

## ORIGINAL ARTICLE

## Genetic Ancestry in Lung-Function Predictions

Rajesh Kumar, M.D., Max A. Seibold, Ph.D., Melinda C. Aldrich, Ph.D., M.P.H., L. Keoki Williams, M.D., M.P.H., Alex P. Reiner, M.D., Laura Colangelo, M.S., Joshua Galanter, M.D., Christopher Gignoux, M.S., Donglei Hu, Ph.D., Saunak Sen, Ph.D., Shweta Choudhry, Ph.D., Edward L. Peterson, Ph.D., Jose Rodriguez-Santana, M.D., William Rodriguez-Cintron, M.D., Michael A. Nalls, Ph.D., Tennille S. Leak, Ph.D., Ellen O'Meara, Ph.D., Bernd Meibohm, Ph.D., Stephen B. Kritchevsky, Ph.D., Rongling Li, M.D., Ph.D., M.P.H., Tamara B. Harris, M.D., Deborah A. Nickerson, Ph.D., Myriam Fornage, Ph.D., Paul Enright, M.D., Elad Ziv, M.D., Lewis J. Smith, M.D., Kiang Liu, Ph.D., and Esteban González Burchard, M.D., M.P.H.

## ABSTRACT

**BACKGROUND**

Self-identified race or ethnic group is used to determine normal reference standards in the prediction of pulmonary function. We conducted a study to determine whether the genetically determined percentage of African ancestry is associated with lung function and whether its use could improve predictions of lung function among persons who identified themselves as African American.

**METHODS**

We assessed the ancestry of 777 participants self-identified as African American in the Coronary Artery Risk Development in Young Adults (CARDIA) study and evaluated the relation between pulmonary function and ancestry by means of linear regression. We performed similar analyses of data for two independent cohorts of subjects identifying themselves as African American: 813 participants in the Health, Aging, and Body Composition (HABC) study and 579 participants in the Cardiovascular Health Study (CHS). We compared the fit of two types of models to lung-function measurements: models based on the covariates used in standard prediction equations and models incorporating ancestry. We also evaluated the effect of the ancestry-based models on the classification of disease severity in two asthma-study populations.

**RESULTS**

African ancestry was inversely related to forced expiratory volume in 1 second ( $FEV_1$ ) and forced vital capacity in the CARDIA cohort. These relations were also seen in the HABC and CHS cohorts. In predicting lung function, the ancestry-based model fit the data better than standard models. Ancestry-based models resulted in the reclassification of asthma severity (based on the percentage of the predicted  $FEV_1$ ) in 4 to 5% of participants.

**CONCLUSIONS**

Current predictive equations, which rely on self-identified race alone, may misestimate lung function among subjects who identify themselves as African American. Incorporating ancestry into normative equations may improve lung-function estimates and more accurately categorize disease severity. (Funded by the National Institutes of Health and others.)

From Children's Memorial Hospital (R.K.) and Northwestern University Feinberg School of Medicine (L.C., L.J.S., K.L.) — both in Chicago; National Jewish Health, Denver (M.A.S.); University of California at San Francisco, San Francisco (M.C.A., J.G., C.G., D.H., S.S., S.C., E.Z., E.G.B.); Henry Ford Health System, Detroit (L.K.W., E.L.P.); University of Washington, Seattle (A.P.R., E.O., D.A.N.); Pediatric Pulmonary Program of San Juan (J.R.-S.) and San Juan Veterans Affairs Medical Center, University of Puerto Rico School of Medicine (W.R.-C.) — both in San Juan; Laboratory of Neurogenetics (M.A.N.) and Laboratory of Epidemiology, Demography, and Biometry (T.B.H.), Intramural Research Program, National Institute on Aging, Bethesda, MD; Graduate School of Public Health, University of Pittsburgh, Pittsburgh (T.S.L.); College of Pharmacy (B.M.) and College of Medicine (R.L.), University of Tennessee Health Science Center, Memphis; Sticht Center on Aging, Wake Forest School of Medicine, Winston-Salem, NC (S.B.K.); School of Public Health, University of Texas Health Sciences Center at Houston, Houston (M.F.); and College of Public Health, University of Arizona, Tucson (P.E.). Address reprint requests to Dr. Kumar at the Division of Allergy and Immunology, Children's Memorial Hospital, 2300 Children's Plaza, Box 60, Chicago IL 60614, or at rkumar@childrensmemorial.org.

Drs. Kumar, Seibold, Aldrich, and Williams contributed equally to this article.

This article (10.1056/NEJMoa0907897) was published on July 7, 2010, at NEJM.org.

N Engl J Med 2010.

Copyright © 2010 Massachusetts Medical Society.

THE USE OF RACIAL OR ETHNIC CLASSIFICATION in medical practice and research has been the subject of debate.<sup>1-3</sup> Race and ethnicity are complex constructs incorporating social, cultural, and genetic factors. Currently, pulmonary-function testing is one of the few clinical applications in which self-reported race or ethnic group is used to define a normal range for a test outcome. Normative equations of lung function have been developed by testing large populations categorized on the basis of self-reported race or ethnic group.<sup>4</sup> However, many populations are racially admixed, and self-identified racial and ethnic categories are crude descriptors of individual genetic ancestry.<sup>5-8</sup> Use of self-reported race or ethnic group may misclassify persons with respect to the normal range for physiological measures, if the measures are dependent on ancestry. Such errors could lead to inaccuracies in evaluating individual pulmonary function and in determining population-specific disease prevalence and severity.

