



**Longitudinal Cardiovascular  
Created Variables Dataset**

**Baseline - Visit 10, and Visits 12, 13 and 15**

**CODEBOOK**

**ARCHIVED DATASET 2020**

**PAGE INTENTIONALLY BLANK**

## DOCUMENTATION FOR THE PUBLIC-USE SWAN LONGITUDINAL CARDIOVASCULAR DATASET

The longitudinal cardiovascular dataset includes created variables related to metabolic syndrome, Framingham Risk Scores, the homeostatic model assessment (HOMA), the AHA Simple 7, the Atherosclerotic Disease Risk score (ASCVD), the General Cardiovascular Risk Score and corresponding Heart Age. There are 35,543 observations in this dataset, including 3,302 unique subjects, 56 variables, and across visits 00 to 10, 12, 13 and 15. Visits 11, 14 and 16 were not included due to the absence of medication data. Visits 02 and 08 do not have cardiovascular results, so any score including them must take this into account.

### CREATED VARIABLES:

#### METABOLIC SYNDROME LONGITUDINAL VARIABLES:

The methodology for calculating metabolic syndrome is based on the following guidelines:

- American Association of Clinical Endocrinologists, where 110 mg/dL fasting blood glucose was set as a cutoff (MS and MSCM)
- National Cholesterol Education Program Adult Treatment Panel (NCEP ATP III) criteria, where the 100 mg/dL fasting blood glucose was set as a cutoff (MS100 and MSCM100)
- The following article arguing for ethnic-specific waist cutoffs for Japanese and Chinese: Tan CE, Ma S, Wai D, Chew SK, Tai ES, **Can we apply the National Cholesterol Education Program Adult Treatment Panel definition of metabolic syndrome to Asians?** *Diabetes Care*, 2004 May;27(5):1182-6 PMID 15111542 DOI: <https://doi.org/10.2337/diacare.27.5.1182>.

The four metabolic syndrome variables MS, MSCM, MS100, and MSCM100, and medication variables BPMEDS and DIABMEDS take into account a longitudinal definition of metabolic syndrome or medication use where any subject who was coded as having satisfied the criteria is coded as having metabolic syndrome or medication use at all time points thereafter.

The presence, or absence, of metabolic syndrome is determined by having at least three of the following five components: hypertension, impaired fasting glucose, obesity, low HDL, and high triglycerides. Please refer to component variables for specific cutoffs. Four different metabolic syndrome variables were created in longitudinal form and categorical form (next section)—each of which differ according to the waist circumference cutoff for obesity and/or the cutoff for impaired fasting glucose.

Cardiovascular lab results were not available for visits 02 and 08, but, in some cases, metabolic syndrome could be calculated from the remaining components, (i.e., waist circumference, diastolic/systolic blood pressure, blood pressure medication, and anti-diabetic medication). Please see code for details on calculation of this variable when no lab results are available.

The four longitudinal versions of metabolic syndrome (MS, MSCM, MS100, and MSCM100), blood pressure and anti-diabetic medications have the following rules:

- Any woman who was coded as having metabolic syndrome (or using BP or diabetes meds) is coded as having metabolic syndrome (or BP and diabetes meds) at every visit thereafter.
- Three rules apply to missing or unclassified metabolic syndrome (or BP and diabetes meds) *Note: metabolic syndrome is mentioned below, but the same applies to BP and diabetes meds longitudinal variables*:
  1. Women *never* coded as having metabolic syndrome, then have unclassified subsequent visit(s) and continue to *not* have metabolic syndrome in the next visit can have metabolic syndrome carried forward.
  2. Women *never* coded as having metabolic syndrome, then have unclassified subsequent visit(s) followed by the next visit have current metabolic syndrome stay unclassified or missing because it is unknown when the subjects started to have metabolic syndrome.
  3. Women *never* coded as having metabolic syndrome up until the last visit where it is missing or unclassified remain missing/unclassified.
- **MS**: This variable indicates a longitudinal definition of metabolic syndrome. The values include 1 (metabolic syndrome), 0 (no metabolic syndrome), and missing. The following component definitions were used:
  - Obesity: ethnic-specific waist cutoff (i.e., if Japanese/Chinese, obese if waist  $\geq 80$  cm; otherwise, obese if waist  $> 88$  cm)

