

Published in final edited form as:

*Cancer Prev Res (Phila)*. 2008 September ; 1(4): 255–265. doi:10.1158/1940-6207.CAPR-08-0082.

## Development and Validation of a Lung Cancer Risk Prediction Model for African-Americans

Carol J. Etzel<sup>1</sup>, Sumesh Kachroo<sup>1,3</sup>, Mei Liu<sup>1,3</sup>, Anthony D'Amelio<sup>1,4</sup>, Qiong Dong<sup>1</sup>, Michele L. Cote<sup>6</sup>, Angela S. Wenzlaff<sup>6</sup>, Waun Ki Hong<sup>2</sup>, Anthony J. Greisinger<sup>5</sup>, Ann G. Schwartz<sup>6</sup>, and Margaret R. Spitz<sup>1</sup>

<sup>1</sup>Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center

<sup>2</sup>Department of Thoracic/Head and Neck Oncology, The University of Texas M. D. Anderson Cancer Center

<sup>3</sup>The University of Texas School of Public Health, Houston Texas

<sup>4</sup>Biomath/Biostatistics Program, The University of Texas Graduate School for Biomedical Sciences, Houston, Texas

<sup>5</sup>Kelsey Research Foundation, Houston, Texas

<sup>6</sup>Karmanos Cancer Institute, Wayne State University, Detroit, Michigan

### Abstract

Because existing risk prediction models for lung cancer were developed in white populations, they may not be appropriate for predicting risk among African-Americans. Therefore, a need exists to construct and validate a risk prediction model for lung cancer that is specific to African-Americans. We analyzed data from 491 African-Americans with lung cancer and 497 matched African-American controls to identify specific risks and incorporate them into a multivariable risk model for lung cancer and estimate the 5-year absolute risk of lung cancer. We performed internal and external validations of the risk model using data on additional cases and controls from the same ongoing multiracial/ethnic lung cancer case-control study from which the model-building data were obtained as well as data from two different lung cancer studies in metropolitan Detroit, respectively. We also compared our African-American model with our previously developed risk prediction model for whites. The final risk model included smoking-related variables [smoking status, pack-years smoked, age at smoking cessation (former smokers), and number of years since smoking cessation (former smokers)], self-reported physician diagnoses of chronic obstructive pulmonary disease or hay fever, and exposures to asbestos or wood dusts. Our risk prediction model for African-Americans exhibited good discrimination [75% (95% confidence interval, 0.67–0.82)] for our internal data and moderate discrimination [63% (95% confidence interval, 0.57–0.69)] for the external data group, which is an improvement over the Spitz model for white subjects. Existing lung cancer prediction models may not be appropriate for predicting risk for African-Americans because (a) they were developed using white populations, (b) level of risk is different for risk factors that African-American share with whites, and (c) unique group-specific risk factors exist for African-Americans. This study developed and validated a risk prediction model for lung cancer that is specific to African-Americans and thus

©2008 American Association for Cancer Research.

**Requests for reprints:** Carol J. Etzel, Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030. Phone: 713-745-2468; Fax: 713-792-8261; E-mail: cetzel@mdanderson.org..

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

more precise in predicting their risks. These findings highlight the importance of conducting further ethnic-specific analyses of disease risk.

Lung cancer risk prediction models are helpful in identifying subgroups of smokers who may exhibit higher risks of lung cancer and may therefore benefit disproportionately from early detection programs and other interventions. Recently developed lung cancer risk models (1–4) have focused predominantly on white populations. Although minority groups such as African-Americans share similar lung cancer risk factors with whites, the level of risk may differ depending on the type and level of exposure. Moreover, African-Americans may exhibit unique exposures (such as occupational exposures) that need to be included in the risk prediction models to improve accuracy. Hence, a need exists to develop and validate group-specific risk prediction models.

The goal of this analysis was to develop and validate a lung cancer risk prediction model that is specific to African-Americans. We present a comprehensive epidemiologic analysis of risk factors for lung cancer in African-Americans that was derived from an ongoing case-control study. Using these risk factors, we further develop and validate (using an independent set of internal data as well as external data) a risk prediction model for lung cancer applicable to African-Americans.

## Materials and Methods

### Study population

As part of a multiracial/ethnic lung cancer case-control study, we recruited study participants from The University of Texas M. D. Anderson Cancer Center and the Michael E. DeBakey VA Medical Center, both in Houston, from 1995 to 2005. All cases with newly diagnosed, histopathologically confirmed, and untreated lung cancer were eligible for the study. Case exclusion criteria for the study included prior chemotherapy or radiotherapy or recent blood transfusion. We recruited our control population from Houston area community centers and the Kelsey-Seybold Clinic, Houston's largest multispecialty physicians group practice. Potential controls were first surveyed with a brief questionnaire about their willingness to participate in research studies and provide preliminary data to assist us in matching demographic characteristics with those of the cases (5). Controls were matched to the cases based on age ( $\pm 5$  y), sex, and ethnicity. To date, the response rate among both the cases and controls has been ~75%.

In this analysis, we focused on a subset of cases and controls who self-reported as being black (African-American). These individuals represented ~14% of the overall study population.

The institutional review boards of the M. D. Anderson Cancer Center, Michael E. DeBakey VA Medical Center, and Kelsey-Seybold Clinic approved this research. Each participant signed an informed consent and completed a personal interview.

Never smokers were defined as those who had smoked <100 cigarettes in their lifetimes, former smokers had quit smoking >1 y before diagnosis (cases) or interview (controls), and current smokers included recent quitters who had quit smoking within the past 12 mo. Pack-years for former and current smokers were calculated as the years smoked times the average number of cigarettes per day divided by 20. Smokers were also asked to report their use of mentholated cigarettes, and former smokers to report the age at which they stopped smoking and the years since smoking cessation.

We classified participants as positive for asbestos exposure if they self-reported such exposure for at least 8 h per week for a year or if they reported having held a job within a documented

asbestos-related industry [according to the Standard Industrial Classification Manual (1972) and the Dictionary of Occupational Titles 1991; ref. 6]. Other environmental exposures included exposures through work or hobbies for  $\geq 8$  h per week for a year to wood dusts (wood dust, sawdust, or sand), fibers (textile fibers or cotton), synthetic vitreous fibers (SVF), toluene and/or xylene (paint, solvents, paint thinners, and printer's inks), and pesticides. We also classified participants according to self-reported physician-diagnosed medical comorbidities, including asthma, chronic obstructive pulmonary disease (COPD), and hay fever.

Participants self-reported the cancer histories of their first-degree relatives (i.e., parents, siblings, and offspring). Specifically, for each relative, we asked participants to provide year of birth, age at the time of study, smoking status (never or ever), presence or absence of cancer (yes or no), type of cancer, age of diagnosis, and year of death. In this analysis, we focused on family history of lung or other smoking-related (lung, bladder, kidney, head and neck, or pancreas) cancers.

We obtained clinical information about the study cases from their medical records, including stage of disease at diagnosis (based on tumor-node-metastasis staging criteria; ref. 7) and histologic characteristics.

