine, Hvidovre Hospital, Copenhagen, Denmark

<sup>8</sup>Manchester Academic Health Sciences Centre, University of Manchester, Manchester, United Kingdom

### Abstract

**Background**—The identification of gene-by-environment interactions is important to understand the genetic basis of chronic obstructive pulmonary disease (COPD). Many COPD genetic association analyses assume a linear relationship between pack-years of smoking exposure and FEV<sub>1</sub>; however, this assumption has not been evaluated empirically in cohorts with a wide spectrum of COPD severity.

**Methods**—We examined the relationship between FEV<sub>1</sub> and pack-years of smoking exposure in 4 large cohorts assembled for the purpose of identifying genetic associations with COPD. Using data from the Alpha-1 Antitrypsin Genetic Modifiers Study, we compared the accuracy and power of two different approaches to model smoking by performing a simulation study of a genetic variant with a range of gene-by-smoking interaction effects.

**Results**—We identified nonlinear relationships between smoking and FEV<sub>1</sub> in 4 large cohorts. We demonstrated that in most situations where the relationship between pack-years and FEV<sub>1</sub> is

---

Corresponding Author: Peter Castaldi, M.D., Tufts Medical Center, 800 Washington St, Box 63, Boston, MA 02111, Phone: 617-636-7359, Fax: 617-636-5560.

#### LICENCE STATEMENT:

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in *Thorax* and any other BMJ PGL products to exploit all subsidiary rights, as set out in our licence (<http://group.bmj.com/products/journals/instructions-for-authors/licence-forms>).

#### COMPETING INTERESTS:

PDP served on the Advisory Board for Talecris Biotherapeutics and received grant support from GSK, Merck, \$100,001 or more, the NIH \$50,001–\$100,000, CIHR (Canada), and AllerGenNCE \$100,001 or more.

nonlinear, a piecewise-linear approach to model smoking and gene-by-smoking interactions is preferable to the commonly used total pack-years approach. We applied the piecewise linear approach to a genetic association analysis of the PI\*Z allele in the Norway case-control cohort and identified a potential PI\*Z-by-smoking interaction ( $p=0.03$  for FEV<sub>1</sub> analysis,  $p=0.01$  for COPD susceptibility analysis).

**Conclusion**—In study samples with subjects having a wide range of COPD severity, a nonlinear relationship between pack-years of smoking and FEV<sub>1</sub> is likely. In this setting, approaches that account for this nonlinearity can be more powerful and less-biased than the commonly-used approach of using total pack-years to model the smoking effect.

### Keywords

smoking; FEV<sub>1</sub>; gene-by-environment interaction; COPD; gene

---

COPD is well-suited to the study of gene-environment interaction, since the major environmental risk factor for COPD, cigarette smoking, is known and quantifiable. With the advent of large, well-powered genome-wide association studies in COPD, the identification of such interactions may be feasible. However, there are a number of challenges to the identification of gene-by-smoking interactions in COPD – the principal genetic risk factors for COPD are still in the process of being identified, a variety of approaches have been employed to model smoking effects, and there is no empiric knowledge of the nature, extent, or functional form of gene-by-smoking interactions in COPD.

While cigarette smoking is easily quantifiable in terms of pack-years ((average daily # of cigarettes smoked/20 cigarettes per pack) × years of smoking), previous work has demonstrated that pack-years alone may be an overly simplistic means of modeling smoking exposure, and nonlinear relations may be present.<sup>1,2</sup> Many COPD genetic association analyses model smoking effects by including a pack-years term in a regression model, which assumes a linear relation between pack-years and FEV<sub>1</sub> or, in the case of logistic regression for COPD status, a linear relation between pack-years and the log odds of having COPD. This practice is supported by seminal work on FEV<sub>1</sub> decline in general population samples.<sup>3-5</sup> However, it is not clear that these findings apply to the types of study samples typically assembled for COPD genetic association studies, namely cross-sectional samples that include subjects with a wide range of lung function impairment, including severe disease. In this setting, a number of factors may result in a nonlinear relation between pack-years and FEV<sub>1</sub>. These factors could include survival bias due to the well-demonstrated association between FEV<sub>1</sub> and mortality<sup>6</sup> and floor effects resulting from a diminished effect of cigarette smoking at very low levels of FEV<sub>1</sub>.

We hypothesized that the relation between FEV<sub>1</sub> and pack-years may be nonlinear in study samples with a wide range of airway obstruction and that, in this setting, methods of modeling smoking that account for nonlinearity may be more accurate and powerful for detecting gene-by-smoking interactions than the traditionally used pack-years approach. We tested this hypothesis in a cohort in which such nonlinear effects had been observed, by simulating a genetic variant with known main effects and gene-by-smoking effects. Finally, we assessed the performance of these modeling approaches in a gene-by-smoking analysis of the alpha-1 antitrypsin (AAT) PI MZ genotype in a case-control sample from Norway.

## METHODS

### Study Samples

We examined the relations between FEV<sub>1</sub> (percent of predicted) and pack-years of cigarette smoking in four large study samples (the Alpha-1 Antitrypsin Genetic Modifiers Study; the International COPD Genetics Network; the Boston Early-Onset COPD Study; and the Bergen, Norway Case-Control Study). The recruitment and inclusion criteria for these studies have been reported previously.<sup>7–10</sup> In brief, the Alpha-1 Antitrypsin Genetic Modifiers Study is a family-based study of individuals with the PI\*ZZ genotype. The International COPD Genetics Network and the Boston Early-Onset COPD Study are family-based studies in which families were identified through a proband affected with COPD. The Bergen, Norway Case-Control Study is a population-based study with a minimum required level of smoking exposure of 2.5 pack-years for both cases and controls. In each of the four studies, subjects underwent spirometric testing in accordance with ATS standards.<sup>11</sup>

### Relation of FEV<sub>1</sub> to Pack-Years

For each of the four studies, we generated scatterplots of the relation between FEV<sub>1</sub> and pack-years and drew smoothing curves through the data using a cubic spline fitting routine. All analyses were performed using SAS version 9.2 (Cary, NC).

### Simulation Studies

Using data from the Alpha-1 Antitrypsin Genetic Modifier Study, we simulated a randomly assigned, biallelic genetic variant in accordance with Hardy-Weinberg proportions. We conducted simulations under multiple scenarios, with each scenario characterized by a particular minor allele frequency, genetic main effect, and gene-by-smoking effect on FEV<sub>1</sub> percent predicted. For each scenario, we conducted 1,000 simulations. The range of allele frequencies was from 10–40%. The main effect of the gene was specified such that each copy of the minor allele decreased FEV<sub>1</sub> percent of predicted by 1 unit, and the gene-by-smoking interaction effect ranged from –0.45 to +0.45 units per-allele per-pack-year. For comparison, the main effect of pack-years in this dataset (after adjusting for age and sex) was approximately –1 unit per pack-year.

