

Data Dictionary and Harmonization Protocol for the NHLBI Pooled Cohorts Study

Version: 1.0

Last Updated: April 10, 2018

The aims and methods for the NHLBI Pooled Cohorts Study are described in the following manuscript:

Oelsner EC, Balte PP, Cassano P, Couper D, Enright PL, Folsom AR, Hankinson J, Jacobs, DR, Jr., Kalhan R, Kaplan R, Kronmal R, Lange L, Loehr LR, London SJ, Navas Acien A, Newman AB, O'Connor GT, Schwartz JE, Smith LJ, Yeh F, Zhang Y, Moran AE, Mwasongwe S, White W, Yende S, Barr RG. Harmonization of Respiratory Data From Nine US Population-Based Cohorts: The NHLBI Pooled Cohorts Study. *Am J Epidemiol.* 2018.

To analyze the NHLBI Pooled Cohorts Study data, please contact:

Dr. Elizabeth C. Oelsner, MD MPH
Division of General Medicine
Columbia University Medical Center
630 West 168th Street, PH9-105
New York, NY 10032
Email: eco7@cumc.columbia.edu

Harmonized data have been transmitted back to the originating cohorts.

NHLBI Pooled Cohorts Study / Funding

This work is funded by numerous grants from the National Institutes of Health (NIH).

The NHLBI Pooled Cohorts Study is funded by NIH/National Heart, Lung, and Blood Institute (NHLBI) K23-HL-130627, R21-HL-129924, and R21-HL121457.

The Atherosclerosis Risk in Communities Study has been funded by NIH/NHLBI contracts. HHSN268201700001I, HHSN268201700003I, HHSN268201700005I, HHSN268201700004I, and HHSN268201700002I.

The Coronary Artery Risk Development in Young Adults Study has been funded by NIH/NHLBI contracts HHSN268201300025C, HHSN268201300026C, HHSN268201300027C, HHSN268201300028C, HHSN268201300029C, and HHSN268200900041C and the Intramural Research Program of the National Institute on Aging.

The Cardiovascular Health Study has been funded by NIH/NHLBI contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and grants U01HL080295, with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629 from the National Institute on Aging (NIA).

The Framingham Heart Study, Offspring Cohort has been funded by NIH/NHLBI contracts N01-HC-25195 and HHSN268201500001I.

The Health, Aging and Body Composition Study has been funded by NIH/NIA contracts N01-AG-6-2101, N01-AG-6-2103, and N01-AG-6-2106; NIA grant R01-AG028050; and National Institute of Nursing Research (NINR) grant R01-NR012459.

The Hispanic Community Health Study/Study of Latinos has been funded by NIH/NHLBI contracts to the University of North Carolina (HHSN268201300001I / N01-HC-65233), University of Miami (HHSN268201300004I / N01-HC-65234), Albert Einstein College of Medicine (HHSN268201300002I / N01-HC-65235), University of Illinois at Chicago – HHSN268201300003I / N01-HC-65236 Northwestern Univ), and San Diego State University (HHSN268201300005I / N01-HC-65237). The following Institutes/Centers/Offices have contributed to the Study through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological Disorders and Stroke, NIH Institution-Office of Dietary Supplements.

The Jackson Heart Study has been funded by NIH/NHLBI and NIH/National Institute for Minority Health and Health Disparities (NIMHD) contracts to Jackson State University (HHSN268201300049C, HHSN268201300050C), Tougaloo College (HHSN268201300048C), and the University of Mississippi Medical Center (HHSN268201300046C, HHSN268201300047C).

The Multi-Ethnic Study of Atherosclerosis has been funded by NIH/NHLBI contracts RC1-HL-100543, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168 and N01-HC-95169; and grants R01-HL-077612 and R01-

HL-093081. This publication was also developed under a STAR research assistance agreement, No. RD831697 (MESA Air), awarded by the U.S Environmental Protection Agency.

The Strong Heart Study has been funded by NIH/NHLBI cooperative agreement grants U01-HL41642, U01-HL41652, U01-HL41654, U01-HL65520, and U01-HL65521 and research grants R01-HL109315, R01HL109301, R01HL109284, R01HL109282 and R01HL109319.

This work was also funded by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.

Table of Contents

Cohort	Page
Atherosclerosis Risk in Communities (ARIC) Study	5
Cardiovascular Risk Development in Young Adults (CARDIA)	15
Cardiovascular Health Study (CHS)	26
Framingham Heart Study – Offspring Cohort (FHS-O)	36
Health, Aging and Body Composition (HABC)	45
Hispanic Community Health Study / Study of Latinos (HCHS/SOL)	54
Jackson Heart Study	61
Multi-Ethnic Study of Atherosclerosis (MESA)	69
Strong Heart Study (SHS)	76

Figure 1. Flow Chart of Longitudinal Lung Function Data in the Atherosclerosis Risk In Communities Study, by Visit, United States, 1989-2013.

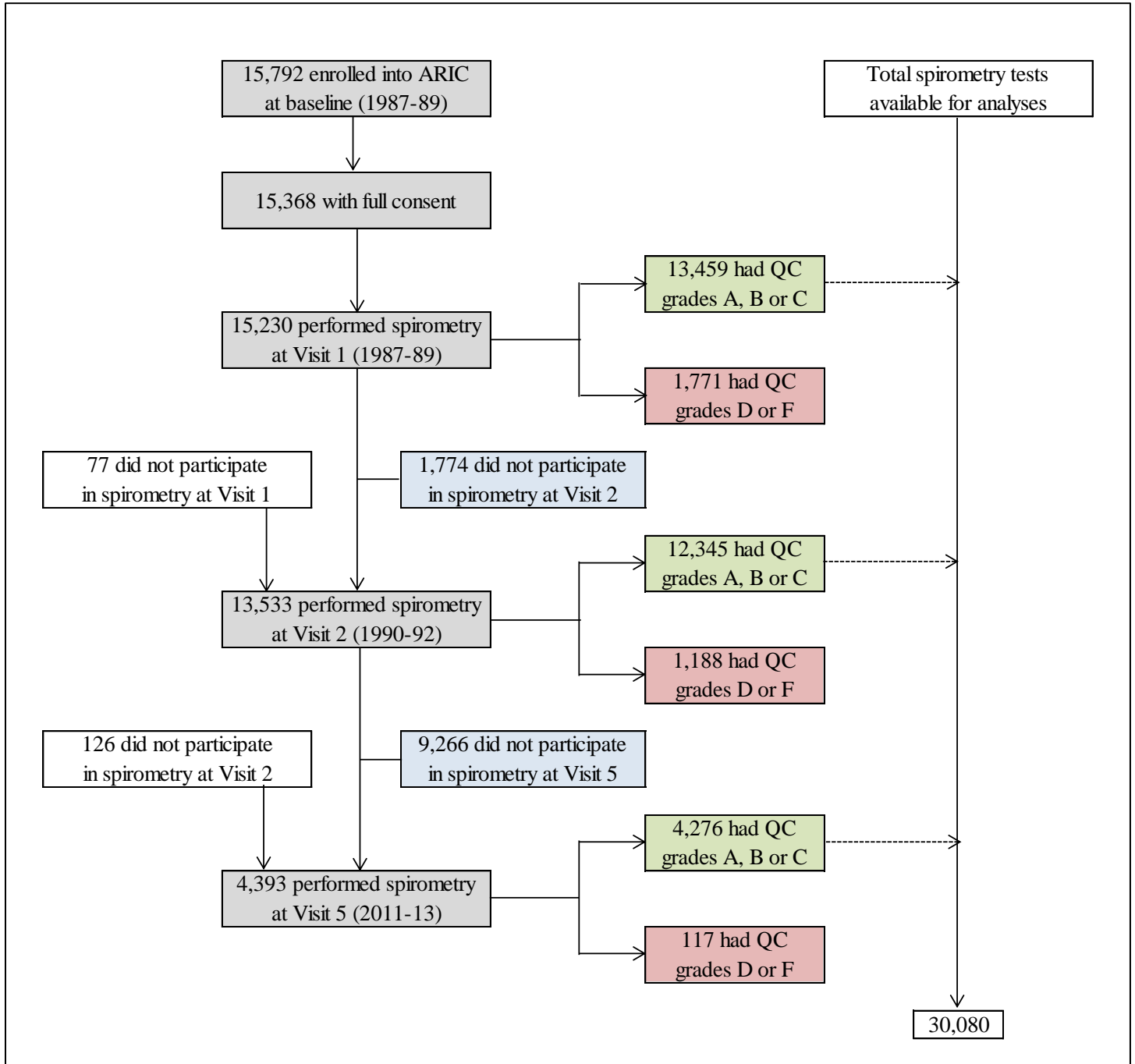


Table 1. Harmonization protocol, Atherosclerosis Risk in Communities Study, United States, 1989-2013.

TOPMed Variable Name (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
AGE (Age (years), time variant)	Years	Calculated	Birthdate: BIRTHDAT Exam dates: Ex1: ANTA09 Ex2: ANTB06 Ex3: ANTC4 Ex4: ANTD9 Ex5: ANT0A	<ul style="list-style-type: none"> • Age was calculated based on birth date and exam dates. • For one participant with missing date for visit 1, date of home interview from dataset 'HOM' was used as visit 1 date. • Usually, home interviews were conducted about 1 week before the clinical visit. • For missing values of age for exams 2, 3, 4 and 5 age was calculated by adding years 3, 6, 9 and 24 to baseline age.
AGE_BASELINE (Age at baseline, exam 1)	Years	Calculated		<ul style="list-style-type: none"> • Age at exam 1 calculated using birth date and exam dates.
BMI (Body mass index, time variant)	Kg/m ²	Calculated		<ul style="list-style-type: none"> • Calculated using BMI formula from height and weight
CLD_EVENT_PRI (Primary CLRD event: ICD code at first position)	-	Calculated		<ul style="list-style-type: none"> • Based on ICD codes
COUNT (Number of observation for each subject)	-	Calculated		<ul style="list-style-type: none"> • Count=1 can be used for cross sectional analysis
EDU_CAT (Highest education at baseline: 0=No schooling, 1=Grades 1-8, 2=Grades 9-11, 3=High school, 4=Some	-	Calculated	HOM54	<ul style="list-style-type: none"> • Education categories were created based on the highest education grade (hom54) completed.

Table 1. Harmonization protocol, Atherosclerosis Risk in Communities Study, United States, 1989-2013.

TOPMed Variable Name <i>(Label)</i>	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
<i>college, 5=Bachelor degree, 6=Graduate degree)</i>				
EXAM <i>(Exam/visit at which spirometry was performed)</i>	-	Calculated		<ul style="list-style-type: none"> Indicates exam when spirometry was performed
GENDER <i>(Gender: 1=Male, 0=Female)</i>	-	Original		<ul style="list-style-type: none"> Recoded: F (female) to 0, M (male) to 1
HT_CM <i>(Height (cms), time variant)</i>	Centimeters	Original	Ex1: ANTA01 Ex2: PFTB03 Ex3: ANTC1 Ex4: ANTD1 Ex5: ANT3	<ul style="list-style-type: none"> Height was not measured in exam 2. So, height from exam 1 was carried forward. For one participant in exam 4 there was data entry error for height (same values for height and weight) so after validating height from previous exams value of height from exam 3 was carried forward.
ID <i>(Original participant id for ARIC study)</i>	-	Original		
PMH_ASTHM_AGE_START_BASE <i>(Age when asthma started at baseline)</i>	Years	Original	RPAA37	<ul style="list-style-type: none"> Age of start asthma was coded to missing if participant answered 'no' to having asthma ever
PMH_ASTHMA_EVER_BASE <i>(Self-reported asthma ever at baseline: 1=Yes, 0=No)</i>	-	Original	RPAA35	<ul style="list-style-type: none"> Recoded: Y (yes) to 1, N (no) to 0 If PMH_ASTHMA_EVER_BASE = missing and PMH_ASTHMA_MD_BASE=1 then PMH_ASTHMA_EVER_BASE=1

Table 1. Harmonization protocol, Atherosclerosis Risk in Communities Study, United States, 1989-2013.

TOPMed Variable Name (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
				<ul style="list-style-type: none"> If PMH_ASTHMA_EVER_BASE = missing and PMH_ASTHMA_MD_BASE=0 then PMH_ASTHMA_EVER_BASE=0
PMH_ASTHMA_MD_BASE <i>(Doctor diagnosed asthma at baseline: 1=Yes, 0=No)</i>	-	Original	RPAA36	<ul style="list-style-type: none"> Recoded: Y (yes) to1, N (no) to 0
PMH_ASTHMA_STILL_BASE <i>(Still have asthma at baseline: 1=Yes, 0=No)</i>	-	Original	RPAA38	<ul style="list-style-type: none"> Recoded: Y (yes) to1, N (no) to 0 If PMH_ASTHMA_MD_BASE =0 then PMH_ASTHMA_STILL_BASE=0
PMH_CBRONCH_AGE_START_BASE <i>(Age when chronic bronchitis started at Baseline)</i>	Years	Original	RPAA30	<ul style="list-style-type: none"> Age of start chronic bronchitis was coded to missing if participant answered 'no' to having chronic bronchitis ever. Participant with age of start chronic bronchitis before age 10 were coded as missing.
PMH_CBRONCHITIS_EVER_BASE <i>(Self-reported chronic bronchitis ever at baseline: 1=Yes, 0=No)</i>	-	Original	RPAA27	<ul style="list-style-type: none"> Recoded: Y (yes) to1, N (no) to 0 If PMH_CBRONCHITIS_EVER_BASE = missing and PMH_CBRONCHITIS_MD_BASE=1 then PMH_CBRONCHITIS_EVER_BASE = 1 If PMH_CBRONCHITIS_EVER_BASE = missing and PMH_CBRONCHITIS_MD_BASE=0 then PMH_CBRONCHITIS_EVER_BASE = 0
PMH_CBRONCHITIS_MD_BASE <i>(Doctor diagnosed chronic bronchitis at baseline: 1=Yes, 0=No)</i>	-	Original	RPAA29	<ul style="list-style-type: none"> Recoded: Y (yes) to1, N (no) to 0

Table 1. Harmonization protocol, Atherosclerosis Risk in Communities Study, United States, 1989-2013.