Advances in genetics have led to the development of ancestry informative markers, which allow genetic ancestry to be easily and inexpensively estimated in admixed populations such as self-identified African Americans and Latinos.<sup>9-11</sup> Ancestry may serve as a proxy for differentially distributed genetic factors that vary according to historical geographic separations.<sup>12</sup> Quantitative traits may vary with ancestry<sup>13-15</sup>; thus, ancestry may have relevance to lung function, which has a genetic component.<sup>16-19</sup> We hypothesized that incorporating measures of individual genetic ancestry into models would improve predictions of pulmonary function in self-identified African Americans.

## METHODS

### STUDY POPULATIONS

Self-identified African-American participants from five independent study populations were included in our analysis. We first analyzed data for participants in the Coronary Artery Risk Development in Young Adults (CARDIA) study (ClinicalTrials.gov number, NCT00005130).<sup>20,21</sup> We subsequently studied the cohorts from the Health, Aging, and Body Composition (HABC) study<sup>22</sup> and the Cardiovascular Health Study (CHS) (NCT00005133).<sup>23</sup> Finally, we evaluated the effect of ancestry-based models on the classification of disease severity in

two study populations of participants with asthma, from the Study of Asthma Phenotypes and Pharmacogenomic Interactions by Race–Ethnicity (SAPPHIRE) (NCT01142947) and the Study of African Americans, Asthma, Genes, and Environments (SAGE).<sup>24</sup> For the CARDIA, HABC, and CHS cohorts, participants who had asthma or pulmonary disease at the time of the baseline examination were excluded. See the Supplementary Appendix (available with the full text of this article at NEJM.org) for details of these five study populations.

The CARDIA, HABC, CHS, SAPPHIRE, and SAGE cohorts consisted of 777, 813, 579, 698, and 95 study participants, respectively; the age ranges at baseline were 18 to 30 years, 68 to 81 years, 65 to 93 years, 18 to 56 years, and 18 to 40 years, respectively. In all studies, spirometry was performed in accordance with American Thoracic Society recommendations.<sup>25</sup> All subjects provided written informed consent, and each study was approved by the institutional review boards at all participating institutions.

### ESTIMATION OF ANCESTRAL ADMIXTURE AND GENOTYPING

To estimate the ancestral admixture in CARDIA participants, we used genomewide genotyping data from the microarray platform (Genome-Wide Human SNP Array 6.0, Affymetrix) used in the Candidate Gene Association Resource (CARE) study. We removed single-nucleotide polymorphisms (SNPs) with more than 5% missing values and those that were not in Hardy–Weinberg equilibrium ( $P < 10^{-5}$ ) as well as SNP pairs in high linkage disequilibrium ( $r^2 \geq 0.8$ ). This left a final sample of 631,243 autosomal SNPs from which to estimate ancestry with the use of the Admixture software program<sup>26</sup> (see the Supplementary Appendix for further details). We genotyped 1332 ancestry informative markers in samples obtained from participants in the HABC study and used the software program Structure to estimate the percent African and European ancestry,<sup>27</sup> assuming two ancestral populations.<sup>28,29</sup> We estimated the percentage of African ancestry in CHS participants by using a maximum-likelihood method incorporating 24 ancestry informative markers,<sup>30</sup> and in SAPPHIRE participants by using a maximum-likelihood estimation<sup>31</sup> and 107 previously described ancestry informative markers.<sup>7</sup> We esti-

ated the percent African ancestry in SAGE participants by using the Structure program and 104 ancestry informative markers.<sup>32</sup>

#### STATISTICAL ANALYSIS

In the CARDIA study, we examined the effect of genetic ancestry on measures of lung function (i.e., the forced expiratory volume in 1 second [FEV<sub>1</sub>], the forced vital capacity [FVC], and the FEV<sub>1</sub>:FVC ratio) ascertained at the baseline examination. Linear regression models, stratified on the basis of sex, included several covariates: percent African ancestry, age, lifetime pack-years of smoking, body-mass index (BMI), height, square of the height, and study site. Age was centered at 25 years, by which time the majority of lung growth has occurred but the onset of lung-function decline in association with age has not.<sup>33,34</sup> Other continuous variables were centered around their study-population means. We also examined differences in the effect of ancestry between the sexes, by adding both sex and interactions between sex and ancestry as covariates to a model analyzing pooled data for men and women.

We evaluated the association between baseline lung function and African ancestry in two additional cohorts, from the HABC study and the CHS. We used linear regression models to assess the associations of genetic ancestry with measures of lung function, after controlling for age, sex, lifetime pack-years of smoking, BMI, height, square of the height, and study site.

To ascertain whether the use of data on genetic ancestry could improve the accuracy of pulmonary-function predictions, we compared models based on self-identified race or ethnic group with models incorporating ancestry as determined by genetic means. To do so, we made use of the age- and sex-specific covariates used by Hankinson and colleagues<sup>4</sup> to estimate FEV<sub>1</sub> in persons identifying themselves as African Americans. These covariates were selected according to their fit in the third National Health and Nutrition Examination Survey (NHANES III) study population, and the models that included them are referred to herein as standard race-based models. In other models, here called ancestry-based models, we used the age- and sex-specific covariates but also added a term for individual ancestry to assess whether ancestry remained independently associated with FEV<sub>1</sub> and FVC and whether the additional covari-

**Table 1. Baseline Characteristics of CARDIA Study Participants, According to Sex.\***

Characteristic	Men (N=309)	Women (N=468)
Age (yr)	24.5±3.7	24.4±3.9
BMI†	24.5±4.1	26.4±6.3
Education (yr)	13.1±2.0	13.2±1.8
Median income category (\$)	25,000–34,999	16,000–24,999
No. of lifetime pack-years of smoking	1.8±3.4	1.7±3.5
Pulmonary function (liters)		
FEV <sub>1</sub>	3.83±0.56	2.85±0.42
FVC	4.63±0.70	3.34±0.50
African ancestry (%)	85.0±10.1	85.1±9.6

\* Plus-minus values are means ±SD. CARDIA denotes the Coronary Artery Risk Development in Young Adults study, FEV<sub>1</sub> forced expiratory volume in 1 second, and FVC forced vital capacity.

