



HCHS/SOL BASELINE DERIVED VARIABLE DICTIONARY

April 2026

**Prepared by the Collaborative Studies Coordinating Center
UNC Chapel Hill, Department of Biostatistics**

FUNDING: The Hispanic Community Health Study/Study of Latinos is a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) 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 HCHS/SOL 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.

Version 5 – Investigator Use Data Sets (_INV5 data file)

HCHS/SOL Baseline Derived Variable Dictionary

Table of Contents

SUMMARY OF UPDATES IN APRIL 2026 DATA RELEASE	10
1. DESIGN	12
1.1 COHORT (Recruitment year from sample address list).....	12
1.2 PSU_ID (Primary sampling unit ID clustering variable- masked).....	12
1.3 STRAT (Sampling design stratification variable - masked).....	12
1.4 LISTNUM (List address record assigned)	12
1.5 PSU_CNT (BG-level Hispanic household frame counts).....	12
1.6 STRAT_CNT (Stratum-level Hispanic household frame count).....	12
1.7 LIST_CNT (Sample list - level Hispanic)	12
1.8 HH_ID (Household ID).....	12
1.9 WEIGHT_FINAL_EXPANDED (Expanded Final Weight)	13
1.10 WEIGHT_FINAL_NORM_OVERALL (Overall Normalized Final Weight)	13
1.11 WEIGHT_FINAL_NORM_CENTER (FC-Specific Normalized Final Weight)	13
1.12 WEIGHT_NONRESP (Non-response adjusted sampling weight).....	13
2. ADMINISTRATIVE	14
2.1 CENTER (Participant's field center)	14
2.2 CENTERNUM – Numeric (Participant's field center)	14
2.3 CLINDATE (Date of the participant's clinic visit).....	14
2.4 CONSENT (Informed consent status for participation in study).....	14
2.5 GENOTYPING (Full consent for genotyping)	15
2.6 DNA_AVAILABLE (DNA available)	15
2.7 NP_PERMIT (Permission for non-profit use of data)	16
2.8 EXTERNAL_PERMIT (Permission for investigators outside HCHS/SOL)	16
2.9 COMMERCIAL_PERMIT (Permission for for-profit use of data).....	17
2.10 ANY_PERMIT (Any permission for public use data).....	17
3. SOCIO-DEMOGRAPHIC	18
3.1 AGE (Age in years at the time of participant's clinic visit).....	18
3.2 AGEGROUP_C2 (2-level grouped age of participant).....	18
3.3 AGEGROUP_C3 (3-level grouped age of participant).....	18
3.4 AGEGROUP_C6 (6-level grouped age of participant).....	19
3.5 AGEGROUP_C6_NHANES (6-level grouped age – NHANES standardization).....	19
3.6 SEX (Sex – Numeric).....	19
3.7 EDUCATION_C2 (2-level group education level).....	20
3.8 EDUCATION_C3 (3-level group education level).....	21
3.9 EDUCATION_C4 (4-level group education level).....	22
3.10 INCOME (Grouped yearly household income)	22
3.11 INCOME_C3 (3-level grouped yearly household income)	23
3.12 INCOME_C5 (5-level grouped yearly household income)	23
3.13 MARITAL_STATUS (3-level grouped marital status).....	24
3.14 EMPLOYED (4-level grouped employment status).....	24
3.15 HC (Current Health Insurance Coverage Excluding Indian Health Services).....	25
3.16 OCCUPATION_CURR (Grouped current occupation for employed participants)	26
3.17 OCCUPATION_LONG (Grouped longest held occupation).....	26

3.18	ADDRESS_YRS (Years lived at address).....	26
4.	ACCULTURATION	27
4.1	BKGRD1 (Re-classification of Hispanic/Latino Background).....	27
4.2	BKGRD1_C7 (7-level re-classification of Hispanic/Latino background).....	28
4.3	BKGRD1_C7_NOMISS (7-level Hispanic/Latino background, with no missing).....	29
4.4	BKGRD1_C6 (6-level re-classification of Hispanic/Latino background).....	29
4.5	SITE_BKGRD (Cross-classification of center and Hispanic/Latino background).....	30
4.6	SITE_BKGRD_45PLUS (Cross-classification of center and Hispanic/Latino background, 45+ years only).....	31
4.7	RACE (Race Group (self-report)).....	32
4.8	US_BORN (Grouped place of birth – 50 US States only).....	32
4.9	US_NATIVE (Grouped place of birth – US states and territories).....	32
4.10	LANG_PREF (Language preference for baseline examination).....	33
4.11	IMGEN_C2 (2-level immigrant generation).....	33
4.12	IMGEN_C4 (4-level immigrant generation).....	34
4.13	YRSUS (Years lived in the US).....	35
4.14	YRSUS_C2 (2-level grouped years lived in the US (50 States)).....	35
4.15	YRSUS_C3 (3-level grouped years lived in the US (50 States)).....	35
4.16	SASH_ALL (Short acculturation scale for Hispanic).....	35
4.17	SASH_LANG (Short acculturation scale for Hispanics – language subscale).....	36
4.18	SASH_MEDIA (Short acculturation scale for Hispanics – media subscale).....	36
4.19	SASH_SOC (Short acculturation scale for Hispanics – social subscale).....	36
4.20	ETHIS (Ethnic Identification Score).....	37
4.21	LANGUAGE_SUBSCORE_MESA (Language Subscore - MESA).....	37
4.22	NATIVITY_SUBSCORE_MESA (Nativity Subscore - MESA).....	38
4.23	ACCULT_MESA (Acculturation Score - MESA).....	38
4.24	AGE_IMMI (Age of immigration among participants not born in US mainland).....	39
4.25	AGE_IMMI_C3 (3-level age of immigration).....	39
5.	ANTHROPOMETRY	40
5.1	BMI (Body Mass Index kg/m ²).....	40
5.2	BMIGRP_C4 (4-level grouped Body Mass Index - WHO).....	40
5.3	BMIGRP_C6 (6-level grouped Body Mass Index - WHO).....	40
5.4	WAIST_HIP (Waist to hip ratio).....	41
5.5	HEIGHT (Height corrected in cm).....	41
6.	RESPIRATORY	42
6.1	ASTHMA_EVER (Ever had asthma).....	42
6.2	ASTHMA_EVER_MD (Ever had asthma with MD diagnosis).....	42
6.3	ASTHMA_CURR (Currently have asthma).....	42
6.4	ASTHMA_CURR_MD (Currently have asthma with ever MD diagnosis).....	43
6.5	ASTHMA_C4 (4-level grouped asthma status).....	43
6.6	ASTHMA_C4_MD (4-level grouped asthma status – by MD diagnosis).....	44
6.7	COPD_SELF (Self-report of ever had COPD or emphysema).....	44
6.8	COPD_EVER (Self-report of ever had COPD, emphysema or chronic bronchitis).....	44
6.9	MRC_CB (Chronic Productive Cough 3+ mths in 2+ yrs).....	45
6.10	WHEEZE_SB (Wheezing ever with shortness of breath).....	45
6.11	WHEEZE_LAST12 (Wheezing in last 12 months).....	45
6.12	PHLEGM (Phlegm day/night 3 months in a row).....	46
6.13	PNEUMONIA (Pneumonia - by MD diagnosis).....	46

6.14	VALID_SPIROMETRY (Valid spirometry studies, pre-BD)	46
6.15	VALID_SPIROMETRY_POSTBD (Valid spirometry studies, post-BD)	46
6.16	AIRFLOW_OBSTRUCTION (Obstructive lung disease)	47
6.17	AIRFLOW_LIMITATION (Airflow limitation (pre FEV1/FVC<70 only))	47
6.18	COPD_BY_BD (COPD by bronchodilator test)	47
6.19	COPD_BY_BD_ALLPRE (COPD by BD test (for all valid pre_BD))	48
6.20	BD_RESPONSE_ABS (Bronchodilator responsive by 15% change)	48
6.21	FEV1_FVC_RATIO (FEV ₁ to FVC %)	48
6.22	POSTBD_PREBD_DIFF (Post(FEV1%pred) - Pre(FEV1%pred))	49
6.23	POSTBD_PREBD_DIFF_GT12 (Relative bronchodilator responsive by 12% change)	49
6.24	POSTBD_PREBD_DIFF_GT15 (Relative bronchodilator responsive by 15% change)	50
6.25	POSTBD_PREBD_DIFF_GT12_2 (Relative bronchodilator responsive by 12% change (among pre FEV1/FVC<70 only))	50
6.26	POSTBD_EXPECTED (Post-BD expected (among valid spirometry))	50
6.27	VALID_SPIROMETRY_POSTBD2 (Valid post-spirometry (A-D grading) adjusted for unexpected BDs)	51
7.	WELL-BEING/QUALITY OF LIFE	52
7.1	CESD10 (10-item CES-D summary score)	52
7.2	STAI10 (10-item state-trait anxiety inventory summary score)	52
8.	SF-12V2 HEALTH SURVEY	54
8.1	AGG_MENT (Aggregate mental health scale)	54
8.2	AGG_PHYS (Aggregate physical health scale)	54
9.	BLOOD PRESSURE MEASURES	55
9.1	LEFT_ABI, RIGHT_ABI, AND ABI_OVERALL (Ankle Brachial Index –occ. failure included)	55
9.2	LEFT_ABIGRP_C4, RIGHT_ABIGRP_C4 (4-level grouped Ankle Brachial Index (occ. failure incl.))	56
9.3	ABIGRP_C4 (4-level grouped Ankle Brachial Index (occ. failure incl.))	56
9.4	LEFT_ABI2, RIGHT_ABI2, AND ABI2_OVERALL (Ankle Brachial Index - version 2)	56
9.5	ABIGRP2_C4 (4-level grouped Ankle Brachial Index - version 2)	57
10.	LABORATORY MEASURES	58
10.1	INSULIN_FAST (Insulin, fasting (calibrated, converted to mU/L))	58
10.2	INSULIN_OGTT (Insulin, post OGTT (calibrated, converted to mU/L))	58
10.3	INSULIN_DATE (Date of insulin measures - All)	58
10.4	DYSLIPIDEMIA (High LDL, low HDL, or high triglycerides - NIH)	59
10.5	DYSLIPIDEMIA_C3 (3-category dyslipidemia)	59
10.6	CARDIAC_RISK_RATIO (Cardiac Risk Ratio (TC/HDL) - AHA)	59
10.7	DYSLIPIDEMIA_TCHDL (Dyslipidemia based on Cardiac Risk Ratio > 5)	60
10.8	DYS_TCHDL_MED (Dyslipidemia based on Cardiac Risk Ratio > 5 or Lipid Lower Drug scanned at baseline)	60
10.9	HIGH_TOTAL_CHOL (Hypercholesterolemia using lab data and lipid meds)	60
10.10	TOTAL_CHOL_C3 (3-category total cholesterol)	61
10.11	FASTTIME (Fasting time in hours)	61
10.12	IFG (Impaired fasting glucose - AHA)	61
10.13	IGT (Impaired glucose tolerance - AHA)	62
10.14	GFR_Y06 (Glomerular filtration rate estimated, 2006)	62
10.15	GFRSCR_Y06_C5 (Chronic kidney disease using eGFR only - NIDDK)	62

10.16	GFRSCYS_Y08 (eGFR based on serum cystatin C w/o demographics, 2008)	63
10.17	GFRSCR_Y09 (eGFR based on serum creatinine, sex, age, and race, 2009)	63
10.18	GFRSCYS_SCR_Y12 (eGFR based on serum cystatin C, serum creatinine, sex, age and race, 2012)	64
10.19	GFRSCR_Y21 (eGFR using serum creatinine only, 2021)	64
10.20	GFRSCYS_SCR_Y21 (eGFR using serum creatinine and cystatin C, 2021)	65
10.21	GFRSCR_Y21_C5 (5-category eGFR using only serum creatinine, 2021)	66
10.22	GFRSCR_Y21_C2 (Binary eGFR using only serum creatinine, 2021)	66
10.23	GFRSCYS_SCR_Y21_C5 (5-category eGFR using serum creatinine and cystatin C, 2021)	66
10.24	GFRSCYS_SCR_Y21_C2 (Binary eGFR using serum creatinine and cystatin C, 2021)	67
10.25	ALBUMINURIA_C3 (Albuminuria based on 2024 KDIGO guidelines)	67

11. GLOBAL PHYSICAL ACTIVITY QUESTIONNAIRE (GPAQ) 68

11.1	GPAQ_REC_MOD (Recreational physical (moderate) activity (min/day))	68
11.2	GPAQ_REC_VIG (Recreational physical (vigorous) activity (min/day))	68
11.3	GPAQ_REC (Recreational physical activity (min/day))	69
11.4	GPAQ_WORK_MOD (Work-related (moderate) physical activity (min/day))	69
11.5	GPAQ_WORK_VIG (Work-related (vigorous) physical activity (min/day))	69
11.6	GPAQ_WORK (Work-related physical activity (min/day))	69
11.7	GPAQ_TOTAL_MOD (Total physical (moderate) activity (min/day))	70
11.8	GPAQ_TOTAL_VIG (Total physical (vigorous) activity (min/day))	70
11.9	GPAQ_TRSPORT (Transportation-related physical activity (min/day))	70
11.10	GPAQ_TOTAL (Total physical activity (min/day))	71
11.11	GPAQ_TOTAL_MET (Total physical activity (MET-min/day))	71
11.12	GPAQ_MOD_WEEK (GPAQ - Total physical (moderate) activity (min/week))	71
11.13	GPAQ_VIG_WEEK (GPAQ - Total physical (vigorous) activity (min/week))	72
11.15	GPAQ_LEVEL (Physical activity level)	72
11.16	GPAQ_PAG2008 (4-level Activity level per 2008 PA guidelines)	73
11.17	GPAQ_PAG2008YN (Meets 2008 activity level guidelines (1=Yes, 0=No))	74

12. CLINICAL CHARACTERISTICS 75

12.1	HYPERTENSION (Hypertension (BP $\geq$ 140/90 and Med Use)	75
12.2	HYPERTENSION2 (Hypertension using NHANES definition)	75
12.3	HYPERTENSION2_C4 (4-level grouped hypertension)	75
12.4	HYPERTENSION2_AHA (using new ACC/AHA Guidelines definition)	76
12.5	HYPERTENSION2_AHA_C5 (5-level using new ACC/AHA Guidelines)	76
12.6	HYPERT_AWARENESS (Awareness of hypertension - NHANES)	77
12.7	HYPERT_TREATMENT (Treatment of hypertension - NHANES)	77
12.8	HYPERT_CONTROL (Control of hypertension - NHANES)	78
12.9	HOMA_IR (HOMA index of insulin resistance)	78
12.10	HOMA_B (HOMA index of Beta Cell Function)	78
12.11	DIABETES_SELF (Diabetes – self-report only)	78
12.12	DIABETES_LAB (3-level Diabetes – Lab ADA Guideline)	79
12.13	DIABETES_LAB_NOOGTT (3-level – Lab ADA Guideline no OGT)	80
12.14	SELFMED_ANTIDIAB (Anti-Diabetics Medication used – Self-reported)	80
12.15	DIABETES1 (3-level Diabetes, definition 1)	81
12.16	DIABETES2 (3-level Diabetes – ADA Guideline plus scanned med, definition 2)	81
12.17	DIABETES2_INDICATOR (Diabetes indicator (ADA), definition 2)	82
12.18	DIABETES3 (3-level Diabetes – ADA Guideline plus self-reported diagnosis)	82
12.19	DIABETES3A (3-level Diabetes - ADA lab no OGTT plus self-reported diagnosis, definition 3A)	83

12.20	DIABETES_C4 (4-level Diabetes (none/pre/treated/untreated)).....	83
12.21	DIAB_DIAG (Diabetes diagnosis)	84
12.22	DIAB_FAMHIST (Family history of diabetes)	84
12.23	DM_AWARE (Diabetes awareness).....	84
12.24	DM_CONTROL (Diabetes controlled: classified at baseline and A1C < 7%).....	85
12.25	DIABETES4 (3-level Diabetes – ADA guideline plus self-reported med, definition 4) .	85
12.26	DIABETES4A (3-level Diabetes, ADA lab no OGTT plus self-reported med use, definition 4A)	86
12.27	DIABETES5 (3-level Diabetes-ADA Guideline plus self-reported diabetes diagnosis or self-reported med, definition 5)	86
12.28	DIABETES5A (3-level Diabetes - ADA Guideline no OGTT plus self-reported diabetes diagnosis or self-reported Med, definition 5A).....	87
12.29	PRECHD_NO_ANGINA (Prevalent cardiovascular heart disease).....	87
12.30	PRECHD_ANGINA (Prevalent cardiovascular heart disease, including angina)	88
12.31	EVER_ANGINA_RELATIVE (At least one relative had report of angina)	88
12.32	EVER_MI_RELATIVE (At least one relative had report of Heart Attack)	88
12.33	EVER_CABG_RELATIVE (At least one relative had report of CABG or angioplasty).	89
12.34	FH_CHD (Family history of CHD including self-reported angina)	89
12.35	CHD_SELF (Self-reported CHD).....	90
12.36	CDCR (Cerebrovascular disease or carotid revascularization, self-report).....	90
12.37	METS_NCEP (Metabolic syndrome – NCEP-ATP III)	91
12.38	METS_NCEP2 (Metabolic syndrome w/ self-reported meds – NCEP-ATP III)	91
12.39	METS_IDF (Metabolic syndrome – IDF)	92
12.40	MI_ECG (MI defined by ECG results)	93
12.41	STROKE (Prevalent stroke, self-report)	94
12.42	STROKE_TIA (Prevalent stroke or TIA, self-report)	94
12.43	FH_STROKE (Family history of stroke).....	94
12.44	PRE_CVD (Prevalent CVD (CHD or stroke, not counting angina or TIA)).....	95
12.45	CIGARETTE_USE (1=Never,2=Former,3=Current)	95
12.46	CURRENT_SMOKER (Current smoker from CIGARETTE_USE).....	95
12.47	EXPOSURE_YEAR (Adjusted years smoked)	96
12.48	CIGARETTES_YEAR (Adjusted cigarette years)	97
12.49	CIGARETTE_PACK_YEARS (Cigarette pack years)	97
12.50	ALCOHOL_USE (1=Never,2=Former,3=Current)	97
12.51	TOTALDRINKS_PER_WEEK (Total weekly alcohol consumption).....	98
12.52	ALCOHOL_USE_DISORDER (Alcohol use disorder risk).....	98
12.53	ALCOHOL_USE_LEVEL (Alcohol use disorder risk - NIAAA).....	99
12.54	DOCTOR_VISIT (Doctor/HCP number of visits in past 12mo)	99
12.55	CLAUDICATION_INT (Intermittent claudication - either leg)	100
12.56	CVD_FRAME (Framingham CVD composite criterion)	100
12.57	FRAME_CVD_RISK_10YR (Framingham CVD - 10 year risk (w/lab predictors)).....	101
12.58	HBA1C_SI (Glycosylated hemoglobin in SI units (mmol/mol))	104
12.59	CIGARETTE_PACK_YEARS_C3 (3-level Cigarette pack years (cutoff =10))	104
12.60	PREVENT Scores.....	104
12.60.1	PREVENT_BASE_10YR_CVD (10-yr risk probability for cardiovascular disease) ...	105
12.60.2	PREVENT_BASE_30YR_CVD (30-yr risk probability for cardiovascular disease) ...	106
12.60.3	PREVENT_BASE_10YR_ASCVD (10-yr risk probability for atherosclerotic cardiovascular disease)	106
12.60.4	PREVENT_BASE_30YR_ASCVD (30-yr risk probability for atherosclerotic cardiovascular disease)	106
12.60.5	PREVENT_BASE_10YR_HF (10-yr risk probability for heart failure)	106
12.60.6	PREVENT_BASE_30YR_HF (30-yr risk probability for heart failure)	106

12.61	CMR (Cardiometabolic Risk Score)	106
12.62	PCE Score (CVD risk using Pooled Cohort 10-Year ASCVD Risk Assessment Equations) 107	
12.62.1	PCE_SCORE (CVD risk in the next 10 years)	108
12.62.2	PCE (High cardiovascular disease risk in the next 10 years)	109
12.63	PARITY (Number of live births).....	109
12.64	Life's Essential 8 (LE8) Scores	109
12.64.1	LE8_SCAN (Life's Essential 8 Score using scanned meds).....	109
12.64.2	LE8_SELF (Life's Essential 8 Score using reported meds).....	110
12.64.3	LE8_BMI (Life's Essential 8 Body Mass Index component score).....	110
12.64.4	LE8_BP_SCAN (Life's Essential 8 Blood Pressure component score with scanned med) 112	
12.64.5	LE8_BP_SELF (Life's Essential 8 Blood Pressure component score with self-reported med) 112	
12.64.6	LE8_DIET (Life's Essential 8 Diet component score)	113
12.64.7	LE8_GLUCOSE (Life's Essential 8 Blood Glucose component score)	113
12.64.8	LE8_LIPIDS_SCAN (Life's Essential 8 Blood Lipids component score with scanned med) 114	
12.64.9	LE8_LIPIDS_SELF (Life's Essential 8 Blood Lipids component Score with self-reported med) 114	
12.64.10	LE8_NICOTINE (Life's Essential 8 Nicotine Exposure component score).....	115
12.64.11	LE8_PA (Life's Essential 8 Physical Activity component score)	116
12.64.12	LE8_SLEEP (Life's Essential 8 Sleep Score using scanned meds)	116
12.65	Life's Simple 7 (LS7) Scores.....	116
12.65.1	LS7_SCAN (LS7 score using scanned medications).....	117
12.65.2	LS7_SELF (LS7 score using reported medications).....	117
12.65.3	LS7_BMI (LS7 Body Mass Index Score)	117
12.65.4	LS7_BP_SCAN (LS7 Blood Pressure component score using scanned meds).....	118
12.65.5	LS7_BP_SELF (LS7 Blood Pressure Component Score using self-reported meds). 118	
12.65.6	LS7_DIET (LS7 Diet Component Score).....	119
12.65.7	LS7_GLUCOSE_SCAN (LS7 Blood Glucose Score using scanned meds)	121
12.65.8	LS7_GLUCOSE_SELF (LS7 Blood Glucose Score using self-reported meds).....	121
12.65.9	LS7_LIPIDS_SCAN (LS7 Total Cholesterol Score using lipid scanned med)	121
12.65.10	LS7_LIPIDS_SELF (LS7 Total Cholesterol Score using lipid self-reported med)	122
12.65.11	LS7_NICOTINE (LS7 Nicotine Exposure Component Score)	122
12.65.12	LS7_PA (LS7 Physical Activity Score)	123
12.65.13	LS7_SCAN_C3 (3-level LS7 score using scanned meds).....	123
12.65.14	LS7_SELF_C3 (3-level LS7 score using self-reported medications)	124
12.65.15	LS7_SCAN_C2 (Indicator of 4+ LS7 components (ideal) using scanned meds)	124
12.65.16	LS7_SELF_C2 (Indicator of 4+ LS7 components (ideal) using self-reported meds). 125	

13. MEDICATIONS – SCANNED/TRANSCRIBED ONLY 126

13.1	MED_CURRENT_HRT (Any Hormone Repl. Therapy)	127
13.2	MED_ESTROGENS (Estrogens)	127
13.3	MED_ESTROGEN_COMBO (Estrogens – Combinations Only).....	127
13.4	MED_CONTRACEPTIVES (Contraceptives)	127
13.5	MED_PROGESTINS (Progestins)	127
13.6	MED_SERM (Selective Estrogen Receptor Modulators).....	128
13.7	MED_FERTILITY_REG (Fertility Regulators).....	128
13.8	MED_LHRH (LHRH/GnRH Only)	128
13.9	MED_LHRH_AG (LHRH/GnRH Agonist Analog)	128
13.10	MED_LHRH_ANT (LHRH/GnRH Antagonists).....	128

13.11	MED_ANTIANGINAL (Antianginals)	128
13.12	MED_ANTIARRHYTHMICS (Antiarrhythmics).....	129
13.13	MED_ANTIASTHMATICS (Antiasthmatic or bronchodilator agents)	129
13.14	MED_ANTICOAG (Anticoagulants).....	129
13.15	MED_ANTICOAG_COUMARIN (Coumarin anticoagulants).....	129
13.16	MED_ANTICOAG_HEPARIN (Heparin anticoagulants)	129
13.17	MED_ANTIDIAB (Antidiabetics).....	129
13.18	MED_ANTIHYPERT (Antihypertensives)	129
13.19	MED_ANTIHYPERT_ACEI (ACE Inhibitors)	129
13.20	MED_ANTIHYPERT_AT2RAS (Angiotensin II Receptor Antagonists)	130
13.21	MED_ANTIPLATELET (Antiplatelets)	130
13.22	MED_ASPIRIN (Aspirin-containing analgesics)	130
13.23	MED_BB (Beta Blockers).....	130
13.24	MED_BB_Ophthalmic (Ophthalmic Beta Blockers)	131
13.25	MED_CACB (Calcium Channel Blockers)	131
13.26	MED_CARDIACGLYCOSIDES (Cardiac Glycosides)	131
13.27	MED_CHEMO (Chemotherapy).....	131
13.28	MED_CLOPID (Clopidogrel (Plavix)).....	131
13.29	MED_COX2 (COX-2 inhibitors).....	131
13.30	MED_DIURETIC (Diuretics).....	132
13.31	MED_DIURETIC_THIAZIDE (Thiazide Diuretics).....	132
13.32	MED_FIBARES_NICOACID (Fibric/Nicotinic Acids (trt of TG and HDL))	132
13.33	MED_INSULIN (Insulins)	132
13.34	MED_LLD (Lipid lowering drugs/Antihyperlipidemics (LLD))	132
13.35	MED_METFORMIN (Metformins)	133
13.36	MED_NSAID (NSAIDs).....	133
13.37	MED_OI_STEROID (Oral/inhalable Corticosteroids).....	133
13.38	MED_STATIN (Statins).....	133
13.39	MED_ANTIAXXI (Antianxiety agents)	133
13.40	MED_ANTIDEPRESS (Antidepressants)	134
13.41	MED_ANTIDEPRE_SSRI (Antidepressants - selective serotonin re-uptake inhibitors).....	134
13.42	MED_ANTIDEPRE_TRICYC (Antidepressants - tricyclic antidepressants)	134
13.43	MED_ANTIPSYCHO (Antipsychotics).....	134
13.44	MED_CORTICOSTEROIDS_INHALED (Inhaled Corticosteroid medication)	134
13.45	MED_PDE4 (Phosphodiesterase IV Inhibitor Medication)	134
13.46	MED_BRONCHODILATOR (Bronchodilator and Antiasthmatic Agents medication)	135
13.47	MED_PULMONARY (Pulmonary medication).....	135
14.	MEDICATIONS – OTHER	137
14.1	ASTHMA_MEDS (Asthma Med Use – self/scanned (curr asthma by MD diag)).....	137
14.2	COPD_MEDS (COPD Med Use – self/scanned (ever COPD by MD diag)).....	137
15.	NEUROCOGNITIVE	138
15.1	ECG_MAJ (Major ECG Abnormalities).....	138
15.2	ECG_MIN (Minor ECG Abnormalities)	138
15.3	NO_DISABILITY_6ITEM (No Disability on NEEA5-7 (1=Yes,0=No))	140
15.4	SEVLT_3TRIALS (SEVLT, Sum of 3 Trials (NEEA8-10)).....	140
15.5	SEVLT_RECALL (SEVLT, Delay Recall (NEEA11 with re-coding))	140
15.6	TOTAL_6ITEM (Total Score on 6 Item Screener (with re-coding))	140
15.7	WORDFREQ (Word Frequency (A+F) (NEEA12-13))	142
15.8	DIGITSYMBOL (Digit Symbol Substitution Score (NEEA14)).....	142

16.	DENTAL	143
16.1	PERIODONTAL_DISEASE (Periodontal Disease - NIDCR).....	143
16.2	PERIO_CDC_AAP (Periodontitis Severity using CDC/AAP Grades).....	144
16.3	PERIO_EXTENT (Percentage Interproximal Sites Positive for Periodontitis)	145
17.	SLEEP	146
17.1	ESS (Epworth Sleepiness Scale).....	146
17.2	ESS_GT10 (Excessive sleepiness (ESS > 10))	146
17.3	AHI_GE15 (Sleep Apnea defined as Apnea/Hypopnea Index (3% desat) ≥ 15).....	146
17.4	SBD5 (Sleep Apnea Syndrome (Apnea/Hypopnea Index (3% desat) ≥ 5 & ESS ≥ 10)) 147	147
17.5	SLPDUR_WKDAY (Weekday sleep duration (hours)).....	147
17.6	SLPDUR_WKEND (Weekend sleep duration (hours))	147
17.7	SLPDUR (Average sleep duration (hours))	148
17.8	MID_SLEEP_TIME_WKDAY (Midpoint sleep time in clock hours; weekdays)	148
17.9	MID_SLEEP_TIME_WKEND (Midpoint sleep time in clock hours; weekend).....	148
17.10	WHIIRS (Women's Health Initiative Insomnia Rating Scale)	149
18.	DIET	150
18.1	DIET_SCORE_JAMA (Diet Score from 1st JAMA paper).....	150
18.2	DIET_SCORE_JAMA_C2 (Categorized Diet Score JAMA (1=Below, 2=Above 60th %- tile))	151
18.3	AHEI2010 (Alternative Healthy Eating Index 2010).....	151
18.4	HEI2010 (Healthy Eating Index 2010).....	152
18.5	DASH (Version 1 – 2003 Eating Plan).....	153
18.6	DASH (Version 2 – 2018 Eating Plan).....	154
18.7	SUPPL_USE (Supplement use in the last 30 days)	155
19.	OTHER VARIABLES	155
19.1	SDI_SCORE. Zip code level neighborhood social deprivation	155
	VARIABLE MODIFICATIONS	156
	APPENDIX	161

Summary of updates in April 2026 data release

New Derived Variables

- ALBUMINURIA_C3
- ADDRESS_YRS (number of years at baseline address)
- AGE_IMMI_C3
- BKGRD1_C7_NOMISS (88 participants with missing background coded as Other/Mixed)
- CMR (cardiometabolic risk score)
- **Diabetes without OGTT:**
 - o DIABETES_LAB_NOOGTT
 - o DIABETES3A – ADA lab no OGTT & self-reported diagnosis
 - o DIABETES4A – ADA lab no OGTT & self-reported med use
 - o DIABETES5A – ADA lab no OGTT & self-reported diagnosis or self-reported meds
- **Estimated Glomerular filtration rate using 2021 equation without race:**
GFRSCR_Y21, GFRSCYS_SCR_Y21, GFRSCR_Y21_C5, GFRSCYS_SCR_Y21_C5, GFRSCR_Y21_C2, GFRSCYS_SCR_Y21_C2
- HC (current health insurance excluding Indian Health Services)
- HH_ID (household identifier)
- **Life Simple 7:** LS7_BMI, LS7_BP_SCAN, LS7_BP_SELF, LS7_DIET, LS7_GLUCOSE_SCAN, LS7_GLUCOSE_SELF, LS7_LIPIDS_SCAN, LS7_LIPIDS_SELF, LS7_NICOTINE_LS7_PA, LS7_SCAN, LS7_SCAN_C2, LS7_SCAN_C3, LS7_SELF, LS7_SELF_C2, LS7_SELF_C3
- **Life Essential 8:** LE8_SCAN and LE8_SELF and its components (LE8_DIET, LE8_PA, LE8_NICOTINE, LE8_SLEEP, LE8_BMI, LE8_LIPIDS_SELF, LE8_LIPIDS_SCAN, LE8_GLUCOSE, LE8_BP_SELF, LE8_BP_SCAN)
- PARITY (number of live births by baseline)
- **Pooled Cohort 10-Year Atherosclerotic CVD Risk Assessment Equation:** PCE
- **PREVENT risk equations.** AHA Predicting Risk of cardiovascular disease and two subtypes (atherosclerotic and heart failure):
 - o PREVENT_BASE_10YR_CVD, PREVENT_BASE_30YR_CVD
 - o PREVENT_BASE_10YR_ASCVD, PREVENT_BASE_30YR_ASCVD
 - o PREVENT_BASE_10YR_HF, PREVENT_BASE_30YR_HF
- SDI_SCORE (neighborhood social deprivation from zip code level)
- WEIGHT_NONRESP (non-response adjusted sampling weight needed for IPW)

Corrected Variables

- CIGARETTE_USE. Corrected values that were missing for 8 participants.
- EDUCATION_C4. Corrected values that were assigned level 3 (some college) that should have been level 4 (University/College).
- NP_PERMIT (version B of ICT form was not being used)
- SUPPL_USE. Set to missing if did not answer NDSR module for 30-day dietary supplement

Variable Name Changed

- GENDERNUM → SEX
- HYPERTENSION_C4 → HYPERTENSION2_C4
- HYPERTENSION_AHA_C5 → HYPERTENSION2_AHA_C5
- CKD → GFRSCR_Y06_C5
- GFR → GFR_Y06
- GFRSCYS → GFRSCYS_Y08
- GFRSCYS_SCR → GFRSCYS_SCR_Y12
- GFRSCR → GFRSCR_Y09

Variables Dropped

- GENDER (“F” or “M”) which is equivalent to numeric variable SEX
- MED_PULMONARY (definition needs to be revised)
- N_HC (current health insurance including Indian Health Services)
- POVPCT, POVPCT_C4

Other

Except for the sampling design variables (sampling weight, strata, and PSU), data from one participant was removed per their request in summer 2025. Design variables were retained solely to preserve the HCHS/SOL multi-stage sampling design. The study will continue to report there were 16,415 participants enrolled at baseline, and data from this participant is treated as missing data.

1. DESIGN

1.1 COHORT (Recruitment year from sample address list)

Wave of household sample address allocation and recruitment.

Cohort= 1, 2, 3 (year 1, year 2, or year 3 of the sample address recruitment cohorts)

1.2 PSU_ID (Primary sampling unit ID clustering variable- masked)

Sample design clustering variable used in statistical analyses as the CLUSTER variable (in SAS) that is a combination of the following:

Field Center and selected block group identifier.

1.3 STRAT (Sampling design stratification variable - masked)

Sample design stratification variable used in statistical analyses as the STRATUM variable (in SAS) that is a combination of the following:

Field Center, Hispanic Household Proportion (high, low), and SES (high, low).

(Hispanic Household Proportion and SES calculated using 2000 Census Data for selected block groups within each field center)

1.4 LISTNUM (List address record assigned)

Sample design variable classifying addresses as originating from Hispanic-surname versus non-Hispanic-surname lists within the original Genesys sampling frame used for recruitment.

Response format: 1 = Non-Hispanic Surname Address List
2 = Hispanic Surname Address List

1.5 PSU_CNT (BG-level Hispanic household frame counts)

Sample design clustering household frame count variable used in statistical analyses.

1.6 STRAT_CNT (Stratum-level Hispanic household frame count)

Sample design stratification household frame variable used in statistical analyses.

1.7 LIST_CNT (Sample list - level Hispanic)

Sample design List household frame count variable used in statistical analyses.

1.8 HH_ID (Household ID)

Household identifier.

For the following Sample Weight variable(s): Please see HCHS/SOL Manuscript “Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos” (Annals of Epidemiology, Volume 20, Issue 8, August 2010, Pages 629-641) for information on how the sample weights were created. Also see the HCHS/SOL Analytic Methods document that describes the proper use of sample weights in statistical analyses.

1.9 WEIGHT_FINAL_EXPANDED (Expanded Final Weight)

The calibrated sample weight is the calibrated, trimmed, non-response adjusted reciprocal of a participant’s probability of selection into the HCHS/SOL study. The trimmed, non-response adjusted sample weights were calibrated using first the 7-category Hispanic Background distribution and then the 12-category age group/sex distribution using the 2010 Decennial Census. This variable is used for analyses purposes only.

1.10 WEIGHT_FINAL_NORM_OVERALL (Overall Normalized Final Weight)

The normalized sample weight is the overall normalized, calibrated, trimmed, non-response adjusted (household and person-level) reciprocal of a participant’s probability of selection into the HCHS/SOL study. The base sample weights, i.e. the reciprocal of a participant’s probability of selection into the HCHS/SOL study, were first adjusted by household-level and person-level non-response. The non-response adjusted sample weights were then trimmed using the selected average percentile that minimizes model MSE within each field center. All sample weights that fell above the selected percentile were trimmed to the selected percentile and the trimmed difference was evenly distributed among all of the non-trimmed sample weights. The trimmed, non-response adjusted sample weights were then calibrated using first the 7-category Hispanic Background distribution and then the 12-category age group/sex distribution using the 2010 Decennial Census. Finally, these calibrated, trimmed, non-response adjusted sample weights were normalized using the weighted mean divided by weighted standard deviation of the entire HCHS/SOL sample to normalize the expanded calibrated sample weights (WEIGHT_FINAL_NORM_OVERALL). This final sample weight variable is used for analyses purposes only.

1.11 WEIGHT_FINAL_NORM_CENTER (FC-Specific Normalized Final Weight)

The FC-specific normalized sample weight is the field center specific normalized, calibrated, trimmed, non-response adjusted (household and person-level) reciprocal of a participant’s probability of selection into the HCHS/SOL study. The calibrated, trimmed, non-response adjusted sample weights were normalized using the weighted mean divided by weighted standard deviation calculated for each field center separately and used to normalize the expanded calibrated sample weights (WEIGHT_FINAL_NORM_CENTER). This final sample weight variable is used for analyses purposes only.

1.12 WEIGHT_NONRESP (Non-response adjusted sampling weight)

The non-response adjusted sampling weight is the base sample weight (i.e. the reciprocal of a participant’s probability of selection into the HCHS/SOL study) only adjusted by household-level and person-level non-response. It should only be used to build IPW sampling weights.

2. ADMINISTRATIVE

2.1 CENTER (Participant's field center)

This is a character variable with four possible values derived from the field center of origin (baseline recruitment): "B"= Bronx, "C"= Chicago, "M"= Miami, "S"= San Diego. Center cannot have missing values because each valid participant ID has an affiliated field center.

Source variable(s):

SUBJECTID. Subject ID

2.2 CENTERNUM – Numeric (Participant's field center)

This is a numeric variable with four possible values derived from the field center of origin: 1= Bronx, 2= Chicago, 3= Miami, 4= San Diego.

Source variable(s):

CENTER. Center (Participant's field center)

2.3 CLINDATE (Date of the participant's clinic visit)

This is a SAS date variable which documents the date of the participant's clinic visit. It is first derived from the Informed Consent tracking form (ICT). If the form or the date of form completion is not present, then it is derived from the Personal Information (PIE), Anthropometry (ANT), or Sitting Blood Pressure (SBP) forms.

CLINDATE= ICTA0A or ICTB0A

or

Minimum (PIEA0A, SBPA0A, ANTA0A) if ICTA0A/ICTB0A is missing

Source variable(s):

PIEA0A. Date of completion of the Personal Information Form

ICTA0A or ICTB0A. Date of completion of the Informed Consent Form (version A or B)

SBPA0A. Date of completion of the Sitting Blood Pressure Form

ANTA0A. Date of completion of the Anthropometry Form

2.4 CONSENT (Informed consent status for participation in study)

This is a binary variable that determines whether or not a study participant consented to participate in the baseline examination study.

If ICTA1=1 or ICTB1=1 then consent=1;

Else consent=0;

Source variable(s):

ICTA1 or ICTB1. Agrees to participate in HCHS/SOL study (yes/no)

2.5 GENOTYPING (Full consent for genotyping)

This is a binary variable that determines whether or not a study participant fully consented to genotyping for use with investigators associated or not associated to HCHS/SOL and specialized laboratories (Note: does NOT reflect consent for use by commercial or for-profit companies that are not part of HCHS/SOL).

If ((ictb4=1 and ictb7=1 and ictb8=1) or (icta4=1 and icta6=1)) then genotyping=1.
Otherwise genotyping=0.

Source variable(s):

ICTA1 or ICTB1. Agrees to participate in HCHS/SOL.

ICTA4 or ICTB4. Agrees to provide a blood sample for DNA and RNA that can be extracted, stored and used

ICTA6. Agrees to share “de-identified” blood, genetic material (DNA/RNA) and other information with HCHS/SOL researchers and their collaborators

ICTB7. Agrees to share non-genetic data, information, and samples available to investigators not associated to HCHS/SOL and specialized laboratories

ICTB8. Agrees to share genetic data, information, and samples available to investigators not associated to HCHS/SOL and specialized laboratories

2.6 DNA_AVAILABLE (DNA available)

This is a binary variable that determines whether or not a study participant had blood draws successfully conducted without incidents or problems for tubes 4 and 5 to make DNA available to investigators associated or not associated to HCHS/SOL and specialized laboratories (Note: does NOT reflect consent for use on DNA samples, i.e. genotyping).

((bioa14a4^=1 & bioa14b4^=1 & bioa14c4^=1 & bioa14d4^=1 & bioa14e4^=1 &
bioa14f4^=1 &
bioa22a4^=1 & bioa22b4^=1 & bioa22c4^=1 & bioa22d4^=1 & bioa22e4^=1) |
(bioa14a5^=1 & bioa14b5^=1 & bioa14c5^=1 & bioa14d5^=1 & bioa14e5^=1 &
bioa14f5^=1 &
bioa22a5^=1 & bioa22b5^=1 & bioa22c5^=1 & bioa22d5^=1 & bioa22e5^=1))

then let DNA_Available=1

Otherwise DNA_Available=0

Source variable(s):

BIOA14A4. Tube 4 – Sample not drawn

BIOA14B4. Tube 4 – Partial sample drawn

BIOA14C4. Tube 4 – Tourniquet reapplied

BIOA14D4. Tube 4 – Fist clenching

BIOA14E4. Tube 4 – Needle movement

BIOA14F4. Tube 4 – Participant reclining

BIOA14A5. Tube 5 – Sample not drawn

BIOA14B5. Tube 5 – Partial sample drawn

BIOA14C5. Tube 5 – Tourniquet reapplied

BIOA14D5. Tube 5 – Fist clenching

BIOA14E5. Tube 5 – Needle movement

BIOA14F5. Tube 5 – Participant reclining
BIOA22A4. Tube 4 – Broken Tube
BIOA22B4. Tube 4 – Sample re-centrifuged
BIOA22C4. Tube 4 – Clotted
BIOA22D4. Tube 4 – Hemolyzed
BIOA22E4. Tube 4 – Lipemic
BIOA22A5. Tube 4 – Broken Tube
BIOA22B5. Tube 4 – Sample re-centrifuged
BIOA22C5. Tube 4 – Clotted
BIOA22D5. Tube 4 – Hemolyzed
BIOA22E5. Tube 4 – Lipemic

2.7 NP_PERMIT (Permission for non-profit use of data)

This is a binary variable that determines whether or not a study participant gave permission for non-profit use of study data.