In each simulation, we calculated an estimated FEV<sub>1</sub> for each individual based on their observed FEV<sub>1</sub>, their simulated genotype, and the strength of the simulated genetic main effect and gene-by-smoking interaction effect. In our primary analysis, we assumed that the gene-by-smoking interaction effect followed the same nonlinear form as the smoking main effect. In each of our analyses, non-smokers were included in the analysis with a value of zero for the pack-years variable. A detailed description of our simulation methods is included in the Supplemental Materials.

Using linear regression, we estimated the genetic main effect and gene-by-smoking effect in each simulated dataset. We ran two regression models, one in which the smoking main effect and gene-by-smoking interaction were modeled using the pack-years approach (inclusion of a pack-years term in the regression equation), and another in which these effects were modeled with a piecewise linear approach (inclusion of separate variables to represent distinct intervals of smoking exposure). In each model, we adjusted for age and sex in addition to the smoking and genetic variables. We recorded the beta-coefficients from each model in each simulation, and we calculated the mean and standard deviation of these values. We quantified the bias of the two approaches by comparing the estimated values of the genetic main effects and gene-by-smoking effects to the actual values, and we estimated power by recording the number of times each beta coefficient was associated with a p-value < 0.05.

For the piecewise linear approach, we determined a breakpoint for the pack-years variable based on the shape of the relation between pack-years and FEV<sub>1</sub>. In the AAT Genetic Modifiers Study, which was the basis for these simulations, a break-point of 20 pack-years was selected based on visual inspection and improvement in model fit. The model fit of the piecewise linear model was compared to the pack-years model using the F-test. This breakpoint was used to code two variables, with one variable representing the first 20 pack-years of exposure and another variable representing all subsequent pack-years. The interaction term in the piecewise linear model included only the “piece” that was statistically significantly associated with FEV<sub>1</sub> in a multivariate context; thus, the interaction term was of the following form: first 20 pack-years of smoking × copies of minor allele.

### Gene-by-Smoking Analysis of the PI\*Z Allele in the Norway Case-Control Study

We compared the two approaches to model smoking in a gene-by-smoking analysis of the PI\*Z allele in the Norway Case-Control study data, using regression methods to test for genetic associations with FEV<sub>1</sub> level and COPD susceptibility (i.e. presence of absence of COPD). For the FEV<sub>1</sub> analysis, we applied sample weights to correct for oversampling of COPD cases, assuming a 10% prevalence of COPD in the general population. We performed one analysis using the traditional approach of modeling smoking with the pack-years approach, and we performed a similar analysis using a piecewise linear approach. Based on inspection and overall model fit for the FEV<sub>1</sub> model, we chose a breakpoint of 40 pack-years for the piecewise linear variable. We tested the main effect of the PI\*Z allele as well as the Z-by-smoking interaction.

### Alpha-1 Antitrypsin Typing

Phenotyping for the PI\*Z allele in the Norway Case-Control study was performed by isoelectric focusing. Individuals with severe AAT deficiency (PI\*Z, null-null, or SZ) were excluded from the Norway case-control study.

## RESULTS

The baseline characteristics of the four study samples are shown in Table 1. Each study had significant numbers of individuals with severe airflow obstruction, though the median FEV<sub>1</sub> level varied substantially between studies.

The relation between pack-years of smoking and FEV<sub>1</sub> (percent of predicted) in each of the study samples is shown in Figure 1. In each study sample, there was a nonlinear relation between FEV<sub>1</sub> and pack-years. For the two study samples in which piecewise linear modeling of smoking was performed (the Alpha-1 Antitrypsin Genetic Modifiers Study and the Norway Case-Control Study), the models with the piecewise linear smoking approach fit the data better than the models with the linear approach ( $p < 0.001$  in both instances). All of the study samples had a similar pattern of an initial strong negative effect of smoking with a subsequent decrease in the negative impact of additional pack-years on FEV<sub>1</sub> level. With the exception of the Norway study, there seemed to be a plateau phase at which additional pack-years were not associated with further FEV<sub>1</sub> decline. In all four samples, the slope of the FEV<sub>1</sub>-pack-years relation decreased at an FEV<sub>1</sub> level of approximately 30–50% of predicted. For three of the samples, this corresponded to a smoking exposure of 40–60 pack-years; however, in the more genetically susceptible Alpha-1 Antitrypsin Deficiency cohort, the leveling of the FEV<sub>1</sub>-pack-years relation occurred at approximately 20 pack-years exposure.

The results of the simulation study are displayed in Table 2 and Supplemental Figure 1. Under most of the simulated scenarios, the piecewise-linear approach yielded more accurate

estimates of genetic main effect size and gene-by-smoking interactions as compared to the pack-years approach. The direction of bias in the estimates generated by the pack-years approach was consistent with what would be expected from an approach that does not fully account for the strength of the gene-by-smoking interaction. When the genetic main effects and gene-by-smoking interaction effects were in the same direction (i.e., both main effect and interaction effect were negative), modeling with pack-years systematically overestimated the magnitude of the genetic main effect and underestimated gene-by-smoking interactions. When the genetic main effects and gene-by-smoking interaction effects were in opposing directions (i.e., main effect negative, interaction effect positive), modeling with pack-years underestimated both genetic main effects and gene-by-smoking interactions. Increasing the strength of the gene-by-smoking interaction led to more bias when pack-years was used to model smoking effects. While in some scenarios, the piecewise-linear approach to smoking yielded biased estimates, in almost all instances the bias was smaller than that of the pack-years approach, and this bias reached statistical significance in only a small number of scenarios.

Graphical depictions of power to detect gene-by-smoking effects are shown in Figure 2. In terms of power to detect gene-by-smoking interactions, the piecewise-linear approach was more powerful than the pack-years approach.

We conducted two sensitivity analyses to examine the robustness of our results. In one sensitivity analysis, we assumed a linear relation between pack-years and the strength of the gene-by-smoking effect (Supplemental Table 1). In this scenario, the piecewise linear approach was often comparable to or superior to the pack-years approach, though there were certain situations in which the pack-years approach performed better. To assess the impact of choice of cutpoint, we performed a sensitivity analysis in which we repeated our simulations using a range of cutpoints for the piecewise-linear transformation of pack years (Supplemental Table 2). As in the primary analysis, the underlying functional form of the gene-by-smoking interaction mirrored the form of the pack-years main effect. These results demonstrate that while the cutpoint of 20 pack-years in this dataset performs better than the extremes, it is difficult to identify a single cutpoint that performs best for genetic main and interaction effects across all scenarios.

We applied these traditional pack-years and piecewise linear methods to case-control candidate gene data, performing genetic association analyses for genetic main effects and gene-by-smoking effects of the PI\*Z allele in individuals from the Norway case-control COPD study. We tested for association between PI MZ and two outcomes, FEV<sub>1</sub> level (percent of predicted) and COPD susceptibility. The characteristics of PI MZ and PI MM subjects are shown in Table 3. There were no statistically significant differences between the two groups in age, gender, pack-years, or FEV<sub>1</sub>. A cutoff of 40 pack-years was selected for the piecewise-linear approach based on visual inspection and model fit. This cutoff was also supported in an examination of the relation between COPD susceptibility and pack-years (Supplemental Figure 2). Using linear regression, we tested for an association between PI genotype and pack-years that might confound the association between genotype and FEV<sub>1</sub> or the gene-by-smoking interaction, and we found no evidence for this association (unadjusted  $p=0.79$ ,  $p$  adjusted for age and sex = 0.96).