TOPMed Variable Name <i>(Label)</i>	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
PMH_CBRONCHITIS_STILL_BASE <i>(Still have chronic bronchitis at baseline: 1=Yes, 0=No)</i>	-	Original	RPAA28	<ul style="list-style-type: none"> Recoded: Y (yes) to1, N (no) to 0 If PMH_CBRONCHITIS_MD_BASE =0 then PMH_CBRONCHITIS_STILL_BASE=0
PMH_DIABETES_MD_BASE <i>(Doctor diagnosed diabetes at baseline: 1=Yes, 0=No)</i>	-	Original	HOM10E	<ul style="list-style-type: none"> Recoded: Y (yes) to1, N (no) to 0
PMH_EMPH_AGE_START_BASE <i>(Age when emphysema started at baseline)</i>	Years	Original	RPAA34	<ul style="list-style-type: none"> Age of start emphysema was coded to missing if participant answered 'no' to having emphysema ever Participant with age of start emphysema before age 10 were coded as missing.
PMH_EMPHYSEMA_EVER_BASE <i>(Self-reported emphysema ever at baseline: 1=Yes, 0=No)</i>	-	Original	RPAA31	<ul style="list-style-type: none"> Recoded: Y (yes) to1, N (no) to 0 If PMH_EMPHYSEMA_EVER_BASE = missing and PMH_EMPHYSEMA_MD_BASE=1 then PMH_EMPHYSEMA_EVER_BASE = 1 If PMH_EMPHYSEMA_EVER_BASE= missing and PMH_EMPHYSEMA_MD_BASE=0 then PMH_EMPHYSEMA_EVER_BASE = 0
PMH_EMPHYSEMA_MD_BASE <i>(Doctor diagnosed emphysema at baseline: 1=Yes, 0=No)</i>	-	Original	RPAA33	<ul style="list-style-type: none"> Recoded: Y (yes) to1, N (no) to 0
PMH_EMPHYSEMA_STILL_BASE <i>(Still have emphysema at baseline: 1=Yes, 0=No)</i>	-	Original	RPAA32	<ul style="list-style-type: none"> Recoded: Y (yes) to1, N (no) to 0 If PMH_EMPHYSEMA_MD_BASE =0 then PMH_EMPHYSEMA_STILL_BASE=0

Table 1. Harmonization protocol, Atherosclerosis Risk in Communities Study, United States, 1989-2013.

TOPMed Variable Name <i>(Label)</i>	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
PMH_HTN_MD_BASE <i>(Doctor diagnosed hypertension at baseline: 1=Yes, 0=No)</i>	-	Original	HOM10A	<ul style="list-style-type: none"> Recoded: Y (yes) to 1, N (no) to 0
PRE_FEV1 <i>(Forced expiratory volume in one sec (L), time variant)</i>	Liters	Original	Ex1: PFTA26 Ex2: PFTB26 Ex5: PULP31	
PRE_FEV1FVC <i>(Ratio of fev1 over fvc (%), time variant)</i>	Percent	Original	Ex1: PFTA31 Ex2: PFTB31	Calculated for exam 5
PRE_FVC <i>(Forced vital capacity (L), time variant)</i>	Liters	Original	Ex1: PFTA24 Ex2: PFTB24 Ex5: PULP27	
RACE <i>(Race: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5= American Indian, 6=Other)</i>	-	Original		<ul style="list-style-type: none"> Recoded: W (White) to 1, B (Black) to 3, A (Asian) to 2, I (American Indian) to 5
RACE_BLACK <i>(Black or African American: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> if race=3 then race_black=1, else race_black=0
RACE_CHINESE <i>(Asian: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> if race=2 then race_chinese=1, else race_chinese=0

Table 1. Harmonization protocol, Atherosclerosis Risk in Communities Study, United States, 1989-2013.

TOPMed Variable Name <i>(Label)</i>	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
RACE_ETHNICITY <i>(Race/Ethnicity: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5= American Indian, 6=Other)</i>	-	Original		<ul style="list-style-type: none"> Recoded: W (White) to 1, B (Black) to 3, A (Asian) to 2, I (American Indian) to 5
RACE_HISPANIC <i>(Hispanic/Latino: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> if race=4 then race_hispanic=1, else race_hispanic=0
RACE_AMIND <i>(American Indian: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> if race=5 then race_amind=1, else race_amind =0
RACE_OTHERS <i>(Other Race: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> if race=6 then race_others=1, else race_others=0
RACE_WHITE <i>(White: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> if race=1 then race_white=1, else race_white=0
SMOKING_CIGS_PERDAY <i>(Cigarettes smoked per day)</i>	-	Original	Ex1: HOM35 Ex2: HHXB47, HHXB48 Ex3: PHXA29, PHXA30 Ex4: PHXB10, PHXB11	

Table 1. Harmonization protocol, Atherosclerosis Risk in Communities Study, United States, 1989-2013.

TOPMed Variable Name (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_CIGS_PERDAY_BASE (Cigarettes smoked per day at baseline visit)	-	Original		
SMOKING_CURRENT (Current smoker, time variant: 1=Yes, 0=No)	-	Calculated	Ex1: CURSMK01 Ex2: CURSMK21 Ex3: CURSMK31 Ex4: CURSMK41 Ex5: CURSMK52	<ul style="list-style-type: none"> • Missing observations were substituted with values based on some basic assumptions. • Any participants with missing values for smoking status at exams 1-4 but recorded as never smoker at exam 5 was considered as never smokers at previous exams too. • Participants who were recorded as former smoker at exam 1 were considered as former smokers at subsequent exams unless specified.
SMOKING_CURRENT_BASELINE (Current smoker at baseline: 1=Yes, 0=No)	-	Calculated	CURSMK01	
SMOKING_EVER (Ever smoker, time variant: 1=Yes, 0=No)	-	Calculated	Ex1: EVRSMK01 Ex2: EVRSMK21 Ex3: EVRSMK31 Ex4: EVRSMK41 Ex5: EVRSMK52	<ul style="list-style-type: none"> • Missing observations were substituted with values based on some basic assumptions. • Any participants with missing values for smoking status at exams 1-4 but recorded as never smoker at exam 5 was considered as never smokers at previous exams too. • Participants who were recorded to be former smoker at exam 1 were considered as ever smokers at subsequent exams unless specified.
SMOKING_EVER_BASELINE (Ever smoker at baseline: 1=Yes, 0=No)	-	Calculated	EVRSMK01	

Table 1. Harmonization protocol, Atherosclerosis Risk in Communities Study, United States, 1989-2013.

TOPMed Variable Name <i>(Label)</i>	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_FORMER <i>(Former smoker, time variant: 1=Yes, 0=No)</i>	-	Calculated	Ex1: EVRSMK01 Ex2: EVRSMK21 Ex3: EVRSMK31 Ex4: EVRSMK41 Ex5: EVRSMK52	<ul style="list-style-type: none"> Missing observations were substituted with values based on some basic assumptions. Any participants with missing values for smoking status at exams 1-4 but recorded as never smoker at exam 5 was considered as never smokers at previous exams too. Participants who were recorded to be former smoker at exam 1 were considered as former smokers at subsequent exams unless specified.
SMOKING_FORMER_BASELINE <i>(Former smoker at baseline: 1=Yes, 0=No)</i>	-	Calculated	FORSMK01	
SMOKING_PACKYEARS <i>(Cigarettes pack years, time variant)</i>	Years	Calculated	CIGTYR01	<ul style="list-style-type: none"> Smoking pack years at baseline were calculated using cigarettes smoking years (CIGTYR01) divided by 20. For exam 2, 3, 4 and 5 pack years were not available. So pack years were calculated using current smoking status, age of start of smoking, current age, number of months/years since quit smoking.
SMOKING_PACKYEARS_BASELINE <i>(Cigarettes pack years at baseline visit)</i>	Years	Calculated	CIGTYR01	
SMOKING_STATUS <i>(3 levels of smoking status, time variant: 0=Never, 1=Former, 2=Current)</i>	-	Calculated	Ex1: CIGT01 Ex2: CIGT21 Ex3: CIGT31 Ex4: CIGT41 Ex5: CIGT52	<ul style="list-style-type: none"> The original variable was modified after taking into consideration the modified current, ever and former status.

Table 1. Harmonization protocol, Atherosclerosis Risk in Communities Study, United States, 1989-2013.

TOPMed Variable Name <i>(Label)</i>	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_STATUS_BASELINE <i>(3 levels of smoking status at baseline: 1=Yes, 0=No)</i>	-	Calculated	CIGT01	
SPIRO_QC_FLAG <i>(Indicates that observation had FEV1 lower than subsequent at least by 1L)</i>	-	Calculated		<ul style="list-style-type: none"> This variable flags the observation with FEV1 < 1L than subsequent exam.
STUDY <i>(Name of cohort, categorical: aric, card, chs, habc, fhs, mesa, hchs/sol, shs)</i>	-			
TIMEFACTOR_SPIRO <i>(Time from baseline spirometry exam, years)</i>	Years	Calculated		<ul style="list-style-type: none"> Calculated using formula: age at spirometry - age at baseline spirometry
TTY_CLD_TOT_PRI <i>(Time to primary CLRD event)</i>	Years	Calculated		<ul style="list-style-type: none"> Based on date of first clinical visit and date when event occurred
WT_KG <i>(Weight (KG), time variant)</i>	Kilograms	Calculated	Ex1: ANTA04 Ex2: ANTB01 Ex3: ANTC2 Ex4: ANTD2 Ex5: ANT4	<ul style="list-style-type: none"> Weight in pounds was converted to kilograms.

Figure 2. Flow Chart of Longitudinal Lung Function Data in the Coronary Artery Risk Development in Young Adults Study, by Visit, United States, 1985-2011.

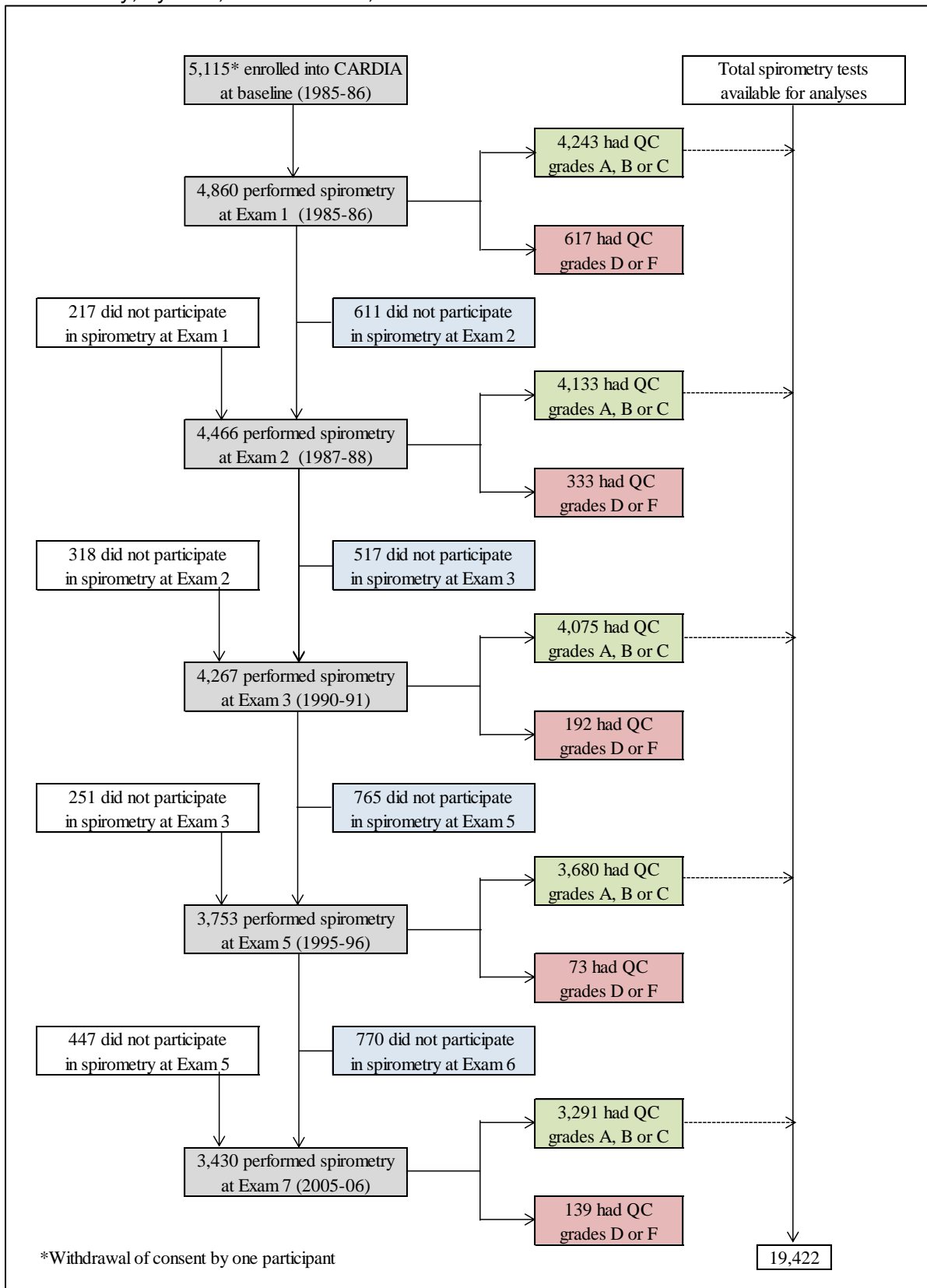


Table 2. Harmonization protocol, Coronary Artery Risk Development in Young Adults Study, United States, 1985-2011.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
AGE (Age (years), time variant)	Years	Original	Yr 0: EXAMAGE Yr 2: EX2_AGE Yr 5: EX3_AGE Yr 7: EX4_AGE Yr 10: EX5_AGE Yr 15: EX6_AGE Yr 20: EX7_AGE Yr 25: EX8_AGE	<ul style="list-style-type: none"> For missing values of age at years 2, 5, 7, 10, 15, 20 and 25 age was calculated by adding years 7, 10, 15, 20 and 25 to baseline age.
AGE_BASELINE (Age at exam 1, years)	Years	Original		
BMI (Body mass index, time variant)	Kg/m ²	Calculated		<ul style="list-style-type: none"> Calculated using BMI formula from height and weight
COUNT (Number of observations for each subject)	-	Calculated		<ul style="list-style-type: none"> Count=1 can be used for cross sectional analysis
EDU_CAT (Highest education at baseline: 0=No schooling, 1=Grades 1-8, 2=Grades 9-11, 3=High school, 4=Some college, 5=Bachelor degree, 6=Graduate degree)	-	Calculated	A03ED	<ul style="list-style-type: none"> New variable for education categories was created based on the highest education grade (A03ED) completed.
EXAM (Exam/visit at which spirometry is performed)	-			<ul style="list-style-type: none"> This variable represents the year at which spirometry exam was conducted- 0, 2, 5, 10 and 20

Table 2. Harmonization protocol, Coronary Artery Risk Development in Young Adults Study, United States, 1985-2011.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
GENDER (Gender: 1=Male, 0=Female)	-	Original	SEX	<ul style="list-style-type: none"> Recoded: 2 (female) to 0
HT_CM (Height (cms), time variant)	Centimeter	Original	Yr 0: A20HGT Yr 2: B20HGT Yr 5: C20HGT Yr 7: D20HGT Yr 10: E20HGT Yr 15: F20HGT Yr 20: G20HGT Yr 25: H20HGT	<ul style="list-style-type: none"> For two participants in year 20 there was (presumably) data entry error for height in cm because the measurements at year 20 did not match with previous measurements. These measurements were correct accordingly.
ID (Original participant long id for CARDIA study)		Original		
PMH_ASTHM_AGE_START_BASE (Age when asthma started at baseline)	Years	Original		<ul style="list-style-type: none"> Age of start of asthma was coded as missing for those with no history of asthma ever
PMH_ASTHMA_EVER_BASE (Self-reported asthma ever at baseline: 1=Yes, 0=No)	-	Original		<ul style="list-style-type: none"> Recoded: 1 (no) to 0, 2 (yes) to 1
PMH_ASTHMA_MD_BASE (Doctor diagnosed asthma at baseline: 1=Yes, 0=No)	-	Original		<ul style="list-style-type: none"> Recoded: 1 (no) to 0, 2 (yes) to 1