ICTA7 = 1 or ICTB9 = 1 → 1

If both ICTA7 and ICTB9 are missing → missing
otherwise 0

Response Format: 0 = use not permitted
1 = permitted

Source variable(s):

ICTA7. Agrees to share de-identified blood, genetic material and other information with nonprofit private companies

ICTB9. Agrees to share with commercial or for-profit companies that are not part of HCHS/SOL.

2.8 EXTERNAL_PERMIT (Permission for investigators outside HCHS/SOL)

This is a binary variable that determines whether or not a study participant gave permission for use for investigators outside HCHS/SOL.

ICTA6 = 1 or ICTB7 = 1 → 1

otherwise 0

Response Format: 0 = use not permitted
1 = permitted

Source variable(s):

ICTA6. Agrees to share de-identified blood, genetic material and other information with HCHS/SOL researchers and other individuals

ICTB7. Agrees to share non-genetic data to investigators not associated to HCHS/SOL and specialized laboratories

2.9 COMMERCIAL_PERMIT (Permission for for-profit use of data)

This is a binary variable that determines whether or not a study participant gave permission for use for for-profit use of study data. Note that for-profit permission was not collected in ICT version A.

ICTB9 = 1 → 1
otherwise 0

Response Format: 0 = use not permitted
1 = permitted

Source variable(s):

ICTB9. Agrees to share genetic and non-genetic data with commercial or for-profit companies

2.10 ANY_PERMIT (Any permission for public use data)

This is a binary variable that combines the 3 permission variables to determine whether or not a study participant gave any permission for public use of data.

NP_Permit=1 or External_Permit=1 or Commercial_Permit=1 → 1
otherwise 0

Response Format: 0 = use not permitted
1 = permitted

Source variable(s):

NP_Permit. Permission for non-profit use of data

External_Permit. Permission for investigators outside HCHS/SOL

Commercial_Permit. Permission for for-profit use of data

3. SOCIO-DEMOGRAPHIC

3.1 AGE (Age in years at the time of participant's clinic visit)

This is the age of the participant in years (an integer variable) at the time of the participant's clinic visit. It is determined from the participant's date of birth and the clinic visit date.

AGE = INTEGER of (PIEA2– PIEA0A)/365.25

When missing use age collected as recruitment:

ELEA2 and if this missing, then use

HSRA5A4-HSRA5N4 or HSRB4A4-HSRB4J4 based on linkage in ELEA of participant to roster row of household screening form

Source variable(s):

PIEA0A. Date of completion of the Personal Information Form

PIEA2. Date of birth

ELEA2. Age of person screened

HSRA5A4-5N4. Age of person rostered from household (rows A-N)

HSRB4A4-4J4. Age of person rostered from household (rows A-J)

3.2 AGEGROUP_C2 (2-level grouped age of participant)

This is the categorical (grouped) age of the participant. It is determined from the derived variable AGE.

Response Format: 1=Ages 18-44

2=Ages 45+

Source variable(s):

AGE. Age of participant at the time of their clinic visit

3.3 AGEGROUP_C3 (3-level grouped age of participant)

This is the categorical (grouped) age of the participant. It is determined from the derived variable AGE.

Response Format: 1=Ages 18-44

2=Ages 45-64

3=Ages 65+

Source variable(s):

AGE. Age of participant at the time of their clinic visit

3.4 AGEGROUP_C6 (6-level grouped age of participant)

This is the categorical (grouped) age of the participant. It is determined from the derived variable AGE.

Response Format: 1=Ages 18-24
 2=Ages 25-34
 3=Ages 35-44
 4=Ages 45-54
 5=Ages 55-64
 6=Ages 65+

Source variable(s):

AGE. age of participant at the time of their clinic visit

3.5 AGEGROUP_C6_NHANES (6-level grouped age – NHANES standardization)

This is the categorical (grouped) age of the participant. It is determined from the derived variable AGE for use with analyses requiring NHANES age standardization (Note: Does not include participants 75+ years old at baseline).

Response Format: 1=Ages 18-29
 2=Ages 30-39
 3=Ages 40-49
 4=Ages 50-59
 5=Ages 60-69
 6=Ages 70-74

Source variable(s):

AGE. Age of participant at the time of their clinic visit

3.6 SEX (Sex – Numeric)

This is a 0/1 variable which describes the participant's sex, female (0) or male (1).

When missing use sex collected at recruitment:

HSRA5A3-HSRA5N3 or HSRB4A3-HSRB4J3 based on linkage in ELEA of participant to roster row of household screening form

Response Format: 0=Female
 1=Male

Source variable(s):

PIEA1. labeled "gender" in the form but meant to be "sex". Gender was NOT asked to the participant, nor sex. Instead, staff assumed the participant's sex.

HSRA5A3-5N3 – Sex of person rostered from household (rows A-N)

HSRB4A3-4J3 – Sex of person rostered from household (rows A-J)

3.7 EDUCATION_C2 (2-level group education level)

The educational status derived variable groups the attainment of a high school diploma or an equivalent degree, or not having that degree. Update 12/2013: categorized missing values based on PIEA21B-PIEA21G, PIEA19B, PIEA21I, and PIEA17.

Ordered categorization for Response Format

1. Highest level of elementary, middle school or high school (PIEA19A=1 or 2 or 3) AND no high school diploma or GED (PIEA21A=0 or missing) → 1
2. [Highest level of elementary, middle school or high school (PIEA19A=1 or 2 or 3) AND high school diploma or GED (PIEA21A=1)] OR Highest level of trade school/vocational school or university/college (PIEA19A=4 or 5) → 2
3. 0 years of education (PIEA17=0) → 1
4. Trade school certificate, associate degree, bachelor's degree, etc. (Any of PIEA21B through PIEA21G=1) → 2
5. Text response for other level indicating No high school diploma or GED (PIEA19B) → 1
6. Text response for other level indicating At least High school diploma or GED (PIEA19B) → 2
7. Text response for other degree indicating At least High school diploma or GED (PIEA21I) → 2
8. 1-10 years of education ($1 \leq \text{PIEA17} \leq 10$) → 1
9. 11+ years of education ($\text{PIEA17} \geq 11$) → 2
10. Otherwise missing

Response format: 1 = No high school diploma or GED
2 = High school diploma/GED or greater

Source variable(s):

PIEA17. How many years of schooling in total have you completed?

PIEA19A. Highest (or closest) grade/level of education achieved (6 levels)

PIEA19B. If other, please specify:

PIEA21A. High school diploma or equivalent (yes/no)

PIEA21B. Trade school/vocational school certificate

PIEA21C. Associate degree (i.e. AA, AS)

PIEA21D. Bachelor's degree (i.e. BA, AB, BS)

PIEA21E. Master's degree (i.e. MA, MS, MEd, MSW, MBA)

PIEA21F. Professional degree (i.e. MD, DDS, DVM, LLB, JD)

PIEA21G. Doctorate degree (i.e. PhD, EdD)

PIEA21I. If other, please specify:

3.8 EDUCATION_C3 (3-level group education level)

The educational status derived variable groups the attainment of a high school diploma or an equivalent degree, not having that high school or equivalent degree, or attainment of education beyond a high school equivalent (i.e. college or vocational). Update 12/2013: categorized missing values based on PIEA21B-PIEA21G, PIEA19B, PIEA21I, and PIEA17.

Ordered categorization

1. Highest level of elementary, middle school or high school (PIEA19A=1 or 2 or 3) AND no high school diploma or GED (PIEA21A=0 or missing) → 1
2. Highest level of elementary, middle school or high school (PIEA19A=1 or 2 or 3) AND high school diploma or GED (PIEA21A=1) → 2
3. Highest level of trade school/vocational school or university/college (PIEA19A=4 or 5) → 3
4. 0 years of education (PIEA17=0) → 1
5. Trade school certificate, associate degree, bachelor's degree, etc. (Any of PIEA21B through PIEA21G=1) → 3
6. Text response for other level indicating No high school diploma or GED (PIEA19B) → 1
7. Text response for other level indicating At most a High school diploma or GED (PIEA19B) → 2
8. Text response for other level indicating Greater than high school (or GED) education (PIEA19B) → 3
9. Text response for other degree indicating At most a High school diploma or GED (PIEA21I) → 2
10. Text response for other degree indicating Greater than high school (or GED) education (PIEA21I) → 3
11. 1-10 years of education ($1 \leq \text{PIEA17} \leq 10$) → 1
12. 11-12 years of education ($11 \leq \text{PIEA17} \leq 12$) → 2
13. 13+ years of education ($\text{PIEA17} \geq 13$) → 3
14. Otherwise missing

Response format: 1 = No high school diploma or GED
2 = At most a High school diploma or GED
3 = Greater than high school (or GED) education

Source variable(s):

PIEA17. How many years of schooling in total have you completed?
PIEA19A. Highest (or closest) grade/level of education achieved (6 levels)
PIEA19B. If other, please specify:
PIEA21A. High school diploma or equivalent (yes/no)
PIEA21B. Trade school/vocational school certificate
PIEA21C. Associate degree (i.e. AA, AS)
PIEA21D. Bachelor's degree (i.e. BA, AB, BS)

PIEA21E. Master's degree (i.e. MA, MS, MEd, MSW, MBA)
 PIEA21F. Professional degree (i.e. MD, DDS, DVM, LLB, JD)
 PIEA21G. Doctorate degree (i.e. PhD, EdD)
 PIEA21I. If other, please specify:

3.9 EDUCATION_C4 (4-level group education level)

Based on Education_C3, Education_C4 was created to differentiate between some colleges vs MORE than college.

If (Education_C3=3 and PIEA19a=5) OR (PIEA21c=1 or PIEA21d=1 or PIEA21e=1 or PIEA21f=1 or PIEA21g=1) then set Education_C4=4.
 Otherwise Education_C4=Education_C3

Response format: 1 = No high school diploma or GED
 2 = At most a High school diploma or GED
 3 = Greater than High school (or GED) education with some college
 4 = University/college education

Source variable(s):

Education_C3. 3-level group education level.
 PIEA19A. Highest (or closest) grade/level of education achieved (6 levels)
 PIEA21C. Associate degree (i.e. AA, AS)
 PIEA21D. Bachelor's degree (i.e. BA, AB, BS)
 PIEA21E. Master's degree (i.e. MA, MS, MEd, MSW, MBA)
 PIEA21F. Professional degree (i.e. MD, DDS, DVM, LLB, JD)
 PIEA21G. Doctorate degree (i.e. PhD, EdD)

3.10 INCOME (Grouped yearly household income)

This derived variable groups each subject to one income category. The income is set to missing if both ECEA3 and ECEA4 are missing. Income is set to ECEA3 if non-missing. Otherwise income is set to ECEA4.

Income	ECEA3	ECEA4
1	1	.
2	2	.
3	3	.
4	4	.
5	5	.
6	.	1
7	.	2
8	.	3
9	.	4
10	.	5
.	.	.

Response format: 1 = Less than \$10,000
2 = \$10,001-\$15,000
3 = \$15,001-\$20,000
4 = \$20,001-\$25,000
5 = \$25,001-\$29,999
6 = \$30,000-\$40,000
7 = \$40,001-\$50,000
8 = \$50,001-\$75,000
9 = \$75,001-\$100,000
10 = More than \$100,000

Source variable(s):

ECEA3. Low income range

ECEA4. High income range

3.11 INCOME_C3 (3-level grouped yearly household income)

This derived variable groups each subject to one income category based exclusively on ECEA2. ECEA2 was answered by some participants who refused to answer the follow up questions ECEA3 and ECEA4. Participants with missing ECEA2 grouped together.

ECEA2=1 → 1

ECEA2=2 → 2

ECEA2 is missing or Don't know/refused → 3

Response format: 1=Less than \$30,000
2=\$30,000 or more
3=Missing

Source variable(s):

ECEA2. Your household income for the year

3.12 INCOME_C5 (5-level grouped yearly household income)

This derived variable groups each subject to one income category. The income is set to missing if both ECEA3 and ECEA4 are missing. Income_C5 is set to ECEA3 if non-missing. Otherwise income is set to ECEA4.

```
if ecea3 in(1) then INCOME_C5=1;
else if ecea3 in(2,3) then INCOME_C5=2;
else if ecea3 in(4,5) | ecea4=1 then INCOME_C5=3;
else if ecea4 in(2,3) then INCOME_C5=4;
else if ecea4 in(4,5) then INCOME_C5=5;
else INCOME_C5=.;
```

Response format: 1 = Less than \$10,000
2 = \$10,001-\$20,000
3 = \$20,001-\$40,000
4 = \$40,001-\$75,000
5 = More than \$75,000

Source variable(s):

ECEA3. Low income range
ECEA4. High income range

3.13 MARITAL_STATUS (3-level grouped marital status)

This derived variable groups each subject's marital status.

if piea3=1 then MARITAL_STATUS=1;
else if piea3 in (2,6) then MARITAL_STATUS=2;
else if piea3 in (3,4,5) then MARITAL_STATUS=3;
else MARITAL_STATUS=.;

Response format: 1=Single
2=Married or living with a partner
3=Separated, divorced, or widow(er)

Source variable(s):

PIEA3. Marital Status

3.14 EMPLOYED (4-level grouped employment status)

This variable groups the employment status of the participant. This variable includes retirees who are or are not employed. Retired participants missing employment are categorized as not currently employed, while unemployed participants missing data on retirement are categorized as not retired.

if (oce1=1 & (oce5=3 | oce5≤.Z)) then EMPLOYED=1;
else if (oce1^=1 & oce5=3) then EMPLOYED=2;
else if (oce5=2) then EMPLOYED=3;
else if (oce5=1) then EMPLOYED=4;
else EMPLOYED=.;

Response Format: 1=Retired and not currently employed (or missing on employment).
2=Not retired (or missing on retirement) and not currently employed
3=Employed part-time (≤35 hours/week)
4=Employed full-time (>35 hours/week)

Source variable(s):

OCEA1. Retired (yes/no)
OCEA5. Current employment status

3.15 HC (Current Health Insurance Coverage Excluding Indian Health Services)

This variable indicates the participant currently has health insurance coverage excluding Indian Health Services. The definition also checks whether any reason was given for stopping coverage. Note that there are two versions of the HCE form (HCEA and HCEAB) due to some wording changes, clearer skip instructions and question 10 was slightly restructured: HCEA (n=10,751) was administered from 3/4/2008 to 6/29/2011 whereas HCEB (n=5,362) was administered from 1/5/2010 to 6/30/2011. Participants only completed it once; in other words, they have one form or the other.

If vers='A' then

HI_IND = sum(HCEA10B, HCEA10C, HCEA10D, HCEA10E, HCEA10F, HCEA10H);

else if vers = 'B' then

HI_IND = sum(HCEB10B, HCEB10C, HCEB10D, HCEB10E, HCEB10F, HCEB10H);

STOP_IND=sum(HCEA12a,HCEA12b,HCEA12c,HCEA12d,HCEA12e,HCEA12f,HCEA12g,HCEA12h,HCEA12i);

If vers = 'A' then

do;

 If (missing(hcea10a) & hi_ind > 0) |

 (hcea10a=0 and not (HI_IND=0 and hcea10g = 1)) then HC = 1;

 else if (missing(hcea10a) & stop_ind>0) |

 (missing(hcea10a) & HCEA11=5) |

 (hcea10a = 1) |

 (hcea10a = 0 and hi_ind = 0 and hcea10g = 1) then HC = 0;

 else if hcea10a = 1 then HC = 0;

end;

else if vers = 'B' then

do;

 If hceb10a = 1 and not (hi_ind=0 and hceb10g=1) then HC = 1;

 else if hceb10a = 1 and hi_ind=0 and hceb10g=1 then HC = 0;

 else if hceb10a = 0 then HC = 0;

end;

Response format: 0 = No current health insurance

1 = Yes, I currently have health insurance

Source variable(s):

Form A:

HCEA10a. None, no insurance and currently not covered (HCEA10a=1 means DO NOT have insurance and currently not covered)

HCEA10b-HCEA10h. What type of health insurance or health care coverage do you have?

HCEA11. About how long has it been since you last had health insurance coverage?

(5=never had insurance)

HCEA12a-HCEA12k. Reasons you stopped being covered by health insurance.

Form B:

HCEB10a. Do you have health insurance or other health care coverage?

HCEB10b-HCEB10h. What type of health insurance or health care coverage do you have?

3.16 OCCUPATION_CURR (Grouped current occupation for employed participants)

For participants who are currently employed (part or full time), this is the type of occupation the participant had at the baseline visit. Occupational codes used in the HCHS/SOL are not based on any standard Bureau of Labor Statistics or Census occupation codes. These codes were developed by HCHS/SOL to reflect common occupational groups in the Hispanic/Latino population and were based on NIH-funded studies in developing countries. These occupational groupings were derived primarily from the Children of Immigrants Longitudinal Survey (Portes, Alejandro, and Rubén G. Rumbaut. Legacies: The story of the immigrant second generation. Univ of California Press, 2001).

oce16a=7 → 1
oce16a=11 → 2
oce16a=6 → 3
1≤oce16a≤4 → 4;
other nonmissing values → 5

Response format: 1=Non-skilled worker
2=Service worker
3=Skilled worker
4=Professional/technical, administrative/executive, or office staff
5=Other occupation

Source variable(s):

OCEA16a. Type of job currently work majority of hours per week.

3.17 OCCUPATION_LONG (Grouped longest held occupation)

This is the type of occupation the participant has held the longest.

oce28a=7 → 1
oce28a=11 → 2
oce28a=6 → 3
1≤oce28a≤4 → 4;
other nonmissing values → 5

Response format: 1=Non-skilled worker
2=Service worker
3=Skilled worker
4=Professional/technical, administrative/executive, or office staff
5=Other occupation

Source variable(s):

OCEA28a. Type of job held the longest.

3.18 ADDRESS_YRS (Years lived at baseline address)

ADDRESS_YRS = year(CLINDATE) – IDEA3k1

if ADDRESS_YRS < 0 or ADDRESS_YRS > 90 then ADDRESS_YRS = missing

Source variable(s):

CLINDATE. Date of clinic visit and

IDEA3k1. Since what year have you lived at this address?

4. ACCULTURATION

4.1 BKGRD1 (Re-classification of Hispanic/Latino Background)

The self-identification with a Hispanic/Latino cultural group allowed for selection of an 'other' designation. This variable groups the text responses of 7=Other (using PIEA5B) for variations of 'Dominican' so that undercounts do not occur for that group. In addition, several 7=Other responses are re-categorized into one of the 7 groups as defined below.

Response format:

- 0 = Dominican
- 1 = Central American
- 2 = Cuban
- 3 = Mexican
- 4 = Puerto Rican
- 5 = South American
- 6 = More than one heritage
- 7 = Other

Where 7=Other, the responses to PIEA5B below have BKGRD1 redefined as follows (as of 9/30/2013):

BKGRD1	PIEA5B text response
0 - Dominican heritage background	'DOIMINICAN', 'DOMENICANA', 'DOMENICANO', 'DOMICANO', 'DOMINCAN', 'DOMINICAN', 'DOMINICAN DESCENT', 'DOMINICAN REPUBLIC', 'DOMINICANA', 'DOMINICANO', 'REPUBLICA DOMINICA', 'SANTO DOMINGO', 'HISPANIC/DOMINICAN', 'DOMINICAN/GERMAN D', 'REPUBLIC SANTO DOM'
1 - Central American heritage background	'HUNDURAS', 'HONDURENIA', 'SALVADORENA', 'GUATEMALAN', 'JUST EL SALVADOR', 'CENTROAMERICANA', 'NICARAGUENCE'
2 - Cuban heritage background	'2' (Reason: '2' = Cuban heritage in PIEA5a options)
3 - Mexican heritage background	'CHICANA', 'CHICANO', 'MEXICAN', 'MEXICAN AMERICAN', 'MEXICAN-AMERICAN', 'MEXICO AMERICANA', 'ASIATICO/MEXICANO'

4 - Puerto-Rican heritage background	'NEW YORK RICAN', 'USA PUERTORICAN', 'PARCIAL PR'
5 - South American heritage background	'BRASIL', 'COLOMBIANA', 'ECUATORIANO', 'PERUANA', 'SPAIN/BRAZIL', 'JUST ECUADORIAN', 'ECUADOR (SEE NOTE)', 'ARGENTINA'
6 - More than one heritage	'CUBAN/COLOMBIAN', 'MEXICAN/PR', 'MIXED RACE', 'MEXICAN/PUERTO RIC', 'EQUADORIAN/P.RICAN', 'PR AND DOMINICAN', 'PR & DOMINICAN', 'PR AND ECUADORIAN', 'BOTH 3 AND 4'

Source variable(s):

PIEA5A. Hispanic/Latino background

PIEA5B. Text response for other Hispanic/Latino background

4.2 BKGRD1_C7 (7-level re-classification of Hispanic/Latino background)

Variable as defined for Bkgrd1. Also, the Other and More than one heritage categories are combined into one group.

Response format:

- 0 = Dominican
- 1 = Central American
- 2 = Cuban
- 3 = Mexican
- 4 = Puerto Rican
- 5 = South American
- 6 = More than one/Other heritage

Source variable(s):

PIEA5A. Hispanic/Latino background

PIEA5B. Text response for other Hispanic/Latino background

4.3 BKGRD1_C7_NOMISS (7-level Hispanic/Latino background, with no missing)

This variable recodes BKGRD1_C7 so that missing is combined with the Other and More than one heritage categories into one group, except for the participant who withdraw consent.

BKGRD1_C7 is missing then BKGRD1_C7_NOMISS = 6;
Else BKGRD1_C7_NOMISS = BKGRD1_C7

Response format: 0=Dominican
1=Central American
2=Cuban
3=Mexican
4=Puerto Rican
5=South American
6=More than one/Other heritage/Missing

Source variable(s):

BKGRD1_C7. 7-level re-classification of Hispanic/Latino Background

4.4 BKGRD1_C6 (6-level re-classification of Hispanic/Latino background)

Variable as defined for Bkgrd1. Also, Central and South American categories are combined into one group and the Other and More than one heritage categories are combined into one group.

Response format: 0 = Dominican
1 = Central or South American
2 = Cuban
3 = Mexican
4 = Puerto Rican
5 = More than one/Other heritage

Source variable(s):

PIEA5A. Hispanic/Latino background

PIEA5B. Text response for other Hispanic/Latino background

4.5 SITE_BKGRD (Cross-classification of center and Hispanic/Latino background)

The self-identification with a Hispanic/Latino cultural group crossed by field center and grouped into the following categories:

Response format:

- 1 = Dominicans - Bronx
- 2 = Central American - Bronx
- 3 = Central American - Chicago
- 4 = Central American - Miami
- 5 = Cubans - Miami
- 6 = Mexicans - Bronx
- 7 = Mexicans - Chicago
- 8 = Mexicans - San Diego
- 9 = Puerto Ricans - Bronx
- 10 = Puerto Ricans - Chicago
- 11 = South American - Bronx
- 12 = South American - Chicago
- 13 = South American - Miami
- 14 = Other - Bronx
- 15 = Other - Chicago
- 16 = Other - Miami
- 17 = Other - San Diego

Source variable(s):

CENTER. Field Center

BKGRD1_C7. Hispanic/Latino background

<i>CENTER (Center)</i>	<i>BKGRD1_C7(7-level re-classification of Hispanic/Latino Background)</i>						
	<i>Dominican</i>	<i>Central American</i>	<i>Cuban</i>	<i>Mexican</i>	<i>Puerto Rican</i>	<i>South American</i>	<i>Mixed/ Other</i>
<i>Bronx</i>	1	2	14	6	9	11	14
<i>Chicago</i>	15	3	15	7	10	12	15
<i>Miami</i>	16	4	5	16	16	13	16
<i>San Diego</i>	17	17	17	8	17	17	17

13 Grey shaded boxes indicate adequate sample size ($n \geq 100$) for comparisons between backgrounds within sites or vice versa. All backgrounds with < 100 within each site are pooled with the Mixed/other category for that site.

4.6 SITE_BKGRD_45PLUS (Cross-classification of center and Hispanic/Latino background, 45+ years only)

The self-identification with a Hispanic/Latino cultural group, based on a sub-cohort of participants 45+ years old, crossed by field center and grouped into the categories below. Participants younger than 45 are classified as missing.

Response format:

- 1 = Dominicans - Bronx
- 2 = Central American - Bronx
- 3 = Central American - Chicago
- 4 = Central American - Miami
- 5 = Cubans - Miami
- 6 = Mexicans - Chicago
- 7 = Mexicans - San Diego
- 8 = Puerto Ricans - Bronx
- 9 = Puerto Ricans - Chicago
- 10 = South American - Bronx
- 11 = South American - Chicago
- 12 = South American - Miami
- 13 = Other – Bronx (including missing on BKGRD1_C7)
- 14 = Other – Chicago (including missing on BKGRD1_C7)
- 15 = Other – Miami (including missing on BKGRD1_C7)
- 16 = Other – San Diego (including missing on BKGRD1_C7)

Source variable(s):

CENTER. Field Center

BKGRD1_C7. Hispanic/Latino background

AGE. Age in years at the time of participant's clinic visit

CENTER (Center)	BKGRD1_C7(7-level re-classification of Hispanic/Latino Background)						
	<i>Dominican</i>	<i>Central American</i>	<i>Cuban</i>	<i>Mexican</i>	<i>Puerto Rican</i>	<i>South American</i>	<i>Mixed/ Other</i>
<i>Bronx</i>	1	2	13	13	8	10	13
<i>Chicago</i>	14	3	14	6	9	11	14
<i>Miami</i>	15	4	5	15	15	12	15
<i>San Diego</i>	16	16	16	7	16	16	16

12 Grey shaded boxes indicate adequate sample size ($n \geq 100$) for comparisons between backgrounds within sites or vice versa. All backgrounds with < 100 within each site are pooled with the Mixed/other category for that site.

4.7 RACE (Race Group (self-report))

The self-identification with a race group. Refusals (.Q) have been coded to a valid category (9) and unknown or missing values have been coded to another valid value (7).

```
If PIEA6 = .Q then RACE = 9
Else if PIEA6 = ( . , 7) then RACE=7
Else then RACE=PIEA6.
```

Response format:

- 1 = "Am. Indian/AK native"
- 2 = "Asian"
- 3 = "HW native/Pac. Islander"
- 4 = "Black"
- 5 = "White"
- 6 = "Multi-racial"
- 7 = "Unknown/NR"
- 9 = "Refused"

Source variable(s):

PIEA6. Participant's self-reported race

4.8 US_BORN (Grouped place of birth – 50 US States only)

This variable groups the place of birth of the participant to the United States (50 states only) or other place of birth (defines refused responses as missing).

```
if piea4='63' then US_BORN=1;
else if piea4^='63' & piea4 not in(' ','Q','q') then US_BORN=0;
else US_BORN=.;
```

Response format:

- 0=Not born in 50 US States.
- 1=Born in 50 US States Only

Source variable(s):

PIEA4. Born in which Country/territory.

4.9 US_NATIVE (Grouped place of birth – US states and territories)

This variable groups the place of birth of the participant to the United States (50 states and US territories which includes Puerto-Rico, Guam, and Virgin Islands) or other place of birth (defines refused responses as missing).

```
if piea4 in(' ','Q','q') then US_NATIVE=.;
else if piea4 in('29','57','63','66') then US_NATIVE=1;
else US_NATIVE=0;
```

Response format:

- 0=Not born in US or US territory
- 1=Born in US or US territory

Source variable(s): PIEA4. Born in which Country/territory.

4.10 LANG_PREF (Language preference for baseline examination)

This variable determines which language was preferred to be used for the baseline examination. Three core forms are used: Personal Information (PIEA), Personal Identifiers (IDEA), and Medical History (MHEA). A majority of those used in one language determines which language was preferred.

Response format: 1=Spanish
2=English

Source variable(s):

Form Version of PIEA/PISA

Form Version of IDEA/IDSA

Form Version of MHEA/MHSA

4.11 IMGEN_C2 (2-level immigrant generation)

This variable determines the participants immigrant generation status based on a combination of the country of origin of the participant, maternal and paternal parents, and/or the maternal and paternal grandparents as follows:

First generation is defined as foreign-born with foreign-born parents (i.e., two foreign-born parents, or one foreign-born parent and one parent missing; same as IMGEN_C4).

The second generation and beyond definition:

If US_BORN in (0, 1) **with both** PIEA11 and PIEA14 are missing, then set IMGEN_C2 as missing.

Else if US_BORN=1 then classified as 2nd + generation.

Else if US_BORN =0 **with at least one US born parent** then classified as 2nd + generation

Else → set IMGEN_C2 as missing.

NOTE: Persons who are born in a US territory are treated as foreign-born or first generation immigrants. This better reflects their migration and acculturation patterns when compared to those born on the US mainland (i.e. US 50).

5/2014 update: recoded 9 to missing

Response format: 1=1st generation
2=2nd generation or higher

Source variable(s):

US_BORN. Grouped place of birth – 50 US states only

PIEA11. Mother's country of origin

PIEA14. Father's country of origin

4.12 IMGEN_C4 (4-level immigrant generation)

This variable determines the participants immigrant generation status based on a combination of the country of origin of the participant, maternal and paternal parents, and/or the maternal and paternal grandparents as follows:

First generation is defined as foreign-born with foreign-born parents (same as IMGEN_C2).

Second generation is defined as US born with **at least one** foreign-born parent. In addition, foreign-born adults with at least one US born parent are classified as second generation to reflect their mixed cultural backgrounds.

The third and fourth generation is defined as US born respondent and both US born mother and father with the following grandparents algorithm shown in chart below:

Grandparent 1	fb	fb	fb	fb	fb	fb	fb	.	.	.	.	.	us b	us b	us b
Grandparent 2	fb	fb	fb	fb	.	fb	fb	fb	fb	.	.	.	us b	us b	us b
Grandparent 3	fb	fb	fb	.	.	usb	us b	us b	us b	us b	.	.	us b	us b	us b
Grandparent 4	fb	usb	.	.	.	.	us b	.	us b	us b	us b	.	.	fb	us b
Generation (IMGEN_C4)	3	3	3	3	3	3	4	4	4	4	4	.	4	4	4

usb= us born, fb= foreign born as defined in IMGEN_C2

Response format: 1=1st generation
2=2nd generation
3=3rd generation
4=4th generation or higher

Source variable(s):

US_BORN. Grouped place of birth – 50 US states only

PIEA11. Mother's country of origin

PIEA12. Maternal grandmother's country of origin

PIEA13. Maternal grandfather's country of origin

PIEA14. Father's country of origin

PIEA15. Paternal grandmother's country of origin

PIEA16. Paternal grandfather's country of origin

4.13 YRSUS (Years lived in the US)

This is a numeric variable that defined years lived in the US as follows:

1. If participant was NOT born in the US, then the years lived in the US is equal to the response of PIEA7.
2. If participant was born in the US, then the years lived in the US is equal to the participant's age (calculated from participant's date of birth).

Source variable(s):

Age. Age of participant in years.

PIEA4. Participant's country of origin

PIEA7. Years lived in mainland US.

4.14 YRSUS_C2 (2-level grouped years lived in the US (50 States))

This is a 2-level grouped numeric variable that defined less than 10years lived in the US versus 10 or more years lived in the US.

Response format: 1=Less than 10 years
2=10 years or more

Source variable(s):

YRSUS. Years lived in the US.

4.15 YRSUS_C3 (3-level grouped years lived in the US (50 States))

This is a 3-level grouped numeric variable that defined less than 10years lived in the US versus 10, more years lived in the US or US born.

Response format: 1=Less than 10 years
2=10 years or more
3= US born

Source variable(s):

YRSUS. Years lived in the US.

US_BORN. Born in US.

4.16 SASH_ALL (Short acculturation scale for Hispanic)

5/2014 update: the variable SASH_ALL was removed from the HCHS participant derived variable file in June 2014 based on recommendations adopted by the steering committee from the HCHS sociocultural working group. SASH_ALL failed to demonstrate adequate psychometric properties in the HCHS cohort data and has been eliminated from the database. Use the two subscales from the SASH instead, SASH_LANG and SASH_SOC.

4.17 SASH_LANG (Short acculturation scale for Hispanics – language subscale)

This is a numeric variable. A 5-point Likert-type scale is assigned to each question with 1 = Only Spanish and 5 = Only English. The average of the 6 questions indicates the degree of language acculturation. If more than 1 item is missing, SASH_LANG is set to missing.

SASH_LANG = average (of scea1-6).

5/2014 update: Added item SCEA6, set to missing if more than 1 item is missing

Source variable(s):

SCEA1. In general, what language(s) do you read and speak?

SCEA2. What was the language(s) you used as a child?

SCEA3. What language(s) do you usually speak at home?

SCEA4. In which language(s) do you usually think?

SCEA5. What language(s) do you usually speak with your friends?

SCEA6. In general, what language(s) are the movies, T.V. and radio programs you prefer to watch and listen to?

4.18 SASH_MEDIA (Short acculturation scale for Hispanics – media subscale)

This derived variable has been dropped on the recommendation of the HCHS socio-cultural working group. Use SCEA6 directly instead of the former SASH_Media scale

4.19 SASH_SOC (Short acculturation scale for Hispanics – social subscale)

This is a numeric variable. A 5-point Likert-type scale is assigned to each question with 1 = All Hispanic/Latino and 5 = All non-Hispanic/non-Latino. The average of the 4 questions indicates the degree of social acculturation. If any item is missing, SASH_SOC is set to missing.

SASH = average (of scea7-10).

5/2014 update: set to missing if any item is missing

Source variable(s):

SCEA7. Your close friends are:

SCEA8. You prefer going to social gatherings/parties at which the people are:

SCEA9. The persons you visit or who visit you are:

SCEA10. If you could choose your children's friends, you would want them to be:

4.20 ETHIS (Ethnic Identification Score)

This is a numeric variable. A 7-point Likert-type scale (values of 1.0-4.0 by 0.5 scale) is assigned to each question with 1.0 = Strongly Disagree and 4.0= Strongly Agree. The average of the 2 questions indicates the degree of ethnic identity. If either item is missing, ETHIS is set to missing.

ETHIS = average (of scea17-18).

5/2014 update: set to missing if either item is missing

Source variable(s):

SCEA17. I have a strong sense of belonging to my own ethnic group.

SCEA18. I have a lot of pride in my ethnic group.

4.21 LANGUAGE_SUBSCORE_MESA (Language Subscore - MESA)

This is a language subscore of the acculturation score similar to the one used in MESA. It is based on the language spoken at home.

This is a numeric variable with possible values from 0 to 2 in 0.5 increments or missing. 0 indicates the lowest level of acculturation, 2 – the highest.

SCEA3	Language_Subscore_MESA
1	0.0
2	0.5
3	1.0
4	1.5
5	2.0
Missing	Missing

Response Format: 0.0 = Only Spanish
0.5 = More Spanish than English
1.0 = Both equally
1.5 = More English than Spanish
2.0 = Only English

Source variable(s):

SCEA3. What language(s) do you usually speak at home?

4.22 NATIVITY_SUBSCORE_MESA (Nativity Subscore - MESA)

This is a nativity subscore of the acculturation score similar to the one used in MESA. It is based on nativity and years of residence in the U.S.

This is a numeric variable. Possible values are integers from 0 to 3 or missing. 0 indicates the lowest level of acculturation, 3 – the highest.

If US_Born is missing or (US_Born=0 and YRSUS is missing) then
Nativity_Subscore_MESA is missing.

Otherwise:

Nativity_Subscore_MESA is 3, if US_Born=1;

2, if US_Born=0 and 20<=YRSUS;

1, if US_Born=0 and 10<=YRSUS<20;

0, if US_Born=0 and .z<YRSUS<10.

Source variable(s):

US_Born. Grouped place of birth – 50 US States only

YRSUS. Years lived in the US

4.23 ACCULT_MESA (Acculturation Score - MESA)

This is a summary acculturation score based on nativity, language spoken at home, and years of residence in the U.S. (similar to the acculturation score used in MESA). It is defined as the sum of nativity subscore and language subscore.

This is a numeric variable with possible values from 0 to 5 in 0.5 increments or missing. 0 indicates the lowest level of acculturation, 5 – the highest.

If Nativity_Subscore_MESA is missing or Language_Subscore_MESA is missing
then Accult_MESA is missing;

Otherwise

Accult_MESA= Nativity_Subscore_MESA + Language_Subscore_MESA.

Source variable(s):

Nativity_Subscore_MESA. Nativity Subscore – MESA

Language_Subscore_MESA. Language Subscore - MESA

4.24 AGE_IMMI (Age of immigration among participants not born in US mainland)

This is a numeric integer valued variable that defined age of immigration as follows:

1. If participant was NOT born in the US 50 states, then the age of immigration is equal to the participants' age minus years lived in the US.
2. If participant was born in the US 50 states, then the age of immigration is equal to missing. Note that 1 person had a missing value for US_BORN and was treated as foreign born.

For participants whose calculated age of immigration is minus 1, shift the value to zero.

```
if us_born=1 then AGE_IMMI=.;  
else AGE_IMMI=(age-piea7);  
if .Z<AGE_IMMI<0 then AGE_IMMI=0;
```

Source variable(s):

Age. Age in years at the time of participant's clinic visit

PIEA7. Years lived in mainland US.

US_Born. Grouped place of birth – 50 US States/DC only

4.25 AGE_IMMI_C3 (3-level age of immigration)

This is a 3-level categorical variable for age of immigration.

```
if .Z < AGE_IMMI < 18 then AGE_IMMI_C3=1;  
else if AGE_IMMI >= 18 then AGE_IMMI_C3=2;  
else if US_BORN=1 then AGE_IMMI_C3=3;  
else AGE_IMMI_C3=.;
```

Response format: 1 = < 18
2 = 18+
3 = US born

Source variable(s):

AGE_IMMI. Age of immigration among participants who were not born in US mainland

US_BORN. Grouped place of birth – 50 US States/DC only

5. ANTHROPOMETRY

5.1 BMI (Body Mass Index kg/m²)

This is a numeric variable.

Missing if the variable, HEIGHT, is missing.

If HEIGHT is not missing then $BMI = ANTA4 / (HEIGHT/100)^2$

Source variable(s):

HEIGHT. Corrected height in centimeters

ANTA4. Weight in kilograms

5.2 BMIGRP_C4 (4-level grouped Body Mass Index - WHO)

This variable is created using the calculated body mass index to define categories based on current WHO classifications.

```
if 0≤BMI<18.5 then BMIGRP_C4=1;
else if 18.5≤BMI<25 then BMIGRP_C4=2;
else if 25≤BMI<30 then BMIGRP_C4=3;
else if 30≤BMI then BMIGRP_C4=4;
else if BMI=.Z then BMIGRP_C4=.;
```

Response format: 1=Underweight (BMI < 18.5)
2=Normal (18.5 ≤ BMI < 25)
3=Overweight (25 ≤ BMI < 30)
4=Obese (BMI ≥ 30)

Source variable(s):

BMI. Body Mass Index in kg/m²

5.3 BMIGRP_C6 (6-level grouped Body Mass Index - WHO)

This variable is created using the calculated body mass index to define categories based on current WHO classifications.

```
if 0≤BMI<18.5 then BMIGRP_C6=1;
else if 18.5≤BMI<25 then BMIGRP_C6=2;
else if 25≤BMI<30 then BMIGRP_C6=3;
else if 30≤BMI<35 then BMIGRP_C6=4;
else if 35≤BMI<40 then BMIGRP_C6=5;
else if 40≤BMI then BMIGRP_C6=6;
else if BMI=.Z then BMIGRP_C6=.;
```


Response format: 1=Underweight (BMI < 18.5)
2=Normal (18.5 ≤ BMI < 25)
3=Overweight (25 ≤ BMI < 30)
4=Obese I (30 ≤ BMI < 35)
5=Obese II (35 ≤ BMI < 40)
6=Obese III (BMI ≥ 40)

Source variable(s):

BMI. Body Mass Index in kg/m²

5.4 WAIST_HIP (Waist to hip ratio)

This is a numeric variable.

Missing if the hip girth is missing

If hip girth is not missing then WAIST_HIP = ANTA10A / ANTA10B

Source variable(s):

ANTA10A. Waist girth measured in centimeters

ANTA10B. Hip girth measured in centimeters

5.5 HEIGHT (Height corrected in cm)

This is a numeric variable. During two time periods the standing height as measured at the Bronx was incorrect due to a mis-mounted stadiometer used in measuring all participants. The derived variable was implemented based on a recommendation from the Quality Control subcommittee.

Missing if the variable, ANTA2 was originally missing.

