Lung function decline over 25 years of follow-up among black and white adults in the ARIC study cohort

Maria C. Mirabelli, PhD, MPH; John S. Preisser, PhD; Laura R Loehr, MD, PhD; Sunil K. Agarwal MD, MPH, PhD; R. Graham Barr, MD, DrPH; David J. Couper, PhD; John L. Hankinson, PhD; Noorie Hyun, PhD; Aaron R. Folsom, MD, MPH; Stephanie J. London, MD, DrPH

Online Supplement

Dropout versus Death in the ARIC Study

The inverse-probability-weighted independence estimating equations conditioning-on-being-alive methods that we used to estimate lung function decline construct weights that depend on the probability of dropout among those participants who are still alive while removing from the pool at risk for dropout those who died before being examined at the next study visit. Table S1 reports the percentages of death, non-death dropout, visit attendance without valid spirometry, and visit attendance with valid spirometry during the follow-up periods between visits 1 and 2 and visits 2 and 5. As shown in Table S1, prior to visit 2, deaths were infrequent in all categories of smoking status. However, in the interval between visit 2 and visit 5, 43.9% of current smokers versus 28.9% of former smokers and 21.4% of never smokers died. So, while FEV_1 decline was estimated to be larger for former and never than current smokers by age 60, a larger percentage of current smokers than former and never smokers did not survive to older ages.

Predicted Probability of Attrition from the ARIC Study: Models of Predicted Probability

To address the scenario in which follow-up time in longitudinal studies is both affected by dropout and truncated by participant death, Kurland and Heagerty (2005) proposed regression methods for inference to populations continuously redefined to consist of members remaining alive at the time of observation. Regression conditioning-on-being-alive incorporates visit-specific weights for each participant determined from a logistic regression model for differential dropout [1]. Specifically, predicted probabilities of not dropping out are determined for participant observations conditional upon being alive at the time of the visit and having not dropped out at the previous visit. All participants in our analysis completed spirometry at visit 1 and we estimated the conditional probabilities of attending and completing visits 2, 3, 4, and 5.

Although spirometry data collection was not included in the ARIC study visits 3 and 4, we estimated predicted probabilities for these visits in order to improve estimation of analytic weights across the approximately 21-year gap between visits 2 and 5. The logistic model included the following four groups of variables: (1) educational attainment, race, sex, study center, and study visit; (2) a combined measure of lung function defined as the average z-scores of FEV_1 and FVC at the last clinic visit that included spirometry data collection, imputed for visits 3 and 4 based on conditional mean (i.e., regression) imputation; (3) time-updated (updated at the previous visit) information for age (linear and quadratic terms), asthma history, body mass index (linear and quadratic terms), chronic obstructive lung disease, height (linear and quadratic), prevalent coronary heart disease, and smoking status; and (4) a participant-level variable for survival indicating the last visit at which the participant was known to be alive [1]. The model also included interaction terms for the following groups of interactions: all two-way interactions between educational attainment, race, sex, smoking status, and study visit; two-way interactions of the average z-score with educational attainment, prevalent coronary heart disease, race, sex, smoking status, and study visit; three-way interaction of the average z-score with race and sex; and interaction of the participant-level survival variable with the average z-score. The model of predicted probability of participation included an array of explanatory variables intended to improve the validity of the predictions. We developed the logistic regression model to improve prediction for dropout with minimal concern about the number of explanatory variables because statistical efficiency of the estimates of lung function decline in the longitudinal model does not diminish when more predictor variables are added to the dropout model [2, 3]; the only concern is that a wide range of weights may cause unstable estimates of lung function decline [3]. In our study, more than 99% of the weights were between 1 and 5, and

the relatively few larger weights were truncated at 5 to reduce the possibility of unstable estimates. Estimates, with standard errors, for all coefficients in the logistic regression model are shown in Table S6.

