Jackson Heart Study

Exam 4

2021-2025

Derived Variable Data Dictionary

November, 2025

Filename: dervar4_20251104

Prepared by the JHS Coordinating Center

JHS Exam 4 Derived Variable Dictionary

Table of Contents

Derived variables dataset Version Change Log

The below table describes the changes made to the JHS derived variables dataset. Each time any variable or its definition is added, removed, changed, or altered in the derived variables dataset it will be recorded in the table here.

Modification Date	Variable Name	Reason for Change
11/04/2025	6.06. diabetes_sr_hx Diet data Cystatin C ECHO	-Was fully missing. Fix made so variable would be included -Several nutrition variables added includes MEPA and HEI diet scores (17.02, 17.03). DASH scores were updated (17.01.). LS7 Components were also updated to reflect dietary measures *Note Cystatin C (9.03), eGFR using Cystatin C (9.08) are still not included -Lab and ECHO values updated
09/03/2025	-	Derived Variables Dataset Published *Some variables are included in the data dictionary but not yet included in the derived variables dataset due to secondary party delays. This includes Cystatin C (9.03), eGFR using Cystatin C (9.08), MEPA and HEI diet scores (17.02, 17.03), LS7 Diet Components (17.03-17.08, 21.07, 21.08), LS7 Score (21.15). These will be included in future releases

0. Overview

The Jackson Heart Exam 4 Derived variables dataset has 1767 records, one for each participant who completed Exam 4 of the Jackson Heart Study. The purpose of this dataset is to provide collaborators widely used and verified derived variables. While some of these variables were also used in past JHS Exam derived variables datasets, many definitions have been updated and changed and thus currently the Exam 4 derived variables dataset should not be used longitudinally. Longitudinal datasets for JHS will be provided in the future to allow cross Exam comparison of variables from Exam 1 to Exam 4, following updates to Exams 1-3 derived variables dataset.

The data set naming convention is as follows: dervar4_YYYYMMDD.sas7bdat or .csv. Version control will be done by freezing the dataset on specific dates (YYYMMDD). Changes made to the derived variables dataset will be noted in the change log section of the data dictionary above. Dataset formats and algorithms listed are in SAS format unless noted otherwise.

Most of the variables are collected at Exam 4 with some exceptions that are noted throughout the document. Variables are named to be easily used in any software using all lower-case characters and underscores '_' to make variable name clear. Variable names also follow a naming convention to clearly identify the category of the variable. Variable naming conventions are listed in the table below. Only participants who had a valid visit are included in the derived variables dataset. We define this as participants who completed either the anthropometry form or the sitting blood pressure form. If for some reason a participant came in twice then we will take the most recent observation that coincides with the most recent consent document. If a participant has a duplicated form then we use the whichever one coincides with the most recent sitting blood pressure measure. Some participants had blood pressure collected twice due to being disqualified on initial visit and returning to complete it later on. Only data collected within 60 days of the participants exam visit date will be included in the derived variables dataset.

Prefix/Suffix	Meaning in variable name
_sr	"_sr" is used to indicate that a variable is self-reported
_hx	"_hx" is used to as history of some condition. For example, "stroke_sr_hx" is self-reported history of stroke.
meds_	"meds_" is indicative of the medication usage variable
#cat	These variables are categorized into # categories. For example, bmi4cat is BMI categorized into 4 categories. trigs4cat_atp3 is triglycerides categorized into 4 categories
#yr	This is used primarily for risk scores that predict risk of some disease within a certain time frame. For example, frs_chd10yr is the Framingham risk score of acquiring coronary heart disease within 10 years.
bp_	These are blood pressure classification labels such as Normal, Pre-hypertensive, Stage 1 HTN, and Stage II HTN. Bp_jnc7 and bp_aha2017 are the JNC7 and AHA 2017 guidelines for BP classifications respectively.
htn_	These are YN for if an individual has hypertension or not based on various guidelines.
ins_	"ins_" is used for insurance status variables

ls7_	Indicates that this variable was used in Life's Simple 7 derivations
le8_	Indicates that this variable was used in Life's Essential 8 derivations
pa_	This represents variables from the physical activity working group
lv_	"lv_" is for Left Ventricular. These variables are used in Echocardiogram data and for some ECG variables as well
_echo	Indicates that the variable was captured using an Echocardiogram
_ecg	Indicates that the variable was captured using an Electrocardiogram. For example, afib_ecg
_pct	"_pct" is a suffix used to indicate that the variable is a percentage
_u_ and _s_	Urinary and Serum assays are taken. For those with both like creatinine _u_ represents urinary while _s_ represents serum blood values.

1. Administrative

1.01. Participant ID: x_id

Variable Name: x_id

Description: JHS X ID. Participant identifier to be used for external JHS use.

Note that public repositories such as BioLINCC will have a different participant ID (**bio_id**).

Format: Character

1.02. Month of Exam 4 Date: visit_month

Variable Name: visit_month

Description: Month of participant exam Exam

Format: 1=”January”

2=”February”

3=”March”

4=”April”

5=”May”

6=”June”

7=”July”

8=”August”

9=”September”

10=”October”

11=”November”

12=”December”

Algorithm: visit_month=month(sbp200a)

Source Variables: SBP2: sbp200a

Note: The date of visit is determined as the date the participant had their most recent blood pressure taken.

1.03. Days since Exam 1: daysfromv1

Variable Name: daysfromv1

Description: Number of days since Exam 1

Format: Numeric (5.0)

Algorithm: sbp200a-v1date

Source Variables: sbp200a, v1date (from V1 derived variables dataset)

Note: Date of Exam 1 was determined as the date that the participant's fasting status was recorded.

1.04. Years since Exam 1: yearsfromv1

Variable Name: yearsfromv1

Description: Number of years since Exam 1

Format: Numeric (5.0)

Algorithm: (sbp200a-v1date)/365.25

Source Variables: sbp200a, v1date (from V1 derived variables dataset)

Note: Date of Exam 1 was determined as the date that the participant's fasting status was recorded.

1.05. Shared ARIC/JHS Only: aric

Variable Name: aric

Description: Participant type. JHS shared ARIC or JHS only at recruitment for Exam 1

Format: 0="JHS Only"

1="Shared ARIC"

1.06. JHS Recruitment Status Type: recruit

Variable Name: recruit

Description: Participant recruitment status type at Exam 1

Format: 1='Shared ARIC'

2='ARIC Household'

3='Random'

4='Family'

5='Volunteer'

Source Variables: plka1

1.07. Fasted for 8 Hours: fast8hr

Variable Name: fast8hr

Description: Indicator for participants that fasted for 8 or more hours before blood draw

Format: 0="No" (Did not fast for more than 8 hours)

1="Yes" (Fasted for 8 or more hours)

Algorithm: In the first version of the BIO form, only the time but not the date that participant started fasting was recorded. So, if the fasting start time was denoted as 'PM' then it was assumed that the fasting started the day before the venipuncture. If the fasting start time was 'AM' then it was assumed that the fasting started the day of the venipuncture.

*Start with calculating date for those who were part of the initial BIO form iteration (BIO1)

*I assume if version='1' and hour(bio204)>=12 then fasting started the night before the venipuncture;

*Otherwise if version='1' and hour(bio204)<12 then fasting probably started the same day as the venipuncture;

if version='1' and hour(bio204)>=**12** then fasttime_e=floor((bio206+(**24***3600)-bio204)/**3600**); *Use floor so rounds down;

else if version='1' and **12**>hour(bio204)>=**0** then fasttime_e=floor((bio206-bio204)/**3600**);

*For Version 2 we need to reformat date and time variables to manually calculate fastinghours;

endfast=dhms(bio200a,**0**,**0**,bio206);

format endfast datetime20.;

startfast=dhms(bio204b,**0**,**0**,bio204);

format startfast datetime20.;

*Make fasting time variable for versions 1 and 2;

if version=**2** then fasthours=floor((endfast-startfast)/**3600**);

if version=**1** then fasthours=fasttime_e;

if fasthours>=8 then fast8hr=1; *Fasted 8 or more hours;

else if fasthours<8 then fast8hr=0; *Did not fast 8 hours;

else fast8hr=.; *Missing;

Source Variables: BIO2: bio200a, bio204, bio204b, bio206, bio105, bio105a, bio107, bio107a

Notes: Several participants had potentially erroneous fasting start times/dates and venipuncture start time/dates. The most common mistake was entering

midnight as 12:00 and entering PM times as AM times. These issues were hardcoded for 10 participants.

2. Demographics

2.01. Age in years (Uncensored): age

Variable Name: age

Description: Age, based on birth date and clinic Exam date.

Note: To protect participant information, this variable will only be given to the JHS coordinating center. For external use, see censored age variable 2.02 or age category variable 2.03.

Format: Numeric, 5.0

Source Variables: Calculated internally by subtracting participant dob with the Exam date

2.02. Age in years (Censored): age_c

Variable Name: age_c

Description: Age, based on birth date and clinic Exam date, censored with top coding at 85 years old.

Note: To protect participant information, this variable will be given to external researchers. Participants 85 or older will be labeled as "85+". Analysts can choose to handle this in various ways such as setting them to a fixed value (85, median or mean of those 85+), using a random distribution to impute such as a "85+ random Poisson(3)". The mean +/- standard deviation of uncensored age is 69.28 +/- 9.75 years. There are 122 individuals labeled as "85+" with mean 87.45 +/- 2.71 years and median 87 years old.

Format: Character, $3

If participant is 85 years are older, they will be set to "85+" in the dataset

Source Variables: dervar4: Age

2.03. Age Categories: age_cat

Variable Name: age_cat

Description: Age categories.

Format: '35-44'

'45-49'

'50-54'

'55-59'

'60-64'

'65-69'

'70-74'

’75-79’

’80-84’

‘85+’

Source Variables: dervar4: Age

2.04. Sex: sex

Variable Name: sex

Description: Participant sex (“Male” or “Female”)

Format: Character, $7.; “Male”,”Female”

Algorithm: = 'Female', if elga4 = 'F'

= 'Male', if elga4 = 'M'

= missing, otherwise

Source Variables: ELGA: elga4 (From Exam 1)

2.05. Male Indicator: male

Variable Name: male

Description: Binary indicator for participant being male

Format: 1=”Male”

0=”Female”

Algorithm: = 0, if elga4 = 'F'

= 1, if elga4 = 'M'

= missing, otherwise

Source Variables: ELGA: elga4 (From Exam 1)

2.06. Alcohol drinking in the past 12 months via FFQ (y/n): alc_ffq

Variable Name: alc_ffq

Description: Indication of whether participant has consumed alcohol within the previous 12 months. This was determined in the Food Frequency Questionnaire.

Format: 1=”Yes”

0=”No”

Algorithm: If 0<ffq2139<=9 or 0<ffq2140<=9 or 0<ffq2141<=9 then alc_ffq=1;

Else If ffq2139=0 and ffq2140=0 and ffq2141=0 then alc_ffq=0;

Else alc_ffq=.;

Source Variables: FFQ2: ffq2139, ffq2140, ffq2141

2.07. Average number of drinks per week via ffq: alc_week_ffq

Variable Name: alc_week_ffq

Description: Average number of drinks a participant drinks in a week. This variable is derived using the Food Frequency Questionnaire

Format: Numeric, 8.2

Algorithm:

```
*Start by setting drinks/week value for each selection;
Array ffq2alc (*) ffq2139 ffq2140 ffq2141;
Array df (*) df1 df2 df3;
Do i=1 to dim(ffq2alc);
If ffq2alc(i)=9 then df{i)=42; *6+ drinks/day 6(7)=42;
If ffq2alc(i)=8 then df(i)=31.5; *4-5 drinks/day 4.5(7)=31.5;
If ffq2alc(i)=7 then df(i)=17.5; *2-3 drinks/day 2.5(7)=17.5;
If ffq2alc(i)=6 then df(i)=7; *1 drinks/day=7/week;
If ffq2alc(i)=5 then df(i)=5.5; *5-6 drinks/week;
If ffq2alc(i)=4 then df(i)=3; *2-4 drinks /week;
If ffq2alc(i)=3 then df(i)=1; *1 drinks/week;
If ffq2alc(i)=2 then df(i)=0.5; *1-3/month 2(1/4)=0.5;
If ffq2alc(i)=1 then df(i)=0.13; *<1 a month 0.5(1/4)=0.13;
If ffq2alc(i)=0 then df(i)=0; *Never=0;
End;

*Now derive alc_week_ffq;
If alc_ffq=0 then alc_week_ffq=0;
Else if alc_ffq=. Then alc_week_ffq=.;
Else alc_week_ffq=(df1+df2+df3)*ffq2142; *(Sum of drinks for each category
times number per session);
```

Source Variables: FFQ2: ffq2139, ffq2140, ffq2141, ffq2142

2.08. Alcohol drinking categories-NIAAA (y/n): alccat_niaaa_ffq

Variable Name: alccat_niaaa_ffq

Description: Classifies participants as non-drinkers, moderate drinkers (1-14 drinks/week for men, 1-7 for women), or heavy drinkers (over 14 drinks for men a week, over 7 drinks for women a week) by NIAAA guidelines.

Format: 0=”None”

1=”Moderate”

2=”Heavy”

Algorithm: if male=**1** and alc_week_ffq>**14** then alccat_niaaa_ffq=**2**;

else if male=**1** and **14**>=alc_week_ffq>**0** then alccat_niaaa_ffq=**1**;

else if male=**0** and alc_week_ffq>**7** then alccat_niaaa_ffq=**2**;

else if male=**0** and **7**>=alc_week_ffq>**0** then alccat_niaaa_ffq=**1**;

else if alc_week_ffq=**0** then alccat_niaaa_ffq=**0**;

else alccat_niaaa_ffq=**.**;

Source Variables: Dervar4: male, alc_week_ffq

2.09. Smoking (Combustible tobacco), self-report: smoking_tobacco

Variable Name: smoking_tobacco

Description: Indication of participant's current smoking status. This includes any combustible tobacco including cigarettes, pipes, cigars, hookah, etc. This does not include e-cigarettes, vape, or chewing tobacco.

Format: ”Never”

”Former”

”Current”

Algorithm: if tob103a in (**4**,**5**,**6**) | tob103b in (**4**,**5**,**6**) | tob103c in (**4**,**5**,**6**) | tob103d in (**4**,**5**,**6**) | tob103e in (**4**,**5**,**6**) then smoking_tobacco="Current";

else if tob103a in (**2**,**3**) | tob103b in (**2**,**3**) | tob103c in (**2**,**3**) | tob103d in (**2**,**3**) | tob103e in (**2**,**3**) then smoking_tobacco="Former";

else if tob103a=**1** and tob103b=**1** and tob103c=**1** and tob103d=**1** and tob103e=**1** then smoking_tobacco="Never";

else smoking_tobacco="";

Source Variables: TOB1: tob103a-tob103e

Note: The majority of participants who are current smokers use cigarettes. Specific information on pipes, hookah, and cigars are available upon request, but their usage is <2%. Likewise, we also have variable assessing use of e-cigarettes and chew tobacco, but less than 1% of participants are using these products

2.10. Family Income Classification: fmlyinc

Variable Name: fmlyinc

Description: Self-reported family income classification.

Format: 'A'="Less than $25,000"

'B'="$25,000-34,999"

'C'="$35,000-49,999"

'D'="$50,000-74,999"

'E'="$75,000-99,999"

'F'="$100,000+"

'K'="Don't Know"

'R'="Refused"

Algorithm:
```
if 0<sds110<=5 then fmlyinc="A";
 else if sds110=6 then fmlyinc="B";
 else if sds110=7 then fmlyinc="C";
 else if sds110=8 then fmlyinc="D";
 else if sds110=9 then fmlyinc="E";
 else if sds110=10 then fmlyinc="F";
 else if sds110=77 then fmlyinc="K";
 else if sds110=88 then fmlyinc="R";
 else fmlyinc="";
```

Source Variables: SDS1: sds110

Notes: Approximately 25% of participants selected 'Don't know' or 'refused' as their family income, so we included separate income categories for these statuses. When doing analysis, we recommend including the 'Don't know' and 'Refused' categories in the analysis (as sensitivity analyses or primary analyses as warranted) because they are not missing at random. We found that those who selected 'Don't know' were more likely to be age>80 years, female, and have lower educational attainment, whereas the people who 'refused' were more likely to have higher educational attainment and age>80 years.

2.11. Number of people family income supports: household_count

Variable Name: household_count

Description: How many people does the family income support. This variable may be used with the Family Income variable to assess inflation adjusted income levels. Note that like the family income variable this variable has a high rate of missingness

Format: Numeric, 6.0

Algorithm: if sds111>**30** then household_count=**.**; *77,88,99 used for missing values;

Else if sds111=0 then household_count=1; *Some participants answered 0 instead of 1 due to misunderstanding the question;

else household_count=sds111;

Source Variables: SDS1: sds111

2.12. Education Attainment Category: edu4cat

Variable Name: edu4cat

Description: Educational attainment derived from two self-reported questions: i) the highest degree or years of school completed, including trade or vocational school or college (pdsa18a, pdsb17a); ii) If high school not completed, indication of GED completion (pdsa18b, pdsb17b).

Note that education was not collected at Exam 4, but was collected at Exams 1 and 3. However; some participants reported a lower education level at Exam 3 than Exam 1. To mitigate this, we have used the maximum educational attainment self-reported at Exam 1 and Exam 3.