For participants from the Bronx,

During the time period from the start of the study until October 15, 2008:

height= ANTA2+3.2 cm

From October 16, 2008 through February 1, 2009:

height= ANTA2+2.9 cm

Otherwise, including participants from other field centers, height= ANTA2

Source variable(s):

ANTA0a. Date of the anthropometry measurement

ANTA2. Height in centimeters

CENTER. Participant's field center

6. RESPIRATORY

6.1 ASTHMA_EVER (Ever had asthma)

This is a 0/1/9 numeric variable. From the self-report on the respiratory history form, if the participant says that they ever had asthma, then:

If RSEA32 is not missing, then ASTHMA_EVER= RSEA32.

Response format: 0 = No
1 = Yes
9 = Don't Know

Source variable(s):

RSEA32. Ever had asthma

6.2 ASTHMA_EVER_MD (Ever had asthma with MD diagnosis)

This is a 0/1 numeric variable. From the self-report on the respiratory history form, if the participant says that they ever had asthma and it was diagnosed by a doctor/HP, then:

If rsea32 is not missing:

If rsea32=0 or (rsea32=1 and rsea34=0) then asthma_ever_md=0.
Else if rsea32 =1 and rsea34=1 then asthma_ever_md=1.
Otherwise asthma_ever_md= missing.

Response format: 0 = No
1 = Yes

Source variable(s):

RSEA32. Ever had asthma

RSEA34. Asthma diagnosed by doctor/HP

6.3 ASTHMA_CURR (Currently have asthma)

This is a 0/1/9 numeric variable. From the self-report on the respiratory history form, if the participant says that they ever had asthma and if they still currently have asthma, then:

If RSEA32 is not missing and RSEA35 is not missing, then
If RSEA32 = 0 then ASTHMA_CURR= 0.
Else if RSEA32=1 and RSEA35=0 then ASTHMA_CURR= 0.
Else if RSEA32=1 and RSEA35=1 then ASTHMA_CURR= 1.
Else if RSEA32=1 and RSEA35 in (.,9) then ASTHMA_CURR= 9.

Response format: 0 = No
1 = Yes
9 = Don't Know

Source variable(s):

RSEA32. Ever had asthma

RSEA35. Still has asthma

6.4 **ASTHMA_CURR_MD (Currently have asthma with ever MD diagnosis)**

This is a 0/1 numeric variable. From the self-report on the respiratory history form, if the participant says that they ever had asthma that was diagnosed by a doctor/HP and if they still currently have asthma, then:

If Asthma_ever_md=0 and rsea35 not =1 or 9 then Asthma_Curr_MD=0

Else if Asthma_ever_md=0 and rsea35=1 then asthma_curr_md=0

Else if Asthma_ever_md=1 and rsea35=0 then Asthma_curr_md=0

Else if Asthma_Ever_MD is non-missing and rsea35=1 then Asthma_curr_md=1

Otherwise asthma_curr_md=.

Response format: 0 = No

1 = Yes

Source variable(s):

ASTHMA_EVER_MD. Ever had asthma with MD diagnosis

RSEA35. Still has asthma

6.5 **ASTHMA_C4 (4-level grouped asthma status)**

This is a 1/2/3/9 numeric variable. From the self-report on the respiratory history form, if the participant says that they ever had asthma or not and whether they still currently have asthma, then:

If ASTHMA_EVER is not missing then

If ASTHMA_EVER= 0 then ASTHMA_C4= 1.

Else if ASTHMA_EVER=1 and ASTHMA_CURR=0 then ASTHMA_C4= 2.

Else if ASTHMA_EVER =1 and ASTHMA_CURR =1 then ASTHMA_C4= 3.

Else if ASTHMA_EVER =1 and ASTHMA_CURR = missing or 9 then ASTHMA_C4= 9.

Response format: 1 = Never

2 = Former

3 = Current

9 = Don't Know

Source variable(s):

ASTHMA_EVER. Ever had asthma

ASTHMA_CURR. Still has asthma

6.6 ASTHMA_C4_MD (4-level grouped asthma status – by MD diagnosis)

This is a 1/2/3/9 numeric variable. From the self-report on the respiratory history form, if the participant says that they ever had asthma or not based on an MD diagnosis and whether they still currently have asthma, then:

If ASTHMA_EVER_MD is not missing then

If ASTHMA_EVER_MD= 0 then ASTHMA_C4_MD= 1.

Else if ASTHMA_EVER_MD =1 and ASTHMA_CURR_MD =0 then ASTHMA_C4_MD = 2.

Else if ASTHMA_EVER_MD =1 and ASTHMA_CURR_MD =1 then ASTHMA_C4_MD = 3.

Else if ASTHMA_EVER_MD =1 and ASTHMA_CURR_MD = missing or 9 then

ASTHMA_C4_MD = 9.

Response format: 1 = Never

2 = Former

3 = Current

9 = Don't Know

Source variable(s):

ASTHMA_EVER_MD. Ever had asthma based on MD diagnosis

ASTHMA_CURR_MD. Still has asthma based on ever MD diagnosis

6.7 COPD_SELF (Self-report of ever had COPD or emphysema)

From the self-report on the respiratory history form, if the participant says that they are told by a doctor that they have COPD or emphysema.

If RSEA44 is not missing, then COPD_Self=(RSEA44=1);

Source variable(s):

RSEA44. Ever had COPD or emphysema

6.8 COPD_EVER (Self-report of ever had COPD, emphysema or chronic bronchitis)

From the self-report on the respiratory history form, if the participant says that they are told by a doctor that they have COPD or emphysema and/or chronic bronchitis.

If RSEA42 and RSEA44 is not missing, then.

If rsea42=0 and rsea44=0 then COPD_EVER=0

Else if rsea42=1 or rsea44=1 then COPD_EVER=1

Source variable(s):

RSEA42. Ever has chronic bronchitis

RSEA44. Ever had COPD or emphysema

6.9 MRC_CB (Chronic Productive Cough 3+ mths in 2+ yrs)

From the self-report on the respiratory history form, if the participant says that they had a cough on most days/nights of the week or when waking up in the morning, both for at least 3 months in a row, for at least the last 2 years, then MRC_CB is set to 1. If participant answered “no” to both having a cough on most days/nights of the week or when waking up in the morning, but answered that he/she had a cough for at least the last 2 years, then MRC_CB is still set to 1.

If $\text{rsea3} \geq 2 \text{ years}$ then $\text{rsea3_cat}=1$
Else if $0 \leq \text{rsea3} < 2 \text{ years}$ then $\text{rsea3_cat}=0$

If $\text{rsea1}=0$ and $\text{rsea2}=0$ and $\text{rsea3_cat} \neq 1$ then $\text{MRC_CB}=0$
Else if $\text{rsea3_cat}=0$ then $\text{MRC_CB}=0$
Else if $\text{rsea3_cat}=1$ then $\text{MRC_CB}=1$

5/2014 update: Count RSEA3=99 as missing

Source variable(s):

RSEA1. Ever had a cough on most days/nights of the week (at least 3 months in a row)
RSEA2. Ever had cough when waking up (at least 4 per week)
RSEA3. Years had cough (yes to rsea1 or rsea2 only)

6.10 WHEEZE_SB (Wheezing ever with shortness of breath)

From the self-report on the respiratory history form, if the participant says that they had wheezing or whistling in chest and if they ever had shortness of breath from wheezing or whistling in chest, then:

If $\text{rsea7}=0$ then $\text{Wheeze_SB}=0$
Else if $\text{rsea9}=0$ then $\text{Wheeze_SB}=0$
Else if $\text{rsea9}=1$ then $\text{Wheeze_SB}=1$

Source variable(s):

RSEA7. Ever had wheezing or whistling in chest
RSEA9. Ever had shortness of breath from wheezing or whistling in chest

6.11 WHEEZE_LAST12 (Wheezing in last 12 months)

From the self-report on the respiratory history form, if the participant says that they had wheezing or whistling in chest and if they had it in the last 12 months, then:

If $\text{rsea7}=0$ then $\text{Wheeze_Last12}=0$
Else if $\text{rsea13}=0$ then $\text{Wheeze_Last12}=0$
Else if $\text{rsea13}=1$ then $\text{Wheeze_Last12}=1$

Source variable(s):

RSEA7. Ever had wheezing or whistling in chest
RSEA13. Ever had wheezing or whistling in chest in the last 12 months

6.12 PHLEGM (Phlegm day/night 3 months in a row)

From the self-report on the respiratory history form, if the participant says that they had phlegm in chest most days or nights and if they had it for at least 3 months in a row, then:

If rsea4=0 then Phlegm=0
Else if resa4=1 then Phlegm=1
Else Phlegm=missing

Source variable(s):

RSEA4. Phlegm from chest on most days/nights of the week during at least three months in a row

6.13 PNEUMONIA (Pneumonia - by MD diagnosis)

From the self-report on the respiratory history form, if the participant says that a doctor has ever told them that they had pneumonia or bronchopneumonia, then:

If rsea40=0 then Pneumonia=0
Else if resa40=1 then Pneumonia=1
Else Pneumonia=missing

Source variable(s):

RSEA40. Ever have pneumonia or bronchopneumonia based on MD diagnosis

6.14 VALID_SPIROMETRY (Valid spirometry studies, pre-BD)

This 0/1 variable is an overall data quality indicator derived by the pulmonary reading center on the adequacy study for use. The value is 0 for not usable vs. 1 for usable for analysis.

Response format: 0 = No acceptable curves for either FEV₁ or FVC quality attribute
1 = At least one acceptable curve for both FEV₁ or FVC quality attributes

Source variable(s):

PRBA3. FVC quality attribute (pre-BD)
PRBA4. FEV₁ quality attribute (pre-BD)

6.15 VALID_SPIROMETRY_POSTBD (Valid spirometry studies, post-BD)

This variable is an overall data quality indicator derived by the pulmonary reading center on the adequacy study for use. The value is 0 for not usable vs. 1 for usable for analysis.

Response format: 0 = No acceptable curves for either FEV₁ or FVC quality attribute
1 = At least one acceptable curve for both FEV₁ or FVC quality attributes

Source variable(s):

POBA3. FVC quality attribute (post-BD)
POBA4. FEV₁ quality attribute (post-BD)

6.16 AIRFLOW_OBSTRUCTION (Obstructive lung disease)

This is a 0/1 variable based on the empirical evidence from the spirometry that the ratio of FEV₁ to FVC is below the level defined by the ATS as indicative of restricted airflow.

If PRBA25 is not missing (and not equal to 0) and PRBA29 is not missing then

AIRFLOW_OBSTRUCTION= 1 if

$(PRBA29 / PRBA25) < 0.70$ or $(PRBA29 / PRBA25) * 100 < prba46$.

Else AIRFLOW_OBSTRUCTION= 0.

5/2014 update: set to missing for invalid spirometries

Source variable(s):

PRBA25. FVC (mL)

PRBA29. FEV₁ (mL)

PRBA46. Lower Limit of Normal FEV₁/FVC%

6.17 AIRFLOW_LIMITATION (Airflow limitation (pre FEV₁/FVC<70 only))

This is a 0/1 variable based on the empirical evidence from the spirometry that the ratio of FEV₁ to FVC is below the level defined by the ATS as indicative of restricted airflow.

Defined only for valid spirometries.

If PRBA25 is not missing and PRBA29 is not missing then

AIRFLOW_LIMITATION= 1 if $(PRBA29 / PRBA25) < 0.70$.

Else AIRFLOW_LIMITATION= 0.

Source variable(s):

PRBA25. FVC (mL)

PRBA29. FEV₁ (mL)

Valid_spirometry. Valid spirometry studies, pre-BD

6.18 COPD_BY_BD (COPD by bronchodilator test)

This indicator of COPD is based on the post-bronchodilator measurement of the FEV₁ to FVC ratio. The airflow remains restricted after the administration of albuterol.

If POBA25 is not missing (or zero) and POBA29 is not missing then COPD_BY_BD = $(POBA29 / POBA25) < 0.70$;

Response format: 0 = No

1 = Yes

Source variable(s):

POBA25. FVC (mL)

POBA29. FEV₁ (mL)

6.19 COPD_BY_BD_ALLPRE (COPD by BD test (for all valid pre_BD))

This is a 0/1 variable based on the post-bronchodilator measurement of the FEV₁ to FVC ratio. The airflow remains restricted after the administration of albuterol. This variable is also defined for all participants who did not conduct a post-spirometry exam. Defined only for valid spirometries.

If valid_spirometry=0 then do the following:

COPD_BY_BD_ALLPRE=.

Else if valid_spirometry=1 & POSTBD_EXPECTED=0 then set

COPD_BY_BD_ALLPRE=0

Else if valid_spirometry_postbd2=2 then set COPD_BY_BD_ALLPRE=.

Otherwise set COPD_BY_BD_ALLPRE=COPD_BY_BD

5/2014 update: set to missing for invalid spirometries

Source variable(s):

COPD_by_BD. COPD by bronchodilator test

VALID_SPIROMETRY. Valid spirometry studies (pre-BD).

Valid_spirometry_postbd2. Valid spirometry (A-D grading) adjusted for unexpected BDs.

Postbd_expected. Post-BD expected among valid spirometry.

6.20 BD_RESPONSE_ABS (Bronchodilator responsive by 15% change)

This is a 0/1 variable based on the post-bronchodilator measurement of the FEV₁. The airflow remains restricted after the administration of albuterol the person did not respond by greater than a 15% improvement over their first measurement.

If PRBA29 is not missing and POBA29 is not missing and PRBA29 > 0 then:

If (((poba29-prba29)/prba29)>0.15) then BD_response_abs= 1.

Otherwise if $0 \leq ((poba29-prba29)/prba29) \leq 0.15$ then BD_Response_abs=0.

Response format: 0 = less than or equal to 15% improvement

1 = greater than 15% improvement

Source variable(s):

PRBA29. FEV₁ (mL) pre-BD measurement

POBA29. FEV₁ (mL) post-BD measurement

6.21 FEV1_FVC_RATIO (FEV₁ to FVC %)

This variable is the ratio of forced expiratory volume in one minute (FEV₁) to forced vital capacity (FVC) expressed as a percentage.

Only defined for VALID_SPIROMETRY=1, else missing

If PRBA25 is not missing (and not equal to 0) and PRBA29 is not missing then

$FEV1_FVC_Ratio = (PRBA29 / PRBA25) * 100$

5/2014 update: set to missing for invalid spirometries

Source variable(s):

VALID_SPIROMETRY. Valid spirometry studies (pre BD).

VALID_SPIROMETRY_POSTBD (Valid spirometry studies, post-BD)

PRBA25. FVC (mL)

PRBA29. FEV₁ (mL)

6.22 POSTBD_PREBD_DIFF (Post(FEV1%pred) - Pre(FEV1%pred))

This variable is based on the ratio of the measured FEV₁ difference from post-bronchodilator and pre-bronchodilator exams over the Predicted FEV₁ calculated pre-bronchodilator (both values must be non-missing, and the post-bronchodilator spirometry examinations must be valid), as follows:

Only defined if valid_spirometry=1 and valid_spirometry_postbd=1, else missing

PostBD_PreBD_Diff = ((POBA29-PRBA29)/PRBA37)*100.

5/2014 update: set to missing for invalid spirometries

Source variable(s):

VALID_SPIROMETRY. Valid spirometry studies (pre BD).

PRBA29. FEV₁ (mL) pre-BD measurement

PRBA37. Predicted FEV₁ (mL) (recalculated)

POBA29. FEV₁ (mL) post-BD measurement

6.23 POSTBD_PREBD_DIFF_GT12 (Relative bronchodilator responsive by 12% change)

This is a 0/1 variable based on the ratio of the measured FEV₁ difference from post-bronchodilator and pre-bronchodilator exams over the Predicted FEV₁ calculated pre-bronchodilator (both values must be non-missing, and the post-bronchodilator spirometry examination must be valid) as follows:

if PostBD_PreBD_Diff $\geq$ 12 then PostBD_PreBD_Diff_GT12 = 1.

else if $0 \leq$ PostBD_PreBD_Diff < 12 then PostBD_PreBD_Diff_GT12 = 0.

Source variable(s):

POSTBD_PREBD_DIFF. Post(FEV1%pred)-Pre(FEV1%pred)

FEV1_FVC_Ratio. FEV₁ to FVC %

PRBA46. Lower Limit of Normal FEV1/FVC% (recalculated)

Valid_spirometry. Valid spirometry studies, pre-BD

6.24 POSTBD_PREBD_DIFF_GT15 (Relative bronchodilator responsive by 15% change)

This is a 0/1 variable based on the ratio of the measured FEV₁ difference from post-bronchodilator and pre-bronchodilator exams over the Predicted FEV₁ calculated pre-bronchodilator (both values must be non-missing, and the post-bronchodilator spirometry examination must be valid) as follows:

if PostBD_PreBD_Diff $\geq$ 15 then PostBD_PreBD_Diff_GT15 = 1.
else if $0 \leq$ PostBD_PreBD_Diff < 15 then PostBD_PreBD_Diff_GT15 = 0.

Source variable(s):

POSTBD_PREBD_DIFF. Post(FEV1%pred)-Pre(FEV1%pred)
FEV1_FVC_Ratio. FEV₁ to FVC %
PRBA46. Lower Limit of Normal FEV1/FVC% (recalculated)
Valid_spirometry. Valid spirometry studies, pre-BD

6.25 POSTBD_PREBD_DIFF_GT12_2 (Relative bronchodilator responsive by 12% change (among pre FEV1/FVC<70 only))

This is a 0/1 variable based on the ratio of the measured FEV₁ difference from post-bronchodilator and pre-bronchodilator exams over the Predicted FEV₁ calculated pre-bronchodilator (both values must be non-missing, and the post-bronchodilator spirometry examination must be valid) **among participants with pre FEV1/FVC<70 only (ignores LLN comparison)**, as follows:

if PostBD_PreBD_Diff $\geq$ 12 then PostBD_PreBD_Diff_GT12_2 = 1.
else if $0 \leq$ PostBD_PreBD_Diff < 12 then PostBD_PreBD_Diff_GT12_2 = 0.

Source variable(s):

POSTBD_PREBD_DIFF. Post(FEV1%pred)-Pre(FEV1%pred)
FEV1_FVC_Ratio. FEV₁ to FVC %
PRBA46. Lower Limit of Normal FEV1/FVC% (recalculated)
Valid_spirometry. Valid spirometry studies, pre-BD

6.26 POSTBD_EXPECTED (Post-BD expected (among valid spirometry))

This variable is a flag for whether a post bronchodilator was expected, based on the pre-measurement. A post measurement was expected if the FEV1 to FVC ratio was less than 70% or less than the lower limit of normal for FEV1/FVC%. Note the flag is missing for participants who did not have a valid pre measurement.

If VALID_SPIROMETRY=1 then do the following:

If $0 \leq (\text{FEV1_FVC_RATIO}/100) < 0.70$ or $0 \leq \text{FEV1_FVC_RATIO} < \text{PRBA46}$ then set POSTBD_EXPECTED=1

Otherwise, set POSTBD_EXPECTED=0.

Response format: 0=Not Expected
1= Expected

Source variable(s):

VALID_SPIROMETRY. Valid Spirometry Studies (Pre-BD)

FEV1_FVC_RATIO. FEV1 to FVC Ratio (%)

PRBA46. Lower Limit of Normal FEV1/FVC% (recalculated)

6.27 VALID_SPIROMETRY_POSTBD2 (Valid post-spirometry (A-D grading) adjusted for unexpected BDs)

This variable is a flag for whether a post bronchodilator was valid (grade A-D). Defined only for participants where a post measurement was expected.

if POSTBD_EXPECTED=1 then set VALID_SPIROMETRY_POSTBD2 =
VALID_SPIROMETRY_POSTBD

Response format: 0=Invalid
1=Valid

Source variable(s):

POSTBD_EXPECTED. Post-BD Expected (among valid spirometry)

VALID_SPIROMETRY_POSTBD. Valid Spirometry Studies (Post-BD)

7. WELL-BEING/QUALITY OF LIFE

7.1 CESD10 (10-item CES-D summary score)

This is a numeric variable with values ranging from 0 to 30.

Missing if more than 2 of the 10 items (WBEA1-WBEA10) are missing

Sum of (WBEA1-WBEA10), where positively worded items (WBEA5, WBEA8) are reverse coded. For participants with 1 or 2 missing items, CESD10 is equal to the weighted average of the nonmissing items*10

5/2014 update: for participants with 1 or 2 missing items, the weighted average of the nonmissing items*10 is used instead of the sum of the nonmissing items

Response format for WBEA1-WBEA10:

- 0 = rarely or none of the time
- 1 = some or a little of the time
- 2 = occasionally or a moderate amount of time
- 3 = all of the time

Source variable(s):

WBEA1. I was bothered by things that usually don't bother me

WBEA2. I had trouble keeping my mind on what I was doing

WBEA3. I felt depressed

WBEA4. I felt that everything I did was an effort

WBEA5. I felt hopeful about the future

WBEA6. I felt fearful

WBEA7. My sleep was restless

WBEA8. I was happy

WBEA9. I felt lonely

WBEA10. I could not "get going"

7.2 STAI10 (10-item state-trait anxiety inventory summary score)

This is a numeric variable with values ranging from 0 to 40.

Missing if more than 2 of the 10 items (WBEA11-WBEA20) are missing

Sum of (WBEA11-WBEA20), where positively worded items (WBEA12, WBEA17, WBEA19) are reverse coded, and 1 is added to all responses (converting from 0-3 to 1-4 scale to match the standardized instrument scoring). For participants with 1 or 2 missing items, STAI10 is equal to the weighted average of the nonmissing items*10

5/2014 update: for participants with 1 or 2 missing items, the weighted average of the nonmissing items*10 is used instead of the sum of the nonmissing items

Response format for WBEA11-WBEA20:

- 0 = almost never
- 1 = sometimes
- 2 = often
- 3 = almost always

Source variable(s):

WBEA11. I feel nervous and restless.

WBEA12. I feel satisfied with myself.

WBEA13. I wish I could be as happy as others seem to be.

WBEA14. I feel like a failure.

WBEA15. I worry too much over something that really doesn't matter.

WBEA16. I lack self-confidence.

WBEA17. I feel secure.

WBEA18. I feel inadequate.

WBEA19. I am a steady person.

WBEA20. I get in a state of tension or turmoil as I think over my recent concerns and interests.

8. SF-12v2 HEALTH SURVEY

8.1 AGG_MENT (Aggregate mental health scale)

This is a numeric variable. The factor scores and general population norms developed by John Ware for scoring the SF-12v2© (see pp. 45-51 of “How to Score Version 2 of the SF-12 Health Survey, QualityMetrics, 2002) are used to compute this summary measure for the mental health domain of quality of life. This sub-scale score is a norm-based transformation of standardized Z scores of the constituent items for the established domains which is scaled to a mean of 50 and standard deviation of 10.

Source variable(s): SFE

8.2 AGG_PHYS (Aggregate physical health scale)

This is a numeric variable. The factor scores and general population norms developed by John Ware for scoring the SF-12v2© (see pp. 45-51 of “How to Score Version 2 of the SF-12 Health Survey, QualityMetrics, 2002) are used to compute this summary measure for the physical health domain of quality of life. This sub-scale score is a norm-based on transformation of a standardized Z scores of the constituent items which is scaled to a mean of 50 and standard deviation of 10.

Source variable(s): SFE

9. BLOOD PRESSURE MEASURES

9.1 LEFT_ABI, RIGHT_ABI, AND ABI_OVERALL (Ankle Brachial Index –occ. failure included)

This is a numeric variable. Ankle Brachial Index (ABI) is the average ankle to arm systolic pressures as measured from the left and right sides. The higher of two ankle measures on the same side is divided by the higher of the two systolic arm pressures to derive the right and left side ABI values. The overall ABI is the mean of those two measurements.

Occlusion failures: missing observation values were set to 300 where occlusion failures occurred as follows:

If ABPA1B was missing and ABPA3A2=1 then ABPA1B was set to 300

If ABPA1C was missing and ABPA3A3=1 then ABPA1C was set to 300

If ABPA1D was missing and ABPA3A4=1 then ABPA1D was set to 300

If ABPA1E was missing and ABPA3A5=1 then ABPA1E was set to 300

There were 2 cases where ABI variables were set to missing:

1. If both ABPA1A and ABPA1F were missing.

2. All three of the variables on the same side of the body are =0 as follows:

ABPA1A, ABPA1B, and ABPA1C

ABPA1D, ABPA1E, and ABPA1F

Left side Ankle Brachial Index:

LEFT_ABI=maximum of (abpa1d, abpa1e) / maximum of (abpa1a,abpa1f)

Right side Ankle Brachial Index:

RIGHT_ABI=maximum of (abpa1b, abpa1c)/maximum of (abpa1a,abpa1f)

Overall Ankle Brachial Index:

ABI_OVERALL=mean of (left_ABI, right_ABI)

5/2014 Update: ABI definitions updated to include occlusion failures

Source variable(s):

ABPA1A. Right brachial

ABPA1B. Right dorsalis pedis

ABPA1C. Right posterior tibial

ABPA1D. Left posterior tibial

ABPA1E. Left dorsalis pedis

ABPA1F. Left brachial

ABPA3A2. Occlusion failure of R. dorsalis pedis

ABPA3A3. Occlusion failure of R. posterior tibial

ABPA3A4. Occlusion failure of L. posterior tibial

ABPA3A5. Occlusion failure of L. dorsalis pedis

9.2 LEFT_ABIGRP_C4, RIGHT_ABIGRP_C4 (4-level grouped Ankle Brachial Index (occ. failure incl.))

This variable groups continuous left and right ABI into 4 categories:

When $0 \leq \text{left_ABI} \leq 0.90$ then set $\text{left_ABIGRP_C4} = 1$
When $0.90 < \text{left_ABI} < 1.0$ then set $\text{left_ABIGRP_C4} = 2$
When $1.0 \leq \text{left_ABI} < 1.4$ then set $\text{left_ABIGRP_C4} = 3$
Otherwise when $\text{left_ABI} \geq 1.40$ then set $\text{left_ABIGRP_C4} = 4$

RIGHT_ABIGRP_C4 is defined similarly.

Response format: 1=Low (PAD)
2=Borderline
3=Normal
4=High

Source variable(s):

LEFT_ABI. Right Ankle Brachial Index (occ. failure incl.)
RIGHT_ABI. Left Ankle Brachial Index (occ. failure incl.)

9.3 ABIGRP_C4 (4-level grouped Ankle Brachial Index (occ. failure incl.))

This variable is based on the calculated ankle brachial index (mean of left/right) as follows:

if $0 \leq \text{ABI_OVERALL} \leq 0.90$ then $\text{ABIGRP_C4} = 1$;
else if $0.90 < \text{ABI_OVERALL} < 1.0$ then $\text{ABIGRP_C4} = 2$;
else if $1.0 \leq \text{ABI_OVERALL} < 1.4$ then $\text{ABIGRP_C4} = 3$;
else if $\text{ABI_OVERALL} \geq 1.40$ then $\text{ABIGRP_C4} = 4$;

Response format: 1=Low (PAD)
2=Borderline
3=Normal
4=High

Source variable(s):

ABI_OVERALL. Overall Ankle Brachial Index (occ. failure incl.)

9.4 LEFT_ABI2, RIGHT_ABI2, AND ABI2_OVERALL (Ankle Brachial Index - version 2)

This is a numeric variable based on calculation methods from 2010 (A/E 2010; 171:368–376). Ankle Brachial Index (ABI) is the average ankle to arm systolic pressures as measured from the left and right sides. The higher of two ankle measures on the same side is divided by either the average of the two systolic arm pressures (if they differ by less than 15 mmHg) or the higher of the two systolic arm pressures (if they differ by more than 15 mmHg) to derive the right and left side ABI values. The overall ABI is the minimum of those two measurements.

If absolute value of (abpa1a - abpa1f) > 15 then:

Left side Ankle Brachial Index:

LEFT_ABI2=maximum of (abpa1d,abpa1e) / maximum of (abpa1a,abpa1f)

Right side Ankle Brachial Index:

RIGHT_ABI2=max(abpa1b,abpa1c)/maximum of (abpa1a,abpa1f)

If absolute value of (abpa1a - abpa1f) ≤ 15 then:

Left side Ankle Brachial Index:

LEFT_ABI2=maximum of (abpa1d,abpa1e) / mean of (abpa1a,abpa1f)

Right side Ankle Brachial Index:

RIGHT_ABI2=maximum of (abpa1b,abpa1c)/mean of (abpa1a,abpa1f)

Overall Ankle Brachial Index:

ABI2_OVERALL=min of (left_ABI2, right_ABI2)

Source variable(s):

ABPA1A. Right brachial

ABPA1B. Right dorsalis pedis

ABPA1C. Right posterior tibial

ABPA1D. Left posterior tibial

ABPA1E. Left dorsalis pedis

ABPA1F. Left brachial

9.5 ABIGRP2_C4 (4-level grouped Ankle Brachial Index - version 2)

This variable is based on the calculated ankle brachial index (min of left/right) as follows:

if $0 \leq \text{ABI2_OVERALL} \leq 0.90$ then ABIGRP2_C4=1;
else if $0.90 < \text{ABI2_OVERALL} < 1.0$ then ABIGRP2_C4=2;
else if $1.0 \leq \text{ABI2_OVERALL} < 1.4$ then ABIGRP2_C4=3;
else if $\text{ABI2_OVERALL} \geq 1.40$ then ABIGRP2_C4=4;

Response format: 1=Low (PAD)

2=Borderline

3=Normal

4=High

Source variable(s):

ABI2_OVERALL. Ankle Brachial Index – version 2.

10. LABORATORY MEASURES

10.1 INSULIN_FAST (Insulin, fasting (calibrated, converted to mU/L))

This variable combines fasting glucose results from the central lab that was processed prior to July 5, 2010 (using units of mU/L) with results processed on or after July 5, 2010 (using units of pmol/L) and the overall units are converted to mU/L (1 mU/L = 6 pmol/L).

In addition, values collected prior to October 29, 2009 have been calibrated using the following regression equation: $y = 1.00494x - 1.4504$

Where y = adjusted insulin value, x = original insulin value using Merdcodia assay.

By doing so, there is no overlap in the adjustment for a change in assay from Merdcodia to Roche Elecsys analyzers which occurred on October 29, 2009. Update 12/2013: changed missing to .L for .L values of LABA73

Source variable(s):

LABA73. Fasting Insulin (mU/L)

LABA96. Fasting Insulin (pmol/L)

10.2 INSULIN_OGTT (Insulin, post OGTT (calibrated, converted to mU/L))

This variable combines post OGTT glucose results from the central lab that were processed prior to July 5, 2010 (using units of mU/L) with results processed on or after July 5, 2010 (using units of pmol/L) and the overall units are converted to mU/L (1 mU/L = 6 pmol/L).

In addition, values collected prior to October 29, 2009 have been calibrated using the following regression equation: $y = 1.00494x - 1.4504$

Where y = adjusted insulin value, x = original insulin value using Merdcodia assay.

By doing so, there is no overlap in the adjustment for a change in assay from Merdcodia to Roche Elecsys analyzers which occurred on October 29, 2009.

5/2014 update: corrected to account for .N values of insulin

Source variable(s):

LABA90. Post OGTT Insulin (mU/L)

LABA97. Post OGTT Insulin (pmol/L)

10.3 INSULIN_DATE (Date of insulin measures - All)

This variable combines date variables for insulin testing from the central lab that were processed prior to July 5, 2010 with those processed on or after July 5, 2010.

Source variable(s):

LABA73A. Date of Insulin measures (before July 5, 2010)

LABA96A. Date of Insulin measures (on or after July 5, 2010)

10.4 DYSLIPIDEMIA (High LDL, low HDL, or high triglycerides - NIH)

This is a 0/1 variable that checks for HDL, LDL, and triglyceride values to determine presence/absence of dyslipidemia as follows:

If laba69 ≤ .Z & laba68 ≤ .Z & laba67 ≤ .Z then DYSLIPIDEMIA= missing.
Else if laba69 ≥ 160 OR 0 ≤ laba68 < 40 OR laba67 ≥ 200 then DYSLIPIDEMIA=1
Else DYSLIPIDEMIA=0.

Response format: 0= Not dyslipidemic (No)
1= Dyslipidemic (Yes)

Source variable(s):

LABA67. Triglycerides (mg/dL)
LABA68. HDL - cholesterol (mg/dL)
LABA69. LDL - Cholesterol (mg/dL)

10.5 DYSLIPIDEMIA_C3 (3-category dyslipidemia)

This is a numeric variable that describes the status of dyslipidemia based on NIH definition and uses of self-reported or scanned medication.

if dyslipidemia ≤ .Z & med_llid = 1 & muea33e = 1 then DYSLIPIDEMIA_C3 = .;
else if dyslipidemia = 0 & med_llid = 1 & muea33e = 1 then DYSLIPIDEMIA_C3 = 1;
else if dyslipidemia = 1 & med_llid = 1 & muea33e = 1 then DYSLIPIDEMIA_C3 = 2;
else if dyslipidemia = 1 | med_llid = 1 | muea33e = 1 then DYSLIPIDEMIA_C3 = 3;

Response format: 1 = Non-dyslipidemia
2 = Dyslipidemia w/o Treatment
3 = Dyslipidemia with Treatment

Source variable(s):

Dyslipidemia. High LDL, low HDL, or high triglycerides - NIH
MED_LLD. Scanned Lipid lowering drugs/Antihyperlipidemics (LLD)
MUEA33e. Self-reported Antihyperlipidemic medication used.

10.6 CARDIAC_RISK_RATIO (Cardiac Risk Ratio (TC/HDL) - AHA)

This is a numeric variable that calculates the ratio of total cholesterol by HDL cholesterol, as follows:

If laba66 and laba68 are both non-missing, then:
CARDIAC_RISK_RATIO = laba66 / laba68.

Source variable(s):

LABA66. Total Cholesterol (mg/dL)
LABA68. HDL - cholesterol (mg/dL)

10.7 DYSLIPIDEMIA_TCHDL (Dyslipidemia based on Cardiac Risk Ratio > 5)

This is a 0/1 variable that checks for Dyslipidemia based on Cardiac Risk Ratio (TC/HDL) > 5 as follows:

```
if cardiac_risk_ratio<=.Z then DYSLIPIDEMIA_TCHDL=.;  
else if cardiac_risk_ratio>5 then DYSLIPIDEMIA_TCHDL=1;  
else DYSLIPIDEMIA_TCHDL=0;
```

Response format: 0= Not dyslipidemic (No)
1= Dyslipidemic (Yes)

Source variable(s):

Cardiac_Risk_Ratio. Cardiac_Risk_Ratio (TC/HDL) AHA

10.8 DYS_TCHDL_MED (Dyslipidemia based on Cardiac Risk Ratio > 5 or Lipid Lower Drug scanned at baseline)

This is a 0/1 variable that checks for Dyslipidemia defined by Cardiac Risk Ratio (TC/HDL) > 5 or Lipid Lower Drug scanned at baseline as follows:

```
if dyslipidemia_tchdl<=.Z & med_llid<=.Z then DYS_TCHDL_MED=.;  
else if dyslipidemia_tchdl=1 | med_llid=1 then DYS_TCHDL_MED=1;  
else DYS_TCHDL_MED=0;
```

Response format: 0= Not dyslipidemic (No)
1= Dyslipidemic (Yes)

Source variable(s):

Cardiac_Risk_Ratio. Cardiac_Risk_Ratio (TC/HDL) AHA
MED_LLD. Lipid lowering drugs/Antihyperlipidemics (LLD)

10.9 HIGH_TOTAL_CHOL (Hypercholesterolemia using lab data and lipid meds)

This is a 0/1 variable that checks for Total cholesterol, HDL, LDL, and antihyperlipidemic medication use values to determine presence/absence of hypercholesterolemia as follows:

```
If missing(laba66) & missing(laba68) & missing(laba69) and missing(med_llid) then  
High_total_Chol=.;  
Else if (laba66>=240) OR (laba69>= 160) OR (0<=laba68<40) OR (MED_LLD=1)  
then High_total_Chol =1;  
Else High_total_Chol =0;
```

5/2014 update: updates due to updated medication variables

Response format: 0= Not Hypercholesterolemic (No)
1= Hypercholesterolemic (Yes)

Source variable(s):

LABA66. Total Cholesterol (mg/dL)

LABA68. HDL - Cholesterol (mg/dL)

LABA69. LDL - Cholesterol (mg/dL)

MED_LLD. Lipid lowering drugs/Antihyperlipidemics (LLD)

10.10 TOTAL_CHOL_C3 (3-category total cholesterol)

This is a numeric variable that describes the stages of total cholesterol based on lab measurement as follow.

```
if .Z<lab66<200 then TOTAL_CHOL_C3=1;  
else if 200<=lab66<=239 then TOTAL_CHOL_C3=2;  
else if lab66>=240 then TOTAL_CHOL_C3=3;  
else TOTAL_CHOL_C3=.;
```

Response format: 1 = Normal Cholesterol
 2 = Borderline High Cholesterol
 3 = High Cholesterol

Source variable(s):

LABA66. Total Cholesterol (mg/dL)

10.11 FASTTIME (Fasting time in hours)

Elapsed time between the time the participant last consumed anything and the blood draw. This is a numeric variable scaled in hours.

Source variable(s):

BIOA7. Day last meal = (1= today, 0hr. adjustment; 2=yesterday, 12hr. adjustment; 3=day before yesterday, 24hr. adjustment.)

BIOA8. time of last meal

BIOA9. date of blood draw

BIOA10. time of blood draw

10.12 IFG (Impaired fasting glucose - AHA)

This is a 0/1 variable that checks for Impaired Fasting Glucose by AHA Guideline as follows:

```
if 100<=LABA70<=125 & (MED_ANTIDIAB NE 1) then IFG=1;  
else if 0<LABA70<100 | LABA70>125 then IFG=0;
```

Response format: 0= Not Impaired Fasting Glucose (No)
 1= Impaired Fasting Glucose (Yes)

Source variable(s):

LABA70. Fasting glucose (mg/dL)

MED_ANTIDIAB. Scanned antidiabetics medication

10.13 IGT (Impaired glucose tolerance - AHA)

This is a 0/1 variable that checks for Impaired Glucose Tolerance by AHA Guideline as follows:

if $0 < \text{LABA70} < 126$ & $140 \leq \text{LABA71} < 200$ then $\text{IGT} = 1$;
else if $\text{LABA70} > .Z$ & $\text{LABA71} > .Z$ then $\text{IGT} = 0$;

Response format: 0= Not Impaired Glucose Tolerance (No)
1= Impaired Glucose Tolerance (Yes)

Source variable(s):

LABA70. Fasting glucose (mg/dL)

LABA71. Glucose, post OGTT (mg/dL)

10.14 GFR_Y06 (Glomerular filtration rate estimated, 2006)

This variable is estimated based on standard reference equations for Hispanics provided by the central clinical laboratory. The GFR value has been adjusted for age and sex.

Glomerular Filtration Rate (mL/min/1.73m²) is estimated using the Modification of Diet in Renal Disease (MDRD) Study equation¹ which can be found at

<https://www.niddk.nih.gov/research-funding/research-programs/kidney-clinical-research-epidemiology/laboratory/glomerular-filtration-rate-equations/adults/previous>

Specifically, the MDRD study equation is:

$\text{GFR (mL/min/1.73 m}^2\text{)} = 175 \times (\text{S}_{\text{Cr}})^{-1.154} \times (\text{Age})^{-0.203} \times (0.742 \text{ if female}) \times (1.212 \text{ if African American})$
(conventional units)

where S_{Cr} is serum creatinine (LABA76)

¹Levey AS, Coresh J, Greene T, Stevens LA, Zhang YL, Hendriksen S, Kusek JW, Van Lente F, Chronic Kidney Disease Epidemiology Collaboration. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. *Ann Intern Med.* 2006;145:247–254.

Source variable(s):

LABA76 – Serum creatinine

SEX. Sex

AGE. Age (in years at the time of participant's clinic visit)

RACE. Race Group (self-report)

10.15 GFRSCR_Y06_C5 (Chronic kidney disease using eGFR only - NIDDK)

This variable is estimated based on the estimated GFR value and is defined using NIDDK guidelines, as follows:

if $\text{GFR_Y06} \geq 90$ then $\text{GFRSCR_Y06_C5} = 1$ ('Normal')
else if $60 \leq \text{GFR_Y06} < 90$ then $\text{GFRSCR_Y06_C5} = 2$ ('Mild')
else if $30 \leq \text{GFR_Y06} < 60$ then $\text{GFRSCR_Y06_C5} = 3$ ('Moderate')
else if $15 \leq \text{GFR_Y06} < 30$ then $\text{GFRSCR_Y06_C5} = 4$ ('Severe')
else if $.Z < \text{GFR_Y06} < 15$ then $\text{GFRSCR_Y06_C5} = 5$ ('End-Stage')

Response format: 1=Normal
2=Mild
3=Moderate
4=Severe
5=End-Stage

Source variable(s):

GFR_Y06. Glomerular Filtration Rate (mL/min/1.73m²), estimated using the Modification of Diet in Renal Disease (MDRD) Study equation by Levey et al, *Ann Intern Med.* 2006; 145:247–254.

10.16 GFRSCYS_Y08 (eGFR based on serum cystatin C w/o demographics, 2008)

Estimated Glomerular filtration rate, based on serum cystatin C and no demographic factors.

Reference:

Estimating GFR using serum cystatin C alone and in combination with serum creatinine: a pooled analysis of 3,418 individuals with CKD. *Am J Kidney Dis.* 2008 Mar ;51(3):395-406

$$\text{GFRSCYS_Y08} = 76.7 * \text{LABA101}^{-1.19}$$

Source variable(s):

LABA101. Cystatin C (mg/L)

10.17 GFRSCR_Y09 (eGFR based on serum creatinine, sex, age, and race, 2009)

Estimated Glomerular filtration rate, based on serum creatinine, and demographics sex, age and race (black vs. nonblack).