- Impaired fasting glucose: impaired if glucose  $\geq 110$  mg/dl (fasting value) or antidiabetic medication use (refer to DIABMEDS categories)
  - Hypertension: hypertensive if systolic blood pressure  $\geq 130$  mmHg or diastolic blood pressure  $\geq 85$  mmHg or taking any blood pressure medication (refer to BPMEDS categories) (cross-sectional use or longitudinal use)
  - Low HDL: low if HDL results  $< 50$  mg/dL
  - High triglycerides: high if triglycerides  $\geq 150$  mg/dL (fasting value)
- **MSCM:** This variable indicates a longitudinal definition of metabolic syndrome. The values include 1 (metabolic syndrome), 0 (no metabolic syndrome), and missing. The following component definitions were used:
    - Obesity: common waist cutoff (obese if waist  $> 88$  cm)
    - Impaired fasting glucose: impaired if glucose  $\geq 110$  mg/dL (fasting value) or antidiabetic medication use (refer to DIABMEDS categories)
    - Hypertension: hypertensive if systolic blood pressure  $\geq 130$  mmHg or diastolic blood pressure  $\geq 85$  mmHg or taking any blood pressure medication (refer to BPMEDS categories) (cross-sectional use or longitudinal use)
    - Low HDL: low if HDL results  $< 50$  mg/dL
    - High triglycerides: high if triglycerides  $\geq 150$  mg/dL (fasting value)
- **MS100:** This variable indicates a longitudinal definition of metabolic syndrome. The values include 1 (metabolic syndrome), 0 (no metabolic syndrome), and missing. The following component definitions were used:
    - Obesity: ethnic-specific waist cutoff (i.e., if Japanese/Chinese, obese if waist  $\geq 80$  cm; otherwise, obese if waist  $> 88$  cm)
    - Impaired fasting glucose: impaired if glucose  $\geq 100$  mg/dL (fasting value) or antidiabetic medication use (refer to DIABMEDS categories)
    - Hypertension: hypertensive if systolic blood pressure  $\geq 130$  mmHg or diastolic blood pressure  $\geq 85$  mmHg or taking any blood pressure medication (refer to BPMEDS categories) (cross-sectional use or longitudinal use)
    - Low HDL: low if HDL results  $< 50$  mg/dL
    - High triglycerides: high if triglycerides  $\geq 150$  mg/dL (fasting value)
- **MSCM100:** This variable indicates a longitudinal definition of metabolic syndrome. The values include 1 (metabolic syndrome), 0 (no metabolic syndrome), and missing. The following component definitions were used:
    - Obesity: common waist cutoff (obese if waist  $> 88$  cm)
    - Impaired fasting glucose: impaired if glucose  $\geq 100$  mg/dL (fasting value) or antidiabetic medication use (refer to DIABMEDS categories)
    - Hypertension: hypertensive if systolic blood pressure  $\geq 130$  mmHg or diastolic blood pressure  $\geq 85$  mmHg or taking any blood pressure medication (cross-sectional use or longitudinal use)
    - Low HDL: low if HDL results  $< 50$  mg/dL
    - High triglycerides: high if triglycerides  $\geq 150$  mg/dL (fasting value)
- **BPMEDS:** This variable represents longitudinal use of blood pressure medication and was used in the coding of metabolic syndrome, Framingham Risk Score, General Cardiovascular Risk Score, Heart Age and ASCVD risk score variables.
    - A value of 1 refers to current use of at least one blood pressure medication (defined as alpha blockers, beta blockers, calcium channel blockers, ACE Inhibitors, Angiotensin II Receptor Blockers (ARBs) and diuretics) or previous use of at least one blood pressure medication.
    - A value of 0 refers to no current or previous use of blood pressure medication.
- **DIABMEDS:** This variable represents longitudinal use of anti-diabetic medication and was used in the coding of metabolic syndrome-related variables.
    - A value of 1 refers to current use of at least one anti-diabetic (defined as metformin, sulfonylurea, meglitinide, thiazolidinedione, DPP-IV Inhibitor, incretin and insulin) or previous use of at least one anti-diabetic medication.
    - A value of 0 refers to no current or previous use of anti-diabetic medication.

## **METABOLIC SYNDROME CATEGORICAL VARIABLES:**

Four-level categorical variables were created for each metabolic syndrome definition (variable names MSCAT, MSCMCAT, MS100CAT, MSCM100CAT) and both blood pressure and diabetes medications (BPMEDS and DIABMEDS) to describe each subject's **current cross-sectional status**.

- **MSCAT:** This is a four-level categorical variable created using the metabolic syndrome *component* cutoffs for the variable MS in order to describe each subject's current metabolic syndrome status at each visit.
  - CURRENT" – meets criteria for current MS (i.e., at least three components)
  - "UNCLASS" - unclassified due to missing information regarding MS at a visit
  - "NEVER" – Not satisfying the criteria for MS, and no previous presence of MS
  - "PRIOR" - No current MS, but previous presence of MS
- **MSCMCAT:** This is a four-level categorical variable created using the metabolic syndrome *component* cutoffs for the variable MSCM in order to describe each subject's current metabolic syndrome status at each visit.
  - CURRENT" – meets criteria for current MSCM (i.e., at least three components)
  - "UNCLASS" - unclassified due to missing information regarding MSCM at a visit
  - "NEVER" – Not satisfying the criteria for MSCM, and no previous presence of MSCM
  - "PRIOR" - No current MSCM, but previous presence of MSCM
- **MS100CAT:** This is a four-level categorical variable created using the metabolic syndrome *component* cutoffs for the variable MS100 in order to describe each subject's current metabolic syndrome status at each visit.
  - CURRENT" – meets criteria for current MS100 (i.e., at least three components)
  - "UNCLASS" - unclassified due to missing information regarding MS100 at a visit
  - "NEVER" – Not satisfying the criteria for MS100, and no previous presence of MS100
  - "PRIOR" - No current MS100, but previous presence of MS100
- **MSCM100CAT:** This is a four-level categorical variable created using the metabolic syndrome *component* cutoffs for the variable MSCM100 in order to describe each subject's current metabolic syndrome status at each visit.
  - CURRENT" – meets criteria for current MSCM100 (i.e., at least three components)
  - "UNCLASS" - unclassified due to missing information regarding MSCM100 at a visit
  - "NEVER" – Not satisfying the criteria for MSCM100, and no previous presence of MSCM100
  - "PRIOR" - No current MSCM100, but previous presence of MSCM100
- **BPMEDCAT:** This is a four-level categorical variable created to describe each subject's usage of blood pressure medication at each visit.
  - CURRENT" - current use of at least one blood pressure medication
  - "UNCLASS" - unclassified due to missing information regarding blood pressure medication usage at a visit (i.e., missing from the interview dataset or no information was gained from the Blood (Re)Contact form)"
  - "NEVER" - No blood pressure medication usage, and no previous blood pressure medication usage
  - "PRIOR" - No current blood pressure medication usage, but prior blood pressure medication usage
- **DIABMEDCAT:** This is a four-level categorical variable created to describe each subject's usage of anti-diabetic medication at each visit.
  - CURRENT" - current use of at least one anti-diabetic medication
  - "UNCLASS" - unclassified due to missing information regarding anti-diabetic medication usage at a visit (i.e., missing from the interview dataset or no information was gained from the Blood (Re)Contact form)"
  - "NEVER" - No anti-diabetic medication usage, and no previous anti-diabetic medication usage
  - "PRIOR" - No current anti-diabetic medication usage, but prior anti-diabetic medication usage

## CREATED SCORES:

- **CONTINUOUS METABOLIC SYNDROME - NHANES method (CONTMS)** provides a way to track metabolic syndrome related risk over time by racial group and sex (here all women). Components are glucose, triglycerides, HDL, waist circumference, and systolic blood pressure. The three racial groups evident in SWAN are non-Hispanic Caucasian, non-Hispanic African-American, and Hispanic. Chinese and Japanese are grouped with non-Hispanic Caucasians.