Format: 0 = "Less than high school"

1 = "High school graduate/GED"

2 = "Attended vocational school, trade school, or college "

3 = ”College degree or higher”

Algorithm: (Exam 3) if pdsb17a<=**11** then edu4catv3=**0**;

if (pdsb17a>=**12** and pdsb17a<=**13**) or pdsb17b=**1** then edu4catv3=**1**;

if pdsb17a>=**14** and pdsb17a<=16 then edu4catv3=**2**;

if pdsb17a>=17 and psdb17a<=19 then edu4catv3=3;

(Exam 1) edu = pdsa18a;

ged = pdsa18b;

if **0**<= edu <=**11** then edu4catv1=**0**;

if (edu in (**12**, **13**)) | (ged = 'Y') then edu4catv1 = **1**;

if (edu >= **14**) & (edu <= **16**) then edu4catv1 = **2**;

if (edu >= **17**) & (edu <= **19**) then edu4catv1 = **3**;

*Now take maximum of V1 and V3;

edu4cat=max(edu4catv1, edu4catv3);

Source Variables: (Exam 3) PDSB: pdsb17a, pdsb17b

(Exam 1) PDSA: pdsa18a, pdsa18b

2.13. Employment Status: employment

Variable Name: employment

Description: Employment status as collected in Jackson Heart Exam 4.

Format: 'A'="Employed, but temporarily laid off"

'B'="Homemaker, not working outside the home"

'C'="Retired from my usual job and not working"

'D'="Retired from my usual job and working for pay"

'E'="Sick or on leave for health reasons"

'F'="Unemployed, looking for work"

'G'="Unemployed, not looking for work"

'H'="Working now, full-time"

'I'="Working now, part-time"

Algorithm: if sds101=**1** then employment="H";

else if sds101=**2** then employment="I";

else if sds101=**3** then employment="A";

else if sds101=**4** then employment="E ";

else if sds101=**5** then employment="F";

else if sds101=**6** then employment="G";

else if sds101=**7** then employment="B ";

else if sds101=**8** then employment="C";

else if sds101=**9** then employment="D";

else if sds101>**9** then employment="";

Source Variables: SDS1: sds101

Notes: We recommend collapsing these variables into categories as needed (e.g., unemployed, retired) based on the research question of interest.

2.14. Marital Status: marital_status

Variable Name: marital_status

Description: Marital Status status as collected in Jackson Heart Visit 4.

Format: ‘M’=“Married/Living With Partner”

‘D’=“Divorced”

‘S’=”Separated”

‘W’=“Widowed”

‘N’=“Never Married”

Algorithm: if sds104=**1** then marital_status='M';

else if sds104=**2** then marital_status='W';

else if sds104=**3** then marital_status='D';

else if sds104=**4** then marital_status='S';

else if sds104=**5** then marital_status='N';

else marital_status=**.**;

Source Variables: SDS1: sds104

3. Anthropometry

3.01. Weight (kg): weight

Variable Name: weight

Description: Participant weight in kg

Format: Numeric, 8.1

Source Variables: ANT2: ant205b

3.02. Height (cm): height

Variable Name: height

Description: Participant height in cm

Format: Numeric, 8.1

Source Variables: ANT2: ant204

3.03. Body Mass Index kg/m^2: bmi

Variable Name: bmi

Description: Participant BMI (Weight(kg) / [Height (m)]^2)

Format: Numeric, 8.2

Algorithm: = weight / ((height / 100)**2)

Source Variables: dervar4: weight, height

3.04. Waist Circumference (cm): waist

Variable Name: waist

Description: Participant waist girth in cm

Format: Numeric, 8.2

Source Variables: ANT2: ant202

3.05. Hip Circumference (cm): hip

Variable Name: hip

Description: Participant hip girth in cm

Format: Numeric, 8.2

Source Variables: ANT2: ant203

3.06. Calculated Body Surface Area (m^2): bsa

Variable Name: bsa

Description: Body surface area estimated from height and weight measures

Format: Numeric, 8.2

Algorithm: '= 0.007184*(height**0.725)*(weight**0.425)

Source Variables: dervar4: weight, height

3.07. BMI 4-way categorization: bmi4cat

Variable Name: bmi4cat

Description: BMI categorized into 4 categories. Underweight (BMI<18.5), Normal (18.5<=BMI<25), Overweight (25<=BMI<30) and Obese (BMI>-30)

Format: 1=”Underweight”

2=”Normal”

3=”Overweight”

4=”Obese”

Algorithm: if **30**<=BMI & ^missing(BMI) then bmi4cat = **4**;

else if **25** <=BMI & BMI<**30** & ^missing(BMI) then bmi4cat = **3**;

else if **18.5**<=BMI & BMI<**25** & ^missing(BMI) then bmi4cat = **2**;

else if BMI<**18.5** & ^missing(BMI) then bmi4cat=**1;**

else bmi4cat=**.**;

Source Variables: dervar4: bmi

4. Medications

4.01. Self-reported blood pressure medication (y/n): meds_bp_sr

Variable Name: meds_bp_sr

Description: Self-reported blood pressure medication (Yes if treated)

Format: 0 = "No"

1 = "Yes"

Algorithm: if msb101a=1 then meds_bp_sr=1;

Else If msb101a=0 then meds_bp_sr=0;

Else meds_bp_sr=.;

Source Variables: MSB: msb101a

4.02. Self-reported diabetes medication (y/n): meds_dm_sr

Variable Name: meds_dm_sr

Description: Self-reported diabetes medication (Yes if treated)

Format: 0 = "No"

1 = "Yes"

Algorithm: if msb101g=1 then meds_dm_sr=1;

Else If msb101g=0 then meds_dm_sr=0;

Else meds_dm_sr=.;

Source Variables: MSB: msb101g

4.03. Self-reported cholesterol medication (y/n): meds_cholesterol_sr

Variable Name: meds_cholesterol_sr

Description: Self-reported being on cholesterol medication (Yes if treated)

Format: 0 = "No"

1 = "Yes"

Algorithm: if msb101b=1 then meds_cholesterol_sr=1;

Else If msb101b=0 then meds_cholesterol_sr=0;

Else meds_cholesterol_sr=.;

Source Variables: MSB: msb101b

4.04. Self-reported Antiarrhythmic medication Status (y/n): meds_antiaryth_sr

Variable Name: meds_antiaryth_sr

Description: Self-reported being on an antiarrhythmic medication (Yes if treated) (msb101d)

Format: 0 = "No"

1 = "Yes"

Algorithm: if msb101d=1 then meds_antiaryth_sr=1;

Else If msb101d=0 then meds_antiaryth_sr=0;

Else meds_antiaryth_sr=.;

Source Variables: MSB: msb101d

4.05. Self-reported Blood Thinner medication Status (y/n): meds_bloodthinner_sr

Variable Name: meds_bloodthinner_sr

Description: Self-reported being on a blood thinner medication (Yes if treated) (msb101f)

Format: 0 = "No"

1 = "Yes"

Algorithm: if msb101f=1 then meds_bloodthinner_sr=1;

Else If msb101f=0 then meds_bloodthinner_sr=0;

Else meds_bloodthinner_sr=.;

Source Variables: MSB: msb101f

4.06. Medication Accountability: medacct

Variable Name: medacct

Description: A measure of completeness of the record of medications taken during the four weeks prior to the clinic Exam. Depends on the responses to two questions on the medication survey form.

Format: 0 = "Incomplete recording of participant's medication use"

1 = "Participant reported no medication use"

2 = "Complete recording of all participant's medication use"

Algorithm: if msa101=0 then medAcct=1; *(Reported no medication use)

if msa102=1 and not missing(msa104a) then medAcct=2; *(Brought all medication in and data entered=complete)

if msa102=0 and missing(msa104a) then medAcct=0; *(Did not bring medications in, no data entered=incomplete)

if msa102=0 and msa103a=1 and not missing(msa104a) then medAcct=2; *Participant didn’t bring medications and said they’d follow-up and did;

if msa102=0 and msa103^=1 then medacct=0; *Participant did not bring in medication and either insisted on listing from memory or did not bring later=incomplete;

if msa102=2 then medAcct=0; *Participant only brought in partial medication;

Source Variables: MSA: msa101, msa102, msa103a; msa104a

4.07. Beta Blocker Medication Status (y/n): meds_betablk

Variable Name: meds_betablk

Description: The variable is yes if any beta blocker medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any beta blocker medication is found in the medication log then meds_betablk=1. If medacct=1 or medacct=2 and no beta blocker med is tracked then meds_betablk=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4F

Source Variables: MSA (Entire form). Dervar4: medacct

4.08. Calcium Channel Blocker Medication Status (y/n): meds_calblk

Variable Name: meds_calblk

Description: The variable is yes if any calcium channel blocker (CCB) medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any calcium channel blocker medication is found in the medication log then meds_calblk=1. If medacct=1 or medacct=2 and no CCB med is tracked then meds_calblk=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4F

Source Variables: MSA (Entire form). Dervar4: medacct

4.09. Thiazide Diuretic Medication Status (y/n): meds_diuretic_thz

Variable Name: meds_diuretic_thz

Description: The variable is yes if any Thiazide diuretic medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any thiazide diuretic medication is found in the medication log then meds_diuretic_thz=1. If medacct=1 or medacct=2 and no med is tracked then meds_diuretic_thz=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4F

Source Variables: MSA (Entire form). Dervar4: medacct

4.10. Potassium Sparing Diuretic Medication Status (y/n): meds_diuretic_k

Variable Name: meds_diuretic_k

Description: The variable is yes if any Potassium Sparing diuretic medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any potassium sparing diuretic medication is found in the medication log then meds_diuretic_k=1. If medacct=1 or medacct=2 and no med is tracked then meds_diuretic_k=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4F

Source Variables: MSA (Entire form). Dervar4: medacct

4.11. Loop Diuretic Medication Status (y/n): meds_diuretic_loop

Variable Name: meds_diuretic_loop

Description: The variable is yes if any loop diuretic medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any loop diuretic medication is found in the medication log then meds_diuretic_loop=1. If medacct=1 or medacct=2 and no med is tracked then meds_diuretic_loop=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4F

Source Variables: MSA (Entire form). Dervar4: medacct

4.12. Diuretic Medication Status (y/n): meds_diuretic

Variable Name: meds_diuretic

Description: The variable is yes if any diuretic medication is found in the medication log. (Thiazide diuretic, potassium sparing diuretic, or loop diuretic)

Format: 0 = "No"

1 = "Yes"

Algorithm: If any diuretic medication is found in the medication log then meds_diuretic=1. If medacct=1 or medacct=2 and no diuretic med is tracked then meds_diuretic=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4F

Source Variables: MSA (Entire form). Dervar4: medacct

4.13. ACE Inhibitor Medication Status (y/n): meds_acei

Variable Name: meds_acei

Description: The variable is yes if any ACE Inhibitor medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any ACE Inhibitor medication is found in the medication log then meds_acei=1. If medacct=1 or medacct=2 and no ACE med is tracked then meds_acei=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4F

Source Variables: MSA (Entire form). Dervar4: medacct

4.14. Angiotensin Receptor Blocker Medication Status (y/n): meds_arb

Variable Name: meds_arb

Description: The variable is yes if any Angiotensin Receptor Blocker (ARB) medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any ARB medication is found in the medication log then meds_arb=1. If medacct=1 or medacct=2 and no ARB med is tracked then meds_arb=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4F

Source Variables: MSA (Entire form). Dervar4: medacct

4.15. Vasodilator Medication Status (y/n): meds_vasodil

Variable Name: meds_vasodil

Description: The variable is yes if any Vasodilator medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any Vasodilator medication is found in the medication log then meds_vasodil=1. If medacct=1 or medacct=2 and no med is tracked then meds_vasodil=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4F

Source Variables: MSA (Entire form). Dervar4: medacct

4.16. Central Acting Agent Medication Status (y/n): meds_caa

Variable Name: meds_caa

Description: The variable is yes if any Central Acting Agent (CAA) medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any CAA medication is found in the medication log then meds_caa=1. If medacct=1 or medacct=2 and no CAA med is tracked then meds_caa=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4F

Source Variables: MSA (Entire form). Dervar4: medacct

4.17. Alpha Blocker Medication Status (y/n): meds_alphablk

Variable Name: meds_alphablk

Description: The variable is yes if any alpha blocker medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any alpha blocker medication is found in the medication log then meds_alphablk=1. If medacct=1 or medacct=2 and no alphablocker med is tracked then meds_alphablk=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4F

Source Variables: MSA (Entire form). Dervar4: medacct

4.18. Blood Pressure Medication from Medication Inventory (y/n): meds_bp

Variable Name: meds_bp

Description: Blood Pressure medication as tracked in the medication log for V4 (Any diuretic, beta blocker, alpha blocker, ARB, ACE Inhibitor, vasodilator, or central acting agent)

Format: 0 = "No"

1 = "Yes"

Algorithm: If any BP medication is detected then meds_bp=1. If medacct=1 or medacct=2 and no BP med is tracked then meds_bp=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4F

Source Variables: MSA (Entire form). Dervar4: medacct

4.19. Insulin Medication Status (y/n): meds_insulin

Variable Name: meds_insulin

Description: The variable is yes if insulin is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any insulin diabetes medication is found, then meds_insulin=1. If medacct=1 or medacct=2 and no non-insulin DM med is tracked then meds_insulin=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4E

Source Variables: MSA (Entire form). Dervar4: medacct

4.20. Metformin Medication Status (y/n): meds_metformin

Variable Name: meds_metformin

Description: The variable is yes if metformin is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If metformin medication is found in the medication log then meds_metformin=1. If medacct=1 or medacct=2 and no metformin is tracked then meds_metformin=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4E

Source Variables: MSA (Entire form). Dervar4: medacct

4.21. DPP-4 Inhibitors Medication Status (y/n): meds_dpp4

Variable Name: meds_dpp4

Description: The variable is yes if any DPP-4 Inhibitor medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any DPP-4 inhibitors medication is found in the medication log then meds_dpp4=1. If medacct=1 or medacct=2 and no med is tracked then meds_dpp4=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4E

Source Variables: MSA (Entire form). Dervar4: medacct

4.22. GLP-1 Receptor Agonists Medication Status (y/n): meds_glp1

Variable Name: meds_glp1

Description: The variable is yes if any GLP-1 Receptor Agonists medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any GLP-1 receptor agonist medication is found in the medication log then meds_glp1=1. If medacct=1 or medacct=2 and no GLP-1 med is tracked then meds_glp1=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4E

Source Variables: MSA (Entire form). Dervar4: medacct

4.23. SGLT-2 Inhibitors Medication Status (y/n): meds_sglt2

Variable Name: meds_sglt2

Description: The variable is yes if any SGLT-2 inhibitor medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any SGLT-2 inhibitors medication is found in the medication log then meds_sglt2=1. If medacct=1 or medacct=2 and no SLGT-2 med is tracked then meds_sglt2=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4E

Source Variables: MSA (Entire form). Dervar4: medacct

4.24. Sulfonylureas Medication Status (y/n): meds_sulfony

Variable Name: meds_sulfony

Description: The variable is yes if any sulfonylureas medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any sulfonylureas medication is found in the medication log then meds_sulfony=1. If medacct=1 or medacct=2 and no med is tracked then meds_sulfony=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4E

Source Variables: MSA (Entire form). Dervar4: medacct

4.25. Thiazolidinediones Medication Status (y/n): meds_thiazolid

Variable Name: meds_thiazolid

Description: The variable is yes if any thiazolidinediones medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any thiazolidinediones medication is found in the medication log then meds_thiazolid=1. If medacct=1 or medacct=2 and no med is tracked then meds_thiazolid=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4E

Source Variables: MSA (Entire form). Dervar4: medacct

4.26. Diabetes Medication Status from Medication Inventory (y/n): meds_dm

Variable Name: meds_dm

Description: The variable is yes if any diabetic medication is found in the medication log. This includes insulin, metformin, DPP-4 Inhibitors, GLP-1 receptor agonists, meglitinides, SGLT-2 inhibitors, sulfonylureas, and thiazolidinediones. Note that some of these medications are used for reasons beyond that of diabetes.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any diabetes medication is found then meds_dm=1. If medacct=1 or medacct=2 and no DM med is tracked then meds_dm=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4E

Source Variables: MSA (Entire form). Dervar4: medacct

Note: Note that variables of other diabetes medication classes are available but not included due to low participant usage. These variables are available upon request. These classes include Alpha Glucosidase Inhibitors and Dopamine 2 Antagonists (No JHS participant has been found to be taking these classes at Exam 4).

4.27. Statin Medication Status (y/n): meds_statin

Variable Name: meds_statin

Description: The variable is yes if any statin medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any statin medication is found in the medication log then meds_statin=1. If medacct=1 or medacct=2 and no statin med is tracked then meds_statin=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4D

Source Variables: MSA (Entire form). Dervar4: medacct

4.28. Ezetimibe Medication Status (y/n): meds_ezetimibe

Variable Name: meds_ezetimibe

Description: The variable is yes if ezetimibe is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If ezetimibe is found in the medication log then meds_ezetimibe=1. If medacct=1 or medacct=2 and no ezetimibe med is tracked then meds_ezetimibe=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4D

Source Variables: MSA (Entire form). Dervar4: medacct

4.29. Fibrate Medication Status (y/n): meds_fibrate

Variable Name: meds_fibrate

Description: The variable is yes if any fibrate medication is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any fibrate medication is found in the medication log then meds_fibrate=1. If medacct=1 or medacct=2 and no fibrate med is tracked then meds_fibrate=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4D

Source Variables: MSA (Entire form). Dervar4: medacct

4.30. Any Cholesterol Medication Status from Medication Inventory (y/n): meds_cholesterol

Variable Name: meds_cholesterol

Description: The variable is yes if any cholesterol medication is found in the medication log. This includes statins, ezetimibe, bile acid sequestrants, PCSK9 Inhibitors, bempedoic acid, fibrates, niacin, and prescribed omega-3 fatty acids.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any cholesterol medication is found in the medication log then meds_cholesterol=1. If medacct=1 or medacct=2 and no antiarrhythmic med is tracked then meds_cholesterol=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4D

Source Variables: MSA (Entire form). Dervar4: medacct

Notes: Bile Acid Sequestrants, PCSK9 Inhibitors, Bempedoic acid, Niacin, and Omega-3 Fatty Acids (prescribed) are not included in the derived variables dataset because <1% of participants reported taking these classes. These variables are available upon request.