The logistic regression model resulted in one predicted probability for each participant at each study visit, representing the probability that the participant completed each study visit (i.e., each person-visit), conditional on being alive at the time of the study visit. These probabilities were used to compute participant-level weights, which were computed as the inverse of the cumulative product of visit-specific probabilities and were used in inverse-probability-weighted independence estimating equations for regression-conditioning-on-being-alive models [1] developed to estimate lung function decline across categories of race-, sex-, and smoking status. Estimating weights in this way allowed data collected from participants with characteristics similar to those of participants without data due to dropout to be up-weighted in the main lung function analyses and therefore to represent the contributions that are both observed and not observed. Satisfying the model assumptions of the inverse-probability-weighted regression conditioning-on-being-alive method requires monotonic patterns of missing data [1, 3], which were achieved in these analyses by treating each participant's observed visit 5 as missing if visit 2 was missing.

Estimation of Lung Function at Baseline and Lung Function Decline

To estimate FEV_1 and FVC at age 45 years, FEV_1/FVC at age 45 years, and changes in FEV_1 and FVC across categories of race, sex, and smoking status for the birth cohort that was 45–64 years of age at baseline, we used linear population-averaged inverse-probability-weighted

regression conditioning-on-being-alive models that incorporated the visit-specific weights described above. The models assumed a working constant variance function over time and were specified with independent working correlation matrices. Use of working independence correlation matrices are required to obtain valid regression coefficient estimates when using standard software [4, 5]. Final regression analyses included data from each visit completed by all 13,896 participants, with the exception that to have monotone patterns of missingness, as required by the weighting methods, visit 5 was omitted if visit 2 was missing. Monotonicity was imposed because weighted GEE methods for intermittent patterns of missingness involve more complex methods that are infrequently used. One notable advantage of the strategy of imposing monotonicity is that the underlying assumptions required are weaker than assumptions required for a model that allows intermittent data patterns [3].

In our study, FEV_1 and FVC were analyzed in separate models using spirometry data from visits 1, 2, and 5 and based on a common set of visit-specific weights described above. Spirometric measures were regressed on linear and quadratic terms for baseline age (centered at 45 years and divided by 10), asthma history, linear and quadratic terms for body mass index, chronic obstructive lung disease history, linear and quadratic terms for height, race, sex, smoking status, study center, and linear and quadratic terms for time (viz., years since the baseline visit divided by 10). Inclusion of both baseline age and time accommodated possible period effects that could occur due to changes in the spirometry protocol, spirometry equipment, or other changes over time. All two-way interactions among race, sex, smoking status, and linear and quadratic time were included, as were the three-way interactions of linear and quadratic terms for time with race and sex, the linear term for time with race and smoking status, the linear term for time with sex

and smoking status, and sex with race and smoking status. Finally, we also included four-way interactions of linear and quadratic terms for time with sex, race, and smoking status. The remaining four-way interactions between linear and quadratic terms for time, race, sex, and smoking status were not statistically significant at α=0.15 and were dropped from the final model. Standard errors for regression coefficients were obtained from empirical sandwich variance estimators as produced by applying SAS Proc Genmod (SAS Institute Inc., Cary, North Carolina) with weights, which is known to overestimate the standard errors [4]. In a simulation study of longitudinal binary data, such standard errors were overestimated by about 15% compared to the standard errors produced by the correct formula, implementation of which is challenging and frequently avoided [6]. In our study, the overestimation (not shown) was less than 4%, so that the conservative formula used when applying SAS Proc Genmod gives reasonable results. Analyses were performed using SAS version 9.3 (SAS Institute Inc., Cary, North Carolina).

Because the youngest participants were 45 years of age at visit 1, we generated population-averaged mean estimates of FEV_1, FVC, and FEV_1/FVC with 95% confidence intervals (CIs) at age 45 years to represent baseline measures of lung function for each of the 12 categories of race, sex, and smoking status. Adjusted estimates of FEV_1, FVC, and FEV_1/FVC at age 45 were generated using baseline mean values of all covariates included in the regression equations. Because of the non-linear specification of time, our results generated annual changes in FEV_1 and FVC that are not constant over the follow-up period. That is, the estimated annual changes are not constant across the range of ages included in our analysis and therefore cannot be summarized with a single set of race-, sex-, and smoking status-specific estimates. To most

clearly present these results, final estimates are presented as estimated annual lung function declines at the *a priori* selected ages of 45, 60, and 75 years. Estimated annual declines at age 45 years were calculated as the change from ages 45 to 46; estimated annual declines at age 60 years were calculated as the change from ages 60 to 61, and estimated annual declines at age 75 years were calculated as the change from ages 75 to 76 years. All estimates were generated using observations from 13,896 participants at up to 3 clinic visits, for a total of 29,111 observations.