The results of these analyses are shown in Table 4. In both the FEV<sub>1</sub> and COPD susceptibility analyses, the main effect and Z allele-by-smoking effects are in opposite directions. In a manner that is consistent with our simulation results, the analyses using the piecewise-linear approach yielded stronger genetic main effect and Z allele-by-smoking interaction estimates than the pack-years approach. In both the FEV<sub>1</sub> and COPD susceptibility analyses, the piecewise linear approach demonstrates a statistically significant

gene-by-smoking effect of the Z-allele ( $p=0.03$  and  $0.01$ , respectively), whereas the pack-years approach did not identify any statistically significant interactions.

## DISCUSSION

We identified a nonlinear relation between smoking and FEV<sub>1</sub> in four large study samples. In simulation studies, we demonstrated that in some scenarios a piecewise linear approach to model smoking is superior to the commonly used pack-years approach in terms of accuracy and power to identify gene-by-smoking interactions. We applied this method in an analysis of the association of the PI MZ genotype with FEV<sub>1</sub> and COPD susceptibility, and we were able to detect statistically significant main and gene-by-smoking interaction effects with the piecewise linear modeling approach that would not have been detected with a pack-years approach. This pattern of results is consistent with the results of our simulations.

Previous work demonstrating a linear relation between FEV<sub>1</sub> and pack-years has generally focused on healthy population samples.<sup>12–16</sup> However, study samples recruited for many genetic association studies are specifically enriched for severe COPD, and our results demonstrate that the relation between pack-years and FEV<sub>1</sub> in these samples can be nonlinear and should be considered when performing gene-by-smoking interaction analyses. A similar nonlinear phenomenon in which risk tapers at higher levels of smoking exposure has been demonstrated with smoking intensity in lung cancer.<sup>17</sup> The two most likely explanations for the nonlinear relations we observed are 1) survival bias, i.e. differential population sampling at higher levels of cigarette exposure, and 2) a physiologic floor in FEV<sub>1</sub> which, once it is reached, results in diminished FEV<sub>1</sub> response to additional cigarette exposure. If these two mechanisms are active, the data points of most interest would be those that occur prior to the plateau phase in the FEV<sub>1</sub>-pack-years relation, since the points on the plateau portion of the curve are likely to be affected either by survival bias or floor effects that may act to dilute the strength of any observed gene-by-smoking interactions. An additional problem with pack-years data is the potential for recall bias, particularly for individuals with extensive smoking histories or for those who have stopped smoking many years prior to the time of smoking ascertainment. If this bias increases with pack-years exposure, it could dilute the association between pack-years and FEV<sub>1</sub> at the extreme end of the pack-years distribution. In the cross-sectional data used in this study, it is difficult to distinguish between these explanations. Further study of this topic using longitudinal data would be useful, though survival bias can affect longitudinal analyses as well.<sup>18</sup> It should also be noted that a nonlinear relation between pack-years and FEV<sub>1</sub> may result from occult interactions of pack-years with other variables. Thus, our proposed modeling approach may not necessarily reflect the true underlying relationship between FEV<sub>1</sub> and other important covariates.

In our analysis of the PI\*Z allele-by-smoking interaction, we noted opposing directions of the main effect of the PI\*Z allele and the PI\*Z-by-smoking interaction. This result suggests that the deleterious effects of the PI\*Z allele may become less prominent as smoking exposure increases. These results are consistent with a previously published report noting increased susceptibility to emphysema in PI\*MZ individuals compared to PI\*MM individuals that was limited to the low-smoking exposure subgroup.<sup>19</sup> It is possible that for individuals with an increased genetic susceptibility to COPD, this difference is most notable at relatively low levels of smoke exposure, and as the smoking burden increases this relative difference becomes more difficult to detect.

Our study has the following strengths. First, we demonstrated the phenomenon of nonlinearity between FEV<sub>1</sub> and pack-years in four large study samples. Second, our simulation strategy allowed us to compare the accuracy and power of two different



approaches to model smoking in a setting in which the true values of genetic main and gene-by-smoking effects were known. Since our simulations were based on actual data, we preserved the natural noise present in FEV<sub>1</sub> measurements. Third, we were able to take the findings of our simulated studies and test them in a genetic association analysis of candidate gene data. Our findings are in line with previous results.<sup>20</sup> The main effects OR of the PI MZ genotype from the piecewise linear analysis for COPD susceptibility is comparable to a recent cumulative meta-analysis estimate, and the OR obtained using the total pack-years approach to these data is within the 95% CI limits of the meta-analysis estimate, suggesting that our sample is comparable to those of other PI MZ studies. Finally, our sample size compares favorably to most previous genetic association studies of PI MZ individuals.

One of the limitations of our study is that we have taken a simple approach, i.e. piecewise linear modeling, to modeling the observed non-linearity of the smoking main effect, but a number of other modeling options could have been pursued, such as multivariate adaptive regression splines (MARS) or generalized additive models. MARS incorporate piecewise linear modeling approaches similar to those used in this study, but MARS automates the cutpoint selection and model building process. MARS is more extensive and explicit in its modeling algorithms, but can also require more degrees of freedom than our manual piecewise linear approach. Generalized additive models can fit highly non-linear curves to data in a piecewise fashion, but interpretation and hypothesis testing for covariates in these models is not straightforward. We also examined transforming the pack-years variable with packs-squared and inverse transformations, but these did not fit the data as well as the piecewise linear approach. Since our purpose was primarily to explore the implications of non-linearity of smoking main effects on the identification of gene-by-smoking interactions, the simplicity and interpretability of the piecewise linear approach were better suited for these purposes. As such, this method is a useful means of demonstrating the potential importance of nonlinear smoking effects for COPD genetic association analyses, but further work is required to identify the optimal approach or set of approaches for handling such nonlinear effects in large scale genetic association analyses.

There are also other sources of complexity to consider regarding the identification of genetic interactions in the setting of non-linear effects that have not been fully explored in this paper. We assume that the functional form of the gene-by-smoking interaction mirrors that of the smoking main effect, but there are no empiric data available regarding the true functional form of gene-by-smoking interactions in COPD, and it is possible that the functional form may vary across different genetic variants. As more COPD-associated variants are identified, more empirical data regarding the form of gene-by-smoking interactions will become available. In addition, while our results support the concept that better fit for the smoking main effect can reduce bias in the gene-by-smoking interaction term, identification of the optimal method for selecting cutpoints for the piecewise linear variable requires further exploration.