### Statistical analysis

We used descriptive statistical analyses to characterize the study population, including Pearson's  $\chi^2$  test to test for distribution differences between the cases and controls for categorical variables and the Student's  $t$  test to determine differences in continuous variables. We used logistic regression analysis to calculate odds ratios (OR) and 95% confidence intervals (95% CI) to identify potential risk factors for lung cancer adjusted by age, sex, and smoking history. All analyses, unless otherwise noted, were completed using Statistical Analysis System version 9.1 software (SAS Institute, Inc.). When necessary, LogXact, available in Cytel Studio 6.3 (Cytel Inc.), was used to do exact logistic regression analyses to estimate ORs and calculate 95% CIs.

### Risk model development

The risk factors (significant at the 5% level in univariate analyses) were incorporated into a multivariable risk model for lung cancer. We evaluated the interactions between smoking variables and environmental exposures (asbestos and wood dusts) and used the likelihood ratio test to test for significant interactions (5%  $\alpha$  level). To identify other possible interactions that we did not specify a priori, we built a classification tree (8) using the recursive partitioning technique available in *rpart* package developed by Therneau and Atkinson<sup>7</sup> for S-Plus (Insightful Corp.). We grew the decision tree such that each subsequent split yielded two daughter nodes with at least 10 participants per node. An unconditional logistic regression model was fit at each recursive split to estimate the risk of lung cancer adjusted by age and sex; any branch that was not deemed to be statistically significant at the 5% level was pruned off the tree.

In the process of developing a final model, we first included statistically significant (via stepwise regression) main effects of smoking status, pack-years of smoking, and age at smoking cessation in the multivariable model. However, we also considered a composite variable that combined these three variable with never smokers as the reference. We also considered a composite variable that was identified through our tree-building process. This particular composite variable included not only smoking status, pack-years of smoking, and age at

---

<sup>7</sup>Technical Report Series No. 61, An introduction to recursive partitioning using the RPART routines. Department of Health Science Research, Mayo Clinic, Rochester, Minnesota, 1997. <http://lib.stat.cmu.edu/general/rpart>.

smoking cessation but also years of cessation and physician diagnosis of emphysema with never smokers and those who had smoked for less than 13.2 pack-years and had no emphysema as the reference group. We compared the Akaike information content (AIC) among the three models and the model with the lowest AIC was chosen as our final model.

To estimate the 5-y absolute risk of lung cancer, we first converted the  $\beta$  coefficients from the final risk model into ORs (which are equivalent to relative risks when prevalence is low), denoted by  $r$ . We then estimated baseline hazards for African-American men and women separately as  $h_{1j} = v_j (1 - s_j)$ , where  $v_j$  is the age, sex, and incidence rate of lung cancer for African-American men ( $j = 1$ ) or women ( $j = 2$ ) from the Surveillance and End Results Program (9) for 2002 and  $s_j$  is the attributable risk derived from the relative risk model as described in Fears et al. (10). The absolute 5-y risk is given by the following equation:

$$P(a, r, j) = \left[ h_{1j} r / (h_{1j} r + h_{2j}) \right] \left[ 1 - \exp \left\{ -5 (h_{1j} r + h_{2j}) \right\} \right] \quad (A)$$

where  $a$  is age in years,  $h_{1j}$  as defined above, and  $h_{2j}$  is the age and sex mortality rate from other causes (such as heart disease, stroke, diabetes, and other cancers excluding lung cancer; ref. 11) for males ( $j = 1$ ) or females ( $j = 2$ ) for African-Americans derived from the National Center for Health Statistics 1999 to 2003 mortality rates (12).

### Model validation

We validated our final model using two independent, internal and external, data sets. The internal validation was based on additional cases and controls from the same ongoing multiracial/ethnic lung cancer case-control study at M. D. Anderson Cancer Center from which the model-building data were obtained. These African-American patients and controls were enrolled between May 2005 and December 2007. The total number of participants analyzed for internal validation was 156. The data for external validation were from two different studies evaluating risk of lung cancer among African-Americans in metropolitan Detroit under the direction of Dr. Ann Schwartz (13,14). The first study was a case-control study of the risk of non-small cell lung cancer among women (13). The second study was a case-control study of lung cancer risk among relatives of early-onset lung cancer cases and frequency-matched controls (14). Due to missing data on the final list of model variables, we only included 325 participants for the external validation analysis. In addition, these two studies did not query specifically for physician diagnosis of hay fever; hence, we used allergies as a proxy for hay fever.

For each validation set, we calculated specificity and sensitivity to construct receiver-operator characteristic curves and calculated the area under the curve (AUC) to estimate the ability of the models to discriminate between African-American patients with lung cancer and controls. Approximate 95% CIs for the AUCs were calculated by using STATA statistical software (Release 8; Stata Corp.), assuming a bivariate exponential distribution. We also compared the AUCs of the African-American model to the published and internally validated Spitz model (15).

### Results

The demographic and clinical characteristics of the case and control test set used to develop the risk model are presented in Table 1. Among the 491 African-American cases and 497 controls, there were fewer male controls (40%) than cases (60%;  $P < 0.001$ ) and the controls were significantly younger than cases (although this difference was still within the 5-year age-matching criterion of the study).

The proportion of ever smokers was predictably higher in the cases compared with the controls ( $P < 0.001$ ), and conversely, only 7% of cases versus 29% of controls were never smokers. Among the study cases, men exhibited the highest percentage (59.9%) of current smokers, whereas 15.2% of women were never smokers (data not shown). Among controls, men had the highest percentage of former smokers (40.2%), whereas 36.8% of women were never smokers (data not shown). Cases who were current smokers smoked an average of 23 cigarettes per day, whereas the controls smoked on average 15 cigarettes per day ( $P < 0.001$ ). Control smokers reported a higher preference (47% versus 41% for cases) for mentholated cigarettes, although this difference was statistically significant only among current smokers ( $P = 0.019$ ; data not shown). Physician-diagnosis of COPD was more frequently reported among cases (16.5%) compared with controls (2.0%;  $P < 0.0001$ ).

The most common histology was adenocarcinoma (38.7%), except in men where squamous cell carcinoma was most frequent (39.7%; data not shown). We also observed that for never and former smokers, adenocarcinoma was the most prevalent histologic type, whereas in current smokers adenocarcinoma and squamous cell carcinoma were equally prevalent (data not shown). The majority of cases presented with stage III (40.7%) or stage IV (40.7%) disease.

### Identification of lung cancer risk factors for African-Americans

The main effects of risk factors are summarized in Table 2. Both current (OR, 6.20) and former smokers (OR, 3.38) exhibited significantly increased risks. Among current smokers with longer smoking duration ( $\geq 30$  years), the risk estimate was 1.97 (95% CI, 1.16–3.33). Current smokers who had smoked  $>20$  cigarettes per day had an approximate 4-fold increased risk (OR, 3.94; 95% CI, 2.26–6.87) of lung cancer and those who had smoked  $>40$  pack-years had 3.44-fold risk (95% CI, 2.15–5.47) with the risk more pronounced among women (OR, 15.91; 95% CI, 4.64–54.59; data not shown). The use of mentholated cigarettes seemed to be protective in current smokers, although the OR did not reach statistical significance ( $P > 0.05$ ) even after stratification by pack-years ( $\leq 40$  versus  $>40$  pack-years; data not shown).