Reference:

Levey AS, et al. (2009). "A New Equation to Estimate Glomerular Filtration Rate." *Annals of Internal Medicine.*

$$\begin{aligned} &\text{If (SEX=F and RACE=4) then } 141 * \min(\text{LABA76}/0.7, 1)^{-0.329} * \\ &\quad \max(\text{LABA76}/0.7, 1)^{-1.209} * 0.993^{\text{AGE}} * 1.018 * 1.159 \\ &\text{If (SEX=F and RACE}\neq 4) \text{ then } 141 * \min(\text{LABA76}/0.7, 1)^{-0.329} * \\ &\quad \max(\text{LABA76}/0.7, 1)^{-1.209} * 0.993^{\text{AGE}} * 1.018 \\ &\text{If (SEX=M and RACE=4) then } 141 * \min(\text{LABA76}/0.9, 1)^{-0.411} * \\ &\quad \max(\text{LABA76}/0.9, 1)^{-1.209} * 0.993^{\text{AGE}} * 1.159 \\ &\text{If (SEX=M and RACE}\neq 4) \text{ then } 141 * \min(\text{LABA76}/0.9, 1)^{-0.411} * \\ &\quad \max(\text{LABA76}/0.9, 1)^{-1.209} * 0.993^{\text{AGE}} \end{aligned}$$

Source variable(s):

SEX. Sex

RACE. Race Group (self-report)

AGE. Age (in years at the time of participant's clinic visit)

LABA76. Creatinine (mg/dL)

10.18 GFRSCYS_SCR_Y12 (eGFR based on serum cystatin C, serum creatinine, sex, age and race, 2012)

Estimated Glomerular filtration rate, based on serum cystatin C, serum creatinine, and demographics sex, age and race (black vs. nonblack).

Reference:

Estimating glomerular filtration rate from serum creatinine and cystatin C. N Engl J Med. 2012 Jul 5 ;367(1):20-9

If (SEX=F and RACE=4) then $135 * \min(\text{LABA76}/0.7, 1)^{-0.248} * \max(\text{LABA76}/0.7, 1)^{-0.601} * \min(\text{LABA101}/0.8, 1)^{-0.375} * \max(\text{LABA101}/0.8, 1)^{-0.711} * 0.995^{\text{AGE}} * 0.969 * 1.08$

If (SEX=F and RACE≠4) then $135 * \min(\text{LABA76}/0.7, 1)^{-0.248} * \max(\text{LABA76}/0.7, 1)^{-0.601} * \min(\text{LABA101}/0.8, 1)^{-0.375} * \max(\text{LABA101}/0.8, 1)^{-0.711} * 0.995^{\text{AGE}} * 0.969$

If (SEX=M and RACE=4) then $135 * \min(\text{LABA76}/0.9, 1)^{-0.207} * \max(\text{LABA76}/0.9, 1)^{-0.601} * \min(\text{LABA101}/0.8, 1)^{-0.375} * \max(\text{LABA101}/0.8, 1)^{-0.711} * 0.995^{\text{AGE}} * 1.08$

If (SEX=M and RACE≠4) then $135 * \min(\text{LABA76}/0.9, 1)^{-0.207} * \max(\text{LABA76}/0.9, 1)^{-0.601} * \min(\text{LABA101}/0.8, 1)^{-0.375} * \max(\text{LABA101}/0.8, 1)^{-0.711} * 0.995^{\text{AGE}}$

Source variable(s):

SEX. Sex

RACE. Race Group (self-report)

AGE. Age (in years at the time of participant's clinic visit)

LABA76. Creatinine (mg/dL)

LABA101. Cystatin C (mg/L)

10.19 GFRSCR_Y21 (eGFR using serum creatinine only, 2021)

Estimated Glomerular filtration rate, based on 2021 definition using serum creatinine, sex and age.

Reference:

Inker LA, Eneanya ND, Coresh J, et al. New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race. N Engl J Med. 2021;385(19):1737-1749.
doi:10.1056/NEJMoa2102953

If SEX=0 then:

if LABA76 ≤ 0.7, then $143.704 * ((\text{LABA76} / 0.7)^{-0.241}) * ((0.9938)^{\text{AGE}})$

if LABA76 > 0.7 then $143.704 * ((\text{LABA76} / 0.7)^{-1.200}) * ((0.9938)^{\text{AGE}})$

If SEX=1 then:

if LABA76 ≤ 0.9 then $142 * ((\text{LABA76} / 0.9)^{-0.302}) * ((0.9938)^{\text{AGE}})$

if LABA76 > 0.9 then $142 * ((\text{LABA76} / 0.9)^{-1.200}) * ((0.9938)^{\text{AGE}})$

Source variable(s):

SEX. Sex

AGE. Age

LABA76. Creatinine (mg/dL)

10.20 GFRSCYS_SCR_Y21 (eGFR using serum creatinine and cystatin C, 2021)

Estimated Glomerular filtration rate, based on 2021 definition using serum creatinine, cystatin C, sex and age.

Reference:

Inker LA, Eneanya ND, Coresh J, et al. New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race. N Engl J Med. 2021;385(19):1737-1749.
doi:10.1056/NEJMoa2102953

If SEX=0 then:

if LABA76 <=0.7 AND LABA101 <= 0.8, then $130.005 * ((\text{LABA76} / 0.7)^{-0.219}) * ((\text{LABA101} / 0.8)^{-0.323}) * ((0.9961)^{\text{AGE}})$

if LABA76 <=0.7 AND LABA101 > 0.8, then $130.005 * ((\text{LABA76} / 0.7)^{-0.219}) * ((\text{LABA101} / 0.8)^{-0.778}) * ((0.9961)^{\text{AGE}})$

if LABA76 > 0.7 AND LABA101 <= 0.8, then $130.005 * ((\text{LABA76} / 0.7)^{-0.544}) * ((\text{LABA101} / 0.8)^{-0.323}) * ((0.9961)^{\text{AGE}})$

if LABA76 > 0.7 AND LABA101 > 0.8, then $130.005 * ((\text{LABA76} / 0.7)^{-0.544}) * ((\text{LABA101} / 0.8)^{-0.778}) * ((0.9961)^{\text{AGE}})$

If SEX=1 then:

if LABA76 <=0.9 AND LABA101 <= 0.8, then $135 * ((\text{LABA76} / 0.9)^{-0.144}) * ((\text{LABA101} / 0.8)^{-0.323}) * ((0.9961)^{\text{AGE}})$

if LABA76 <=0.9 AND LABA101 > 0.8, then $135 * ((\text{LABA76} / 0.9)^{-0.144}) * ((\text{LABA101} / 0.8)^{-0.778}) * ((0.9961)^{\text{AGE}})$

if LABA76 > 0.9 AND LABA101 <= 0.8, then $135 * ((\text{LABA76} / 0.9)^{-0.544}) * ((\text{LABA101} / 0.8)^{-0.323}) * ((0.9961)^{\text{AGE}})$

if LABA76 > 0.9 AND LABA101 > 0.8, then $135 * ((\text{LABA76} / 0.9)^{-0.544}) * ((\text{LABA101} / 0.8)^{-0.778}) * ((0.9961)^{\text{AGE}})$

Source variable(s):

SEX. Sex

AGE. Age

LABA76. Creatinine (mg/dL)

LABA101. Cystatin C (mg/L)

10.21 GFRSCR_Y21_C5 (5-category eGFR using only serum creatinine, 2021)

Estimated Glomerular filtration rate 5-category variable, based on 2021 definition using serum creatinine, sex and age.

If GFRSCR_Y21 >= 90, then GFRSCR_Y21_C5 = 1
If 60 <= GFRSCR_Y21 < 90, then GFRSCR_Y21_C5 = 2
If 30 <= GFRSCR_Y21 < 60, then GFRSCR_Y21_C5 = 3
If 15 <= GFRSCR_Y21 < 30, then GFRSCR_Y21_C5 = 4
If .Z < GFRSCR_Y21 < 15, then GFRSCR_Y21_C5 = 5

Response format: 1 = Normal
2 = Mild
3 = Moderate
4 = Severe
5 = End-Stage

Source variable(s):

GFRSCR_Y21. eGFR based on 2021 definition using serum creatinine only

10.22 GFRSCR_Y21_C2 (Binary eGFR using only serum creatinine, 2021)

Estimated Glomerular filtration rate binary variable, based on 2021 definition using serum creatinine only.

IF GFRSCR_Y21 >=60 AND LABA81<30 then GFRSCR_Y21_C2 = 0
else if GFRSCR_Y21 <60 OR LABA81 >=30 then GFRSCR_Y21_C2 = 1
if GFRSCR_Y21 =missing AND LABA81 = missing then GFRSCR_Y21_C2 = missing
if GFRSCR_Y21 >=60 AND LABA81 = missing then GFRSCR_Y21_C2 = missing
if GFRSCR_Y21 =missing AND LABA81 <30 AND LABA81 not missing then
GFRSCR_Y21_C2 =missing

Source variable(s):

GFRSCR_Y21. eGFR based on 2021 definition using serum creatinine only
LABA81. Urine Albumin/creatinine ratio

10.23 GFRSCYS_SCR_Y21_C5 (5-category eGFR using serum creatinine and cystatin C, 2021)

Estimated Glomerular filtration rate 5-category variable, based on 2021 definition using serum creatinine, cystatin C, sex and age.

If GFRSCYS_SCR_Y21 >= 90, then GFRSCYS_SCR_Y21_C5 = 1
If 60 <= GFRSCYS_SCR_Y21 < 90, then GFRSCYS_SCR_Y21_C5 = 2
If 30 <= GFRSCYS_SCR_Y21 < 60, then GFRSCYS_SCR_Y21_C5 = 3
If 15 <= GFRSCYS_SCR_Y21 < 30, then GFRSCYS_SCR_Y21_C5 = 4
If .Z < GFRSCYS_SCR_Y21 < 15, then GFRSCYS_SCR_Y21_C5 = 5

Response format: 1 = Normal
2 = Mild
3 = Moderate
4 = Severe
5 = End-Stage

Source variable(s):

GFRSCYS_SCR_Y21. eGFR based on 2021 definition using serum creatinine & cystatin C

10.24 GFRSCYS_SCR_Y21_C2 (Binary eGFR using serum creatinine and cystatin C, 2021)

Estimated Glomerular filtration rate binary variable, based on 2021 definition using serum creatinine only.

IF GFRSCYS_SCR_Y21 >=60 AND LABA81<30 then GFRSCYS_SCR_YR21_C2 = 0
else if GFRSCYS_SCR_Y21 <60 OR LABA81 >=30 then GFRSCYS_SCR_YR21_C2 = 1
if GFRSCYS_SCR_Y21 =missing AND LABA81 = missing then
GFRSCYS_SCR_YR21_C2 = missing
if GFRSCYS_SCR_Y21>=60 AND LABA81 = missing then GFRSCYS_SCR_YR21_C2 =
missing
if GFRSCYS_SCR_Y21=missing AND LABA81 <30 AND LABA81 not missing then
GFRSCYS_SCR_YR21_C2 =missing

Source variable(s):

GFRSCYS_SCR_Y21. eGFR based on 2021 definition using serum creatinine & cystatin C
LABA81. Urine Albumin/creatinine ratio

10.25 ALBUMINURIA_C3 (Albuminuria based on 2024 KDIGO guidelines)

Albuminuria 3-category variable, based on 2024 KDIGO definition.

Reference:

Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney Int. 2024;105(4S): S117–S314. PMID: 38490803

If LABA81 not missing and LABA81 < 30 then ALBUMINURIA_C3 = 1
If LABA81 >= 30 and LABA81 <= 300 then ALBUMINURIA_C3 = 2
If LABA81 > 300 then ALBUMINURIA_C3 = 3

Response format: 1 = Normal
2 = Microalbuminuria
3 = Macroalbuminuria

Source variable(s):

LABA81. Albumin/creatinine ratio (mg/g)

11. GLOBAL PHYSICAL ACTIVITY QUESTIONNAIRE (GPAQ)

The GPAQ was developed by the World Health Organization (WHO). It collects information on physical activity participation in three settings (or domains: work, travel, leisure) and sedentary behavior. The variables below were defined using the WHO GPAQ Analysis Guide:

<https://www.who.int/docs/default-source/ncds/ncd-surveillance/gpaq-analysis-guide.pdf>

A MET (Metabolic Equivalent) is the ratio of the metabolic rate for a given activity relative to resting metabolic rate. One MET is defined as 1 kcal/kg/hour and is equivalent to the energy cost of sitting quietly. A MET is also defined as oxygen uptake in ml/kg/min with one MET equal to the oxygen cost of sitting quietly, around 3.5 ml/kg/min.

For calculations described below, MET values were assigned value corresponding to an average MET for typical types and intensities of activity.

Domain	METS value
Work	- Moderate MET value = 4.0 - Vigorous MET value = 8.0
Transport	Cycling and walking MET value = 4.0
Recreation	- Moderate MET value = 4.0 - Vigorous MET value = 8.0

11.1 GPAQ_REC_MOD (Recreational physical (moderate) activity (min/day))

This is a numeric variable that calculates the average amount of time spent per day doing a moderate recreational physical activity. Update 12/2013: set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.

If days of activity and time spent at different intensity levels are not missing, then
$$\text{GPAQ_rec_mod} = (\text{PAEA16} \times \text{PAEA17A_MIN})/7$$

Source variable(s): In a typical week, ...

PAEA16. Days of moderate-intensity sports, fitness or recreational activities?

PAEA17A_MIN. Time spent doing moderate-intensity recreational activities (min)

11.2 GPAQ_REC_VIG (Recreational physical (vigorous) activity (min/day))

This is a numeric variable that calculates the average amount of time spent per day doing a vigorous recreational physical activity. Update 12/2013: set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.

If days of activity and time spent at different intensity levels are not missing, then
$$\text{GPAQ_rec_vig} = (\text{PAEA12} \times \text{PAEA13A_MIN})/7$$

Source variable(s): In a typical week, ...

PAEA12. Days of vigorous-intensity sports, fitness or recreational (leisure) activities?

PAEA13A_MIN. Time spent doing vigorous-intensity recreational activities (min)

11.3 GPAQ_REC (Recreational physical activity (min/day))

This is a numeric variable that finds the total time per day doing some form of recreational physical activity. Update 12/2013: set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.

If days of activity and time spent at different intensity levels are not missing, then
$$\text{GPAQ_REC} = \text{sum}(\text{GPAQ_REC_MOD}, \text{GPAQ_REC_VIG}).$$

Source variable(s):

GPAQ_rec_mod. Recreational physical (moderate) activity (min/day)

GPAQ_rec_vig. Recreational physical (vigorous) activity (min/day)

11.4 GPAQ_WORK_MOD (Work-related (moderate) physical activity (min/day))

This is a numeric variable that calculates the average amount of time spent per day doing a moderate work-related physical activity. Update 12/2013: set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.

If days of activity and time spent at different intensity levels are not missing, then
$$\text{GPAQ_WORK_MOD} = (\text{PAEA5} \times \text{PAEA6A_MIN})/7$$

Source variable(s): In a typical week, ...

PAEA5. Days of moderate-intensity activities as part of your work?

PAEA6A_MIN. Time spent doing moderate-intensity activities at work (min)

11.5 GPAQ_WORK_VIG (Work-related (vigorous) physical activity (min/day))

This is a numeric variable that calculates the average amount of time spent per day doing a vigorous work-related physical activity. Update 12/2013: set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.

If days of activity and time spent at different intensity levels are not missing, then
$$\text{GPAQ_WORK_VIG} = (\text{PAEA2} \times \text{PAEA3A_MIN})/7$$

Source variable(s): In a typical week, ...

PAEA2. Days of vigorous-intensity activities as part of your work?

PAEA3A_MIN. Time spent doing vigorous-intensity activities at work (min)

11.6 GPAQ_WORK (Work-related physical activity (min/day))

This is a numeric variable that finds the total time per day doing some form of work-related physical activity. Update 12/2013: set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.

If days of activity and time spent at different intensity levels are not missing, then
$$\text{GPAQ_WORK} = \text{sum}(\text{GPAQ_WORK_MOD}, \text{GPAQ_WORK_VIG}).$$

Source variable(s):

GPAQ_work_mod. Work-related (moderate) physical activity (min/day)

GPAQ_work_vig. Work-related (vigorous) physical activity (min/day)

11.7 GPAQ_TOTAL_MOD (Total physical (moderate) activity (min/day))

This is a numeric variable that finds the average amount of time spent per day doing moderate physical activity. Update 12/2013: set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.

If days of activity and time spent at different intensity levels are not missing, then
$$\text{GPAQ_TOTAL_MOD} = \text{sum}((\text{PAEA5} \times \text{PAEA6A_MIN}), (\text{PAEA8} \times \text{PAEA9A_MIN}), (\text{PAEA16} \times \text{PAEA17A_MIN}))/7$$

Source variable(s): In a typical week, ...

PAEA5. Days of moderate-intensity activities as part of your work?

PAEA6A_MIN. Time spent doing moderate-intensity activities at work (min)

PAEA8. Days walk or bicycle to get to and from places?

PAEA9_MIN. Time spent walking or bicycling for travel (min)

PAEA16. Days of moderate-intensity sports, fitness or recreational activities?

PAEA17A_MIN. Time spent doing moderate-intensity recreational activities (min)

11.8 GPAQ_TOTAL_VIG (Total physical (vigorous) activity (min/day))

This is a numeric variable that finds the average amount of time spent per day doing vigorous physical activity. Update 12/2013: set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.

If days of activity and time spent at different intensity levels are not missing, then
$$\text{GPAQ_TOTAL_VIG} = \text{sum}((\text{PAEA2} \times \text{PAEA3A_MIN}), (\text{PAEA12} \times \text{PAEA13A_MIN}))/7$$

Source variable(s): In a typical week, ...

PAEA2. Days of vigorous-intensity activities as part of your work?

PAEA3A_MIN. Time spent doing vigorous-intensity activities at work (min)

PAEA12. Days of vigorous-intensity sports, fitness or recreational (leisure) activities?

PAEA13A_MIN. Time spent doing vigorous-intensity recreational activities (min)

11.9 GPAQ_TRSPORT (Transportation-related physical activity (min/day))

This is a numeric variable that calculates the average amount of time spent per day walking or biking to and from a place. Update 12/2013: set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.

If days of activity and time spent at different intensity levels are not missing, then
$$\text{GPAQ_TRSPORT} = (\text{PAEA8} \times \text{PAEA9A_MIN})/7$$

Source variable(s): In a typical week, ...

PAEA8. Days walk or bicycle to get to and from places?

PAEA9_MIN. Time spent walking or bicycling for travel (min)

11.10 GPAQ_TOTAL (Total physical activity (min/day))

This is a numeric variable that shows the average amount of time doing some form of physical activity per day. Update 12/2013: set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.

If activity time for each domain is non-missing, then

$$\text{GPAQ_TOTAL} = \text{sum}(\text{GPAQ_WORK}, \text{GPAQ_TRSPORT}, \text{GPAQ_REC})$$

Source variable(s):

GPAQ_work. Work-related physical activity (min/day)

GPAQ_trsport. Transportation-related physical activity (min/day)

GPAQ_rec. Recreational physical activity (min/day)

11.11 GPAQ_TOTAL_MET (Total physical activity (MET-min/day))

This is a numeric variable that calculates the total amount of time spent doing some form of physical activity in a week. Update 12/2013: set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.

If activity time for each domain is non-missing, then

$$\begin{aligned} \text{GPAQ_TOTAL_MET} = & ((\text{PAEA2} \times \text{PAEA3A_MIN} \times 8) + (\text{PAEA5} \times \text{PAEA6A_MIN} \times 4) \\ & + (\text{PAEA8} \times \text{PAEA9A_MIN} \times 4) \\ & + (\text{PAEA12} \times \text{PAEA13A_MIN} \times 8) + (\text{PAEA16} \times \text{PAEA17A_MIN} \times 4))/7 \end{aligned}$$

Source variable(s):

PAEA2. Days per week of vigorous-intensity activities as part of your work

PAEA3A_MIN. Time (min) per day of vigorous-intensity activities at work

PAEA5. Days per week of moderate-intensity activities as part of your work

PAEA6A_MIN. Time (min) per day of moderate-intensity activities at work

PAEA8. Days per week of walk or bicycle to get to and from places

PAEA9A_MIN. Time (min) per day of walk or bicycle to get to and from places

PAEA12. Days per week of vigorous-intensity sports, fitness or recreational (leisure) activities

PAEA13A_MIN. Time (min) per day of vigorous-intensity sports, fitness or recreational (leisure) activities

PAEA16. Days per week of moderate-intensity sports, fitness or recreational (leisure) activities

PAEA17A_MIN. Time (min) per day of moderate-intensity sports, fitness or recreational (leisure) activities

11.12 GPAQ_MOD_WEEK (GPAQ - Total physical (moderate) activity (min/week))

This is a numeric variable that finds the average amount of time spent per week doing moderate physical activity.

$$\text{GPAQ_MOD_WEEK} = \text{GPAQ_TOTAL_MOD} * 7$$

Source variable(s):

GPAQ_TOTAL_MOD. Total physical (moderate) activity (min/day)

11.13 GPAQ_VIG_WEEK (GPAQ - Total physical (vigorous) activity (min/week))

This is a numeric variable that finds the average amount of time spent per week doing vigorous physical activity.

$$\text{GPAQ_VIG_WEEK} = \text{GPAQ_TOTAL_VIG} * 7$$

Source variable(s):

GPAQ_TOTAL_VIG. Total physical (vigorous) activity (min/day)

11.15 GPAQ_LEVEL (Physical activity level)

This is a categorical variable which assigns a value of 1 assigned to those with high total physical activity levels, a value of 2 to those with moderate total physical activity levels and a value of 3 to those with low total physical activity levels. Update 12/2013: corrected a typo in code such that participants who were actually "moderate" were being set to "low". Set participants to missing if they reported more than 1440 total minutes per day of active + sedentary. Level of total physical activity

Physical activity cutoff value	
High	• If vigorous PA (work+rec) $\geq$ 3 days & GPAQ_total_MET $\geq$ 1500 - OR • If moderate+vigorous PA $\geq$ 7 days & GPAQ_total_MET $\geq$ 3000
Moderate	• If vigorous PA (work+rec) $\geq$ 3 days & GPAQ_total_vig $\times$ 7 $\geq$ 60 min - OR • If moderate PA (work+transport+rec) $\geq$ 5 days & GPAQ_total_mod $\times$ 7 $\geq$ 150 min - OR • If moderate+vigorous PA $\geq$ 5 days & GPAQ_total $\times$ 7 $\geq$ 600 min
Low	If the value does not reach the criteria for either high or moderate levels of physical activity

Source variable(s):

PAEA2. Days per week of vigorous-intensity activities as part of your work

PAEA5. Days per week of moderate-intensity activities as part of your work

PAEA8. Days per week of walk or bicycle to get to and from places

PAEA12. Days per week of vigorous-intensity sports, fitness or recreational (leisure) activities

PAEA16. Days per week of moderate-intensity sports, fitness or recreational (leisure) activities

GPAQ_TOTAL_VIG. Total physical (vigorous) activity (min/day)

GPAQ_TOTAL_MET. Total physical activity (MET-min/day)

GPAQ_TOTAL_MOD. Total physical (moderate) activity (min/day)

GPAQ_TOTAL. Total physical activity (min/day)

11.16 GPAQ_PAG2008 (4-level Activity level per 2008 PA guidelines)

This is an ordinal variable that contains the level of physical activity per the 2008 PA guidelines (<http://www.health.gov/paguidelines/guidelines/default.aspx>). There are 4 mutually exclusive levels: high, medium, low or inactive based on all counts. Set participants to missing if they were missing GPAQ_MOD_WEEK and GPAQ_VIG_WEEK or if they reported more than 1440 total minutes per day of active + sedentary.

GPAQ_PAG2008	Definition from 2008 Physical Activity Guidelines (for adults)
4 = High activity	"more than the equivalent of 300 minutes of moderate-intensity physical activity a week" or more than 150 minutes of vigorous activity, or an equivalent combination of both. Activity should be performed in episodes of at least 10 minutes.
3 = Medium activity	"150 minutes to 300 (5 hours) minutes of moderate-intensity activity a week (or 75 to 150 minutes of vigorous-intensity physical activity a week)" or the equivalent combination of moderate and vigorous activity. Activity should be performed in episodes of at least 10 minutes.
2 = Low activity	"activity beyond baseline but fewer than 150 minutes (2 hours and 30 minutes) of moderate-intensity physical activity a week or the equivalent amount (75 minutes, or 1 hour and 15 minutes) of vigorous-intensity activity" or the equivalent combination of moderate and vigorous activity.
1 = Inactive	"no activity beyond baseline activities of daily living"

Note that all borderline cases are given the higher level of activity. If none of these criteria are met, GPAQ_PAG2008 is set to missing.

```
if GPAQ_MOD_WEEK>=300 | GPAQ_VIG_WEEK>=150 |
GPAQ_MV_WEEK>=300 then GPAQ_PAG2008=4;

else if 150<=gpaq_mod_week<300 | 75<=gpaq_vig_week<150 |
150<=gpaq_mv_week<300 then GPAQ_PAG2008=3;

else if 1<=gpaq_mod_week<150 | 1<=gpaq_vig_week<75 |
1<=gpaq_mv_week<150 then GPAQ_PAG2008=2;

else if (0<=gpaq_mod_week<1 & 0<=gpaq_vig_week<1) then
GPAQ_PAG2008=1;

else GPAQ_PAG2008=.;
```

Response format: 1 = Inactive
2 = Low Activity
3 = Medium Activity
4 = High Activity

Variable GPAQ_MV_WEEK is an intermediate variable created for the purposes of this definition only. It is defined as below:

$GPAQ_MV_WEEK = (GPAQ_TOTAL_MOD * 7) + (2 * (GPAQ_TOTAL_VIG * 7))$

Source variable(s):

GPAQ_MOD_WEEK. GPAQ - Total physical (moderate) activity (min/week)

GPAQ_VIG_WEEK. GPAQ - Total physical (vigorous) activity (min/week)

GPAQ_MV_WEEK. GPAQ - Total physical activity (min/week)

11.17 GPAQ_PAG2008YN (Meets 2008 activity level guidelines (1=Yes, 0=No))

This is a 0/1 categorical variable that determines whether or not high or medium activity level has been met based on GPAQ_PAG2008 (all counts). Update 12/2013: changes reflected from GPAQ_PAG2008

GPAQ_PAG2008YN = 1 if GPAQ_PAG2008 = 3 **or** GPAQ_PAG2008 = 4
Else GPAQ_PAG2008YN = 0

Response format: 0 = No
1 = Yes

Source variable(s):

GPAQ_PAG2008. Activity level per 2008 PA guidelines (categorical - 4 levels)

12. CLINICAL CHARACTERISTICS

12.1 HYPERTENSION (Hypertension (BP $\geq$ 140/90 and Med Use))

This is a 0/1 numeric variable. Hypertension is defined here if the systolic or diastolic BP is greater than or equal to 140/90 or if the participant is currently taking antihypertensive medications (scanned meds only). Participants without a blood pressure measurement and no medication use were assumed to be not hypertensive. This variable does NOT include self-report doctor's diagnosis of hypertension or self-report antihypertensive medication use.

5/2014 update: updates due to updated medication variables

Response format: 0=Not hypertensive (No)
1=Hypertensive (Yes)

Source variable(s):

SBPA5. Systolic blood pressure.

SBPA6. Diastolic blood pressure.

MED_ANTIHYPERT. Scanned medication was for treatment of hypertension.

12.2 HYPERTENSION2 (Hypertension using NHANES definition)

This is a 0/1 numeric variable. Hypertension is defined here using the following NHANES definition: if the systolic or diastolic BP is greater than or equal to 140/90 or if the participant self-reported as currently taking antihypertensive medications. Participants without a blood pressure measurement and no medication use were assumed to be not hypertensive.

Response format: 0=Not hypertensive (No)
1=Hypertensive (Yes)

Source variable(s):

SBPA5. Systolic blood pressure.

SBPA6. Diastolic blood pressure.

MUEA33D. Were any of the medications you took during the last 4 weeks for high blood pressure or hypertension?

12.3 HYPERTENSION2_C4 (4-level grouped hypertension)

This is a numeric variable that classifies participants into 4 categories based on hypertension and hypertensive medication use (including both self-reported and scanned medications): no hypertension, pre-hypertension, treated hypertension, and untreated hypertension.

HYPERTENSION2=0 and ((120 $\leq$ SPBA5<140) or (80 $\leq$ SPBA6<90) $\rightarrow$ 2

Other HYPERTENSION2=0 $\rightarrow$ 1

HYPERTENSION2=1 and (MUEA33D=1 OR MED_ANTIHYPERT=1) $\rightarrow$ 3

Other HYPERTENSION2=1 $\rightarrow$ 4

Missing HYPERTENSION2 $\rightarrow$ missing

Response format: 1 = No hypertension
2 = Pre-hypertension
3 = Treated hypertension
4 = Untreated hypertension

Source variable(s):

HYPERTENSION2. Hypertension using NHANES definition.

SBPA5. Systolic blood pressure.

SBPA6. Diastolic blood pressure.

MUEA33D. Self-reported medication for high blood pressure or hypertension.

MED_ANTIHYPERT. Scanned Antihypertensive medication use.

12.4 HYPERTENSION2_AHA (using new ACC/AHA Guidelines definition)

This is a 0/1 numeric variable. Hypertension is defined here using new ACC/AHA Guidelines definition (New ACC/AHA Guideline reference: <https://www.ahajournals.org/doi/epub/10.1161/HYP.0000000000000065>): if the systolic or diastolic BP is greater than or equal to 130/80 or if the participant self-reported as currently taking antihypertensive medications. Participants without a blood pressure measurement and no medication use were assumed to be not hypertensive.

Response format: 0=Not hypertensive (No)
1=Hypertensive (Yes)

Source variable(s):

SBPA5. Systolic blood pressure.

SBPA6. Diastolic blood pressure.

MUEA33D. Were any of the medications you took during the last 4 weeks for high blood pressure or hypertension?

12.5 HYPERTENSION2_AHA_C5 (5-level using new ACC/AHA Guidelines)

This is a numeric variable that describes the stages of hypertension based on new ACC/AHA definition at 5 levels (None/Elevated/treated/untreated Stage 1/ untreated Stage 2). This definition takes into account systolic and diastolic blood pressure measured, and self-reported medications for high blood pressure or hypertension.

IF HYPERTENSION2_AHA=0 and ((120<=SBPA5<130) AND SBPA6<80) → 2

Otherwise HYPERTENSION2_AHA=0 → 1

HYPERTENSION2_AHA=1 and MUEA33D=1 → 3

Otherwise if HYPERTENSION2_AHA=1 and ((SBPA5>=140) OR (SBPA6>=90)) → 5

Otherwise if HYPERTENSION2_AHA=1 → 4

Missing HYPERTENSION2_AHA → missing

Response format: 1=Not hypertension
2=Elevated hypertension
3=Treated hypertension
4=Untreated stage 1 hypertension
5=Untreated stage 2 hypertension

Source variable(s):

Hypertension2. Hypertension using new ACC/AHA definition.

SBPA5. Systolic blood pressure.

SBPA6. Diastolic blood pressure.

MUEA33D. Self-reported medication for high blood pressure or hypertension.

12.6 HYPERT_AWARENESS (Awareness of hypertension - NHANES)

This is a 0/1 numeric variable and is defined by the self-report of a doctor's diagnosis of hypertension (MHEA1=1) for those who were defined as having hypertension using the NHANES definition (HYPERTENSION2=1) only. Note that participants who were missing MHEA1 were set to 0 (not aware).

Response format: 0=Not aware (No), or missing MHEA1
1=Aware (Yes)

Source variable(s):

MHEA1. Has a doctor ever said that you have high blood pressure or hypertension?

HYPERTENSION2. Hypertension using NHANES definition

12.7 HYPERT_TREATMENT (Treatment of hypertension - NHANES)

This is a 0/1 numeric variable and is defined by the participant's self-report awareness of hypertension (HYPERT_AWARENESS) and self-report use of antihypertensive medications (MUEA33D=1) for those who were defined as having hypertension using the NHANES definition (HYPERTENSION2=1) only. Note that participants who had missing data for hypertensive medication use were set to 0 (no medication use).

Response format: 0=No (or missing data for) antihypertensive med use (No)
1=Hypertensive med use (Yes)

Source variable(s):

MUEA33D. Were any of the medications you took during the last 4 weeks for high blood pressure or hypertension?

HYPERTENSION2. Hypertension using NHANES definition

HYPERT_AWARENESS. Awareness of Hypertension - NHANES

12.8 HYPERT_CONTROL (Control of hypertension - NHANES)

This is a 0/1 numeric variable and is defined for participants whose blood pressure are below the cutoff of 140/90 (both systolic and diastolic must be less than the cutoff values) for those who were defined as having hypertension using the NHANES definition (HYPERTENSION2=1) only and also indicated whether/not they were receiving treatment for hypertension (HYPERT_TREATMENT= 0 or 1).

Response format: 0=Hypertension not under control (No)
1=Hypertension is under control (Yes)

Source variable(s):

SBPA5. Systolic blood pressure.

SBPA6. Diastolic blood pressure.

HYPERTENSION2. Hypertension using NHANES definition

HYPERT_TREATMENT. Treatment of Hypertension – NHANES

12.9 HOMA_IR (HOMA index of insulin resistance)

This is a numeric variable that defined the Homeostasis model assessment of insulin resistance as follows (LABA70 and INSULIN_FAST must be non-missing):

$$\text{HOMA_IR} = \text{LABA70} * \text{INSULIN_FAST} / 405.$$

Source variable(s):

LABA70. Fasting Glucose in mg/dL

INSULIN_FAST. Insulin, fasting (converted to mU/L)

12.10 HOMA_B (HOMA index of Beta Cell Function)

This is a numeric variable that defined the Homeostasis model assessment of insulin resistance as follows:

If $\text{LABA70} \leq 63$ then set HOMA_B as missing,
Otherwise
$$\text{HOMA_B} = 360 * \text{INSULIN_FAST} / (\text{LABA70} - 63).$$

Source variable(s):

LABA70. Fasting Glucose in mg/dL

INSULIN_FAST. Insulin, fasting (converted to mU/L)

12.11 DIABETES_SELF (Diabetes – self-report only)

This is a 0/1 numeric variable that indicates a self-report of diabetes.

If the MHEA form is missing then Diabetes_self = .
If mhea16=1 and mhea16b=(0 or Missing) then Diabetes_self =1.
Else if mhea16 is missing then Diabetes_Self= missing.
Otherwise Diabetes_Self=0

5/2014 update: revised to code missing MHEA16 as no and MHEA16=.Q as missing instead of zero

Response format: 0 = No
1 = Yes

Source variable(s):

MHEA16. Has a doctor ever said that you have diabetes?

MHEA16b. Was this during pregnancy?

12.12 DIABETES_LAB (3-level Diabetes – Lab ADA Guideline)

This is a numeric variable that describes the stages of diabetes. This definition takes into account serum glucose levels adjusted for fasting time and, if available, post-OGTT glucose levels, and A1C percentages. This variable does NOT include any medication use variables.

This variable is missing if fasting glucose, post-OGTT, and A1C are missing.

If (fasting time > 8 AND fasting glucose $\geq$ 126 mg/dL) or (fasting time $\leq$ 8 AND fasting glucose $\geq$ 200 mg/dL) or (post-OGTT glucose $\geq$ 200 mg/dL) or (A1C $\geq$ 6.5%) or then DIABETES_Lab=3

Else if (fasting time > 8 AND fasting glucose in range 100-125 mg/dL) or (post-OGTT glucose in range 140-199 mg/dL) or (5.7% $\leq$ A1C<6.5%) then DIABETES_Lab=2

Else DIABETES_Lab=1

Response format: 1 = normal glucose regulation
2 = impaired glucose tolerance
3 = diabetes

Source variable(s):

LABA70. Fasting glucose (mg/dL)

LABA71. Glucose, post OGTT (mg/dL)

LABA72. % Glycosylated Hemoglobin (A1C)

FASTTIME. Elapsed time (hrs.) between the time the participant last consumed anything and the blood draw

12.13 DIABETES_LAB_NOOGTT (3-level – Lab ADA Guideline no OGT)

This is a numeric variable that describes the stages of diabetes. This definition takes into account serum glucose levels adjusted for fasting time and A1C percentages. This variable does NOT include any medication use variables.

This variable is missing if fasting glucose and A1C are missing.

If (fasting time > 8 AND fasting glucose $\geq$ 126 mg/dL) or (fasting time $\leq$ 8 AND fasting glucose $\geq$ 200 mg/dL) or (A1C $\geq$ 6.5%) or then DIABETES_Lab=3

Else if (fasting time > 8 AND fasting glucose in range 100-125 mg/dL) or (5.7% $\leq$ A1C<6.5%) then DIABETES_Lab=2

Else DIABETES_Lab=1

Response format: 1 = normal glucose regulation
2 = impaired glucose tolerance
3 = diabetes

Source variable(s):

LABA70. Fasting glucose (mg/dL)

LABA72. % Glycosylated Hemoglobin (A1C)

FASTTIME. Elapsed time (hrs.) between the time the participant last consumed anything and the blood draw

12.14 SELFMED_ANTIDIAB (Anti-Diabetics Medication used – Self-reported)

This is a 0/1 numeric variable and is defined by the participant's self-report use of anti-diabetics medications (Yes if MUEA33C=1, No if MUEA33c=0, after reset missing to zero when confirmed not taking any medications, MUEA2=1). Note that participants who had missing data or unsure (MUEA33c=9) for antidiabetics medication use were set to missing.

Response format: 0=No antidiabetics med reported (No)
1=Self-reported use antidiabetics med (Yes)

Source variable(s):

MUEA33C. Were any of the medications you took during the last 4 weeks for high blood sugar or diabetes?

MUEA2. Not bringing any of the medications you took during the last 4 weeks, because you forgot/unable to, or took no medication?

12.15 DIABETES1 (3-level Diabetes, definition 1)

This is a numeric variable that describes the stages of diabetes. This definition takes into account serum glucose levels adjusted for fasting time and, if available, post-OGTT glucose levels and scanned/transcribed anti-diabetic medication use.

This variable is missing if fasting glucose, post-OGTT and Med_Antidiab are missing. Note that participants with no glucose lab data and no anti-diabetic medication use were assumed to have normal glucose regulation.

If (fasting time > 8 AND fasting glucose $\geq$ 126) or (fasting time $\leq$ 8 AND fasting glucose $\geq$ 200) or (post-OGTT glucose $\geq$ 200) or (Med_Antidiab=1) then DIABETES1=3
Else if (fasting time > 8 AND fasting glucose in range 100-125) or (post-OGTT glucose in range 140-199) then DIABETES1=2
Else DIABETES1=1

5/2014 update: updates due to updated medication variables

Response format: 1 = normal glucose regulation
2 = impaired glucose tolerance
3 = diabetes

Source variable(s):

LABA70. Fasting glucose (mg/dL)

LABA71. Glucose, post OGTT (mg/dL)

FASTTIME. Elapsed time (hrs.) between the time the participant last consumed anything and the blood draw

Med_Antidiab. Scanned/transcribed anti-diabetic medication use

12.16 DIABETES2 (3-level Diabetes – ADA Guideline plus scanned med, definition 2)

This is a numeric variable that describes the stages of diabetes based on American Diabetes Association definition (*Diabetes Care* 2010;33:S62-69). This definition takes into account serum glucose levels adjusted for fasting time and, if available, post-OGTT glucose levels, A1C percentages, and scanned/transcribed anti-diabetic medication use.

This variable is missing if fasting glucose, post-OGTT, A1C, and Med_Antidiab are missing. Note that participants with no glucose lab data and no anti-diabetic medication use were assumed to have normal glucose regulation.

If (fasting time > 8 AND fasting glucose $\geq$ 126 mg/dL) or (fasting time $\leq$ 8 AND fasting glucose $\geq$ 200 mg/dL) or (post-OGTT glucose $\geq$ 200 mg/dL) or (A1C $\geq$ 6.5%) or (Med_Antidiab=1) then DIABETES2=3
Else if (fasting time > 8 AND fasting glucose in range 100-125 mg/dL) or (post-OGTT glucose in range 140-199 mg/dL) or (5.7% $\leq$ A1C<6.5%) then DIABETES2=2
Else DIABETES2=1

Response format: 1 = Non-diabetic
2 = Pre-diabetic
3 = Diabetic

Source variable(s):

LABA70. Fasting glucose (mg/dL)

LABA71. Glucose, post OGTT (mg/dL)

LABA72. % Glycosylated Hemoglobin (A1C)

FASTTIME. Elapsed time (hrs.) between the time the participant last consumed anything and the blood draw

Med_Antidiab. Scanned/transcribed anti-diabetic medication use

12.17 DIABETES2_INDICATOR (Diabetes indicator (ADA), definition 2)

This is a 0/1 indicator variable based on the Diabetes variable (version 2) stages of diabetes from the American Diabetes Association (*Diabetes Care* 2010;33:S62-69).

Missing, if Diabetes2 is missing;
1, if Diabetes2=3;
0, otherwise.

Response format: 0 = Non-diabetic (No)
1 = Diabetic (Yes)

Source variable(s):

DIABETES2. Diabetes2 (3-level grouped Diabetes – ADA)

12.18 DIABETES3 (3-level Diabetes – ADA Guideline plus self-reported diagnosis)

This is a numeric variable that describes the stages of diabetes based on American Diabetes Association definition (*Diabetes Care* 2010;33:S62-69) with an additional self-report criterion. This definition takes into account serum glucose levels adjusted for fasting time and, if available, post-OGTT glucose levels, A1C percentages, and self-report of diabetes. 5/2014 update: updates due to updated DIABETES_SELF

If Diabetes_Self=1 or Diabetes_Lab=3 then Diabetes3=3.

Else if Diabetes_Lab=2 then Diabetes3=2.

Else if Diabetes_Self=missing and Diabetes_Lab=(1 or missing) then Diabetes3=missing.
Otherwise Diabetes3=1.