Table 2. Harmonization protocol, Coronary Artery Risk Development in Young Adults Study, United States, 1985-2011.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
PMH_ASTHMA_STILL_BASE <i>(Still have asthma at baseline: 1=Yes, 0=No)</i>	-	Original		<ul style="list-style-type: none"> Recoded: 1 (no) to 0, 2 (yes) to 1
PMH_CBRONCH_AGE_START_BASE <i>(Age when chronic bronchitis started at baseline)</i>	Years	Original		<ul style="list-style-type: none"> Age of start of chronic bronchitis was coded as missing for those with no history of chronic bronchitis ever
PMH_CBRONCHITIS_EVER_BASE <i>(Self-reported chronic bronchitis ever at baseline: 1=Yes, 0=No)</i>	-	Original		<ul style="list-style-type: none"> Recoded: 1 (no) to 0, 2 (yes) to 1
PMH_CBRONCHITIS_MD_BASE <i>(Doctor diagnosed chronic bronchitis at baseline: 1=Yes, 0=No)</i>	-	Original		<ul style="list-style-type: none"> Recoded: 1 (no) to 0, 2 (yes) to 1
PMH_CBRONCHITIS_STILL_BASE <i>(Still have chronic bronchitis at baseline: 1=Yes, 0=No)</i>	-	Original		<ul style="list-style-type: none"> Recoded: 1 (no) to 0, 2 (yes) to 1
PMH_DIABETES_MD_BASE <i>(Doctor diagnosed diabetes at baseline: 1=Yes, 0=No)</i>	-	Original		<ul style="list-style-type: none"> Recoded: 1 (no) to 0, 2 (yes) to 1
PMH_EMPH_AGE_START_BASE <i>(Age when emphysema started at baseline)</i>	Years	Original		<ul style="list-style-type: none"> Age of start of emphysema was coded as missing for those with no history of emphysema ever

Table 2. Harmonization protocol, Coronary Artery Risk Development in Young Adults Study, United States, 1985-2011.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
PMH_EMPHYSEMA_EVER_BASE <i>(Self-reported emphysema ever at baseline: 1=Yes, 0=No)</i>	-	Original		<ul style="list-style-type: none"> Recoded: 1 (no) to 0, 2 (yes) to 1
PMH_EMPHYSEMA_MD_BASE <i>(Doctor diagnosed emphysema at baseline: 1=Yes, 0=No)</i>	-	Original		<ul style="list-style-type: none"> Recoded: 1 (no) to 0, 2 (yes) to 1
PMH_EMPHYSEMA_STILL_BASE <i>(Still have emphysema at baseline: 1=Yes, 0=No)</i>	-	Original		<ul style="list-style-type: none"> Recoded: 1 (no) to 0, 2 (yes) to 1
PMH_HTN_MD_BASE <i>(Doctor diagnosed hypertension at baseline: 1=Yes, 0=No)</i>	-	Original		<ul style="list-style-type: none"> Recoded: 1 (no) to 0, 2 (yes) to 1
PRE_FEV1 <i>(Forced expiratory volume in one sec (L), time variant)</i>	Liters	Original	Yr 0: A12FE1 Yr 2: B12FE1 Yr 5: C12FE1 Yr 10: E12FE1 Yr 20: G12FE1	
PRE_FEV1FVC <i>(Ratio of fev1 over fvc (%), time variant)</i>	Percent	Calculated		<ul style="list-style-type: none"> We did not use the original variables because there was discrepancy between FEV₁/FVC ratios calculated from best FEV₁ and FVC and originally provided FEV₁/FVC ratio.
PRE_FVC <i>(Forced vital capacity (L), time variant)</i>	Liters	Original	Yr 0: A12FVC Yr 2: B12FVC Yr 5: C12FVC Yr 10: E12FVC Yr 20: G12FVC	

Table 2. Harmonization protocol, Coronary Artery Risk Development in Young Adults Study, United States, 1985-2011.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
RACE (Race: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5= American Indian, 6=Other)	-	Original		<ul style="list-style-type: none"> Recoded: 5 (Whites) to 1, 4 (Blacks) to 3
RACE_BLACK (Black or African American: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=3 then race_black=1, else race_black=0
RACE_CHINESE (Asian:1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=2 then race_chinese=1, else race_chinese=0
RACE_ETHNICITY (Race/ethnicity: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5= American Indian, 6=Other)	-	Original		<ul style="list-style-type: none"> Recoded: 5 (Whites) to 1, 4 (Blacks) to 3
RACE_HISPANIC (Hispanic/Latino:1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=4 then race_hispanic=1, else race_hispanic=0
RACE_AMIND (American Indian:1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=5 then race_amind=1, else race_amind =0
RACE_OTHERS (Other race:1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=6 then race_others=1, else race_others=0

Table 2. Harmonization protocol, Coronary Artery Risk Development in Young Adults Study, United States, 1985-2011.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
RACE_WHITE (White: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • if race=1 then race_white=1, • else race_white=0
SHORT_ID (Original participant short id for CARDIA study)	-	Original		
SMOKING_CIGS_PERDAY (Cigarettes smokers per day, time variant)	-	Original	Yr 0: A09CGTDY Yr 2: B09CGTDY Yr 5: C09CGTDY Yr 7: D09CGTDY Yr 10: E09CGTDY Yr 15: F09CGTDY Yr 20: G09CGTDY Yr 25: H09CGTDY	
SMOKING_CIGS_PERDAY_BASE (Cigarettes smokers per day at baseline)	-	Original		
SMOKING_CURRENT (Current smoker, time variant: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • Calculated from smoking status: if smoking status=2 then smoking_current=1; else smoking_current=0
SMOKING_CURRENT_BASELINE (Current smoker at baseline: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • Calculated from smoking status: if smoking status=2 then smoking_current=1; else smoking_current=0

Table 2. Harmonization protocol, Coronary Artery Risk Development in Young Adults Study, United States, 1985-2011.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_EVER <i>(Ever smoker, time variant: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=0 then smoking_ever=0; else smoking_ever=1
SMOKING_EVER_BASELINE <i>(Ever smoker at baseline: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=0 then smoking_ever=0; else smoking_ever=1
SMOKING_FORMER <i>(Former smoker, time variant: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=1 then smoking_former=1; else smoking_former=0
SMOKING_FORMER_BASELINE <i>(Former smoker at baseline: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=1 then smoking_former=1; else smoking_former=0
SMOKING_PACKYEARS <i>(Cigarettes pack years, time variant)</i>	Years	Calculated	Yr 0: A09CGTDY, A09SMKYR Yr 2: B09CGTDY, B09SMKYR Yr 5: C09CGTDY, C09SMKYR Yr7: D09CGTDY, D09SMKYR Yr 10: E09CGTDY, E09SMKYR Yr 15: F09CGTDY, F09SMKYR Yr 20: G09CGTDY, G09SMKYR Yr 25: H09CGTDY, H09SMKYR	<ul style="list-style-type: none"> Calculated using following formula: $\text{packyears} = (\# \text{ of cigarettes smoked per day}) \times (\# \text{ of years cigarretes smoked}) / 20$

Table 2. Harmonization protocol, Coronary Artery Risk Development in Young Adults Study, United States, 1985-2011.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
			H09SMKYR	
SMOKING_PACKYEARS_BASELINE (Pack years at baseline visit)	Years	Calculated		<ul style="list-style-type: none"> Calculated using following formula: $\text{packyears} = \frac{(\# \text{ of cigarettes smoked per day}) \times (\# \text{ of years cigarettes smoked})}{20}$
SMOKING_STATUS (3 levels of smoking status, time variant: 0=Never, 1=Former, 2=Current)	-	Calculated	Yr 0: A10TOBAC, A10CIGS, A09SMKNW Yr 2: B10TOBAC, B10CIGS, B09SMKNW Yr 5: C10TOBAC, C10CIGS, C09SMKNW Yr 7: D10TOBAC, D10CIGS, D09SMKNW Yr 10: E10TOBAC, E10CIGS, E09SMKNW Yr 15: F10TOBAC, F10CIGS, F09SMKNW Yr 20: G10TOBAC, G10CIGS, G09SMKNW Yr 25: H10TOBAC, H10CIGS, H09SMKNW	<ul style="list-style-type: none"> Calculated using following formula for each year: <ol style="list-style-type: none"> if A10TOBAC (have you ever used tobacco product) = 1 (no) or/and A10CIGS (have you ever smoked cigarettes regularly)= 1 (no) then smoking status=0 (never smoker) if A10TOBAC (have you ever used tobacco product) = 2 (yes) and A10CIGS (have you ever smoked cigarettes regularly) = 2 (yes) and A09SMKNW (do you still smoke cigarettes regularly)= 1 (no) then smoking status=1 (former smoker) if A10TOBAC (have you ever used tobacco product) = 2 (yes) and A10CIGS (have you ever smoked cigarettes regularly) = 2 (yes) and A09SMKNW (do you still smoke cigarettes regularly)= 2 (yes) then smoking status=2 (current smoker)

Table 2. Harmonization protocol, Coronary Artery Risk Development in Young Adults Study, United States, 1985-2011.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
<p>SMOKING_STATUS_BASELINE (3 levels of smoking status at baseline: 1=Yes, 0=No)</p>	-	Calculated	Yr 0: A10TOBAC, A10CIGS, A09SMKNW	<ul style="list-style-type: none"> Calculated using following formula: <ol style="list-style-type: none"> if A10TOBAC (have you ever used tobacco product) = 1 (no) or/and A10CIGS (have you ever smoked cigarettes regularly)= 1 (no) then smoking status=0 (never smoker) if A10TOBAC (have you ever used tobacco product) = 2 (yes) and A10CIGS (have you ever smoked cigarettes regularly) = 2 (yes) and A09SMKNW (do you still smoke cigarettes regularly)= 1 (no) then smoking status=1 (former smoker) if A10TOBAC (have you ever used tobacco product) = 2 (yes) and A10CIGS (have you ever smoked cigarettes regularly) = 2 (yes) and A09SMKNW (do you still smoke cigarettes regularly)= 2 (yes) then
<p>SPIRO_QC_FLAG (Difference between last and first spirometry > 1 L)</p>	-	Calculated		<ul style="list-style-type: none"> This variable flags the observation with FEV1 < 1L than subsequent exam.
<p>STUDY (Name of cohort, categorical: aric, card, chs, habc, fhs, mesa, hchs/sol, shs)</p>				
<p>TIMEFACTOR_SPIRO (Time from baseline spirometry exam, years)</p>	Years	Calculated		<ul style="list-style-type: none"> Calculated using formula: age at spirometry - age at baseline spirometry

Table 2. Harmonization protocol, Coronary Artery Risk Development in Young Adults Study, United States, 1985-2011.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
WT_KG <i>(Weight (KG), time variant)</i>	Kilogram	Calculated	Yr 0: A20WGT Yr 2: B20WGT Yr 5: C20WGT Yr 7: D20WGT Yr 10: E20WGT Yr 15: F20WGT Yr 20: G20WGT Yr 25: H20WGT	<ul style="list-style-type: none"> • For one participant there was (presumably) data entry error for weight at year 10 because the measurement at year 10 did not match with consequent measurements. • The measurement was correct accordingly.

Figure 3. Flow Chart of Longitudinal Lung Function Data in the Cardiovascular Health Study, by visit, United States, 1989-2006.

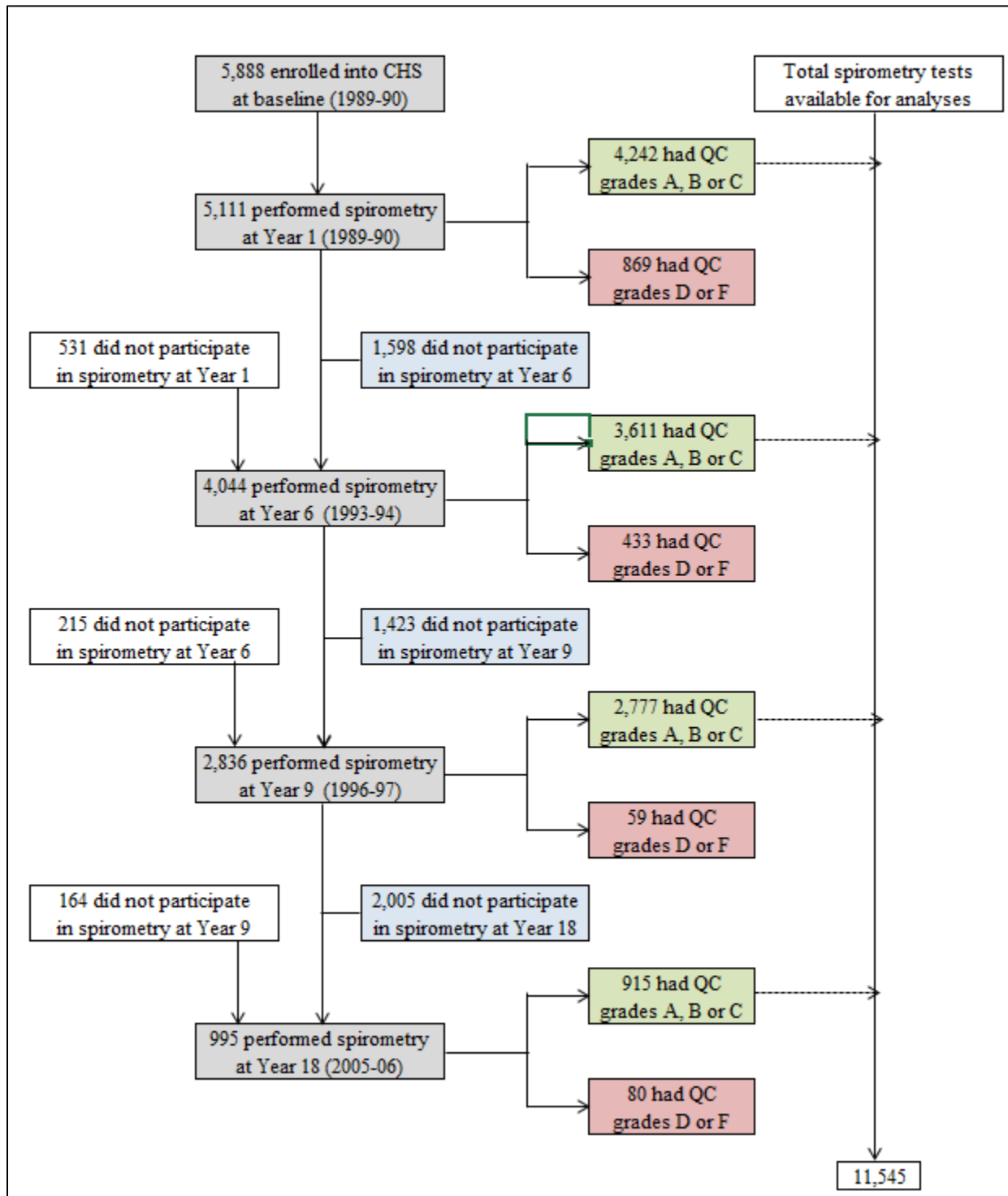


Figure 3. Harmonization Protocol, Cardiovascular Health Study, United States, 1989-2006.