† The body-mass index (BMI) is the weight in kilograms divided by the square of the height in meters.

ate improved the overall fit of the model. Model fit was assessed with the use of the mean squared error, the coefficient of determination (R<sup>2</sup>), and an adjusted coefficient of determination (adjusted R<sup>2</sup>). The adjusted coefficient is a measure of the proportion of variance in FEV<sub>1</sub> explained by the model, accounting for the number of covariates used. We tested these models in the three population-based cohorts: the CARDIA, HABC, and CHS cohorts.

We also categorized asthma severity on the basis of the percentage of the predicted FEV<sub>1</sub> cutoff values specified in current U.S. guidelines.<sup>35</sup> We assessed the number of participants whose asthma-severity categorization differed when the predicted FEV<sub>1</sub> value was obtained from models that incorporated genetic ancestry rather than from models that did not. We derived parameter estimates for these models from data for the CARDIA cohort, using the covariates from Hankinson et al.<sup>4</sup> for self-identified African-American men 20 years of age or older and women 18 years of age or older. We then applied these models to the same age groups in the SAPPHIRE and SAGE cohorts.

## RESULTS

#### STUDY POPULATIONS

The CARDIA cohort consisted of 777 participants who identified themselves as African American,

**Table 2. Characteristics of the CHS and HABC Study Participants.\***

Characteristic	CHS (N=579)	HABC (N=813)
Age — yr		
Mean	72.5	73.8
Range	65–93	68–81
Female sex — no. (%)	367 (63.4)	435 (53.5)
BMI†	28.4±5.0	28.5±5.2
Education — yr	13.0±5.2	11.6±3.5
Median income category — \$	12,000–15,999	10,000–25,000
Smoking status — no./total no. (%)		
Current smoker	81/576 (14.1)	126/811 (15.5)
Former smoker	209 (36.3)	336 (41.4)
Never smoked	286 (49.7)	349 (43.0)
Pulmonary function — liters		
FEV <sub>1</sub>	1.79±0.6	2.0±0.6
FVC	2.52±0.8	2.7±0.7
African ancestry — %	72.8±20.7	78.0±12.8

\* Plus-minus values are means ±SD. CHS denotes the Cardiovascular Heart Study, FEV<sub>1</sub> forced expiratory volume in 1 second, FVC forced vital capacity, and HABC the Health, Aging, and Body Composition study.

† The body-mass index (BMI) is the weight in kilograms divided by the square of the height in meters.

with a mean age of 24.5 years at the time of enrollment (Table 1). The mean percentage of African ancestry did not differ significantly between men and women (85.0% and 85.1%, respectively). Participants in the CHS and HABC cohorts were older (Table 2).

#### RELATION BETWEEN ANCESTRY AND LUNG FUNCTION

Figure 1 shows the distribution of, and considerable variation in, the percentage of genetically determined African ancestry among male and female CARDIA participants (Panels A and B, respectively). Table 3 presents the results of the test for association between pulmonary-function variables (FEV<sub>1</sub>, FVC, and the FEV<sub>1</sub>:FVC ratio) and African ancestry among CARDIA subjects, with adjustment for the following potential confounders: age, lifetime pack-years of smoking, BMI, height, square of the height, and study site. The coefficient for African ancestry represents the change in each lung-function variable for each percentage-point increase in African ancestry, with adjustment for all other variables shown. In both men and women, there were significant inverse