Response format: 1 = Non-diabetic
2 = Pre-diabetic
3 = Diabetic

Source variable(s):

Diabetes_Lab. 3-level grouped Diabetes – Lab criterion only

Diabetes_Self. Diabetes – self-report only

12.19 DIABETES3A (3-level Diabetes - ADA lab no OGTT plus self-reported diagnosis, definition 3A)

This is a numeric variable that describes the stages of diabetes based on ADA lab criteria (serum glucose levels adjusted for fasting time, post-OGTT glucose levels if available, and A1C percentages) and self-reported diagnosis. Since post-OGTT glucose levels are not available at Visit 3, the letter 'A' is added to the definition name to highlight this distinction. This diabetes variable was defined by the algorithm strictly. Per Diabetes SIG decision and memo, Diabetes classifications at previous visits are not part of definition specs and the HCHS/SOL study lab classification was not considered as a formal diagnosis.

If Diabetes_Self=1 or Diabetes_Lab_NoOGTT=3 then Diabetes3A=3.
Else if Diabetes_Lab_NoOGTT=2 then Diabetes3A=2.
Else if Diabetes_Self=missing and Diabetes_Lab_NoOGTT=(1 or missing) then Diabetes3A=missing.
Otherwise Diabetes3A=1.

Response format: 1=Non-diabetic
2=Pre-diabetic
3=Diabetic

Source variable(s):

DIABETES_SELF. Diabetes – self-report only
SELFMED_ANTIDIAB. Diabetes – self-reported med use
DIABETES_LAB_NOOGTT. 3-level grouped diabetes – ADA lab no OGTT

12.20 DIABETES_C4 (4-level Diabetes (none/pre/treated/untreated))

This 4 category variable assesses diabetes using DIABETES3 and medication use based on MUEA33C (self-reported medication for high blood sugar or diabetes) and MED_ANTIDIAB (scanned antidiabetic medication). 5/2014 update: updates due to updated DIABETES3 and medication variables.

DIABETES3=1 → 1
DIABETES3=2 → 2
DIABETES3=3 and (MUEA33C=1 or MED_ANTIDIAB=1) → 3
Other DIABETES3=3 → 4
Missing DIABETES3 → missing

Response format: 1=No diabetes
2=Pre-diabetes
3=Treated diabetes
4=Untreated diabetes

Source variable(s):

DIABETES3. 3-level grouped Diabetes - includes self-report
MUEA33C. Self-reported medication for high blood sugar or diabetes
MED_ANTIDIAB. Scanned antidiabetics

12.21 DIAB_DIAG (Diabetes diagnosis)

This is a categorical variable that describes the diagnoses of diabetes and whether it was diagnosed before baseline, not diagnosed before baseline but possibly detected from lab data, or as non-diabetic. Note that participants who were diabetic but had missing data for self-reported diabetes were assumed to be undiagnosed diabetics. 5/2014 update: updates due to updated DIABETES_SELF and DIABETES3.

If Diabetes_Self=1 then Diab_diag=1.

Else if Diabetes_Self not =1 and Diabetes3=3 then Diab_diag=2.

Else if Diabetes_Self=missing and Diabetes3= missing then Diab_diag=missing.

Otherwise Diab_diag=3

Response format: 1 = Diagnosed diabetic before baseline
2 = Undiagnosed diabetic (or missing self-report data)
3 = Non-diabetic

Source variable(s):

Diabetes3. 3-level grouped Diabetes – includes self-report

Diabetes_Self. Diabetes – self-report only

12.22 DIAB_FAMHIST (Family history of diabetes)

This is a 0/1 numeric variable that describes the family history of diabetes.

If (MHEA16e and MHEA16f and MHEA16g) all missing then Diab_FamHist= missing.

Else If MHEA16e=1 or MHEA16f=1 or MHEA16g=1 then Diab_FamHist=1;

Otherwise Diab_FamHist=0.

Response format: 0=No family history of diabetes (No)
1=Family history of diabetes (Yes)

Source variable(s):

MHEA16e. Mother had diabetes

MHEA16f. Father had diabetes

MHEA16g. Brother/sister had diabetes

12.23 DM_AWARE (Diabetes awareness)

This is a 0/1 numeric variable that describes the awareness of diabetes. Defined only for self-reported diabetics.

if DIABETES_SELF missing AND (MUEA33c missing or =9) then DM_Aware=.;

else if DIABETES_SELF=1 OR MUEA33c=1 then DM_Aware=1;

else DM_Aware=0;

Set to missing for nondiabetics.

Response format: 0=Not aware of diabetes
1=Aware of diabetes

Source variable(s):

MUEA33c. High blood sugar or diabetes – self-reported med use

DIABETES_SELF. Diabetes – self-report only

DIABETES3. Diabetes ADA with self-reported DM

12.24 DM_CONTROL (Diabetes controlled: classified at baseline and A1C < 7%)

This is a 0/1 numeric variable that indicates controlled diabetes. Defined only for diabetics.
5/2014 update: set to missing for nondiabetics.

```
if DIABETES3^=3 then DM_CONTROL=.;  
else if DIABETES3=3 & .Z<lab72<7 then DM_CONTROL=1;  
else DM_CONTROL=0;
```

Response format: 0=Uncontrolled diabetes
 1=Controlled diabetes

Source variable(s):

Diabetes3. 3-level grouped Diabetes – includes self-report

LABA72. A1C

12.25 DIABETES4 (3-level Diabetes – ADA guideline plus self-reported med, definition 4)

This is a numeric variable that describes the stages of diabetes based on American Diabetes Association definition (*Diabetes Care* 2010;33:S62-69). This definition uses same definition as Diabetes2, except using self-reported medication instead of scanned medication.

```
If selfmed_antidiab=1 | diabetes_lab=3 then DIABETES4=3;  
Else if diabetes_lab=2 then DIABETES4=2;  
Else if selfmed_antidiab=. & diabetes_lab=. then DIABETES4=.;  
Else DIABETES4=1;
```

Response format: 1 = Non-diabetic
 2 = Pre-diabetic
 3 = Diabetic

Source variable(s):

Diabetes_lab. ADA Guideline Lab criteria based on FPG, OGTT and A1c

SelfMed_Antidiab. Self-reported use antidiabetics medication.

12.26 DIABETES4A (3-level Diabetes, ADA lab no OGTT plus self-reported med use, definition 4A)

This is a numeric variable that describes the stages of diabetes based on ADA lab criteria (taking into account serum glucose levels adjusted for fasting time, post-OGTT glucose levels if available, and A1C percentages) and self-reported antidiabetics medication use. Since post-OGTT glucose levels are not available at Visit 3, the letter 'A' is added to the definition name to highlight this distinction. This diabetes variable was defined by the algorithm strictly. Per Diabetes SIG decision and memo, Diabetes classifications at previous visits are not part of definition specs and the HCHS/SOL study lab classification was not considered as a formal diagnosis.

If SELFMED_ANTIDIAB=1 or DIABETES_LAB_NOOGTT=3 then DIABETES4A=3;
Else if DIABETES_LAB_NOOGTT=2 then DIABETES4A=2;
Else DIABETES4A=1;
If all source variables are missing then DIABETES4A is missing;

Response format: 1=Non-diabetic
2=Pre-diabetic
3=Diabetic

Source variable(s):

SELFMED_ANTIDIAB. Diabetes – self-reported med use

DIABETES_LAB_NOOGTT. 3-level grouped diabetes – ADA lab no OGTT

12.27 DIABETES5 (3-level Diabetes-ADA Guideline plus self-reported diabetes diagnosis or self-reported med, definition 5)

This is a numeric variable that describes the stages of diabetes based on ADA Guideline lab criteria, Self-reported diabetes diagnosis or Self-reported antidiabetics med.

If diabetes_self=1 | selfmed_antidiab=1 | diabetes_lab=3 then DIABETES5=3;
Else if diabetes_lab=2 then DIABETES5=2;
Else if diabetes_self=. & selfmed_antidiab=. & diabetes_lab=. then DIABETES5=.;
Else DIABETES5=1;

Response format: 1 = Non-diabetic
2 = Pre-diabetic
3 = Diabetic

Source variable(s):

Diabetes_lab. ADA Guideline Lab criteria based on FPG, OGTT and A1c

Diabetes_Self. Self-reported diabetes diagnosis.

SelfMed_Antidiab. Self-reported use antidiabetics medication.

12.28 DIABETES5A (3-level Diabetes - ADA Guideline no OGTT plus self-reported diabetes diagnosis or self-reported Med, definition 5A)

This is a numeric variable that describes the stages of diabetes based on ADA lab criteria (taking into account serum glucose levels adjusted for fasting time, post-OGTT glucose levels if available, and A1C percentages) and self-reported antidiabetics medication use. Since post-OGTT glucose levels are not available at Visit 3, the letter 'A' is added to the definition name to highlight this distinction. This diabetes variable was defined by the algorithm strictly. Per Diabetes SIG decision and memo, Diabetes classifications at previous visits are not part of definition specs and the HCHS/SOL study lab classification was not considered as a formal diagnosis.

```
If diabetes_self=1 | selfmed_antidiab=1 | diabetes_lab_noogtt=3 then DIABETES5A=3;
Else if diabetes_lab_noogtt=2 then DIABETES5A=2;
Else if diabetes_self=. & selfmed_antidiab=. & diabetes_lab_noogtt=. then
DIABETES5A=.;
Else DIABETES5A=1;
```

Response format: 1 = Non-diabetic
2 = Pre-diabetic
3 = Diabetic

Source variable(s):

Diabetes_lab_NoOGTT. ADA Guideline Lab criteria no OGTT

Diabetes_Self. Self-reported diabetes diagnosis.

SelfMed_Antidiab. Self-reported use antidiabetics medication.

12.29 PRECHD_NO_ANGINA (Prevalent cardiovascular heart disease)

This is a 0/1 variable that combines ECG reports of possible old MI as well as self-report of heart attack, or procedure (angioplasty, stent, bypass).

```
if ecga267^=' ' | ecga268^=' ' | mhea4=1 | mhea9=1 then PreCHD_No_Angina=1;
else if ecga267=' ' & ecga268=' ' & mhea4 in(.,.Q) & mhea9 in(.,.Q) then
PreCHD_No_Angina=.;
else PreCHD_No_Angina=0;
```

Response format: 0=No Prevalent CHD (No)
1=Prevalent CHD (Yes)

Source variable(s):

ECGA267. Major Q Wave abnormalities [old Myocardial Infarction (MI)]

ECGA268. Minor Q, QS waves with ST, T abnormalities (possible old MI)

MHEA4. Has a doctor ever said that you had a heart attack?

MHEA9. Have you had a balloon angioplasty, a stent, or bypass surgery to the arteries in your heart to improve the blood flow to your heart?

12.30 PRECHD_ANGINA (Prevalent cardiovascular heart disease, including angina)

This is a 0/1 variable that combines ECG reports of possible old MI as well as self-report of angina, heart attack, or procedure (angioplasty, stent, bypass).

```
if ecga267^=' ' | ecga268^=' ' | mhea3=1 | mhea4=1 | mhea9=1 then PreCHD_Angina=1;  
else if ecga267=' ' & ecga268=' ' & mhea3 in(.,.Q) & mhea4 in(.,.Q) & mhea9 in(.,.Q) then  
PreCHD_Angina =.;  
else PreCHD_Angina =0;
```

Response format: 0=No Prevalent CHD (No)
1=Prevalent CHD (Yes)

Source variable(s):

ECGA267. Major Q Wave abnormalities [old Myocardial Infarction (MI)]

ECGA268. Minor Q, QS waves with ST, T abnormalities (possible old MI)

MHEA3. Has a doctor ever said that you have angina?

MHEA4. Has a doctor ever said that you had a heart attack?

MHEA9. Have you had a balloon angioplasty, a stent, or bypass surgery to the arteries in your heart to improve the blood flow to your heart?

12.31 EVER_ANGINA_RELATIVE (At least one relative had report of angina)

This is a 0/1 variable that combines self-report of angina for immediate family relatives, i.e. mother, father, and sibling(s).

If MHEA3 is non-missing then:

IF MHEA3b, MHEA3c, and MHEA3d are all missing then Ever_angina_relative = missing.

IF MHEA3b=1 or MHEA3c=1 or MHEA3d=1 then Ever_angina_relative = 1

Otherwise Ever_angina_relative = 0.

Response format: 0=No family history of angina (No)
1=Family history of angina (Yes)

Source variable(s):

MHEA3. Has a doctor ever said that you have angina?

MHEA3b. Has a doctor ever said that these relatives have angina: Mother

MHEA3c. Has a doctor ever said that these relatives have angina: Father

MHEA3d. Has a doctor ever said that these relatives have angina: Brother(s) or Sister(s)

12.32 EVER_MI_RELATIVE (At least one relative had report of Heart Attack)

This is a 0/1 variable that combines self-report of heart attack for immediate family relatives, i.e. mother, father, and sibling(s).

If MHEA4 is non-missing then:

IF MHEA4B1, MHEA4C1, and MHEA4D1 are all missing then Ever_MI_relative = missing.

IF MHEA4B1=1 or MHEA4C1=1 or MHEA4D1=1 then Ever_MI_relative = 1

Otherwise Ever_MI_relative = 0.

Response format: 0=No family history of heart attack (No)
1=Family history of heart attack (Yes)

Source variable(s):

MHEA4. Has a doctor ever said that you have had a heart attack?

MHEA4B1. Has a doctor ever said that these relatives have had a heart attack: Mother

MHEA4C1. Has a doctor ever said that these relatives have had a heart attack: Father

MHEA4D1. Has a doctor ever said that these relatives have had a heart attack: Brother(s)
or Sister(s)

12.33 EVER_CABG_RELATIVE (At least one relative had report of CABG or angioplasty)

This is a 0/1 variable that combines self-report of CABG or angioplasty for immediate family relatives, i.e. mother, father, and sibling(s).

If MHEA9 is non-missing then:

IF MHEA9a, MHEA9b, and MHEA9c are all missing then Ever_CABG_relative = missing.

IF MHEA9a=1 or MHEA9b=1 or MHEA9c=1 then Ever_CABG_relative = 1

Otherwise Ever_angina_relative = 0.

Response format: 0=No family history of CABG or angioplasty (No)
1=Family history of CABG or angioplasty (Yes)

Source variable(s):

MHEA9. Have you had a balloon angioplasty, a stent, or bypass surgery to the arteries in your heart to improve the blood flow to your heart?

MHEA9a. Has a doctor ever said that these relatives have CABG or angioplasty: Mother

MHEA9b. Has a doctor ever said that these relatives have CABG or angioplasty: Father

MHEA9c. Has a doctor ever said that these relatives have CABG or angioplasty:
Brother(s) or Sister(s)

12.34 FH_CHD (Family history of CHD including self-reported angina)

This is a 0/1 variable that combines the self-reports of angina, heart attack, or procedure (angioplasty, stent, bypass) for immediate family relatives, i.e. mother, father, and sibling(s).

IF Ever_angina_relative, Ever_MI_relative, and Ever_CABG_relative are all missing then FH_CHD = missing.

Else if Ever_angina_relative=1 or Ever_MI_relative=1 or Ever_CABG_relative=1 then FH_CHD = 1.

Otherwise FH_CHD = 0

Response format: 0=No family history of CHD (No)
1=Family history of CHD (Yes)

Source variable(s):

Ever_Angina_relative. At least one relative had report of angina

Ever_MI_relative. At least one relative had report of Heart Attack

Ever_CABG_relative. At least one relative had report of CABG or angioplasty

12.35 CHD_SELF (Self-reported CHD)

This is a 0/1 variable that combines the self-reports of coronary revascularization or heart attack.

```
If mhea4=1 | mhea9=1 then CHD_SELF=1;  
Else if mhea4 in(.,.Q) & mhea9 in(.,.Q) then CHD_SELF=.;  
Else CHD_SELF=0;
```

Response format: 0=No CHD (No)
1= CHD (Yes)

Source variable(s):

MHEA4. Has a doctor ever said that you have had a heart attack?

MHEA9. Have you ever had a CABG, stent, or balloon angioplasty?

12.36 CDCR (Cerebrovascular disease or carotid revascularization, self-report)

This is a 0/1 variable that checks for medical history of stroke, mini-stroke or TIA, or balloon angioplasty or surgery to the arteries in the neck, as follows:

```
If MHEA10=1 | MHEA11=1 | MHEA12=1 then CDCR=1;  
Else if MHEA10≤.Z & MHEA11≤.Z & MHEA12≤.Z then CDCR=.;  
Else CDCR=0;
```

Response format: 0 = No CDCR (No)
1 = CDCR (Yes)

Source variable(s):

MHEA10. Has a doctor ever said that you had a stroke?

MHEA11. Has a doctor ever said that you had a mini-stroke or TIA (transient ischemic attack)?

MHEA12. Have you had a balloon angioplasty or surgery to the arteries of your neck to prevent or correct a stroke?

12.37 METS_NCEP (Metabolic syndrome – NCEP-ATP III)

This is a 0/1 numeric variable with 1=metabolic syndrome present and 0=no metabolic syndrome. 5/2014 update: revised to use SEX instead of PIEA1.

The Adult Treat Panel III (ATP III) of the National Cholesterol Education Program (NCEP) issued an evidence-based set of guidelines on cholesterol management in 2001 (*JAMA* 2001;285:2486-97). The following definition of metabolic syndrome is based on the updated ATP III NCEP guidelines (*Circulation* 2004;109:433-438, *Circulation* 2004;110:227-239).

The diagnosis of metabolic syndrome is made when three or more of the following risk factors are present:

Elevated BP plusMeds: SBPA5 $\geq$ 130 or SBPA6 $\geq$ 85 mm Hg or use of anti-hypertension medications (Med_Antihypert=1) (note: requires non-missing SPB, DBP, or medication use).

High Trig: LABA67 $\geq$ 150 mg/dL

Low HDL: LABA68 < 40 mg/dl (men), LABA68 < 50 mg/dl (women)

IFG NCEP: LABA70 $\geq$ 100 mg/dl or use of anti-diabetic medications (MED_ANTIDIAB=1)

Abdominal obesity NCEP: ANTA10A $\geq$ 102 cm (men), ANTA10A $\geq$ 88 cm (women)

First, each of the components is evaluated (0/1 or missing).

If 3 or more components are missing components, then METS_NCEP should be missing.

Otherwise, METS_NCEP = 0 if sum of five components is 0, 1, 2

= 1 if sum of five components is 3, 4, 5

Source variable(s):

SEX. Sex (5/2014 update: revised to use SEX instead of PIEA1)

ANTA10A. Girth - waist

LABA67. Triglycerides (mg/dL)

LABA68. HDL cholesterol (mg/dL)

LABA70. Fasting glucose (mg/dL)

SBPA5. Systolic blood pressure

SBPA6. Diastolic blood pressure

MED_ANTIHYPERT. Anti-hypertension medication use

MED_ANTIDIAB. Anti-diabetic medication use

12.38 METS_NCEP2 (Metabolic syndrome w/ self-reported meds – NCEP-ATP III)

This is a 0/1 numeric variable with 1=metabolic syndrome present and 0=no metabolic syndrome. This variable incorporates the original METS_NCEP definition with changes to include self-reported medication use instead of scanned/transcribed medication use.

The diagnosis of metabolic syndrome is made when three or more of the following risk factors are present (*indicates variable substitutions from original definition):

*Elevated BP SelfMeds: SBPA5 $\geq$ 130 or SBPA6 $\geq$ 85 mm Hg or self-report use of anti-hypertension medications (MUEA33D=1) (note: requires non-missing SPB, DBP, or self-report medication use).

High Trig: LABA67 $\geq$ 150 mg/dL

Low HDL: LABA68 < 40 mg/dl (men), LABA68 < 50 mg/dl (women)

*IFG NCEP SelfMeds: LABA70 $\geq$ 100 mg/dl or self-report use of anti-diabetic medications (MUEA33C=1)

Abdominal obesity NCEP: ANTA10A $\geq$ 102 cm (men), ANTA10A $\geq$ 88 cm (women)

First, each of the components is evaluated (0/1 or missing).

If 3 or more components are missing components, then METS_NCEP2 should be missing.

Otherwise, METS_NCEP2 = 0 if sum of five components is 0, 1, 2

= 1 if sum of five components is 3, 4, 5

Source variable(s):

SEX. Sex (12/2019 update: revised to use SEX instead of PIEA1)

ANTA10A. Girth - waist

LABA67. Triglycerides (mg/dL)

LABA68. HDL cholesterol (mg/dL)

LABA70. Fasting glucose (mg/dL)

SBPA5. Systolic blood pressure

SBPA6. Diastolic blood pressure

MUEA33C. Self-report anti-diabetic medication use

MUEA33D. Self-report anti-hypertensive medication use

12.39 METS_IDF (Metabolic syndrome – IDF)

This is a 0/1 numeric variable with 1=metabolic syndrome present and 0=no metabolic syndrome. 5/2014 update: corrected code to require abdominal obesity.

The International Diabetes Federation (IDF) provides the following definition of metabolic syndrome. The IDF definition emphasizes central obesity because it is “independently” associated with each of the other metabolic syndrome components, as well as insulin resistance. Another difference between the IDF and NCEP-ATP III definitions is the recommended use of ethnic-specific waist girth thresholds (thresholds for Asian populations are lower than those for European and North American groups). Further, the ATP III definition includes 110 mg/dL cutoff for fasting plasma glucose (although the American Diabetes Association recently decided to lower the threshold for impaired fasting glucose to 100 mg/dL).

The IDF diagnosis of metabolic syndrome is made:

- a) Abdominal obesity IDF: ANTA10A $\geq$ 94 (men), ANTA10A $\geq$ 80 cm (women)
- b) Plus any 2 of the following:
 - High Trig plusMeds: LABA67 $\geq$ 150 mg/dL or treatment for this lipid abnormality (Med_Fibares_Nicoacid =1)

- Low HDL plusMeds: LABA68 < 40 mg/dL (men), LABA68 < 50 mg/dL (women) or specified treatment for this lipid abnormality (Med_Fibares_Nicoacid =1)
- Elevated BP plusMeds: SBPA5 ≥ 130 or SBPA6 ≥ 85 mm Hg or use of anti-hypertension medications (Med_Antihypert =1) (note: requires non-missing SPB, DBP, or medication use).
- IFG IDF: LABA70 ≥ 100 mg/dl, or previous diagnosis of diabetes (MHEA16=1), or use of anti-diabetic medications (Med_Antidiab=1).

First, each of the components is evaluated (0/1 or missing).

If abdominal obesity is missing or 2 or more or the other components are missing components, then METS_IDF should be missing.

Otherwise, METS_IDF = 0 if abdominal obesity is 0 or abdominal obesity is 1 and sum of four remaining components is 0 or 1.
 = 1 if abdominal obesity is 1 and the sum of the four remaining components is 2, 3, or 4.

Source variable(s):

SEX. Sex (5/2024: revised to use SEX instead of PIEA1)

MHEA16. Has a doctor ever said that you have diabetes (high sugar in blood or urine)

ANTA10A. Girth – waist

LABA67. Triglycerides (mg/dL)

LABA68. HDL cholesterol (mg/dL)

LABA70. Fasting glucose (mg/dL)

SBPA5. Systolic blood pressure

SBPA6. Diastolic blood pressure

MED_FIBARES_NICOACID. Fibric/Nicotinic Acids (trt of TG and HDL)

MED_ANTIDIAB. Anti-diabetic medication use

MED_ANTIHYPERT. Anti-hypertensive medication use

12.40 MI_ECG (MI defined by ECG results)

This is a 0/1 variable that checks some ECG measures to detect possible old MIs. Participants without ECGs are set to missing.

```
else if ECGA267 ne '' | ECGA268 ne '' then MI_ECG=1;
else if ECGA267 = '' & ECGA268 = '' then MI_ECG=0;
```

Source variable(s):

ECGA267. Major Q Wave abnormalities [old Myocardial Infarction (MI)]

ECGA268. Minor Q,QS waves with ST,T abnormalities (possible old MI)

12.41 STROKE (Prevalent stroke, self-report)

This is a 0/1 variable that checks for a self-report medical history of stroke as follows:

STROKE=MHEA10;

Response format: 0=No Prevalent Stroke (No)
1= Prevalent Stroke (Yes)

Source variable(s):

MHEA10. Has a doctor ever said that you had a stroke?

12.42 STROKE_TIA (Prevalent stroke or TIA, self-report)

This is a 0/1 variable that checks for medical history of stroke or mini-stroke or TIA as follows:

if MHEA10=1 | MHEA11=1 then STROKE_TIA=1;
else if MHEA10≤.Z & MHEA11≤.Z then STROKE_TIA=.;
else STROKE_TIA=0;

Response format: 0 = No Prevalent Stroke/TIA (No)
1 = Prevalent Stroke/TIA (Yes)

Source variable(s):

MHEA10. Has a doctor ever said that you had a stroke?

MHEA11. Has a doctor ever said that you had a mini-stroke or TIA (transient ischemic attack)?

12.43 FH_STROKE (Family history of stroke)

This is a 0/1 variable that combines self-report of stroke for immediate family relatives, i.e. mother, father, and sibling(s).

If MHEA10 is non-missing then:

IF MHEA10a, MHEA10b, and MHEA10c are all missing then FH_Stroke = missing.

IF MHEA10a=1 or MHEA10b=1 or MHEA10c=1 then FH_Stroke = 1

Otherwise FH_Stroke = 0.

Response format: 0=No family history of stroke (No)
1=Family history of stroke (Yes)

Source variable(s):

MHEA10. Has a doctor ever said that you had a stroke?

MHEA10a. Has a doctor ever said that these relatives had a stroke: Mother

MHEA10b. Has a doctor ever said that these relatives had a stroke: Father

MHEA10c. Has a doctor ever said that these relatives had a stroke: Brother(s) or Sister(s)

12.44 PRECVD (Prevalent CVD (CHD or stroke, not counting angina or TIA))

This is a 0/1 variable that indicates prevalent cardiovascular disease, including CHD and stroke, without counting angina or TIA. as follows:

```
If PRECHD_NO_ANGINA<=.Z & STROKE<=.Z then PRECVD=.;  
Else if PRECHD_NO_ANGINA=1 | STROKE=1 then PRECVD=1;  
Else PRECVD=0;
```

Response format: 0=No Prevalent CVD (No)
1= Prevalent CVD (Yes)

Source variable(s):

PRECHD_NO_ANGINA. Prevalent CHD, not counting Angina.
STROKE. Prevalent Stroke, self-report.

12.45 CIGARETTE_USE (1=Never,2=Former,3=Current)

Current, former, and never cigarette use.

Response format: 1 = "Never"
2 = "Former"
3 = "Current"

```
If TBEA1=1 & TBEA3 in(1,2) then CIGARETTE_USE=3;  
else if (TBEA1=1 OR TBEA1<=.z) & TBEA3 in(3) then CIGARETTE_USE=2;  
else if TBEA1=0 then CIGARETTE_USE=1;  
else if (TBEA1<=.z OR TBEA3<=.z) AND TBEA4>1 then CIGARETTE_USE=3;
```

10/2025 update: revised to use TBEA4 if missing TBEA1 or TBEA3 to rescue 8 missing values.

Source variable(s):

TBEA1. Have you ever smoked at least 100 cigarettes in your entire life?
TBEA3. Do you now smoke daily, some days or not at all?
TBEA4. How many cigarettes do you smoke per day now?

12.46 CURRENT_SMOKER (Current smoker from CIGARETTE_USE)

This is a 0/1 indicator variable to denote current smoker (CIGARETTE_USE=3) status using the variable CIGARETTE_USE.

Response format: 0 = No (i.e., never or former smoker)
1 = Yes (i.e., current smoker)

Source variable(s):

CIGARETTE_USE. Cigarette_Use (1=Never,2=Former,3=Current)

12.47 EXPOSURE_YEAR (Adjusted years smoked)

This is a numeric variable that calculates the years a participant smoked cigarettes. This is calculated using the difference from the age the participant began smoking to either:

- a. the participant's age at baseline examination or
- b. the age that the participant quit smoking (if applicable)

This variable is also adjusted for any years that the participant quit smoking during the above calculated smoking time period (Note: TBEA4B, 5C, and 6B are = 0 if the participant did not quit smoking during this time period). If the final calculation is between 0 and 1, then the variable is set to =1. If the number of years quit was missing, it was assumed to be zero.

Five cases for computing years of cigarette smoking exposure:

- 1) Current smoker, Daily: If (TBEA1=1 and TBEA3=1) then Exposure_year = (Age – TBEA2 – TBEA4B)
- 2) Current smoker, Some days: If (TBEA1=1 and TBEA3=2) then Exposure_year = (Age – TBEA2 – TBEA5C)
- 3) Former smoker: If (TBEA1=1 and TBEA3=3) then Exposure_Year = (TBEA6 – TBEA2 – TBEA6B)
- 4) Never Smoked: If (TBEA1=0) then Exposure_year = 0
- 5) All Other cases: Exposure_year = .

5/2014 update: revised to treat TBEA2=99 or 1 as missing. Corrected code to set never smokers to zero, former and current smokers with negative values to missing and to subtract TBEA2 for former smokers

Source variable(s):

TBEA1. Have you ever smoked at least 100 cigarettes in your entire life?

TBEA2. How old were you when you first started to smoke cigarettes fairly regularly?

TBEA3. Do you now smoke daily, some days or not at all?

TBEA4A. Did you ever quit smoking for 6 months or longer? (Daily)

TBEA4B. For how many years in total did you quit smoking? (Daily)

TBEA5B. Did you ever quit smoking for 6 months or longer? (Some days)

TBEA5C. For how many years in total did you quit smoking? (Some days)

TBEA6. How old were you when you completely stopped smoking?

TBEA6A. When you were a smoker, did you ever quit smoking for 6 months or longer before you

TBEA6B. During the time that you were a smoker, for how many years in total did you quit smoking? (Former)

Age. Age in years at the time of participant's clinic visit

12.48 CIGARETTES_YEAR (Adjusted cigarette years)

This is a numeric variable that calculates the cigarette years as the number of exposure years multiplied by the average number of cigarettes smoked per day. Set to missing for participants who said yes to TBEA1 and 0 for TBEA9.

5/2014 update: updates due to changes in EXPOSURE_YEAR, set current smokers with TBEA9=0 to missing

Source variable(s):

TBEA1. Have you ever smoked at least 100 cigarettes in your entire life?

TBEA9. Of the entire time you have or had smoked, on average how many cigarettes do you or did you smoke per day?

Exposure_Year. Adjusted Years Smoked

12.49 CIGARETTE_PACK_YEARS (Cigarette pack years)

This is a numeric variable that calculates the cigarette pack years as the number of exposure years multiplied by the average number of cigarettes smoked per day (Cigarettes_Year), divided by 20.

$$\text{Cigarette_Pack_years} = \text{Cigarettes_Year} / 20$$

5/2014 update: updates due to changes in Exposure_Year

Source variable(s):

Cigarettes_Year. Adjusted Cigarette Years

12.50 ALCOHOL_USE (1=Never,2=Former,3=Current)

Current, former, and never alcohol use.

Response format: 1 = "Never"
 2 = "Former"
 3 = "Current"

if ALEA1=1 then ALCOHOL_USE=3;
else if ALEA1=0 & ALEA7=1 then ALCOHOL_USE=2;
else if ALEA1=0 & ALEA7=0 then ALCOHOL_USE=1;

Source variable(s):

ALEA1. Do you presently drink alcoholic beverages?

ALEA7. Did you ever drink alcohol?

12.51 TOTALDRINKS_PER_WEEK (Total weekly alcohol consumption)

This is a numeric variable that determines the total alcohol Intake of the numbers of drinks per week only for current drinkers.

```
if ALEA1^=1 then ALCOHOL_INTAKE=.;  
else ALCOHOL_INTAKE=sum(of ALEA2-ALEA5);
```

Source variable(s):

ALEA1. Do you presently drink alcoholic beverages?

ALEA2. How many glasses of red wine do you usually have per week?

ALEA3. How many glasses of white wine do you usually have per week?

ALEA4. How many cans, bottles, or glasses of beer do you usually have per week?

ALEA5. How many drinks of liquor, spirits, or mixed drinks do you usually have per week?

12.52 ALCOHOL_USE_DISORDER (Alcohol use disorder risk)

Based on SEX-specific cutoffs for weekly alcohol use (see below) provided by the National Institute for Alcohol Abuse and Alcoholism (www.niaaa.nih.gov), ALCOHOL_USE_DISORDER provides a risk level for developing an alcohol use disorder (craving, loss of control, physical dependence, tolerance, etc.). Note that the NIAAA criteria includes binge drinking (i.e. drinks on any single day) levels that were not collected in HCHS/SOL. ALCOHOL_USE_DISORDER sets former drinkers to missing with the reasoning that they may meet criteria for an alcohol use disorder, but the assessment of that risk is not possible based on current use.

SEX	Weekly alcohol use	Alcohol_Use_Disorder
F	never used alcohol	Non-drinker
F	current use <7 drinks/week	Low risk drinker
F	7+ drinks/week	At-risk drinker
M	never used alcohol	Non-drinker
M	current use <14 drinks/week	Low risk drinker
M	14+ drinks/week	At-risk drinker

Response format: 0 = Non-drinker
1 = Low risk drinker
2 = At-risk drinker

Source variable(s):

Sex. Sex

Alcohol_Use. (1=Never,2=Former,3=Current)

TotalDrinks_Per_Week. Total Weekly Alcohol Consumption

12.53 ALCOHOL_USE_LEVEL (Alcohol use disorder risk - NIAAA)

Using the weekly alcohol use cutoffs from ALCOHOL_USE_DISORDER, ALCOHOL_USE_LEVEL combines never users former users into a group called "no current use". Definitions of "low level use" and "high level use" correspond to those of "low risk drinker" and "at-risk drinker" from ALCOHOL_USE_DISORDER

SEX	Weekly alcohol use	Alcohol_Use_Level
F	no current use (never or former users)	No current use
F	current use <7 drinks/week	Low level use
F	7+ drinks/week	High level use
M	no current use (never or former users)	No current use
M	current use <14 drinks/week	Low level use
M	14+ drinks/week	High level use

Response format: 0 = No current use
1 = Low level use
2 = High level use

Source variable(s):

SEX. Sex

ALCOHOL_USE. (1=Never,2=Former,3=Current)

TOTALDRINKS_PER_WEEK. Total Weekly Alcohol Consumption

12.54 DOCTOR_VISIT (Doctor/HCP number of visits in past 12mo)

This variable defines grouped categories for the number of times a participant has visited the doctor in the last 12 months.

Update 12/2013: DOCTOR_VISIT was corrected to calculate values for version B for the health care use form (HCE).

```
if hcea5=0 | hceb5=0 then doctor_visit=1;  
else if hcea5=1 | hceb5=1 then doctor_visit=2;  
else if hcea5 in(2,3) | hceb5 in(2,3) then doctor_visit=3;  
else if hcea5>=4 | hceb5>=4 then doctor_visit=4;
```

Response format: 1 = None
2 = Once
3 = 2-3 Times
4 = More than 3 times

Source variable(s):

HCEA5. During the past 12 months, how many times did you see a physician for your health care?

HCEB5. During the past 12 months, how many times did you see a physician for your health care?

12.55 CLAUDICATION_INT (Intermittent claudication - either leg)

This is a 0/1/9 variable that defines intermittent claudication being present in either leg as follows:

```
if clea2=. then CLAUDICATION_INT=.;  
else if clea2=1 and (clea7=1 | clea12=1) then CLAUDICATION_INT=1;  
else if clea2=1 and (clea7=2 | clea12=2) then CLAUDICATION_INT=0;  
else CLAUDICATION_INT=9;
```

Response format: 0 = Intermittent claudication not present (No)
1 = Intermittent claudication present (Yes)
9 = Don't Know

Source variable(s):

CLEA2. Do you get pain or discomfort in either leg on walking?

CLEA7. What happens to the pain if you stand still? (Right Leg)

CLEA12. What happens to the pain if you stand still? (Left Leg)

12.56 CVD_FRAME (Framingham CVD composite criterion)

This is a 0/1 indicator variable that defines a composite CVD definition based on the Framingham Study criterion (see the following link for reference documentation:

<https://pubmed.ncbi.nlm.nih.gov/3568338/>):

The Framingham Heart Study defines CVD as a composite of CHD (coronary death, myocardial infarction, coronary insufficiency, and angina), cerebrovascular events (including ischemic stroke, hemorrhagic stroke, and transient ischemic attack), peripheral artery disease (intermittent claudication), and heart failure].

If any of the following highlighted variables = 1, then let CVD_FRAME=1, otherwise =0. Note that this variable is meant to classify positive cases, participants with incomplete information were classified as no/unknown

CHD:

Coronary death – n/a in baseline HCHS

Myocardial infarction – MI_ECG=1 or MHEA4=1

Coronary insufficiency – considering all of the following:

Rheumatic heart disease – MHEA6=1

Atrial fibrillation- MHEA7=1

Angio/stent used to improve:

Blood flow to heart – MHEA9=1

Correct stroke – MHEA12=1

Aortic aneurysm, ballooning, or AAA – MHEA13=1

Angina – MHEA3=1

Cerebrovascular events:

including ischemic stroke - MHEA10=1

hemorrhagic stroke – MHEA10=1

transient ischemic attack - MHEA11=1

Peripheral artery disease (intermittent claudication) -

MHEA14=1 or CLAUDICATION_INT=1 (from CLEA form) or ABIGRP2_C4 = 1 (defined PAD).

Heart failure – MHEA5=1

Other heart problems - MHEA8=1

Response format: 0 = No or unknown CVD
1 = CVD risk (yes)

Source variable(s):

MHEA3. Has a doctor ever said that you have angina?

MHEA4. Has a doctor ever said that you have had a heart attack?

MHEA5. Has a doctor ever said that you had heart failure?

MHEA6. Has a doctor ever said that you had rheumatic heart disease?

MHEA7. Has a doctor ever told you that you had atrial fibrillation?

MHEA8. Has a doctor ever said that you had some other kind of heart problem?

MHEA9. Have you had a balloon angioplasty, a stent, or bypass surgery to the arteries in your heart to improve the blood flow to your heart?

MHEA10. Has a doctor ever said that you had a stroke?

MHEA11. Has a doctor ever said that you had a mini-stroke of TIA (transient ischemic attack)?

MHEA12. Have you had a balloon angioplasty or surgery to the arteries of your neck to prevent or correct a stroke?

MHEA13. Has a doctor ever said that you have an aortic aneurysm, an AAA, or ballooning of your aorta?

MHEA14. Has a doctor ever said that you have peripheral arterial disease (problems with circulation, blocked arteries to the legs)?

ABIGRP2_C4. 4-level grouped Ankle Brachial Index – version 2

CLAUDICATION_INT. Intermittent Claudication - either leg

MI_ECG. MI defined by ECG results

12.57 FRAME_CVD_RISK_10YR (Framingham CVD - 10 year risk (w/lab predictors))

This variable defines the Framingham CVD 10 year risk using lab predictors based on the Framingham Study criterion for participants Individuals 30 to 74 years old only (see the following link for reference documentation: <https://www.framinghamheartstudy.org/fhs-risk-functions/cardiovascular-disease-10-year-risk/>):

****Note:** In the Framingham definition, the beta coefficients were calculated only among those who did NOT have a previous diagnosis of CVD. This exclusion was applied to obtain the beta coefficients only and should not be, and therefore was not, used as an

exclusion criterion for calculating the Framingham 10-year risk of CVD for the HCHS/SOL participants. This CVD_FRAME variable has been provided separately for potential use. The 10-year risk for women can be calculated as $1 - 0.95012^{\exp(\sum \beta X - 26.1931)}$ where β is the regression coefficient and X is the level for each risk factor; the risk for men is given as $1 - 0.88936^{\exp(\sum \beta X - 23.9802)}$

Diabetes definition as outlined by Framingham:

If LABA70 >= 126 OR MED_ANTIDIAB=1 OR MUEA33C=1 OR MHEA16c=1
then DIABETES_FRAME=1. Otherwise =0.

Hypertension medication use as outlined by Framingham:

If MED_ANTIHYPERT=1 OR MUEA33D=1 then HYPERT_MED_FRAME=1.
Otherwise =0.

The final definition will be based on Framingham beta coefficient and variable definitions as follows:

Men* (10-year Baseline Survival: So(10) = 0.88936)	
Variable	Beta**
Log of Age (AGE)	3.06117
Log of Total Cholesterol (LABA66)	1.12370
Log of HDL Cholesterol (LABA68)	-0.93263
Log of SBP if not treated (SBPA5 where HYPERT_MED_FRAME=0)	1.93303
Log of SBP if treated (SBPA5 where HYPERT_MED_FRAME=1)	1.99881
Smoking (Current_smoker)	0.65451
Diabetes (DIABETES_FRAME)	0.57367

Women* (10-year Baseline Survival: So(10) = 0.95012)	
Variable	Beta**
Log of Age (AGE)	2.32888
Log of Total Cholesterol (LABA66)	1.20904
Log of HDL Cholesterol (LABA68)	-0.70833
Log of SBP if not treated (SBPA5 where HYPERT_MED_FRAME=0)	2.76157
Log of SBP if treated (SBPA5 where HYPERT_MED_FRAME=1)	2.82263
Smoking (Current_smoker)	0.52873
Diabetes (DIABETES_FRAME)	0.69154

Take the log of each continuous variable as indicated above and multiply it by the indicated beta coefficient. For the categorical response variables above simply multiply the variable by the indicated beta coefficient. Then sum all of the values calculated above to define LOG_RISK, setting participants to missing if they are missing any of the above components. Finally, calculate the desired variable, FRAME_CVD_RISK_10YR, for individuals age 30-74 only as follows:

Where $30 \leq \text{AGE} \leq 74$:

For women (SEX='F'):

let FRAME_CVD_RISK_10YR = $1 - 0.95012^{\exp(\text{LOG_RISK} - 26.1931)}$

For men (SEX='M'):

let FRAME_CVD_RISK_10YR = $1 - 0.88936^{\exp(\text{LOG_RISK} - 23.9802)}$

Source variable(s):

AGE. Age in years at the time of participant's clinic visit

LABA66. Total Cholesterol (mg/dL)

LABA68. HDL cholesterol (mg/dL)

LABA70. Fasting glucose (mg/dL)

SBPA5. Systolic blood pressure

MUEA33C. Self-report anti-diabetic medication use

MUEA33D. Self-report anti-hypertensive medication use

MHEA3. Has a doctor ever said that you have angina?