4.31. Antiarrhythmic Medication Status from Medication Inventory (y/n): meds_antiaryth

Variable Name: meds_antiaryth

Description: The variable is yes if any antiarrhythmic medication is found in the medication log. This includes potassium blockers and sodium channel blockers.

Note: Beta Blockers and Calcium Channel Blockers may be used for arrhythmia but they do not convert AFIB and are often used for reasons beyond arrythmia such as hypertension.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any antiarrhythmic medication is found in the medication log then meds_antiaryth=1. If medacct=1 or medacct=2 and no antiarrhythmic med is tracked then meds_antiaryth=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4C

Source Variables: MSA (Entire form). Dervar4: medacct

4.32. Aspirin Medication Status (y/n): meds_aspirin

Variable Name: meds_aspirin

Description: The variable is yes if aspirin is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If aspirin is found in the medication log then meds_aspirin=1. If medacct=1 or medacct=2 and no aspirin med is tracked then meds_aspirin=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4B

Source Variables: MSA (Entire form). Dervar4: medacct

4.33. NSAID Medication Status (y/n): meds_nsaid

Variable Name: meds_nsaid

Description: The variable is yes if NSAID is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If NSAID is found in the medication log then meds_nsaid=1. If medacct=1 or medacct=2 and no NSAID is tracked then meds_nsaid=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4B

Source Variables: MSA (Entire form). Dervar4: medacct

4.34. Acetaminophen Medication Status (y/n): meds_acetamin

Variable Name: meds_acetamin

Description: The variable is yes if acetaminophen is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If acetaminophen is found in the medication log then meds_acetamin=1. If medacct=1 or medacct=2 and no acetaminophen is tracked then meds_acetamin=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4B

Source Variables: MSA (Entire form). Dervar4: medacct

4.35. Antiplatelet Medication Status from Medication Inventory (y/n): meds_antiplat

Variable Name: meds_antiplat

Description: The variable is yes if any Antiplatelet is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any antiplatelet is found in the medication log then meds_antiplat=1. If medacct=1 or medacct=2 and no antiplatelet med is tracked then meds_antiplat=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4B

Source Variables: MSA (Entire form). Dervar4: medacct

4.36. Anticoagulant Medication Status from Medication Inventory (y/n): meds_anticoag

Variable Name: meds_anticoag

Description: The variable is yes if any Anticoagulant is found in the medication log.

Format: 0 = "No"

1 = "Yes"

Algorithm: If any anticoagulant is found in the medication log then meds_anticoag=1. If medacct=1 or medacct=2 and no anticoagulant med is tracked then meds_anticoag=0. Else it will be missing.

Details for how medications were categorized can be found in Appendix 4A

Source Variables: MSA (Entire form). Dervar4: medacct

4.37. Blood Thinner Medication Status from Medication Inventory (y/n): meds_bloodthinner

Variable Name: meds_bloodthinner

Description: The variable is yes if any Blood Thinner is found in the medication log. (Defined as being on either an antiplatelet or an anticoagulant)

Format: 0 = "No"

1 = "Yes"

Algorithm: (Exam 4) If any blood thinner (antiplatelet or anticoagulant) is found in the medication log then meds_bloodthinner=1. If medacct=1 or medacct=2 and no blood thinner med is tracked then meds_bloodthinner=0. Else it will be missing.

Source Variables: MSA (Entire form). Dervar4: medacct

5. Hypertension (HTN)

5.01. Systolic Blood Pressure (mmHg): sbp

Variable Name: sbp

Description: Average of two SBP readings in V4 (Omron HEM-907XL). Note: The first two BP readings were used here to be consistent with past JHS Exams.

Format: Numeric, 8.2

Algorithm: sbp = mean(sbp205, sbp208);

Source Variables: SBP2: sbp205, sbp208

5.02. Diastolic Blood Pressure (mmHg): dbp

Variable Name: dbp

Description: Average of two DBP readings in V4 (Omron HEM-907XL). Note: The first two BP readings were used here to be consistent with past JHS Exams.

Format: Numeric, 8.2

Algorithm: dbp = (sbp206, sbp209);

Source Variables: SBP2: sbp206, sbp209

5.03. JNC7 BP classification: bp_jnc7

Variable Name: bp_jnc7

Description: Blood Pressure Classification using JNC7 guidelines

Normal: 0 < SBP < 120 and 0 < DBP < 80

Pre-HTN: 120 ≤ SBP < 140 or 80 ≤ DBP < 90

Stage I HTN: 140 ≤ SBP < 160 or 90 ≤ DBP < 100

Stage II HTN: SBP ≥ 160 or DBP ≥ 100

Format: 0 = "Normal"

1 = "Pre-HTN"

2 = "Stage I HTN"

3 = "Stage II HTN"

Algorithm: = 0, if 0 < sbp < 120 and 0 < dbp < 80

= 1, if 120 ≤ sbp < 140 or 80 ≤ dbp < 90

= 2, if 140 ≤ sbp < 160 or 90 ≤ dbp < 100

= 3, if sbp ≥ 160 or dpb ≥ 100

= missing, if missing(sbp) or missing(dbp)

Source Variables: Dervar4: sbp, dbp

Reference: Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, et al. The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure: the JNC 7 report. *JAMA*. 2003; 289(19): 2560-2571.

5.04. Hypertension Status JNC7: htn_bp_srmeds_jnc7

Variable Name: htn_bp_srmeds_jnc7

Description: HTN status is defined as blood pressure > 140/90 mmHg (per JNC 7) or use of blood pressure lowering medication.

Format: 0='No'

1='Yes'

Algorithm: if (0<=bp_jnc7 <= **1**) & meds_bp_sr = **0** then htn_bp_srmeds_jnc7 = **0**;

Else if bp_jnc7 >= **2** | meds_bp_sr = **1** then htn_bp_srmeds_jnc7 = **1**;

else htn_bp_srmeds_jnc7 = **.**; *(bp_jnc7<=1 and missing(meds_bp_sr)) or (missing(bp_jnc7) and meds_bp_sr=0) or (missing(bp_jnc7) and missing(meds_bp_sr));

Source Variables: Dervar4: bp_jnc7, meds_bp_sr

5.05. AHA/ACC 2017 BP classification: bp_aha2017

Variable Name: bp_aha2017

Description: Blood Pressure Classification using American Heart Association (AHA) and American College of Cardiology (ACC) 2017 guidelines

Normal: 0 < SBP < 120 and 0 < DBP < 80

Elevated: 120 ≤ SBP < 130 and DBP<80

Stage I HTN: 130 ≤ SBP < 140 or 80 ≤ DBP < 90

Stage II HTN: SBP ≥ 140 or DBP ≥ 90

Format: 0 = "Normal"

1 = "Elevated"

2 = "Stage I HTN"

3 = "Stage II HTN"

Algorithm: if sbp<**120** and dbp<**80** then bp_aha2017=**0**;

if **120**<=sbp<**130** and dbp<**80** then bp_aha2017=**1**;

if **140**>sbp>=**130** or **90**>dbp>=**80** then bp_aha2017=**2**;

if sbp>=**140** or dbp>=**90** then bp_aha2017=**3**;

if missing(sbp) or missing(dbp) then bp_aha2017=**.**;

Source Variables: Dervar4: sbp, dbp

Reference: Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol. 2018; 71*(19): e127-e248.

5.06. Hypertension Status AHA/ACC 2017: htn_bp_srmeds_aha2017

Variable Name: htn_bp_srmeds_aha2017

Description: HTN status is defined as blood pressure > 130/80 mmHg (per AHA/ACC 2017) or use of blood pressure lowering medication.

Format: 0='No'

1='Yes'

Algorithm: if (0<=bp_aha2017 <= **1**) & meds_bp_sr = **0** then htn_bp_srmeds_aha2017 = **0**;

if bp_aha2017 >= **2** | meds_bp_self = **1** then htn_bp_srmeds_aha2017 = **1**;

else then htn_bp_srmeds_aha2017 = **.**; *(bp_aha2017<=1 and missing(meds_bp_sr)) or (missing(bp_aha2017) and meds_bp_sr=0) or (missing(bp_aha2017) and missing(meds_bp_sr));

Source Variables: Dervar4: bp_aha2017, meds_bp_sr

5.07. Self-Reported History of Hypertension: htn_sr_hx

Variable Name: htn_sr_hx

Description: Self-reported history of hypertension at Exam 4

Format: 1='Yes'

0='No'

Algorithm: if phx103=**1** then htn_sr_hx=**1**;

else if phx103=**0** then htn_sr_hx=**0**;

else htn_sr_hx=**.**;

Source Variables: PHX1: phx103

Notes: The Personal Health Form (PHX1) asked participants *Have you ever been told by a doctor or health professional that you had/have: high blood pressure.*

However, the initial Exam 4 participants were asked '***Since your last visit*** *has a doctor or health professional told you that you have had a high blood pressure?'.* This was changed due to participants not remembering what had occurred since their last Exam (10+ years ago). For these first participants if the participant reported high blood pressure then we set their phx103=1 (Yes). If the participant did not report high blood pressure since their last Exam then we set phx103=. (missing) because we do not know for sure if they had high blood pressure or not.

6. Diabetes (DM)

6.01. Fasting Blood Glucose (mg/dL): fbg_s

Variable Name: fbg_s

Description: Fasting Blood Glucose (mg/dL). This is the serum glucose value from the lab for participants who fasted 8 or more hours prior to the blood draw.

Format: Numeric, 8.2

Algorithm: if fast8hr=1 then fbg_s=sgl;

Else fbg_s=.;

Source Variables: Dervar4: fast8hr; Central Lab (compiledresult): sgl

Notes: Instrument: Orthos Vitro 5600

Clinical Reportable Range: 20-6250 mg/dL

The raw lab data has negative value codes to represent missingness and out of range values. These have been set to missing in the derived variables dataset.

6.02. Non-Fasting Blood Glucose (mg/dL): nfbg_s

Variable Name: nfbg_s

Description: Non-Fasting Blood Glucose (mg/dL). This is the serum glucose value from the lab for participants who did not fast for 8 hours prior to blood draw.

Format: Numeric, 8.2

Algorithm: if fast8hr=0 then nfbg_s=sgl;

Else nfbg_s=.;

Source Variables: Dervar4: fast8hr; Central Lab (compiledresult): sgl

Notes: Instrument: Orthos Vitro 5600

Clinical Reportable Range: 20-6250 mg/dL

The raw lab data has negative value codes to represent missingness and out of range values. These have been set to missing in the derived variables dataset.

6.03. NGSP Hemoglobin HbA1c (%): hba1c

Variable Name: hba1c

Description: Glycated Hemoglobin Test, measures the average blood glucose level for the past 2 to 3 months and is determined by measuring the percentage of glycated hemoglobin, or HbA1c, in the blood, reported in National Glycohemoglobin Standardization Program (NGSP) units (%)

Format: Numeric, 8.2

Algorithm: Lab: HbA1C

Source Variables: Central Lab (compiledresult): HbA1C

Reference: Sacks D.B. (2012). Measurement of hemoglobin HbA1c: A new twist on the path to harmony. Diabetes Care, 35, 2667 - 2680.

Notes: Instrument: Selia Capillarys 3 Tera

Clinical Reportable Range: 4.0-16.5%

The raw lab data has negative value codes to represent missingness and out of range values. These have been set to missing in the derived variables dataset.

6.04. Diabetes Categorization: diab3cat

Variable Name: diab3cat

Description: Categorizes participants as having diabetes, pre-diabetes or no diabetes

Participant is labeled as having diabetes if any of the following is true

1). Fasting glucose>=126 mg/dL

2). Non-fasting glucose>=200 mg/dL

3). HbA1c>=6.5%

4). Participant is taking 2 or more diabetes class medications

5). Participant is taking 1 DM medication and that medication is either insulin, DPP4 inhibitor, or sulfonylureas [medications used only for diabetes]

6). Participant is taking 1 DM medication that is a SGLT2, GLP1, Thiazolidinediones, or Metformin and self-reported taking this medication for diabetes [Medication may not be used for diabetes, but participant said they were]

Otherwise participant is labeled as Pre-diabetes or at risk of diabetes if: (1) 5.7% <= HbA1c < 6.5% ; or (2) 100 mg/dL ≤ Fasting Glucose < 126 mg/dL; AND any of the following are true:

3). Participant is taking 1 DM medication that is a SGLT2, GLP1, Thiazolidinediones, or Metformin but NO self-report of taking diabetes medications or

4). No diabetes medication taken in medication inventory

Otherwise participant is labeled as having No diabetes if: (1) HbA1c < 5.7%; and (2) fasting plasma glucose < 100 mg/dL; AND any of the following are true:

3). Participant is taking 1 DM medication that is a SGLT2, GLP1, Thiazolidinediones, or Metformin but NO self-report of taking diabetes medications or

4). No diabetes medication taken in medication inventory

Otherwise diab3cat will be missing.

Note: Flowchart of diabetes categorization can be found in Appendix 6A

Format:

0="No Diabetes"

1="Pre-diabetes"

2="Diabetes"

Algorithm:

*Subvar meds_dm_sum;

meds_dm_sum=meds_insulin+meds_metformin+meds_dpp4+meds_glp1+meds_sglt2+meds_sulfony+meds_thiazolid;

*Diabetes categories;

if hba1c>=**6.5** | fbg_s>=**126**| nfbg_s>=**200** then diab3cat=**2**; *IF Biomarkers=DM then DM;

else if meds_dm_sum>=**2** then diab3cat=**2**; *More than 2 DM med then DM;

else if **0**<hba1c<**5.7** & **0**<fbg_s<**100** & meds_dm=**0** then diab3cat=**0**; *Normal biomarkers and no DM medication then NO DM;

else if (**6.5**>hba1c>=**5.7** | **126**>fbg_s >=**100**) & meds_dm = **0** then diab3cat=**1**; *No DM med, but pre-diabetes assays then PRE-DM;

else if meds_insulin=**1** or meds_dpp4=**1** or meds_sulfony=**1** then diab3cat=**2**; *These medications are DM specific so DM;

else if (meds_metformin=**1**|meds_glp1=**1**|meds_sglt2=**1**|meds_thiazolid=**1**) and meds_dm_sum=**1** and meds_dm_sr=**1** then diab3cat=**2**; *Other meds and self-reported DM THEN DM;

else if (meds_metformin=**1**|meds_glp1=**1**|meds_sglt2=**1**|meds_thiazolid=**1**) and meds_dm_sum=**1** and meds_dm_sr=**0** and (**6.5**>hba1c>=**5.7** | **126**>fbg_s >=**100**) then diab3cat=**1**; *Pre-diabetes assays, no SR med, some med classes then PRE-DM;

else if (meds_metformin=**1**|meds_glp1=**1**|meds_sglt2=**1**|meds_thiazolid=**1**) and meds_dm_sum=**1** and meds_dm_sr=**0** and (**0**<hba1c<**5.7** & **0**<fbg_s<**100**) then diab3cat=**0**; *No diabetes assays, no SR med, some med classes then No DM;

else diab3cat=**.**;

*reasons for missing: SR No DM Med and/or No DM Med Inventory, but missing HbA1C and Glucose;

*HBA1C and Glucose are Normal/Pre-diabetes, but missing SR DM Med OR missing medication inventory with SR DM Med='Yes';

*Non-fasting with non-fasting glucose <200 and taking 1 secondary DM medication with NO SR DM MED or taking 0 DM Medication with HbA1c<6.5%;

Variables: dervar4: fbg_s, nfbg_s, hba1c, meds_dm_sr, meds_dm, meds_insulin, meds_dpp4, meds_sulfony, meds_sglt2, meds_glp1, meds_thiazolid, meds_metformin

6.05. Diabetes Status: diabetes

Variable Name: diabetes

Description: Diabetes status as defined by having fasting blood glucose>=126 or non-fasting blood glucose>=200 or HbA1c>=6.5% or taking 2 or more DM medications or taking 1 DM medication that is insulin, DPP4, or sulfonylureas or taking any other DM medication with participant self-reporting taking medication for diabetes.

Format: 1=”Yes”

0=”No”

Algorithm: if diab3cat in (**0**,**1**) then diabetes=**0**;

else if diab3cat=**2** then diabetes=**1**;

else diabetes=**.**;

Source Variables: dervar4: diab3cat

6.06. Self-Reported History of Diabetes (y/n): diabetes_sr_hx

Variable Name: diabetes_sr_hx

Description: Indicator for if the participant self-reported having diabetes

Format: 1=”Yes”

0=”No”

Algorithm: if phx110=**1** then diabetes_sr_hx=**1**;

else if phx110=**0** then diabtes_sr_hx=**0**;

else diabetes_sr_hx=**.**;

Source Variables: PHX1: phx110

Notes: The Personal Health Form (PHX1) asked participants *Have you ever been told by a doctor or health professional that you had/have: diabetes.* However, the intial

Exam 4 participants were asked '***Since your last visit*** *has a doctor or health professional told you that you have had diabetes?'.* This was changed due to participants not remembering what had occurred since their last Exam (10+ years ago). For these first participants if the participant reported diabetes, then we set their phx110=1 (Yes). If the participant did not report diabetes since their last Exam then we set phx110=. (missing) because we do not know for sure if they had diabetes or not.