Former smokers who had quit smoking for  $>10$  years had a significantly reduced risk of lung cancer (OR, 0.42; 95% CI, 0.27–0.67) compared with those who quit for less than 10 years. In addition, those former smokers who had quit smoking after age 30 had a significantly higher risk of lung cancer (OR, 2.60; 95% CI, 1.03–6.58) compared with those who quit before the age of 30.

We observed significantly elevated risks among those African-Americans who self-reported being exposed to asbestos (OR, 1.58; 95% CI, 1.18–2.12), wood dusts (OR, 1.50; 95% CI, 1.09–2.05), SVFs (OR, 2.27; 95% CI, 1.44–3.60), and toluene and/or xylene (OR, 1.40; 95% CI, 1.02–1.92). A prior history of physician-diagnosed COPD was associated with a substantially elevated risk (OR, 6.38; 95% CI, 3.24–12.59), whereas those with a reported history of hay fever had protection from lung cancer (OR, 0.68; 95% CI, 0.47–0.99). Previous diagnoses of asthma did not account for elevated risk in either smokers or nonsmokers (data not shown).

Approximately 16% of the cases and 12.9% of controls (Table 1) reported a family history of lung cancer in a first-degree relative (OR, 1.11; 95% CI, 0.76–1.61). More than 21% of the cases reported a family history of any smoking-related cancer compared with 19.1% of controls ( $P = 0.375$ ), but no significant risks were observed among these cases.

We also evaluated interactions between smoking variables and asbestos and dusts, but none of these interactions reached statistical significance (data not shown).



Physician diagnosis of COPD (selected as the first split of the data), pack-years smoked, age at smoking cessation, and years since smoking cessation were all identified as important risk factors for lung cancer in the classification tree model (Fig. 1). Although no higher-order interactions were evident, we did observe that African-Americans with COPD who smoked the heaviest ( $\geq 26.4$  pack-years) were at the highest risk (crude OR, 31.6; 95% CI, 13.6–73.4; data not shown) for lung cancer compared with never smokers and those former or current smokers who smoked the lightest amount of cigarettes ( $< 13.2$  pack-years). Even those African-Americans who did not have COPD, who quit smoking before age 56 and had not smoked within 3 years, but still had more than 26.4 pack-years of smoking still exhibited elevated risk (crude OR, 2.2; 95% CI, 1.3–3.7; data not shown) compared with the reference group of never and light smokers without COPD.

### Lung cancer risk model and estimation of 5-year absolute risk of lung cancer

In our multivariable risk modeling, the main effects for exposure to toluene and/or xylene were nonsignificant (data not shown). Those variables retained in the multivariable risk model included smoking status, pack-years of smoking ( $\leq 40$  versus  $> 40$  pack-years), age at smoking cessation ( $\leq 30$  versus  $> 30$  years), exposure to asbestos or dusts, and history of COPD or hay fever, with an AIC of 1,075.7. For the second model that included the composite variable never smokers (reference), former smokers who quit by age  $\leq 30$  years, former smokers who quit by age  $> 30$  years, current smokers who smoked  $\leq 40$  pack-years, and current smokers who smoked  $> 40$  pack-years along with main effects for exposure to asbestos or dusts and history of COPD or hay fever, the AIC was 1,080.1. In the model including the composite variable of smoking and COPD identified from the tree modeling (Fig. 1) and the main effects for the exposure and other medical history variables, hay fever and dusts were borderline significant ( $P = 0.05$ ). The AIC for this model was 1,045.5. Therefore, we chose this model as our final model (Table 3) and calculated attributable risks for men and women.

Using the methods described above, we estimated 5-year absolute risks of lung cancer and illustrated our approach with two different risk profile scenarios. First, for an African-American man aged 71 years who smoked for 40 pack-years but quit smoking after age 30, had no previous diagnoses of COPD or hay fever, and reported exposure to both asbestos and wood dusts, his relative risk of lung cancer compared with an African-American man of the same age with no risk factors (Table 3) is  $r = 10.1 \times 1.46 \times 1.38 = 20.35$ . The baseline hazard (Table 4) for this African-American man was  $h_{1,1} = 572.0534/100,000 \times (1 - 0.79) = 0.001201$ . The mortality rate from the Centers for Disease Control and Prevention for African-American men between 70 and 74 years old from causes other than lung cancer is  $h_{2,1} = 4,818.6/100,000 = 0.048186$  (Table 4). By substituting these numbers into the absolute risk equation (Eq. A), the estimated 5-year absolute risk for this man is 10.25%. Next, the relative risk of lung cancer for an African-American woman aged 56 who never smoked and has hay fever is  $1.00 \times 0.66 = 0.66$ . The baseline hazard (Table 4) for this African-American women is  $h_{1,2} = 160.2701/100,000 \times (1 - 0.59) = 0.0006571$ , and the mortality rate from the Centers for Disease Control and Prevention from causes other than lung cancer is  $h_{2,2} = 971.3/100,000 = 0.00971$  (Table 4). Therefore, the estimated 5-year absolute risk for this woman is 0.21%.

### Internal and external validation

Summary statistics on the variables in the final model for the internal and external validation groups have been summarized in Table 5. The internal validation set included 89 lung cancer cases and 67 controls. Fifty-five percent of the cases were male, whereas only 39% of the controls were male. Controls were much younger than cases (47.6 versus 62.9 years for males and 50.9 versus 65.2 years for females). Forty percent of the controls were never smokers, whereas only 11.3% of the cases were never smokers. The discriminatory power (as measured by the AUC) for the internal validation group was 75% (95% CI, 0.67–0.82), indicating good

discrimination. When we stratified the internal group by gender, the resulting AUCs for males and females were 0.70 (95% CI, 0.58–0.82) and 0.75 (95% CI, 0.64–0.85), respectively. The external validation set included 172 lung cancer cases and 153 controls. The majority of the cases (87%) and the controls (78%) were female. Controls were younger than cases (44.1 versus 46.1 years for males and 51.5 versus 55.2 years for females). The majority (86.1%) of the cases were former smokers, whereas most (56.9%) of the controls were current smokers. The discriminatory power of the model for the external validation data was more moderate with an overall AUC of 63% (95% CI, 0.57–0.69). Among women and early-onset lung cancer cases (the focus of each of the studies in the external data group), the model had discriminatory power of 61% (95% CI, 0.54–0.67) and 64% (95% CI, 0.64–0.74), respectively. The Spitz model (3) was based on white cases and controls who were frequency matched to cases on age, sex, and smoking status (never, former, and current). Separate models were presented for never, former, and current smokers, respectively. We were able to validate the Spitz model only for ever smokers because we had limited data on environmental tobacco smoke exposure (a key risk factor in the Spitz model for never smokers) among the African-American never smokers. Using our validation data, the AUCs for the Spitz model were 67% (95% CI, 0.57–0.77) and 59% (95% CI, 0.52–0.65) for African-American smokers in the internal and external validation groups, respectively. In comparison, the discriminatory power using the risk model for African-Americans was 79% (95% CI, 0.70–0.88) and 66% (95% CI, 0.61–0.71), respectively, for the two groups of African-American smokers.