MHEA4. Has a doctor ever said that you had a heart attack?

MHEA5. Has a doctor ever said that you had heart failure?

MHEA6. Has a doctor ever said that you had rheumatic heart disease?

MHEA7. Has a doctor ever told you that you had atrial fibrillation?

MHEA8. Has a doctor ever said that you had some other kind of heart problem?

MHEA9. Have you had a balloon angioplasty, a stent, or bypass surgery to the arteries in your heart to improve the blood flow to your heart?

MHEA10. Has a doctor ever said that you had a stroke?

MHEA11. Has a doctor ever said that you had a mini-stroke or TIA?

MHEA12. Have you had a balloon angioplasty or surgery to the arteries of your neck to prevent or correct a stroke?

MHEA13. Has a doctor ever said that you have an aortic aneurysm, an AAA, or ballooning of your aorta?

MHEA14. Has a doctor ever said that you have peripheral arterial disease?

MHEA16c. Are you being treated with insulin?

MED_ANTIDIAB. Anti-diabetic medication use

MED_ANTIHYPERT. Antihypertensives

CURRENT_SMOKER. Current Smoker (from CIGARETTE_USE)

12.58 HBA1C_SI (Glycosylated hemoglobin in SI units (mmol/mol))

$HBA1C_SI = (LABA72 * 10.93) - 23.50$, rounded to the nearest integer

Beginning with manuscripts submitted after Jan. 1, 2013, Diabetes requires authors to report HbA1c levels in both traditional, DCCT-derived units (as %) and SI, IFCC-recommended units (as mmol/mol).

Change of HbA1c reporting to the new SI units, Jones G, Barker G, et al. MJA 2011; 195: 45–46

Source variable(s):

LABA72. % Glycosylated Hemoglobin

12.59 CIGARETTE_PACK_YEARS_C3 (3-level Cigarette pack years (cutoff =10))

Groups cigarette use into 3 categories: 0, between 0 and 10 and 10+.

If CIGARETTE_PACK_YEARS =0 then set CIGARETTE_PACK_YEARS_C3=1.

If CIGARETTE_PACK_YEARS is >0 and <10 set CIGARETTE_PACK_YEARS_C3=2.

If CIGARETTE_PACK_YEARS is >=10 then set CIGARETTE_PACK_YEARS_C3=3

Response format: 1 = 0 years
2 = >0 and <10 years
3 = 10+ years

Source variable(s):

CIGARETTE_PACK_YEARS. Cigarette Pack Years

12.60 PREVENT Scores

The American Heart Association Predicting Risk of cardiovascular disease EVENTS (PREVENT) algorithms calculate risk estimates for cardiovascular disease (CVD) and two subtypes (atherosclerotic CVD [ASCVD] and heart failure [HF]) based on specified input conditions. The macro provides 10-year risk estimates for individuals 30-79 years of age and provides 30-year risk estimates for individuals 30-59 years of age. If there are missing or out-of-range values for sex, age (<30 or >79), SBP (<90 or >200), diabetes, smoker status, hypertension treatment, or eGFR (≤ 0), then all 10- and 30-year risk probabilities are set to missing. An R package was requested on 12/12/2024 at <https://www.jotform.com/240774577352161>, downloaded from the AHA PREVENT Git repository (<https://github.com/AHA-DS-Analytics/PREVENT>), and converted to a SAS macro.

Reference:

Khan SS, Matsushita K, Sang Y, Ballew SH, Grams ME, Surapaneni A, Blaha MJ, Carson AP, Chang AR, Ciemins E, Go AS, Gutierrez OM, Hwang SJ, Jassal SK, Kovesdy CP, Lloyd-Jones DM, Shlipak MG, Palaniappan LP, Sperling L, Virani SS, Tuttle K, Neeland IJ, Chow SL, Rangaswami J, Pencina MJ, Ndumele CE, Coresh J; Chronic Kidney Disease Prognosis Consortium and the American Heart Association Cardiovascular-Kidney-Metabolic Science Advisory Group. Development and Validation of the American Heart Association's PREVENT Equations. *Circulation*. 2024 Feb 6;149(6):430-449. doi: 10.1161/CIRCULATIONAHA.123.067626.


```

if HYPERTENSION2 = 0 then HYPTREAT = 0;
  else if missing(HYPERTENSION2) then HYPTREAT = .;
  else HYPTREAT = HYPERT_TREATMENT;

if MUEA2 = 1 then CHOLTREAT= 0;
  else if (MUEA33E = 9) or (MISSING(MUEA33E) and MISSING(MUEA2))
    then CHOLTREAT = .;
  else CHOLTREAT = MUEA33E;

%prevent_base_values (sex=SEX,
                      age=AGE,
                      tc=LABA66,
                      hdl=LABA68,
                      sbp=SBPA5,
                      diab=DIABETES_SELF,
                      smoker=CURRENT_SMOKER,
                      bmi=BMI,
                      egfr=GFRSCR_Y21,
                      hyptreat= HYPTREAT,
                      choltreat=CHOLTREAT, suffix=);

```

Source variable(s):

SEX. Sex

AGE. Age

SBPA5. Systolic blood pressure

DIABETES_SELF. Diabetes self-report prevalent cases

CURRENT_SMOKER. Current smoker status

GFRSCR_Y21. eGFR based on 2021 definition using serum creatinine only

HYPERTENSION2. Hypertension

HYPERT_TREATMENT

LABA66. Total cholesterol (mg/dL)

LABA68. HDL (mg/dL)

MUEA33E. Self-reported use lipid lowering drugs/antihyperlipidemic medication. Note: MUEA33E missing had been reset to 0 when MUEA2=1 (confirmed not taking any medications).

BMI. Body mass index

12.60.1 PREVENT_BASE_10YR_CVD (10-yr risk probability for cardiovascular disease)

The 10-year risk probability for cardiovascular disease. The input variables are sex, age, systolic blood pressure, diabetes status, smoker status, eGFR, hypertension treatment, total cholesterol level, HDL level, and high cholesterol treatment. If total cholesterol is missing, <130 or >320, or HDL is missing, <20 or >100, or missing cholesterol treatment, then the probability is set to missing.

12.60.2 PREVENT_BASE_30YR_CVD (30-yr risk probability for cardiovascular disease)

The 30-year risk probability for cardiovascular disease. The input variables are sex, age, systolic blood pressure, diabetes status, smoker status, eGFR, hypertension treatment, total cholesterol level, HDL level, and high cholesterol treatment. If total cholesterol is missing, <130 or >320, or HDL is missing, <20 or >100, or missing cholesterol treatment, or age > 59 then the probability is set to missing.

12.60.3 PREVENT_BASE_10YR_ASCVD (10-yr risk probability for atherosclerotic cardiovascular disease)

The 10-year risk probability for atherosclerotic cardiovascular disease. The input variables are sex, age, systolic blood pressure, diabetes status, smoker status, eGFR, hypertension treatment, total cholesterol level, HDL level, and high cholesterol treatment. If total cholesterol is missing, <130 or >320, or HDL is missing, <20 or >100, or missing cholesterol treatment, then the probability is set to missing.

12.60.4 PREVENT_BASE_30YR_ASCVD (30-yr risk probability for atherosclerotic cardiovascular disease)

The 30-year risk probability for atherosclerotic cardiovascular disease. The input variables are sex, age, systolic blood pressure, diabetes status, smoker status, eGFR, hypertension treatment, total cholesterol level, HDL level, and high cholesterol treatment. If total cholesterol is missing, <130 or >320, or HDL is missing, <20 or >100, or missing cholesterol treatment, or age > 59 then the probability is set to missing.

12.60.5 PREVENT_BASE_10YR_HF (10-yr risk probability for heart failure)

The 10-year risk probability for heart failure. The input variables are sex, age, systolic blood pressure, diabetes status, smoker status, eGFR, hypertension treatment, and BMI. If BMI is missing, <18.5, or ≥40 then the probability is set to missing.

12.60.6 PREVENT_BASE_30YR_HF (30-yr risk probability for heart failure)

The 30-year risk probability for heart failure. The input variables are sex, age, systolic blood pressure, diabetes status, smoker status, eGFR, hypertension treatment, and BMI. If BMI is missing, <18.5, or ≥40, or age > 59 then the probability is set to missing.

12.61 CMR (Cardiometabolic Risk Score)

The cardiometabolic risk score (CMR) is designed to characterize overall functioning across multiple measures of cardiovascular activity, similar to MetS and consistent with concepts of allostatic load. To compare to prior studies among sexual minorities, the SGM Ancillary Study created a cumulative biological risk score by counting the number (range: 0–5) of biological markers that met a clinically defined high-risk criterion. The criteria for high risk was defined as: (a) SBP ≥ 140 mmHg; (b) DBP ≥ 90 mmHg; (c) glycosylated hemoglobin ≥ 6.4%; (d) C-reactive protein (CRP) ≥ 3 mg/dL; and (e) WC ≥ 102 cm for men and ≥ 88 cm for women. Individuals received a value of 1 if they are above the threshold of risk. It is recommended to analyze this as a continuous variable since there is no established cut-off value for high risk.

Algorithm:

```
If (SEX = 0 and ANTA10a >= 88) or  
  (SEX = 1 and ANTA10a >= 102) then CENTROBESE = 1;  
Else if SEX <= .z or ANTA10a <= .z then CENTROBESE = .;  
Else CENTROBESE = 0;
```

CMR = 0;

```
If SBPA5 >= 140 then CMR = CMR + 1;  
If SBPA6 >= 90 then CMR = CMR + 1;  
If LABA72 >= 6.4 then CMR = CMR + 1;  
If LABA91 >= 3 then CMR = CMR + 1;  
If CENTROBESE = 1 then CMR = CMR + 1;
```

If any source variable is missing then CMR is missing.

Source variable(s):

SEX. Sex
ANTA10a. Waist circumference (cm)
SBPA5. Systolic blood pressure
SBPA6. Diastolic blood pressure
LABA72. Glycosylated Hemoglobin (%)
LABA91. High-sensitivity C-Reactive Protein (mg/dL)

References:

Beckie TM. A systematic review of allostatic load, health, and health disparities. Biological research for nursing. 2012;14(4):311-346.

Hatzenbuehler ML, McLaughlin KA, Slopen N. Sexual orientation disparities in cardiovascular biomarkers among young adults. American journal of preventive medicine. 2013;44(6):612-621.

Hatzenbuehler ML, Slopen N, McLaughlin KA. Stressful life events, sexual orientation, and cardiometabolic risk among young adults in the United States. Health psychology: official journal of the Division of Health Psychology, American Psychological Association. 2014;33(10):1185-1194.

12.62 PCE Score (CVD risk using Pooled Cohort 10-Year ASCVD Risk Assessment Equations)

The PCE, Pooled Cohort 10-Year Atherosclerotic Cardiovascular Disease Risk Assessment Equation, uses a complex algorithm to estimate the risk of heart attack or stroke over the following ten years for adults aged 40-79 years. The measures used to calculate this risk include sex assigned at birth, age, socially-ascribed race, total cholesterol, LDL-cholesterol, HDL-cholesterol, systolic and diastolic blood pressure (SBP), hypertension treatment (i.e., anti-hypertensives), type-2 diabetes diagnosis, statin treatment (i.e., anti-cholesterol), and tobacco use (current smoking only). We include the PCE measure for all participants because it is commonly used in clinical practice and is likely to have real-world applicability. Since the reference algorithm was not used in Hispanic populations, all participants have the race parameter set to white.

References:

Goff DC, Jr., Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Journal of the American College of Cardiology. 2014;63(25 Pt B):2935-2959.

Topel ML, Shen J, Morris AA, et al. Comparisons of the Framingham and Pooled Cohort Equation Risk Scores for Detecting Subclinical Vascular Disease in Blacks Versus Whites. The American journal of cardiology. 2017.

12.62.1 PCE_SCORE (CVD risk in the next 10 years)

A PCE score is first calculated using the ASCVD macro with the following parameters:

%ASCVD

```
(MALE      = SEX,
WHITE      = 1,
AGE        = AGE,
TOT_CHOL   = LABA66,
HDL        = LABA68,
SYSTOLIC   = SBPA5,
ANTI_HYP   = HYPERT_TREATMENT2,
CURSMOKE   = CURRENT_SMOKER,
DIABETES   = DIABETES_SELF)
```

where intermediate variables HYPERT_TREATMENT2 and HYPERTMED_SELF are defined as:

```
if MUEA2 = 1 and MUEA33d < .Z or MUEA33d in (0,9) then HYPERTMED_SELF = 0;
else if MUEA33d = 1 then HYPERTMED_SELF = 1;
```

```
if HYPERTENSION2 = 0 & missing(HYPERT_TREATMENT) then
HYPERT_TREATMENT2 = 0;
else if missing(HYPERTMED_SELF) then HYPERT_TREATMENT2 = .;
else HYPERT_TREATMENT2 = HYPERT_TREATMENT;
```

Source variables:

SEX. Sex

AGE. Age

LABA66. Total cholesterol (mg/dL)

LABA68. HDL-cholesterol (mg/dL)

SBPA5. Systolic blood pressure

HYPERT_TREATMENT2. Treatment of Hypertension (recoding missing because of not hypertensive (NHANES definition) to zero)

CURRENT_SMOKER. Current smoker status

DIABETES_SELF. Self-reported diabetes diagnosis

12.62.2 PCE (High cardiovascular disease risk in the next 10 years)

PCE is an indicator variable for those participants with PCE score > 7.5% who are considered to have high CVD risk.

If PCE_SCORE > 7.5 and all components are non-missing then PCE = 1;
Else PCE = 0;

Response format: 0 = No risk
1 = High cardiovascular disease risk over the following 10 years

Source variable:

PCE_SCORE. CVD risk using Pooled Cohort 10-Year ASCVD Risk Assessment Equations

12.63 PARITY (Number of live births)

The number of live births a participant has had.

if MHEA34=. then PARITY=.
else if MHEA34=0 then PARITY=0
else if MHEA36>19 then PARITY=.
else PARITY=minimum of MHEA35 and MHEA36

Source variables:

MHEA34. Have you ever been pregnant?
MHEA35. How many times have you been pregnant?
MHEA36. How many live births have you had?

12.64 Life's Essential 8 (LE8) Scores

12.64.1 LE8_SCAN (Life's Essential 8 Score using scanned meds)

The Life's Essential 8 (LE8) is a composite cardiovascular health score, ranging from 0 to 100 points, developed by the American Heart Association. It is constructed as the unweighted average of 8 component scores: diet, physical activity, nicotine exposure, sleep health, body mass index, blood lipids, blood glucose, and blood pressure. Each component takes values from 0 to 100. This variable used scanned medication data for blood pressure and lipids components. If more than 1 of the components are missing, then the composite score is set to missing.

Reference:

Lloyd-Jones DM, Allen NB, Anderson CAM, Black T, Brewer LC, Foraker RE, Grandner MA, Lavretsky H, Perak AM, Sharma G, Rosamond W; American Heart Association. Life's Essential 8: Updating and Enhancing the American Heart Association's Construct of Cardiovascular Health: A Presidential Advisory From the American Heart Association. *Circulation*. 2022 Aug 2;146(5):e18-e43. doi: 10.1161/CIR.0000000000001078. Epub 2022 Jun 29. PMID: 35766027; PMCID: PMC10503546.

Response Format: Continuous score from 0 to 100

Source variable(s):

LE8_BMI. Body Mass Index Component Score.
LE8_BP_SCAN. Blood Pressure Component Score with Scanned Medication.
LE8_DIET. Diet Component Score using HEI-2010.
LE8_GLUCOSE. Blood Glucose Component Score.
LE8_LIPIDS_SCAN. Blood Lipids Component Score with Scanned Medication.
LE8_NICOTINE. Nicotine Exposure Component Score.
LE8_PA. Physical Activity (self-reported) Component Score.
LE8_SLEEP. Sleep Health Component Score.

12.64.2 LE8_SELF (Life's Essential 8 Score using reported meds)

The Life's Essential 8 (LE8) is a composite cardiovascular health score, ranging from 0 to 100 points, developed by the American Heart Association. It is constructed as the unweighted average of 8 component scores: diet, physical activity, nicotine exposure, sleep health, body mass index, blood lipids, blood glucose, and blood pressure. Each component takes values from 0 to 100. This variable used self-reported medication data for blood pressure and lipids components. If more than 1 of the components are missing, then the composite score is set to missing.

Reference:

Lloyd-Jones DM, Allen NB, Anderson CAM, Black T, Brewer LC, Foraker RE, Grandner MA, Lavretsky H, Perak AM, Sharma G, Rosamond W; American Heart Association. Life's Essential 8: Updating and Enhancing the American Heart Association's Construct of Cardiovascular Health: A Presidential Advisory From the American Heart Association. *Circulation*. 2022 Aug 2;146(5):e18-e43. doi: 10.1161/CIR.0000000000001078. Epub 2022 Jun 29. [PMID: 35766027](#); [PMCID: PMC10503546](#).

Response Format: Continuous score from 0 to 100

Source variable(s):

LE8_BMI. Body Mass Index Component Score.
LE8_BP_SELF. Blood Pressure Component Score with Self-Reported Medication.
LE8_DIET. Diet Component Score.
LE8_GLUCOSE. Blood Glucose Component Score.
LE8_LIPIDS_SELF. Blood Lipids Component Score with Self-Reported Medication.
LE8_NICOTINE. Nicotine Exposure Component Score.
LE8_PA. Physical Activity Component Score.
LE8_SLEEP. Sleep Health Component Score.

12.64.3 LE8_BMI (Life's Essential 8 Body Mass Index component score)

Score based on BMI (kg/m²) using LE8 scoring thresholds.

If BMIGRP_C6 in (1,2) then LE8_BMI = 100
Else if BMIGRP_C6 = 3 then LE8_BMI = 70
Else if BMIGRP_C6 = 4 then LE8_BMI = 30
Else if BMIGRP_C6 = 5 then LE8_BMI = 15
Else if BMIGRP_C6 = 6 then LE8_BMI = 0
Else if BMIGRP_C6 = . then LE8_BMI = .

Response Format: Continuous score from 0 to 100

Source variable(s):

BMIGRP_C6. 6-level grouped Body Mass Index – WHO.

12.64.4 LE8_BP_SCAN (Life's Essential 8 Blood Pressure component score with scanned med)

Score based on systolic/diastolic BP and scanned antihypertensive medication use.

```
If SBPA5 = . or SBPA6 = . or MED_ANTIHYPERT = . then LE8_BP_SCAN = .  
Else if SBPA5 >= 160 or SBPA6 >= 100 then LE8_BP_SCAN = 0  
Else if MED_ANTIHYPERT = . then LE8_BP_SCAN = .  
Else if 140 <= SBPA5 <= 159 or 90 <= SBPA6 <= 99 then LE8_BP_SCAN = 25  
Else if 130 <= SBPA5 <= 139 or 80 <= SBPA6 <= 89 then LE8_BP_SCAN = 50  
Else if 120 <= SBPA5 <= 129 and SBPA6 < 80 then LE8_BP_SCAN = 75  
Else if SBPA5 < 120 and SBPA6 < 80 then LE8_BP_SCAN = 100  
If MED_ANTIHYPERT = 1 then subtract 20 points (unless score is 0)
```

Response Format: Continuous score from 0 to 100

Source variable(s):

SBPA5. Systolic blood pressure.

SBPA6. Diastolic blood pressure.

MED_ANTIHYPERT. Scanned medication was for treatment of hypertension.

12.64.5 LE8_BP_SELF (Life's Essential 8 Blood Pressure component score with self-reported med)

Score based on systolic/diastolic BP and self-reported antihypertensive medication use.

```
If SBPA5 = . or SBPA6 = . or LE8_BP_MED_SELF = . then LE8_BP_SELF = .  
Else if SBPA5 >= 160 or SBPA6 >= 100 then LE8_BP_SELF = 0  
Else if LE8_BP_MED_SELF = . then LE8_BP_SELF = .  
Else if (140 <= SBPA5 <= 159 or 90 <= SBPA6 <= 99) then LE8_BP_SELF = 25  
Else if (130 <= SBPA5 <= 139 or 80 <= SBPA6 <= 89) then LE8_BP_SELF = 50  
Else if 120 <= SBPA5 <= 129 and SBPA6 < 80 then LE8_BP_SELF = 75  
Else if SBPA5 < 120 and SBPA6 < 80 then LE8_BP_SELF = 100  
If LE8_BP_MED_SELF = 1 then subtract 20 points (unless score is 0)
```

Where LE8_BP_MED_SELF is defined as:

```
If MUEA33D = 1 then LE8_BP_MED_SELF = 1
```

```
Else if MUEA33D in(0,.S) then LE8_BP_MED_SELF = 0
```

```
Else LE8_BP_MED_SELF = .
```

Response Format: Continuous score from 0 to 100

Source variable(s):

SBPA5. Systolic blood pressure.

SBPA6. Diastolic blood pressure.

MUEA33D. Self-report anti-hypertensive medication use.

12.64.6 LE8_DIET (Life's Essential 8 Diet component score)

Score based on adherence to Healthy Eating Index (HEI) 2010 recommendations.

```
If LE8_HEI_GROUP = 5 then LE8_DIET = 100
Else if LE8_HEI_GROUP = 4 then LE8_DIET = 80
Else if LE8_HEI_GROUP = 3 then LE8_DIET = 50
Else if LE8_HEI_GROUP = 2 then LE8_DIET = 25
Else if LE8_HEI_GROUP = 1 then LE8_DIET = 0
Else if LE8_HEI_GROUP = . then LE8_DIET = .
```

Where LE8_HEI_GROUP is defined as:

```
If HEI2010 = . then LE8_HEI_Group = .
Else if weighted HEI2010 percentile < 25 then LE8_HEI_Group = 1
Else if weighted HEI2010 percentile < 50 then LE8_HEI_Group = 2
Else if weighted HEI2010 percentile < 75 then LE8_HEI_Group = 3
Else if weighted HEI2010 percentile < 95 then LE8_HEI_Group = 4
Else LE8_HEI_Group = 5
```

Response Format: Continuous score from 0 to 100

Source variable(s): HEI2010. Healthy Eating Index 2010.

12.64.7 LE8_GLUCOSE (Life's Essential 8 Blood Glucose component score)

Score based on fasting blood glucose or HbA1c levels.

```
For participants with DIABETES2=3 or DIABETES_SELF=1:
  LABA72 GE 10 then LE8_GLUCOSE = 0
  9.0 LE LABA72 LE 9.9 then LE8_GLUCOSE = 10
  8.0 LE LABA72 LE 8.9 then LE8_GLUCOSE = 20
  7.0 LE LABA72 LE 7.9 then LE8_GLUCOSE = 30
  LABA72 LE 7.0 then LE8_GLUCOSE = 40
```

```
For participants with DIABETES2 NE 3 AND DIABETES_SELF NE 1:
  5.7 LE LABA72 LE 6.4 then LE8_GLUCOSE = 60
  LABA72 < 5.7 then LE8_GLUCOSE = 100
```

Response Format: Continuous score from 0 to 100

Source variable(s):

LABA72. % Glycosylated Hemoglobin (A1C).

DIABETES2. Diabetes2 (3-level grouped Diabetes – ADA).

DIABETES_SELF. Diabetes – self-report only.

12.64.8 LE8_LIPIDS_SCAN (Life's Essential 8 Blood Lipids component score with scanned med)

Score based on non-HDL cholesterol and scanned lipid-lowering medication use.

```
If LE8_NON_HDL = . or MED_LLD = . then LE8_LIPIDS_SCAN = .  
Else if LE8_NON_HDL >= 220 then LE8_LIPIDS_SCAN = 0  
Else if MED_LLD = . then LE8_LIPIDS_SCAN = .  
Else if 190 <= LE8_NON_HDL <= 219 then LE8_LIPIDS_SCAN = 20  
Else if 160 <= LE8_NON_HDL <= 189 then LE8_LIPIDS_SCAN = 40  
Else if 130 <= LE8_NON_HDL <= 159 then LE8_LIPIDS_SCAN = 60  
Else if LE8_NON_HDL < 130 then LE8_LIPIDS_SCAN = 100  
If MED_LLD = 1 then subtract 20 points (unless score is 0)
```

Where LE8_NON_HDL is defined as:

LE8_NON_HDL = LABA66 – LABA68

If LABA66 = . or LABA68 = . then LE8_NON_HDL = .

Response Format: Continuous score from 0 to 100

Source variable(s):

LABA66. Total Cholesterol (mg/dL).

LABA68. HDL cholesterol (mg/dL).

MED_LLD. Scanned Lipid lowering drugs/Antihyperlipidemics (LLD).

12.64.9 LE8_LIPIDS_SELF (Life's Essential 8 Blood Lipids component Score with self-reported med)

Score based on non-HDL cholesterol and self-reported lipid-lowering medication use.

```
If LE8_NON_HDL = . or CHOL_MED_SELF = . then LE8_LIPIDS_SELF = .  
Else if LE8_NON_HDL >= 220 then LE8_LIPIDS_SELF = 0  
Else if LE8_CHOL_MED_SELF = . then LE8_LIPIDS_SELF = .  
Else if 190 <= LE8_NON_HDL <= 219 then LE8_LIPIDS_SELF = 20  
Else if 160 <= LE8_NON_HDL <= 189 then LE8_LIPIDS_SELF = 40  
Else if 130 <= LE8_NON_HDL <= 159 then LE8_LIPIDS_SELF = 60  
Else if LE8_NON_HDL < 130 then LE8_LIPIDS_SELF = 100  
If LE8_CHOL_MED_SELF = 1 then subtract 20 points (unless score is 0)
```

Where LE8_CHOL_MED_SELF is defined as:

If MUEA33E = 1 then LE8_CHOL_MED_SELF = 1

Else if MUEA33E in(0,S) then LE8_CHOL_MED_SELF = 0

Else LE8_CHOL_MED_SELF = .

And LE8_NON_HDL is defined as:

LE8_NON_HDL = LABA66 – LABA68

If LABA66 = . or LABA68 = . then LE8_NON_HDL = .

Response Format: Continuous score from 0 to 100

Source variable(s):

LABA66. Total Cholesterol (mg/dL).

LABA68. HDL cholesterol (mg/dL).

MUEA33E. Self-reported Antihyperlipidemics medication used.

12.64.10 LE8_NICOTINE (Life's Essential 8 Nicotine Exposure component score)

Score based on smoking status and secondhand smoke exposure.

```
If CIGARETTE_USE = . then LE8_NICOTINE = .
Else if CIGARETTE_USE = 3 then LE8_NICOTINE = 0
Else if LE8_NICOTINE_QUIT = . or LE8_NICOTINE_ACTIVE = . then
LE8_NICOTINE = .
Else if CIGARETTE_USE = 2 and LE8_NICOTINE_QUIT < 1 then LE8_NICOTINE
= 25
Else if CIGARETTE_USE = 2 and 1 <= LE8_NICOTINE_QUIT < 5 then
LE8_NICOTINE = 50
Else if CIGARETTE_USE = 2 and LE8_NICOTINE_QUIT >= 5 then
LE8_NICOTINE = 75
Else if CIGARETTE_USE = 1 then LE8_NICOTINE = 100
If LE8_NICOTINE_ACTIVE = 1 then subtract 20 points (unless score is 0)
```

Where LE8_NICOTINE_QUIT is defined as:

If CIGARETTE_USE=1 (never smoked) → Age

If CIGARETTE_USE=2 (former smoker) → Age – TBEA6

If CIGARETTE_USE=3 (current smoker) → 0

Where LE8_NICOTINE_ACTIVE is defined as:

If TBEA14 = . then LE8_NICOTINE_ACTIVE = .

Else if TBEA14 = 0 then LE8_NICOTINE_ACTIVE = 0

Else if TBEA14 in (1:4) then LE8_NICOTINE_ACTIVE = 1

Response Format: Continuous score from 0 to 100

Source variable(s):

CIGARETTE_USE. Cigarette use (1=Never,2=Former,3=Current).

AGE. Age in years at the time of participant's clinic visit.

TBEA6. How old were you when you completely stopped smoking?

TBEA14. Number of smokers in household.

12.64.11 LE8_PA (Life's Essential 8 Physical Activity component score)

Score based on minutes of self-reported moderate/vigorous physical activity per week.

```
If GPAQ_MV_WEEK >= 150 then LE8_PA = 100
Else if 120 <= GPAQ_MV_WEEK <= 149 then LE8_PA = 90
Else if 90 <= GPAQ_MV_WEEK <= 119 then LE8_PA = 80
Else if 60 <= GPAQ_MV_WEEK <= 89 then LE8_PA = 60
Else if 30 <= GPAQ_MV_WEEK <= 59 then LE8_PA = 40
Else if 1 <= GPAQ_MV_WEEK <= 29 then LE8_PA = 20
Else if GPAQ_MV_WEEK = 0 then LE8_PA = 0
Else if GPAQ_MV_WEEK = . then LE8_PA = .
```

Response Format: Continuous score from 0 to 100

Source variables:

GPAQ_MV_WEEK. GPAQ - Total physical activity (min/week).

12.64.12 LE8_SLEEP (Life's Essential 8 Sleep Score using scanned meds)

Score based on average hours of sleep per night.

```
If SLPDUR = . then LE8_SLEEP = .
Else if 7 <= SLPDUR < 9 then LE8_SLEEP = 100
Else if 9 <= SLPDUR < 10 then LE8_SLEEP = 90
Else if 6 <= SLPDUR < 7 then LE8_SLEEP = 70
Else if 5 <= SLPDUR < 6 or SLPDUR >= 10 then LE8_SLEEP = 40
Else if 4 <= SLPDUR < 5 then LE8_SLEEP = 20
Else if SLPDUR < 4 then LE8_SLEEP = 0
```

Response Format: Continuous score from 0 to 100

Source variable(s): SLPDUR. Average sleep duration (hours)

12.65 Life's Simple 7 (LS7) Scores

The Life's Simple 7 (LS7) is a composite cardiovascular health score, ranging from 0 to 14 points, developed by the American Heart Association in 2010. It is constructed as the sum of 7 component scores: diet, physical activity, nicotine exposure, body mass index, blood lipids, blood glucose, and blood pressure. Each component takes integer values 0, 1 or 2.

Reference:

Lloyd-Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, Van Horn L, Greenlund K, Daniels S, Nichol G, Tomaselli GF, Arnett DK, Fonarow GC, Ho PM, Lauer MS, Masoudi FA, Robertson RM, Roger V, Schwamm LH, Sorlie P, Yancy CW, Rosamond WD; American Heart Association Strategic Planning Task Force and Statistics Committee. Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. *Circulation*. 2010 Feb;121(4):586-613. doi:10.1161/CIRCULATIONAHA.109.192703. Epub 2010 Jan 20. [PMID: 20089546](#).

12.65.1 LS7_SCAN (LS7 score using scanned medications)

This version of the composite score used scanned medication data for blood pressure, lipids, and glucose components. If more than 1 of the components are missing, then the composite score is set to missing.

Response Format: Integer score from 0 to 14

Source variables:

LS7_BMI. Body Mass Index Score.

LS7_BP_SCAN. Blood Pressure Score using Scanned Medications.

LS7_DIET. Diet Score.

LS7_GLUCOSE_SCAN. Blood Glucose Score using Scanned Medications.

LS7_LIPIDS_SCAN. Total Cholesterol Score using Lipid Scanned Medications.

LS7_NICOTINE. Nicotine Exposure Score.

LS7_PA. Physical Activity (Self-Reported) Score.

12.65.2 LS7_SELF (LS7 score using reported medications)

This version of the variable used self-reported medication data for blood pressure, lipids, and glucose components. If more than 1 of the components are missing, then the composite score is set to missing.

Response Format: Integer score from 0 to 14

Source variables:

LS7_BMI. Body Mass Index Score.

LS7_BP_SELF. Blood Pressure Score using Self-Reported Medication.

LS7_DIET. Diet Score.

LS7_GLUCOSE_SELF. Blood Glucose Score using Self-Reported Medication.

LS7_LIPIDS_SELF. Total Cholesterol Score using Lipid Self-Reported Medication.

LS7_NICOTINE. Nicotine Exposure Score.

LS7_PA. Physical Activity (Self-Reported) Score.

12.65.3 LS7_BMI (LS7 Body Mass Index Score)

Score based on BMI (kg/m²) using LS7 scoring thresholds.

If BMIGRP_C4 in (1,2) then LS7_BMI = 2
else if BMIGRP_C4 = 3 then LS7_BMI = 1
else if BMIGRP_C4 = 4 then LS7_BMI = 0
else if BMIGRP_C4 = . then LS7_BMI = .

Response Format: 0=poor cardiovascular health
1=intermediate cardiovascular health
2=ideal cardiovascular health

Source variables:

BMIGRP_C4. 4-level grouped Body Mass Index – WHO.

12.65.4 LS7_BP_SCAN (LS7 Blood Pressure component score using scanned meds)

Score based on systolic/diastolic BP and scanned antihypertensive medication use.

```
If SBPA5 = . or SBPA6 = . or MED_ANTIHYPERT = . then LS7_BP_SCAN = .  
Else if SBPA5 >= 140 or SBPA6 >= 90 then LS7_BP_SCAN = 0  
Else if 120 <= SBPA5 < 140 or 80 <= SBPA6 < 90 then LS7_BP_SCAN = 1  
Else if MED_ANTIHYPERT = 1 and (SBPA5 < 120 or SBPA6 < 80) then  
LS7_BP_SCAN = 1  
Else if MED_ANTIHYPERT = 0 and (SBPA5 < 120 or SBPA6 < 80) then  
LS7_BP_SCAN = 2
```

Response Format: 0=poor cardiovascular health
1=intermediate cardiovascular health
2=ideal cardiovascular health

Source variables:

SBPA5. Systolic blood pressure.

SBPA6. Diastolic blood pressure.

MED_ANTIHYPERT. Scanned medication was for treatment of hypertension.

12.65.5 LS7_BP_SELF (LS7 Blood Pressure Component Score using self-reported meds)

Score based on systolic/diastolic BP and self-reported antihypertensive medication use.

```
If SBPA5 = . or SBPA6 = . or LS7_BP_MED_SELF = . then LS7_BP_SELF=.  
Else if SBPA5 >= 140 or SBPA6 >= 90 then LS7_BP_SELF=0  
Else if 120 <= SBPA5 < 140 or 80 <= SBPA6 < 90 then LS7_BP_SELF=1  
Else if LS7_BP_MED_SELF = 1 and (SBPA5 < 120 or SBPA6 < 80) then  
LS7_BP_SELF=1  
Else if LS7_BP_MED_SELF = 0 and (SBPA5 < 120 or SBPA6 < 80) then  
LS7_BP_SELF=2
```

Where LS7_BP_MED_SELF is defined as:

```
If MUEA33D = 1 then LS7_BP_MED_SELF=1  
Else if MUEA33D in(0,.S) then LS7_BP_MED_SELF=0  
Else LS7_BP_MED_SELF = .
```

Response Format: 0=poor cardiovascular health
1=intermediate cardiovascular health
2=ideal cardiovascular health

Source variables:

SBPA5. Systolic blood pressure.

SBPA6. Diastolic blood pressure.

MUEA33D. Self-report anti-hypertensive medication use.

12.65.6 LS7_DIET (LS7 Diet Component Score)

Score based on adherence to LS7 recommendations for five food/nutrient subcomponents: fish, fruits & vegetables, grains, sugar-sweetened beverages, and sodium. The score was only calculated among reliable recalls (DTIA16=0) and after excluding those recalls with extreme energy intake (DTIA20) values (below the 1st percentile or above the 99th percentile by sex and RECALLNUM).

To score the five diet subcomponents, we first averaged the values of the source variables (over 1 or 2 dietary recalls) for each participant. Then, we summed up the averaged values (daily intake) for fish and sugar-sweetened beverages, we also multiplied by 7 to obtain weekly intake. We then created the diet subcomponents which are binary variables following Lloyd-Jones DM et al (Circulation, 2010) and as shown in table below.

Then, we created another intermediate variable called LS7_DIET_SUM (range 0 to 5) by summing up the five subcomponents. Participants missing both dietary recalls are assigned missing values for the diet component score for participants with at least one dietary recall. There is no missing data for the sub-components.

Lastly, LS7_DIET was scored as

IF LS7_DIET_SUM in (0,1) then LS7_DIET = 0
 ELSE IF LS7_DIET_SUM in (2,3) then LS7_DIET = 1
 ELSE IF LS7_DIET_SUM in (4,5) then LS7_DIET = 2

DIET SUBCOMPONENT	SOURCE VARIABLES	INTERMEDIATE VARIABLE	SUBCOMPONENT SCORING
LS7_DIET_FISH (0 or 1)	DTSA69, DTSA70, DTSA71, DTSA72, DTSA73	DIET_FISH_WEEK (Average fish intake, serv/wk)	If 0 <= DIET_FISH_WEEK < 7 then LS7_DIET_FISH = 0 Else if DIET_FISH_WEEK >= 7 then LS7_DIET_FISH = 1
LS7_DIET_FV (0 or 1)	DTSA4, DTSA5, DTSA6, DTSA7, DTSA8, DTSA11, DTSA12, DTSA13, DTSA16, DTSA17, DTSA18, DTSA19, DTSA20	DIET_FV (Average fruit and vegetables intake, serv/day)	If 0 LE DIET_FV < 4.5 then LS7_DIET_FV = 0 Else if DIET_FV >= 4.5 then LS7_DIET_FV = 1
LS7_DIET_GRAINS (0 or 1)	DTSA22, DTSA23, DTSA25, DTSA26, DTSA28, DTSA31, DTSA34, DTSA37, DTSA40, DTSA46, DTSA47	DIET_GRAINS (Average grains intake, serv/day)	if 0 LE DIET_GRAINS < 3 then LS7_DIET_GRAINS = 0 when DIET_GRAINS GE 3 then LS7_DIET_GRAINS = 1
LS7_DIET_SSB (0 or 1)	DTSA134, DTSA137, DTSA139, DTSA142, DTSA145, DTSA148, DTSA151	DIET_SSB_WEEK (Average sugar- sweetened beverages intake, serv/wk)	if 0 LE DIET_SSB_WEEK < 4.5 then LS7_DIET_SSB = 1 when DIET_SSB_WEEK GE 4.5 then LS7_DIET_SSB = 0
LS7_DIET_SODIUM (0 or 1)	DTIA67	DIET_SODIUM (Average sodium intake, mg/day)	if 0 LE DIET_SODIUM < 1500 then LS7_DIET_SODIUM =1 when DIET_SODIUM GE 1500 then LS7_DIET_SODIUM = 0

Response Format: 0=poor cardiovascular health
1=intermediate cardiovascular health
2=ideal cardiovascular health

Source variables:

DTIA16. Intake Reliability.
DTIA20. Energy (kcal).
DTSA4. FRU0100 Citrus Juice.
DTSA5. FRU0200 Fruit Juice excluding Citrus Juice.
DTSA6. FRU0300 Citrus Fruit.
DTSA7. FRU0400 Fruit excluding Citrus Fruit.
DTSA8. FRU0500 Avocado and Similar.
DTSA11. VEG0100 Dark-green Vegetables.
DTSA12. VEG0200 Deep-yellow Vegetables.
DTSA13. VEG0300 Tomato.
DTSA16. VEG0450 Other Starchy Vegetables.
DTSA17. VEG0700 Legumes (cooked dried beans).
DTSA18. VEG0600 Other Vegetables.
DTSA19. VEG0900 Fried Vegetables.
DTSA20. VEG0500 Vegetable Juice.
DTSA22. GRW0100 Grains, Flour and Dry Mixes - Whole Grain.
DTSA23. GRS0100 Grains, Flour and Dry Mixes - Some Whole Grain.
DTSA25. GRW0200 Loaf-type Bread and Plain Rolls - Whole Grain.
DTSA26. GRS0200 Loaf-type Bread and Plain Rolls - Some Whole Grain.
DTSA28. GRW0300 Other Breads (quick breads, corn muffins, tortillas) - Whole Grain.
DTSA31. GRW0400 Crackers - Whole Grain.
DTSA34. GRW0500 Pasta - Whole Grain.
DTSA37. GRW0600 Ready-to-eat Cereal (not presweetened) - Whole Grain.
DTSA40. GRW0700 Ready-to-eat Cereal (presweetened) - Whole Grain.
DTSA46. GRW1000 Snack Bars - Whole Grain.
DTSA47. GRS1000 Snack Bars - Some Whole Grain.
DTSA69. MFF0100 Fish - Fresh and Smoked.
DTSA70. MFL0100 Lean Fish - Fresh and Smoked.
DTSA71. MFF0200 Fried Fish - Commercial Entrée and Fast Food.
DTSA72. MSL0100 Shellfish.
DTSA73. MSF0100 Fried Shellfish - Commercial Entrée and Fast Food.
DTSA134. BVS0400 Sweetened Soft Drinks.
DTSA137. BVS0300 Sweetened Fruit Drinks.
DTSA139. BVS0500 Sweetened Tea.
DTSA142. BVS0100 Sweetened Coffee.
DTSA145. BVS0200 Sweetened Coffee Substitutes.
DTSA148. BVS0600 Sweetened Water.
DTSA151. BVS0700 Nondairy-based Sweetened Meal Replacement/Supplement.