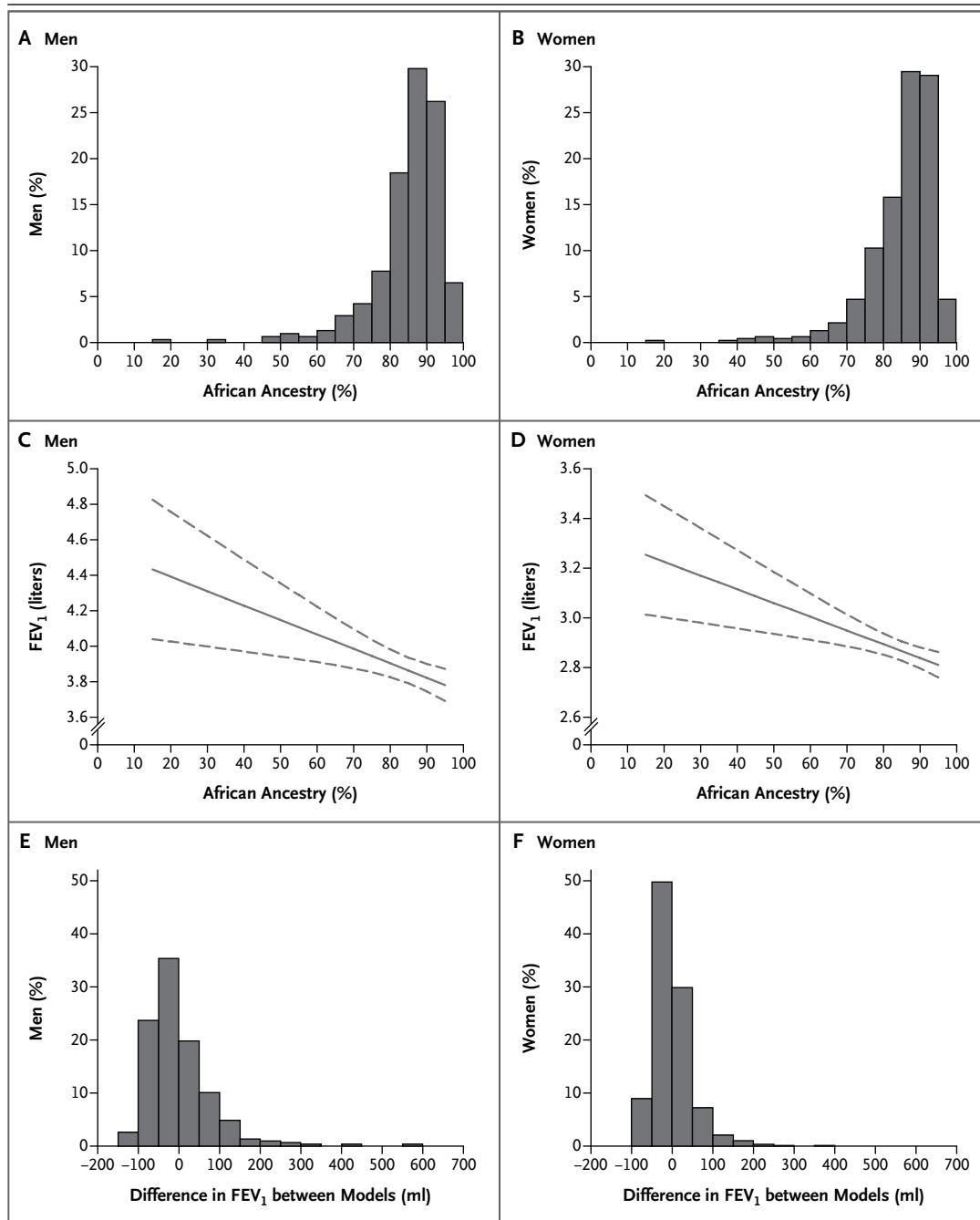
relations between African ancestry and FEV<sub>1</sub> and between African ancestry and FVC. The association of ancestry with FEV<sub>1</sub> among men and among women in the CARDIA cohort is shown in Figures 1C and 1D, respectively. In an analysis of the pooled data for men and women, interaction terms for sex and ancestry neared significance (P=0.09 for FEV<sub>1</sub>:FVC and P=0.07 for FVC), suggesting that the negative relation between ancestry and FVC was greater for men than for women (data not shown).

We also performed a post hoc analysis in the CARDIA cohort, incorporating additional measures of smoking exposure, including smoking status (current smoker, former smoker, or never smoked), the number of cigarettes smoked per day, and the number of years of smoking. The effect size and the significance of the association between ancestry and lung function were not substantially changed in this analysis (data not shown).

We evaluated the difference in the CARDIA cohort between the predicted FEV<sub>1</sub> estimated from the ancestry-based models (Table 3) and the predicted FEV<sub>1</sub> estimated from the standard race-based models, which do not include genetically determined ancestry (Table 1 in the Supplementary Appendix). Figure 1 shows the differences in predicted FEV<sub>1</sub> between the standard race-based and ancestry-based models for men and women (Panels E and F, respectively). The mean (±SD) absolute difference in predicted FEV<sub>1</sub> between the two models for each participant was 55.5±58.3 ml for men and 34.1±36.9 ml for women.

#### ANALYSES OF ADDITIONAL COHORTS

We did not observe a significant interaction between sex and ancestry with respect to the effect on lung-function outcomes in either the HABC cohort or the CHS cohort (data not shown). Therefore, in each cohort, we analyzed the data for men and women together. In both cohorts, we found that African ancestry was significantly associated with both FEV<sub>1</sub> and FVC (Table 4). For each percentage-point increase in African ancestry, the associated changes in FEV<sub>1</sub> and FVC were –3.99 ml and –5.50 ml, respectively, among HABC participants, and –2.39 ml and –3.46 ml, respectively, among CHS participants. We performed an additional analysis of the HABC data wherein the height while seated was used in lieu of the height while standing. There was minimal attenuation



**Figure 1. Ancestry, Forced Expiratory Volume in 1 Second (FEV<sub>1</sub>), and Differences in Predicted FEV<sub>1</sub> among CARDIA Study Participants.**

The percentage of African ancestry, as estimated by means of genetic markers, in men and women who identified themselves as African American is shown in Panels A and B, respectively. The relation between African ancestry and FEV<sub>1</sub> (indicated as a solid curve, with 95% confidence intervals as dashed curves) is also shown for men and women, in Panels C and D, respectively. The extent and distribution of the differences in predicted FEV<sub>1</sub> between the ancestry-based models and the standard race-based models (not including ancestry) in men and women are shown in Panels E and F, respectively.