Because participants who were aged 45 years at the baseline visit were more likely to be followed through visit 5 than were older participants, we focused our inference on this youngest birth cohort. Our statistical models include both baseline age and time effects. In these models, time is a period effect. The basic model for the *i*-th person at time t is $y_{it} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{2i1} + \beta_3(x_{it} - x_{i1}) + \beta_4(x_{it} - x_{i1})^2 + \varepsilon_{it}$, where x_{i1} is age at visit t (t=1,2,5) centered at 45 years and divided by 10. Thus, x_{i1} is baseline (visit 1) age in decades centered at 4.5. The period effect is reflected by the time since baseline ($x_{it} - x_{i1}$). Our inference focuses on lung function decline in the birth cohort that is aged 45 at baseline, reducing the model above to $E(y_{it}) = \beta_0 + \beta_3 d + \beta_4 d^2$, where $d=(x_{it} - 4.5)$ equals the time (number of decades) since baseline. Annual changes vary by age due to the quadratic effect, for example:

Expected change from Age 45 to 46: $E(y_{it}|d=0.1) - E(y_{it}|d=0) = 0.1\ \beta_3 + 0.01\ \beta_4$

Expected change from Age 60 to 61: $E(y_{it}|d=1.6) - E(y_{it}|d=1.5) = 0.1\ \beta_3 + 0.31\ \beta_4$

Expected change from Age 75 to 76: $E(y_{it}|d=3.1) - E(y_{it}|d=3.0) = 0.1\ \beta_3 + 0.61\ \beta_4$

Moreover, these values will vary according to race, sex, and smoking group according to the interactions of these variables with time (Tables S3 and S4). While ARIC participants who were 45 years at visit 1 would be no older than 70 years of age at visit 5, Table S2 shows that there are

individuals of other ages (i.e., other birth cohorts) at visit 5. Information from these other birth cohorts is extrapolated through use of our statistical model to generate annual changes at age 75 reported in Tables 3 and 4 for the cohort that was age 45 years at baseline.

REFERENCES

1. Kurland BF, Heagerty PJ. Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. Biostatistics 2005; 6(2):241-58.

2. Preisser JS, Galecki AT, Lohman KK, Wagenknecht LE. Analysis of smoking trends with incomplete longitudinal binary responses. Journal of the American Statistical Association 2000; 95(452):1021-1031.

3. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association 1995; 90(429):106-121.

4. Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis, 2nd edition. 2004, Hoboken, N.J.: Wiley-Interscience.

5. Tchetgen Tchetgen EJ, Glymour MM, Weuve J, Robins J. Specifying the correlation structure in inverse-probability- weighting estimation for repeated measures. Epidemiology 2012; 23(4):644-6.

6. Preisser JS, Lohman KK, Rathouz PJ. Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random. Statistics in Medicine 2002; 21(20):3035-3054.

Table S1. Participant dropout and death during follow-up periods between study visits 1, 2, and 5

	All participants	Smoking status[1]		
		Current	Former	Never
Follow-up Period	No. (%)[2,3]	No. (%)[3]	No. (%)[3]	No. (%)[3]
Visit 1 → Visit 2 (approximately 3 years)				
Population at visit 1	13,896 (100.0)[4]	3,603 (25.9)[4]	4,500 (32.4)[4]	5,793 (41.7)[4]
Death before visit 2	256 (1.8)	114 (3.2)	74 (1.6)	68 (1.2)
Non-death dropout before visit 2	1,244 (9.0)	510 (14.2)	328 (7.3)	406 (7.0)
Attended visit 2, but no valid spirometry data at visit 2	938 (6.8)	272 (7.6)	300 (6.7)	366 (6.3)
Valid spirometry data at visit 2	11,458 (82.5)	2,707 (75.1)	3,798 (84.4)	4,953 (85.5)
Visit 2 → Visit 5 (approximately 21 years)				
Population at visit 2	11,458 (100.0)[4]	2,492 (21.7)[4]	4,474 (39.0)[4]	4,492 (39.2)[4]
Death before visit 5	3,348 (29.2)	1,095 (43.9)	1,291 (28.9)	962 (21.4)
Non-death dropout before visit 5	2,720 (23.7)	880 (35.3)	527 (11.8)	1,313 (29.2)
Attended visit 5, but no valid spirometry data at visit 5	1,593 (13.9)	301 (12.1)	591 (13.2)	701 (15.6)
Valid spirometry data at visit 5	3,797 (33.1)	216 (8.7)	2065 (46.2)	1,516 (33.7)