A further limitation of our study is that it used self-reported smoking history. It is likely that this is relatively accurate for the interval of smoke exposure. It is much less clear how well it serves as a measure of exposure.<sup>21</sup> Smokers vary greatly in their smoking behavior. The exposure to smoke-derived toxins, therefore, can vary greatly from one smoker to the next despite similar numbers of cigarettes smoked. In addition, smoke chemistry is exceedingly complex.<sup>22</sup> Changes in smoke topography, i.e. the way in which a cigarette is smoked that includes puff volume, puff time, dwell time and number of puffs per cigarette all have profound effects on toxin exposure.<sup>23</sup> Even within a single individual, cigarettes are smoked differently and yields of toxin will vary, and it is likely that there will be differential exposure among the many toxins contained in smoke.<sup>24</sup> At present, there are limited means

to measure exposure to specific smoke-derived toxins, but methodologies in this regard are advancing.

With the advent of large COPD GWAS studies, well-powered examinations for moderate to large gene-by-smoking interactions will be feasible, and gene-by-smoking interaction will likely be an important aspect of future COPD genetic association analyses. We demonstrate that in cross-sectional data of populations with a wide range of airflow obstruction, nonlinear relations between FEV<sub>1</sub> and pack-years may be observed. In these situations, a piecewise linear approach to model the smoking main effect and gene-by-smoking interactions is preferable to modeling smoking as total pack-years since it reduces bias and can be more powerful for detecting gene-by-smoking interactions.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

EKS received grant support and consulting fees from GlaxoSmithKline for studies of COPD genetics. EKS received honoraria and consulting fees from AstraZeneca. SR has consulted or participated in advisory boards for: Able Associates, Adelphia Research, Almirall/Prescott, APT Pharma/Britnall, Aradigm, AstraZeneca, Boehringer Ingelheim, Chiesi, Common Health, Consult Complete, COPDForum, DataMonitor, Decision Resources, Defined Health, Dey, Dunn Group, Eaton Associates, Equinox, Gerson, GlaxoSmithKline, Infomed, KOL Connection, M. Pankove, MedaCorp, MDRx Financial, Mpex, Novartis, Nycomed, Oriel Therapeutics, Otsuka, Pennside Partners, Pfizer (Varenicline), PharmaVentures, Pharmaxis, Price Waterhouse, Propagate, Pulmatrix, Reckner Associates, Recruiting Resources, Roche, Schlesinger Medical, Scimed, Sudler and Hennessey, TargeGen, Theravance, UBC, Uptake Medical, VantagePoint Management. SR has given lectures for: American Thoracic Society, AstraZeneca, Boehringer Ingelheim, California Allergy Society, Creative Educational Concept, France Foundation, Information TV, Network for Continuing Ed, Novartis, Pfizer, SOMA. SR has received industry-sponsored grants from: AstraZeneca, Biomarck, Centocor, Mpex, Nabi, Novartis, Otsuka.

DAL has received grant support, consultancy fees and honoraria from GlaxoSmithKline, consultancy fees from Talecris Biotherapeutics, Genzyme and Amicus Therapeutics and honoraria from LKB.

Jørgen Vestbo has received honoraria for consulting and presenting for pharmaceutical companies with an interest in COPD. He is also an investigator on the ECLIPSE study and the International COPD Genetics Network, both sponsored by GlaxoSmithKline. The other authors declare no competing interests.

### FUNDING:

The authors were supported by the following grants: K08HL102265, UL1 RR025752, R01 HL084323, R01 HL075478, U01 089856, and P01 HL083069. The International COPD Genetics Network is funded by a grant from GlaxoSmithKline.

## Reference List

1. Rachtel B, Siemiatycki J, Abrahamowicz M, Leffondre K. A flexible modeling approach to estimating the component effects of smoking behavior on lung cancer. *Journal of Clinical Epidemiology*. 2004; 57(10):1076–1085. [PubMed: 15528059]
2. Hoffmann K, Bergmann MM. Re: “Modeling smoking history: a comparison of different approaches”. *American Journal of Epidemiology*. 2003; 158(4):393–394. [PubMed: 12915508]
3. Dockery DW, Speizer FE, Ferris BG Jr, Ware JH, Louis TA, Spiro A III. Cumulative and reversible effects of lifetime smoking on simple tests of lung function in adults. *American Review of Respiratory Disease*. 1988; 137(2):286–292. [PubMed: 3341623]
4. Burrows B, Knudson RJ, Cline MG, Lebowitz MD. Quantitative relationships between cigarette smoking and ventilatory function. *American Review of Respiratory Disease*. 1977; 115(2):195–205. [PubMed: 842934]



5. Xu X, Dockery DW, Ware JH, Speizer FE, Ferris BG Jr. Effects of cigarette smoking on rate of loss of pulmonary function in adults: a longitudinal assessment. *American Review of Respiratory Disease*. 1992; 146(5 Pt 1):1345–1348. [PubMed: 1443894]
6. Stavem K, Aaser E, Sandvik L, Bjornholt JV, Erikssen G, Thaulow E, et al. Lung function, smoking and mortality in a 26-year follow-up of healthy middle-aged males. *European Respiratory Journal*. 2005; 25(4):618–625. [PubMed: 15802334]
7. DeMeo DL, Sandhaus RA, Barker AF, Brantly ML, Eden E, McElvaney NG, et al. Determinants of airflow obstruction in severe alpha-1-antitrypsin deficiency. *Thorax*. 2007; 62(9):806–813. [PubMed: 17389752]
8. Patel BD, Coxson HO, Pillai SG, Agusti AG, Calverley PM, Donner CF, et al. Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. *American Journal of Respiratory & Critical Care Medicine*. 2008; 178(5):500–505. [PubMed: 18565956]
9. Silverman EK, Chapman HA, Drazen JM, Weiss ST, Rosner B, Campbell EJ, et al. Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. Risk to relatives for airflow obstruction and chronic bronchitis. *American Journal of Respiratory & Critical Care Medicine*. 1998; 157(6 Pt 1):1770–1778. [PubMed: 9620904]
10. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genetics*. 2009; 5(3):e1000421. [PubMed: 19300482]
11. American Thoracic Society. Standardization of Spirometry, 1994 Update. *American Journal of Respiratory & Critical Care Medicine*. 1995; 152(3):1107–1136. [PubMed: 7663792]
12. Fletcher C, Peto R. The natural history of chronic airflow obstruction. *British Medical Journal*. 1977; 1(6077):1645–1648. [PubMed: 871704]
13. Burrows B, Knudson RJ, Cline MG, Lebowitz MD. Quantitative relationships between cigarette smoking and ventilatory function. *American Review of Respiratory Disease*. 1977; 115(2):195–205. [PubMed: 842934]
14. Kerstjens HA, Rijcken B, Schouten JP, Postma DS. Decline of FEV1 by age and smoking status: facts, figures, and fallacies. *Thorax*. 1997; 52(9):820–827. [PubMed: 9371217]
15. Xu X, Dockery DW, Ware JH, Speizer FE, Ferris BG Jr. Effects of cigarette smoking on rate of loss of pulmonary function in adults: a longitudinal assessment. *American Review of Respiratory Disease*. 1992; 146(5 Pt 1):1345–1348. [PubMed: 1443894]
16. Dockery DW, Speizer FE, Ferris BG Jr, Ware JH, Louis TA, Spiro A III. Cumulative and reversible effects of lifetime smoking on simple tests of lung function in adults. *American Review of Respiratory Disease*. 1988; 137(2):286–292. [PubMed: 3341623]
17. Rachet B, Siemiatycki J, Abrahamowicz M, Leffondre K. A flexible modeling approach to estimating the component effects of smoking behavior on lung cancer. *Journal of Clinical Epidemiology*. 2004; 57(10):1076–1085. [PubMed: 15528059]
18. Xu X, Dockery DW, Ware JH, Speizer FE, Ferris BG Jr. Effects of cigarette smoking on rate of loss of pulmonary function in adults: a longitudinal assessment. *American Review of Respiratory Disease*. 1992; 146(5 Pt 1):1345–1348. [PubMed: 1443894]
19. Sorheim IC, Bakke P, Gulsvik A, Pillai SG, Johannessen A, Gaarder PI, et al. Alpha-1 antitrypsin PI MZ heterozygosity is associated with airflow obstruction in two large cohorts. *Chest*. 2010
20. Hersh CP, Dahl M, Ly NP, Berkey CS, Nordestgaard BG, Silverman EK. Chronic obstructive pulmonary disease in alpha1-antitrypsin PI MZ heterozygotes: a meta-analysis. *Thorax*. 2004; 59(10):843–849. [PubMed: 15454649]
21. Hatsukami DK, Benowitz NL, Rennard SI, Oncken C, Hecht SS. Biomarkers to assess the utility of potential reduced exposure tobacco products. *Nicotine & Tobacco Research*. 2006; 8(4):600–622. [PubMed: 16920658]
22. Borgerding M, Klus H. Analysis of complex mixtures--cigarette smoke. *Experimental & Toxicologic Pathology*. 2005; 57 (Suppl 1):43–73. [PubMed: 16092717]
23. O'Connor RJ, Kozlowski LT, Hammond D, Vance TT, Stitt JP, Cummings KM. Digital image analysis of cigarette filter staining to estimate smoke exposure. *Nicotine & Tobacco Research*. 2007; 9(8):865–871. [PubMed: 17654299]