The methodology for calculating the continuous metabolic syndrome is described in the following articles: Gurka MJ, Lilly CL, Oliver MN, DeBoer MD, **An examination of sex and racial/ethnic differences in the metabolic syndrome among adults: a confirmatory factor analysis and a resulting continuous severity score.**, *Metabolism*, 2014 Feb;63(2):218-25, PMID: PMC4071942.

DeBoer MD, Gurka MJ, **Clinical utility of metabolic syndrome severity scores: considerations for practitioners.**, *Diabetes Metab Syndr Obes*, 2017 Feb 20;10:65-72 PMID: PMC5325095.

- **FRAMINGHAM RISK SCORE** is comprised of five components (i.e., age, smoking status, HDL, cholesterol, and blood pressure). Each woman in Visits 00, 01, 03-07, 09, 12, 13 and 15 was assigned points based on values according to these five components. The points were summed in order to assign a 10-year coronary heart disease (CHD) risk score. Visits 02, 08, 11, 14 and 16 were excluded in the FRS-related calculations.

The methodology for calculating the Framingham risk score is described in the following article:

**Executive Summary of the Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III).** *JAMA*, 2001 May 16;285(19):2486-97. <https://jamanetwork.com/journals/jama/fullarticle/193847>

- **FRPT1:** This variable is a summation of Framingham Risk-related points (from age, blood pressure, cholesterol, HDL, and smoking). The manner in which missing values of these points are handled is as follows: if any of the point categories are missing, the value of FRPT1 will be missing (i.e., all point categories are summed). This variable is recommended for use in analyses.
- **FRPT2:** This variable is also a summation of Framingham Risk-related points (from age, blood pressure, cholesterol, HDL, and smoking). However, the manner in which missing values of these points are handled is that only non-missing values of the point categories are considered (i.e., the sum function in SAS was used).
- **FRSCAT:** This variable is the Framingham Risk Score categorization. It is determined using the Framingham Risk-related points variable, FRPT1 (also FRPT2 when determining >20% risk).
  - 1 refers to  $\geq 23$  points, or a >20% 10-year hard coronary heart disease risk.
  - 2 refers to between 20 and 22 points, or an 11-17% 10-year hard coronary heart disease risk.
  - 3 refers to  $\leq 19$  points, or a 0-8% 10-year hard coronary heart disease risk.
- **ASCVD RISK SCORE (ASCVDRSK):** The ASCVD Risk Score is a measure of 10-year risk of developing a first atherosclerotic cardiovascular disease event (nonfatal myocardial infarction or coronary heart disease (CHD) death or fatal or nonfatal stroke). It was developed by a working group of the American Heart Association and the American College of Cardiology. The components of the score are: use of blood pressure medications, current smoking status, current diabetes status and log-transformed values of age, total cholesterol, HDL cholesterol, systolic blood pressure. Each component is multiplied by a race-specific coefficient and summed for a total score and transformed by an overall race-specific formula.

The methodology for calculating the risk score are described by Goff et al in this paper:

Goff DC, 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.** *Circulation* 2014;129(25 suppl 2):S49-73. [http://circ.ahajournals.org/content/129/25\\_suppl\\_2/S49](http://circ.ahajournals.org/content/129/25_suppl_2/S49).

This variable is available for visits 00, 01, 03-07, 09, 12, 13 and 15. It was set to missing if any of the original values of age, cholesterol, HDL, SBP or smoking were missing. The score can range from 0-1.

- **GENERAL CARDIOVASCULAR RISK SCORE (GENCVDRIK):** The General Cardiovascular Risk Score and corresponding Heart Age were developed by the Framingham Heart Study as a single multivariable risk function that predicts risk of developing all CVD and its constituents. The model used to create the score uses

age, total cholesterol, HDL cholesterol, systolic blood pressure, smoking status and diabetes status. Each component has certain point values assigned to different regions, with a lower point value corresponding to a lower risk. The component scores are added together (using “+” rather than the SUM() function to assure all components are present). The value of the score is mapped to an “age,” which corresponds to the age of a woman with the same predicted risk, but with all other risk factor levels in normal ranges. The “heart age” term actually represents “vascular age,” and the two are used interchangeably.

The General Cardiovascular Risk score and corresponding Heart Age was calculated using methods described here:

D’Agostino RB, Sr, Vasani RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. **General cardiovascular risk profile for use in primary care: the Framingham Heart Study.** *Circulation*. 2008;117(6):743–753. doi: 10.1161/CIRCULATIONAHA.107.699579. <http://circ.ahajournals.org/content/117/6/743.full>

- **HEARTAGE:** This is the “heart age” that corresponds to the sum of the General Cardiovascular Risk Score, according to the paper above (D’Agostino et al.)
- **HOMA** represents a computer model of the glucose-insulin feedback system during a fasting state, specifically regarding the functions of the tissues and organs related to glucose regulation. A software license for the HOMA2 model that was revised in 2004 was obtained from the publisher and involves specifically installed software

For more information on the origins of HOMA, see the two following articles:

Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC **Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man.** *Diabetologia*. 1985 Jul;28(7):412-9. DOI: <https://doi.org/10.1007/bf00280883>

Due A, Larsen TM, Hermansen K, Stender S, Holst JJ, Toubro S, Martinussen T, Astrup A **Comparison of the effects on insulin resistance and glucose tolerance of 6-mo high-monounsaturated-fat, low-fat and control diets.** *Am J Clin Nutr*, 2008 Apr;87(4):855-62. DOI:<https://doi.org/10.1093/ajcn/87.4.855>

The following three created variables use fasting glucose and insulin results for calculations:

- **HOMA\_IR:** This variable is HOMA insulin resistance (R), which is calculated using fasting insulin and glucose results through the following equation: (insulin X glucose)/22.5. Glucose (mg/dL) was converted to millimoles per liter (mmol/l) through multiplying by 0.0555, and insulin remained in the uIU/ml units.
  - **HOMA\_IS:** This variable is HOMA insulin sensitivity (%S), which is the reciprocal of insulin resistance.
  - **HOMA\_BCF:** This variable is HOMA beta ( $\beta$ )-cell function (%B).
- **AMERICAN HEART ASSOCIATION’S LIFE’S SIMPLE 7 SCORE:** This score outlines seven key risk factors for heart disease: smoking, BMI, physical activity, healthy diet, total cholesterol, blood pressure and fasting serum glucose. Each risk factor has guidelines for management to achieve the greatest benefits for heart health. **NOTE:** *This score is only available at visits where diet information from the Food Frequency Questionnaire (FFQ) was available (Visits 00, 05, and 09).*