[1] For the follow-up period between visits 1 and 2, smoking status is based on smoking status at visit 1; for the follow-up period between visits 2 and 5, smoking status is based on smoking status at visit 2.

[2] Percentages during follow-up between visits 1 and 2 are based on the population at visits 1 (n=13,896) and percentages during follow-up between visits 2 and 5 are based on the population at visit 2 (n=11,458)

[3] Column percentages, unless otherwise specified

[4] Row percentages

Table S2. Demographic and health-related characteristics of the study sample at visit 5

	All Participants	Race	
		Black	White
	No. (%)[1]	No. (%)[1]	No. (%)[1]
Total	3,797 (100.0)	721 (19.0)[2]	3,076 (81.0)[2]
Demographic characteristics			
Age			
65–69	472 (12.4)	128 (17.8)	344 (11.1)
70–74	1,457 (38.4)	306 (42.4)	1,151 (37.4)
75–79	1,052 (27.7)	170 (23.6)	882 (28.7)
80–84	628 (16.5)	95 (13.2)	533 (17.3)
85–90	188 (5.0)	22 (3.1)	166 (5.4)
Sex			
Female	2,263 (59.6)	503 (69.8)	1,760 (57.2)
Male	1,534 (40.4)	218 (30.2)	1,316 (42.8)
Health-related characteristics			
Asthma			
No	3,446 (90.8)	653 (90.6)	2,793 (90.8)
Yes	351 (9.2)	68 (9.4)	283 (9.2)
Body mass index			
≤18.4	30 (0.8)	5 (0.7)	25 (0.8)
18.5–24.9	953 (25.1)	116 (16.1)	837 (27.2)
25.0–29.9	1,521 (40.1)	266 (36.9)	1,255 (40.8)
≥30.0	1,293 (34.1)	334 (46.3)	959 (31.2)
Chronic obstructive lung disease[3]			
No	3,469 (91.4)	674 (93.5)	2,795 (90.9)
Yes	328 (8.6)	47 (6.5)	281 (9.1)
Smoking status			
Current smoker	216 (5.7)	47 (6.5)	169 (5.5)
Former smoker	2,065 (54.4)	351 (48.7)	1,714 (55.7)
Never smoker	1,516 (39.9)	323 (44.8)	1,193 (38.8)
Pack-years of smoking at visit 1[4]			
0	1,479 (39.0)	265 (36.8)	1,214 (39.5)
0.01–10.99	475 (12.5)	99 (13.7)	376 (12.2)
11.00–24.74	539 (14.2)	99 (13.7)	440 (14.3)
24.75–38.99	576 (15.2)	120 (16.6)	456 (14.8)
≥39.00	626 (16.5)	113 (15.7)	513 (16.7)
Unknown	102 (2.7)	25 (3.5)	77 (2.5)

[1] Column percentages, unless otherwise noted
[2] Row percentages
[3] Other than asthma
[4] Pack-years not recorded at visit 5

Table S3. Regression coefficients generated from inverse-probability-weighted regression conditioning-on-being-alive model of annual change in FEV_1, in liters