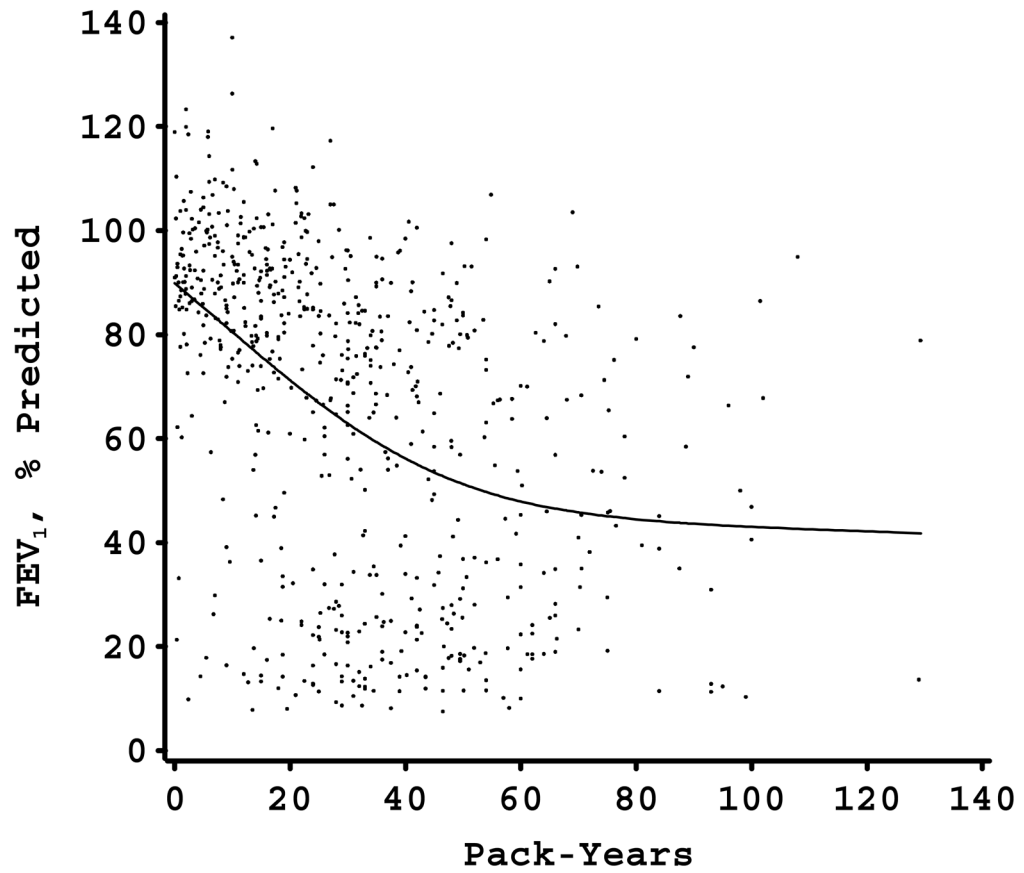
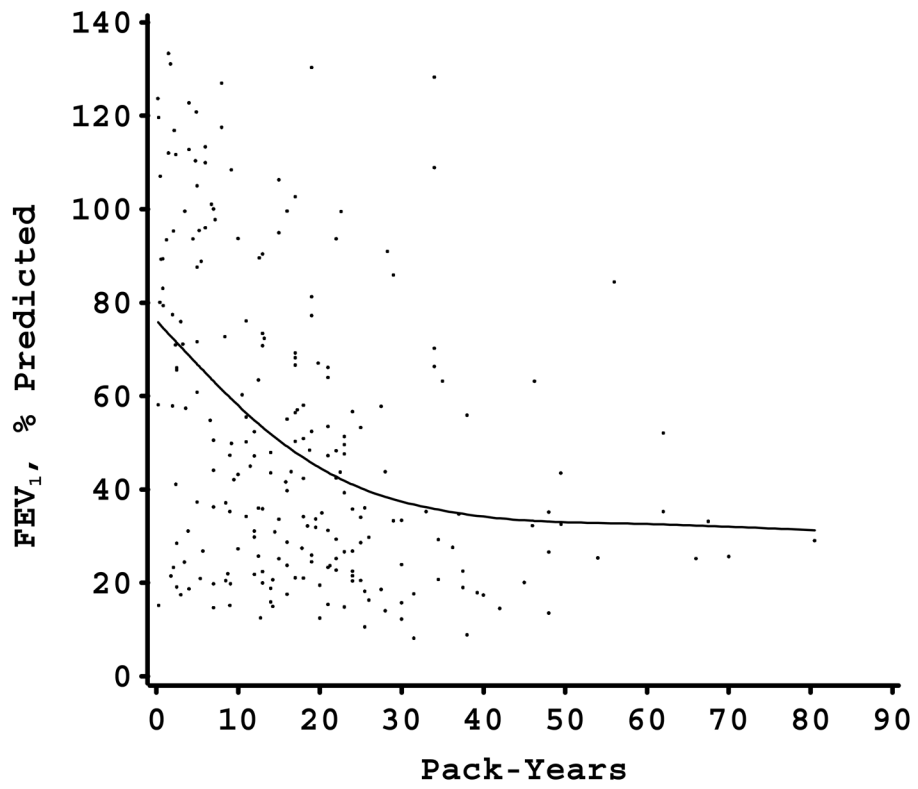
24. Mooney M, Green C, Hatsukami D. Nicotine self-administration: cigarette versus nicotine gum diurnal topography. *Human Psychopharmacology*. 2006; 21(8):539–548. [PubMed: 17066490]