**Table 3. Associations between Lung Function and Covariates of Ancestry-Based Models among Participants in the CARDIA Study, According to Sex.\***

Covariate	Men (N=309)			Women (N=468)		
	FEV <sub>1</sub>	FVC	FEV <sub>1</sub> :FVC	FEV <sub>1</sub>	FVC	FEV <sub>1</sub> :FVC
	milliliters			milliliters		
Intercept†	<b>3885.6</b>	<b>4763.9</b>	<b>81.8</b>	<b>2892.0</b>	<b>3368.4</b>	<b>86.1</b>
Age (per yr)‡	3.75	<b>20.57</b>	<b>-0.29</b>	<b>-9.55</b>	0.23	<b>-0.31</b>
African ancestry (per percentage point)§	<b>-8.14</b>	<b>-12.12</b>	0.04	<b>-5.54</b>	<b>-5.41</b>	-0.03
Lifetime smoking (per pack-yr)¶	-9.13	-7.82	0.00	<b>-14.95</b>	-8.72	<b>-0.22</b>
BMI (per unit)‖	9.11	<b>20.17</b>	-0.15	-0.69	-0.12	-0.03
Height (per cm)**	<b>38.53</b>	<b>53.70</b>	<b>-0.12</b>	<b>33.62</b>	<b>42.58</b>	<b>-0.07</b>
Square of height (per cm <sup>2</sup> )††	-0.37	0.06	-0.01	0.004	0.16	<b>-0.01</b>
Study site						
Chicago‡‡	-91.64	<b>-206.26</b>	1.61	4.12	60.64	<b>-1.67</b>
Minnesota	56.15	-22.97	1.46	-56.82	-57.75	-0.18
Oakland§§	-44.11	<b>-193.85</b>	<b>2.29</b>	-58.77	-91.08	0.57

\* The data are the milliliters (for forced expiratory volume in 1 second [FEV<sub>1</sub>] and forced vital capacity [FVC]) or ratios (for FEV<sub>1</sub>:FVC) associated with each listed covariate, after adjustment for all other covariates listed. The intercept value is added to the covariate value (after appropriate weighting) to obtain the lung-function estimate. Bold type indicates a significant association (P<0.05). CARDIA denotes the Coronary Artery Risk Development in Young Adults study.

† P<0.001 for all coefficients for intercept.

‡ For coefficients for age, P=0.03 and P=0.007 for FVC and FEV<sub>1</sub>:FVC among men, respectively, and P=0.03 and P<0.001 for FEV<sub>1</sub> and FEV<sub>1</sub>:FVC among women, respectively.

§ For coefficients for African ancestry, P=0.004 and P<0.001 for FEV<sub>1</sub> and FVC among men, respectively, and P=0.001 and P=0.007 for FEV<sub>1</sub> and FVC among women, respectively.

¶ For coefficients for pack-years of smoking, P=0.003 and P=0.004 for FEV<sub>1</sub> and FEV<sub>1</sub>:FVC among women, respectively.

‖ The body-mass index (BMI) is the weight in kilograms divided by the square of the height in meters. For the coefficient for BMI, P=0.01 for FVC among men.

\*\* For coefficients for height, P<0.001 for FEV<sub>1</sub> and FVC among men and women, and P=0.03 and P=0.04 for FEV<sub>1</sub>:FVC among men and among women, respectively.

†† For the coefficient for the square of height, P=0.051 for FEV<sub>1</sub>:FVC among women.

‡‡ For the coefficients for Chicago, P=0.03 for FVC among men and P=0.02 for FEV<sub>1</sub>:FVC among women.

§§ For the coefficients for Oakland, P=0.02 for FVC and for FEV<sub>1</sub>:FVC among men.

in the relation between African ancestry and the lung-function variables FEV<sub>1</sub> and FVC, and these associations remained significant (Table 2 in the Supplementary Appendix).

#### PREDICTING FEV<sub>1</sub> WITH THE USE OF STANDARD RACE-BASED MODELS AND ANCESTRY-BASED MODELS

To determine whether the inclusion of ancestry improves the fit of predictive models of pulmonary function, we regressed data for FEV<sub>1</sub> on the covariates used in the standard race-based models (i.e., those used elsewhere<sup>4</sup> for self-identified African-American women ≥18 years of age and men ≥20 years of age) in the three population-based cohorts (CARDIA, HABC, and CHS). We also used ancestry-based models containing the age-, sex-, and race-specific covariates as well as an

additional term for individual genetic African ancestry. The term for ancestry was significantly associated with both FVC and FEV<sub>1</sub> in all groups, with the exception of men in the CHS cohort (Tables 3 and 4 in the Supplementary Appendix). The ancestry-based models explained more of the variance than the standard race-based models (except for FEV<sub>1</sub> among women in the CHS cohort), even after we had accounted for the number of variables included in the model (reflected in the adjusted R<sup>2</sup>).

#### CLASSIFYING ASTHMA SEVERITY

We assessed the severity of asthma in 698 SAPPHERE participants and 95 SAGE participants. Severity was categorized according to the percent of the predicted FEV<sub>1</sub> cutoff values specified in current U.S. guidelines,<sup>35</sup> and these per-

**Table 4. Associations between Lung Function and Covariates of Ancestry-Based Models among Participants in the CHS and HABC Studies.\***