## Discussion

This study developed and validated a specific risk prediction model of lung cancer for African-Americans. Smoking cessation (former smokers), pack-years smoked, prior COPD, no history of hay fever, and exposure to asbestos and wood dust were significant risk factors in this minority population. Internal and external validation results showed that our model discriminates well between cases and controls, except for a subgroup of early-onset lung cancer cases. The model that we propose for African-Americans outperformed a similar lung cancer risk model developed primarily for white populations.

In our comparison of risk factors in the African-American model and those developed using white-only data, we observed similarities and differences in distribution of risk factors among the two racial groups. In our study, 41% of study cases presented with advanced (stage IV) lung cancer. In contrast, 24.3% of the white cases in the entire parent study population had stage IV lung cancer at their initial examination. These differences in stage at time of presentation may be influenced not only by disparities in socio-economic factors, health insurance or access to health care in African-American populations (16), but by biologic factors as well. Difference exist in pulmonary function among different racial groups (17): African-Americans have lower forced vital capacity, lower forced expiratory volume, and higher percentage of forced expiratory capacity expired in 1 second compared to whites. Adenocarcinoma was the most common histology diagnosed in African-American women, whereas African-American men presented with squamous cell carcinoma as frequently as adenocarcinoma. Previously (3), we showed that 51% of white cases presented with adenocarcinoma followed by squamous cell carcinoma (22%).

Most of our African-American current smokers smoked between a half pack and a full pack (10–27 cigarettes) per day, similar to levels reported by Haiman et al. (18) and less than levels reported for whites. In our parallel case-control study in whites, the average number of cigarettes per day was 28.1 in cases compared with 26.4 in controls ( $P \leq 0.001$ ; ref. 3). In this study, cases who smoked  $\geq 20$  cigarettes per day exhibited about a 4-fold increase in risk (OR, 3.94; 95% CI, 2.26–6.87). Stellman et al. (19) reported that the risk of lung cancer for African-

Americans who smoked >20 cigarettes per day was five times higher than that for African-American women and six times higher than men who smoked  $\leq 20$  cigarettes per day.

The use of mentholated cigarettes is believed to increase the risk of lung cancer because these cigarettes have higher tar content, generate higher levels of carcinogens, and facilitate the absorption of carbon monoxide (20). It has also been suggested that the anesthetic properties and cooling effects of menthol may lead to deeper inhalation and longer retention, which may result in increased exposure to carcinogenic substances present in tobacco smoke (21,22). But the association between mentholated cigarette consumption and lung cancer is controversial (19,23–25). In a recent review, Werley et al. (21) reported that 70% to 80% of African-American men and women chose mentholated cigarettes. We observed lower percentages of menthol use (41–60%) in our study among African-American men and women smokers. In our analysis, we observed no significant risks of lung cancer among former or current smokers who reported smoking mentholated cigarettes (OR range, 0.69–0.99) and our data suggested a possible protective effect of mentholated cigarettes for current smokers. Sidney et al. (26) reported an increased risk in men who smoke mentholated cigarettes, whereas others (19,24, 25) reported a nonsignificant protective effect in both whites and African-Americans. A recent study by Murray et al. (27) also found no evidence of a higher risk of death from lung cancer in smokers who smoked mentholated cigarettes.

African-American men and women who reported longer durations of smoking cessation had lower risks of lung cancer. Others have reported similar declines in risk in African-American populations (19,28). We further observed that men who quit smoking after age 30 remained at substantially higher risks of lung cancer. A cohort study of 34,439 male British physicians reported that the maximum years of life expectancy gained were when cases stopped smoking before the age of 30 (29). Between 10% and 15% of all cases with lung cancer are never smokers (30); however, we observed a much lower proportion (2.4%) in our male cases.

We identified exposures to asbestos and wood dusts as significant risk factors for lung cancer in African-Americans. Such an association in African-Americans has been well documented previously (31,32). Our observations were similar to those in previous reports of asbestos exposure in whites (33) and of exposure to wood dusts in a subset of this population previously studied (34,35).

We also noted that some specific occupational exposures (SVFs and toluene and/or xylene) were associated with lung cancer among African-Americans, although these risks were not significant in our final multivariable model. Pohlabeln et al. (36) reported an association between exposure to man-made vitreous fibers (i.e., SVFs) and lung cancer after adjusting for smoking and asbestos exposure. All SVFs have essentially the same oncogenic potential. Some studies, however, may not show a statistically significant association because they lack sufficient statistical power to detect the small level of incremental risk associated with a fiber that clears rapidly from the lungs (37–39). Cohort studies have shown an increase in mortality due to lung cancer in whites after exposure to toluene (40,41). However, in our parallel study in whites (3), we identified a nonsignificant association between exposure to toluene and/or xylene and risk of lung cancer, although the prevalence of such exposure in cases with lung cancer (30.1%) was similar to that in African-American (34.4%) cases.

The risks associated with COPD that we report for African-Americans in our current study are substantially higher than those we reported previously for whites (OR, 1.7–2.65; refs. 3,32). For white cases, the prevalence of COPD was 22.4% and 9.4% in controls (3). However, in this analysis, the prevalence of COPD was 16.5% among the African-American cases and only 2% among controls. These differences reflect national rates. The 2002 COPD data from the American Lung Association report that the prevalence rates for chronic bronchitis in the U.S.



population are 45.7/1,000 for whites and 46.1/1,000 for blacks (42). National incidence rates for emphysema are 17.1/1,000 for whites and 8.3/1,000 for blacks (42). Hnizdo et al. (43) reported that the percent prevalence of airflow obstruction for the total working population was 10.7% for Caucasians and 7.5% for African-Americans. Littman et al. (44) reported that independent of smoking history, smokers (mostly white) with a history of chronic bronchitis and/or COPD have a higher risk of lung cancer, a finding confirmed in other studies (45,46).

The protective effect of hay fever that we observed in the minority group is similar to our previous findings for whites (33) and also previously reported in a mixed ethnic population of never smokers (47). Borderline significant protective effects of hay fever were also noted by Osann (48) in women with lung cancer. However, a later study by Osann et al. (49) found no significant association between hay fever and the risk of small cell lung cancer among women, although the Osann et al. study focused only on women with small cell lung cancer (49).

In this analysis, we did observe an elevated but not statistically significant association between family history and lung cancer. In our previous analyses of white cases and controls, we observed increased risks of lung cancer for never smokers (OR, 2.00) and former smokers (OR, 1.59) with a family history of any cancer and among current smokers (OR, 1.51) with a family history of smoking-related cancers (3). Cote et al. (13) reported that first-degree relatives of African-American individuals with early-onset lung cancer have a greater risk of lung cancer than whites, and cigarette smoking further increased these risks. We similarly stratified African-American cases by age at onset of disease (data not shown) but observed no increased risk among cases with early-onset (age  $\leq 50$  years) disease.

### Risk model

Cassidy et al. (50) reviewed the history of risk prediction models and particularly emphasize lung cancer risk prediction models and their importance to lung cancer research. In our lung cancer risk model for white lung cancer patients and controls (3), the controls, unlike our minority controls, were matched to their respective patients on smoking status. Because of the large number of white participants, we were able to construct and internally validate separate models for never, former, and current smokers. Some of the risk factors included in the three models previously published (1–3) overlapped with those found to be statistically significant in our minority risk models (i.e., smoking-related measures, exposures to asbestos or wood dusts, emphysema, and no previous hay fever). However, the level of risk may differ depending on the type and level of exposure. Therefore, these three risk models may not be applicable to minority populations.