Variable (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
AGE (Age (years), time variant)	Years	Original	Yr 2: AGEYR2 Yr 5: AGEY5 Yr 6: AGEY6 Yr 9: AGEY9 Yr 18: AGE18	<ul style="list-style-type: none"> For missing values of age at years 5, 6, 9 and 18 age was calculated by adding years 3, 4, 7 and 16 to baseline age.
AGE_BASELINE (Age at first spirometry exam)	Years	Original	AGEYR2, AGEY5 (cohort 2)	<ul style="list-style-type: none"> For cohort 2 (N=687): age at year 5 was considered as baseline age
BMI (Body mass index, time variant)	Kg/m ²	Calculated		<ul style="list-style-type: none"> Calculated using BMI formula from height and weight
CLD_EVENT_PRI (Primary CLRD event: ICD code at first position)	-	Calculated		<ul style="list-style-type: none"> Based on ICD codes
COUNT (Number of observation for each subject)	-			
EDU_CAT (Highest education at baseline: 0=No schooling, 1=Grades 1-8, 2=Grades 9-11, 3=High school, 4=Some college, 5=Bachelor degree, 6=Graduate degree)	-	Calculated	GRADE01	<ul style="list-style-type: none"> New variable for education categories was created based on the highest education grade (A03ED) completed.
EXAM (Exam/visit at which spirometry is performed)	-	Original		

Figure 3. Harmonization Protocol, Cardiovascular Health Study, United States, 1989-2006.

Variable (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
GENDER (Gender: 1=Male, 0=Female)	-	Original	GEND01	
HT_CM (Height (cms), time variant)	Centimeter	Original	Yr 2: STHT13 Yr 5: STHT5 Yr 9: STHT9 Yr 18: HEIGHT18	<ul style="list-style-type: none"> Height for exam 6 was not measured so height at exam 5 was carried forward.
IDNO (Original participant id for CHS study)	-	Original		
PMH_ASTHM_AGE_START_BASE (Age when asthma started at baseline)	Years	Original	ATHT07, ASTHAG57	<ul style="list-style-type: none"> If pmh_asthma_ever_base=0 then pmh_asthm_age_start_base=.
PMH_ASTHMA_EVER_BASE (Self-reported asthma ever at baseline: 1=Yes, 0=No)	-	Original	ATH07, ASTH57	
PMH_ASTHMA_MD_BASE (Doctor diagnosed asthma at baseline: 1=Yes, 0=No)	-	Original	ATHDR07, ASTHDR57	<ul style="list-style-type: none"> If pmh_asthma_ever_base=0 then pmh_asthma_md_base=0
PMH_ASTHMA_STILL_BASE (Still have asthma at baseline: 1=Yes, 0=No)	-	Original	ATHH07, ASTHH57	<ul style="list-style-type: none"> If pmh_asthma_ever_base=0 then pmh_asthma_still_base=0

Figure 3. Harmonization Protocol, Cardiovascular Health Study, United States, 1989-2006.

Variable (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
PMH_CBRONCH_AGE_START_BASE (Age when chronic bronchitis started at baseline)	Years	Original	BRNCT07, CBRNAG57	<ul style="list-style-type: none"> If pmh_cbronchitis_ever_base=0 then pmh_cbronch_age_start_base=.
PMH_CBRONCHITIS_EVER_BASE (Self-reported chronic bronchitis ever at baseline: 1=Yes, 0=No)	-	Original	CRNC07, CBRN57	
PMH_CBRONCHITIS_MD_BASE (Doctor diagnosed chronic bronchitis at baseline: 1=Yes, 0=No)	-	Original	BRNCDR07, CBRNDR57	<ul style="list-style-type: none"> If pmh_cbronchitis_ever_base=0 then pmh_cbronchitis_md_base=0
PMH_CBRONCHITIS_STILL_BASE (Still have chronic bronchitis at baseline: 1=Yes, 0=No)	-	Original	BRNCHV07, CBRNHV57	<ul style="list-style-type: none"> If pmh_cbronchitis_ever_base=0 then pmh_cbronchitis_still_base=0
PMH_DIABETES_MD_BASE (Doctor diagnosed diabetes at baseline: 1=Yes, 0=No)	-	Original	DIAB01	
PMH_EMPH_AGE_START_BASE (Age when emphysema started at baseline)	Years	Original	EPHT07, EMPHAG57	<ul style="list-style-type: none"> If pmh_emphysema_ever_base=0 then pmh_emph_age_start_base=.
PMH_EMPHYSEMA_EVER_BASE (Self-reported emphysema ever at baseline: 1=Yes, 0=No)	-	Original	EPH07, EMPH57	

Figure 3. Harmonization Protocol, Cardiovascular Health Study, United States, 1989-2006.

Variable (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
PMH_EMPHYSEMA_MD_BASE (Doctor diagnosed emphysema at baseline: 1=Yes, 0=No)	-	Original	EPHDR07, EMPHDR57	<ul style="list-style-type: none"> If pmh_emphysema_ever_base=0 then pmh_emphysema_md_base=0
PMH_EMPHYSEMA_STILL_BASE (Still have emphysema at baseline: 1=Yes, 0=No)	-	Original	EPHH07, EMPHH57	<ul style="list-style-type: none"> If pmh_emphysema_ever_base=0 then pmh_emphysema_still_base=0
PMH_HTN_MD_BASE (Doctor diagnosed hypertension at baseline: 1=Yes, 0=No)	-	Original	HIBP01	
PRE_FEV1 (Forced expiratory volume in one sec (L), time variant)	Liters	Original	Yr 2: FEV2 Yr 6: FEV6 Yr 9: FEV9 Yr 18: FEV1_18	
PRE_FEV1FVC (Ratio of pre_fev1 over pre_fvc (%), time variant)	Percent	Calculated		<ul style="list-style-type: none"> Calculated using formula: pre_fev1fvc=pre_fev1/pre_fvc*100
PRE_FVC (Forced vital capacity (L), time variant)	Liters	Original	Yr 2: FVC2 Yr 6: FVC6 Yr 9: FVC9 Yr 18: FVC_18	
RACE (Race: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5= American Indian, 6=Other)	-	Original	RACE01	<ul style="list-style-type: none"> Recoded: 1 (Blacks) to 3, 3 (American Indians) to 5, 4 (Asians) to 2, 5 (Others) to 0

Figure 3. Harmonization Protocol, Cardiovascular Health Study, United States, 1989-2006.

Variable (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
RACE_BLACK (Black or African American: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=3 then race_black=1, else race_black=0
RACE_CHINESE (Asian: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=2 then race_chinese=1, else race_chinese=0
RACE_ETHNICITY (Race/ethnicity: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5= American Indian, 6=Other)	-	Original	RACE01	<ul style="list-style-type: none"> Recoded: 1 (Blacks) to 3, 3 (American Indians) to 5, 4 (Asians) to 2, 5 (Others) to 0
RACE_HISPANIC (Hispanic/Latino: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=4 then race_hispanic=1, else race_hispanic=0
RACE_AMIND (American Indian: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=5 then race_amind=1, else race_amind=0
RACE_OTHERS (Other race: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=6 then race_others=1, else race_others=0
RACE_WHITE (White: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=1 then race_white=1, else race_white=0

Figure 3. Harmonization Protocol, Cardiovascular Health Study, United States, 1989-2006.

Variable (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_CIGS_PERDAY (Cigarettes smoked per day, time variant)	-	Original	Yr 2: NUMCIG02 Yr 5: NUMCIG05 Yr 6: NUMCIG06 Yr 9: NUMCIG09 Yr 18: NUMCIG18	
SMOKING_CIGS_PERDAY_BASE (Cigarettes smoked per day at baseline visit)	-	Original	Yr 2: NUMCIG02 Yr 5: NUMCIG05 (Cohort 2)	
SMOKING_CURRENT (Current smoker, time variant: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=2 then smoking_current=1; else smoking_current=0
SMOKING_CURRENT_BASELINE (Current smoker at baseline: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=2 then smoking_current=1; else smoking_current=0
SMOKING_EVER (Ever smoker, time variant: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=0 then smoking_ever=0; else smoking_ever=1
SMOKING_EVER_BASELINE (Ever smoker at baseline: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=0 then smoking_ever=0; else smoking_ever=1
SMOKING_FORMER (Former smoker, time variant: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=1 then smoking_former=1; else smoking_former=0

Figure 3. Harmonization Protocol, Cardiovascular Health Study, United States, 1989-2006.

Variable (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_FORMER_BASELINE (Former smoker at baseline:1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=1 then smoking_former=1; else smoking_former=0
SMOKING_PACKYEARS (Cigarettes pack years, time variant)	Years	Calculated		<ul style="list-style-type: none"> Cigarettes packyears were not provided for years 5, 6, 9 and 18. Packyears were calculated by using numbers of cigarettes smoked per day and total number of smoking years. Packyears=(numbers of cigarettes smoked per day) * (total number of smoking years)/20 For current smokers who have smoked continuously from baseline to last visit, total number of smoking years were calculated using following formulas: (age at current visit) - (age of start of smoking). For former smokers who were current smokers at previous visit, total number of smoking years were calculated using following formulas {(age at current visit - [(age at current visit) - (age at previous visit)/2]} - {age of start of smoking} For former smokers who had been former smoker at previous visit, total number of smoking years were carried forward from previous visit.
SMOKING_PACKYEARS_BASELINE (Cigarettes pack years at baseline visit)	Years	Original	PKYRSBL	
SMOKING_STATUS (3 levels of smoking status, time variant:0=Never, 1=Former, 2=Current)	-	Calculated	Yr 2: SMOKE2 Yr 5: SMOKE5 Yr 6: SMOKE06 Yr 9: SMOKE9 Yr 18: SMOKING18	<ul style="list-style-type: none"> Recoded: 1. <u>smoke2, smoke5, smoke9</u>: 1 (never) to 0, 2 (former) to 1, 3 (current) to 2 2. <u>smoke06, smoking18</u>: 1 (never) to 0, 2 (former, quit > 1 year ago) to 1, 3 (former, quit < 1 year ago) to 1, 4 (current) to 2, 9 (don't know) to missing

Figure 3. Harmonization Protocol, Cardiovascular Health Study, United States, 1989-2006.

Variable (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
				<ul style="list-style-type: none"> Missing observations were substituted with values based on some basic assumptions. Any participant with missing values for smoking status at any exam but recorded as never smoker in subsequent was considered as never smokers at previous exams too. Participants who were recorded as former smoker at exam 2 were considered as former smokers at subsequent exams unless specified.
<p>SMOKING_STATUS_BASELINE (3 levels of smoking status at baseline: 1=Yes, 0=No)</p>	-	Calculated	<p>Yr 2: SMOKE2 Yr 5: SMOKE5 (Cohort 2)</p>	<ul style="list-style-type: none"> Recoded: <ol style="list-style-type: none"> <u>smoke2, smoke5, smoke9</u>: 1 (never) to 0, 2 (former) to 1, 3 (current) to 2 <u>smoke06, smoking18</u>: 1 (never) to 0, 2 (former, quit > 1 year ago) to 1, 3 (former, quit < 1 year ago) to 1, 4 (current) to 2, 9 (don't know) to missing Missing observations were substituted with values based on some basic assumptions. Any participant with missing values for smoking status at any exam but recorded as never smoker in subsequent was considered as never smokers at previous exams too. Participants who were recorded as former smoker at exam 2 were considered as former smokers at subsequent exams unless specified.
<p>SPIRO_QC_FLAG (Indicates that observation had FEV1 lower than subsequent at least by 1L)</p>	-	Calculated		<ul style="list-style-type: none"> This variable flags the observation with FEV1 < 1L than subsequent exam.
<p>STUDY (Name of cohort, categorical: aric, card, chs, habc, fhs, mesa, hchs/sol, shs)</p>	-	Calculated		

Figure 3. Harmonization Protocol, Cardiovascular Health Study, United States, 1989-2006.

Variable (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
TIMEFACTOR_SPIRO (Time from baseline spirometry exam, years)	Years	Calculated		<ul style="list-style-type: none"> Calculated using formula: (age at spirometry) – (age at baseline spirometry)
TTY_CLD_TOT_PRI (Time to primary CLRD event)	-	Calculated		<ul style="list-style-type: none"> Based on date at which event occurred and censored
WT_KG (Weight (KG), time variant)	Kilogram	Calculated	Yr 2: WEIGHT2 Yr 5: WEIGHT5 Yr 6: WEIGHT6 Yr 9: WEIGHT9 Yr 18: WEIGHT18	<ul style="list-style-type: none"> Originally weight was recorded in pounds which were converted to kgs.

Figure 4. Flow Chart of Longitudinal Lung Function Data in the Framingham Heart Study-Offspring cohort, by Visit, United States, 1983-2014.