Variable[1]	β (SE)[1]
Intercept	3.7265 (0.0201)
Study Center=Jackson, MS	0.0110 (0.0281)
Study Center=Suburbs of Minneapolis, MN	0.0664 (0.0118)
Study Center=Washington County, MD	-0.0356 (0.0119)
Black	-0.5595 (0.0392)
Female	-0.6229 (0.0179)
Age	-0.3256 (0.0257)
Age^2	-0.0083 (0.0133)
Time	-0.5228 (0.0373)
$Time^2$	0.0422 (0.0148)
Height	0.0297 (0.0006)
$Height^2$	0.0003 (0.0000)
Body Mass Index	-0.0041 (0.0012)
Body Mass $Index^2$	-0.0004 (0.0001)
Current Smoker	-0.5504 (0.0223)
Former Smoker	-0.1992 (0.0180)
History of Asthma	-0.2470 (0.0186)
History of Chronic Obstructive Lung Disease Other than Asthma	-0.3734 (0.0207)
Time*Black	0.2317 (0.0799)
Time*Female	0.1356 (0.0383)
Time*Current Smoker	-0.1602 (0.0502)
Time*Former Smoker	0.0192 (0.0411)
$Time^2$*Black	-0.0784 (0.0304)
$Time^2$*Female	-0.0231 (0.0151)
$Time^2$*Current Smoker	0.0846 (0.0183)
$Time^2$*Former Smoker	-0.0052 (0.0159)
Black*Female	0.1754 (0.0333)
Black*Current Smoker	0.2819 (0.0440)
Black*Former Smoker	0.1082 (0.0409)
Female*Current Smoker	0.2032 (0.0266)
Female*Former Smoker	0.1221 (0.0212)
Time*Female*Black	-0.0914 (0.0916)
Time*Black*Current Smoker	0.0469 (0.0584)
Time*Black*Former Smoker	0.0285 (0.0395)
Time*Female*Current Smoker	0.0060 (0.0360)
Time*Female*Former Smoker	-0.0074 (0.0158)
$Time^2$*Female*Black	0.0517 (0.0353)
Female*Black*Current Smoker	-0.0650 (0.0513)
Female*Black*Former Smoker	-0.0467 (0.0475)
Time*Female*Black*Current Smoker	-0.1403 (0.0708)
$Time^2$*Female*Black*Former Smoker	-0.0247 (0.0431)

[1] Categories not shown (e.g., white race, male, and never smoker) are referent categories

Table S4. Regression coefficients generated from inverse-probability-weighted regression conditioning-on-being-alive model of annual change in FVC, in liters

Variable[1]	β (SE)[1]
Intercept	4.7806 (0.0233)
Study Center=Jackson, MS	-0.0655 (0.0325)
Study Center=Suburbs of Minneapolis, MN	0.0767 (0.0134)
Study Center=Washington County, MD	-0.0475 (0.0133)
Black	-0.7954 (0.0460)
Female	-0.8114 (0.0208)
Age	-0.3360 (0.0293)
Age^2	0.0002 (0.0152)
Time	-0.6763 (0.0416)
$Time^2$	0.0809 (0.0168)
Height	0.0447 (0.0007)
$Height^2$	0.0004 (0.0000)
Body Mass Index	-0.0200 (0.0013)
Body Mass $Index^2$	-0.0001 (0.0001)
Current Smoker	-0.3788 (0.0247)
Former Smoker	-0.1143 (0.0204)
History of Asthma	-0.1453 (0.0191)
History of Chronic Obstructive Lung Disease Other than Asthma	-0.2683 (0.0203)
Time*Black	0.2246 (0.0959)
Time*Female	0.2265 (0.0423)
Time*Current Smoker	-0.1385 (0.0541)
Time*Former Smoker	-0.0140 (0.0452)
$Time^2$*Black	-0.0615 (0.0353)
$Time^2$*Female	-0.0629 (0.0168)
$Time^2$*Current Smoker	0.0890 (0.0213)
$Time^2$*Former Smoker	0.0072 (0.0176)
Black*Female	0.3443 (0.0392)
Black*Current Smoker	0.2643 (0.0498)
Black*Former Smoker	0.1326 (0.0463)
Female*Current Smoker	0.0906 (0.0294)
Female*Former Smoker	0.0829 (0.0242)
Time*Female*Black	-0.1728 (0.1085)
Time*Black*Current Smoker	-0.0271 (0.0732)
Time*Black*Former Smoker	0.0310 (0.0486)
Time*Female*Current Smoker	-0.0111 (0.0377)
Time*Female*Former Smoker	0.0025 (0.0186)
$Time^2$*Female*Black	0.0735 (0.0407)
Female*Black*Current Smoker	-0.0736 (0.0580)
Female*Black*Former Smoker	-0.0904 (0.0538)
Time*Female*Black*Current Smoker	-0.0905 (0.0864)
$Time^2$*Female*Black*Former Smoker	-0.0240 (0.0525)