Covariate	CHS (N=579)			HABC (N=813)		
	FEV <sub>1</sub> milliliters	FVC milliliters	FEV <sub>1</sub> :FVC	FEV <sub>1</sub> milliliters	FVC milliliters	FEV <sub>1</sub> :FVC
Intercept†	<b>2050.7</b>	<b>2902.2</b>	<b>70.3</b>	<b>2165.8</b>	<b>2755.3</b>	<b>77.8</b>
Age (per yr)‡	<b>-19.78</b>	<b>-22.71</b>	-0.09	<b>-17.27</b>	<b>-19.86</b>	-0.13
Sex§	<b>-398.04</b>	<b>-640.25</b>	<b>3.24</b>	<b>-348.66</b>	<b>-442.68</b>	-0.29
African ancestry (per percentage point)¶	<b>-2.39</b>	<b>-3.46</b>	0.02	<b>-3.99</b>	<b>-5.50</b>	0.01
Lifetime smoking (per pack-yr)	<b>-3.66</b>	-1.75	<b>-0.09</b>	<b>-4.80</b>	<b>-3.45</b>	<b>-0.06</b>
BMI (per unit)**	-1.46	-4.56	0.06	-1.36	<b>-10.25</b>	<b>0.21</b>
Height (per cm) ††	<b>19.73</b>	<b>28.94</b>	-0.01	<b>21.12</b>	<b>33.96</b>	<b>-0.07</b>
Square of height (per cm <sup>2</sup> )	0.07	0.02	0.01	-0.15	0.09	-0.004
Study site						
Site 1 †††	26.74	54.55	-0.12	<b>65.86</b>	<b>121.31</b>	-0.28
Site 2	54.33	50.67	1.13			
Site 3	59.88	71.89	0.53			

\* The data are the milliliters (for forced expiratory volume in 1 second [FEV<sub>1</sub>] and forced vital capacity [FVC]) or ratios (for FEV<sub>1</sub>:FVC) associated with each listed covariate, after adjustment for all other covariates listed. The intercept value is added to the covariate value (after appropriate weighting) to obtain the lung-function estimate. Bold type indicates a significant association (P<0.05). CHS denotes the Cardiovascular Heart Study, and HABC the Health, Aging, and Body Composition study.

† P<0.001 for all coefficients for intercept.

‡ For coefficients for age, P<0.001 for FEV<sub>1</sub> and FVC in the CHS and HABC cohorts.

§ For coefficients for sex, P<0.001 for FEV<sub>1</sub> and FVC in the CHS and HABC cohorts, and P=0.01 for FEV<sub>1</sub>:FVC in the CHS cohort.

¶ For coefficients for African ancestry, P=0.009 and P=0.004 for FEV<sub>1</sub> and FVC in the CHS cohort, respectively, and P<0.001 for FEV<sub>1</sub> and FVC in the HABC cohort.

|| P<0.001 for coefficients for pack-years of smoking for FEV<sub>1</sub> and FEV<sub>1</sub>:FVC in the CHS cohort and for FEV<sub>1</sub>, FVC, and FEV<sub>1</sub>:FVC in the HABC cohort.

\*\* The body-mass index (BMI) is the weight in kilograms divided by the square of the height in meters. For coefficients for BMI, P=0.002 and P<0.001 for FVC and FEV<sub>1</sub>:FVC in the HABC cohort, respectively.

†† For coefficients for height, P<0.001 for FEV<sub>1</sub> and FVC in the CHS cohort and the HABC cohort, and P=0.045 for FEV<sub>1</sub>:FVC in the HABC cohort.

††† For coefficients for site 1, P=0.03 for FEV<sub>1</sub> and P<0.001 for FVC in the HABC cohort.

percentages were derived from predicted FEV<sub>1</sub> values from models with and those without a term for genetic African ancestry. A total of 28 participants (4.0%) in SAPPHIRE and 5 (5.3%) in SAGE would have had a change in their asthma-severity category had ancestry been included in the model used to predict the FEV<sub>1</sub>. Specifically, the asthma would have been classified as less severe in 14 (2.0%) and more severe in 14 (2.0%) of SAPPHIRE participants, with corresponding reclassifications in 1 (1.1%) and 4 (4.2%) of SAGE participants.

## DISCUSSION

We found an association between genetic ancestry and lung function among subjects who identified themselves as African American; specifi-

cally, the percentage of African ancestry was inversely associated with lung function in three independent cohorts across a wide range of ages. However, there were differences in the magnitude of the effect across the cohorts, possibly because of the older age of subjects in the HABC and CHS cohorts. The relative contribution of genetic ancestry is likely to be smaller among older persons who have had a greater cumulative exposure to environmental factors that adversely affect lung function.

Extrapolation of the distribution of percentage of African ancestry among the men and women in the CARDIA study to persons of the same sex and age group (18 to 34 years) suggests that for approximately 6.4% of persons in the United States who identify themselves as African Americans

(i.e., 0.65 million persons, according to 2008 U.S. Census estimates)<sup>36</sup> the percentage of African ancestry would be 15% higher or lower than the mean, suggesting they would be misrepresented by the use of standard race-based models. In these persons, the degree of misestimation of FEV<sub>1</sub> could be greater than 122.1 ml in men and 83.1 ml in women, the equivalent of an age-associated loss of lung function of approximately 15 ml per year over a 5-to-8-year period.<sup>4</sup>

The inverse association between lung function and African ancestry may be especially important when predicted lung-function values are used to assess the severity of lung disease.<sup>35,37,38</sup> Current clinical practice guidelines use the FEV<sub>1</sub>:FVC ratio or percentages of predicted lung function in the assessment of the severity of chronic obstructive pulmonary disease (COPD),<sup>39,40</sup> the severity of asthma,<sup>35</sup> and the degree of overall lung impairment.<sup>41</sup> Accordingly, inaccurate estimates of percentages of predicted lung function may result in the misclassification of disease severity and impairment.<sup>41</sup> For example, an estimated 2.1 million self-identified African Americans have asthma.<sup>42</sup> On the basis of the percentage of the predicted FEV<sub>1</sub> value, the severity of the asthma would be misclassified for approximately 4% of these patients (i.e., 84,000 patients), were ancestry not taken into account. Misclassification of severity might affect even more patients with COPD, a condition more common than asthma. An improvement in the accuracy of predicted lung function may lead to more appropriate treatment for the level of impairment, resulting in more effective and efficient care.