12.65.7 LS7_GLUCOSE_SCAN (LS7 Blood Glucose Score using scanned meds)

Score based on fasting blood glucose or scanned diabetes medication use.

If MED_ANTIDIAB or LABA70 is missing then LS7_GLUCOSE_SCAN = .
Else if LABA70 >= 126 then LS7_GLUCOSE_SCAN = 0
Else If 100 <= LABA70 <= 125 then LS7_GLUCOSE_SCAN = 1
Else If MED_ANTIDIAB = 1 and LABA70 < 100 then LS7_GLUCOSE_SCAN = 1
Else If MED_ANTIDIAB = 0 and LABA70 < 100 then LS7_GLUCOSE_SCAN = 2

Response Format: 0=poor cardiovascular health
1=intermediate cardiovascular health
2=ideal cardiovascular health

Source variables:

LABA70. Fasting glucose (mg/dL).

MED_ANTIDIAB. Scanned antidiabetics medication.

12.65.8 LS7_GLUCOSE_SELF (LS7 Blood Glucose Score using self-reported meds)

Score based on fasting blood glucose or self-reported diabetes medication use.

If SELFMED_ANTIDIAB or LABA70 is missing then LS7_GLUCOSE_SELF = .
Else if LABA70 >= 126 then LS7_GLUCOSE_SELF = 0
Else If 100 <= LABA70 <= 125 then LS7_GLUCOSE_SELF = 1
Else If SELFMED_ANTIDIAB = 1 and LABA70 < 100 then LS7_GLUCOSE_SELF = 1
Else If SELFMED_ANTIDIAB = 0 and LABA70 < 100 then LS7_GLUCOSE_SELF = 2

Response Format: 0=poor cardiovascular health
1=intermediate cardiovascular health
2=ideal cardiovascular health

Source variables:

LABA70. Fasting glucose (mg/dL).

SELFMED_ANTIDIAB. Anti-Diabetics Medication used – Self-reported.

12.65.9 LS7_LIPIDS_SCAN (LS7 Total Cholesterol Score using lipid scanned med)

Score based on total cholesterol and scanned lipid-lowering medication use.

If LABA66 or MED_LLD is missing then LS7_LIPIDS_SCAN = .
Else if LABA66 >= 240 then LS7_LIPIDS_SCAN = 0
Else If 200 <= LABA66 < 240 then LS7_LIPIDS_SCAN = 1
Else If MED_LLD = 1 and LABA66 < 200 then LS7_LIPIDS_SCAN = 1
Else If MED_LLD = 0 and LABA66 < 200 then LS7_LIPIDS_SCAN = 2

Response Format: 0=poor cardiovascular health
1=intermediate cardiovascular health
2=ideal cardiovascular health

Source variables:

LABA66. Total Cholesterol (mg/dL).

MED_LLD. Scanned Lipid lowering drugs/Antihyperlipidemics (LLD).

12.65.10 LS7_LIPIDS_SELF (LS7 Total Cholesterol Score using lipid self-reported med)

Score based on total cholesterol and self-reported lipid-lowering medication use.

If LABA66 or LS7_CHOL_MED_SELF is missing then LS7_LIPIDS_SELF = .

Else if LABA66 >= 240 then LS7_LIPIDS_SELF = 0

Else If 200 <= LABA66 < 240 then LS7_LIPIDS_SELF = 1

Else If LS7_CHOL_MED_SELF = 1 and LABA66 < 200 then LS7_LIPIDS_SELF = 1

Else If LS7_CHOL_MED_SELF = 0 and LABA66 < 200 then LS7_LIPIDS_SELF = 2

Where LS7_CHOL_MED_SELF is defined as:

If MUEA33E = 1 then LS7_CHOL_MED_SELF = 1

Else if MUEA33E in(0,.S) then LS7_CHOL_MED_SELF = 0

Else LS7_CHOL_MED_SELF = .

Response Format: 0=poor cardiovascular health
1=intermediate cardiovascular health
2=ideal cardiovascular health

Source variables:

LABA66. Total Cholesterol (mg/dL).

MUEA33E. Self-reported Antihyperlipidemics medication used.

12.65.11 LS7_NICOTINE (LS7 Nicotine Exposure Component Score)

Score based on smoking status and secondhand smoke exposure.

If CIGARETTE_USE = . or LS7_NICOTINE_QUIT = . then .

Else if CIGARETTE_USE = 3 then 0

Else if CIGARETTE_USE = 2 and LS7_NICOTINE_QUIT <= 1 then 1

Else if CIGARETTE_USE = 1 or LS7_NICOTINE_QUIT > 1 then 2

Where LS7_NICOTINE_QUIT is defined as:

If CIGARETTE_USE=1 (never smoked) → Age

If CIGARETTE_USE=2 (former smoker) → Age – TBEA6

If CIGARETTE_USE=3 (current smoker) → 0

Response Format: 0=poor cardiovascular health
1=intermediate cardiovascular health
2=ideal cardiovascular health

Source variables:

CIGARETTE_USE. Cigarette use (1=Never,2=Former,3=Current).

AGE. Age in years at the time of participant's baseline clinic visit.

TBEA6. How old were you when you completely stopped smoking?

12.65.12 LS7_PA (LS7 Physical Activity Score)

Score based on minutes of self-reported moderate/vigorous physical activity per week.

If (GPAQ_MOD_WEEK or GPAQ_VIG_WEEK) and GPAQ_MV_WEEK are missing
then LS7_PA = .
Else if GPAQ_MOD_WEEK >= 150 or GPAQ_MV_WEEK >= 150 or
GPAQ_VIG_WEEK >= 75 then LS7_PA = 2
Else if 1 <= GPAQ_MOD_WEEK < 150 or 1 <= GPAQ_MV_WEEK < 150 or 1 <=
GPAQ_VIG_WEEK < 75 then LS7_PA = 1
Else LS7_PA = 0

Response Format: 0=poor cardiovascular health
1=intermediate cardiovascular health
2=ideal cardiovascular health

Source variables:

GPAQ_MV_WEEK. GPAQ - Total physical activity (min/week).

12.65.13 LS7_SCAN_C3 (3-level LS7 score using scanned meds)

Three category score for Life's Simple 7 (LS7), taking integer values 0, 1 or 2. This metric serves as an analog from the individual component ideal/intermediate/ideal classification to the entire cardiovascular health score, with scores of 0, 1, and 2 corresponding to poor, intermediate, and ideal, respectively. This version of the variable used scanned medication data for blood pressure, lipids, and glucose components. If more than 1 of the components are missing, then the composite score is set to missing.

Reference:

Hasbani NR, Ligthart S, Brown MR, Heath AS, Bebo A, Ashley KE, Boerwinkle E, Morrison AC, Folsom AR, Aguilar D, de Vries PS; American Heart Association. American Heart Association's Life's Simple 7: Lifestyle Recommendations, Polygenic Risk, and Lifetime Risk of Coronary Heart Disease. *Circulation*. 2022 Apr;145(10):808-818. doi:10.1161/CIRCULATIONAHA.121.053730. Epub 2022 Mar 1. PMID: 35221349.

If LS7_SCAN is missing then LS7_SCAN_C3 = .
If the sum of LS7_SCAN is in 0 to 4 then LS7_SCAN_C3 = 0
If the sum of LS7_SCAN is in 5 to 9 then LS7_SCAN_C3 = 1
If the sum of LS7_SCAN is in 10 to 14 then LS7_SCAN_C3 = 2

Response Format: 0=poor cardiovascular health
1=intermediate cardiovascular health
2=ideal cardiovascular health

Source variables:

LS7_SCAN. LS7 score using scanned medications.

12.65.14 LS7_SELF_C3 (3-level LS7 score using self-reported medications)

Three category score for Life's Simple 7 (LS7), taking integer values 0, 1 or 2. This metric serves as an analog from the individual component ideal/intermediate/ideal classification to the entire cardiovascular health score, with scores of 0, 1, and 2 corresponding to poor, intermediate, and ideal, respectively. This version of the variable used self-reported medication data for blood pressure, lipids, and glucose components. If more than 1 of the components are missing, then the composite score is set to missing.

Reference:

Hasbani NR, Ligthart S, Brown MR, Heath AS, Bebo A, Ashley KE, Boerwinkle E, Morrison AC, Folsom AR, Aguilar D, de Vries PS; American Heart Association. American Heart Association's Life's Simple 7: Lifestyle Recommendations, Polygenic Risk, and Lifetime Risk of Coronary Heart Disease. *Circulation*. 2022 Apr;145(10):808-818. doi:10.1161/CIRCULATIONAHA.121.053730. Epub 2022 Mar 1. PMID: 35221349.

If LS7_SELF is missing then LS7_SELF_C3 = .
If the sum of LS7_SELF is in 0 to 4 then LS7_SELF_C3 = 0
If the sum of LS7_SELF is in 5 to 9 then set to LS7_SELF_C3 = 1
If the sum of LS7_SELF is in 10 to 14 then set to LS7_SELF_C3 = 2

Response Format: 0=poor cardiovascular health
1=intermediate cardiovascular health
2=ideal cardiovascular health

Source variables:

LS7_SELF. LS7 score using reported medications.

12.65.15 LS7_SCAN_C2 (Indicator of 4+ LS7 components (ideal) using scanned meds)

Binary score for Life's Simple 7 (LS7) based on number of ideal components. A component is considered "ideal" if it has a score of 2. This version of the variable used scanned medication data for blood pressure, lipids, and glucose components. If more than 1 of the components are missing, then the composite score is set to missing.

Response Format: 0=less than 4 ideal components
1=at least 4 ideal components

Source variables:

LS7_BMI. Body Mass Index Score.

LS7_BP_SCAN. Blood Pressure Score using Scanned Medication.

LS7_DIET. Diet Score.

LS7_GLUCOSE_SCAN. Blood Glucose Score using Scanned Medication.

LS7_LIPIDS_SCAN. Total Cholesterol Score using Lipid Scanned Medication.

LS7_NICOTINE. Nicotine Exposure Score.

LS7_PA. Physical Activity (Self-Reported) Score.

12.65.16 LS7_SELF_C2 (Indicator of 4+ LS7 components (ideal) using self-reported meds)

Binary score for Life's Simple 7 (LS7) based on number of ideal components. A component is considered "ideal" if it has a score of 2. A value of 0 corresponds to less than 4 ideal components and 1 to at least 4. This version of the variable used self-reported medication data for blood pressure, lipids, and glucose components. If more than 1 of the components are missing, then the composite score is set to missing.

Response Format: 0=less than 4 ideal components
1=at least 4 ideal components

Source variables:

LS7_BMI. Body Mass Index Score.

LS7_BP_SELF. Blood Pressure Score using Self-Reported Medication.

LS7_DIET. Diet Score.

LS7_GLUCOSE_SELF. Blood Glucose Score using Self-Reported Medication.

LS7_LIPIDS_SELF. Total Cholesterol Score using Lipid Self-Reported Medication.

LS7_NICOTINE. Nicotine Exposure Score.

LS7_PA. Physical Activity (Self-Reported) Score.

13. MEDICATIONS – SCANNED/TRANSCRIBED ONLY

All participants are asked to bring in all prescribed or over-the-counter medications taken in the past 4 weeks to the baseline examination for inventory. MUEA1 and MUEA2 are used to define individuals who did/did not bring in any medications as well as those who have/have not taken any medications in the past 4 weeks in order to determine absence of medication use. Medications are then scanned or transcribed into medication coding fields MUEA5A-MUEA5D to MUEA29A-MUEA29D and MUEA39A-MUEA39D (up to 26 possible medications). For the derived indicator variables below, participants were set to missing unless the reported MUEA2=1 (took no medication) or reported at least 1 medication in the fields above.

The medication information is then compared to the HCHS master drug file which draws data from several sources:

1. Medi-Span's Master Drug Data Base (MDDDB ®, version 2, 2003)
2. Lexi-Comp ® Online value-added version of Index Nominum Drug Data Base (version 2009)
3. Ovid's Martindale Complete Drug Reference (version 2009)

Mexican (MX) and Dominican (DO) medications from (2-3) above were extracted and then assigned therapeutic classification codes associated with their generic ingredients in (1). Assignments were automatic when data retrieval from (2-3) was fully electronic and generic ingredients in the source files matched exactly. Assignments were manual when data retrieval from (2-3) was semi-manual or generic ingredients in the source files did not match exactly and required review by a health professional.

The HCHS master drug file “contains basic drug names, National Drug Codes (NDCs), and assorted descriptive information used to uniquely identify drug products. From this data, generic product identifiers (GPI) are obtained which groups drug products by a hierarchical therapeutic classification scheme. The 14-character GPI consists of a hierarchy of seven subsets, each providing increasingly more specific information about drug products. These subsets are structured and identified below” (from MDDDB ®, version 2, 2003):

GPI Subset	Record Type	Size	Representation	Example
12-xx-xx-xx-xx-xx-xx	1	2	Drug Group	*MISC. ENDOCRINE*
12-34-xx-xx-xx-xx-xx	2	4	Drug Class	*Posterior Pituitary**
12-34-56-xx-xx-xx-xx	3	6	Drug Subclass	*Vasopressin***
12-34-56-78-xx-xx-xx		8	Drug Name	Desmopressin
12-34-56-78-90-xx-xx	4	10	Drug Name	Acetate
12-34-56-78-90-12-xx		12	Dosage Form	Tablet
12-34-56-78-90-12-34	5	14	Strength	0.1MG

*NOTE: The below medications were indicated for use within the last 4 weeks from baseline examination.

Hormone Replacement Therapies:

13.1 MED_CURRENT_HRT (Any Hormone Repl. Therapy)

This is a 0/1 numeric variable that indicates use of at least one of the following HRT medications:

Med_Estrogens
Med_Estrogen_Combo
Med_Contraceptives
Med_Progestins
Med_SERM
Med_Fertility_Reg
Med_LHRH
Med_LHRH_AG
Med_LHRH_ANT

13.2 MED_ESTROGENS (Estrogens)

This is a 0/1 numeric variable.

Medication Codes:

The “24xxxx” group of medications: *ESTROGENS*

13.3 MED_ESTROGEN_COMBO (Estrogens – Combinations Only)

This is a 0/1 numeric variable.

Medication Codes:

The “2499xx” class of medications: *Estrogen Combinations**

13.4 MED_CONTRACEPTIVES (Contraceptives)

This is a 0/1 numeric variable.

Medication Codes:

The “25xxxx” group of medications: *CONTRACEPTIVES*

13.5 MED_PROGESTINS (Progestins)

This is a 0/1 numeric variable.

Medication Codes:

The “26xxxx” group of medications: *PROGESTINS*

13.6 MED_SERM (Selective Estrogen Receptor Modulators)

This is a 0/1 numeric variable.

Medication Codes:

The “300530” subclass of medications: *Selective Estrogen Receptor Modulators (SERMs)***

13.7 MED_FERTILITY_REG (Fertility Regulators)

This is a 0/1 numeric variable.

Medication Codes:

The “3006xx” class of medications:

13.8 MED_LHRH (LHRH/GnRH Only)

This is a 0/1 numeric variable.

Medication Codes:

The “300700” class of medications: *Luteinizing Hormone Releasing-Hormones (LHRH/GnRH)**

13.9 MED_LHRH_AG (LHRH/GnRH Agonist Analog)

This is a 0/1 numeric variable.

Medication Codes:

The “300800” subclass of medications: *LHRH/GnRH Agonist Analog Pituitary Suppressants**

13.10 MED_LHRH_ANT (LHRH/GnRH Antagonists)

This is a 0/1 numeric variable.

Medication Codes:

The “300900” subclass of medications: *GnRH/LHRH Antagonists**

Other Medications:

13.11 MED_ANTIANGINAL (Antianginals)

This is a 0/1 numeric variable.

Medication Codes:

The “32xxxx” group of medications: *ANTIANGINAL AGENTS*

13.12 MED_ANTIARRHYTHMICS (Antiarrhythmics)

This is a 0/1 numeric variable.

Medication Codes:

The “35xxxx” group of medications: *ANTIARRHYTHMICS*

13.13 MED_ANTIASTHMATICS (Antiasthmatic or bronchodilator agents)

This is a 0/1 numeric variable.

Medication Codes:

The “44xxxx” group of medications: *ANTIASTHMATIC AND BRONCHODILATOR AGENTS*

13.14 MED_ANTICOAG (Anticoagulants)

This is a 0/1 numeric variable.

Medication Codes:

Any “83xxxx” group of medications: *ANTICOAGULANTS*

13.15 MED_ANTICOAG_COUMARIN (Coumarin anticoagulants)

This is a 0/1 numeric variable.

Medication Codes:

Any “8320xx” class of medications: *Coumarin Anticoagulants**

13.16 MED_ANTICOAG_HEPARIN (Heparin anticoagulants)

This is a 0/1 numeric variable.

Medication Codes:

Any “8310xx” class of medications: *Heparins And Heparinoid-Like Agents**

13.17 MED_ANTIDIAB (Antidiabetics)

This is a 0/1 numeric variable.

Medication Codes:

Any “27xxxx” group of medications: *ANTIDIABETICS*

13.18 MED_ANTIHYPERT (Antihypertensives)

This is a 0/1 numeric variable.

Medication Codes:

Any “36xxxx” group of medications: *ANTIHYPERTENSIVES*

13.19 MED_ANTIHYPERT_ACEI (ACE Inhibitors)

This is a 0/1 numeric variable.

Medication Codes:

Any "3610xx" class of medications: *ACE Inhibitors**

Any "369915" subclass of medications: *ACE Inhibitor & Calcium Channel Blocker Combinations***

Any "369918" subclass of medications: *ACE Inhibitors & Thiazide/Thiazide-Like***

13.20 MED_ANTIHYPERT_AT2RAS (Angiotensin II Receptor Antagonists)

This is a 0/1 numeric variable.

Medication Codes:

Any "361500" subclass of medications: *Angiotensin II Receptor Antagonists**

Any "369940" subclass of medications: *Angiotensin II Receptor Antagonists & Thiazides***

13.21 MED_ANTIPLATELET (Antiplatelets)

This is a 0/1 numeric variable.

Medication Codes:

Any "8515xx" class of medications: *Platelet Aggregation Inhibitors**

13.22 MED_ASPIRIN (Aspirin-containing analgesics)

This is a 0/1 numeric variable.

Medication Codes:

Any "6410xx" class of medications: *Salicylates**

Any "6420xx" class of medications: *Analgesics Other**

Any "6499xx" class of medications: *Analgesic Combinations**

Any "6599xx" class of medications: *Narcotic Combinations**

13.23 MED_BB (Beta Blockers)

This is a 0/1 numeric variable.

Medication Codes:

Any "33xxxx" group of medications: *BETA BLOCKERS*

Any "369920" subclass of medications: *Beta Blocker & Diuretic Combinations***

Any "369925" subclass of medications: *Beta Blocker & Calcium Channel Blocker Combinations***

12/2019 update: Any "369988" subclass of medications: *Beta Blocker-Nutritional supplement combinations***

13.24 MED_BB_Ophthalmic (Ophthalmic Beta Blockers)

This is a 0/1 numeric variable.

Medication Codes:

Any "862500" subclass of medications: *Beta-blockers - Ophthalmic**

Any "862599" subclass of medications: *Beta-blockers - Ophthalmic Combinations***

13.25 MED_CACB (Calcium Channel Blockers)

This is a 0/1 numeric variable.

Medication Codes:

Any "34xxxx" group of medications:

Any "369915" subclass of medications: *CALCIUM CHANNEL BLOCKERS*

Any "369925" subclass of medications: *Beta Blocker & Calcium Channel Blocker Combinations***

Any "409925" subclass of medications: *Calcium Channel Blocker & HMG CoA Reductase Inhibit Comb***

12/2019 update: Any "369930" subclass of medications: *Angiotensin II Receptor Antagonist & Calcium Channel Blocker Combinations***

13.26 MED_CARDIACGLYCOSIDES (Cardiac Glycosides)

This is a 0/1 numeric variable.

Any "312000" subclass of medications: *Cardiac Glycosides**

13.27 MED_CHEMO (Chemotherapy)

This is a 0/1 numeric variable.

Medication Codes:

Any "21xxxx" group of medications: *ANTINEOPLASTICS*

13.28 MED_CLOPID (Clopidogrel (Plavix))

This is a 0/1 numeric variable.

Medication Codes:

The "851580201" subclass of medications: *Thienopyridine Derivatives***

13.29 MED_COX2 (COX-2 inhibitors)

This is a 0/1 numeric variable.

Medication Codes:

Any "661005" subclass of medications: *Cyclooxygenase 2 (COX-2) Inhibitors***

13.30 MED_DIURETIC (Diuretics)

This is a 0/1 numeric variable.

Medication Codes:

Any "369918" subclass of medications: *ACE Inhibitors & Thiazide/Thiazide-Like***

Any "369920" subclass of medications: *Beta Blocker & Diuretic Combinations***

Any "369940" subclass of medications: *Angiotensin II Receptor Antagonists & Thiazides***

Any "369950" subclass of medications: *Adrenolytics-Central & Thiazide Combinations***

Any "369955" subclass of medications: *Adrenolytics-Peripheral & Thiazides***

Any "369970" subclass of medications: *Antihypertensives-MAOIs & Thiazides***

Any "369980" subclass of medications: *Antihypertensives-MISC & Thiazides***

Any "369990" subclass of medications: *Vasodilators & Thiazides***

Any "37xxxx" group of medications: *DIURETICS*

13.31 MED_DIURETIC_THIAZIDE (Thiazide Diuretics)

This is a 0/1 numeric variable.

Medication Codes:

Any "369918" subclass of medications: *ACE Inhibitors & Thiazide/Thiazide-Like***

Any "369920" subclass of medications: *Beta Blocker & Diuretic Combinations***

Any "369940" subclass of medications: *Angiotensin II Receptor Antagonists & Thiazides***

Any "369950" subclass of medications: *Adrenolytics-Central & Thiazide Combinations***

Any "369955" subclass of medications: *Adrenolytics-Peripheral & Thiazides***

Any "369970" subclass of medications: *Antihypertensives-MAOIs & Thiazides***

Any "369980" subclass of medications: *Antihypertensives-MISC & Thiazides***

Any "369990" subclass of medications: *Vasodilators & Thiazides***

Any "376000" subclass of medications: *Thiazides and Thiazide-Like Diuretics**

13.32 MED_FIBARES_NICOACID (Fibric/Nicotinic Acids (trt of TG and HDL))

This is a 0/1 numeric variable.

Medication Codes:

Any "392000" subclass of medications: *Fibric Acid Derivatives**

Any "394500" subclass of medications: *Nicotinic Acid Derivatives**

Any "399920" subclass of medications: *Fibric Acid Derivative Combinations***

13.33 MED_INSULIN (Insulins)

This is a 0/1 numeric variable.

Medication Codes:

Any "2710xx" class of medications: *Insulin**

13.34 MED_LLD (Lipid lowering drugs/Antihyperlipidemics (LLD))

This is a 0/1 numeric variable.

Medication Codes:

Any "39xxxx" group of medications: *ANTIHYPERTENSIVES* plus the class of "409925" (*Calcium Channel Blocker & HMG CoA Reductase Inhibitor Combination***).

13.35 MED_METFORMIN (Metformins)

This is a 0/1 numeric variable.

Medication Codes:

Any "2725xx" class of medications: *Biguanides**

Any "2799xx" class of medications: *Antidiabetic Combinations**

13.36 MED_NSAID (NSAIDs)

This is a 0/1 numeric variable.

Medication Codes:

Any "6610xx" class of medications: *Nonsteroidal Anti-inflammatory Agents (NSAIDs)**

13.37 MED_OI_STEROID (Oral/inhalable Corticosteroids)

This is a 0/1 numeric variable. Does not include not including mineralocorticoids, androgens, estrogens, and progestagens.

Medication Codes:

The "221000" subclass of medications: *Glucocorticosteroids**

The "221099" subclass of medications: *Steroid Combinations***

The "444000" subclass of medications: *Steroid Inhalants**

The "449947" subclass of medications: *Sympathomimetic-Steroid w/ Anti-infective***

The "449980" subclass of medications: *Steroid-Antihistamines***

The "449982" subclass of medications: *Steroid-Antihistamine-Expectorants***

The "449984" subclass of medications: *Steroid-Sympathomimetic-Antihistamine-&/or Expectorant***

The "449988" subclass of medications: *Xanthine-Steroids***

13.38 MED_STATIN (Statins)

This is a 0/1 numeric variable.

Medication Codes:

Any "3940xx" class of medications HMG CoA Reductase Inhibitors

12/2019 update: Any "399940" subclass of medications: *Intestinal Cholesterol Absorption Inhibitor-HMG CoA Reductase Inhibitor Combinations***

12/2019 update: Any "409925" subclass of medications: *Calcium Channel Blocker & HMG CoA Reductase Inhibitor Combinations***

13.39 MED_ANTIANXI (Antianxiety agents)

This is a 0/1 numeric variable.

Medication Codes:

Any "57xxxx" class of medications: "ANTIANXIETY AGENTS"

13.40 MED_ANTIDEPRESS (Antidepressants)

This is a 0/1 numeric variable.

Medication Codes:

Any "58xxx" class of medications: "ANTIDEPRESSANTS"

13.41 MED_ANTIDEPRE_SSRI (Antidepressants - selective serotonin re-uptake inhibitors)

This is a 0/1 numeric variable.

Medication Codes:

Any "5816xx" class of medications: "SELECTIVE SEROTONIN REUPTAKE INHIBITORS"

13.42 MED_ANTIDEPRE_TRICYC (Antidepressants - tricyclic antidepressants)

This is a 0/1 numeric variable.

Medication Codes:

Any "5820xx" class of medications: "TRICYCLIC AGENTS"

13.43 MED_ANTIPSYCHO (Antipsychotics)

This is a 0/1 numeric variable.

Medication Codes:

Any "59xxx" class of medications: "ANTIPSYCHOTICS"

13.44 MED_CORTICOSTEROIDS_INHALED (Inhaled Corticosteroid medication)

This is a 0/1 numeric variable.

Medication Codes:

The "221000" subclass of medications: "GLUCOCORTICOSTEROIDS"

The "221099" subclass of medications: "STEROID COMBINATIONS"

13.45 MED_PDE4 (Phosphodiesterase IV Inhibitor Medication)

This is a 0/1 numeric variable.

Medication Codes:

Any "311000" subclass of medications: "PHOSPHODIESTERASE INHIBITORS"

13.46 MED_BRONCHODILATOR (Bronchodilator and Antiasthmatic Agents medication)

This is a 0/1 numeric variable.

Medication Codes:

The "441000" subclass of medications: "BRONCHODILATORS - ANTICHOLINERGICS"
The "442010" subclass of medications: "BETA ADRENERGICS"
The "442020" subclass of medications: "MIXED ADRENERGICS"
The "443000" subclass of medications: "XANTHINES"
The "444000" subclass of medications: "STEROID INHALANTS"
The "445000" subclass of medications: "LEUKOTRIENE MODULATORS"
The "445040" subclass of medications: "5-LIPOXYGENASE INHIBITORS"
The "445050" subclass of medications: "LEUKOTRIENE RECEPTOR ANTAGONISTS"
The "446000" subclass of medications: "ANTIASTHMATIC - MONOCLONAL ANTIBODIES"
The "446030" subclass of medications: "ANTI-IGE MONOCLONAL ANTIBODIES"
The "449900" subclass of medications: "ASTHMA AND BRONCHODILATOR AGENT COMBINATIONS"
The "449905" subclass of medications: "XANTHINE MIXTURES"
The "449910" subclass of medications: "XANTHINE - EXPECTORANTS"
The "449920" subclass of medications: "XANTHINE - SYMPATHOMIMETICS"
The "449922" subclass of medications: "XANTHINE – SYMPATHOMIMETIC - EXPECTORANT"
The "449925" subclass of medications: "XANTHINE - BARBITURATES"
The "449927" subclass of medications: "SYMPATHOMIMETIC - BARBITURATE"
The "449928" subclass of medications: "XANTHINE-ANTIHISTAMINE-EXPECTORANT"
The "449930" subclass of medications: "XANTHINE – SYMPATHOMIMETIC - BARBITURATE"
The "449931" subclass of medications: "Xanthine-Sympathomimetic-Barbiturate-Antihistamine"
The "449932" subclass of medications: "XANTHINE – SYMPATHO – BARBIT - EXPECTOR"
The "449950" subclass of medications: "XANTHINE - ANTITUSSIVE"
The "449952" subclass of medications: "XANTHINE-ANTITUSSIVE-ANTIHISTAMINE"
The "449955" subclass of medications: "XANTHINE-ANTITUSSIVE-EXPECTORANT"
The "449985" subclass of medications: "XANTHINE-ANTIHISTAMINES"
The "449988" subclass of medications: "Xanthine-Steroids"
The "449990" subclass of medications: "Xanthine-Sympathomimetic-Antihistamine-&/or Expectorant"
The "449993" subclass of medications: "Xanthine-Sympathomimetic-Antitussive-Expectorant"
The "449995" subclass of medications: "Xanthine-Sympath-Antihistamine-Antitussive-Expectorant"

13.47 MED_PULMONARY (Pulmonary medication)

This is a 0/1 numeric variable.

Medication Codes:

The "221000" subclass of medications: "GLUCOCORTICOSTEROIDS"
The "221099" subclass of medications: "STEROID COMBINATIONS"
The "311000" subclass of medications: "PHOSPHODIESTERASE INHIBITORS"
The "441000" subclass of medications: "BRONCHODILATORS - ANTICHOLINERGICS"
The "442010" subclass of medications: "BETA ADRENERGICS"
The "442020" subclass of medications: "MIXED ADRENERGICS"
The "443000" subclass of medications: "XANTHINES"

The "444000" subclass of medications: "STEROID INHALANTS"
 The "445000" subclass of medications: "LEUKOTRIENE MODULATORS"
 The "445040" subclass of medications: "5-LIPOXYGENASE INHIBITORS"
 The "445050" subclass of medications: "LEUKOTRIENE RECEPTOR ANTAGONISTS"
 The "446000" subclass of medications: "ANTIASTHMATIC - MONOCLONAL ANTIBODIES"
 The "446030" subclass of medications: "ANTI-IGE MONOCLONAL ANTIBODIES"
 The "449900" subclass of medications: "ASTHMA AND BRONCHODILATOR AGENT COMBINATIONS"
 The "449905" subclass of medications: "XANTHINE MIXTURES"
 The "449910" subclass of medications: "XANTHINE - EXPECTORANTS"
 The "449920" subclass of medications: "XANTHINE - SYMPATHOMIMETICS"
 The "449922" subclass of medications: "XANTHINE – SYMPATHOMIMETIC - EXPECTORANT"
 The "449925" subclass of medications: "XANTHINE - BARBITURATES"
 The "449927" subclass of medications: "SYMPATHOMIMETIC - BARBITURATE"
 The "449928" subclass of medications: "XANTHINE-ANTIHISTAMINE-EXPECTORANT"
 The "449930" subclass of medications: "XANTHINE – SYMPATHOMIMETIC - BARBITURATE"
 The "449931" subclass of medications: "Xanthine-Sympathomimetic-Barbiturate-Antihistamine"
 The "449932" subclass of medications: "XANTHINE – SYMPATHO – BARBIT - EXPECTOR"
 The "449950" subclass of medications: "XANTHINE - ANTITUSSIVE"
 The "449952" subclass of medications: "XANTHINE-ANTITUSSIVE-ANTIHISTAMINE"
 The "449955" subclass of medications: "XANTHINE-ANTITUSSIVE-EXPECTORANT"
 The "449985" subclass of medications: "XANTHINE-ANTIHISTAMINES"
 The "449988" subclass of medications: "Xanthine-Steroids"
 The "449990" subclass of medications: "Xanthine-Sympathomimetic-Antihistamine-&/or Expectorant"
 The "449993" subclass of medications: "Xanthine-Sympathomimetic-Antitussive-Expectorant"
 The "449995" subclass of medications: "Xanthine-Sympath-Antihistamine-Antitussive-Expectorant"

14. MEDICATIONS – OTHER

14.1 ASTHMA_MEDS (Asthma Med Use – self/scanned (curr asthma by MD diag))

This is a 0/1 numeric variable. This variable combines scanned/transcribed anti-asthmatic medication use with self-report asthma medication use. This variable is defined as 0 for those that do not currently have asthma based on an MD diagnosis.

If ASTHMA_CURR_MD=0 then ASTHMA_MEDS=0.
Else if ASTHMA_CURR_MD=1 and (RSEA37=1 or MED_ANTIASTHMATICS=1)
then ASTHMA_MEDS=1.
Else if ASTHMA_CURR_MD=1 and (RSEA37=0 or MED_ANTIASTHMATICS=0)
then ASTHMA_MEDS=0.
Otherwise, ASTHMA_MEDS=missing.

Response format: 0 = No current asthma or no anti-asthmatic medication use
1 = Anti-asthmatic medication use

Source Variable(s):

RSEA37. Ever take asthma medication in last 12 months
ASTHMA_CURR_MD. Currently have asthma with ever MD diagnosis
MED_ANTIASTHMATICS. Anti-asthmatic Medication Use

14.2 COPD_MEDS (COPD Med Use – self/scanned (ever COPD by MD diag))

This is a 0/1 numeric variable. This variable combines scanned/transcribed anti-asthmatic medication use with self-report COPD medication use. This variable is defined as 0 for those that have never had COPD/emphysema or chronic bronchitis based on an MD diagnosis.

If COPD_EVER=1 and (RSEA46=1 or MED_ANTIASTHMATICS=1) then
COPD_MEDS=1.
Else if COPD_EVER = non-missing then COPD_MEDS=0.
Otherwise, COPD_MEDS=missing.

Response format: 0 = Never had COPD/emphysema or no indication of COPD/anti-asthmatic medication use
1 = COPD/anti-asthmatic medication use

Source Variable(s):

RSEA46. Ever take COPD/Emph medication in last 12 months.
COPD_EVER. Ever had COPD/Emph or CB
MED_ANTIASTHMATICS. Anti-asthmatic Medication Use

15. NEUROCOGNITIVE

15.1 ECG_MAJ (Major ECG Abnormalities)

Defined only for participants that have ECG data.

First, derived corresponding Binary indicator derived variables: ECG_266 – ECG_280:

For the 15 source variables (from ECGA266 to ECGA280):

If the variable has missing value (no text), then set a Binary indicator variable as 0;

If the variable has descriptive text as the value, then set a Binary indicator variable as 1.

Second, derive the count variables for the major ECG abnormalities:

ECG_Maj_N (count of Major ECG Abnormalities) = the sum of Binary indicator variables ECG_266 – ECG_280

Then define the Major ECG abnormalities variables:

If ECG_Maj_N > 0 then set Major ECG abnormalities ECG_Maj as 1, otherwise if ECG_Maj_N = 0 then set ECG_Maj as 0;

Response format: 0 = No
1 = Yes

Source Variable(s):

ECGA266. one of the following

ECGA267. Major Q Wave abnormalities [old Myocardial Infarction (MI)]

ECGA268. Minor Q, QS waves with ST, T abnormalities (possible old MI)

ECGA269. Major Isolated ST_T abnormalities

ECGA270. Left Ventricular Hypertrophy

ECGA271. Atrial Fibrillation or Flutter

ECGA272. one of the following

ECGA273. Complete (third degree) A-V block

ECGA274. one of the following

ECGA275. Wolff-Parkinson-White pattern

ECGA276. Major QT Prolongation, QT_i ≥ 116%

ECGA277. Pacemaker

ECGA278. one of the following

ECGA279. one of the following

ECGA280. Supraventricular tachycardia intermittent

15.2 ECG_MIN (Minor ECG Abnormalities)

Defined only for participants that have ECG data.

First, derived corresponding Binary indicator derived variables: ECG_281 – ECG_300:

For the 20 source variables (from ECGA281 to ECGA300):

If the variable has missing value (no text), then set a Binary indicator variable as 0;

If the variable has descriptive text as the value, then set a Binary indicator variable as 1.

Second, derive the count variables for the minor ECG abnormalities:

ECG_Min_N (count of Minor ECG Abnormalities) = the sum of Binary indicator variables
ECG_281 – ECG_300

Then define the Minor ECG abnormalities variable:

If ECG_Min_N > 0 then set Minor ECG abnormalities ECG_Min as 1, otherwise if
ECG_Min_N = 0 then set ECG_Min as 0;

Response format: 0 = No
1 = Yes

Source Variable(s):

ECGA281. MINOR ISOLATED Q, QS waves
ECGA282. MINOR ISOLATED ST, T Abnormalities
ECGA283. HIGH R waves
ECGA284. ST Segment Elevation, Anterolateral Site
ECGA285. ST Segment Elevation, Posterior Site
ECGA286. ST Segment Elevation, Anterior Site
ECGA287. Incomplete RBBB
ECGA288. Incomplete LBBB
ECGA289. Minor QT Prolongation, QT_i ≥ 112%
ECGA290. Short PR
ECGA291. Left axis deviation
ECGA292. Right axis deviation
ECGA293. Frequent Ventricular Premature Beats
ECGA294. P-R Interval > .22 sec
ECGA295. one of the following
ECGA296. Supraventricular rhythm
ECGA297. Sinus Tachycardia
ECGA298. Sinus Bradycardia
ECGA299. Low QRS amplitude
ECGA300. P-wave amplitude high

15.3 NO_DISABILITY_6ITEM (No Disability on NEEA5-7 (1=Yes,0=No))

This is a 0/1 numeric variable..

1, if NEEA1 in (1,2,3);
0, if NEEA1 not in (1,2,3), excluding permanently missing forms and forms with all fields marked "unresolvable";
Missing otherwise.

5/2014 update: revised to set participants missing NEE forms to missing

Response format: 0 = Disability
1 = No disability

Source Variable(s):

NEEA1. Number of presentations necessary for the participant to repeat the words.

15.4 SEVLT_3TRIALS (SEVTL, Sum of 3 Trials (NEEA8-10))

This is a continuous variable for the sum of the scores of the first 3 trials of the SEVLT test. Only defined for participants who took a neurocognitive assessment (NEE form).

SEVLT_3trials = NEEA8+NEEA9+NEEA10
SEVLT_3trials is missing if any of NEEA8, NEEA9, NEEA10 is missing.

Source variable(s):

NEEA8. Words Recalled from Part B (Trial 1).

NEEA9. Words Recalled from Part B (Trial 2).

NEEA10. Words Recalled from Part B (Trial 3).

15.5 SEVLT_RECALL (SEVLT, Delay Recall (NEEA11 with re-coding))

This is a continuous variable for the scores of the 5th delayed recall trial of the SEVLT test. Definition per Dr. Hector Gonzalez. Only defined for participants who took a neurocognitive assessment (NEE form).

SEVLT_recall = 0, if NEEA11 is missing and SEVLT_3trials is not missing,
Otherwise SEVLT_recall = NEEA11.

Source variable(s):

NEEA11. Words Recalled from Part B (Trial 5)

SEVLT_3Trials. SEVTL, Sum of 3 Trials (NEEA8=10)

15.6 TOTAL_6ITEM (Total Score on 6 Item Screener (with re-coding))

This is a continuous variable for the total score of the 6-item screener, which uses recoding of the original responses per definitions by Dr. Hector Gonzalez. It can take integer values from 0 to 6. Only defined for participants who took a neurocognitive assessment (NEE form).

The computation of recoded variables is according to the table below:

Variable	Definition
No_Disability_6item	1, if NEEA1 in (1,2,3); 0, if NEEA1 not in (1,2,3), excluding permanently missing forms and forms with all fields marked "unresolvable"; Missing otherwise.
NEEA2_r1	=2, if NEEA2 in (2,3,4); = NEEA2 otherwise.
NEEA3_r1	=2, if NEEA3 in (2,3,4); = NEEA3 otherwise.
NEEA4_r1	=2, if NEEA4 in (2,3,4); = NEEA4 otherwise.
NEEA5_r1	=2, if NEEA5 in (2,3,4) and No_Disability_6item=1; = NEEA5 otherwise.
NEEA6_r1	=2, if NEEA6 in (2,3,4) and No_Disability_6item=1; = NEEA6 otherwise.
NEEA7_r1	=2, if NEEA7 in (2,3,4) and No_Disability_6item=1; = NEEA7 otherwise.
NEEA2_recode	= 0, if NEEA2_r1=2; =1, if NEEA2_r1=1; = missing, if NEEA2_r1 in (3,4) or missing.
NEEA3_recode	= 0, if NEEA3_r1=2; =1, if NEEA3_r1=1; = missing, if NEEA3_r1 in (3,4) or missing.
NEEA4_recode	= 0, if NEEA4_r1=2; =1, if NEEA4_r1=1; = missing, if NEEA4_r1 in (3,4) or missing.
NEEA5_recode	= 0, if NEEA5_r1=2; =1, if NEEA5_r1=1; = missing, if NEEA5_r1 in (3,4) or missing.
NEEA6_recode	= 0, if NEEA6_r1=2; =1, if NEEA6_r1=1; = missing, if NEEA6_r1 in (3,4) or missing.
NEEA7_recode	= 0, if NEEA7_r1=2; =1, if NEEA7_r1=1; = missing, if NEEA7_r1 in (3,4) or missing.

Total_6item = NEEA2_recode + NEEA3_recode + NEEA4_recode + NEEA5_recode +
NEEA6_recode + NEEA7_recode

Total_6item should be missing if any of the variables in the sum is missing.

Source variable(s):

NEEA2. "What year is this?"

NEEA3. "What month is this?"

NEEA4. "What is the day of the week?"