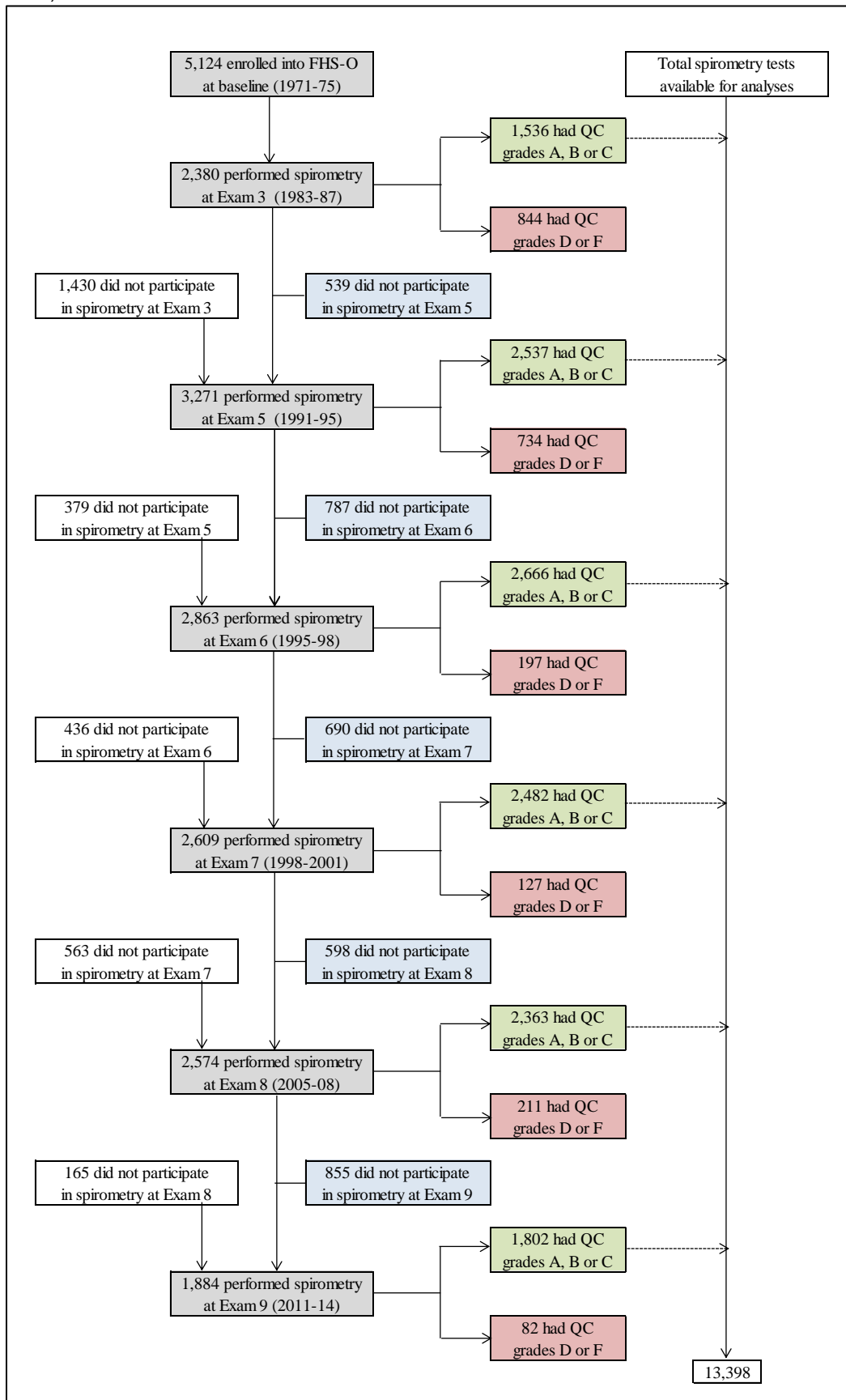


Table 4. Harmonization Protocol, Framingham Heart Study-Offspring Cohort, United States, 1983-2014.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
AGE <i>(Age (years), time variant)</i>	Years	Calculated	Ex 1: A42, A43, A44, A45, A46, A47 Ex 5: E217 Ex 6: F179 Ex 7: G007 Ex 8: H009 Ex 9: J009	<ul style="list-style-type: none"> • Age was calculated based on birth date and exam dates. • However, we were not able to get exam date for exam 3 or even 'age' variable in any of the datasets. • Therefore, age at exam 3 was calculated using age at exam 1 + 12 (as exam 3 was conducted 12 years after exam 1).
AGE_BASELINE <i>(Age at baseline, exam 3)</i>	Years	Calculated		<ul style="list-style-type: none"> • We were not able to get exam date for exam 3 or even 'age' variable in any of the datasets. • Therefore, age at exam 3 was calculated using age at exam 1 + 12 (as exam 3 was conducted 12 years after exam 1).
BMI <i>(Body mass index, time variant)</i>	Kg/m ²	Calculated		<ul style="list-style-type: none"> • Calculated using BMI formula from height and weight
COUNT <i>(Number of observation for each subject)</i>	-	Calculated		<ul style="list-style-type: none"> • Count=1 can be used for cross sectional analysis
EDU_CAT <i>(Highest education at baseline:0=No schooling, 1=Grades 1-8, 2=Grades 9-11, 3=High school, 4=Some college, 5=Bachelor degree, 6=Graduate degree)</i>	-	Calculated	B43	<ul style="list-style-type: none"> • New variable for education categories was created based on the highest education grade (B43) completed.

Table 4. Harmonization Protocol, Framingham Heart Study-Offspring Cohort, United States, 1983-2014.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
EXAM (Exam/visit at which spirometry is performed)	-	Calculated		
GENDER (Gender: 1=Male, 0=Female)	-	Original	C2	<ul style="list-style-type: none"> Recoded: 2 (Female) to 0
HT_CM (Height (cms), time variant)	Centimeter	Calculated	Ex 3: C417 Ex 5: E025 Ex 6: F008 Ex 7: G441 Ex 8: H399 Ex 9: J472	<ul style="list-style-type: none"> Originally height was recorded in inches which were converted to cms.
PMH_ASTHMA_MD_BASE (Doctor diagnosed asthma at baseline: 1=Yes, 0=No)	-		Ex 3: C374	<ul style="list-style-type: none"> Recoded: 0 (no) and 2 (maybe) as 0
PMH_CBRONCHITIS_MD_BASE (Doctor diagnosed chronic bronchitis at baseline: 1=Yes, 0=No)	-		Ex 3: C371	<ul style="list-style-type: none"> Recoded: 0 (no) and 2 (maybe) as 0
PMH_DIABETES_MD_BASE (Doctor diagnosed diabetes at baseline: 1=Yes, 0=No)	-		Ex 3: C363	<ul style="list-style-type: none"> Recoded: 0 (no) and 2 (maybe) as 0

Table 4. Harmonization Protocol, Framingham Heart Study-Offspring Cohort, United States, 1983-2014.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
PMH_EMPHYSEMA_MD_BASE <i>(Doctor diagnosed emphysema at baseline: 1=Yes, 0=No)</i>	-		Ex 3: C370	<ul style="list-style-type: none"> Recoded: 0 (no) and 2 (maybe) as 0
PMH_HTN_MD_BASE <i>(Doctor diagnosed hypertension at baseline: 1=Yes, 0=No)</i>	-		Ex 3: C332	<ul style="list-style-type: none"> Recoded: 0 (no) and 2 (borderline) as 0
PRE_FEV1 <i>(Forced expiratory volume in one sec (L), time variant)</i>	Liters	Original	Ex 3: FV1_3_1 Ex 5: bestfev1 Ex 6: bestfev1 Ex 7: bestfev1 Ex 8: FV1_8_1 Ex 9: FV1_9_1	
PRE_FEV1FVC <i>(Ratio of fev1 over fvc (%), time variant)</i>	Percent	Calculated		<ul style="list-style-type: none"> Calculated using best FEV1 and FVC.
PRE_FVC <i>(Forced vital capacity (L), time variant)</i>	Liters	Original	Ex 3: FVC_3_1 Ex 5: bestfvc Ex 6: bestfvc Ex 7: bestfvc Ex 8: FVC_8_1 Ex 9: FVC_9_1	

Table 4. Harmonization Protocol, Framingham Heart Study-Offspring Cohort, United States, 1983-2014.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
RACE (Race: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5=American Indian, 6=Other)	-	Calculated		<ul style="list-style-type: none"> • Framingham heart study was conducted only in Caucasians. • Therefore, race was not asked in exam 1. So all participants in this study are coded as white.
RACE_BLACK (Black or African American: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • Since race=1 (white), race_black=0
RACE_CHINESE (Asian: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • Since race=1 (white), race_chinese=0
RACE_ETHNICITY (Race/ethnicity: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5= American Indian, 6=Other)	-	Calculated		<ul style="list-style-type: none"> • Framingham heart study was conducted only in Caucasians. • Therefore, race was not asked in exam 1. So all participants in this study are coded as white.
RACE_HISPANIC (Hispanic/Latino: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • Since race=1 (white), race_hispanic=0
RACE_AMIND (American Indian: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • Since race=1 (white), race_amind=0
RACE_OTHERS (Other race: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • Since race=1 (white), race_others=0

Table 4. Harmonization Protocol, Framingham Heart Study-Offspring Cohort, United States, 1983-2014.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
RACE_WHITE (White: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • Since race=1 (white), race_white=1
RANID (Original participant id for Framingham study)	-	Original	ranid	
SMOKING_CIGS_PERDAY (Cigarettes smokers per day, time variant)	-	Original	Ex 3: C68 Ex 5: E320 Ex 6: F289 Ex 7: G117 Ex 8: H064 Ex 9: J066	
SMOKING_CIGS_PERDAY_BASE (Cigarettes smokers per day at baseline visit)	-	Original	Ex 3: C68	
SMOKING_CURRENT (Current smoker, time variant: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • Calculated from smoking status: if smoking status=2 then smoking_current=1; else smoking_current=0
SMOKING_CURRENT_BASELINE (Current smoker at baseline: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • Calculated from smoking status: if smoking status=2 then smoking_current=1; else smoking_current=0

Table 4. Harmonization Protocol, Framingham Heart Study-Offspring Cohort, United States, 1983-2014.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_EVER (Ever smoker, time variant: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=0 then smoking_ever=0; else smoking_ever=1
SMOKING_EVER_BASELINE (Ever smoker at baseline: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=0 then smoking_ever=0; else smoking_ever=1
SMOKING_FORMER (Former smoker, time variant: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=1 then smoking_former=1; else smoking_former=0
SMOKING_FORMER_BASELINE (Former smoker at baseline: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=1 then smoking_former=1; else smoking_former=0
SMOKING_PACKYEARS (Cigarettes pack years, time variant)	Years	Calculated		<ul style="list-style-type: none"> Pack years for exams 1, 2, 8 and 9 were calculated using # of cigarettes per day, age when started smoking and age when stopped smoking. Age when started smoking and age when stopped smoking were not asked for exams 3, 4, 5, 6 and 7. For these years, pack years calculated by using numbers of cigarettes smoked per day and total number of smoking years. Packyears=(numbers of cigarettes smoked per day) * (total number of smoking years)/20 For current smokers who have smoked continuously from baseline to last visit, total number of smoking years were calculated using following formulas: (age at current visit) - (age of start of smoking).

Table 4. Harmonization Protocol, Framingham Heart Study-Offspring Cohort, United States, 1983-2014.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
				<ul style="list-style-type: none"> For former smokers who were current smokers at previous visit, total number of smoking years were calculated using following formulas {age at current visit - [(age at current visit) - (age at previous visit)/2]} - {age of start of smoking} For former smokers who had been former smoker at previous visit, total number of smoking years were carried forward from previous visit.
SMOKING_PACKYEARS_BASELINE <i>(Cigarettes pack years at baseline visit)</i>	Years	Calculated		<ul style="list-style-type: none"> Pack years for exam 3 was calculated by using numbers of cigarettes smoked per day and total number of smoking years. Packyears=(numbers of cigarettes smoked per day) * (total number of smoking years)/20 For current smokers who have smoked continuously from baseline to last visit, total number of smoking years were calculated using following formulas: (age at current visit) - (age of start of smoking). For former smokers who were current smokers at previous visit, total number of smoking years were calculated using following formulas {age at current visit - [(age at current visit) - (age at previous visit)/2]} - {age of start of smoking} For former smokers who had been former smoker at previous visit, total number of smoking years were carried forward from previous visit.
SMOKING_STATUS <i>(3 levels of smoking status, time variant:0=Never, 1=Former, 2=Current)</i>	-	Calculated	Ex 3: C67 Ex 5: E319 Ex 6: F288 Ex 7: G116 Ex 8: H060, H061, H062	<ul style="list-style-type: none"> Smoking status was calculated using based on whether the participant smoked regularly in past year and smoking status at previous year.

Table 4. Harmonization Protocol, Framingham Heart Study-Offspring Cohort, United States, 1983-2014.

Variable (Label)	Units	Calculated/ Original	Original Variable Name	Details regarding classification, calculation, and/or correction
			Ex 9: J062, J063, J064	
SMOKING_STATUS_BASELINE (3 levels of smoking status at baseline: 1=Yes, 0=No)	-	Calculated	Ex 3: C67	<ul style="list-style-type: none"> Smoking status was calculated using based on whether the participant smoked regularly in past year and smoking status at previous year.
SPIRO_QC_FLAG (Indicates that observation had FEV1 lower than subsequent at least by 1L)	-	Calculated		<ul style="list-style-type: none"> This variable flags the observation with FEV1 < 1L than subsequent exam.
STUDY (Name of cohort, categorical: aric, card, chs, habc, fhs, mesa, hchs/sol, shs)	-	Calculated		
TIMEFACTOR_SPIRO (Time from baseline spirometry exam, years)	Years	Calculated		<ul style="list-style-type: none"> Calculated using formula: age at spirometry - age at baseline spirometry
WT_KG (Weight (KG), time variant)	Kilogram	Calculated	Ex 3: C416 Ex 5: E024 Ex 6: F007 Ex 7: G440 Ex 8: H393 Ex 9: J474	<ul style="list-style-type: none"> Originally weight was recorded in pounds which were converted to kgs.

Figure 5. Flow Chart of Longitudinal Lung Function Data in the Health, Aging and Body Composition Study, by Visit, United States, 1997-2006.

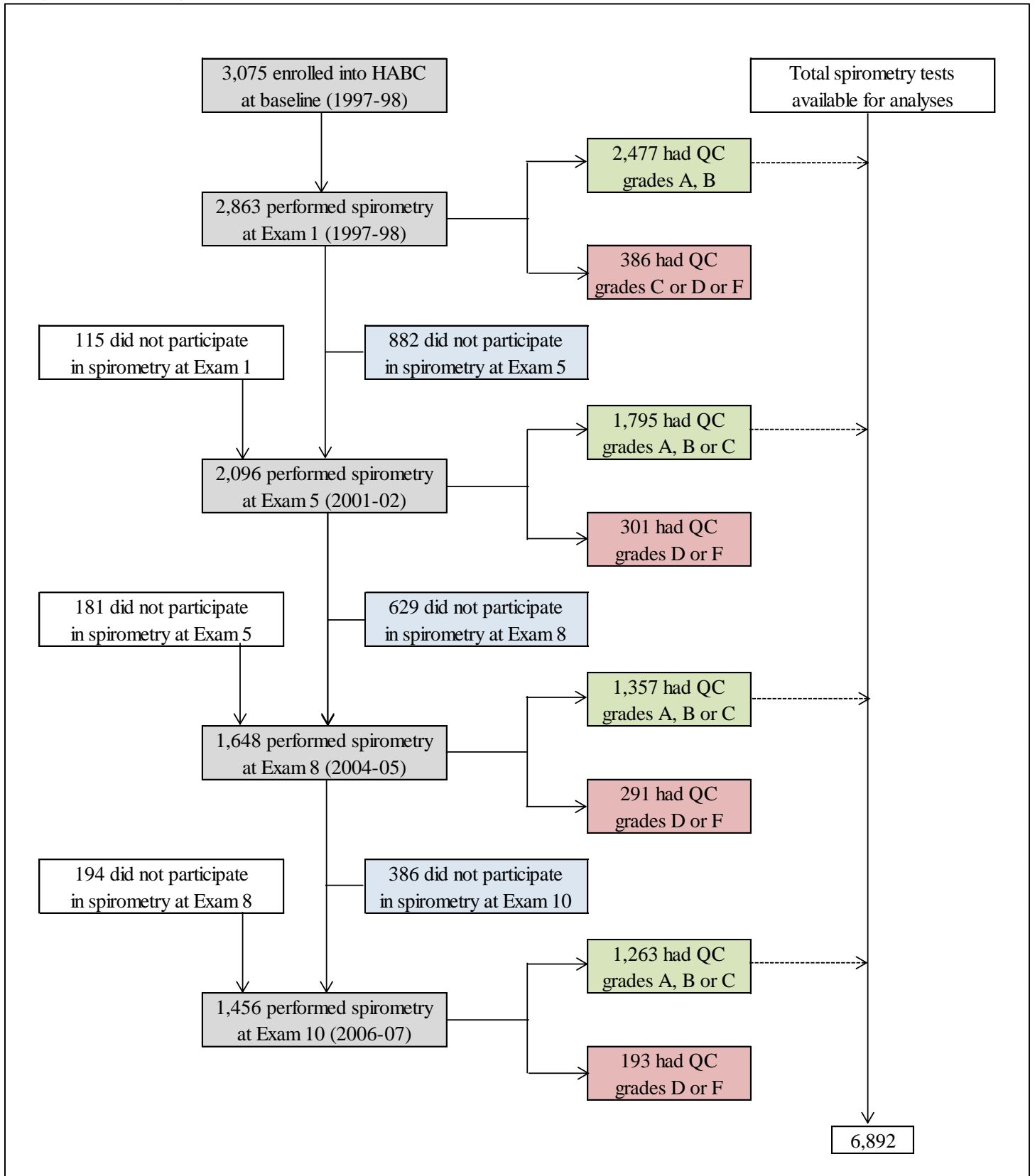


Table 5. Harmonization Protocol, Health, Aging and Body Composition Study, by Visit, United States, 1997-2006.