Each risk factor is assigned a score indicating if a person has an “optimal” level of that risk factor (2 points), an “average” level (1 point), or an “inadequate” level (0 points). Three variables were created for this score:

**AHA7SCR:** The scores for each component are summed to a total score, which can range from 0 to 14.

**AHA7OPT:** counts the total number of components with ‘optimal’ scores, and ranges from 0 to 7.

**AHA7CAT:** categorizes the simple 7 score as optimal if the range is 10-14, average if the range is 5-9, and inadequate if the range is 0-4.

The methodology for calculating the AHA Life’s Simple 7 score is described in the following article:

Folsom AR, Shah AM, Lutsey PL, et al. **American Heart Association’s Life’s Simple 7: Avoiding Heart Failure and Preserving Cardiac Structure and Function.** *The American Journal of Medicine*. 2015;128(9):970-976.e2. doi:10.1016/j.amjmed.2015.03.027. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4554769/>

## COMPONENT VARIABLES:

- **AHA\_BMI:** Body Mass Index pulled from the physical measures dataset is used to assess the weight component of the AHA Simple 7 score. An optimal score (2) is  $<25 \text{ kg/m}^2$ , average (1) is  $25\text{-}29.99 \text{ kg/m}^2$ , and inadequate (0) is  $\geq 30 \text{ kg/m}^2$ .
- **AHA\_BP:** The blood pressure component of the AHA Simple 7 score measures the management of high blood pressure, and is pulled from the physical measures dataset and the BPMEDCAT created in the metabolic syndrome step. Optimal is  $<120/<80 \text{ mm Hg}$ , untreated by BP meds. Average is a SBP 120-139 or DBP 80-89 mm Hg or treated blood pressure via BP meds in the  $<120/<80 \text{ mm Hg}$  range. Inadequate is a SBP  $\geq 140$  or DBP  $\geq 90 \text{ mm Hg}$ .
- **AHA\_CHOL:** The control of high cholesterol is measured in the AHA Simple 7 cholesterol component, and is pulled from the CVR results. The SWAN drug group 9 from the SWAN drug group dataset was used to determine cholesterol drug use. Optimal is  $<200 \text{ mg/dl}$ , untreated by cholesterol controlling drugs. Average is 200-239 mg/dl untreated or  $<200 \text{ mg/dl}$  treated via a cholesterol controlling statin, and inadequate is  $\geq 240 \text{ mg/dl}$ .
- **AHA\_DIET:** The diet component was calculated using the food frequency questionnaire, which was collected at baseline, visits 05, and visit 09. Optimal is 4-5 components, average is 2-3 components, and inadequate is 0-1 components. Five components contributed to the score:
  - 1) Fruits and veg: this variable was available in “servings per day,” we assumed 1 serving is  $\frac{1}{2}$  a cup.
  - 2) Fiber-rich whole grains: The requirement was “3 1-oz servings of fiber-rich whole grains per day” where the grains had at least 1.1 g fiber per 10 g carbohydrates. The NutritionQuest software gives total fiber from grains in grams.
    - a. Our estimate is in grams, but the requirement is in ounces. We want to estimate the total fiber in grams for 3 1-oz servings of whole grains. In order to determine if the criteria were met for this, we calculate the amount of grams of fiber in the recommended daily servings in the following way:
    - b. Converting units: 1 oz fiber-rich whole grains = 28.3495 grams
    - c. According to the requirements, “Fiber-rich whole grains” have at least 1.1 g fiber/10 g carbohydrates = 0.11 fiber/total ratio or 11% fiber
    - d.  $28.3495 \text{ grams/oz fiber-rich whole grains} * 0.11 \text{ fiber/total ratio} = 3.1 \text{ g fiber} / 28.3495 \text{ g total} = 3.11 \text{ g fiber/1 oz serving}$
    - e. Requirement is 3 1-oz servings per day :  $3*3.1 \text{ g} = 9.3 \text{ g fiber from grains per day}$
    - f. Thus, if a woman’s daily fiber from grains was  $\geq 9.3 \text{ g}$ , she was considered having that component of a healthy diet.
  - 3) Fish consumption was asked for two different types of preparation: “fried fish or fish sandwich, at home or in a restaurant” and “other fish, broiled or baked.” We used the raw data which includes daily amount in grams. To convert this to weekly ounces, we added the two amounts in grams, multiplied by 0.035274 g /oz and 7 days/week. We then compared this with the  $2*3.5 \text{ oz}$  weekly recommendation to determine if a woman met the dietary component for fish.
  - 4) Sugar-sweetened beverages: We used the following items from the raw food frequency questionnaire data: Regular cola soft drinks (not diet, not ginger-ale type); Snapple, Calistoga, sweetened bottled waters or iced teas; Kool-Aid, Hi-C, or other drinks with added vitamin C; orange juice or grapefruit juice; apple juice, grape juice; or “sugar or honey in coffee or tea or on cereal.” For the liquids, we converted grams into fluid ounces and then derived the daily calories using the 450 kcal/36 oz assumption. For the daily calories from added sugar, we assumed a conversion of 100 calories per 24 grams, as per the AHA website ([http://www.heart.org/HEARTORG/HealthyLiving/HealthyEating/Nutrition/Sugar-101\\_UCM\\_306024\\_Article.jsp#.WJD-YIMrK70](http://www.heart.org/HEARTORG/HealthyLiving/HealthyEating/Nutrition/Sugar-101_UCM_306024_Article.jsp#.WJD-YIMrK70)). We added the liquid calories and solid calories and multiplied by 7 to get each woman’s estimated weekly SSB consumption. We compared with the recommended 450 oz/week to determine if a woman met the recommendation or not.
  - 5) Sodium: daily sodium intake in milligrams was directly estimated using the output from the Nutritionquest software.
- **AHA\_GLUC:** The control of blood sugar measured in the AHA Simple 7 fasting glucose component is pulled from the CVR results for the visits and the DIABMEDCAT created in the metabolic syndrome step. Optimal is  $<100 \text{ mg/dl}$ , untreated by diabetes meds, average is 100-125 mg/dl, or treated via diabetes meds in the  $<100 \text{ mg/dl}$  range, and inadequate is  $\geq 126 \text{ mg/dl}$ .