NEEA5. "Now, what were those three words I asked you to remember?" – Blue

NEEA6. "Now, what were those three words I asked you to remember?" – Pear

NEEA7. "Now, what were those three words I asked you to remember?" – Sofa

No_Disability_6Item. No Disability on NEEA5-7 (1=Yes,0=No)

15.7 WORDFREQ (Word Frequency (A+F) (NEEA12-13))

This is a continuous variable for the sum of the score of the A and F parts of the Word Fluency test. Only defined for participants who took a neurocognitive assessment (NEE form).

=NEEA12 + NEEA13

- should be missing if any of them are missing.

Source Variable(s):

NEEA12. Words Produced on Part C: Letter F

NEEA13. Words Produced on Part C: Letter A

15.8 DIGITSYMBOL (Digit Symbol Substitution Score (NEEA14))

This is a continuous variable for the score of the digit symbol substitution test. Only defined for participants who took a neurocognitive assessment (NEE form).

DigitSymbol = NEEA14

Source variable(s):

NEEA14. Total Correct Symbols on PART D: DIGIT SYMBOL SUBSTITUTION.

16. DENTAL

16.1 PERIODONTAL_DISEASE (Periodontal Disease - NIDCR)

This is a 0/1 numeric variable. The NIDCR cites the prevalence of periodontal disease based on dental examination probing measurements. A person must have at least one site with 3mm or more of attachment loss and 4mm or more of pocket depth in order to be positive for periodontal disease.

Periodontal_Disease = there is at least one tooth with Attachment Loss (AL) 3mm or greater AND Pocket Depth (PD) 4mm or greater.

Response format: 0 = No periodontal disease (No)
1 = Periodontal disease detected (Yes)

```
** pocket depth (PD) measures;  
array pd(6,32) perapmb1-perapmb32  
           perapbu1-perapbu32  
           perapdb1-perapdb32  
           perapdl1-perapdl32  
           perapln1-perapln32  
           perapml1-perapml32;  
array pdx(32) pdx1-pdx32;
```

```
** attachment loss (AL) measures = PD - CEJ (must be >= 0);  
array al(6,32) almb1-almb32  
           albu1-albu32  
           aldb1-aldb32  
           aldl1-aldl32  
           alln1-alln32  
           alml1-alml32;  
array alx(32) alx1-alx32;
```

```
array chx(32) chx1-chx32;
```

```
do k=1 to 32;  
  do l=1 to 6;  
    if pd(l,k) >=0 then do;  
      if pd(l,k) >=4 then pdx(k)=1;  
    end;  
  
    if al(l,k) >=-100 then do;  
      if al(l,k) >=3 then alx(k)=1;  
    end;  
  end;  
end;
```

```
if pdx(k)=1 & alx(k)=1 then chx(k)=1;
```

```
else chx(k)=0;  
end;
```

```
if sum(of chx1-chx32)>=1 then PERIODONTAL_DISEASE=1;  
else PERIODONTAL_DISEASE=0;
```

Source variable(s):

Pocket Depth and Attachment loss variable arrays from Dental Derive File (see above code).

16.2 PERIO_CDC_AAP (Periodontitis Severity using CDC/AAP Grades)

We want to extend the definition of periodontal disease to include a derived variable that uses the most recent Centers for Disease Control (CDC) and American Academy of Periodontology (AAP) severity grading guidelines. These definitions are different than the existing one for PERIODONTAL_DISEASE in the current PART_DERV file. The following specifications were provided by Jim Beck at the UNC Dental School. Clinical measures to assess periodontal health include probing pocket depth (PD) and gingival recession; where PD refers to the depth of the pocket measured from the gingival margin (GM) to the depth of the gingival sulcus, and gingival recession is the distance from the GM to the cemento-enamel junction (CEJ). Clinical attachment level (CAL) is calculated from the sum of PD and GM-CEJ measurements and forms the basis for the currently accepted periodontal disease classification system. For attachment loss, (AL) measures = PD - CEJ (must be ≥ 0).

The HCHS/SOL used a comprehensive full mouth examination protocol examining six sites per tooth in up to 28 teeth yielding up to 168 sites per person. For the purpose of this study, four outcome measures will be reported based on the Centers for Disease Control and Prevention and the American Academy of Periodontology (CDC/AAP) case definition of periodontitis at four severity thresholds: (1) mild; (2) moderate; (3) severe; (4) total. These classifications (described below) are designed for population-based surveillance of periodontal disease and are based on measures of clinical periodontal clinical attachment level (CAL) and probing pocket depth (PD) thus defined:

Mild: [2 or more interproximal sites with 3 mm or more CAL AND 2 or more sites with 4 mm or more PD] (not on same tooth) OR [At least 1 site with 5mm or more PD]

Moderate: [2 or more interproximal sites with CAL 4 mm or more (not on same tooth)]
OR [2 or more interproximal sites with PD 5 mm or more (not on same tooth)]

Severe: [2 or more interproximal sites with CAL 6 mm or more (not on same tooth)] AND [At least 1 tooth with an interproximal site with PD of 5 mm or more]

Response format: . = Missing if all values for pocket depth and attachment loss are missing
0 = Normal
1 = Mild periodontitis
2 = Moderate periodontitis
3 = Severe periodontitis

16.3 PERIO_EXTENT (Percentage Interproximal Sites Positive for Periodontitis)

This is an additional method of describing the extent of disease in the mouth is the Extent of disease, which is the percent of sites within an individual that have a CAL score ≥ 3 mm or a PD score ≥ 4 , 5, or 6 mm. Each tooth has 4 potential measurement sites for evaluation so the individual denominator for extent of disease would be (4 x number of teeth present). The measures taken on the buccal and the lingual sites are not used in this scoring method, so the array variables below have those variable ranges deleted for the buccal and lingual measures on all 32 tooth positions.

PERIO_EXTENT = (Sum positive sites per person / Number of possible sites) * 100

Note the number of possible sites = 4 * (total tooth count)

Attachment loss and the CEJ and pocket depth measurements as follows:

Pocket depths:

perapmbu1 to perapmb32
~~perapbu1-perapbu32~~
perapdb1-perapdb32
perapdl1-perapdl32
~~perapln1-perapln32~~
perapml1-perapml32;

Cemento-enamel junction (CEJ) measurements:

peracmb1 to peracmb32
~~peracbu1-peracbu32~~
peracdb1-peracdb32
peracdl1-peracdl32
~~peracln1-peracln32~~
peracml1-peracml32;

Attachment loss:

almb1 to almb32
~~albu1-albu32~~
aldb1-aldb32
aldl1-aldl32
~~alln1-alln32~~
alml1-alml32

Note that the “bu1-bu32”, and the “ln1-ln32” rows in the arrays above are not used in the CDC definition of periodontal disease.

The strike-through character above means ignore these sites when evaluation is done for the CDC severity scoring method.

17. SLEEP

17.1 ESS (Epworth Sleepiness Scale)

This is an integer value with range from 0-24. The Epworth Sleepiness Scale assesses the daytime sleepiness of patients. Dr Murray W Johns owns the copyright in the ESS (Copyright © M.W.Johns 1990-1997). The official website is <http://epworthsleepinessscale.com/>

Algorithm: The ESS is computed by adding ALL 8 ITEMS (SLEA12A TO SLEA12H). If any item is missing then ESS is missing.

Warning: The levels of each item need to be recoded as follows: No chance (from 1 to 0), slight chance (from 2 to 1), moderate chance (from 3 to 2) and high chance (from 4 to 3).

Source variable(s):

What is the chance that you would doze off or fall asleep (not just "feel tired") in each of the following situations?

SLEA12a. Sitting and reading

SLEA12b. Watching TV

SLEA12c. Sitting inactive in a public place (such as a theater or a meeting)

SLEA12d. Riding as a passenger in a car for an hour without a break

SLEA12e. Lying down to rest in the afternoon when circumstances permit

SLEA12f. Sitting and talking to someone

SLEA12g. Sitting quietly after a lunch without alcohol

SLEA12h. In a car, while stopped for a few minutes in traffic

17.2 ESS_GT10 (Excessive sleepiness (ESS > 10))

This is a 0/1 numeric variable. If ESS is non-missing, then:

If $ESS > 10$ then $ESS_GT10 = 1$.

Otherwise if $0 \leq ESS \leq 10$ then $ESS_GT10 = 0$.

REFERENCE: Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep. 1991 Dec;14(6):540-5. PMID: [1798888](#)

Source variable(s):

ESS. Epworth Sleepiness Scale

17.3 AHI_GE15 (Sleep Apnea defined as Apnea/Hypopnea Index (3% desat) ≥ 15)

This is a 0/1 numeric variable. If SLPA54 is non-missing, then:

If $SLPA54 \geq 15$ then $AHI_GE15 = 1$.

Otherwise if $0 \leq SLPA54 < 15$ then $AHI_GE15 = 0$

Source variable(s):

SLPA54. Apnea/Hypopnea Index (3% desat)

17.4 SBD5 (Sleep Apnea Syndrome (Apnea/Hypopnea Index (3% desat) >= 5 & ESS >=10))

This is a 0/1 numeric variable. If SLPA54 and ESS are non-missing, then:
If SLPA54>=5 & ESS>=10 then SBD5=1
Otherwise SBD5=0.

Source variable(s):

SLPA54. Apnea/Hypopnea Index (3% desat)
ESS. Epworth Sleepiness Scale

17.5 SLPDUR_WKDAY (Weekday sleep duration (hours))

This is a numeric variable that calculates the self-reported sleep duration during weekdays. Sleep data was cleaned for creation of sleep duration variables. Sleep durations less than 3 hours or greater than 14 hours were set to missing.

```
IF SLEA2A_2401 <= .z OR SLEA1A_2401 <= .z then SLPDUR_WKDAY = .;  
ELSE IF SLEA2A_2401 > .z AND SLEA1A_2401 > .z then do;  
    IF SLEA2A_2401 >= SLEA1A_2401  
        then SLPDUR_WKDAY = (SLEA2A_2401 - SLEA1A_2401)/(60*60);  
    ELSE IF SLEA2A_2401 < SLEA1A_2401  
        then SLPDUR_WKDAY = (SLEA2A_2401+86400 -  
        SLEA1A_2401)/(60*60);  
end;
```

Source variable(s):

SLEA1A_2401. What time do you usually go to bed? (on weekdays)
SLEA2A_2401. What time do you usually wake up? (on weekdays)

17.6 SLPDUR_WKEND (Weekend sleep duration (hours))

This is a numeric variable that calculates the self-reported sleep duration during weekend days. Sleep data was cleaned for creation of sleep duration variables. Sleep durations less than 3 hours or greater than 14 hours were set to missing.

```
IF SLEA2C_2401 <= .z OR SLEA1C_2401 <= .z then SLPDUR_WKEND = .;  
ELSE IF SLEA2C_2401 > .z AND SLEA1C_2401 > .z then do;  
    IF SLEA2C_2401 >= SLEA1C_2401  
        then SLPDUR_WKEND = (SLEA2C_2401 - SLEA1C_2401)/(60*60);  
    ELSE IF SLEA2C_2401 < SLEA1C_2401  
        then SLPDUR_WKEND = (SLEA2C_2401+86400 -  
        SLEA1C_2401)/(60*60);  
end;
```

Source variable(s):

SLEA1C_2401. What time do you usually go to bed? (on weekends)
SLEA2C_2401. What time do you usually wake up? (on weekends)

17.7 SLPDUR (Average sleep duration (hours))

This is a numeric variable that calculates the weighted average daily sleep duration during an entire week (five weekdays and two weekend days), using self-reported sleep data. Sleep data was cleaned for creation of sleep duration variables. Sleep durations less than 3 hours or greater than 14 hours were set to missing.

$$\text{SLPDUR} = (\text{SLPDUR_WKEND} * 2 + \text{SLPDUR_WKDAY} * 5) / 7;$$

Source variable(s):

SLPDUR_WKDAY. Weekday sleep duration (hours)

SLPDUR_WKEND. Weekend sleep duration (hours)

17.8 MID_SLEEP_TIME_WKDAY (Midpoint sleep time in clock hours; weekdays)

This is a numeric variable that calculates the midpoint of self-reported sleep time during weekdays. It is in clock hours in decimal point (e.g. 3.5 is 3:30am and 15.75 is 5:45pm). Sleep data was cleaned for creation of midpoint sleep time variables.

```
IF SLEA2A_2401 <= .z OR SLEA1A_2401 <= .z then MID_SLEEP_TIME_WKDAY = .;
  ELSE IF SLEA2A_2401 > .z AND SLEA1A_2401 > .z then do;
IF SLEA2A_2401 >= SLEA1A_2401
  then MID_SLEEP_TIME_WKDAY = ((SLEA2A_2401 + SLEA1A_2401)/2) / (60*60);
ELSE IF SLEA2A_2401 < SLEA1A_2401
  then MID_SLEEP_TIME_WKDAY = ((SLEA2A_2401+86400 + SLEA1A_2401)/2)
/(60*60);
IF MID_SLEEP_TIME_WKDAY >= 24
  then MID_SLEEP_TIME_WKDAY = MID_SLEEP_TIME_WKDAY - 24;
end;
```

Source variable(s):

SLEA1A_2401. What time do you usually go to bed? (on weekdays)

SLEA2A_2401. What time do you usually wake up? (on weekdays)

17.9 MID_SLEEP_TIME_WKEND (Midpoint sleep time in clock hours; weekend)

This is a numeric variable that calculates the midpoint of self-reported sleep time during weekend. It is in clock hours in decimal point (e.g. 3.5 is 3:30am and 15.75 is 5:45pm). Sleep data was cleaned for creation of midpoint sleep time variables.

```
IF SLEA2C_2401 <= .z OR SLEA1C_2401 <= .z then MID_SLEEP_TIME_WKEND = .;
  ELSE IF SLEA2C_2401 > .z AND SLEA1C_2401 > .z then do;
IF SLEA2C_2401 >= SLEA1C_2401
  then MID_SLEEP_TIME_WKEND = ((SLEA2C_2401 + SLEA1C_2401)/2) / (60*60);
ELSE IF SLEA2C_2401 < SLEA1C_2401
  then MID_SLEEP_TIME_WKEND = ((SLEA2C_2401+86400 + SLEA1C_2401)/2)
/(60*60);
IF MID_SLEEP_TIME_WKEND >= 24
  then MID_SLEEP_TIME_WKEND = MID_SLEEP_TIME_WKEND - 24;
end;
```

Source variable(s):

SLEA1C_2401. What time do you usually go to bed? (on weekends)

SLEA2C_2401. What time do you usually wake up? (on weekends)

17.10 WHIIRS (Women's Health Initiative Insomnia Rating Scale)

The score is calculated from non-missing values for any of the following five items: SLEA4, SLEA5, SLEA6, SLEA7 and SLEA11. The original items range from 0 to 4 and the total score from 0 to 20. In HCHS/SOL, SLEA form the scoring for questions 4 to 7 each ranged from 1 to 5 so 1 should be subtracted from each of these items to obtain the score for use in calculating WHIIRS. For example, No, not in the past 4 weeks should give you a score of 0, etc. The scoring for SLEA11 ranges from 0 to 4 and does not need to be modified. Another way to think about it is:

$$\text{WHIIRS} = \text{SLEA4} + \text{SLEA5} + \text{SLEA6} + \text{SLEA7} + \text{SLEA11} - 4$$

Reference:

Levine DW, Kripke DF, Kaplan RM, Lewis MA, Naughton MJ, Bowen DJ, Shumaker SA. Reliability and validity of the Women's Health Initiative Insomnia Rating Scale. Psychol Assess. 2003 Jun;15(2):137-48.

Source variable(s):

SLEA4. Did you have trouble falling asleep?

SLEA5. Did you wake up several times at night?

SLEA6. Did you wake up earlier than you planned to?

SLEA7. Did you have trouble getting back to sleep after you woke up too early?

SLEA11. Overall, was your typical night's sleep during the past 4 weeks:

18. DIET

18.1 DIET_SCORE_JAMA (Diet Score from 1st JAMA paper)

Daviglus ML, Talavera GA, Avilés-Santa ML, et al. Prevalence of major cardiovascular risk factors and cardiovascular diseases among Hispanic/Latino individuals of diverse backgrounds in the United States. JAMA. 2012 Nov 7 ;308(17):1775-84.

Name:	Label:	Definition:
Subscore_SatFat	Subscore for Saturated Fatty Acids	Missing, if PRED_TOTSFA or SEX is missing; Else: 5, if .z<PRED_TOTSFA<=20%-tile (SEX-specific) for Saturated Fatty Acids intake 4, if 20%-tile< PRED_TOTSFA<=40%-tile; 3, if 40%-tile< PRED_TOTSFA<=60%-tile; 2, if 60%-tile< PRED_TOTSFA<=80%-tile; 1, if 80%-tile< PRED_TOTSFA.
Subscore_Fib	Subscore for Dietary Fiber	Missing, if PRED_FIBER or SEX is missing; Else: 1, if .z<PRED_FIBER<=20%-tile (SEX-specific) for Dietary Fiber intake 2, if 20%-tile< PRED_FIBER<=40%-tile; 3, if 40%-tile< PRED_FIBER<=60%-tile; 4, if 60%-tile< PRED_FIBER<=80%-tile; 5, if 80%-tile< PRED_FIBER. (the order is reverse compared to Subscore_SatFat)
Subscore_Ca	Subscore for Dietary Calcium	Missing, if pred_CA or SEX is missing; Else: 1, if .z<pred_CA<=20%-tile (SEX-specific) for Calcium intake 2, if 20%-tile< pred_CA<=40%-tile; 3, if 40%-tile< pred_CA<=60%-tile; 4, if 60%-tile< pred_CA<=80%-tile; 5, if 80%-tile< pred_CA. (the order is same as in Subscore_Fib)
Subscore_K	Subscore for Dietary Potassium	Missing, if pred_K or SEX is missing; Else: 1, if .z<pred_K<=20%-tile (SEX-specific) for Potassium intake 2, if 20%-tile< pred_K<=40%-tile; 3, if 40%-tile< pred_K<=60%-tile; 4, if 60%-tile< pred_K<=80%-tile; 5, if 80%-tile< pred_K. (the order is same as in Subscore_Fib)
Diet_Score_JAMA	Diet Score from 1st JAMA paper	Missing, if any of Subscore_SatFat, Subscore_Fib, Subscore_Ca, Subscore_K is missing; Else Diet_Score_JAMA=Score_SatFat + Score_Fib + Score_Ca + Score_K.

Source variables (from PRED_NUTR_DERV):

PRED_TOTSFA. Total Saturated Fatty Acids (SFA) (g) - NCI predicted intake (based on DTIA28)

PRED_FIBER. Total dietary fiber (g) - NCI predicted intake (based on DTIA38)

PRED_CA. Calcium (mg) - NCI predicted intake (based on DTIA60)

PRED_K. Potassium (mg) - NCI predicted intake (based on DTIA68)

18.2 DIET_SCORE_JAMA_C2 (Categorized Diet Score JAMA (1=Below, 2=Above 60th %-tile))

Diet_score_JAMA_c2 =1, if .z<Diet_score_JAMA<60th sex-specific percentile;
Diet_score_JAMA_c2 =2, if Diet_score_JAMA>=60th sex-specific percentile;
Missing, if Diet_Score_JAMA is missing.

Source variable(s):

Diet_Score_JAMA. Diet Score from 1st JAMA paper

18.3 AHEI2010 (Alternative Healthy Eating Index 2010)

The alternative healthy eating index (AHEI-2010) is a measure of diet quality based on foods and nutrients predictive of chronic disease risk (Chiuve SE et al, 2012). It is derived using the following 11 components. The overall score as well as its 11 components (AHEI1 to AHEI11) are available.

- 1) Vegetables without potatoes, servings/day
- 2) Whole Fruit (i.e. does not include fruit juice), servings/day
- 3) Whole grains, servings/day
- 4) Sugar sweetened beverages and fruit juice, servings/day
- 5) Nuts and legumes, servings/day
- 6) Red/processed meat, servings/day
- 7) Trans Fat, % energy
- 8) Long-chain (n-3) fats (EPA+DHA), mg/day
- 9) Polyunsaturated fatty acids (PUFA), % energy;
- 10) Sodium, mg/day;
- 11) Alcohol, drinks/day

AHEI-2010 score is the sum of these 11 individual components' scores, each with a range from 0 (worst) and 10 (best). Hence, **AHEI-2010 takes values from 0 to 110**. Higher scores represent healthy eating habits and lower scores represent unhealthy eating habits. Only scores for components "whole grains", sodium and "alcohol" are sex-specific.

In HCHS/SOL, AHEI-2010 was calculated from available 24hr dietary recall data (one or two recalls per participant) using the NCI method to predict usual intake for each component. A participant has an AHEI-2010 score if he/she has at least one 24-hr dietary recall after exclusions. Recalls with daily energy intake (DTIA20) below the sequence-sex specific 1st percentile or above the 99th percentile or unreliable according to the interviewer (DTIA16) were excluded.

For more details on its derivation please see document "HCHS/SOL Dietary Overview, Methods and Guidelines".

Reference: S E Chiuve et al. (2012) "Alternative Dietary Indices Both Strongly Predict Risk of Chronic Disease"; J. Nutr. 142: 1009 – 1018.

18.4 HEI2010 (Healthy Eating Index 2010)

The Healthy Eating Index-2010 (HEI-2010; Guenther et al., J Acad Nutr Diet 2013) is a measure of overall diet quality, independent of quantity, which can be used to assess compliance with the 2010 Dietary Guidelines for Americans and to monitor changes in dietary patterns. It includes twelve dietary components, nine adequacy and three moderation components. The overall score as well as its 12 components (HEI1 to HEI12) are available.

Adequacy

1. Total Fruit
2. Whole Fruit
3. Total Vegetables
4. Greens and Beans
5. Whole Grains
6. Dairy
7. Total Protein Foods
8. Seafood and Plant Proteins
9. Fatty Acids

Moderation

10. Refined Grains
11. Sodium
12. Empty Calories

HEI-2010 score is the sum of these 12 components' scores, that range from 0-5, 0-10, or 0-20. HEI score ranges from 0 to 100 with a higher score indicating greater consistency of the diet with the 2010 Dietary Guidelines for Americans. The performance of the HEI-2010 has been evaluated through an assessment of its content validity, construct validity, and reliability (Guenther et al., J Nutr 2014).

In HCHS/SOL, the HEI-2010 was calculated as the average of available 24hr dietary recalls. A participant has a HEI-2010 score if he/she has at least one 24-hr dietary recall after exclusions. Recalls with daily energy intake (DTIA20) below the sequence-sex specific 1st percentile or above the 99th percentile or unreliable according to the interviewer (DTIA16) were excluded. We used the NDSR Guide to create variables needed to calculate scores for each component of the HEI-2010 developed by the Nutrition Coordinating Center (NCC) at University of Minnesota, Minneapolis, MN.

For more details on its derivation please see document "HCHS/SOL Dietary Overview, Methods and Guidelines".

References:

- Guenther PM, Casavale KO, Reedy J, Kirkpatrick SI, Hiza HA, Kuczynski KJ, Kahle LL, Krebs-Smith SM. Update of the Healthy Eating Index: HEI-2010. J Acad Nutr Diet. 2013 Apr;113(4):569-80.
- Guenther PM, Kirkpatrick SI, Reedy J, Krebs-Smith SM, Buckman DW, Dodd KW, Casavale KO, Carroll RJ. The Healthy Eating Index-2010 is a valid and reliable measure of diet quality according to the 2010 Dietary Guidelines for Americans. J Nutr. 2014 Mar;144(3):399-407.

- NDSR Guide to Creating Variables Needed to Calculate Scores for Each Component of the Healthy Eating Index-2010 (HEI-2010) developed by the Nutrition Coordinating Center (NCC), University of Minnesota, Minneapolis, MN.

18.5 DASH (Version 1 – 2003 Eating Plan)

The dietary approaches to stop hypertension (DASH) diet was initially designed to control blood pressure, and it consists of a diet higher in potassium, calcium, magnesium, fiber, and protein, and lower in total fat, saturated fat, and cholesterol (Karanja et al., 1999). **In HCHS/SOL, we created two versions of the DASH score, both based in Gunther's DASH index and using the average of the two 24hr dietary recalls.**

Version 1. Based on 2003 DASH Eating Plan to match Günther *et al.*, 2009.

Version 2. Based on 2018 DASH Eating Plan which recommends whole grains and fat-free or low-fat dairy products and limits sodium intake.

Günther's DASH index was defined based on the 2003 DASH Eating Plan from the National Institutes of Health, National Heart, Lung, and Blood Institute (NHLBI) (<https://www.ahajournals.org/doi/full/10.1161/HYPERTENSIONAHA.108.116665>). This version has 8 components, each ranging from 0 (worst) to 10 (best). The grains component is the sum of the scores for the total grains and whole grains sub-components, and the dairy component is the sum of the scores for total dairy and low-fat dairy sub-components. Each of the four sub-components ranges from 0 (worst) to 5 (best). **The version 1 of the total DASH score is the sum of all 8 components, and it ranges from 0 to 80.** The overall score as well as all components (DASH1A_2003 to DASH8_2003) are available.

1. Grains:
 - total grains
 - whole grains
2. Vegetables
3. Fruits
4. Dairy:
 - all dairy
 - low-fat dairy
5. Meat/Poultry/Fish/Eggs
6. Nuts/Seeds/Legumes
7. Fats/Oils
8. Sweets

For the scoring of both versions of the DASH, we used as cut-off values the serving sizes for 7 different levels of energy intake (1200/1400/1600/1800/2000/2600/3100) recommended by the NHLBI 2018 DASH Eating Plan. Each participant was assigned an estimated energy requirement (EER) based on sex, age, and physical level activity following the estimated daily calorie needs to maintain calorie balance provided by the

Dietary Guidelines for Americans 2015-2020 (Table A2-1 from https://odphp.health.gov/sites/default/files/2019-09/2015-2020_Dietary_Guidelines.pdf).

For more details on its derivation please see document “HCHS/SOL Dietary Overview, Methods and Guidelines”.

References:

- Günther AL, Liese AD, Bell RA, Dabelea D, Lawrence JM, Rodriguez BL, Standiford DA, Mayer-Davis EJ. Association between the dietary approaches to hypertension diet and hypertension in youth with diabetes mellitus. *Hypertension*. 2009 Jan;53(1):6-12.

- Karanja NM, Obarzanek E, Lin PH, McCullough ML, Phillips KM, Swain JF, Champagne CM, Hoben KP. Descriptive characteristics of the dietary patterns used in the Dietary Approaches to Stop Hypertension Trial. DASH Collaborative Research Group. *J Am Diet Assoc*. 1999 Aug;99(8 Suppl):S19-27.

- *DASH Eating Plan*, NHLBI website. Retrieved from <https://www.nhlbi.nih.gov/education/dash-eating-plan>

18.6 DASH (Version 2 – 2018 Eating Plan)

Because an updated DASH Eating Plan became available in 2018 (<https://www.nhlbi.nih.gov/education/dash-eating-plan>), we created a second version of the DASH score, updating Günther’s DASH index using the most recent recommendations. The 2018 DASH Eating Plan recommends whole grains and fat-free or low-fat dairy products and limits sodium intake. Thus, compared to DASH version 1 we excluded subcomponents total grains and total dairy, and we included sodium as a component. This version of the DASH has 9 components, each ranging from 0 (worst) to 10 (best); **the total DASH score is the sum of the 9 components, and it ranges from 0 to 90.** The overall score as well as all components (DASH1_2018 to DASH9_2018) are available.

1. Whole grains
2. Vegetables
3. Fruits
4. Low-fat dairy
5. Meat/Poultry/Fish/Eggs
6. Nuts/Seeds/Legumes
7. Fats/Oils
8. Sweets
9. Sodium

For more details on its derivation please see document “HCHS/SOL Dietary Overview, Methods and Guidelines”.

18.7 SUPPL_USE (Supplement use in the last 30 days)

This is a 0/1 numeric variable.

If SMIA34>0 then SUPPL_USE = 1 (yes)

If SMIA34=0 then SUPPL_USE = 0 (no)

Source variable(s):

SMIA34. (NDSR File Supplement recall for the last 30 days)

19. OTHER VARIABLES

19.1 SDI_SCORE. Zip code level neighborhood social deprivation

The social deprivation index (SDI) is from the American Community Survey SDI score using the baseline address zip code level. It is a composite measure of seven demographic characteristics collected in the American Community Survey: percent living in poverty, percent with less than 12 years of education, percent single-parent households, the percentage living in rented housing units, the percentage living in the overcrowded housing unit, percent of households without a car, and percentage unemployed adults under 65 years of age. A CSV file with SDI data from 2008 to 2012 was downloaded from the Robert Graham Center (<https://www.graham-center.org/maps-data-tools/social-deprivation-index.html>) on March 1st, 2024.

Variable Modifications

Data release	Variable	# participants changed	Notes
12/2013	DOCTOR_VISIT	5335	Corrected to calculate values for version B for the health care use form (HCE)
12/2013	EDUCATION_C2	295	Categorized missing values based on PIEA21B-PIEA21G, PIEA19B, PIEA21I, and PIEA17
12/2013	EDUCATION_C3	295	Categorized missing values based on PIEA21B-PIEA21G, PIEA19B, PIEA21I, and PIEA17
12/2013	GPAQ_REC_MOD	104	Set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.
12/2013	GPAQ_REC_VIG	106	Set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.
12/2013	GPAQ_REC	106	Set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.
12/2013	GPAQ_WORK_MOD	105	Set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.
12/2013	GPAQ_WORK_VIG	106	Set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.
12/2013	GPAQ_WORK	106	Set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.
12/2013	GPAQ_TOTAL_MOD	106	Set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.
12/2013	GPAQ_TOTAL_VIG	106	Set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.
12/2013	GPAQ_TRSPORT	106	Set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.
12/2013	GPAQ_TOTAL	106	Set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.
12/2013	GPAQ_TOTAL_MET	106	Set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.
12/2013	GPAQ_LEVEL	2516	Corrected a typo in code such that participants who were actually "moderate" were being set to "low". Set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.
12/2013	GPAQ_PAG2008	245	Set participants to missing if they were missing GPAQ_MOD_WEEK and GPAQ_VIG_WEEK. Set participants to missing if they reported more than 1440 total minutes per day of active + sedentary.
12/2013	GPAQ_PAG2008YN	245	Changes reflected from GPAQ_PAG2008
12/2013	INSULIN_FAST	32	Changed missing to .L for .L values of LABA73
5/2014	AGG_PHYS	1	Set to missing due to invalid data for 1 item
5/2014	AGG_MENT	1	Set to missing due to invalid data for 1 item
5/2014	LEFT_ABI	173	ABI definition updated to include occlusion failures
5/2014	RIGHT_ABI	173	ABI definition updated to include occlusion failures
5/2014	MRC_CB	18	Updated to count RSEA3=99 as missing
5/2014	FEV1_FVC_RATIO	122	Set to missing for invalid spirometries
5/2014	AIRFLOW_OBSTRUCTION	122	Set to missing for invalid spirometries
5/2014	POSTBD_PREBD_DIFF	2	Set to missing for invalid spirometries
5/2014	COPD_BY_BD_ALLPRE	335	Set to missing for invalid spirometries
5/2014	CESD10	55	For participants with 1 or 2 missing items, the weighted average of the nonmissing items*10 is used instead of the sum of the nonmissing items
5/2014	STAI10	137	For participants with 1 or 2 missing items, the weighted average of the nonmissing items*10 is used instead of the sum of the nonmissing items
5/2014	MED_ESTROGENS	394	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_ESTROGEN_COMBO	394	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_CONTRACEPTIVES	394	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_PROGESTINS	394	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_SERM	394	Revised criteria for missing values of medication variables and corrected to use MUEA39

Data release	Variable	# participants changed	Notes
5/2014	MED_FERTILITY_REG	394	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_LHRH	394	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_LHRH_AG	394	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_LHRH_ANT	394	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_CURRENT_HRT	394	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_LLD	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_STATIN	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_ASPIRIN	57	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_ANTICOAG_COUMARIN	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_ANTICOAG_HEPARIN	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_ANTIPLATELET	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_CLOPID	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_ANTIDIAB	52	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_ANTICOAG	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_COX2	55	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_OI_STEROID	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_CHEMO	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_NSAID	196	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_METFORMIN	52	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_INSULIN	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_ANTIASTHMATICS	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_ANTIHYPERT	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_CACB	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_ANTIARRHYTHMICS	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_ANTIHYPERT_ACEI	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_CARDIACGLYCOSIDES	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_ANTIANGINAL	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_BB	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_BB_OPTHALMIC	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_ANTIHYPERT_AT2RAS	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_DIURETIC	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_DIURETIC_THIAZIDE	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	MED_FIBARES_NICOACID	51	Revised criteria for missing values of medication variables and corrected to use MUEA39
5/2014	METS_NCEP	2	Revised to use SEX instead of PIEA1

Data release	Variable	# participants changed	Notes
5/2014	METS_IDF	447	Corrected code to require abdominal obesity, revised to use SEX instead of P1EA1
5/2014	SASH_LANG	11277	Added item SCEA6, set to missing if more than 1 item is missing
5/2014	SASH_SOC	636	Set to missing if any item is missing
5/2014	ETHIS	93	Set to missing if either item is missing
5/2014	INSULIN_OGTT	2	Corrected to account for .N values of insulin
5/2014	HYPERTENSION	1	Updates due to updated medication variables
5/2014	DIABETES1	2	Updates due to updated medication variables
5/2014	DIABETES_SELF	18	Revised to code missing MHEA16 as no and MHEA16=.Q as missing instead of zero
5/2014	DIABETES3	4	Updates due to updated DIABETES_SELF
5/2014	DIABETES_C4	5	Updates due to updated DIABETES3 and medication variables
5/2014	DIAB_DIAG	9	Updates due to updated DIABETES_SELF and DIABETES3
5/2014	DM_AWARE	14608	Set to missing for nondiabetics
5/2014	DM_CONTROL	13027	Set to missing for nondiabetics
5/2014	IMGEN_C2	94	Recoded 9 to missing
5/2014	EXPOSURE_YEAR	11210	Revised to treat TBEA2=99 or 1 as missing. Corrected code to set never smokers to zero, former and current smokers with negative values to missing and to subtract TBEA2 for former smokers
5/2014	CIGARETTES_YEAR	1323	Updates due to changes in Exposure_Year, set current smokers with TBEA9=0 to missing
5/2014	CIGARETTE_PACK_YEARS	1271	Updates due to changes in Exposure_Year
5/2014	NO_DISABILITY_6ITEM	124	Revised to set participants missing NEE forms to missing
5/2014	HIGH_TOTAL_CHOL	2	Updates due to updated medication variables
5/2014	LEFT_ABI2		Removed, replaced by LEFT_ABI
5/2014	RIGHT_ABI2		Removed, replaced by RIGHT_ABI
5/2014	ABI2_OVERALL		Removed, replaced by ABI_OVERALL
5/2014	ABIGRP2_C4		Removed, replaced by ABIGRP_C4
5/2014	ELEVATED_BP_PLUSMEDS		Removed – intermediate variable
5/2014	HIGH_TRIG		Removed – intermediate variable
5/2014	LOW_HDL		Removed – intermediate variable
5/2014	IFG_NCEP		Removed – intermediate variable
5/2014	ABDOMINAL_OBESITY_NCEP		Removed – intermediate variable
5/2014	HIGH_TRIG_PLUSMEDS		Removed – intermediate variable
5/2014	LOW_HDL_PLUSMEDS		Removed – intermediate variable
5/2014	ABDOMINAL_OBESITY_IDF		Removed – intermediate variable
5/2014	IFG_IDF		Removed – intermediate variable
5/2014	ELEVATED_BP_SELFMEDS		Removed – intermediate variable
5/2014	IFG_NCEP_SELFMEDS		Removed – intermediate variable
5/2014	AIRFLOW_LIMITATION_LL_N		Removed – intermediate variable
5/2014	ECG_MAJ_N		Removed – intermediate variable
5/2014	ECG_MIN_N		Removed – intermediate variable
5/2014	IFG		Removed, no scientific value
5/2014	IGT		Removed, no scientific value
5/2014	SASH_ALL		Removed based on SCAS recommendations
12/2019	DM_AWARE	1609	Algorithm correction. Positive cases increase n=479, missing cases decrease n=1212.
12/2019	ESS_GT10	626	Correction on >10 instead >=10 from ESS_GE10, N=626 switched from 1 to 0.
12/2019	HYPERTENSION2_AHA	NA	New baseline derived variable, specs from V2
12/2019	HYPERTENSION_AHA_C5	NA	New baseline derived variable, specs from V2
12/2019	ABIGRP2_C4	NA	Rename from ABIGRP_C4_V2 (V2 was version 2 which is now used to denote visit 2)
12/2019	LEFT_ABI2	NA	Rename from Left_ABI_V2

Data release	Variable	# participants changed	Notes
12/2019	RIGHT_ABI2	NA	Rename from Right ABI V2
12/2019	ABI2_OVERALL	NA	Rename from ABI Overall V2
12/2019	POSTBD_PREBD_DIFF_GT12_2	NA	Rename from POSTBD_PREBD_DIFF_GT12_V2
12/2019	VALID_SPIROMETRY_POSTBD2	NA	Rename from VALID_SPIROMETRY_POSTBD_V2
12/2019	PRECVD	NA	New baseline derived variable, specs from V2
12/2019	FASTTIME, ELEVATED_BP_PLUSMEDS, ABDOMINAL_OBESITY_IDF, HIGH_TRIG_PLUSMEDS, LOW_HDL_PLUSMEDS, IFG_IDF, IDF_NCEP GFR, IFG_NCEP_SELFMEDS	NA	Keep intermediate variables for Diabetes, Metabolic Syndrome and GFR
12/2019	SUPPL_USE	NA	Added new derived variable
12/2019	EDUCATION_C4	NA	Added new derived variable
12/2019	MID_SLEEP_TIME_WKDAY MID_SLEEP_TIME_WKEND	NA	Added new derived variable
12/2019	HEI1 to HEI12, HEI2010	NA	Added Healthy Eating Index variables
12/2019	DASH2003 and components: DASH1A_2003, DASH1B_2003, DASH2_2003, DASH3_2003, DASH4A_2003, DASH4B_2003, DASH5_2003, DASH6_2003, DASH7_2003, DASH8_2003	NA	Added DASH variable (VERSION 1)
12/2019	DASH2018 and components: DASH1_2018, DASH2_2018, DASH3_2018, DASH4_2018, DASH5_2018, DASH6_2018, DASH7_2018, DASH8_2018, DASH9_2018	NA	Added DASH variable (VERSION 2)
12/2019	SELFMED_ANTIDIAB DIABETES4 DIABETES5	NA	Added per Diabetes SIG decision
12/2019	AGEGROUP_C3 YRSUS_C3	NA	
12/2019	PERIODONTAL_DISEASE	3008	Corrected algorithm
12/2019	PERIO_EXTENT	4	Corrected algorithm
12/2019	METS_NCEP2	66	Revised to use SEX instead of PIEA1
12/2019	MED_STATIN	79	Two more codes added: 399940 and 409925
12/2019	MED_CACB	34	One more code added: 369930
12/2019	MED_BB	5	One more code added: 369988
4/2026	ALBUMINURIA_C3 ADDRESS_YRS AGE_IMMI_C3 BKGRD1_C7R CMR DIABETES_LAB_NOOGTT DIABETES3A DIABETES4A DIABETES5A GFRSCR_Y21, GFRSCYS_SCR_Y21, GFRSCR_Y21_C5, GFRSCYS_SCR_Y21_C5, GFRSCR_Y21_C2, GFRSCYS_SCR_Y21_C2 HC HH_ID LE8_DIET, LE8_PA, LE8_NICOTINE, LE8_SLEEP, LE8_BMI, LE8_LIPIDS_SELF, LE8_LIPIDS_SCAN, LE8_GLUCOSE, LE8_BP_SELF, LE8_BP_SCAN, LE8_SCAN, LE8_SELF, LS7_BMI, LS7_BP_SCAN, LS7_BP_SELF, LS7_DIET, LS7_GLUCOSE_SCAN, LS7_GLUCOSE_SELF, LS7_LIPIDS_SCAN, LS7_LIPIDS_SELF, LS7_NICOTINE LS7_PA, LS7_SCAN, LS7_SCAN_C2, LS7_SCAN_C3, LS7_SELF, LS7_SELF_C2, LS7_SELF_C3, MED_CORTICOSTEROIDS_INHALED, MED_PDE4, MED_BRONCHODILATOR PARITY, PCE_SCORE, PCE,	NA	Added new derived variables

Data release	Variable	# participants changed	Notes
	PREVENT_BASE_10YR_CVD, PREVENT_BASE_30YR_CVD, PREVENT_BASE_10YR_ASCVD, PREVENT_BASE_30YR_ASCVD, PREVENT_BASE_10YR_HF, PREVENT_BASE_30YR_HF SUPPL_USE		
4/2026	WEIGHT_NONRESP	NA	Being released for the 1 st time because needs to be used to address missing data in Visits 2 and 3 using IPW
4/2026	NP_PERMIT	25	Corrected algorithm to include ICTB9=1
4/2026	CIGARETTE_USE	8	Corrected algorithm
4/2026	EDUCATION_C4	277	Corrected algorithm
4/2026	SUPPL_USE	630	Corrected algorithm
4/2026	SEX	NA	Rename from GENDERNUM
4/2026	HYPERTENSION2_C4	NA	Rename from HYPERTENSION_C4
4/2026	HYPERTENSION2_AHA_C5	NA	Rename from HYPERTENSION_AHA_C5
4/2026	GFRSCR_Y06_C5	NA	Rename from CKD
4/2026	GFR_Y06	NA	Rename from GFR
4/2026	GFRSCYS_Y08	NA	Rename from GFRSCYS
4/2026	GFRSCYS_SCR_Y12	NA	Rename from GFRSCYS_SCR
4/2026	GFRSCR_Y09	NA	Rename from GFRSCR
4/2026	GENDER (F, M), N_HC, POVPCT, POVPCT_C4	NA	Variables removed