TOPMed Variable Name (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
AGE (Age (years), time variant)	Years	Calculated	Birthdate: DOB Exam dates: Ex1: CV1DATE, Ex2: CV2DATE, ..., Ex14: CV14DATE	<ul style="list-style-type: none"> Age was calculated based on birth date and exam dates.
AGE_BASELINE (Age at baseline, exam 1)	Years	Calculated		<ul style="list-style-type: none"> Age at exam 1 calculated using birth date and exam dates.
BMI (Body mass index, time variant)	Kg/m ²	Calculated		<ul style="list-style-type: none"> Calculated using BMI formula from height and weight
CLD_EVENT_PRI (Primary CLRD event: ICD code at first position)	-	Calculated		<ul style="list-style-type: none"> Based on ICD codes
COUNT (Number of observation for each subject)	-	Calculated		<ul style="list-style-type: none"> Count=1 can be used for cross sectional analysis
EDU_CAT (Highest education at baseline: 0=No schooling, 1=Grades 1-8, 2=Grades 9-11, 3=High school, 4=Some college, 5=Bachelor degree, 6=Graduate degree)	-	Calculated	LPSCHOOL	<ul style="list-style-type: none"> Education categories were created based on the highest education grade (lpschool) completed.

Table 5. Harmonization Protocol, Health, Aging and Body Composition Study, by Visit, United States, 1997-2006.

TOPMed Variable Name <i>(Label)</i>	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
EXAM <i>(Exam/visit at which spirometry was performed)</i>	-	Calculated		<ul style="list-style-type: none"> Indicates exam when spirometry was performed
GENDER <i>(Gender: 1=Male, 0=Female)</i>	-	Original		<ul style="list-style-type: none"> Recoded: 2 (female) to 0
HT_CM <i>(Height (cms), time variant)</i>	Centimeters	Original	Ex1: P2SH Ex4: D2SH Ex6: F3SH Ex8: S4SH Ex10: S4SH Ex11: Y11SH	<ul style="list-style-type: none"> Original variables had height recorded in millimeters. Height was converted to centimeters. In exams (2, 3, 5, 7 and 9) with missing data on height, height was calculated from BMI and weight.
HABCID <i>(Original participant id for ARIC study)</i>	-	Original		
PMH_ASTHMA_MD_BASE <i>(Doctor diagnosed asthma at baseline: 1=Yes, 0=No)</i>	-	Original	MHLCASTH	<ul style="list-style-type: none"> Recoded: 8 (don't know) to missing
PMH_ASTHMA_STILL_BASE <i>(Still have asthma at baseline: 1=Yes, 0=No)</i>	-	Original	MHLCSHA	<ul style="list-style-type: none"> Recoded: 8 (don't know) to missing

Table 5. Harmonization Protocol, Health, Aging and Body Composition Study, by Visit, United States, 1997-2006.

TOPMed Variable Name <i>(Label)</i>	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
PMH_CBRONCHITIS_MD_BASE <i>(Doctor diagnosed chronic bronchitis at baseline: 1=Yes, 0=No)</i>	-	Original	MHLCCHBR	<ul style="list-style-type: none"> Recoded: 8 (don't know) to missing
PMH_CBRONCHITIS_STILL_BASE <i>(Still have chronic bronchitis at baseline: 1=Yes, 0=No)</i>	-	Original	MHLCSHCB	<ul style="list-style-type: none"> Recoded: 8 (don't know) to missing
PMH_COPD_MD_BASE <i>(Doctor diagnosed COPD at baseline: 1=Yes, 0=No)</i>	-	Original	MHLCOPD	<ul style="list-style-type: none"> Recoded: 8 (don't know) to missing
PMH_DIABETES_MD_BASE <i>(Doctor diagnosed diabetes at baseline: 1=Yes, 0=No)</i>	-	Original	LBSGDIAB	<ul style="list-style-type: none"> Recoded: 8 (don't know) to missing
PMH_EMPHYSEMA_MD_BASE <i>(Doctor diagnosed emphysema at baseline: 1=Yes, 0=No)</i>	-	Original	MHLCEMPH	<ul style="list-style-type: none"> Recoded: 8 (don't know) to missing
PMH_HTN_MD_BASE <i>(Doctor diagnosed hypertension at baseline: 1=Yes, 0=No)</i>	-	Original	MHHCHBP	<ul style="list-style-type: none"> Recoded: 8 (don't know) to missing
PRE_FEV1 <i>(Forced expiratory volume in one sec (L), time variant)</i>	Liters	Original	Ex1-Ex10: BES_FEV1	<ul style="list-style-type: none"> FEV1 values were converted to liters from milliliters For one participant, recorded FEV1 at exam 1 was 0.27 and at exam 5 was 2.473. Therefore, FEV1 at exam 1 was changed to 2.7.

Table 5. Harmonization Protocol, Health, Aging and Body Composition Study, by Visit, United States, 1997-2006.

TOPMed Variable Name <i>(Label)</i>	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
PRE_FEV1FVC <i>(Ratio of fev1 over fvc (%), time variant)</i>	Percent	Calculated		<ul style="list-style-type: none"> Calculated using formula: pre_fev1fvc=pre_fev1/ pre_fvc*100
PRE_FVC <i>(Forced vital capacity (L), time variant)</i>	Liters	Original	Ex1-Ex10: BES_FVC	<ul style="list-style-type: none"> FVC values were converted to liters from milliliters
RACE <i>(Race: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5=American Indian, 6=Other)</i>	-	Original		<ul style="list-style-type: none"> Recoded: 2 (Black) to 3
RACE_BLACK <i>(Black or African American: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> if race=3 then race_black=1, else race_black=0
RACE_CHINESE <i>(Asian: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Study recruited only Caucasians and African Americans; so race_chinese=0
RACE_ETHNICITY <i>(Race/ethnicity: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5= American Indian, 6=Other)</i>	-	Original		<ul style="list-style-type: none"> Recoded: 2 (Black) to 3
RACE_HISPANIC <i>(Hispanic/Latino: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Study recruited only Caucasians and African Americans; so race_hispanic=0

Table 5. Harmonization Protocol, Health, Aging and Body Composition Study, by Visit, United States, 1997-2006.

TOPMed Variable Name <i>(Label)</i>	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
RACE_AMIND <i>(American Indian: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Study recruited only Caucasians and African Americans; so race_amind =0
RACE_OTHERS <i>(Other race: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Study recruited only Caucasians and African Americans; so race_others=0
RACE_WHITE <i>(White: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> if race=1 then race_white=1, else race_white=0
SMOKING_CIGS_PERDAY <i>(Cigarettes smoked per day)</i>	-	Original	Ex1: BQSCCGAV Ex3: CMSMOKAV Ex5: EBSMOKAV Ex8: RGSMOKAV Ex10: RGSMOKAV Ex11: Y11SMOKAV	
SMOKING_CIGS_PERDAY_BASE <i>(Cigarettes smoked per day at baseline visit)</i>	-	Original	BQSCCGAV	
SMOKING_CURRENT <i>(Current smoker, time variant: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=2 then smoking_current=1; else smoking_current=0

Table 5. Harmonization Protocol, Health, Aging and Body Composition Study, by Visit, United States, 1997-2006.

TOPMed Variable Name <i>(Label)</i>	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_CURRENT_BASELINE <i>(Current smoker at baseline: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=2 then smoking_current=1; else smoking_current=0
SMOKING_EVER <i>(Ever smoker, time variant: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=0 then smoking_ever=0; else smoking_ever=1
SMOKING_EVER_BASELINE <i>(Ever smoker at baseline: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=0 then smoking_ever=0; else smoking_ever=1
SMOKING_FORMER <i>(Former smoker, time variant: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=1 then smoking_former=1; else smoking_former=0
SMOKING_FORMER_BASELINE <i>(Former smoker at baseline: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=1 then smoking_former=1; else smoking_former=0
SMOKING_PACKYEARS <i>(Cigarettes pack years, time variant)</i>	Years	Calculated	PACKYR1	<ul style="list-style-type: none"> Cigarettes packyears were not provided for exam1. For subsequent exams packyears were calculated by using numbers of cigarettes smoked per day and total number of smoking years. Packyears= (numbers of cigarettes smoked per day) * (total number of smoking years)/20 For current smokers who have smoked continuously from baseline to last visit, total

Table 5. Harmonization Protocol, Health, Aging and Body Composition Study, by Visit, United States, 1997-2006.

TOPMed Variable Name (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
				<p>number of smoking years were calculated using following formulas: (age at current visit) - (age of start of smoking).</p> <ul style="list-style-type: none"> For former smokers who were current smokers at previous visit, total number of smoking years were calculated using following formulas {age at current visit - [(age at current visit) - (age at previous visit)/2]} - {age of start of smoking} For former smokers who had been former smoker at previous visit, total number of smoking years were carried forward from previous visit.
<p>SMOKING_PACKYEARS_BASELINE (Cigarettes pack years at baseline visit)</p>	Years	Calculated	PACKYR1	
<p>SMOKING_STATUS (3 levels of smoking status, time variant: 0=Never, 1=Former, 2=Current)</p>	-	Calculated	<p>Ex1: SMK1 Ex3: SMK3 Ex5: SMK5 Ex8: SMK8 Ex9: SMK9 Ex10: SMK10 Ex11: SMK11 Ex12: SMK12 Ex13: SMK13</p>	<ul style="list-style-type: none"> Recoded: 1 (current) to 2, 2 (former) to 1 Any participant with missing values for smoking status at any exam but recorded as never smoker in subsequent was considered as never smokers at previous exams too. Participants who were recorded as former smoker at exam 1 were considered as former smokers at subsequent exams unless specified. Missing observations were substituted with values based on some basic assumptions.

Table 5. Harmonization Protocol, Health, Aging and Body Composition Study, by Visit, United States, 1997-2006.

TOPMed Variable Name <i>(Label)</i>	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_STATUS_BASELINE <i>(3 levels of smoking status at baseline: 1=Yes, 0=No)</i>	-	Calculated	SMK1	
SPIRO_QC_FLAG <i>(Indicates that observation had FEV1 lower than subsequent at least by 1L)</i>	-	Calculated		<ul style="list-style-type: none"> • This variable flags the observation with FEV1 < 1L than subsequent exam.
STUDY <i>(Name of cohort, categorical: aric, card, chs, habc, fhs, mesa, hchs/sol, shs)</i>	-			
TIMEFACTOR_SPIRO <i>(Time from baseline spirometry exam, years)</i>	Years	Calculated		<ul style="list-style-type: none"> • Calculated using formula: age at spirometry - age at baseline spirometry
TTY_CLD_TOT_PRI <i>(Time to primary CLRD event)</i>	Years	Calculated		<ul style="list-style-type: none"> • Based on date of first clinical visit and date when event occurred
WT_KG <i>(Weight (KG), time variant)</i>	Kilograms	Calculated	Ex1: Y1WTK Ex2: Y2WTK Ex3: Y3WTK Ex4: Y4WTK Ex5: Y5WTK Ex6: Y6WTK Ex8: Y8WTK Ex10: Y10WTK Ex11: Y11WTK	<ul style="list-style-type: none"> • Weight in pounds was converted to kilograms.

Table 6. Harmonization protocol, Hispanic Community Health Study/Study of Latinos, United States, 2008-2011.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
AGE (Age (years), time variant)	Years	Original	AGE	
AGE_BASELINE (age at baseline exam)	Years	Original		
BMI (Body mass index, time variant)	Kg/m ²	Original	BMI	
COUNT (Number of observation for each subject)	-	Calculated		
EDU_CAT (Highest education at baseline:0=No schooling, 1=Grades 1-8, 2=Grades 9-11, 3=High school, 4=Some college, 5=Bachelor degree, 6=Graduate degree)	-	Calculated	EDUCATION_C3	<ul style="list-style-type: none"> • Since source variable was not available we used 3-level group education level (Education_C3: 1 = No high school diploma or GED 2 = At most a High school diploma or GED 3 = Greater than high school (or GED) education). • However, to be consistent with other cohorts values between 0 and 2 were imputed for Education_C3=1 and between 4 and 6 for Education_C3=3.

Table 6. Harmonization protocol, Hispanic Community Health Study/Study of Latinos, United States, 2008-2011.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
				<ul style="list-style-type: none"> Education_C3=2 was recoded as 3
ETHNICITY (Ethnicity: 1=Hispanic, 0=Non-Hispanic)	-	Calculated		
EXAM (Exam/visit at which spirometry is performed)	-	Calculated		
GENDER (Gender: 1=Male, 0=Female)	-	Original	GENDERNUM	
HT_CM (Height (cms), time variant)	Centimeter	Original	HEIGHT	
PMH_ASTHMA_EVER_BASE (Self-reported asthma ever, time variant: 1=Yes, 0=No)	-	Original	ASTHMA_EVER	<ul style="list-style-type: none"> Recoded: 9 (don't know) to . (missing)
PMH_ASTHMA_MD_BASE (Doctor confirmed asthma ever, time variant: 1=Yes, 0=No)	-	Original	ASTHMA_EVER_MD	

Table 6. Harmonization protocol, Hispanic Community Health Study/Study of Latinos, United States, 2008-2011.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
PMH_ASTHMA_STILL_BASE <i>(Still have asthma, time variant: 1=Yes, 0=No)</i>	-	Original	ASTHMA_CURR	<ul style="list-style-type: none"> Recoded: 9 (don't know) to . (missing)
PMH_DIABETES_MD_BASE <i>(Doctor diagnosed diabetes, time variant: 1=Yes, 0=No)</i>	-	Original	DIABETES_SELF	
PMH_HTN_MD_BASE <i>(Self-reported awareness of HTN in those with HTN defined by NHANES, time variant: 1=Aware, 0=Non-aware)</i>	-	Original	HYPERT_AWARENESS	
PRE_FEV1 <i>(Forced expiratory volume in one sec (L), time variant)</i>	Liters	Original	PRBA29	<ul style="list-style-type: none"> Converted to litres
PRE_FEV1FVC <i>(Ratio of fev1 over fvc (%), time variant)</i>	Percent	Original	PRBA39	<ul style="list-style-type: none"> Converted to percent
PRE_FVC <i>(Forced vital capacity (L), time variant)</i>	Liters	Original	PRBA25	<ul style="list-style-type: none"> Converted to litres
RACE <i>(Race: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5=American Indian, 6=Other)</i>	-	Calculated		<ul style="list-style-type: none"> HCHS was conducted only in Hispanics. So all participants in this study are coded as hispanics.