- **AHA\_PA:** The physical activity component of the AHA Simple 7 component was calculated using the Kaiser Permanente Activity Scale questionnaire. The variables used were from the “sports” section of the questionnaire. Women were classified as optimal if they exercised for more than 9 months a year, more than once a week, with a moderate or large increase in heart rate and breathing, on average 2 or more hours per week. A woman was classified as average if she met all of the previous criteria, except she exercised for less than 2 hours per week. All other women were classified as “non-exercisers” and given an inadequate score. NOTE: Women at site 16 did an abbreviated interview at Visit 09 that did not collect the physical activity questions.
- **AHA\_SMOKE:** The smoking component of the AHA Simple 7 score was derived from the Self-A form where women in SWAN were asked at follow-up visits if she smoked since her past visit, or if she stopped smoking, the month and year that she last smoked. To construct the smoking variable at Visits 05 and 09, all data about smoking habits were used from prior visits, comparing dates from prior visits to determine if former smokers quit at least 12 months before the fifth or ninth study visit. At baseline, the age that smoking stopped was subtracted from the current age to compute the score. Optimal is never or quitting smoking greater than 12 months ago, average is a recent former smoker who quit within the past 12 months, and inadequate is a current smoker.
- **HTN:** This variable represents the hypertension component of metabolic syndrome. A value of 1 refers to a systolic blood pressure  $\geq 130$  mmHg, or a diastolic blood pressure  $\geq 85$  mmHg, or cross-sectional or longitudinal use of at least one blood pressure medication. A value of 0 refers to no cross-sectional or longitudinal use of blood pressure medication.
- **HTN\_D80:** This variable is not used in any risk scores present in the dataset, but refers to an updated indicator of hypertension with a cut point of a diastolic BP  $\geq 80$  mmHg for use in analyses. A value of 1 refers to a systolic blood pressure  $\geq 130$  mmHg, or a diastolic blood pressure  $\geq 80$  mmHg, or cross-sectional or longitudinal use of at least one blood pressure medication. A value of 0 refers to no cross-sectional or longitudinal use of blood pressure medication.
- **GLUC100:** This variable represents the impaired glucose component of metabolic syndrome. A value of 1 refers to a fasting glucose result of  $\geq 100$  mg/dl OR currently taking diabetes meds, which artificially lower glucose. A value of 0 refers to a fasting glucose result of  $< 100$  mg/dl. This variable was used in the coding of the variables MS100 and MSCM100.
- **GLUC110:** This variable represents the impaired glucose component of metabolic syndrome. A value of 1 refers to a fasting glucose result of  $\geq 110$  mg/dl OR currently taking diabetes meds, which artificially lower glucose. A value of 0 refers to a fasting glucose result of  $< 110$  mg/dl. This variable was used in the coding of the variables MS and MSCM.
- **OBESE:** This variable represents the obesity component of metabolic syndrome *tailored to ethnicity*. A value of 1 refers to a waist circumference  $\geq 80$  cm if Japanese or Chinese, or a waist circumference  $> 88$  cm if Caucasian, Black, or Hispanic. A value of 0 refers to a waist circumference  $< 80$  cm if Japanese or Chinese, or a waist circumference  $\leq 88$  cm if Caucasian, Black, or Hispanic. This variable was used in the coding of the variables MS and MS100.
- **OBESECM:** This variable represents the obesity component of metabolic syndrome. A value of 1 refers to a waist circumference  $> 88$  cm—*irrespective of ethnicity*. A value of 0 refers to a waist circumference  $\leq 88$  cm. This variable was used in the coding of the variables MSCM and MSCM100.
- **HIGHTG:** This variable represents the high triglycerides component of metabolic syndrome. A value of 1 refers to a fasting triglycerides result of  $\geq 150$  mg/dl. A value of 0 refers to a fasting triglycerides result of  $< 150$  mg/dl.
- **LOWHDL:** This variable represents the low HDL component of metabolic syndrome. A value of 1 refers to an HDL result of  $< 50$  mg/dl. A value of 0 refers to an HDL result of  $\geq 50$  mg/dl.
- **STERFLAG:** This flag represents use of at least one oral or injectable corticosteroid [i.e., LDSG categories 15a and 15c), had a glucose level above 110 or 100 mg/dl, and were coded as having metabolic syndrome. This flag indicates women whose use of a steroid could inflate their glucose results.