7. Lipids

7.01. HDL Cholesterol Level (mg/dL): hdl

Variable Name: hdl

Description: High Density Lipoprotein level (mg/dL)

Format: Numeric, 6.2

.H=missing out of range high (investigators will need to address out of range values on their own)

Algorithm: =hdl;

=.H, if hdl is out of range high;

Source Variables: Central Lab (compiledresult): hdl

Notes: Instrument: Orthos Vitro 5600

Clinical Reportable Range: 5-200 mg/dL

The raw lab data has negative value codes to represent missingness and out of range values. These have been set to missing in the derived variables dataset or to .L and .H for out of range low and high scores.

7.02. HDL Categorization (ATP-III): hdl3cat_atp3

Variable Name: hdl3cat_atp3

Description: A categorical variable for high density lipoprotein cholesterol level based on ATP III. NOTE: A high value is good (protective against heart disease)

Format: 0=”Low”

1=”Normal”

2=”High”

Algorithm: if sex = 'Male' & **0**<hdl < **40** then hdl3cat_atp3 = **0**;

else if sex = 'Female' & **0**<hdl < **50** then hdl3cat_atp3 = **0**;

else if sex = 'Male' & **40** <= hdl < **60** then hdl3cat_atp3 = **1**;

else if sex = 'Female' & **50** <= hdl < **60** then hdl3cat_atp3 = **1**;

else if hdl >= **60** or hdl=**.H** then hdl3cat_atp3 = **2**;

else if missing(hdl) then hdl3cat_atp3 = **.**;

Source Variables: Analysis4: sex, hdl

Reference: Expert Panel on Detection, Evaluation, and Treatment of high Blood Cholesterol in Adults. 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; 285(19): 2486 – 2497.

7.03. Fasting Triglycerides Cholesterol Level (mg/dL): trigs_fast

Variable Name: trigs_fast

Description: Based on triglycerides and fasting time. If fasting time < 8 hours then fasting triglyceride level is set to missing.

Format: Numeric, 7.2

Algorithm: if fast8hr=1 then trigs_fast=trig;

Else trigs_fast=.;

Source Variables: Central Lab (compiled result): trig; dervar4: fast8hr

Notes: Instrument: Orthos Vitro 5600

Clinical Reportable Range: 20-2250 mg/dL

The raw lab data has negative value codes to represent missingness and out of range values. These have been set to missing in the derived variables dataset or to .L and .H for out of range low and high scores.

7.04. Fasting Triglyceride Categorization (ATP-III): trigs4cat_atp3

Variable Name: trigs4cat_atp3

Description: A categorical variable for Fasting triglyceride level based on ATP III.

Format: 0=”Normal”

1=”Borderline High”

2=”High”

3=”Very High”

Algorithm: if trigs_fast < **150** then trigs4cat_atp3 = **0**;

if **150** <= trigs_fast < **200** then trigs4cat_atp3 = **1**;

if **200** <= trigs_fast < **500** then trigs4cat_atp3 = **2**;

if trigs_fast >= **500** then trigs4cat_atp3 = **3**;

if missing(trigs_fast) then trigs4cat_atp3 = **.**;

Source Variables: dervar4: trigs_fast

Reference: Expert Panel on Detection, Evaluation, and Treatment of high Blood Cholesterol in Adults. (2001). 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*, 285(19), 2486 – 2497.

7.05. Total Cholesterol Level (mg/dL): totchol

Variable Name: totchol

Description: Total cholesterol (mg/dL)

Format: Numeric, 6.2

Algorithm: totchol=chol

Source Variables: Central Lab (compiledresult): chol

Notes: Instrument: Orthos Vitro 5600

Clinical Reportable Range: 60-1625 mg/dL

The raw lab data has negative value codes to represent missingness and out of range values. These have been set to missing in the derived variables dataset or to .L and .H for out of range low and high scores.

7.06. Fasting LDL Cholesterol Level (mg/dL) Friedewald: ldl_fried

Variable Name: ldl_fried

Description: Based on Low Density Lipoprotein (mg/dL) cholesterol and fasting time. If fasting time < 8 hours then fasting LDL cholesterol is set to missing. Derivations were calculated by the central lab using Friedewald's method. When Friedewald's method did not apply (trigs>400), direct LDL was measured (<5 participants).

Format: Numeric, 6.2

.L=missing out of range low

Algorithm: if fast8hr=1 and ldl>0 then ldl_fried=ldl;

Else if fast8hr=1 and ldl=-333 then ldl_fried=.L; *Out of range low;

Else ldl_fried=missing;

Source Variables: Central Lab (compiledresult): ldl; dervar4: fast8hr

Reference: Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem*. 1972;18(6):499-502.

7.07. Fasting LDL Categorization (ATP-III)-Friedewald: ldl5cat_atp3

Variable Name: ldl5cat_atp3

Description: A categorical variable for fasting low density lipoprotein cholesterol level based on ATP III. Friedewald LDL used.

Format: 0=”Optimal”

1=”Near/Above Optimal”

2=”Borderline High”

3=”High”

4=”Very High”

Algorithm: if **0**<ldl_fried < **100** or ldl_fried=**.L** then ldl5cat_atp3 = **0**;

else if **100** <= ldl_fried < **130** then ldl5cat_atp3 = **1**;

else if **130** <= ldl_fried < **160** then ldl5cat_atp3 = **2**;

else if **160** <= ldl_fried < **190** then ldl5cat_atp3 = **3**;

else if ldl_fried >= **190** then ldl5cat_atp3 = **4**;

else if missing(ldl_fried) then ldl5cat_atp3 = **.**;

Source Variables: dervar4: ldl

Reference: Expert Panel on Detection, Evaluation, and Treatment of high Blood Cholesterol in Adults. (2001). 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*, 285(19), 2486 – 2497.

8. Other Biomarkers

8.01. High Sensitivity C-Reactive Protein Serum mg/dL: crp_hs

Variable Name:	crp_hs
Description:	High Sensitivity C-Reactive Protein (Serum mg/dL). This is an inflammation biomarker.
Format:	Numeric, 6.3
	=.L if out of range low
Algorithm:	if CRP=-**333** then crp_hs=**.L**;
	else if CRP=-**999** then crp_hs=**.**;
	else crp_hs=CRP/**10**;
Source Variables:	Central Lab (researchdata): crp (measured in ug/mL)
Notes:	Instrument: Siemens BNII
	Clinical Reportable Range: 0.16-51.5 ug/mL
	The raw lab data has negative value codes to represent missingness and out of range values. These have been set to missing in the derived variables dataset or to .L and .H for out of range low and high scores.

9. Renal

9.01. IDMS Tracebale Serum Creatinine (mg/dL): creat_s_idms

Variable Name: creat_s_idms

Description: IDMS Tracebale Serum Creatinine (mg/dL)

Format: Numeric, 8.2

Algorithm: if creat=-**999** then creat_s_idms=**.**;

else creat_s_idms = creat;

Source Variables: Central Lab (compiledresult): creat

Notes: Instrument: Orthos Vitro 5600

Clinical Reportable Range: 0.2-65.00 mg/dL

The raw lab data has negative value codes to represent missingness and out of range values. These have been set to missing in the derived variables dataset or to .L and .H for out of range low and high scores.

Note: Standards for measurement of blood creatinine have changed over the span of Jackson Heart. In the future longitudinal derived dataset there will be calibrated and uncalibrated versions of this variable to try to make it comparable across exams.

9.02. Spot Urinary Creatinine (mg/dL): creat_u_spot

Variable Name: creat_u_spot

Description: Random spot urine creatinine (mg/dL)

Format: Numeric, 8.2

Algorithm: if UCRR=-**999** then creat_u_spot=**.**;

else creat_u_spot = UCRR;

Source Variables: Central Lab (compiledresult): ucrr

Notes: Instrument: Orthos Vitro 5600

Clinical Reportable Range: 3.2-693 mg/dL

-999 is the code for lab values not available

The raw lab data has negative value codes to represent missingness and out of range values. These have been set to missing in the derived variables dataset or to .L and .H for out of range low and high scores.

9.03. Cystatin C concentration (Serum mg/L): cystatinc

Variable Name: cystatinc

Description: Cystatin C is a protein encoded by the CST3 gene that is mainly used as a biomarker of kidney function.

NOTE: Variable is not included in current version of the derived variables dataset, but will be included in forthcoming releases.

Format: Numeric, 6.2

Algorithm: if cys=-**999** then cystatinc=**.**;

else cystatinc = cys;

Source Variables: Central Lab (researchdata): cys

Notes: Instrument: Siemens BNII

Clinical Reportable Range: 0.064-10.1 mg/L

-999 is the code for lab values not available

The raw lab data has negative value codes to represent missingness and out of range values. These have been set to missing in the derived variables dataset or to .L and .H for out of range low and high scores.

9.04. Self-Reported dialysis or transplant (y/n): dialy_transplant_sr

Variable Name: dialy_transplant_sr

Description: Indicator for if the participant reported being on dialysis or kidney transplant at JHS V4

Note: The question used at Exam 4 asks about dialysis or transplant status together. We do not have a variable for dialysis alone.

Format: 0=”No”

1=”Yes”

Algorithm: =1, if phx111a=1;

=0, if phx111=0 or phx111a=0;

=missing, otherwise

Source Variables: PHX: phx111, phx111a

Note: This question was added shortly after exam 4 started, so the initial participants will be missing their dialysis status variable

9.05. Urinary Albumin to Creatinine Ratio (mg/g): uacr

Variable Name: uacr

Description: Urinary Albumin to Creatinine Ratio (mg/g)

Format: Numeric, 8.2

Algorithm: **For albumin we set those with out of range low values to 0.5 mg/dL (lower detection limit is 0.6 mg/dL) for purposes of ACR calculation. This approach was reviewed and approved by the JHS renal working group and upon consultations with the central lab and biostatisticians. **

if UALB=-**999** then album_u_spot_e=**.**; *This means Albumin was not collected;

else if UALB=-**555** then album_u_spot_e=**102***(**2****0.5**); *This is for out of range high albumin;

else if UALB=-**333** then album_u_spot_e=**0.5; ***This is for out of range low albumin;

else album_u_spot _e= UALB;

**Need to convert creatinine to g/dL for ACR calculation*;*

creat_u_spot_g=creat_u_spot/**1000**;

uacr=album_u_spot_e/creat_u_spot_g;

Source Variables: dervar4: creat_u_spot, Central Lab (compiledresult): UALB

Note: Due to high amount of out of range values of urinary albumin, urinary albumin was not included in the V4 derived variables dataset. It can be provided upon request, but should be used with caution.

9.06. Albuminuria Categorization-NKD: albuminuria3cat

Variable Name: albuminuria3cat

Description: Albuminuria Categorization via National Kidney Foundation guidelines using ACR. This 3-way categorization classifies ACR as Normal (<30), Moderately Elevated-microalbuminuria (30<=UACR<=300), and Severely Elevated-macroalbuminuria (>300)

Format: 'A1'="A1. Normal"

'A2'="A2. Microalbuminuria"

'A3'="A3. Macroalbuminuria"

Algorithm: if **0**<=uacr<**30** then albuminuria3cat=**'A1'**;

else if **30**<=uacr<=**300** then albuminuria3cat=**'A2'**;

else if uacr>**300** then albuminuria3cat=**'A3'**;

else albuminuria3cat=**""**;

Source Variables: dervar4: uacr

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. https://doi.org/10.1016/j.kint.2023.10.018

9.07. eGFR CKD-EPI 2021 using Creatinine: egfr_ckdepi2021_creat

Variable Name: egfr_ckdepi2021_creat

Description: Estimated eGFR (ml/min/1.73 m2) based on the CKD-EPI 2021 Formula using blood creatinine, age, and sex

Format: Numeric, 8.2

Algorithm: egfr_ckdepi2021_creat = **.**;

if ^missing(creat_s_idms) then do;

if male = **1** then egfr_ckdepi2021_creat = **142** * (min((creat_s_idms/**0.9**), **1**)**(-**0.302**)) * (max((creat_s_idms/**0.9**), **1**)**(-**1.200**)) * (**0.9938****age);

if male = **0** then egfr_ckdepi2021_creat = **142** * (min((creat_s_idms/**0.7**), **1**)**(-**0.241**)) * (max((creat_s_idms/**0.7**), **1**)**(-**1.200**)) * (**0.9938****age) * **1.012** ;

end;

Source Variables: dervar4: male, age, creat_s_idms

Reference: Inker LA, Eneanya ND, Coresh J, Tighiouart H, Wang D, 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

9.08. eGFR CKD-EPI 2021 using Creatinine and Cystatin C: egfr_ckdepi2021_creatcc

Variable Name: egfr_ckdepi2021_creatcc

Description: Estimated eGFR (ml/min/1.73 m2) based on the CKD-EPI 2021 Formula using blood creatinine, age, sex, and cystatin c.

NOTE: Variable is not included in current version of the derived variables dataset, but will be included in forthcoming releases.

Format: Numeric, 8.2

Algorithm: egfr_ckdepi2021_creatcc = **.**;

if ^missing(creat_s_idms) and ^missing(cystatinc) then do;

if male = **1** then egfr_ckdepi2021_creatcc = **135** * (min((creat_s_idms/**0.9**), **1**)**(-**0.144**)) * (max((creat_s_idms/**0.9**), **1**)**(-**0.544**))* (min((cystatinc/**0.8**), **1**)**(-**0.323**)) * (max((creat_s_idms/**0.8**), **1**)**(-**0.778**))* (**0.9961****age);

if male = **0** then egfr_ckdepi2021_creatcc = **135** * (min((creat_s_idms/**0.7**), **1**)**(-**0.219**)) * (max((creat_s_idms/**0.7**), **1**)**(-**0.544**)) *(min((cystatinc/**0.8**), **1**)**(-**0.323**)) * (max((creat_s_idms/**0.8**), **1**)**(-**0.778**))* (**0.9961****age) * **0.963** ;

end;

Source Variables: dervar4: male, age, creat_s_idms, cystatinc

Reference: Inker LA, Eneanya ND, Coresh J, Tighiouart H, Wang D, 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

9.09. eGFR Categorization: egfr6cat

Variable Name: egfr6cat

Description: eGFR Categorization via National Kidney Foundation guidelines **using eGFR with the CKD-Epi 2021 variable using age, sex, and blood creatinine**. This 6-way categorization classifies eGFR as Normal or High (>=90), mildly decreased (60-89), mildly to moderately decreased (45-59), moderately to severely decreased (30-44), severely decreased (15-29), and kidney failure (<15).

Format: ‘G1’=”G1. Normal or High”

‘G2’=”G2. Mildly decreased”

‘G3a’=”G3a. Mildly to Moderately decreased”

‘G3b’=”G3b. Moderately to Severely decreased”

‘G4’=”G4. Severely decreased”

‘G5’=”G5. Kidney Failure”

Algorithm: if egfr_ckdepi2021_creat>=90 then egfr6cat=’G1’;

Else if 90>egfr_ckdepi2021_creat>=60 then egfr6cat=’G2’;

Else if 60>egfr_ckdepi2021_creat>=45 then egfr6cat=’G3a’;

Else if 45>egfr_ckdepi2021_creat>=30 then egfr6cat=’G3b’;

Else if 30>egfr_ckdepi2021_creat>=15 then egfr6cat=’G4’;

Else if 15>egfr_ckdepi2021_creat>0 then egfr6cat=’G5’;

Else egfr6cat=’’;

Source Variables: dervar4: egfr_ckdepi2021_creat

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. https://doi.org/10.1016/j.kint.2023.10.018

10. Respiratory

10.01. Self-Reported History of Asthma (y/n): asthma_sr_hx

Variable Name: asthma_sr_hx

Description: Indicator for self-reported history of asthma.

Format: 1="Yes"

0="No"

Algorithm: if phx114=**1** then asthma_sr_hx=**1**;

else if phx114=**0** then asthma_sr_hx=**0**;

else asthma_sr_hx=**.**;

Source Variables: PHX1: phx114

Notes: The Personal Health Form (PHX1) asked participants *Have you ever been told by a doctor or health professional that you had/have: asthma.* However; the initial Exam 4 participants were asked '***Since your last visit*** *has a doctor or health professional told you that you have had asthma?'.* This was changed due to participants not remembering what has occurred since their last Exam (10+ years ago). For these first participants if the participant reported asthma then we set their phx114=1 (Yes). If the participant did not report asthma since their last Exam then we set phx110=. (missing) because we do not know for sure if they had asthma or not.

11. ECHO

11.01. Left Ventricular Mass (g) from ECHO: lvm_echo

Variable Name: lvm_echo

Description: Left Ventricular mass (LVM) in grams from ECHO based on linear measures from 2D images

Format: Numeric, 8.2

Algorithm: =LVM

Source Variables: ECHO(echcum): LVM

11.02. LVM Indexed by Height(m)^2.7: lvm_index_ht

Variable name: lvm_index_ht

Description: Left vetricular mass (LVM) in grams indexed by height in meters to 2.7 (height^2.7)

Format: Numeric, 8.2

Algorithm: lvm_index_ht=LVM/((height/100)**2.7)

Source Variables: ECHO(echcum): LVM; Dervar4: height

Reference: Nunez E, Arnett DK, Benjamin EJ, Liebson PR, Skelton TN, et al. Optimal threshold value for left ventricular hypertrophy in blacks: The Atherosclerosis Risk in Communities Study. *Hypertension*. 2005; 45:58 - 63.

11.03. LVM Indexed by Body Surface Area (BSA): lvm_index_bsa

Variable Name: lvm_index_bsa

Description: Left ventricular mass (LVM) in grams/m**2 indexed by BSA

Format: Numeric, 8.2

Algorithm: =LVMI

Source Variables: ECHO(echcum): LVMI

Reference: DuBois D. A formula to estimate the approximate surface area if height and weight be known. *Arch Intern Med*. 1916; 17: 863-871.