Table 6. Harmonization protocol, Hispanic Community Health Study/Study of Latinos, United States, 2008-2011.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
RACE_BLACK (Black or African American: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • Since race=4 (hispanic), race_black=0
RACE_CHINESE (Asian: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • Since race=4 (hispanic), race_chinese=0
RACE_ETHNICITY (Race/ethnicity: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5= American Indian, 6=Other)	-	Calculated		<ul style="list-style-type: none"> • HCHS was conducted only in Hispanics. So all participants in this study are coded as Hispanics.
RACE_HISPANIC (Hispanic/Latino: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • Since race=4 (hispanic), race_hispanic=1
RACE_AMIND (American Indian: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • Since race=4 (hispanic), race_amind=0
RACE_OTHERS (Other race: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • Since race=4 (hispanic), race_others=0
RACE_WHITE (White: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> • Since race=4 (v), race_white=0

Table 6. Harmonization protocol, Hispanic Community Health Study/Study of Latinos, United States, 2008-2011.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_CIGS_PERDAY <i>(Cigarettes smoked per day)</i>	-	Calculated		<ul style="list-style-type: none"> Number of cigarettes per day are calculated using formula: $\text{Cigarettes_Year (number of exposure years multiplied by the average number of cigarettes smoked per day)} / \text{Exposure_Year (total number of years a participant smoked ciigarettes)}$
SMOKING_CIGS_PERDAY_BASE <i>(Cigarettes smoked per day at baseline visit)</i>	-	Calculated		<ul style="list-style-type: none"> Number of cigarettes per day are calculated using formula: $\text{Cigarettes_Year (number of exposure years multiplied by the average number of cigarettes smoked per day)} / \text{Exposure_Year (total number of years a participant smoked ciigarettes)}$
SMOKING_CURRENT <i>(Current smoker, time variant: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=2 then smoking_current=1; else smoking_current=0
SMOKING_CURRENT_BASELINE <i>(Current smoker at baseline: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=2 then smoking_current=1; else smoking_current=0
SMOKING_EVER <i>(Ever smoker, time variant: 1=Yes, 0=No)</i>	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=0 then smoking_ever=0; else smoking_ever=1

Table 6. Harmonization protocol, Hispanic Community Health Study/Study of Latinos, United States, 2008-2011.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_EVER_BASELINE (Ever smoker at baseline: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=0 then smoking_ever=0; else smoking_ever=1
SMOKING_FORMER (Former smoker, time variant: 1=Yes, 0=Nov)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=1 then smoking_former=1; else smoking_former=0
SMOKING_FORMER_BASELINE (Former smoker at baseline: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=1 then smoking_former=1; else smoking_former=0
SMOKING_PACKYEARS (Cigarettes pack years, time variant)	Years	Original	CIGARETTE_PACK_ YEARS	
SMOKING_PACKYEARS_BASELINE (Cigarettes pack years at baseline visit)	Years	Original	CIGARETTE_PACK_ YEARS	
SMOKING_STATUS (3 levels of smoking status, time variant:0=Never, 1=Former, 2=Current)	-	Original	CIGARETTE_USE	<ul style="list-style-type: none"> Recoded: 1 (never) to 0, 2 (former) to 1, 3 (current) to 2

Table 6. Harmonization protocol, Hispanic Community Health Study/Study of Latinos, United States, 2008-2011.

Variable <i>(Label)</i>	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_STATUS_BASELINE <i>(3 levels of smoking status at baseline: 1=Yes, 0=No)</i>	-	Original	CIGARETTE_USE	<ul style="list-style-type: none"> Recoded: 1 (never) to 0, 2 (former) to 1, 3 (current) to 2
SPIRO_QC_FLAG <i>(Indicates that observation had FEV1 lower than subsequent at least by 1L)</i>	-	Calculated		<ul style="list-style-type: none"> This variable flags the observation with FEV1 < 1L than subsequent exam.
TIMEFACTOR_SPIRO <i>(Time from baseline spirometry exam, years)</i>	Years	Calculated		<ul style="list-style-type: none"> Calculated using formula: age at spirometry - age at baseline spirometry
WT_KG <i>(Weight (KG), time variant)</i>	Kilogram	Original	ANTA4	

Table 7. Harmonization protocol, Jackson Heart Study, United States, 2000-2004.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
AGE (Age (years), time variant)	Years	Original	AGE	
AGE_BASELINE (age at baseline exam)	Years	Original	AGE	
BMI (Body mass index, time variant)	Kg/m ²	Original		
COUNT (Number of observation for each subject)	-	Calculated		<ul style="list-style-type: none"> Count=1 can be used for cross sectional analysis
EDU_CAT (Highest education at baseline:0=No schooling, 1=Grades 1-8, 2=Grades 9-11, 3=High school, 4=Some college, 5=Bachelor degree, 6=Graduate degree)	-	Calculated	PDSA18A	<ul style="list-style-type: none"> Education categories were created based on the highest education grade (PDSA18A) completed.
EXAM (Exam/visit at which spirometry is performed)	-	Calculated		<ul style="list-style-type: none"> Indicates exam when spirometry was performed
GENDER (Gender: 1=Male, 0=Female)	-	Original	MALE	

Table 7. Harmonization protocol, Jackson Heart Study, United States, 2000-2004.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
HT_CM (Height (cms), time variant)	Centimeter	Centimeters	JH157701_JP0792	HEIGHT
PMH_ASTHM_AGE_START_BASE (Age when asthma started at baseline)	Years	-		
PMH_ASTHMA_EVER_BASE (Self-reported asthma ever at baseline: 1=Yes, 0=No)	-	Years	JH157701_JP0792	RPAA15
PMH_ASTHMA_MD_BASE (Doctor diagnosed asthma at baseline: 1=Yes, 0=No)	-	-	JH157701_JP0792	RPAA13
PMH_ASTHMA_STILL_BASE (Still have asthma at baseline: 1=Yes, 0=No)	-	-	JH157701_JP0792	RPAA14
PMH_DIABETES_MD_BASE (Doctor diagnosed diabetes at baseline: 1=Yes, 0=No)	-	Original	DIABETES	
PMH_HTN_MD_BASE (Doctor diagnosed hypertension at baseline: 1=Yes, 0=No)	-	Original	HTN	

Table 7. Harmonization protocol, Jackson Heart Study, United States, 2000-2004.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
PRE_FEV1 (Forced expiratory volume in one sec (L), time variant)	Liters	Original	PULA18	<ul style="list-style-type: none"> 4 observations with negative FEV1 values were deleted.
PRE_FEV1FVC (Ratio of fev1 over fvc (%), time variant)	Percent	Calculated		<ul style="list-style-type: none"> Calculated for best FEV1 and FVC
PRE_FEV1FVC_QC (Quality assessment grades for FEV1/FVC: 1=pre_fev1_qc and pre_fvc_qc=1, 0=pre_fev1_qc or pre_fvc_qc=0)	-	Calculated		<ul style="list-style-type: none"> Data for 3 accepted trials was provided. Difference between wo largest trials was calculated. Indicates that both FEV1 and FVC have two acceptable curves and reproducibility within 200 ml
PRE_FEV1FVC_QC_150 (Quality assessment grades for FEV1/FVC: 1=pre_fev1_qc_150 and pre_fvc_qc_150=1, 0=pre_fev1_qc_150 or pre_fvc_qc_150=0)	-	Calculated		<ul style="list-style-type: none"> Data for 3 accepted trials was provided. Difference between wo largest trials was calculated. Indicates that both FEV1 and FVC have two acceptable curves and reproducibility within 150 ml
PRE_FVC (Forced vital capacity (L), time variant)	Liters	Original	PULA17	
RACE (Race: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5= American Indian, 6=Other)	-	Calculated		<ul style="list-style-type: none"> Jackson Heart Study was conducted only in African American. Therefore, race was not asked in exam 1. So all participants in this study are coded as black.

Table 7. Harmonization protocol, Jackson Heart Study, United States, 2000-2004.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
RACE_BLACK (Black or African American: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Since race=3 (black), race_black=1
RACE_CHINESE (Asian: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Since race=3 (black), race_chinese=0
RACE_ETHNICITY (Race/ethnicity: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5= American Indian, 6=Other)	-	Calculated		<ul style="list-style-type: none"> Jackson Heart Study was conducted only in African American. Therefore, race was not asked in exam 1. So all participants in this study are coded as black.
RACE_HISPANIC (Hispanic/Latino: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Since race=3 (black), race_hispanic=0
RACE_AMIND (American Indian: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Since race=3 (black), race_amind=0
RACE_OTHERS (Other Race: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Since race=3 (black), race_others=0
RACE_WHITE (White: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Since race=3 (black), race_white=0

Table 7. Harmonization protocol, Jackson Heart Study, United States, 2000-2004.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_CIGS_PERDAY (Cigarettes smoked per day)	-	Original	TOBA5	
SMOKING_CIGS_PERDAY_BASE (Cigarettes smoked per day at baseline visit)	-	Original	TOBA5	
SMOKING_CURRENT (Current smoker, time variant: 1=Yes, 0=No)	-	Original	CURRENTSMOKER	
SMOKING_CURRENT_BASELINE (Current smoker at baseline: 1=Yes, 0=No)	-	Original	CURRENTSMOKER	
SMOKING_EVER (Ever smoker, time variant: 1=Yes, 0=No)	-	Original	EVERSMOKER	
SMOKING_EVER_BASELINE (Ever smoker at baseline: 1=Yes, 0=No)	-	Original	EVERSMOKER	
SMOKING_FORMER (Former smoker, time variant: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> If smoking_current=0 and smoking_ever=1 then smoking_former=1

Table 7. Harmonization protocol, Jackson Heart Study, United States, 2000-2004.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_FORMER_BASELINE (Former smoker at baseline: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> If smoking_current_baseline =0 and smoking_ever_baseline =1 then smoking_former_baseline =1
SMOKING_PACKYEARS (Cigarettes pack years, time variant)	Years	Calculated	TOBA2, TOBA4A, TOBA4B, TOBA5, AGE	<ul style="list-style-type: none"> Smoking pack years were calculated using cigarettes smoked per day, age of start of smoking, current age and number of months/years since quit smoking. First total number of years of cigarette smoking were calculated as: For current smokers= (current age) – (age of start of smoking) For former smokers= (current age - months/years since quit smoking) – (age of start of smoking) Pack years were calculated using formula: ((years of cigarette smoking) x (cigarettes smoked per day)) / 20
SMOKING_PACKYEARS_BASELINE (Cigarettes pack years at baseline visit)	Years	Calculated	TOBA2, TOBA4A, TOBA4B, TOBA5, AGE	<ul style="list-style-type: none"> Smoking pack years were calculated using cigarettes smoked per day, age of start of smoking, current age and number of months/years since quit smoking. First total number of years of cigarette smoking were calculated as: For current smokers= (current age) – (age of start of smoking) For former smokers= (current age - months/years since quit smoking) – (age of start of smoking)

Table 7. Harmonization protocol, Jackson Heart Study, United States, 2000-2004.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
				<ul style="list-style-type: none"> Pack years were calculated using formula: ((years of cigarette smoking) x (cigarettes smoked per day)) / 20
SMOKING_STATUS (3 levels of smoking status, time variant: 0=Never, 1=Former, 2=Current)	-	Calculated		<ul style="list-style-type: none"> if smoking_ever=0 then smoking_status=0; if smoking_former=1 then smoking_status=1; if smoking_current=1 then smoking_status=2;
SMOKING_STATUS_BASELINE (3 levels of smoking status at baseline: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if smoking_ever=0 then smoking_status=0; if smoking_former=1 then smoking_status=1; if smoking_current=1 then smoking_status=2;
SPIRO_QC_FLAG (Indicates that observation had FEV1 lower than subsequent at least by 1L)	-	Calculated		<ul style="list-style-type: none"> This variable flags the observation with FEV1 < 1L than subsequent exam.
STUDY (Name of cohort, categorical: aric, card, chs, habc, fhs, mesa, hchs/sol, shs)	-			
SUBJID (Original participant id for JHS)	-	Calculated		<ul style="list-style-type: none"> Calculated using formula: age at spirometry - age at baseline spirometry

Table 7. Harmonization protocol, Jackson Heart Study, United States, 2000-2004.

Variable <i>(Label)</i>	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
TIMEFACTOR_SPIRO <i>(Time from baseline spirometry exam, years)</i>	Years	Original	WEIGHT	
WT_KG <i>(Weight (KG), time variant)</i>	Kilograms			

Figure 6. Flow Chart of Longitudinal Lung Function Data in the Multi-Ethnic Study of Atherosclerosis, by Visit, United States, 2003-2012.