[1] Categories not shown (e.g., white race, male, and never smoker) are referent categories

Table S5. Unweighted generalized estimating equation (GEE) estimates of FEV_1, FVC, and FEV_1/FVC at age 45 years, by race, sex, and smoking status: the ARIC Study

	FEV_1, in ml	FVC, in ml	FEV_1/FVC, %
	Mean (95% CI)[1]	Mean (95% CI)[1]	Mean (95% CI)[1]
Black			
Female			
Current smoker	2,578 (2,528–2,627)	3,401 (3,345–3,456)	75.6 (74.7–76.4)
Former smoker	2,692 (2,638–2,746)	3,501 (3,442–3,560)	76.8 (75.9–77.7)
Never smoker	2,711 (2,666–2,757)	3,498 (3,446–3,551)	77.9 (77.2–78.6)
Male			
Current smoker	2,872 (2,809–2,935)	3,828 (3,759–3,897)	74.6 (73.7–75.6)
Former smoker	3,054 (2,991–3,118)	3,964 (3,894–4,034)	76.6 (75.7–77.6)
Never smoker	3,139 (3,075–3,204)	3,942 (3,867–4,016)	79.2 (78.3–80.1)
White			
Female			
Current smoker	2,746 (2,713–2,779)	3,665 (3,629–3,702)	74.2 (73.7–74.7)
Former smoker	3,017 (2,986–3,047)	3,919 (3,885–3,954)	77.1 (76.6–77.6)
Never smoker	3,096 (3,069–3,122)	3,953 (3,923–3,984)	78.8 (78.4–79.2)
Male			
Current smoker	3,149 (3,108–3,190)	4,359 (4,313–4,405)	72.7 (72.0–73.3)
Former smoker	3,497 (3,462–3,532)	4,616 (4,576–4,656)	76.2 (75.7–76.7)
Never smoker	3,696 (3,660–3,733)	4,732 (4,689–4,774)	78.8 (78.3–79.3)

[1]Adjusted for baseline characteristics (age, height, study center) and time-varying characteristics collected at baseline and updated at visits 2 and 5 (asthma history, body mass index, chronic obstructive lung disease, follow-up time)

Table S6. Regression coefficients generated from the logistic model developed to estimate predicted probabilities of completing spirometry, conditioning-on-being-alive

Variable[1]	β (SE)
Intercept	1.2741 (0.2931)
Study Center=Jackson, MS	0.5524 (0.0882)
Study Center=Suburbs of Minneapolis, MN	0.4311 (0.0426)
Study Center=Washington County, MD	0.5493 (0.0431)
Age	-0.1138 (0.0168)
Age^2	-0.0178 (0.0159)
Height	-0.0741 (0.0246)
Height^2	-0.0281 (0.0128)
Body Mass Index	0.0151 (0.0164)
Body Mass Index^2	-0.0003 (0.0003)
History of Asthma	0.1851 (0.0639)
History of Chronic Obstructive Lung Disease Other than Asthma	0.1643 (0.0650)
Black	-1.1705 (0.1408)
Female	0.1301 (0.1329)
Current Smoker	-0.4825 (0.1243)
Former Smoker	0.1488 (0.1164)
Educational Attainment=12–16 years	0.4315 (0.1192)
Educational Attainment=17–21 years	0.5618 (0.1236)
Prevalent Coronary Heart Disease	-0.1598 (0.0618)
Z-score	0.0431 (0.0804)
Visit 3 Completed	1.3143 (0.1536)
Visit 4 Completed	0.7616 (0.1495)
Visit 5 Completed	-2.4443 (0.1305)
Survival at Visit 2	-0.9562 (0.1294)
Survival at Visit 3	-0.8062 (0.1217)
Survival at Visit 4	-0.3161 (0.0576)
Survival at Visit 3*Visit 3 Completed	-0.7251 (0.1942)
Survival at Visit 4*Visit 3 Completed	-0.0012 (0.0984)
Survival at Visit 4*Visit 4 Completed	-0.1048 (0.0969)
Black*Female	0.1597 (0.0973)
Current Smoker*Black	0.3491 (0.0908)
Former Smoker*Black	0.1909 (0.0836)
Black*Educational Attainment=12–16 years	-0.1327 (0.0847)
Black*Educational Attainment=17–21years	0.0315 (0.0878)
Current Smoker*Female	0.0677 (0.1143)
Former Smoker*Female	0.0963 (0.1003)
Female*Educational Attainment=12–16 years	0.0164 (0.1075)
Female*Educational Attainment=17–21 years	-0.0639 (0.1104)