11.04. LVH based on LVM indexed by height (Nunez 2005): lvh_ht2005_echo

Variable Name: lvh_ht2005_echo

Description: Indicator for LV hypertrophy based on LVM indexed by Height

Format: 0=”Absent”

1=”Present”

Algorithm: if lvm_index_ht <= **51** then lvh_ht2005_echo = **0**;

if lvm_index_ht > **51** then lvh_ht2005_echo = **1**;

if missing(lvm_index_ht) then lvh_ht2005_echo = **.**;

Source Variables: dervar4: lvm_index_ht

Reference: Nunez E, Arnett DK, Benjamin EJ, Liebson PR, Skelton TN, et al. Optimal threshold value for left ventricular hypertrophy in blacks: The Atherosclerosis Risk in Communities Study. *Hypertension*. 2005; 45:58 - 63.

11.05. LVH based on LVM indexed by BSA (ASE): lvh_ase_echo

Variable Name: lvh_ase_echo

Description: Indicator for LV hypertrophy based on LVM indexed by body surface area. This definition is the recommended definition from the American Society of Echocardigraphy

Format: 0="Absent"

1="Present"

Algorithm: if lvm_index_bsa >115 and male=1 then lvh_ase_echo = **1**;

If 0<lvm_index_bsa<=115 and male=1 then lvh_ase_echo=0

if lvm_index_bsa > 95 and male=0 then lvh_ase_echo = **1**;

if 0<lvm_index_bsa<=95 and male=0 then lvh_ase_echo=0;

if missing(lvm_index_bsa) then lvh_ase_echo = **.**;

Source Variables: dervar4: lvm_index_bsa, male

Reference: DuBois D. A formula to estimate the approximate surface area if height and weight be known. *Arch Intern Med.* 1916; 17: 863-871.

11.06. Left Ventricular Ejection Fraction %: lv_ef

Variable Name: lv_ef

Description: LVEF for participant letters, includes qualitative LVEF where quantitative LVEF by Simpson's method is not possible

Format: Numeric, 8.2 (values are percents)

Algorithm: Measured directly by ECHO reading center

=LVEFREPORT

Source Variables: ECHO(echcum): LVEFREPORT

11.07. Left Ventricular Ejection Fraction Categorization: lv_ef3cat

Variable Name: lv_ef3cat

Description: Classification of ejection fraction into 'normal', 'preserved' or 'reduced'

Format: 0=”Normal”

1=”Preserved”

2=”Reduced”

Algorithm: if lv_ef>**55** then lv_ef3cat = **0**;

if **55** >= lv_ef >= **40** then lv_ef3cat = **1**;

if **40** > lv_ef then lv_ef3cat = **2**;

if missing(lv_ef) then lv_ef3cat = **.**;

Source Variables: dervar4: lv_ef

11.08. Diastolic LV Diameter (mm): lv_diasdiam

Variable Name: lv_diasdiam

Description: Internal diameter of the left ventricle in diastole

Format: Numeric, 8.2

Algorithm: lv_diasdiam = LVEDD*10

Source Variables: ECHO(echcum): LVEDD (measured in cm)

11.09. Systolic LV Diameter (mm): lv_sysdiam

Variable Name: lv_sysdiam

Description: Internal diameter of the left ventricle in systole

Format: Numeric, 8.2

Algorithm: lv_sysdiam = LVESD*10

Source Variables: ECHO(echcum): LVESD (measured in cm)

11.10. LV Fractional Shortening: lv_fs

Variable Name: lv_fs

Description: Proportion of the difference between diastolic and systolic LV diameter to diastolic diameter

Format: 0=”Normal”

1=”Abnormal”

Algorithm: if ((lv_diasdiam - lv_sysdiam)/lv_diasdiam) >= **0.29** then lv_fs = **0**;

if ((lv_diasdiam - lv_sysdiam)/lv_diasdiam) < **0.29** then lv_fs = **1**;

if missing(lv_diasdiam) | missing(lv_sysdiam) then lv_fs = **.**;

Source Variables: dervar4: lv_diasdiam, lv_sysdiam

Reference: Devereux RB, Alonso DR, Lutas EM, Gottlieb GJ, Campo E, et al. Echocardiographic assessment of left ventricular hypertrophy: Comparison to necropsy findings. *Am J Cardiology*. 1986; 57(6): 450 - 458.

11.11. Relative Wall Thickness: lv_rwt

Variable Name: lv_rwt

Description: Left ventricular relative wall thickness

Format: Numeric, 8.2

Algorithm: =LVWT

Source Variables: ECHO(echcum): LVWT

Reference: Gaasch WH and Zile MR. Left ventricular structural remodeling in health and disease. *J Am Coll Cardiol*. 2011; 58: 1733 - 40.

11.12. Left Atrial Volume (mL): lav_echo

Variable Name: lav_echo

Description: Left Atrial Volume (mL)

Format: Numeric, 8.2

Algorithm: =LAV

Source Variables: ECHO(echcum): LAV

11.13. E/e' Septal Ratio: e_em_sept

Variable Name: e_em_sept

Description: E/e' ratio from the septal mitral annulus

Format: Numeric, 8.2

Algorithm: =EEPRIMESEPT

Source Variables: ECHO(echcum): EEPRIMESEPT

11.14. E/e' Lateral Ratio: e_em_lat

Variable Name: e_em_lat

Description: E/e' ratio from the lateral mitral annulus

Format: Numeric, 8.2

Algorithm: =EEPRIMELAT

Source Variables: ECHO(echcum): EEPRIMELAT

12. Electrocardiogram (ECG)

12.01. MI determined by ECG: mi_ecg

Variable Name: mi_ecg

Description: Myocardial Infraction as determined by Minnesota code ECG.

Format: 0=”Absent”

1=”Present”

Algorithm: Given directly by ECG reading center

Source Variables: ECG(combined): mc_mi

12.02. Heart Rate determined by ECG (bpm): hr_ecg

Variable Name: hr_ecg

Description: Heart Rate measured in bpm by ECG tracing

Format: Numeric, 8.3

Algorithm: Given directly by ECG reading center

Source Variables: ECG(combined): vrate

12.03. AFIB determined by ECG: afib_ecg

Variable Name: afib_ecg

Description: Atrial Fibrillation (AFIB) as determined by ECG. Defined as Minnesota code 8-3-1

Format: 0=”Absent”

1=”Present”

Algorithm: = 0, if [missing(AR83X) & NOT missing(hr_ecg)] or AR83X = 2 or AR83X=**9** or AR83X=**0**

= 1, if AR83X = 1

= missing, otherwise

Source Variables: ECG(combined): AR83X; dervar4: hr_ecg

12.04. Atrial Flutter determined by ECG: aflutter_ecg

Variable Name: aflutter_ecg

Description: Atrial Flutter as determined by ECG. Defined as Minnesota code 8-3-2

Format: 0=”Absent”

1=”Present”

Algorithm: = 0, if [missing(AR83X) & NOT missing(hr_ecg)] or AR83X = 1 or AR83X=**9** or AR83X=**0**

= 1, if AR83X = 2

= missing, otherwise

Source Variables: ECG(combined): AR83X; dervar4: hr_ecg

12.05. QRS Interval (msec): qrs_ecg

Variable Name: qrs_ecg

Description: QRS Interval (msec)

Format: Numeric, 8.0

Algorithm: Given directly by ECG reading center

Source Variables: ECG(combined): QRSDUR

12.06. QT Interval (msec): qt_ecg

Variable Name: qt_ecg

Description: QT Interval (msec)

Format: Numeric, 8.0

Algorithm: Given directly by ECG reading center

Source Variables: ECG(combined): QTDUR

12.07. Framingham Corrected QT Interval (msec): qt_framc_ecg

Variable Name: qt_framc_ecg

Description: QT interval corrected for heart rate using the Framingham linear regression formula

Format: Numeric, 8.2

Algorithm: qt_framc_ecg = qt_ecg + 154 * (1 - (60/hr_ecg))

Source Variables: dervar4: qt_ecg, hr_ecg

Source: Rautaharju PM, Surawicz B, Gettes LS, et al. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part IV: the ST segment, T and U waves, and the QT interval: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society. Endorsed by the International Society for Computerized Electrocardiology. *J Am Coll Cardiol*. 2009;53(11):982-991. doi:10.1016/j.jacc.2008.12.014

Soliman EZ, Howard G, Cushman M, Kissela B, Kleindorfer D, et al. Prolongation of QTc and risk of stroke: The REGARDS (REasons for Geographic and Racial Differences in Stroke) Study. *J Am Coll Cardiol*. 2012; 59 (16): 1460 – 1467.

12.08. Bazett Corrected QT Interval (msec): qt_bazc_ecg

Variable Name: qt_bazc_ecg

Description: QT interval corrected for heart rate using the Bazett formula

Format: Numeric, 8.2

Algorithm: qt_bazc_ecg= qt_ecg * (hr_ecg / 60)**(1/2)

Source Variables: dervar4: qt_ecg, hr_ecg

Source: Bazett H. An analysis of the time-relations of. Heart. 1920;7:353.

Soliman EZ, Howard G, Cushman M, Kissela B, Kleindorfer D, et al. Prolongation of QTc and risk of stroke: The REGARDS (REasons for Geographic and Racial Differences in Stroke) Study. *J Am Coll Cardiol*. 2012; 59 (16): 1460 – 1467.

12.09. Hodge Corrected QT Interval (msec): qt_hodc_ecg

Variable Name: qt_hodc_ecg

Description: QT interval corrected for heart rate using the Hodge formula

Format: Numeric, 8.2

Algorithm: qt_hodc_ecg = qt_ecg + 1.75*(hr_ecg - 60)

Source Variables: dervar4: qt_ecg, hr_ecg

Reference: Hodges M. Bazett's QT correction reviewed-Evidence that a linear QT correction for heart is better. Journal of the American College of Cardiology. 1983;1:694.

Soliman EZ, Howard G, Cushman M, Kissela B, Kleindorfer D, et al. Prolongation of QTc and risk of stroke: The REGARDS (REasons for Geographic and Racial Differences in Stroke) Study. *J Am Coll Cardiol*. 2012; 59 (16): 1460 – 1467.

12.10. Fridericia Corrected QT Interval (msec): qt_fridc_ecg

Variable Name: qt_fridc_ecg

Description: QT interval corrected for heart rate using the Fridericia formula

Format: Numeric, 8.2

Algorithm: qt_frdc_ecg = qt_ecg * (hr_ecg / 60)**(1/3)

Source Variables: dervar4: qt_ecg, hr_ecg

Reference: Fridericia LS. sense.-ED. L. Acta Medica Scandinavica. 1920;53:469.

Soliman EZ, Howard G, Cushman M, Kissela B, Kleindorfer D, et al. Prolongation of QTc and risk of stroke: The REGARDS (REasons for Geographic and Racial Differences in Stroke) Study. *J Am Coll Cardiol*. 2012; 59 (16): 1460 – 1467.

12.11. Cornell Voltage (microvolts): cv_ecg

Variable Name: cv_ecg

Description: Cornell voltage (μV)

Format: Numeric, 8.0

Algorithm: Given directly by the ECG reading center

Source Variables: ECG(combined): CV

Reference: Casale PN, Devereux RB, Alonso DR, Campo E, Kligfield P. Improved sex-specific criteria of left ventricular hypertrophy for clinical and computer interpretation of electrocardiograms: validation with autopsy findings. *Circulation*. 1987; 75:565-572

12.12. LVH-Cornell Voltage Criteria: lvh_cv_ecg

Variable Name: lvh_cv_ecg

Description: Cornell voltage criteria used in defining LVH

Format: "NOLVH"=No Hypertrophy

"CONDEF"=Conduction Defect

"LVH_CV"=Hypertrophy via Cornell Voltage Criteria

Algorithm: Given directly by the ECG reading center

Source Variables: ECG(combined): LVH_CV

Reference: Casale PN, Devereux RB, Alonso DR, Campo E, Kligfield P. Improved sex-specific criteria of left ventricular hypertrophy for clinical and computer interpretation of electrocardiograms: validation with autopsy findings. *Circulation*. 1987; 75:565-572

12.13. LVH-Cornell Voltage Product: lvh_cvp_ecg

Variable Name: lvh_cvp_ecg

Description: Cornell voltage product used in defining LVH

Format: "NOLVH"=No Hypertrophy

"CONDEF"=Conduction Defect

"LVH_CVP"=Hypertrophy via Cornell Voltage Product

Algorithm: Given directly by the ECG reading center

Source Variables: ECG(combined): LVH_CVP

Reference: Molloy TJ, Okin PM, Devereux RB, Kligfield P. Electrocardiographic detection of left ventricular hypertrophy by the simple QRS voltage-duration product. *J Am Coll Cardiol*. 1992; 20(5): 1180-1186.

12.14. LVH-Minnesota Code: lvh_mc_ecg

Variable Name: lvh_mc_ecg

Description: Minnesota Code used in defining LVH

Format: "NOLVH"=No Hypertrophy

"CONDEF"=Conduction Defect

"LVH_MC"=Hypertrophy via Minnesota Code

Algorithm: Given directly by the ECG reading center

Source Variables: ECG(combined): LVH_MC

Reference: Prineas RJ, Crow RS, Zhang ZM. The Minnesota code manual of electrocardiographic findings. 2009; Springer Science & Business Media.

12.15. Sokolow-Lyon Voltage (microvolts): slv_ecg

Variable Name: slv_ecg

Description: Sokolow-Lyon voltage (μV)

Format: Numeric, 8.0

Algorithm: =sv

Source Variables: ECG(combined): SV

Reference: Sokolow M, Lyon TP. The ventricular complex in left ventricular hypertrophy as obtained by unipolar precordial and limb leads. *Am Heart J.* 1949; 37(2): 161-186.

12.16. LVH-Sokolow-Lyon Voltage: lvh_slv_ecg

Variable Name: lvh_slv_ecg

Description: Sokolow Lyon Voltage used in defining LVH

Format: "NOLVH"=No Hypertrophy

"CONDEF"=Conduction Defect

"LVH_SL"=Hypertrophy via Sokolow-Lyons Criteria

Algorithm: =lvh_slv_ecg

Source Variables: ECG(combined): LVH_SL

Reference: Sokolow M, Lyon TP. The ventricular complex in left ventricular hypertrophy as obtained by unipolar precordial and limb leads. *Am Heart J.* 1949; 37(2): 161-186.

12.17. Minnesota Code Serial Change: mc_serchange

Variable Name: mc_serchange

Description: Minnesota Code Serial Change

Format: Q1 to Q7= New incident MI

ST1 to ST7 = New ST/T abnormalities

QX= Undetermined

Algorithm: Given directly by the ECG reading center

=INCI_MC

Source Variables: ECG(combined): INCI_MC

Reference: Prineas RJ, Crow RS, Zhang ZM. The Minnesota code manual of electrocardiographic findings. 2009; Springer Science & Business Media.

13. Stroke History

13.01. Self-Reported History of Stroke: stroke_sr_hx

Variable Name: stroke_sr_hx

Description: Indicator for self-reported history of stroke.

Note that this is typically not an appropriate outcome variable for statistical analyses. If your proposal involves event outcomes, inquire about adjudicated outcome data from the JHS coordinating center.

Format: 1='Yes'

0='No'

Algorithm: if phx108=**1** then stroke_sr_hx=**1**;

else if phx108=**0** then stroke_sr_hx=**0**;

else stroke_sr_hx=**.**;

Source Variables: PHX1: phx108

Notes: The Personal Health Form (PHX1) asked participants *Have you ever been told by a doctor or health professional that you had/have: a stroke.* However; the intial Exam 4 participants were asked '***Since your last visit*** *has a doctor or health professional told you that you have had a stroke?'.* This was changed due to participants not remembering what has occurred since their last Exam (10+ years ago). For these first participants if the participant reported a stroke then we set their phx108=1 (Yes). If the participant did not report a stroke since their last Exam then we set phx108=. (missing) because we do not know for sure if they had a stroke or not.

14. CVD History

14.01. Self-Reported History of Myocardial Infarction: mi_sr_hx

Variable Name: mi_sr_Hx

Description: Indicator for self-reported history of myocardial infarction (MI)

Note that this is typically not an appropriate outcome variable for statistical analyses. If your proposal involves event outcomes, inquire about adjudicated outcome data from the JHS coordinating center.

Format: 1='Yes'

0='No'

Algorithm: if phx105=**1** then mi_sr_hx=**1**;

else if phx105=**0** then mi_sr_hx=**0**;

else mi_sr_hx=**.**;

Source Variables: PHX1: phx105

Notes: The Personal Health Form (PHX1) asked participants *Have you ever been told by a doctor or health professional that you had/have: a heart attack.* However; the initial Exam 4 participants were asked '***Since your last visit*** *has a doctor or health professional told you that you have had a heart attack?'.* This was changed due to participants not remembering what has occurred since their last Exam (10+ years ago). For these first participants if the participant reported a heart attack then we set their phx108=1 (Yes). If the participant did not report a heart attack since their last Exam then we set phx108=. (missing) because we do not know for sure if they had a heart attack or not.

14.02. Myocardial Infarction Status from ECG and Self-Report: mi_sr_ecg

Variable Name: mi_sr_ecg

Description: MI status at Exam 4 as defined by having self-reported MI or diagnostic ecg findings of MI. Note: this variable was previously labeled as Coronary Heart Disease History, but has been changed to MI for clarity.