Indeed, our model validation results showed that the ability of the Spitz model to discriminate between minority patients and controls was moderate (67% for African-Americans smokers), whereas the group-specific models showed better discriminatory power (79%). The Spitz model included smoking-related variables (age at smoking cessation for former smokers and pack-years smoked for current smokers) but could not include a risk estimate for smoking status because it was a matching variable. The inclusion of smoking status in the minority risk models could have further increased the discriminatory power that we observed. In the model based on white patients, asbestos and wood dust occupational exposures were significant. In the African-American model, asbestos and wood dusts were also included. Moreover, family histories of any cancer and of smoking-related cancers were included in the Spitz model for whites (3), but family history of cancer was not a statistically significant factor in our minority models.

The results from the validation analyses indicate that our lung cancer prediction model for African-Americans can prove helpful by providing more precise estimates compared with the

existing risk prediction models and further highlight the need for ethnic-specific lung cancer risk prediction models. However, the sample size for validation was small.

The other limitations of our study include the fact that our study was hospital based and the controls were drawn only from the metropolitan area of Houston, Texas; therefore, the results may vary in other geographic locations. Because the sample size of the study was small, the precision of our OR estimates is not as small compared with those in our white cases of lung cancer. We may be able to see better results with larger data sets or data sets where participants are not selectively enrolled based on particular criteria (such as gender or early age at diagnosis). Further, we were unable to confirm self-reported comorbidities or exposures. However, as noted by Cassidy et al. (50), a good risk prediction tool for lung cancer should include other factors in addition to smoking and age and our minority risk models included occupational exposures, comorbidities, and smoking phenotypes. Therefore, our results provide a foundation for lung cancer risk models for minority populations.

This study shows that African-Americans have risk factors for lung cancer that also occur in whites. However, the level of risk may differ depending on the type and level of exposure. African-Americans have a higher mortality due to lung cancer compared with whites and will benefit from early detection programs, including those using risk prediction tools. These findings highlight the importance of conducting further ethnic-specific analyses of disease risk.

## Acknowledgments

We thank Tamara Locke for her assistance in editing this manuscript.

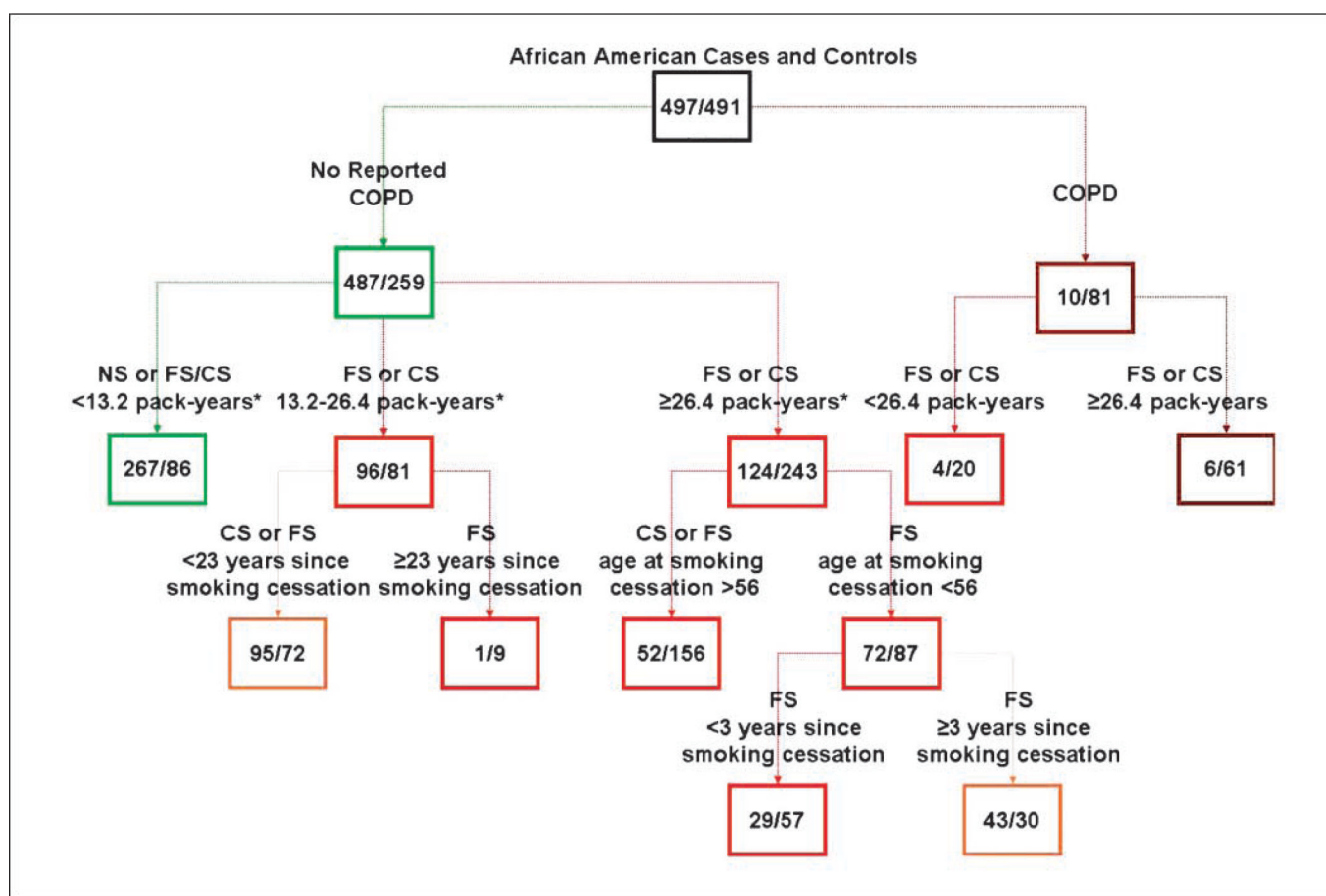
**Grant support:** Cancer prevention fellowship funded by the National Cancer Institute grant K07CA093592; National Cancer Institute grants CA55769, CA123208, CA60691, and CA87895; National Cancer Institute contract N01-PC35745; and Flight Attendant Medical Research Institute.