There are some important limitations of our study. First, our analysis does not address population groups other than self-identified African Americans, such as Latinos, who have more complex patterns of ancestral admixture. Second, the association between lung function and ancestry found in our study may be the result of factors other than genetic variation, such as premature birth, prenatal nutrition, socioeconomic status, and other environmental factors.<sup>43,44</sup> Third, we did not study a replication population with the same age range as that of the CARDIA cohort. Thus, we may have overestimated the association between ancestry and lung function in the CARDIA participants, who were young adults. Finally, some researcher groups used different statistical approaches to estimate ancestry in their respective

study populations. We have found previously, however, that different approaches (e.g., Markov models and maximum-likelihood estimation) produce highly correlated results from the same set of markers.<sup>8,45</sup> The consistency of our findings across three cohorts, despite the different methods for estimating ancestry, underscores the robustness of the association with ancestry.

In conclusion, our study shows that incorporating measures of individual genetic ancestry into normative equations of lung function in persons who identify themselves as African Americans may provide more accurate predictions than formulas based on self-reported ancestry alone. The same argument may apply to other ancestrally defined groups; further studies in this area are necessary. Further studies are also needed to determine whether estimates informed by genetic ancestry are associated with health outcomes. It remains to be seen whether differences associated with race or ethnic group in the response to medications that control asthma<sup>46,47</sup> are more tightly associated with estimates of ancestry. Although measures of individual genetic ancestry may foster the development of personalized medicine, large clinical trials and cohort studies that include assessments of genetic ancestry are needed to determine whether measures of ancestry are more useful clinically than a reliance on self-identified race alone.

Supported by contracts (N01-HC-48047 through 48050 and N01-HC-95095) from the National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health (NIH), as well as grants from the NIH (HL078885, HL088133, AI077439, and ES015794, to Dr. Burchard; K23HL093023-01, to Dr. Kumar; R01 HL71862, R01 HL71017, and R01 AG032136, to Dr. Reiner; and R01 AI79139, R01 AI61774, and R01 HL79055, to Dr. Williams), the Robert Wood Johnson Foundation Amos Medical Faculty Development Program and the Flight Attendant Medical Research Institute (to Dr. Burchard), the American Asthma Foundation (to Dr. Williams), and the Tobacco-Related Disease Research Program (New Investigator Award 15KT-0008, to Dr. Choudhry). The CHS was supported by NHLBI grants (N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, and U01 HL080295), with additional funding from the National Institute of Neurological Disorders and Stroke and a grant from the National Institute on Aging (NIA) (U19 AG023122 from the Longevity Consortium). The CARE Consortium was supported in part by a grant from the NHLBI (with a full description of the funding available at <http://public.nhlbi.nih.gov/GeneticsGenomics/home/care.aspx>). Health ABC was supported in part by the Intramural Research Program of the NIH and by the NIA, through contracts (N01-AG-6-2101, N01-AG-6-2103, and N01-AG-6-2106), and by a grant from the National Center for Research Resources (U54-RR020278) for the genotyping.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank the participants in the studies (CARDIA, HABC, CHS, SAGE, and SAPPHIRE) for their involvement; the study coordinators and staff for their dedication and commitment; the research institutions, study investigators, field staff, and study par-

ticipants for their contributions in creating the CARE Consortium, a resource for biomedical research; the UCSF Biostatistics High Performance Computing System, which was used to perform some computations; and Dr. Albert Levin for his input.