Format: 1='Yes'

0='No'

Algorithm: if mi_ecg=1 or mi_sr_hx=1 then mi_sr_ecg=1;

else if mi_ecg=0 and mi_sr_hx=0 then mi_sr_ecg=0;

else mi_sr_ecg=.;

Source Variables: Dervar4: mi_sr_hx, mi_ecg

14.03. Stroke or MI Status: stroke_mi

Variable Name: stroke_mi

Description: Stroke or MI status as determined by having history of myocardial infarction (self-reported and ECG) or stroke (self-reported). Participant is labeled as 'yes' if they have either stroke or MI history.

Format: 1='Yes'

0='No'

Algorithm: if stroke_sr_hx=1 or mi_sr_ecg=1 then stroke_mi=1;

if stroke_sr_hx=0 and mi_sr_ecg=0 then stroke_mi=0;

else stroke_mi=.;

Source Variables: Dervar4: stroke_sr_hx, mi_sr_ecg

14.04. Self-Reported History of Heart Failure: hf_sr_hx

Variable Name: hf_sr_hx

Description: Indicator for history of heart failure.

Note that this is typically not an appropriate outcome variable for statistical analyses. If your proposal involves event outcomes, inquire about adjudicated outcome data from the JHS coordinating center.

Format: 1='Yes'

0='No'

Algorithm: if phx106=**1** then hf_sr_hx=**1**;

else if phx106=**0** then hf_sr_hx=**0**;

else hf_sr_hx=**.**;

Source Variables: PHX1: phx106

Notes: The question pertaining to heart failure was added shortly after exam 4 began, so the initial participants seen (N=134) were not asked this question.

14.05. Self-Reported History of Atrial Fibrillation (AFIB): afib_sr_hx

Variable Name: afib_sr_hx

Description: Indicator for history of AFIB

Format: 1='Yes'

0='No'

Algorithm: if phx109=**1** then afib_sr_hx=**1**;

else if phx109=**0** then afib_sr_hx=**0**;

else afib_sr_hx=**.**;

Source Variables: PHX1: phx109

Notes: The question pertaining to AFIB was added shortly after exam 4 began, so the initial participants were not asked this question.

14.06. Self-Reported History of Peripheral Artery Disease (PAD): pad_sr_hx

Variable Name: pad_sr_hx

Description: Indicator for history of peripheral artery disease.

Format: 1='Yes'

0='No'

Algorithm: if phx115=**1** then pad_sr_hx=**1**;

else if phx115=**0** then pad_sr_hx=**0**;

else pad_sr_hx=**.**;

Source Variables: PHX1: phx115

Notes: The question pertaining to peripheral artery disease was added a couple months after exam 4 began. The initial participants were asked if they experienced 'blood circulation problems' in general, so the pad_sr_hx variable is missing for these participants as this question is not necessarily comparable to the past version.

15. Healthcare

*Note: Questions related to healthcare in V4 were added shortly after exam 4 began and are not available for the initial participants.

15.01. Private Insurance Status: ins_private

Variable Name: ins_private

Description: Indicator for reporting having private insurance.

Format: 0 = "No"

1 = "Yes"

Algorithm: if sds113___1=**1** or sds113___2=**1** then ins_private=**1**;

*Hidden code here for manual adjudications for other (specify) insurance entries;

Else if sds112=**0** then ins_private=**0**;

else if sds112=**1** and (sds113___1=**.** or sds113___2=**.**) then ins_private=**0**; *sds113 was a checkbox in redcap, so it will either be 'Yes' or missing for each question;

else ins_private=**.**;

Note: Some participants selected 'Other' instead of the correct insurance type. These were manually adjudicated (See Appendix 15A).

Source Variables: SDS1: sds112, sds113__1, sds113__2 (Adjudicated variables: sds113__6, sds113f1)

15.02. Medicaid Insurance Status: ins_medicaid

Variable Name: ins_medicaid

Description: Indicator for reporting having Medicaid/public-aid insurance

Format: 0 = "No"

1 = "Yes"

Algorithm: if sds113___4=**1** then ins_medicaid=**1**;

else if sds112=**0** then ins_medicaid=**0**;

*Hidden code here for manual adjudications for other (specify) insurance entries;

else if sds112=**1** and sds113___4=**.** then ins_medicaid=**0**; *sds113 was a checkbox in redcap, so it will either be 'Yes' or missing for each question;

else ins_medicaid=**.**;

Note: Some participants selected 'Other' instead of the correct insurance type. These were manually adjudicated (See 15A).

Source Variables: SDS1: sds113__4 (Adjudicated variables: sds113__6, sds113f1)

15.03. Medicare Insurance Status: ins_medicare

Variable Name: ins_medicare

Description: Indicator for reporting having medicare insurance

Format: 0 = "No"

1 = "Yes"

Algorithm: if sds113___3=**1** then ins_medicare=**1**;

*Hidden code here for manual adjudications for other (specify) insurance entries;

else if sds112=**0** then ins_medicare=**0**;

else if sds112=**1** and sds113___3=**.** then ins_medicare=**0**; *sds113 was a checkbox in redcap, so it will either be 'Yes' or missing for each question;

else ins_medicare=**.**;

Note: Some participants selected 'Other' instead of the correct insurance type. These were manually adjudicated (See Appendix 15A).

Source Variables: SDS1: sds113__3 (Adjudicated variables: sds113__6, sds113f1)

15.04. VA/Champus Insurance Status: ins_va

Variable Name: ins_va

Description: Indicator for reporting having VA insurance

Format: 0 = "No"

1 = "Yes"

Algorithm: if sds113___5=**1** then ins_va=**1**;

*Hidden code here for manual adjudications for other (specify) insurance entries;

else if sds112=**0** then ins_va=**0**;

else if sds112=**1** and sds113___5=**.** then ins_va=**0**; *sds113 was a checkbox in redcap, so it will either be 'Yes' or missing for each question;

else ins_va=**.**;

Note: Some participants selected 'Other' instead of the correct insurance type. These were manually adjudicated (See Appendix 15A).

Source Variables: SDS1: sds112, sds113__5 (Adjudicated variables: sds113__6, sds113f1)

15.05. Health Insurance Status: insured

Variable Name: insured

Description: Indicator for reporting having any kind of health insurance

Format: 0 = "No"

1 = "Yes"

Algorithm: if sds112 in (0,1) then insured=sds112

Source Variables: SDS1: sds112

16. Validated Instruments and Psychosocial

16.01. Depressive Symptoms Score (Scale 0-60): cesd

Variable Name: cesd

Description: Center for Epidemiologic Studies Depression (CES-D) scale score. Scores of 16 or greater indicate potential depressive symptoms.

Format: Numeric, 8.2 (Range 0-60)

Algorithm: If any of the CES1 questions are >3 then they are missing. Q 4, 8, 12, and 16 of the CESD should be reverse coded

ces104i=**3**-ces104;

ces108i=**3**-ces108;

ces112i=**3**-ces112;

ces116i=**3**-ces116;

*If over 4 questions from the CESD are missing then there is not a valid score possible for CESD;

nmiss=cmiss(of ces101 -- ces120);

cesd=sum(ces101,ces102,ces103, ces104i,ces105, ces106, ces107, ces108i, ces109, ces110, ces111, ces112i, ces113, ces114, ces115, ces116i, ces117, ces118, ces119, ces120);

if nmiss>**4** then cesd=**.**;

Source Variables: CES1: ces101-ces120

Reference: Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. *Appl Psychol Meas*. 1977; 1(3): 385-401.

16.02. Total Global Stress Score (Scale 0-24): stress_perceived

Variable Name: stress_perceived

Description: Global perceived stress (GPSS) score. Also called global chronic stress. Higher scores indicate higher stress.

Format: Numeric, 8.2 (Range: 0-24)

Algorithm: If any of the GCS1 questions are >4 then they are missing. Then

nmiss=cmiss(of gcs101 -- gcs108);

stress_perceived=sum(gcs101,gcs102,gcs103, gcs104,gcs105, gcs106, gcs107, gcs108)-8; *Values are 1-4, but should be 0-3. So we subtract 1 from each question;

if nmiss>**0** then stress_perceived=**.**;

Source Variables: GCS1: gcs101-gcs108

Reference: Bromberger JT, Matthews KA. A longitudinal study of the effects of pessimism, trait anxiety, and life stress on depressive symptoms in middle-aged women. *Psychol Aging*. 1996; *11*(2): 207-213.

Cohen S, Kamarck T, Mermelstein, R. A global measure of perceived stress. *J Health Soc Behav*. 1983; 24(4): 385-396.

Payne TJ, Wyatt SB, Mosley TH, Dubbert PM, Guiterrez-Mohammed ML, et al. Sociocultural methods in the Jackson Heart Study: conceptual and descriptive overview. *Ethn Dis*. 2005; 15 (4 Suppl 6): S6-38-48.

16.03. Chronic Burden Score (Scale 0-5): chronic_burden

Variable Name: chronic_burden

Description: Chronic Burden scale score. The instrument has 5 items with subitems for duration and intensity of the stress. The item score is 0 or 1. The chronic burden score is derived by adding up the item scores. If any of the items are missing then the chronic burden score is missing. Higher scores indicate higher burden and stress in a participant's health, job, finances, and relationships.

Format: Numeric, 8.2 (Range: 1-5)

Algorithm: if psi101=**1** and psi101a=**1** and psi101b in (**2**,**3**) then cbf1=**1**;

if psi101=**0** or psi101a=**0** or psi101b=**1** then cbf1=**0**;

if psi102=**1** and psi102a=**1** and psi102b in (**2**,**3**) then cbf2=**1**;

if psi102=**0** or psi102a=**0** or psi102b=**1** then cbf2=**0**;

if psi103=**1** and psi103a=**1** and psi103b in (**2**,**3**) then cbf3=**1**;

if psi103=**0** or psi103a=**0** or psi103b=**1** then cbf3=**0**;

if psi104=**1** and psi104a=**1** and psi104b in (**2**,**3**) then cbf4=**1**;

if psi104=**0** or psi104a=**0** or psi104b=1 then cbf4=**0**;

if psi105=**1** and psi105a=**1** and psi105b in (**2**,**3**) then cbf5=**1**;

if psi105=**0** or psi105a=**0** or psi105b=**1** then cbf5=**0**;

if cmiss(cbf1--cbf5)>**0** then chronic_burden=**.**;

else chronic_burden=sum(cbf1, cbf2, cbf3, cbf4, cbf5);

Source Variables: PSI1: psi101-psi105b

Reference: Diez Roux AV, Ranjit N, Powell L, Jackson S, Lewis TT, Shea S, Wu C. Psychosocial factors and coronary calcium in adults without clinical cardiovascular disease.

Ann Intern Med. 2006;144(11):822-31. doi: 10.7326/0003-4819-144-11-200606060-00008. PMID: 16754924.

16.04. Life Satisfaction Score (Scale 5-35): life_satisfaction

Variable Name: life_satisfaction

Description: Life satisfaction (SWLS: Satisfaction with Life Scale) is a 5-item scale measuring life satisfaction (Diener et al., 1985). The item score ranges from 1 to 7. The life satisfaction score is derived by adding up the item scores. If any of the items are missing then the life satisfaction score is missing. Higher scores indicate higher life Satisfaction with 31-35: Extremely Satisfied; 26-30: Satisfied; 21-25: Slightly Satisfied; 20: Neutral; 15-19: Slightly Dissatisfied; 10-14: Dissatisfied; 5-9: Extremely Dissatisfied.

Format: Numeric, 8.2 (Range 5-35)

Algorithm: If psi112-psi116 are >=7 (refused/don't know) then they are set to missing. Note that life satisfaction scores should be 1 (Strongly disagree) to 7 (Strongly agree), but JHS scores are from 0-6 so 5 was added to the sum of the options to account for this.

```
nmiss_ls=cmiss(of psi112 -- psi116);

life_satisfaction=sum(psi112,psi113, psi114, psi115, psi116)+5;

if nmiss_ls>0 then life_satisfaction=.;
```

Source Variables: PSI1: psi112-psi116

Reference: Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction with Life Scale. *J Pers Assess*. 1985;49(1):71-75.

16.05. Brief Resilience Scale (Scale 1-5): resilience

Variable Name: resilience

Description: Resilience (BRS: Brief Resilience Scale) is a 6-item scale measuring an individual's ability to bounce back, resist illness, and adapt to stress or thrive in the face of adversity (Smith et al., 2017). Higher scores indicate higher resilience. The item score ranges from 1 to 5. Items 1, 3, and 5 are positively worded while items 2, 4, and 6 are negatively worded. The BRS is derived by reverse coding items 2, 4, and 6, and finding the mean of the 6 items. If any of the items are missing then the BRS score is missing.

Format: Numeric, 8.2 (Range 1-5)

Algorithm: he Likert Scale for the BRS is as follows: 1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree.

If psi124-psi129 are >=7 then they are set to missing. For psi124, psi126, and psi128 a value of 1 was added to the score to get the scale on a 1-5 scale instead of 0-4 as recorded in JHS. For psi125, psi127, and psi129, the value was subtracted from 5 to reverse code and obtain on a 1-5 scale.

```
nmiss_r=cmiss(of psi124-- psi129);

psi124_e=psi124+1;

psi125_e=5-psi125;

psi126_e=psi126+1;

psi127_e=5-psi127;

psi128_e=psi128+1;

psi129_e=5-psi129;

resilience=mean(psi124_e, psi125_e, psi126_e, psi127_e, psi128_e, psi129_e);

if nmiss_r>0 then resilience=.;
```

Source Variables: PSI1: psi124-psi129

Reference: Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. *Int J Behav Med*. 2008; 15(3): 194-200.

16.06. Health Literacy Score (Scale 4-20): health_lit

Variable Name: health_lit

Description: The Health Literacy (LIT) form consists of 4 questions to assess how well participants are able to understand information obtain from their doctors about their health (Anderson et al., 2021; Chew et al., 2004; Woloshin et al., 2005). Prior studies have dichotomized health literacy as adequate (score >10) or limited (score ≤10). The item score for items 1, 2 and 3 is 1 to 5 and the item score for item 4 is 1 to 4. The health literacy score is derived by adding up the item scores. If any of the items are missing then the health literacy score is missing.

Format: Numeric, 8.2 (Range 4-20)

Algorithm: If lit101-lit104 are >=7 then they are set to missing. The 4 question scores are summed to obtain the health literacy score

```
nmiss_lit=cmiss(of lit101-- lit104);

health_lit=sum(lit101, lit102, lit103, lit104);

if nmiss_lit>0 then health_lit=.;
```

Source Variables: LIT1: lit101-lit104

Reference: Anderson MD, Merkin SS, Everson-Rose SA, Widome R, Seeman T, et al. Health literacy within a diverse community-based cohort: the Multi-Ethnic Study of Atherosclerosis. *J Immigr Minor Health*. 2021; 23(4): 659-667.

Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. *Fam Med*. 2004; 36(8): 588-594.

Woloshin S, Schwartz LM, Welch HG. (Patients and medical statistics: Interest, confidence, and ability. *J Gen Intern Med*. 2005; 20(11): 996-1000.

16.07. Epworth Sleepiness Scale (Scale 0-24): epworth_sleep

Variable Name: epworth_sleep

Description: The Epworth Sleepiness scale is a part of the Sleep Questionnaire. This scale is widely used in the field of sleep medicine as a subjective measure of a patient's sleepiness. The test is a list of eight situations in which the tendency to become sleepy is rated on a scale of 0, no chance of dozing, to 3, high chance of dozing. The total score is derived by adding up the values, and ranges from 0 to 24.

Scores of 0-7 indicate unlikely to be abnormally sleepy, 8-9 indicate average amount of daytime sleepiness, 10-15 indicates potential excessive sleepiness, and 16-24 indicates excessive sleepiness and recommendations of medical attention.

Format: Numeric, 8.2 (Range 0-24)

Algorithm: If sle207a-sle207h are >=7 then they are set to missing. Note that the scale values should be on a range of 0-3, but JHS collected on a scale of 1-4. To account for this, we have subtracted 8 from the scale's score to obtain a 0-24 scale as guided.

```
nmiss_sle=cmiss(of sle207a-- sle207h);

epworth_sleep=sum(sle207a, sle207b, sle207c, sle207d, sle207e, sle207f,
sle207g, sle207h)-8;

if nmiss_sle>0 then epworth_sleep=.;
```

Source Variables: SLE2: sle207a-sle207h

Reference: Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep*. 1991;14(6):540-545.

16.08. Food Security Scale (Scale 0-6): food_sec

Variable Name: food_sec

Description: Data on food security is collected on the Food Security Questionnaire (FSQ) using the U.S. Household Food Security Survey Module, a six-item short form developed by the National Center for Health Statistics to assess household food insecurity (i.e., lacking access to food for a healthy and active life) (Blumberg et

al., 1999). Scoring is based on the number of affirmative responses to the 6 questions. Two or more affirmative responses indicates food insecurity, 5 or more affirmative responses indicates hunger.