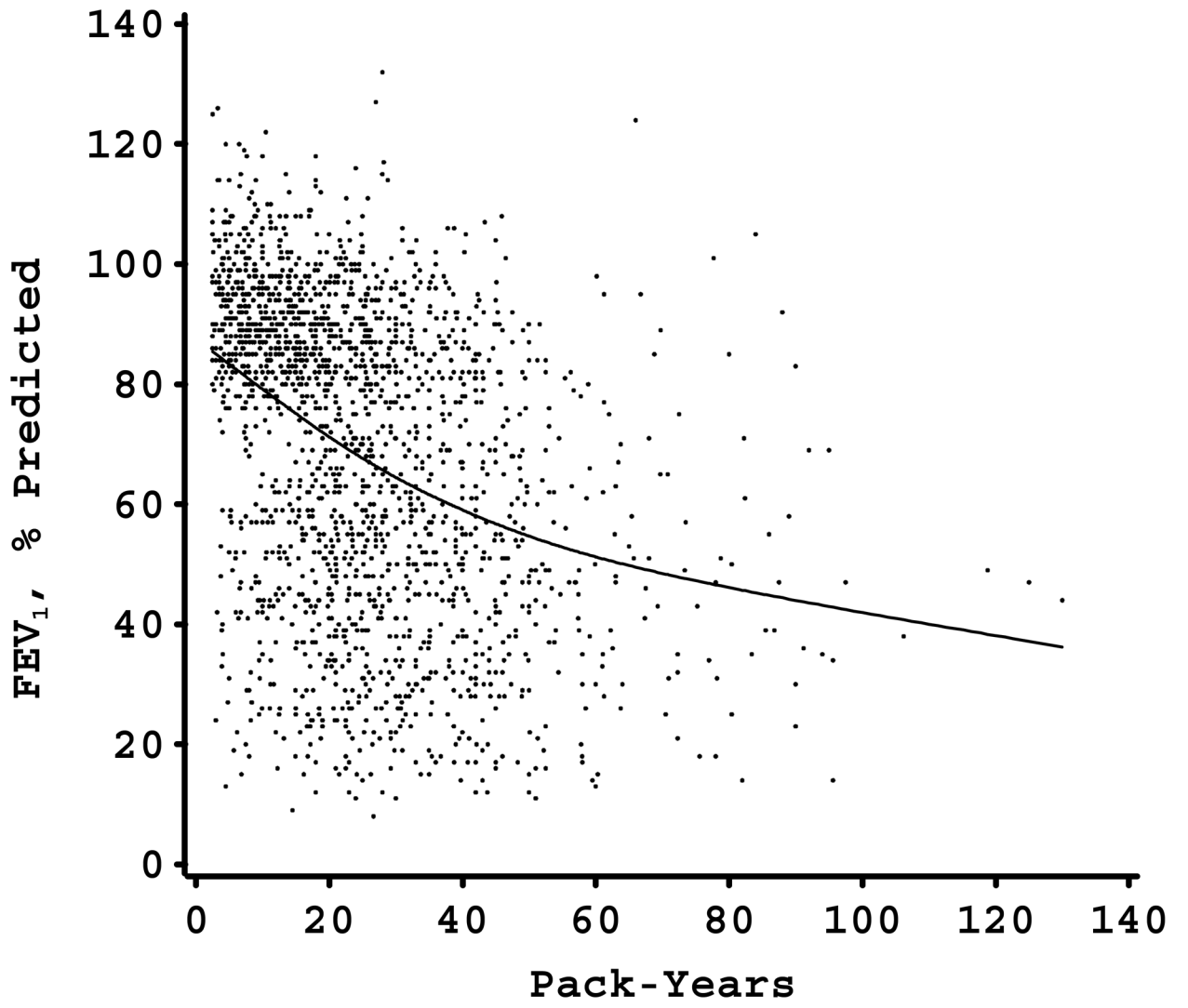
## Appendix

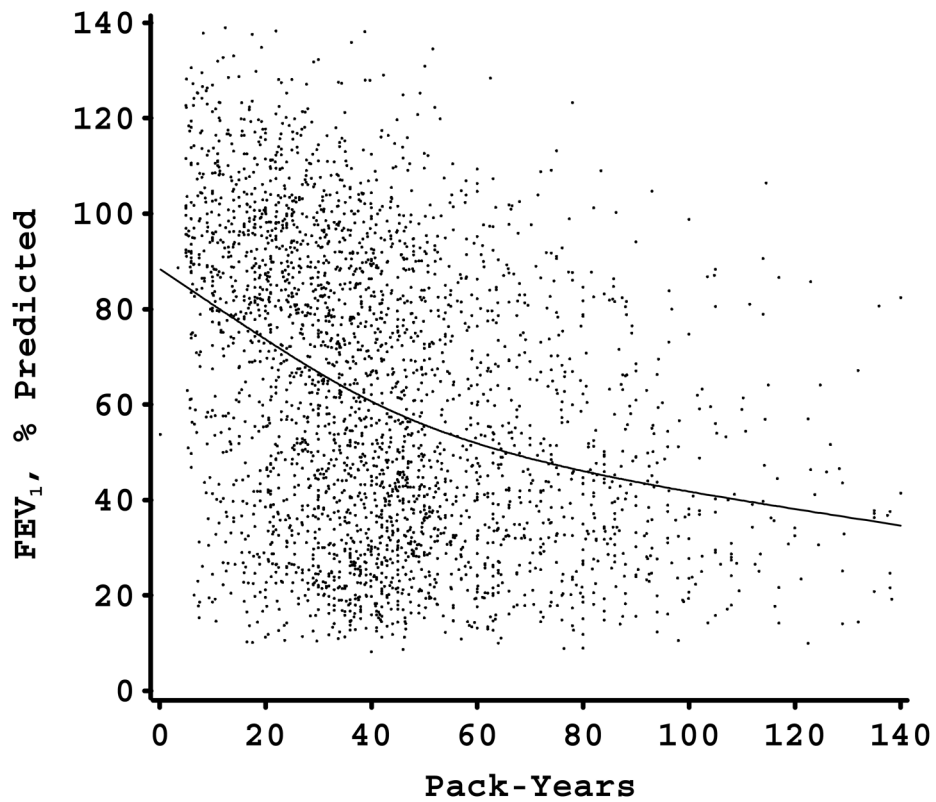
The ICGN (International COPD Genetics Network) Investigators are: Alvar Agusti (Hospital Universitari Son Dureta, Mallorca, Spain), Peter Calverley (University of Liverpool, Liverpool, UK), Claudio F. Donner (S. Maugeri Foundation, Veruno, Novara, Italy), Robert D. Levy (James Hogg iCAPTURE Centre, University of British Columbia, Vancouver), David Lomas (University of Cambridge, Cambridge, UK), Barry J. Make (National Jewish Health, Denver, Colorado), Wayne Anderson (GlaxoSmithKline, Research Triangle Park, North Carolina); Peter Pare (James Hogg iCAPTURE Centre, University of British Columbia, Vancouver), Sreekumar Pillai (GlaxoSmithKline, Research Triangle Park, North Carolina), Stephen Rennard (University of Nebraska, Omaha, Nebraska), Emiel Wouters (University Hospital Maastricht, Maastricht, The Netherlands), Edwin K Silverman (The Channing Laboratory and Pulmonary and Critical Care Division, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts), and Jørgen Vestbo (Hvidovre Hospital, Copenhagen, Denmark; <sup>8</sup> Manchester Academic Health Sciences Centre, University of Manchester, Manchester, United Kingdom).