- **FRS\_AGEPT**: This variable represents the age-related points regarding the Framingham Risk Score assigned to women with information in the cross-sectional cardiovascular datasets from visits 00, 01, 03-07, 09, 12, 13 and 15. Values were calculated to only accommodate the age groups present in the data.
  - 0 = those aged 40 to 44 years
  - 3 = those aged 45 to 49 years
  - 6 = those aged 50 to 54 years
  - 8 = those aged 55 to 59 years
  - 10 = those aged 60 to 64 years
  - 12 = those aged 65 to 69 years
  - 14 = those aged 70 to 74 years
- **FRS\_BPPT**: This variable represents the blood pressure-related points regarding the Framingham Risk Score assigned to women with information in the cross-sectional cardiovascular datasets from visits 00, 01, 03-07, 09, 12, 13 and 15. Please note that both **cross-sectional and longitudinal blood pressure medication use** was taken into account with respect to this variable.
  - 0 = systolic blood pressure <120 mmHg, irrespective of blood pressure medication use
  - 1 = systolic blood pressure between 120 and 129 mmHg not using any blood pressure medication
  - 2 = systolic blood pressure between 130 and 139 mmHg not using any blood pressure medication
  - 3 = systolic blood pressure between 120 and 129 mmHg using at least one blood pressure medication OR systolic blood pressure between 140 and 159 mmHg not using any blood pressure medication
  - 4 = systolic blood pressure between 130 and 139 mmHg using at least one blood pressure medication OR systolic blood pressure  $\geq$ 160 mmHg not using any blood pressure medication
  - 5 = systolic blood pressure between 140 and 159 mmHg using at least one blood pressure medication
  - 6 = systolic blood pressure  $\geq$ 160 mmHg using at least one blood pressure medication
- **FRS\_CHOLPT**: This variable represents the cholesterol-related points regarding the Framingham Risk Score assigned to women with information in the cross-sectional cardiovascular datasets from visits 00, 01, 03-07, 09, 12, 13 and 15.
  - 0 = a cholesterol result <160 mg/dl, irrespective of age.
  - 1 = a cholesterol result between 160 and 199 mg/dl, aged between 60 and 74 years OR a cholesterol result between 200 and 239 mg/dl, aged between 70 and 74 years
  - 2 = a cholesterol result between 160 and 199 mg/dl, aged between 50 and 59 years OR a cholesterol result between 200 and 239 mg/dl, aged between 60 and 69 years OR a cholesterol result  $\geq$ 240 mg/dl, aged between 70 and 74 years
  - 3 = a cholesterol result between 160 and 199 mg/dl, aged between 40 and 49 years OR a cholesterol result between 240 and 279 mg/dl, aged between 60 and 69 years
  - 4 = a cholesterol result between 200 and 239 mg/dl, aged between 50 and 59 years OR a cholesterol result  $\geq$ 280 mg/dl, aged between 60 and 69 years
  - 5 = a cholesterol result between 240 and 279 mg/dl, aged between 50 and 59 years
  - 6 = a cholesterol result between 200 and 239 mg/dl, aged between 40 and 49 years
  - 7 = a cholesterol result  $\geq$ 280 mg/dl, aged between 50 and 59 years
  - 8 = a cholesterol result between 240 and 279 mg/dl, aged between 40 and 49 years,
  - 10 = a cholesterol result  $\geq$ 280 mg/dl, aged between 40 and 49 years
- **FRS\_HDLPT**: This variable represents the HDL-related points regarding the Framingham Risk Score assigned to women with information in the cross-sectional cardiovascular datasets from visits 00, 01, 03-07, 09, 12, 13 and 15.
  - 1 = an HDL result of  $\geq$ 60 mg/dl
  - 0 = an HDL result between 50 and 59 mg/dl
  - 1 = an HDL result between 40 and 49 mg/dl
  - 2 = an HDL result <40 mg/dl
- **FRS\_SMOKEPT**: This variable represents the smoking-related points regarding the Framingham Risk Score assigned to a subject with information in the cross-sectional cardiovascular datasets from visits 00, 01, 03-07, 09, 12, 13 and 15.
  - 0 = did not currently smoke
  - 1 = currently smoked, aged between 70 and 74 years
  - 2 = currently smoked, aged between 60 and 69 years

4 = currently smoked, aged between 50 and 59 years  
7 = currently smoked, aged between 40 and 49 years

- **HA\_AGEPT:** This variable represents the age-related points regarding the Heart Age and General Cardiovascular Risk Scores assigned to women with information in the cross-sectional cardiovascular datasets from visits 00, 01, 03-07, 09, 12, 13 and 15.
  - 4 = those aged 40 to 44 years
  - 5 = those aged 45 to 49 years
  - 7 = those aged 50 to 54 years
  - 8 = those aged 55 to 59 years
  - 9 = those aged 60 to 64 years
  - 10 = those aged 65 to 69 years
  - 11 = those aged 70 to 74 years
- **HA\_CHOLPT:** This variable represents the cholesterol-related points regarding the Heart Age and General Cardiovascular Risk Scores assigned to women with information in the cross-sectional cardiovascular datasets from visits 00, 01, 03-07, 09, 12, 13 and 15.
  - 0 = a cholesterol result <160 mg/dl
  - 1 = a cholesterol result between 160 and 199 mg/dl
  - 3 = a cholesterol result between 200 and 239 mg/dl
  - 4 = a cholesterol result between 240 and 279 mg/dl
  - 5 = a cholesterol result ≥280 mg/dl
- **HA\_HDLPT:** This variable represents the HDL-related points regarding the Heart Age and General Cardiovascular Risk Scores assigned to a subject with information in the cross-sectional cardiovascular datasets from visits 00, 01, 03-07, 09, 12, 13 and 15.
  - 2 = an HDL result of ≥60 mg/dl
  - 1 = an HDL result between 50 and 59 mg/dl
  - 0 = an HDL result between 45 and 49 mg/dl
  - 1 = an HDL result between 35 and 44 mg/dl
  - 2 = an HDL result <35 mg/dl
- **HA\_SMOKEPT:** This variable represents the smoking-related points regarding the Heart Age and General Cardiovascular Risk Scores assigned to a subject with information in the cross-sectional cardiovascular datasets from visits 00, 01, 03-07, 09, 12, 13 and 15.
  - 0 = did not currently smoke
  - 3 = currently smoking
- **HA\_BPPT:** This variable represents the blood pressure-related points regarding the Heart Age and General Cardiovascular Risk Scores assigned to women with information in the cross-sectional cardiovascular datasets from visits 00, 01, 03-07, 09, 12, 13 and 15.
  - 3 = systolic blood pressure <120 mmHg, not using any blood pressure medication
  - 1 = systolic blood pressure <120 mmHg, using at least one blood pressure medication
  - 0 = systolic blood pressure between 120 and 129 mmHg not using any blood pressure medication
  - 1 = systolic blood pressure between 130 and 139 mmHg not using any blood pressure medication
  - 2 = systolic blood pressure between 120 and 129 mmHg using at least one blood pressure medication OR systolic blood pressure between 140 and 149 mmHg not using any blood pressure medication
  - 3 = systolic blood pressure between 130 and 139 mmHg using at least one blood pressure medication
  - 4 = systolic blood pressure between 150 and 159 mmHg not using any blood pressure medication
  - 5 = systolic blood pressure between 140 and 149 mmHg using at least one blood pressure medication OR systolic blood pressure ≥160 mmHg not using any blood pressure medication
  - 6 = systolic blood pressure between 150 and 159 mmHg using at least one blood pressure medication
  - 7 = systolic blood pressure ≥160 mmHg using at least one blood pressure medication
- **HA\_DBPT:** This variable represents the diabetes points regarding the Heart Age and General Cardiovascular Risk Scores assigned to women with information in the longitudinal diabetes dataset indicating whether they were classified as diabetic or not at that visit.
  - 0 = not diabetic
  - 4 = diabetic