## REFERENCES

- Burchard EG, Ziv E, Coyle N, et al. The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med* 2003;348:1170-5.
- Phimister EG. Medicine and the racial divide. *N Engl J Med* 2003;348:1081-2.
- Cooper RS, Kaufman JS, Ward R. Race and genomics. *N Engl J Med* 2003;348:1166-70.
- Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999;159:179-87.
- Race, Ethnicity and Genetics Working Group. The use of racial, ethnic, and ancestral categories in human genetics research. *Am J Hum Genet* 2005;77:519-32.
- Goldstein DB, Hirschhorn JN. In genetic control of disease, does 'race' matter? *Nat Genet* 2004;36:1243-4.
- Yaeger R, Avila-Bront A, Abdul K, et al. Comparing genetic ancestry and self-described race in African Americans born in the United States and in Africa. *Cancer Epidemiol Biomarkers Prev* 2008;17:1329-38.
- Yang JJ, Burchard EG, Choudhry S, et al. Differences in allergic sensitization by self-reported race and genetic ancestry. *J Allergy Clin Immunol* 2008;122(4):820.e9-827.e9.
- Hoggart CJ, Parra EJ, Shriver MD, et al. Control of confounding of genetic associations in stratified populations. *Am J Hum Genet* 2003;72:1492-504.
- Rosenberg NA, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet* 2003;73:1402-22.
- Tang H, Peng J, Wang P, Risch NJ. Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol* 2005;28:289-301.
- Rosenberg NA, Pritchard JK, Weber JL, et al. Genetic structure of human populations. *Science* 2002;298:2381-5.
- Divers J, Moossavi S, Langefeld CD, Freedman BI. Genetic admixture: a tool to identify diabetic nephropathy genes in African Americans. *Ethn Dis* 2008;18:384-8.
- Reiner AP, Carlson CS, Ziv E, Iribarren C, Jaquish CE, Nickerson DA. Genetic ancestry, population sub-structure, and cardiovascular disease-related traits among African-American participants in the CARDIA Study. *Hum Genet* 2007;121:565-75.
- Tang H, Jorgenson E, Gadde M, et al. Racial admixture and its impact on BMI and blood pressure in African and Mexican Americans. *Hum Genet* 2006;119:624-33.
- Hancock DB, Eijgelsheim M, Wilk JB, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 2010;42:45-52.
- Hubert HB, Fabsitz RR, Feinleib M, Gwinn C. Genetic and environmental influences on pulmonary function in adult twins. *Am Rev Respir Dis* 1982;125:409-15.
- McCleary GE, Svartengren M, Pedersen NL, Heller DA, Plomin R. Genetic and environmental influences on pulmonary function in aging Swedish twins. *J Gerontol* 1994;49:264-8.
- Redline S, Tishler PV, Rosner B, et al. Genotypic and phenotypic similarities in pulmonary function among family members of adult monozygotic and dizygotic twins. *Am J Epidemiol* 1989;129:827-36.
- Friedman GD, Cutter GR, Donahue RP, et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol* 1988;41:1105-16.
- Cutter GR, Burke GL, Dyer AR, et al. Cardiovascular risk factors in young adults: the CARDIA baseline monograph. *Control Clin Trials* 1991;12:Suppl:1S-77S.
- Visser M, Newman AB, Nevitt MC, et al. Reexamining the sarcopenia hypothesis: muscle mass versus muscle strength. *Ann N Y Acad Sci* 2000;904:456-61.
- Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* 1991;1:263-76.
- Tsai HJ, Shaikh N, Kho JY, et al. Beta 2-adrenergic receptor polymorphisms: pharmacogenetic response to bronchodilator among African American asthmatics. *Hum Genet* 2006;119:547-57.
- American Thoracic Society. Standardization of spirometry, 1994 update. *Am J Respir Crit Care Med* 1995;152:1107-36.
- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009;19:1655-64.
- Nalls MA, Wilson JG, Patterson NJ, et al. Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart studies. *Am J Hum Genet* 2008;82:81-7. [Erratum, *Am J Hum Genet* 2008;82:532.]
- McKeigue PM, Carpenter JR, Parra EJ, Shriver MD. Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. *Ann Hum Genet* 2000;64:171-86.
- Parra EJ, Kittles RA, Argyropoulos G, et al. Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *Am J Phys Anthropol* 2001;114:18-29.
- Reiner AP, Ziv E, Lind DL, et al. Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study. *Am J Hum Genet* 2005;76:463-77.
- Wu B, Liu N, Zhao H. PSMIX: an R package for population structure inference via maximum likelihood method. *BMC Bioinformatics* 2006;7:317.
- Galanter J, Choudhry S, Eng C, et al. ORMDL3 gene is associated with asthma in three ethnically diverse populations. *Am J Respir Crit Care Med* 2008;177:1194-200.
- Fletcher C, Peto R. The natural history of chronic airflow obstruction. *BMJ* 1977;1:1645-8.
- Tager IB, Segal MR, Speizer FE, Weiss ST. The natural history of forced expiratory volumes: effect of cigarette smoking and respiratory symptoms. *Am Rev Respir Dis* 1988;138:837-49.
- National Asthma Education and Prevention Program. Expert Panel Report 3 (EPR-3): guidelines for the diagnosis and management of asthma — summary report 2007. *J Allergy Clin Immunol* 2007;120:Suppl:S94-S138.
- U.S. Census Bureau. Annual estimates of the resident population by sex, race, and Hispanic origin for the United States: April 1, 2000 to July 1, 2008. (Accessed June 21, 2010, at <http://www.census.gov/popest/national/asrh/NC-EST2008/NC-EST2008-03.xls>.)
- Marsh SE, Travers J, Weatherall M, et al. Proportional classifications of COPD phenotypes. *Thorax* 2008;63:761-7.
- Shirtcliffe P, Weatherall M, Marsh S, et al. COPD prevalence in a random population survey: a matter of definition. *Eur Respir J* 2007;30:232-9.
- Rabe KF, Hurd S, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007;176:532-55.
- Celli BR, MacNee W. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J* 2004;23:932-46. [Erratum, *Eur Respir J* 2006;27:242.]
- Guides to the evaluation of permanent impairment. 6th ed. Chicago: American Medical Association, 2008.
- Arif AA, Delclos GL, Lee ES, Tortolero

- SR, Whitehead LW. Prevalence and risk factors of asthma and wheezing among US adults: an analysis of the NHANES III data. *Eur Respir J* 2003;21:827-33.
43. Raby BA, Celedon JC, Litonjua AA, et al. Low-normal gestational age as a predictor of asthma at 6 years of age. *Pediatrics* 2004;114(3):e327-e332.
44. Rona RJ, Gulliford MC, Chinn S. Effects of prematurity and intrauterine growth on respiratory health and lung function in childhood. *BMJ* 1993;306:817-20.
45. Aldrich MC, Selvin S, Hansen HM, et al. Comparison of statistical methods for estimating genetic admixture in a lung cancer study of African Americans and Latinos. *Am J Epidemiol* 2008;168:1035-46.
46. Bloche MG. Race-based therapeutics. *N Engl J Med* 2004;351:2035-7.
47. Lemanske RF Jr, Mauger DT, Sorkness CA, et al. Step-up therapy for children with uncontrolled asthma receiving inhaled corticosteroids. *N Engl J Med* 2010;362:975-85.

*Copyright © 2010 Massachusetts Medical Society.*