The AATGM (Alpha-1 Antitrypsin Genetic Modifiers Study) Investigators are: Alan Barker (University of Oregon), Mark Brantly (University of Florida), Edward J. Campbell (Utah Valley Pulmonary Clinic), Edward Eden (St. Luke's/Roosevelt Hospital) N. Gerard McElvaney (Beaumont Hospital, Dublin), Stephen Rennard (University of Nebraska), Robert Sandhaus (National Jewish Health), Edwin K. Silverman (Brigham and Women's Hospital), James Stocks (University of Texas Health Center at Tyler), James Stoller (Cleveland Clinic), Charlie Strange (Medical University of South Carolina), Gerard Turino (St. Luke's/Roosevelt Hospital).

We thank John Ioannidis and David Kent for their discussions and input.

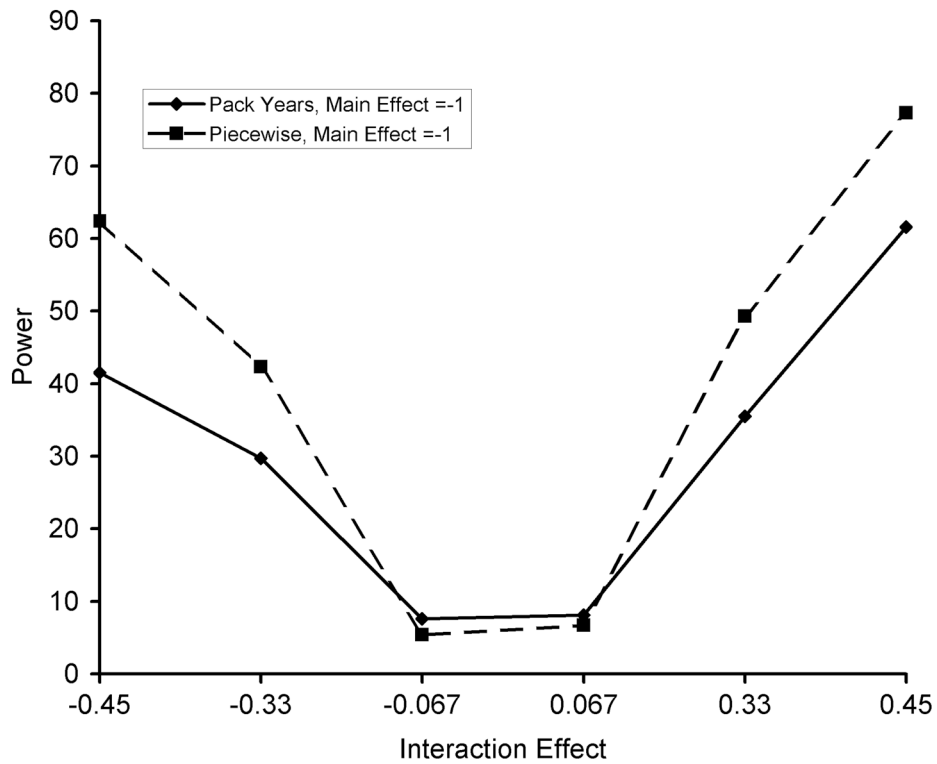






**Figure 1.**

Figures 1-D. FEV<sub>1</sub>, % predicted by pack-years scatterplot in 4 large cohorts (Panel A, Alpha-1 Antitrypsin Genetic Modifiers Study; Panel B, Boston Early-Onset COPD Study; Panel C, Norway Case-Control Study; Panel D, International COPD Genetics Network) with smoothing curve. Flattening of curve occurs at FEV<sub>1</sub> levels between 30 and 50% of predicted.



**Figure 2.** Observed power to detect gene-by-smoking interactions for the pack-years versus piecewise linear approach to model smoking. Simulation study based on data from the Genetic Modifiers of Alpha-1 Antitrypsin Disease Study. Simulation parameters are as follows: minor allele frequency = 25%, genetic main effect = -1 unit from observed FEV<sub>1</sub> percent predicted per copy of minor allele, gene-by-smoking effect varies as shown. For this power analysis, the threshold for detecting an effect was set at alpha < 0.05 for the null hypothesis that the gene-by-smoking effect is equal to zero. For gene-by-smoking effects the piecewise linear model is more powerful. At low values of the gene-by-smoking interaction, the total pack-years approach appears more powerful due to upwardly biased estimates of the gene-by-smoking interaction (values shown in Table 2).



**Table 1**

## Characteristics of Study Samples

	<b>AAT</b>	<b>EOCOPD</b>	<b>ICGN</b>	<b>Norway</b>
Subjects	372	972	3058	1909
Age (SD)	52 (10)	46 (18)	58 (8)	61 (11)
Female, (%)	202 (54)	567 (58)	1374 (45)	847 (44)
Ever Smokers, (%)	231 (62)	659 (68)	3058 (100)	1909 (100)
Pack-Years Smoking, median (IQR)	5 (0–19)	14 (0–35)	39 (25–55)	23 (13–34)
FEV <sub>1</sub> , % Predicted, median (IQR)	58 (33–93)	84 (60–96)	58 (35–87)	76 (48–90)

**Table 2**

Simulation Study Results of The Impact of Two Different Methods of Coding Smoking on the Accuracy of Estimation for Genetic Main Effect and Gene-by-Smoking Interaction Effects in the Alpha-1-Antitrypsin Genetic Modifiers Study. Simulations were performed for same direction and opposite direction main and interaction effects. Strength of interaction effect is based on the multivariate smoking main effect of -1 unit from FEV<sub>1</sub> percent predicted per pack-year.