- **FLAGFAS:** This variable is the flag for any fasting uncertainty that would necessitate the results for insulin, glucose, triglycerides, and LDL, be set to missing. This variable is now calculated for each visit with cardiovascular results. Glucose and triglyceride results are used in the coding for metabolic syndrome. Glucose and insulin results are used in the coding of HOMA-related variables. A value of 1 represents a subject who was:
  - not fasting or missing information regarding the fasting status at the time of the blood draw, or
  - missing the completion date of the blood draw (depending on the visit and time of blood draw, from the Blood (Re)Contact Form or follow-up interview), or
  - the completion date of the blood draw does not match any collection dates of blood draw derived from each blood draw tube/vacutainer that is recorded on the Specimen Collection Record during specimen processing, or
  - the laboratory result collection dates do not match the completion date of the blood draw, or
  - any of the collection dates from the Specimen Collection Record are different from each other, or
  - any of laboratory result collection dates do not match the dates from the Specimen Collection Record.
 Otherwise, FLAGFAS is equal to 0 or missing if during a visit without a blood draw (i.e., visits 02, 08, or 10).
  
- **RACE:** Participant race/ethnicity is provided from the Screener dataset, coded as:
  - 1= Black
  - 2= Chinese/Chinese American
  - 3= Japanese/Japanese American
  - 4= White Non-Hispanic
  - 5= Hispanic
  
- **SITE:** Participant study site is provided from the Screener dataset, coded as:
  - 11= Detroit, MI
  - 12= Boston, MA
  - 13= Chicago, IL
  - 14= Oakland, CA
  - 15= Los Angeles, CA
  - 16= Newark, NJ
  - 17= Pittsburgh, PA

## DATA DICTIONARY

Variable	Label	Values
ID	Cohort ID	7 character study ID
VISIT	Study Visit	"00" to "15"
AHA_BMI	AHA Simple 7 BMI Component	0 = $\geq 30$ kg/m <sup>2</sup> (inadeq) 1 = 24 - 29.99 kg/m <sup>2</sup> (avg) 2 = $< 25$ kg/m <sup>2</sup> (optimal)
AHA_BP	AHA Simple 7 Blood Pressure Component	0 = SBP $\geq 140$ or DBP $\geq 90$ mm Hg (inadeq) 1 = SBP 120-139 or DBP 80-89 mm Hg or treated to $<120/<80$ mm Hg (avg) 2 = $<120/<80$ mm Hg, untreated (optimal)
AHA_CHOL	AHA Simple 7 Cholesterol Component	0 = $\geq 240$ mg/dl (inadeq) 1 = 200-239 mg/dl (avg) 2 = $< 200$ mg/dl (optimal)
AHA_DIET	AHA Simple 7 Healthy Diet Component	0 = 0-1 components (inadeq) 1 = 2-3 components (avg) 2 = 4-5 components (optimal)
AHA_GLUC	AHA Simple 7 Fasting Serum Glucose Component	0 = $\geq 126$ mg/dl (inadeq) 1 = 100-125 mg/dl, or treated to $< 100$ mg/dl (avg) 2 = $< 100$ mg/dl, untreated (optimal)
AHA_PA	AHA Simple 7 Physical Activity Component	0 = None (inadeq) 1 = 1-149 min/wk moderate or 1-74 min/wk vigorous or 1-149 min/wk moderate+vigorous (avg) 2 = $\geq 150$ min/wk moderate or $\geq 75$ min/wk vigorous or $\geq 150$ min/wk moderate+vigorous (optimal)
AHA_SMOKE	AHA Simple 7 Smoking Component	0 = Current (inadeq) 1 = Former $<12$ mos ago (avg) 2 = Never or quit $>12$ mos ago (optimal)
AHA7CAT	AHA Simple 7 Categories	0 = Inadequate (0-4) 1 = Average (5-9) 2 = Optimal (10-14)
AHA7OPT	AHA Simple 7 # of Optimal Components	0 - 7
AHA7SCR	AHA Simple 7 Total Score	0 - 14
ASCVDRSK	Atherosclerotic Cardiovascular Disease (ASCVD) Risk Score For Women With Complete Risk Data ( <i>set to missing if original values of AGECONT, SBPAVG, HDL, CHOL, BPMEDS, Smoking or Diabetes are missing</i> )	numeric
BPMEDCAT	BP Med Use   Categorical	NEVER CURRENT PRIOR UNCLASS
BPMEDS	Longitudinal BP Med Use	0 = no 1 = yes
CONTMS	Continuous Metabolic Syndrome ( <i>as per the Matthew Gurka 2014 Metabolism article</i> )	numeric
DIABMEDCAT	Anti-Diabetic Med Use   Categorical	NEVER CURRENT PRIOR UNCLASS
DIABMEDS	Longitudinal Anti-Diabetic Med Use	0 = no 1 = yes
FLAGFAS	Fasting Uncertainty Flag ( <i>pulled from CVRs</i> )	0 = no 1 = yes, there is uncertainty