## References

1. Bach PB, Kattan MW, Thornquist MD, et al. Variations in lung cancer risk among smokers. *J Natl Cancer Inst* 2003;95:470–8. [PubMed: 12644540]
2. Colditz GA, Atwood KA, Emmons E, et al. Harvard report on cancer prevention volume 4: Harvard Cancer Risk Index. Risk Index Working Group, Harvard Center for Cancer Prevention. *Cancer Causes Control* 2000;11:477–88. [PubMed: 10880030]
3. Spitz MR, Hong WK, Amos CI, et al. A risk model for prediction of lung cancer. *J Natl Cancer Inst* 2007;99:715–26. [PubMed: 17470739]
4. Cassidy A, Myles JP, van Tongeren M, et al. The LLP risk model: an individual risk prediction model for lung cancer. *Br J Cancer* 2008;98:270–6. [PubMed: 18087271]
5. Hudmon KS, Honn SE, Jiang H, et al. Identifying and recruiting healthy control subjects from a managed care organization: a methodology for molecular epidemiological case-control studies of cancer. *Cancer Epidemiol Biomarkers Prev* 1997;6:565–71. [PubMed: 9264268]
6. Schabath MB, Spitz MR, Delclos GL, Gunn GB, Whitehead LW, Wu X. Association between asbestos exposure, cigarette smoking, myeloperoxidase (MPO) genotypes, and lung cancer risk. *Am J Ind Med* 2002;42:29–37. [PubMed: 12111688]
7. Fleming, ID.; Cooper, JS.; Henson, DE., et al. *AJCC cancer staging manual*. 5th ed. Lippincott-Raven; Philadelphia (PA): 1997.
8. Breiman, L.; Friedman, JN.; Olsen, RA.; Stone, CJ. *Classification and regression trees*. Wadsworth and Brooks/Cole; Monterey: 1984.
9. National Cancer Institute. SEER Scientific Systems, 2005. [2007 June 29]. Available from: <http://www-seer.ims.nci.nih.gov>
10. Fears TR, Guerry D, Pfeiffer RM, et al. Identifying individuals at high risk of melanoma: a practical predictor of absolute risk. *J Clin Oncol* 2006;24:3590–6. [PubMed: 16728488]

11. National Center for Health Statistics. Health, United States, 2007 with chartbook on trends in the health of Americans. National Center for Health Statistics; Hyattsville (MD): 2007.
12. National Center for Health Statistics. [2007 June 29]. Available from: [http://www.cdc.gov/nchs/datawh/statab/unpubd/mortabs/gmwk210\\_10.htm](http://www.cdc.gov/nchs/datawh/statab/unpubd/mortabs/gmwk210_10.htm)
13. Van Dyke AL, Cote ML, Prysak G, Claeys GB, Wenzlaff AS, Schwartz AG. Regular adult aspirin use decreases the risk of non-small cell lung cancer among women. *Cancer Epidemiol Biomarkers Prev* 2008;17:148–57. [PubMed: 18187393]
14. Coté ML, Kardia SL, Wenzlaff AS, Ruckdeschel JC, Schwartz AG. Risk of lung cancer among white and black relatives of individuals with early-onset lung cancer. *JAMA* 2005;293:3036–42. [PubMed: 15972566]
15. Hanley JA, McNeil BJ. A method of comparing the areas under the receiver operating characteristic curves derived from the same cases. *Radiology* 1983;148:839–43. [PubMed: 6878708]
16. Smedley, BD.; Stith, AY.; Nelson, AR., editors. Unequal treatment: confronting racial and ethnic disparities in health care. National Academy Press; Washington, DC: 2003.
17. Oscherwitz M, Edlavitch SA, Baker TR, Jarboe T. Differences in pulmonary functions in various racial groups. *Am J Epidemiol* 1972;95:319–27. [PubMed: 5015582]
18. Haiman CA, Stram DO, Wilkens LR, et al. Ethnic and racial differences in the smoking-related risk of lung cancer. *N Engl J Med* 2006;354:333–42. [PubMed: 16436765]
19. Stellman SD, Chen Y, Muscat JE, et al. Lung cancer risk in white and black Americans. *Ann Epidemiol* 2003;13:294–302. [PubMed: 12684197]
20. Gadgeel SM, Kalemkerian GP. Racial differences in lung cancer. *Cancer Metastasis Rev* 2003;22:39–46. [PubMed: 12716035]
21. Werley MS, Coggins CR, Lee PN. Possible effects on smokers of cigarette mentholation: a review of the evidence relating to key research questions. *Regul Toxicol Pharmacol* 2007;47:189–203. [PubMed: 17097785]
22. Abidoye O, Ferguson MK, Salgia R. Lung carcinoma in African Americans. *Nat Clin Pract Oncol* 2007;4:118–29. [PubMed: 17259932]
23. Carpenter CL, Jarvik ME, Morgenstern H, McCarthy WJ, London SJ. Mentholated cigarette smoking and lung-cancer risk. *Ann Epidemiol* 1999;9:114–20. [PubMed: 10037555]
24. Brooks DR, Palmer JR, Strom BL, Rosenberg L. Menthol cigarettes and risk of lung cancer. *Am J Epidemiol* 2003;158:609–16. [PubMed: 14507595]
25. Kabat GC, Hebert JR. Use of mentholated cigarettes and lung cancer risk. *Cancer Res* 1991;51:6510–3. [PubMed: 1742723]
26. Sidney S, Tekawa IS, Friedman GD, Sadler MC, Tashkin DP. Mentholated cigarette use and lung cancer. *Arch Intern Med* 1995;155:727–32. [PubMed: 7695461]
27. Murray RP, Connett JE, Skeans MA, Tashkin DP. Menthol cigarettes and health risks in Lung Health Study data. *Nicotine Tob Res* 2007;9:101–7. [PubMed: 17365741]
28. Pathak DR, Samet JM, Humble CG, Skipper BJ. Determinants of lung cancer risk in cigarette smokers in New Mexico. *J Natl Cancer Inst* 1986;76:597–604. [PubMed: 3457198]
29. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. *BMJ* 2004;328:1519. [PubMed: 15213107]
30. U.S. Department of Health and Human Services. The health consequences of smoking: a report of the Surgeon General. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; Rockville (MD): 2004.
31. Muscat JE, Stellman SD, Richie JP Jr, Wynder EL. Lung cancer risk and workplace exposures in black men and women. *Environ Res* 1998;76:78–84. [PubMed: 9515062]
32. Loomis D, Schulz M. Mortality from six work-related cancers among African Americans and Latinos. *Am J Ind Med* 2000;38:565–75. [PubMed: 11025498]
33. Schabath MB, Delclos GL, Martynowicz MM, et al. Opposing effects of emphysema, hay fever, and select genetic variants on lung cancer risk. *Am J Epidemiol* 2005;161:412–22. [PubMed: 15718477]