Format: Numeric, 8.2 (Range 0-6)

Algorithm: Scoring: If any question >7= then they were set to missing. Fsq101 and fsq102 are affirmative if the participant answered often (1) or sometimes (2) true. Fsq103, fsq105, and fsq106 are affirmative if the participant responded 'yes'. Fsq104 is only asked if fsq103 was 'yes' and is considered affirmative if the participant responded almost every month (1) or some months, but not every month (2). Indicators for each question were coded and summed up to get the scale's score.

```
if fsq101=1 or fsq101=2 then fsq101_score=1;

else if fsq101=3 then fsq101_score=0;

else fsq101_score=.;

if fsq102=1 or fsq102=2 then fsq102_score=1;

else if fsq102=3 then fsq102_score=0;

else fsq102_score=.;

if fsq103=0 or fsq103=1 then fsq103_score=fsq103;

else fsq103_score=.;

if fsq104=1 or fsq104=2 then fsq104_score=1;

else if fsq104=3 or fsq103_score=0 then fsq104_score=0;

else fsq104_score=.;

if fsq105=0 or fsq105=1 then fsq105_score=fsq105;

else fsq105_score=.;

if fsq106=0 or fsq106=1 then fsq106_score=fsq106;

else fsq106_score=.;

nmiss_fsq=cmiss(of fsq101_score-- fsq106_score);

food_sec=sum(fsq101_score, fsq102_score, fsq103_score, fsq104_score,
                                    fsq105_score, fsq106_score);

if nmiss_fsq>0 then food_sec=.;
```

Source Variables: FSQ1: fsq101-fsq106

Reference: Blumberg SJ, Bialostosky K, Hamilton WL, Briefel RR. The effectiveness of a short form of the Household Food Security Scale. *Am J Public Health*. 1999;89(8):1231-1234.

16.09. QoL Physical Health T-Score SF-12v2: qol_pcs_t

Variable Name: qol_pcs_t

Description: Physical Health T-Score based on the Quality of Life SF-12v2 Health Assessment Survey. Average scores are between 45-55 with low scores indicative of substantial limitations in self-care, physical, social, and role activities; severe bodily pain; frequent tiredness; and heath rated as poor.

Format: Numeric, 6.2 (Range 0-100)

Algorithm: Scoring is proprietary. See reference for more information

Source Variables: QOL2

Reference: Ware JE, Kosinski M, Bjorner JB, Turner-Bowker DM, Gandek B, Maruish ME. (2008). sf-36v2 health survey: Administration guide for clinical trial Investigators. *Lincoln, RI: Quality Metric Incorporated*, 1-34.

Ware JE, Kosinski M, Turner-Bowker DM, Gandek B. (2002). *How to score Version 2 of the SF-12 Health Survey (with a supplement documenting Version 1).* Lincoln, RI: QualityMetric Incorporated.

Ware JE, Kosinski M, Turner-Bowker DM, Sundaram M, Gandek B, Maruish ME. (2009). *User's manual for the SF-12v2 Health Survey* (2nd ed.). Lincoln, RI: QualityMetric Incorporated.

16.10. QoL Mental Health T-Score SF-12v2: qol_mcs_t

Variable Name: qol_mcs_t

Description: Mental Health T-Score based on the Quality of Life SF-12v2 Health Assessment Survey. Average scores are between 45-55 with low scores indicative of frequent psychological distress, substantial social and role disability due to emotional problems, and health rated as poor.

Format: Numeric, 6.2 (Range 0-100)

Algorithm: Scoring is proprietary. See reference for more information

Source Variables: QOL2

Reference: Ware JE, Kosinski M, Bjorner JB, Turner-Bowker DM, Gandek B, Maruish ME. (2008). sf-36v2 health survey: Administration guide for clinical trial Investigators. *Lincoln, RI: Quality Metric Incorporated*, 1-34.

Ware JE, Kosinski M, Turner-Bowker DM, Gandek B. (2002). *How to score Version 2 of the SF-12 Health Survey (with a supplement documenting Version 1).* Lincoln, RI: QualityMetric Incorporated.

Ware JE, Kosinski M, Turner-Bowker DM, Sundaram M, Gandek B, Maruish ME. (2009). *User's manual for the SF-12v2 Health Survey* (2nd ed.). Lincoln, RI: QualityMetric Incorporated.

16.11. Anger In Score-Anger Expression Inventory (Scale 8-32): anger_in

Variable Name: anger_in

Description: Score measuring Anger-In from the Anger Expression Inventory. Higher scores are reflective of a pattern of holding anger inside, potentially leading to internal stress and physical health concerns

Format: Numeric, 8.2 (Range 8-32)

Algorithm: All scores labeled as >4 (refused or don't know) are set to missing. The Anger-In score is derived by taking the sum of items 2, 3, 4, 7, 9, 11, 12, and 14 of the Anger Expression Inventory. If any item in the scale is missing, then we cannot obtain a valid Anger-In score.

```
nmiss_ain=cmiss(aei102, aei103, aei104, aei107, aei109, aei111, aei112,
aei114);

anger_in=sum(aei102, aei103, aei104, aei107, aei109, aei111, aei112, aei114);

if nmiss_ain>0 then anger_in=.;
```

Source Variables: AEI1: aei101-aei116

Reference: Spielberger CD. The experience and expression of anger: Construction and validation of an anger expression scale. Editors: Chesney MA, Rosenman RH. Anger and Hostility in Cardiovascular and Behavioral Disorders. pgs. 5-30. Washington, DC: Hemisphere, 1985.

Note: The above citation is difficult to obtain. For interpretation of the scale, we recommend reviewing prior JHS references that have used the Anger Expression Inventory.

16.12. Anger Out Score-Anger Expression Inventory (Scale 8-32): anger_out

Variable Name: anger_out

Description: Score measuring Anger-Out from the Anger Expression Inventory. Higher scores are reflective of a tendency to express anger outwardly through yelling, arguing, or physical aggression

Format: Numeric, 8.2 (Range 8-32)

Algorithm: All scores labeled as >4 (refused or don't know) are set to missing. The Anger-Out score is derived by taking the sum of items 1, 5, 6, 8, 10, 13, 15, and 16 of the Anger Expression Inventory. If any item in the scale is missing then we cannot obtain an Anger-Out score.

nmiss_aout=cmiss(aei101, aei105, aei106, aei108, aei110, aei113, aei115, aei116);

anger_out=sum(aei101, aei105, aei106, aei108, aei110, aei113, aei115, aei116);

if nmiss_aout>**0** then anger_out=**.**;

Source Variables: AEI1: aei101-aei116

Reference: Spielberger CD. The experience and expression of anger: Construction and validation of an anger expression scale. Editors: Chesney MA, Rosenman RH. Anger and Hostility in Cardiovascular and Behavioral Disorders. pgs. 5-30. Washington, DC: Hemisphere, 1985.

Note: The above citation is difficult to obtain. For interpretation of the scale. we recommend reviewing prior JHS references that have used the Anger Expression Inventory

16.13. Anger Total Score-Anger Expression Inventory (Scale 16-64): anger_tot

Variable Name: anger_tot

Description: Score measuring the total anger from the Anger Expression Inventory. This is the sum of the Anger-In and Anger-Out subscales. Higher scores are indicative of having higher anger (in and out).

Format: Numeric, 8.2 (Range 16-64)

Algorithm: All scores labeled as >4 (refused or don't know) are set to missing. The Total Anger score is derived by taking the sum of all 16 items of the Anger Expression Inventory. If any item on the AEI scale are missing then we cannot get a valid anger score.

nmiss_atot=cmiss(of aei101--aei116);

anger_tot=sum(aei102, aei103, aei104, aei107, aei109, aei111, aei112, aei114, aei101, aei105, aei106, aei108, aei110, aei113, aei115, aei116);

if nmiss_atot>**0** then anger_tot=**.**;

Source Variables: AEI1: aei101-aei116

Reference: Spielberger CD. The experience and expression of anger: Construction and validation of an anger expression scale. Editors: Chesney MA, Rosenman RH. Anger and Hostility in Cardiovascular and Behavioral Disorders. pgs. 5-30. Washington, DC: Hemisphere, 1985.

Note: The above citation is difficult to obtain. For interpretation of the scale, we recommend reviewing prior JHS references that have used the Anger Expression Inventory.

17. Nutrition

17.01. DASH Diet Score: dash_score

Variable Name: dash_score

Description: The Dietary Approaches to Stop Hypertension (DASH) diet score is a measure of diet quality that indicates how well someone adheres to the DASH diet. The DASH diet is based on a clinical trial that found that eating more fruits, vegetables, and low-fat dairy, and less saturated fat, cholesterol, and total fat, can reduce blood pressure.

Note: We use Fung's DASH Index

Format: Numeric, 8.2 (Range 8-40)

Algorithm: Derived internally by the FFQ reading center

Source Variables: dash_clean: dash_4th

Reference: Appel LJ, Moore TJ, Obarzanek E, Vollmer WM, Svetkey LP, Sacks FM, Bray GA, Vogt TM, Cutler JA, Windhauser MM, Lin PH, and Karanja N. A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group. *N Engl J Med*. 1997; 336(16): 1117-1124.

Fung, TT, Chiuve, SE, McCullough, ML, Rexrode, KM, Logroscino, G, and Hu, FB. Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. *Archives of internal medicine*. 2008; 168(7), 713-720.

17.02. Mediterranean Eating Pattern for Americans score: mepa_score

Variable Name: mepa_score

Description: The Mediterranean Eating Pattern for Americans (MEPA) screener is a tool that assesses how well a person's diet adheres to the Mediterranean diet.

Format: Numeric, 8.2 (Range 0-16)

Algorithm: Derived internally by the FFQ reading center

Source Variables: mepa_clean: mepa

Reference: Cerwinske LA, Rasmussen HE, Lipson S, Volgman AS, Tangney CC. Evaluation of a dietary screener: the Mediterranean Eating Pattern for Americans tool. *J Hum Nutr Diet. 2017;* 30(5): 596-603.

17.03. Healthy Eating Index Score: hei_score

Variable Name: hei_score

Description: Score for the Healthy Eating Index. The Healthy Eating Index (HEI) is a measure of diet quality used to assess how well a set of foods aligns with key recommendations and dietary patterns

Format: Numeric, 8.2 (Range 0-100)

Algorithm: Derived internally by the FFQ reading center

Source Variables: hei_clean: hei2010_4th

Reference: Guenther, Patricia M., et al. "Update of the healthy eating index: HEI-2010." *Journal of the Academy of Nutrition and Dietetics* 113.4 2013: 569-580.

17.04. Fruits and Vegetable Indicator for LS7: fruitveg

Variable Name: fruitveg

Description: An indicator for if the participant consumed over 4.5 cups for fruits/vegetables a day. This is used for Life's Simple 7.

Format: 1='Yes'

0='No'

Algorithm: Derived internally by the FFQ reading center

Source Variables: lifesimple7_clean: fruit_vegs_score

17.05. Fish Indicator for LS7: fish

Variable Name: fish

Description: An indicator for if the participant consumed over 3.5 ounces of fish twice a week. This is used for Life's Simple 7.

Format: 1='Yes'

0='No'

Algorithm: Derived internally by the FFQ reading center

Source Variables: lifesimple7_clean: fish_score

17.06. Sodium Indicator for LS7: sodium

Variable Name: sodium

Description: An indicator for if the participant consumed under 1500mg of sodium a day on average. This is used for Life's Simple 7.

Format: 1='Yes'

0='No'

Algorithm: Derived internally by the FFQ reading center

Source Variables: lifesimple7_clean: sodium_score

17.07. Sugary Beverages Indicator for LS7: sugbeverage

Variable Name: sugbeverage

Description: An indicator for if the participant consumed under 450 kcal of sugary beverages a week on average. This is used for Life's Simple 7.

Format: 1='Yes'

0='No'

Algorithm: Derived internally by the FFQ reading center

Source Variables: lifesimple7_clean: ssb_score

17.08. Whole Grain Indicator for LS7: whlgrain

Variable Name: whlgrain

Description: An indicator for if the participant consumed over 3 servings of whole grains a day. This is used for Life's Simple 7.

Format: 1='Yes'

0='No'

Algorithm: Derived internally by the FFQ reading center

Source Variables: lifesimple7_clean: wg_score

18. Physical Activity

18.01. Sports Index: pa_sportindex

Variable Name: pa_sportindex

Description: Physical Activity Index that rates physical activities from sports and formal exercise.

Format: Numeric, 8.2 (Range 1-5)

Algorithm: General Equation: =(i4+i19+SSS)/3

=1 if pac118=0

i4=pac104;

i19=pac119;

Steps to Calculate SSS (Simple Sports Score):

1). Recoding of Sport and Exercise Variables (PAC120-PAC131)

a). pac120, pac124, and pac128 are recoded using a list of sports and value of low, moderate and vigorous for each derived from the JHS METs score, Compendium of Physical Activities, and CDC/ACSM sport intensity list. Intensity codes are recoded as follows:

0=N/A

0.76='Low'

1.26='Moderate'

1.76='Vigorous'

Below is a table with list of sports codes associated with each level.

Physical Activity Level CDC/ACSM	Code (PAC120, PAC124, PAC128)
Low	15, 47, 50, 57, 58, 72, 73, 74
Moderate	3, 10, 12, 13, 20, 21, 22, 24, 25, 26, 27, 28, 32, 33, 34, 35, 36, 37, 38, 41, 42, 43, 44, 46, 48, 49, 51, 53, 54, 56, 59, 60, 66, 68, 69, 71, 75, 76, 77
Vigorous	1, 2, 4, 5, 6, 7, 8, 9, 11, 14, 16, 17, 18, 19, 23, 29, 30, 31, 39, 40, 45, 52, 55, 61, 62, 63, 64, 65, 67, 70, 78, 79. 80
Other-Adjudicated	499 • if activity in ("chair aerobics", "yoga", "shooting pool", "stretch bands", "leg lifts", "other activities") then intensity="Low"; • if activity in ("isometrics","knee bends, arm lifts, toe to heel exercises","physical therapy exercises","qigong","tv exercise", "dancing","strength training","foot pedal","new step seated elliptical","hulu hooping","rowing") then intensity="Mod"; • if activity in ("pickle ball","pickleball", "jumping on trampoline","exercise","line dancing","aerobics","light aerobic","jumping rope", "trampolining","squats") then intensity="Vig";

Note that participants who select other will have their selection adjudicated by the JHS coordinating center and the Diet and Physical Activity Working Group.

Repeat for 3 sports to get intens1, intens2, intens3

b). Recode time per week values (f1=pac122, f2=pac126, f3=pac130)

0.5=Less than 1 month (1)

1.5=1 to 3 months (2)

2.5=4 to 6 months (3)

3.5=7 to 9 months (4)

4.5=More than 9 months (5)

c). Recode time per week values (d1=pac121, d2=pac125, d3=pac129)

0.04=Less than 1 hour (1)

0.17=At least 1 but less than 2 hours

0.42=At least 2 but less than 3 hours

0.67=At least 3 but less than 4 hours

0.92=4 hours or more

2). Calculate the Simple Sports Score

a). Multiply recoded intensity, time, and proportion of year activity performed

ss1=intens1*f1*d1

ss2=intens2*f2*d2

ss3=intens3*f3*d3

ss1,ss2,ss3=0 if pac118=0 or intensity, f, or d are missing

b). Sum products and assign Simple Sports Score (SSS)

sss_sum=ss1+ss2+ss3

	Simple Sport Score
SSS_sum = 0	SSS=1
SSS_sum > 0.01 but < 4	SSS=2
SSS_sum ≥ 4 but < 8	SSS=3
SSS_sum ≥ 8 but < 12	SSS=4
SSS_sum > 12	SSS=5

*More clear instructions for the physical activity variables can be found in the JHS JPAC scoring guide (Appendix 18A). Further appendices with documentation for METs coding and labeling of low, moderate, and vigorous activity are available upon request.

*Specific SAS coding is available upon request. It was not included here due to length.

Source Variables: PAC1: pac104, pac118-pac131

Reference: Smitherman TA, Dubbert PM, Grothe KB, Sung JH, Kendzor DE, Reis JP, Ainsworth BE, Newton Jr RL, Lesniak KT, Taylor Jr, HA. Validation of the Jackson Heart Study physical activity survey in African Americans. *J Phys Act Health*. 2009; 6(1): S124-132.

18.02. Home/Yard Index: pa_hyindex

Variable Name: pa_hyindex

Description: Physical activity index that rates physical activities in the home that involve cleaning, caretaking and home repairs

Format: Numeric, 8.2 (Range 1-5)

Algorithm: if pac111 = **1** then i12=**1**;

else if pac111 = **2** then i12=**3**;

else if pac111 = **3** then i12=**5**;

i13=pac112; i14=pac113; i15=pac114; i16=pac115; i17=pac116; i18=pac117;

pa_hyindex=(i12+i13+i14+i15+i16+i17+i18)/7

=missing if any of above missing

Source Variables: PAC1: pac111, pac112, pac113, pac114, pac115, pac116, pac117

Reference: Smitherman TA, Dubbert PM, Grothe KB, Sung JH, Kendzor DE, Reis JP, Ainsworth BE, Newton Jr RL, Lesniak KT, Taylor Jr, HA. Validation of the Jackson Heart Study physical activity survey in African Americans. *J Phys Act Health*. 2009; 6(1): S124-132.

18.03. Active Living Index: pa_activeindex

Variable Name: pa_activeindex

Description: Physical activity during leisure time excluding sports.

Format: Numeric, 8.2 (Range 1-5)

Algorithm: i1=pac101; i2=pac102; i3=pac103; tlv=pac105

pa_activeindex=(i1+i2+i3+ [6-tlv])/4

= missing if one of the above I scores is missing

Source Variables: PAC1: pac101, pac102, pac103, pac105

Reference: Smitherman TA, Dubbert PM, Grothe KB, Sung JH, Kendzor DE, Reis JP, Ainsworth BE, Newton Jr RL, Lesniak KT, Taylor Jr, HA. Validation of the Jackson Heart Study physical activity survey in African Americans. *J Phys Act Health*. 2009; 6(1): S124-132.

18.04. Total Physical Activity Score: pa_score

Variable Name: pa_score

Description: Total physical activity score (Sum of sports, home/yard, and active living index). Details for all the physical activity scores are documented in Appendix 19A

Format: Numeric, 8.2 (Range 3-15)

Algorithm: pa_score=pa_sportindex+pa_hyindex+pa_activeindex

Source Variables: Dervar4: pa_sportindex, pa_hyindex, pa_activeindex

18.05. Total moderate/vigorous physical activity time: mvlpa

Variable Name: mvlpa

Description: Total moderate/vigourous physical activity time (min/week)

Format: Numeric

Algorithm: Coding for MVLPA is lengthy and complicated so we do not provide directly in the data dictionary. Code is available upon request.