Variable	Label	Values
FRS_AGEPT	Framingham Age Points	0 = 40 to 44 years 3 = 45 to 49 years 6 = 50 to 54 years 8 = 55 to 59 years 10 = 60 to 64 years 12 = 65 to 69 years 14 = 70 to 74 years
FRS_BPPT	Framingham BP Points	0 = systolic BP <120 mmHg 1 = systolic BP btw 120-129 mmHg, no BP med 2 = systolic BP btw 130-139 mmHg, no BP med 3 = systolic BP btw 120-129 mmHg, ≥1 BP med OR systolic BP btw 140-159 mmHg, no BP med 4 = systolic BP btw 130-139 mmHg, ≥1 BP med OR systolic BP ≥160 mmHg no BP med 5 = systolic BP btw 140-159 mmHg, ≥1 BP med 6 = systolic BP ≥160 mmHg, ≥1 BP med
FRS_CHOLPT	Framingham Cholesterol Points	0 = <160 mg/dl 1 = 160 - 199 mg/dl, age 60-74 yrs OR 200 - 239 mg/dl, age 70-74 years 2 = 160 - 199 mg/dl, age 50-59 yrs OR 200 - 239 mg/dl, age 60-69 yrs OR ≥240 mg/dl, age 70-74 yrs 3 = 160 - 199 mg/dl, age 40-49 yrs OR 240 - 279 mg/dl, age 60-69 yrs 4 = 200 - 239 mg/dl, age 50-59 yrs ≥280 mg/dl, age 60-69 yrs 5 = 240 - 279 mg/dl, age 50-59 yrs 6 = 200 - 239 mg/dl, age 40-49 yrs 7 = ≥280 mg/dl, age 50-59 yrs 8 = 240 - 279 mg/dl, age 40-49 yrs, 10 = ≥280 mg/dl, age 40-49 years
FRS_HDLPT	Framingham HDL Points	-1 = ≥60 mg/dl 0 = 50 - 59 mg/dl 1 = 40 - 49 mg/dl 2 = <40 mg/dl
FRS_SMOKEPT	Framingham Smoking Points	0 = no smoking 1 = smoking, age 70-74 yrs 2 = smoking, age 60-69 yrs 4 = smoking, age 50-59 yrs 7 = smoking, age 40-49 yrs
FRPT1	Framingham 10-Year Risk Score 1 (Recommended)	Summation of components (missing if any components are missing)
FRPT2	Framingham 10-Year Risk Score 2	Summation of components (including all components regardless of missing status)
FRSCAT	Framingham Risk Categorization	1 = ≥23 points 2 = 20 - 22 points 3 = ≤19 points
GENCVDRISK	General Cardiovascular Risk Profile Risk Score	Numeric?
GLUC100	Impaired Glucose Met Syndrome Component (100 mg/dl)	0 = no 1 = yes

Variable	Label	Values
GLUC110	Impaired Glucose Met Syndrome Component (110 mg/dl)	0 = no 1 = yes
HA_AGEPT	Heart Age/General Cardiovascular Risk Profile Age Component	4 = 40 to 44 yrs 5 = 45 to 49 yrs 7 = 50 to 54 yrs 8 = 55 to 59 yrs 9 = 60 to 64 yrs 10 = 65 to 69 yrs 11 = 70 to 74 yrs
HA_BPPT	Heart Age/General Cardiovascular Risk Profile Blood Pressure Component	-3 = systolic BP <120 mmHg, no BP med -1 = systolic BP <120 mmHg, ≥1 BP med 0 = sys BP btw 120-129 mmHg, no BP med 1 = sys BP btw 130-139 mmHg, no BP med 2 = sys BP btw 120-129 mmHg, ≥1 BP med OR sys BP btw 140-149 mmHg, no BP med 3 = sys BP btw 130-139 mmHg, ≥1 BP med 4 = sys BP btw 150-159 mmHg, no BP med 5 = sys BP btw 140-149 mmHg, ≥1 BP med OR sys BP ≥160 mmHg, no BP med 6 = sys BP btw 150-159 mmHg, ≥1 BP med 7 = systolic BP ≥160 mmHg ≥1 BP med
HA_CHOLPT	Heart Age/General Cardiovascular Risk Cholesterol Component	0 = <160 mg/dl 1 = 160 - 199 mg/dl 3 = 200 - 239 mg/dl 4 = 240 - 279 mg/dl 5 = ≥280 mg/dl
HA_DBPT	Heart Age/General Cardiovascular Risk Diabetes Component	0 = Not diabetic 4 = Diabetic
HA_HDLPT	Heart Age/General Cardiovascular Risk HDL Component	-2 = ≥60 mg/dl -1 = 50 - 59 mg/dl 0 = 45 - 49 mg/dl 1 = 35 - 44 mg/dl 2 = <35 mg/dl
HA_SMOKEPT	Heart Age/General Cardiovascular Risk	0 = not smoking 3 = currently smoking
HEARTAGE	Heart Age (Years)	numeric
HIGHTG	High Triglycerides Met Syndrome Component	0 = no 1 = yes
HOMA_BCF	HOMA2 Beta-Cell Function	numeric
HOMA_IR	HOMA Insulin Resistance	numeric
HOMA_IS	HOMA2 Insulin Sensitivity	numeric
HTN	Hypertension Met Syndrome Component	0 = no 1 = yes
HTN_D80	Hypertension Met Syndrome – updated diastolic criteria ( <i>not used in any scoring, but available for use</i> )	0 = no 1 = yes
LOWHDL	HDL Met Syndrome Component	0 = no 1 = yes
MS	Met Syn: Gluc110, Ethnic Waist   Long	0 = no 1 = yes
MS100	Met Syn: Gluc100, Ethnic Waist   Long	0 = no 1 = yes
MS100CAT	Met Syn: Gluc100, Ethnic Waist   Cat	NEVER CURRENT PRIOR UNCLASS

Variable	Label	Values
MSCAT	Met Syn: Gluc110, Ethnic Waist   Cat	NEVER CURRENT PRIOR UNCLASS
MSCM	Met Syn: Gluc110, Common Waist   Long	0 = no 1 = yes
MSCM100	Met Syn: Gluc100, Common Waist   Long	0 = no 1 = yes
MSCM100CAT	Met Syn: Gluc100, Common Waist   Cat	NEVER CURRENT PRIOR UNCLASS
MSCMCAT	Met Syn: Gluc110, Common Waist   Cat	NEVER CURRENT PRIOR UNCLASS
OBESE	Obesity Met Syn Component (Ethnic)	0 = no 1 = yes
OBESECM	Obesity Met Syn Component (Common)	0 = no 1 = yes
STERFLAG	Corticosteroid/High Glucose Flag	0 = no 1 = yes