34. Wu X, Delclos GL, Annegers JF, et al. A case-control study of wood dust exposure, mutagen sensitivity, and lung cancer risk. *Cancer Epidemiol Biomarkers Prev* 1995;4:583–8. [PubMed: 8547823]
35. Barcenas CH, Delclos GL, El-Zein R, Tortolero-Luna G, Whitehead LW, Spitz MR. Wood dust exposure and the association with lung cancer risk. *Am J Ind Med* 2005;47:349–57. [PubMed: 15776474]
36. Pohlabeln H, Jockel KH, Bruske-Hohlfeld I, et al. Lung cancer and exposure to man-made vitreous fibers: results from a pooled case-control study in Germany. *Am J Ind Med* 2000;37:469–77. [PubMed: 10723041]
37. Brown RC, Bellmann B, Muhle H, Davis JM, Maxim LD. Survey of the biological effects of refractory ceramic fibres: overload and its possible consequences. *Ann Occup Hyg* 2005;49:295–307. [PubMed: 15668258]
38. Moolgavkar SH, Luebeck EG, Turim J, Brown RC. Lung cancer risk associated with exposure to man-made fibers. *Drug Chem Toxicol* 2000;23:223–42. [PubMed: 10711399]
39. Moolgavkar SH, Turim J, Brown RC, Luebeck EG. Long man-made fibers and lung cancer risk. *Regul Toxicol Pharmacol* 2001;33:138–46. [PubMed: 11350196]
40. Walker JT, Bloom TF, Stern FB, Okun AH, Fingerhut MA, Halperin WE. Mortality of workers employed in shoe manufacturing. *Scand J Work Environ Health* 1993;19:89–95. [PubMed: 8316784]
41. Svensson BG, Nise G, Englander V, Attewell R, Skerfving S, Moller T. Deaths and tumours among rotogravure printers exposed to toluene. *Br J Ind Med* 1990;47:372–9. [PubMed: 2378814]
42. American Lung Association. Lung disease data at a glance: COPD. [2007 June 15]. Available from: <http://www.lungusa.org/site/pp.asp?c=dvLUK9O0E&b=315487>
43. Hnizdo E, Sullivan PA, Bang KM, Wagner G. Airflow obstruction attributable to work in industry and occupation among U.S. race/ethnic groups: a study of NHANES III data. *Am J Ind Med* 2004;46:126–35. [PubMed: 15273964]
44. Littman AJ, Thornquist MD, White E, Jackson LA, Goodman GE, Vaughan TL. Prior lung disease and risk of lung cancer in a large prospective study. *Cancer Causes Control* 2004;15:819–27. [PubMed: 15456995]
45. Brenner AV, Wang Z, Kleinerman RA, et al. Previous pulmonary diseases and risk of lung cancer in Gansu Province, China. *Int J Epidemiol* 2001;30:118–24. [PubMed: 11171871]
46. Ramanakumar AV, Parent ME, Menzies D, Siemiatycki J. Risk of lung cancer following nonmalignant respiratory conditions: evidence from two case-control studies in Montreal, Canada. *Lung Cancer* 2006;53:5–12. [PubMed: 16733074]
47. Gorlova OY, Zhang Y, Schabath MB, et al. Never smokers and lung cancer risk: a case-control study of epidemiological factors. *Int J Cancer* 2006;118:1798–804. [PubMed: 16217766]
48. Osann KE. Lung cancer in women: the importance of smoking, family history of cancer, and medical history of respiratory disease. *Cancer Res* 1991;51:4893–7. [PubMed: 1654203]
49. Osann KE, Lowery JT, Schell MJ. Small cell lung cancer in women: risk associated with smoking, prior respiratory disease, and occupation. *Lung Cancer* 2000;28:1–10. [PubMed: 10704703]
50. Cassidy A, Duffy SW, Myles JP, Liloglou T, Field JK. Lung cancer risk prediction: a tool for early detection. *Int J Cancer* 2007;120:1–6. [PubMed: 17058200]



**Fig. 1.** Classification tree analysis of risk predictions in African-Americans. Nodes of the classification tree are formed by recursive splits of lung cancer case/control status by predictor variables. The numbers within each node indicate the number of control subjects/number of case patients. \*, trinary split resulting from two binary splits of pack-years. *NS*, never smokers; *FS*, former smokers; *CS*, current smokers.



**Table 1**

Distribution of demographic, smoking, and clinical variables

	Cases (n = 491)	Controls (n = 497)	P
Sex, n (%)			
Males	294 (60.0)	244 (49.1)	
Females	197 (40.0)	253 (50.9)	<0.001
Age (mean ± SD)			
Overall	61.3 ± 10.5	57.7 ± 11.0	<0.0001
Males	62.1 ± 9.9	58.1 ± 9.8	<0.001
Females	60.0 ± 11.2	57.2 ± 12.0	0.005
Smoking status, n (%)			
Never	37 (7.5)	144 (29.0)	
Former	176 (35.9)	174 (35.0)	
Current	278 (56.6)	179 (36.0)	<0.0001
Current smokers			
Smoking duration (mean ± SD)	40.2 ± 11.3	35.4 ± 11.5	<0.001
No. cigarettes/day (mean ± SD)	23.3 ± 12.5	14.6 ± 9.0	<0.001
Pack-years (mean ± SD)	49.8 ± 64.3	32.0 ± 75.6	0.009
Mentholated cigarette use, * n (%)	128 (48.5)	106 (59.9)	0.019
Former smokers			
Smoking duration (mean ± SD)	33.3 ± 14.0	25.1 ± 12.9	<0.0001
No. cigarettes/day (mean ± SD)	20.4 ± 12.8	17.4 ± 13.0	0.029
Pack-years (mean ± SD)	36.0 ± 27.8	23.4 ± 21.9	<0.0001
Years since smoking cessation (mean ± SE)	6.4 ± 10.0	10.5 ± 11.6	0.029
Age at smoking cessation (mean ± SE)	53.5 ± 12.2	45.1 ± 11.4	<0.001
Mentholated cigarette use, * n (%)	70 (40.7)	82 (47.4)	0.210
Environmental exposures (%+)			
Asbestos	210 (44.4)	151 (30.8)	<0.0001
Wood dusts	178 (39.7)	130 (27.1)	<0.0001
Fibers	51 (11.5)	37 (7.8)	0.058
SVF	77 (17.5)	35 (7.5)	<0.0001
Benzene	122 (27.3)	97 (20.6)	0.020
Toluene/xylene	155 (34.4)	118 (25.0)	0.002
Dry cleaning	22 (5.2)	13 (2.8)	0.069
Vehicle exhaust	122 (27.9)	113 (23.9)	0.178
Pesticides	44 (10.4)	32 (6.9)	0.061
Glues and plastics	55 (12.4)	41 (8.6)	0.117
Comorbidities (%+)			
Asthma	64 (13.0)	62 (12.5)	0.697
COPD	81 (16.5)	10 (2.0)	<0.0001
Hay fever	65 (13.2)	89 (17.9)	0.070
Family history among first degree			

	Cases ( <i>n</i> = 491)	Controls ( <i>n</i> = 497)	<i>P</i>
Relatives (%) lung cancer	78 (15.9)	64 (12.9)	0.178
Smoking related (%)	105 (21.4)	95 (19.1)	0.375
Histology, * <i>n</i> (%)			
Adenocarcinoma	168 (38.7)		
Squamous cell	144 (33.2)		
Other NSCLC	75 (17.3)		
Small cell	26 (6.0)		
Other lung cancer	21 (4.8)		
Stage, * <i>n</i> (%)			
I	32 (11.2)		
II	21 (7.4)		
III	116 (40.7)		
IV	116 (40.7)		

Abbreviations: NSCLC, non-small cell lung cancer; %+, percent positive.

\* Sum may not add up to total due to missing data.