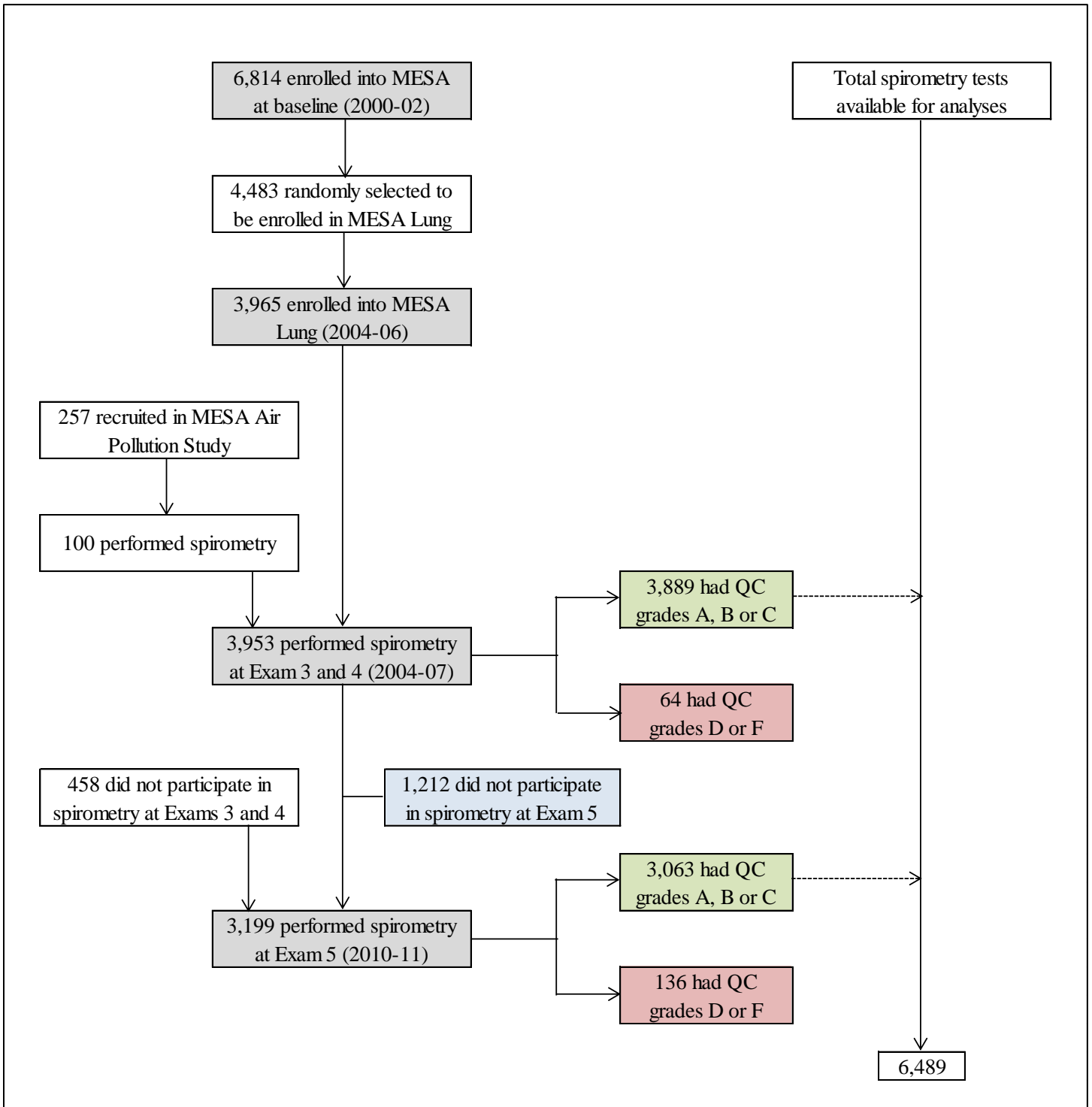


Table 8. Harmonization Protocol, Multi-Ethnic Study of Atherosclerosis, United States, 2003-2012.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
AGE (Age (years), time variant)	Years	Original	Ex 1: AGE1C Ex 2: AGE2C Ex 3: AGE3C Ex 4: AGE4C Ex 5: AGE5C MESA AIR: AGE1C	
AGE_BASELINE (Age at baseline, exam 1)	Years	Original	AGE1C	
BMI (Body mass index, time variant)	Kg/m ²	Original	Ex 1: BMI1C Ex 2: BMI2C Ex 3: BMI3C Ex 4: BMI4C Ex 5: BMI5C	
CLD_EVENT_PRI (Primary CLRD event: ICD) code at first position	-	Calculated		<ul style="list-style-type: none"> Based on ICD codes
COUNT (Number of observation for each subject)	-	Calculated		<ul style="list-style-type: none"> Count=1 can be used for cross sectional analysis
EDU_CAT (Highest education at baseline:0=No schooling, 1=Grades 1-8, 2=Grades 9- 11, 3=High school, 4=Some college, 5=Bachelor degree, 6=Graduate degree)	-	Calculated	EDUC1	<ul style="list-style-type: none"> New variable for education categories was created based on the highest education grade (hom54) completed.

Table 8. Harmonization Protocol, Multi-Ethnic Study of Atherosclerosis, United States, 2003-2012.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
EXAM (Exam/visit at which spirometry is performed)	-	Calculated		<ul style="list-style-type: none"> Exam=1 represents baseline where no spirometry was performed
GENDER (Gender: 1=Male, 0=Female)	-	Original	GENDER1	
HT_CM (Height (cms), time variant)	Centimeter	Original	Ex 1: HTCM1 Ex 2: HTCM2 Ex 3: HTCM3 Ex 4: HTCM4 Ex 5: HTCM5	
PMH_ASTHMA_MD_BASE (Doctor diagnosed asthma ever, time variant: 1=Yes, 0=No)	-	Original	ASTHMA1	
PMH_DIABETES_MD_BASE (Doctor diagnosed diabetes, time variant: 1=Yes, 0=No)	-	Original	DIABET1	
PMH_EMPHYSEMA_MD_BASE (Doctor diagnosed emphysema, time variant: 1=Yes, 0=No)	-	Original	EMPHYS1	
PMH_HTN_MD_BASE (Doctor diagnosed hypertension, time variant: 1=Yes, 0=No)	-	Original	HIGHBP1	

Table 8. Harmonization Protocol, Multi-Ethnic Study of Atherosclerosis, United States, 2003-2012.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
PRE_FEV1 (Forced expiratory volume in one sec (L), time variant)	Liters	Original		
PRE_FEV1FVC (Ratio of pre_fev1 over pre_fvc (%), time variant)	Percent	Original		
PRE_FVC (Forced vital capacity (L), time variant)	Liters	Original		
RACE (Race: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5= American Indian, 6=Other)	-	Original	RACE1C	
RACE_BLACK (Black or African American: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=3 then race_black=1, else race_black=0
RACE_CHINESE (Asian: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=2 then race_chinese=1, else race_chinese=0
RACE_ETHNICITY (Race/ethnicity: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5= American Indian, 6=Other)	-	Original	RACE1C	

Table 8. Harmonization Protocol, Multi-Ethnic Study of Atherosclerosis, United States, 2003-2012.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
RACE_HISPANIC (Hispanic/Latino: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=4 then race_hispanic=1, else race_hispanic=0
RACE_AMIND (American Indian: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=5 then race_amind=1, else race_amind =0
RACE_OTHERS (Other race: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=6 then race_others=1, else race_others=0
RACE_WHITE (White: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=1 then race_white=1, else race_white=0
SMOKING_CIGS_PERDAY (Cigarettes smoked per day, time variant)	--	Original	Ex 1: CIGSDAY1 Ex2: CIGSDAY2 Ex 3: CIGSDAY3 Ex 4: CIGSDAY4 Ex 5: CIGSDAY5 MESA AIR: CIGSDAY1	
SMOKING_CIGS_PERDAY_BASE (Cigarettes smoked per day at baseline)	-	Original	CIGSDAY1	
SMOKING_CURRENT (Current smoker, time variant: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=2 then smoking_current=1; else smoking_current=0

Table 8. Harmonization Protocol, Multi-Ethnic Study of Atherosclerosis, United States, 2003-2012.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_CURRENT_BASELINE (Current smoker at exam 1:1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=2 then smoking_current=1; else smoking_current=0
SMOKING_EVER (Ever smoker, time variant:1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=0 then smoking_ever=0; else smoking_ever=1
SMOKING_EVER_BASELINE (Ever smoker at exam 1:1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=0 then smoking_ever=0; else smoking_ever=1
SMOKING_FORMER (Former smoker, time variant: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=1 then smoking_former=1; else smoking_former=0
SMOKING_FORMER_BASELINE (Former smoker at exam 1:1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=1 then smoking_former=1; else smoking_former=0
SMOKING_PACKYEARS (Cigarettes pack years ,time variant)	Years	Original	Ex 1: PKYRS1C Ex 2: PKYRS2C Ex3: PKYRS3C Ex 4: PKYRS4C Ex 5: PKYRS5C MESA AIR: PKYRS1C	
SMOKING_PACKYEARS_BASELINE (Cigarettes pack years at baseline visit)	Years	Original	PKYRS1C	

Table 8. Harmonization Protocol, Multi-Ethnic Study of Atherosclerosis, United States, 2003-2012.

Variable (Label)	Units	Original/ Calculated	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_STATUS <i>(3 levels of smoking status, time variant: 0=Never, 1=Former, 2=Current)</i>	-	Original	Ex 1: CIG1C Ex 2: CIG2C Ex 3: CIG3C Ex 4: CIG4C Ex 5: CIG5C MESA AIR: CIG1C	
SMOKING_STATUS_BASELINE <i>(3 levels of smoking status at exam 1: 1=Yes, 0=No)</i>	-	Original	CIG1C	
SPIRO_QC_FLAG <i>(Indicates that observation had FEV1 lower than subsequent at least by 1L)</i>	-	Calculated		<ul style="list-style-type: none"> This variable flags the observation with FEV1 < 1L than subsequent exam.
TIMEFACTOR_SPIRO <i>(Time from baseline spirometry exam, years)</i>	Years	Calculated		<ul style="list-style-type: none"> Calculated using formula: age at spirometry - age at baseline spirometry
TTY_CLD_TOT_PRI <i>(Time to primary CLRD event)</i>	-	Calculated		<ul style="list-style-type: none"> Based on date at which event occurred and censored
WT_KG <i>(Weight (KG), time variant)</i>	Kilogram	Calculated	Ex 1: WTLB1 Ex 2: WTLB2 Ex 3: WTLB3 Ex 4: WTLB4 Ex 5: WTLB5	<ul style="list-style-type: none"> Originally weight was recorded in pounds which were converted to kgs.

Table 9. Harmonization Protocol, Strong Heart Study, United States, 1989-1991.

Variable (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
AGE (Age (years), time variant)	Years	Original	Phase 1: S1AGE Phase 2: S2AGE	
AGE_BASELINE (Age at first spirometry exam)	Years	Original	Phase 1: S1AGE	•
BMI (Body mass index, time variant)	Kg/m ²	Original	Phase 1: S1BMI Phase 2: S2BMI	•
COUNT (Number of observation for each subject)	-			
EDU_CAT (Highest education at baseline: 0=No schooling, 1=Grades 1-8, 2=Grades 9-11, 3=High school, 4=Some college, 5=Bachelor degree, 6=Graduate degree)	-	Calculated	Phase 1: S1EDU	• New variable for education categories was created based on the highest education grade (S1EDU) completed.
EXAM (Exam/visit at which spirometry is performed)	-	Original		
GENDER (Gender: 1=Male, 0=Female)	-	Original	Phase 1: SEX	• Recoded: F (female) to 0, M (male) to 1

Table 9. Harmonization Protocol, Strong Heart Study, United States, 1989-1991.

Variable (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
HT_CM (Height (cms), time variant)	Centimeter	Original	Phase 1: S1HT Phase 2: S2HT	
IDNO (Original participant id for SHS study)	-	Original		
PMH_ASTHMA_MD_BASE (Doctor diagnosed asthma at baseline: 1=Yes, 0=No)	-	Original	Phase 2: s2med2_22	<ul style="list-style-type: none"> • Question asked only in Phase 2
PMH_ASTHMA_STILL_BASE (Still have asthma at baseline: 1=Yes, 0=No)	-	Original	Phase 2: s2med2_23	<ul style="list-style-type: none"> • Question asked only in Phase 2
PMH_CBRONCHITIS_MD_BASE (Doctor diagnosed chronic bronchitis at baseline: 1=Yes, 0=No)	-	Original	Phase 2: s2med2_21	<ul style="list-style-type: none"> • Question asked only in Phase 2
PMH_DIABETES_MD_BASE (Doctor diagnosed diabetes at baseline: 1=Yes, 0=No)	-	Original	Phase 1: s1dmhx_n	<ul style="list-style-type: none"> • Recoded: 1(no) to 0, 2 (borderline) to 0, 3(yes) to 1
PMH_EMPHYSEMA_MD_BASE (Doctor diagnosed emphysema at baseline: 1=Yes, 0=No)	-	Original	Phase 1: s1med31	

Table 9. Harmonization Protocol, Strong Heart Study, United States, 1989-1991.

Variable (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
PMH_HTN_MD_BASE (Doctor diagnosed hypertension at baseline: 1=Yes, 0=No)	-	Original	Phase 1: s1htnhx_n	
PRE_FEV1 (Forced expiratory volume in one sec (L), time variant)	Liters	Original	s2fev1	
PRE_FEV1FVC (Ratio of pre_fev1 over pre_fvc (%), time variant)	Percent	Calculated	s2fev1fvc_ratio	
PRE_FVC (Forced vital capacity (L), time variant)	Liters	Original	s2fvc	
RACE (Race: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5=American Indian, 6=Other)	-	Calculated		<ul style="list-style-type: none"> Race was coded as 5 for American Indians
RACE_BLACK (Black or African American: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Since race=1 (American Indians), race_black=0
RACE_CHINESE (Asian: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Since race=1 (American Indians), race_chinese=0

Table 9. Harmonization Protocol, Strong Heart Study, United States, 1989-1991.

Variable (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
RACE_ETHNICITY (Race/ethnicity: 1=White, 2=Asian, 3=Black, 4=Hispanic/Latino, 5= American Indian, 6=Other)	-	Calculated		<ul style="list-style-type: none"> Race was coded as 5 for American Indians
RACE_HISPANIC (Hispanic/Latino: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Since race=1 (American Indians), race_hispanic=0
RACE_AMIND (American Indian: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> if race=5 then race_amind=1, else race_amind =0
RACE_OTHERS (Other race: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Since race=1 (American Indians), race_others=0
RACE_WHITE (White: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Since race=1 (American Indians), race_white=1
SMOKING_CURRENT (Current smoker, time variant: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=2 then smoking_current=1; else smoking_current=0
SMOKING_CURRENT_BASELINE (Current smoker at baseline: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=2 then smoking_current=1; else smoking_current=0

Table 9. Harmonization Protocol, Strong Heart Study, United States, 1989-1991.

Variable (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
SMOKING_EVER (Ever smoker, time variant: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=0 then smoking_ever=0; else smoking_ever=1
SMOKING_EVER_BASELINE (Ever smoker at baseline: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=0 then smoking_ever=0; else smoking_ever=1
SMOKING_FORMER (Former smoker, time variant: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=1 then smoking_former=1; else smoking_former=0
SMOKING_FORMER_BASELINE (Former smoker at baseline: 1=Yes, 0=No)	-	Calculated		<ul style="list-style-type: none"> Calculated from smoking status: if smoking status=1 then smoking_former=1; else smoking_former=0
SMOKING_PACKYEARS (Cigarettes pack years, time variant)	Years	Original	Phase 1: s1ppy Phase 2: s2ppy	<ul style="list-style-type: none"> For former smokers who had been former smoker at previous visit, total number of pack years were carried forward from previous visit.
SMOKING_PACKYEARS_BASELINE (Cigarettes pack years at baseline visit)	Years	Original	Phase 1: s1ppy	
SMOKING_STATUS (3 levels of smoking status, time variant: 0=Never, 1=Former, 2=Current)	-	Original	Phase 1: s1smoke_n Phase 2: s2smoke_n	<ul style="list-style-type: none"> Recoded: 1 (never) to 0, 2 (former) to 1, 3 (current) to 0 Missing observations were substituted with values based on some basic assumptions. Any participant with missing values for smoking status at any exam but recorded as

Table 9. Harmonization Protocol, Strong Heart Study, United States, 1989-1991.

Variable (Label)	Units	Calculated / Original	Original Variable Name	Details regarding classification, calculation, and/or correction
				never smoker in subsequent was considered as never smokers at previous exams too. <ul style="list-style-type: none"> Participants who were recorded as former smoker at exam 1 were considered as former smokers at subsequent exams unless specified.
SMOKING_STATUS_BASELINE <i>(3 levels of smoking status at baseline: 1=Yes, 0=No)</i>	-	Original	Phase 1: s1smoke_n	
STUDY <i>(Name of cohort, categorical: aric, card, chs, habc, fhs, mesa, hchs/sol, shs)</i>	-	Calculated		
TIMEFACTOR_SPIRO <i>(Time from baseline spirometry exam, years)</i>	Years	Calculated		<ul style="list-style-type: none"> Timefactor is zero because of only one spirometry exam.
WT_KG <i>(Weight (KG), time variant)</i>	Kilogram	Original	Phase 1: S1WT Phase 2: S2WT	