(Continued on next page)

Table S6 (continued)

Variable[1]	β (SE)
Current Smoker*Educational Attainment=12–16 years	-0.1071 (0.1011)
Current Smoker*Educational Attainment=17–21 years	0.0226 (0.1079)
Former Smoker*Educational Attainment=12–16 years	0.0014 (0.0947)
Former Smoker*Educational Attainment=17–21 years	-0.0117 (0.0958)
Visit 3 Completed *Black	-0.0249 (0.0958)
Visit 4 Completed *Black	0.2561 (0.0989)
Visit 5 Completed *Black	0.9846 (0.0919)
Visit 3 Completed *Female	-0.4154 (0.1222)
Visit 4 Completed *Female	-0.2294 (0.1242)
Visit 5 Completed *Female	0.2366 (0.1090)
Visit 3 Completed*Educational Attainment=12–16 years	-0.1426 (0.1083)
Visit 3 Completed *Educational Attainment=17–21 years	-0.0797 (0.1152)
Visit 4 Completed *Educational Attainment=12–16 years	-0.0472 (0.1110)
Visit 4 Completed *Educational Attainment=17–21 years	-0.0894 (0.1164)
Visit 5 Completed *Educational Attainment=12–16 years	-0.0672 (0.1037)
Visit 5 Completed *Educational Attainment=17–21 years	0.0167 (0.1063)
Current Smoker*Visit 3 Completed	-0.5883 (0.1096)
Current Smoker*Visit 4 Completed	-0.1868 (0.1153)
Current Smoker*Visit 5 Completed	-0.2574 (0.1162)
Former Smoker*Visit 3 Completed	-0.3579 (0.1098)
Former Smoker*Visit 4 Completed	-0.2871 (0.1061)
Former Smoker*Visit 5 Completed	0.1697 (0.0864)
Z-score*Visit 3 Completed	-0.1674 (0.0628)
Z-score*Visit 4 Completed	0.0040 (0.0663)
Z-score*Visit 5 Completed	0.3342 (0.0601)
Z-score*Black	0.0643 (0.1120)
Z-score*Female	0.1451 (0.0553)
Current Smoker*Z-score	0.0521 (0.0608)
Former Smoker*Z-score	-0.0180 (0.0534)
Z-score*Educational Attainment=12–16 years	0.0115 (0.0583)
Z-score*Educational Attainment=17–21 years	0.0142 (0.0607)
Z-score*Survival at Visit 2	0.1451 (0.1186)
Z-score*Survival at Visit 3	0.2053 (0.0934)
Z-score*Survival at Visit 4	0.1084 (0.0404)
Z-score*Study Center=Jackson, MS	0.0116 (0.0974)
Z-score*Study Center=Suburbs of Minneapolis, MN	0.0077 (0.0416)
Z-score*Study Center=Washington County, MD	-0.0365 (0.0443)
Prevalent Coronary Heart Disease*Z-score	0.0985 (0.0649)
Z-score*Black*Female	-0.1684 (0.1004)

[1] Categories not shown (e.g., white race, male, and never smoker) are referent categories