**Table 2**

ORs and 95% CIs for smoking-related and other lung cancer risk factors

Risk factor	Reference group	OR * (95% CI)
Smoking status		
Former	Never smokers	3.38 (2.20–5.20)
Current		6.20 (4.06–9.47)
Current smokers		
Duration >30 y	≤30 y	1.97 (1.16–3.33)
No. cigarettes/day >20	≤20/d	3.94 (2.26–6.87)
Pack-years >40	<40 pack-years	3.44 (2.15–5.47)
Menthol cigarette use	None	0.69 (0.46–1.03)
Former smokers		
Duration >30 y	30 y	2.43 (1.58–3.75)
No. cigarettes/day >20	≤20/d	1.75 (1.04–2.93)
Pack-years >40	<40 pack-years	2.74 (1.67–4.52)
Years since cessation >10 y	≤10 y	0.42 (0.27–0.67)
Age at cessation >30 y	≤30 y	2.60 (1.03–6.58)
Menthol cigarette use	None	0.99 (0.62–1.56)
Environmental exposures		
Asbestos	No exposure	1.58 (1.18–2.12)
Wood dusts	No exposure	1.50 (1.09–2.05)
Fibers	No exposure	1.33 (0.83–2.13)
SVF	No exposure	2.27 (1.44–3.60)
Benzene	No exposure	1.17 (0.83–1.65)
Toluene/xylene	No exposure	1.40 (1.02–1.92)
Dry cleaning	No exposure	1.60 (0.76–3.37)
Vehicle exhaust	No exposure	1.01 (0.73–1.43)
Pesticides	No exposure	1.36 (0.82–2.25)
Glues and plastics	No exposure	1.22 (0.78–1.91)
Comorbidities		
Asthma	No self-reported diagnosis	1.11 (0.74–1.65)
COPD	No self-reported diagnosis	6.38 (3.24–12.59)
Hay fever	No self-reported diagnosis	0.68 (0.47–0.99)
Family history		
Lung cancer in ≥1 FDR	No family history	1.11 (0.76–1.61)
Smoking related cancer in ≥1 FDR	No family history	1.02 (0.73–1.42)

Abbreviation: FDR, first-degree relative.

\* Adjusted for age, sex, and smoking status when possible.

**Table 3**

Relative risks for each variable and attributable risk for the final multivariable epidemiologic risk model for African-Americans

<b>Risk factor</b>	<b>RR* (95% CI)</b>
Composite variable for smoking and COPD <sup>†</sup>	
NS or pack-years<13.3 and no COPD	1.00
13.3≤pack-years<26.4 and no COPD and years since cessation<23	2.56 (1.69–3.90)
Pack-years≥26.4 and no COPD and years since cessation≥3 and age at cessation<56	2.69 (1.47–4.92)
Pack-years≥26.4 and no COPD and years since cessation<3 and age at cessation<56	6.06 (3.66–10.06)
Pack-years≥26.4 and no COPD and age at cessation≥56	10.1 (6.24–16.22)
Pack-years<26.4 and COPD	20.00 (5.62–71.11)
13.3≤pack-years<26.4 and no COPD and years since cessation≥23	21.38 (2.52–181.24)
Pack-years≥26.4 and COPD	34.29 (12.94–90.91)
Asbestos	
Unexposed	1.00
Exposed	1.46 (1.04–2.04)
Dusts	
Unexposed	1.00
Exposed	1.38 (0.97–1.96)
Hay fever	
No diagnosis	1.00
Diagnosis	0.66 (0.43–1.01)
Attributable risk	
Males	0.79
Females	0.59

Abbreviation: RR, relative risk.

\* Age and sex adjusted.

<sup>†</sup> Identified from tree model in Fig. 1.

**Table 4**

Estimated rates per 100,000 for incidence of lung cancer and all-cause mortality without lung cancer by age and sex in African-Americans

Age group (y)	African-American			
	Men		Women	
	Incidence <sup>*</sup>	Mortality <sup>†</sup>	Incidence <sup>*</sup>	Mortality <sup>†</sup>
20–24	0.2015	224.5	0.2000	70.6
25–29	0.4540	249.6	0.4212	93.6
30–34	1.1557	262.5	0.6739	132.5
35–39	3.9880	334.8	4.9258	213.1
40–44	19.2688	507.1	15.5789	325.1
45–49	53.1998	798.1	33.6673	486.7
50–54	114.9708	1,220.7	64.0670	697.2
55–59	230.2280	1,678.0	108.8192	971.3
60–64	336.7714	2,437.0	160.2701	1,445.8
65–69	452.6758	3,312.0	236.0585	2,047.3
70–74	572.0534	4,818.6	272.7912	3,020.5
75–79	690.5254	6,952.6	302.7210	4,495.8
80–84	647.7919	9,796.1	273.0735	6,654.1
85+	650.5089	14,876.3	229.61176	1,3616.7

\* Lung cancer incidence for African-Americans from Surveillance and End Results (SEER; ref. 9).

† Mortality from all causes excluding lung cancer for African-Americans (11).



Table 5

Distribution of demographic, smoking, and clinical variables for internal and external validation data sets

Characteristic	Internal validation			External validation		
	Cases ( <i>n</i> = 89)	Controls ( <i>n</i> = 67)	<i>P</i> <sup>*</sup>	Cases ( <i>n</i> = 172)	Controls ( <i>n</i> = 153)	<i>P</i> <sup>*</sup>
% Male	55	39	0.004	13	22	0.04
Age (mean ± SD)						
Males	62.9 ± 8.9	47.6 ± 12.6	<0.0001	46.1 ± 3.0	44.1 ± 4.3	0.05
Females	65.2 ± 8.7	50.9 ± 15.1	<0.0001	55.2 ± 9.7	51.5 ± 10.7	0.003
Smoking status, <i>n</i> (%)						
Never	10 (11.3)	27 (40.3)	<0.0001	-	-	
Former	38 (42.7)	12 (17.9)		148 (86.1)	66 (43.1)	
Current	41 (46.0)	28 (41.8)		24 (13.9)	87 (56.9)	<0.0001
Asbestos, <i>n</i> (%)	35 (39.3)	15 (22.4)	0.025	37 (21.5)	20 (13.1)	0.05
Dusts, <i>n</i> (%)	20 (22.5)	16 (23.9)	0.84	11 (6.4)	11 (7.2)	0.78
COPD, <i>n</i> (%)	19 (21.3)	1 (1.5)	<0.001	14 (8.1)	4 (2.6)	0.03
Hay fever, <sup>‡</sup> <i>n</i> (%)	7 (7.9)	10 (14.9)	0.16	50 (29.1)	50 (32.7)	0.48
<b>Risk model validation</b>						
Overall		AUC (95% CI)	<i>P</i> <sup>‡</sup>		AUC (95% CI)	<i>P</i> <sup>‡</sup>
Gender		0.75 (0.67–0.82)			0.63 (0.57–0.69)	
Males		0.70 (0.58–0.82)				
Females		0.75 (0.64–0.85)			0.61 (0.54–0.67)	
Early onset					0.64 (0.64–0.74)	
Ever smokers						
African American model		0.79 (0.70–0.88)			0.66 (0.61–0.71)	
Spitz model		0.67 (0.57–0.77)	0.001		0.59 (0.52–0.65)	0.028

<sup>\*</sup> *P*-value from Pearson's Chi-square test (categorical variable) or two-sample *t*-test (continuous variable).

<sup>‡</sup> For the external validation, group allergies were used as a proxy for hay fever.

<sup>‡</sup> *P*-value for the comparison of two AUCs (16).