Simulation Parameters	Regression Results <sup>†</sup>											
	Linear Pack Years						Piecewise Linear					
	Main Effects			G * S			Main Effects			G * S		
SNP Main Effect	G * S	MAF <sup>††</sup>	Bias <sup>‡</sup>	p-value <sup>‡</sup>	Bias <sup>‡</sup>	p-value <sup>‡</sup>	Bias <sup>‡</sup>	p-value <sup>‡</sup>	Bias <sup>‡</sup>	p-value <sup>‡</sup>	Bias <sup>‡</sup>	p-value <sup>‡</sup>
0	0	0.10	0.17	0.024	-0.02	<0.001	0.05	0.512	-0.004	0.523	-0.004	0.523
0	0	0.25	0.04	0.464	-0.01	<0.001	0.03	0.597	-0.01	0.040	-0.01	0.040
0	0	0.40	-0.07	0.265	0.01	0.002	-0.04	0.466	0.01	0.251	0.01	0.251
-1	0	0.10	-0.04	0.578	-0.01	0.029	-0.19	0.014	0.01	0.027	0.01	0.027
-1	0	0.25	0.04	0.464	-0.004	0.282	0.08	0.198	-0.01	0.061	-0.01	0.061
-1	0	0.40	0.05	0.382	0.002	0.520	0.08	0.205	-0.003	0.588	-0.003	0.588
-1	-0.07	0.10	-0.09	0.250	0.02	<0.001	-0.05	0.523	0.003	0.664	0.003	0.664
-1	-0.07	0.25	-0.18	0.002	0.03	<0.001	-0.02	0.770	0.001	0.839	0.001	0.839
-1	-0.07	0.40	-0.09	0.124	0.04	<0.001	0.11	0.054	0.003	0.503	0.003	0.503
-1	-0.33	0.10	-0.73	<0.001	0.16	<0.001	-0.11	0.157	0.03	<0.001	0.03	<0.001
-1	-0.33	0.25	-0.78	<0.001	0.18	<0.001	0.03	0.665	0.03	<0.001	0.03	<0.001
-1	-0.33	0.40	-0.83	<0.001	0.18	<0.001	-0.07	0.288	0.03	<0.001	0.03	<0.001
-1	-0.45	0.10	-0.87	<0.001	0.24	<0.001	-0.03	0.651	0.08	<0.001	0.08	<0.001
-1	-0.45	0.25	-1.24	<0.001	0.27	<0.001	-0.23	<0.001	0.08	<0.001	0.08	<0.001
-1	-0.45	0.40	-1.17	<0.001	0.27	<0.001	-0.15	0.015	0.08	<0.001	0.08	<0.001
-1	0.07	0.10	0.23	0.001	-0.04	<0.001	-0.04	0.601	0.01	0.305	0.01	0.305
-1	0.07	0.25	0.21	<0.001	-0.03	<0.001	0.004	0.943	0.004	0.457	0.004	0.457
-1	0.07	0.40	0.04	0.507	-0.02	<0.001	-0.05	0.429	<0.001	0.988	<0.001	0.988
-1	0.33	0.10	0.91	<0.001	-0.17	<0.001	-0.06	0.470	0.002	0.775	0.002	0.775
-1	0.33	0.25	0.94	<0.001	-0.17	<0.001	0.02	0.726	<0.001	0.919	<0.001	0.919

Simulation Parameters	Regression Results <sup>†</sup>											
	Linear Pack Years						Piecewise Linear					
	Main Effects			G * S			Main Effects			G * S		
SNP Main Effect	G * S	MAF <sup>††</sup>	Bias <sup>‡</sup>	p-value <sup>‡</sup>	Bias <sup>‡</sup>	p-value <sup>‡</sup>	Bias <sup>‡</sup>	p-value <sup>‡</sup>	Bias <sup>‡</sup>	p-value <sup>‡</sup>	Bias <sup>‡</sup>	p-value <sup>‡</sup>
-1	0.33	0.40	0.81	<0.001	-0.15	<0.001	0.01	0.821	<0.001	0.993	<0.001	0.993
-1	0.45	0.10	1.18	<0.001	-0.23	<0.001	-0.08	0.273	-0.004	0.459	-0.004	0.459
-1	0.45	0.25	1.26	<0.001	-0.22	<0.001	0.12	0.043	-0.01	0.059	-0.01	0.059
-1	0.45	0.40	1.16	<0.001	-0.22	<0.001	0.04	0.534	-0.002	0.756	-0.002	0.756

G \* S = gene-by-smoking interaction term. MAF = minor allele frequency.

<sup>†</sup> Mean bias of beta coefficient for main effect and gene-by-smoking interaction term from 1,000 regressions on 1,000 simulated datasets (i.e. mean of the observed beta coefficients – the simulated value of the pertinent effect).

Model is FEV1 (% predicted) = Age + Sex + Pack-Years + Gene + G\*S, where the pack-years variable and the gene-by-smoking interaction term are calculated using either the total number of pack-years smoked (linear pack-years adjustment), or a piecewise-linear representation of pack-years.

<sup>††</sup> Minor Allele Frequency of the simulated biallelic genetic variant.

<sup>‡</sup> p-value for test of null hypothesis that bias = 0.

**Table 3**  
 Characteristics of PI MZ and PI MM Subjects in the Norway Case-Control Study

	PI MZ	PI MM	p-value
N	78	1591	--
Age (SD)	62 (10)	60 (11)	0.33
Female, n (%)	43 (55)	875 (55)	0.97
Pack-Years, median (IQR)	25 (16–31)	22 (13–34)	0.24
FEV <sub>1</sub> , % Predicted, median (IQR)	72 (46–95)	81 (54–94)	0.38

Results from an Analysis of Main Effect and Gene-by-Smoking Interaction Effects of the Z-Allele in PI MZ and PI MM Subjects from the Bergen Norway Case-Control Study Using Two Different Methods of Adjustment for Smoking.

**Table 4**

	Z Allele	p-value	Z* S Interaction *	p-value	Pack- Years **	p-value
Model for FEV <sub>1</sub> , % predicted	Linear	-1.87	0.62	0.16	-0.33	<0.001
	Piecewise Linear	-5.59	0.14	0.35	-0.38	<0.001
Model for COPD Status	Linear	2.78	0.07	0.97	1.05	<0.001
	Piecewise Linear	4.87	0.01	0.94	1.08	<0.001

Regression models adjusted for age, sex, and smoking (using either a linear or piecewise linear form of the pack-years variable). The FEV<sub>1</sub> model includes a sample weight adjustment to reflect oversampling of COPD cases.

Results reported as beta-coefficients for FEV<sub>1</sub> model, odds ratios for COPD model, and their respective p-values.

\* The Z\*S interaction term is of the form (# copies of Z-allele × pack-years) for the linear smoking adjustment and (# copies of Z-allele × first 40 pack-years) for the linear and piecewise linear adjustments, respectively.

\*\* Smoking adjustment for the linear pack-years model is done by including a numerical term for total pack-years of smoking. For the piecewise-linear model, the smoking main effect is represented by 2 variables representing the first 40 pack-years of smoking and subsequent exposure.