Summary: As was the case with variable 18.01, the Diet and Physical Activity Working Group went through JHS sports and adjudicated each sport with a METs score via the 2024 Compendium of Physical Activities. Each sport was assigned to be either low activity (METs score <3), moderate activity (METS score 3-5.9), and vigorous activity (METS score 6+).

We then multiply the frequency of the activity by the duration to get minutes of physical activity. We then separate those minutes into categories by intensity with subvaraibles labels lightPA, moderatePA, and vigorousPA.

Mvlpa=moderatePA+vigorousPA

Source Variables: PAC1: pac118-pac130

19. Genetics

Notes for all genetic variables: There is also a csv file, JHS_selected_variant_details.csv, with rsid, related phenotype, chromosome, position, originating freeze, ref/alt alleles, and allele frequency for the alt allele. Original genotype data is available upon request.

19.01. Sickle Cell Disease/Trait-rs334: sicklecell

Variable Name: sicklecell

Description: Binary indicator for sickle cell trait/disease (rs334). Note: no JHS participant has sickle cell disease.

Format: 0=No risk allele

1=Any risk allele (Sickle cell trait, AT, or Sickle Cell Disease, AA)

Algorithm: R Code:
```
raw_genotypes <- raw_genotypes %>%
  mutate(sicklecell = case_when(
    rs334_T == 2 ~ 0,
    rs334_T == 1 ~ 1,
    rs334_T == 0 ~ 1,
    TRUE ~ as.numeric(NA)))
```

SAS Code:
```
if rs334_T=2 then sicklecell=0;
Else if rs334_T=1 or rs334_T=0 then sicklecell=1;
Else sicklecell=.;
```

Source Variables: genetics (raw_genotype): rs334_T

Notes: rs334 samples were dropped if they were homozygous for the risk allele but with phenotypic data showing no obvious treatment for or complications from sickle cell disease. We suspect these are sequencing errors.

19.02. APOL1 G1 Risk Allele-rs60910145: apol1g1a

Variable Name: apol1g1a

Description: APOL1 G1 risk from allele rs60910145.

Format: 0=no risk allele, TT

1=one risk allele, GT

2=two risk alleles, GG

Algorithm: R Code:
```
raw_genotypes <- raw_genotypes %>%
  mutate(apol1g1a = case_when(
    rs60910145_T == 2 ~ 0,
    rs60910145_T == 1 ~ 1,
    rs60910145_T == 0 ~ 2,
    TRUE ~ as.numeric(NA)))
```

SAS Code: if rs60910145_T =2 then apol1g1a=0;

Else if rs60910145_T=1 then apol1g1a=1;

Else if rs60910145_T=0 then apol1g1a=2;

Else apol1g1a=.;

Source Variables: genetics (raw genotype): rs60910145_T

19.03. APOL1 G1 Risk Allele- rs73885319: apol1g1b

Variable Name: apol1g1b

Description: APOL1 G1 risk from allele rs73885319.

Format: 0=no risk allele, AA

1=one risk allele, AG

2=two risk alleles, GG

Algorithm: R Code: raw_genotypes <- raw_genotypes %>%
mutate(apol1g1b = case_when(
rs73885319_A == 2 ~ 0,
rs73885319_A == 1 ~ 1,
rs73885319_A == 0 ~ 2,
TRUE ~ as.numeric(NA)))

SAS Code: SAS Code: if rs73885319_A=2 then apol1g1b=0;

Else if rs73885319_A=1 then apol1g1b=1;

Else if rs73885319_A =0 then apol1g1b=2;

Else apol1g1b=.;

Source Variables: genetics (raw genotype): rs73885319_A

19.04. APOL1 G2 Risk Allele- rs71785313: apol1g2

Variable Name: apol1g2

Description: APOL 1 G2 risk from allele rs71785313. Note: (rs12106505 is in prior literature also sometimes used as a proxy for this deletion variant- we here use the deletion to define G2 status)

Format: 0=no risk allele, RR

1=one risk allele, RD

2=two risk alleles, DD

Algorithm: R Code: raw_genotypes <- raw_genotypes %>%
mutate(apol1g2 = case_when(
rs71785313_AATAATT == 2 ~ 0,

rs71785313_AATAATT == 1 ~ 1,
rs71785313_AATAATT == 0 ~ 2,
TRUE ~ as.numeric(NA)))

SAS Code: SAS Code: SAS Code: if rs71785313_AATAATT =2 then apol1g2=0;

Else if rs71785313_AATAATT=1 then apol1g2=1;

Else if rs71785313_AATAATT =0 then apol1g2=2;

Else apol1g2=.;

Source Variables: genetics (raw_genotype): rs71785313_AATAATT

19.05. APOL1 CVD Risk Genotype: apolrisk

Variable Name: apolrisk

Description: APOL1 CVD Risk Genotype determined by APOL1 G1 and APOL1G2 alleles. Risk is increased when any of the SNPs are homozygous for the risk allele or when there's at least one risk allele in both g1 and g2. Two variants were included for apol1g1 as these variants are in high, but not perfect, linkage disequilibrium in the JHS cohort. This lack of perfect LD has been observed in other cohorts (PMID: 30586505 for example).

Format: 0 =low risk (as specified in description)

1 = increased risk (as specified in description)

Algorithm: R Code: raw_genotypes <- raw_genotypes %>%
mutate(apolrisk = case_when(
apol1g1a == 2 | apol1g1b == 2 | apol1g2 == 2 ~ 1,
apol1g2 == 0 ~ 0,
apol1g1a == 0 & apol1g1b == 0 & apol1g2 == 1 ~ 0,
apol1g2 == 1 ~ 1,
TRUE ~ as.numeric(NA)))

SAS Code: if apol1g1a=2 or apol1g1b=2 or apol1g2=2 then apolrisk=1;

Else if apol1g2=0 then apolrisk=0;

Else if apol1g1a=0 and apol1g1b=0 and apol1g2=1 then apolrisk=0;

Else if apol1g2=1 then apolrisk=1;

Else apolrisk=.;

Source Variables: genetics (raw genotype): apol1g1a, apol1g1b, apol1g2

Reference: Nadkarni G N, Gignoux CR, Sorokin EP, Daya M, Rahman R, Barnes KC et al. Worldwide frequencies of APOL1 renal risk variants. *N Engl J Med.* 2018; 379(26): 2571-2572. doi:10.1056/NEJMc1800748

19.06. APOE2 haplotype- rs7412: apoe2

Variable Name: apoe2

Description: Number of copies of rs7412 alt allele (T) corresponding to number of copies of APOE2 haplotype (decreased Alzheimer's disease risk)

Format: 0=no apoe2 allele, CC

1=one apoe2 allele, CT

2=two apoe2 alleles, TT

Algorithm: R Code: raw_genotypes <- raw_genotypes %>%
mutate(apoe2 = case_when(
rs7412_C == 2 ~ 0,
rs7412_C == 1 ~ 1,
rs7412_C == 0 ~ 2,
TRUE ~ as.numeric(NA)))

SAS Code: if rs7412_C =2 then apoe2=0;

Else if rs7412_C=1 then apoe2=1;

Else if rs7412_C =0 then apoe2=2;

Else apoe2=.;

Source Variables: genetics (raw genotype): rs7412_C

19.07. APOE4 haplotype- rs429358: apoe4

Variable Name: apoe4

Description: Number of copies of rs429358 alt allele (C) corresponding to number of copies of APOE4 haplotype (increased Alzheimer's disease risk).

Format: 0=no risk allele, TT

1=one risk allele, CT

2=two risk alleles, CC

Algorithm: R Code: raw_genotypes <- raw_genotypes %>%
mutate(apoe4 = case_when(
rs429358_T == 2 ~ 0,
rs429358_T == 1 ~ 1,
rs429358_T == 0 ~ 2,
TRUE ~ as.numeric(NA)))

SAS Code: SAS Code: SAS Code: if rs429358_T =2 then apoe4=0;

Else if rs429358_T =1 then apoe4=1;

Else if rs429358_T =0 then apoe4=2;

Else apoe4=.;

Source Variables: genetics (raw genotype): rs429358_T

19.08. Duffy blood group antigen- rs2814778: duffy

Variable Name: duffy

Description: Duffy blood group antigen (rs2814778)

Format: 0=TT, duffy positive

1=CT, duffy positive

2=CC, duffy negative/null

Algorithm: R Code: raw_genotypes <- raw_genotypes %>%
mutate(duffy = case_when(
rs2814778_T == 2 ~ 0,
rs2814778_T == 1 ~ 1,
rs2814778_T == 0 ~ 2,
TRUE ~ as.numeric(NA)))

SAS Code: if rs2814778_T =2 then duffy=0;

Else if rs2814778_T =1 then duffy=1;

Else if rs2814778_T =0 then duffy=2;

Else duffy=.;

Source Variables: genetics (raw genotype): rs2814778_T

Notes: From JHS Genetics Working Group: Duffy status is likely best modelled under recessive model by making binary variable based on duffy positive (TT and CT) and duffy null (CC).

19.09. PCSK9-C679X Low density lipoprotein cholesterol level quantitative trait locus 1-rs28362286: pcsk9

Variable Name: pcsk9

Description: PCSK9-C679X Low density lipoprotein cholesterol level quantitative trait locus 1 (rs28362286)

Format: 0=CC-common

1=AC-lowered LDL-C and lower risk for coronary events

2=AA-lowered LDL-C and lower risk for coronary events

Algorithm: R Code: raw_genotypes <- raw_genotypes %>%
mutate(pcsk9 = case_when(
rs28362286_C == 2 ~ 0,
rs28362286_C == 1 ~ 1,

rs28362286_C == 0 ~ 2,
TRUE ~ as.numeric(NA)))

SAS Code: if rs28362286_C =2 then pcsk9=0;

Else if rs28362286_C =1 then pcsk9=1;

Else if rs28362286_C =0 then pcsk9=2;

Else pcsk9=.;

Source Variables: genetics (raw genotype): rs28362286_C

19.10. Serine Tyrosine Substitution at AA1103- rs7626962: scn5a_s1103y

Variable Name: scn5a_s1103y

Description: Serine Tyrosine Substitution at AA1103- rs7626962

Format: 0=GG-common

1=GT-increased susceptibility to long QT syndrome

2=TT-increases susceptibility to long QT syndrome

Algorithm: R Code: raw_genotypes <- raw_genotypes %>%
mutate(scn5a_s1103y = case_when(
rs7626962_G == 2 ~ 0,
rs7626962_G == 1 ~ 1,
rs7626962_G == 0 ~ 2,
TRUE ~ as.numeric(NA)))

SAS Code: if rs7626962_G =2 then scn5a_s1103y =0;

Else if rs7626962_G =1 then scn5a_s1103y =1;

Else if rs7626962_G =0 then scn5a_s1103y =2;

Else scn5a_s1103y =.;

Source Variables: genetics (raw genotype): rs7626962_G

19.11. Hemoglobin C (HbC) locus-rs33930165: hbc

Variable Name: hbc

Description: Hemoglobin C (HbC) locus (rs33930165).

Format: 0=CC-common

1=AC, reduction in malarial risk

2-AA, larger reduction in malarial risk

Algorithm: R Code: raw_genotypes <- raw_genotypes %>%
mutate(hbc = case_when(

rs33930165_C == 2 ~ 0,
rs33930165_C == 1 ~ 1,
rs33930165_C == 0 ~ 2,
TRUE ~ as.numeric(NA)))

SAS Code: SAS Code: if rs33930165_C =2 then hbc =0;

Else if rs33930165_C =1 then hbc =1;

Else if rs33930165_C =0 then hbc =2;

Else hbc =.;

Source Variables: genetics (raw genotype): rs33930165_C

19.12. Derived Ancestry Principal Component 1 (PC1): pc1_avg

Variable Name: pc1_avg

Description: PCair (derived ancestry principal components) should be considered for adjustment in genetic analyses (particularly PC1). These were derived using frequency and LD pruned variants from freeze 8, but should capture ancestry variability relevant to all genetic analyses.

Based on scree plot, PC1 is by far the most influential PC. There are 9 others available (upon request), but PC1 is included in the derived variables dataset.

Format: Numeric

Algorithm: Derived by JHS Genetics working group

Source Variables: genetics (raw genotype): pc1_avg

20. Risk Scores

Note that for all risk scores, participants who have self-reported CVD events at Exam 4 will have missing risk scores. These self-reported CVD events include heart failure (hf_sr_hx), stroke, and myocardial infarction (stroke_mi).

20.01. Framingham 10-year Risk Score for coronary heart disease (CHD): frs_chd10yr

Variable Name: frs_chd10yr

Description: A score that predicts 10-year risk of angina, MI, coronary insufficiency, and death from CHD.

Format: 6. *if 0.56-1.0, then labeled as ">=0.56"

(0.11 would reflect 11% risk)

Algorithm: The equations use sex, age, total cholesterol, hdl-c, systolic blood pressure, current smoking status and diabetes status.

See Appendix 20A for code.

Source Variables: Dervar4: sex, age, totchol, hdl, sbp, smoker_current, diabetes

Reference: Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation.* 1998; 97(18), 1837-1847.

20.02. Framingham 10-year Risk Score for cardiovascular disease (CVD): frs_cvd10yr

Variable Name: frs_cvd10yr

Description: A score that predicts 10-year risk of angina, MI, stroke, TIA, peripheral vascular disease, heart failure, and death from CHD.

Format: Numeric, 6. *if 0.00-0.01 then label as "<=0.01" *if 0.30-1.00 then label as ">=0.30"

(0.11 would reflect 11% risk)

Algorithm: The equations use sex, age, total cholesterol, hdl-c, systolic blood pressure, current smoking status, blood pressure medication use, and diabetes status.

See Appendix 20A for code.

Source Variables: Dervar4: sex, age, totchol, hdl, sbp, smoker_current, diabetes, bp_meds_sr

Reference: D'Agostino Sr RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation*. 2008; 117(6): 743-753.

20.03. Framingham Adult Treatment Panel 10-year Risk Score for coronary heart disease (CHD): frs_atpiii_chd10yr

Variable Name: frs_atpiii_chd10yr

Description: A score that predicts 10-year risk of MI, and death from CHD.

Format: 6. *if 0.00- 0.01 then label as "<=0.01" *if 0.30-1.00 then label as ">=0.30"

(0.11 would reflect 11% risk)

Algorithm: The equations use sex, age, total cholesterol, hdl-c, systolic blood pressure, current smoking status, blood pressure medication use.

See Appendix 20A for code.

Source Variables: Dervar4: sex, age, totchol, hdl, sbp, smoker_current, meds_bp_sr

Reference: D'Agostino RB, Grundy S, Sullivan LM, Wilson P, CHD Risk Prediction Group. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA.* 2001; 286(2): 180-187.

20.04. ACC/AHA Pooled Cohort Equation 10-year Risk Score for Atherosclerotic Cardiovascular Disease (ASVCD): accaha_pce_ascvd10yr

Variable Name: accaha_pce_ascvd10yr

Description: A score that predicts 10-year risk of MI, death from CHD, and stroke.

Format: Numeric, 6.2

(0.11 would reflect 11% risk)

Algorithm: The equations use sex, age, total cholesterol, hdl-c, systolic blood pressure, current smoking status, blood pressure medication use, and diabetes status.

See Appendix 20A for code

Source Variables: Dervar4: sex, age, totchol, hdl, sbp, smoker_current, diabetes, bp_meds_sr

Reference: Goff Jr DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, 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-S73.

20.05. PREVENT™ Equation 10-year Risk Score for Atherosclerotic Cardiovascular Disease (ASCVD): prevent_b_ascvd10yr

Variable Name: prevent_b_ascvd10yr

Description: A score that predicts 10-year risk of MI, death from CHD, and stroke using the PREVENT Equations (using only base variables). Low risk (<5%). Borderline risk (5% to 7.4%), Intermediate risk (7.5% to 19.9%), High risk (≥20%).

Format: Numeric, 6.2

(0.11 would reflect 11% risk)

Algorithm: Code is based on equations mentioned in the reference paper. To derive variables an R package created by the PREVENT others was used from their

github. Special permission is needed to access the github. Base variables needed for the equations include sex, age (30-79), total cholesterol, hdl-c, systolic blood pressure, current smoking status, diabetes status, use of statins, use of blood pressure medications, egfr (CKD-EPI 2021), and BMI.

Source Variables: Dervar4: sex, age, totchol, hdl, sbp, smoker_current, diabetes, bp_meds_sr, egfr_ckdepi2021_creat, meds_statin, bmi

Reference: Khan SS, Matsushita K, Sang Y, Ballew SH, Grams ME, et al. Development and validation of the American Heart Association's PREVENT equations. *Circulation.* 2024; 149(6): 430-449.

Khan SS, Coresh J, Pencina MJ, Ndumele CE, Rangaswami J, et al. Novel prediction equations for absolute risk assessment of total cardiovascular disease incorporating cardiovascular-kidney-metabolic health: a scientific statement from the American Heart Association. *Circulation.* 2024; 148(24): 1982-2004.

20.06. PREVENT™ Equation 10-year Risk Score for cardiovascular disease (CVD): prevent_b_cvd10yr

Variable Name: prevent_b_cvd10yr

Description: A score that predicts 10-year risk of Angina, MI, stroke, TIA, peripheral vascular disease, heart failure, and death from CHD using the PREVENT Equations (using only base variables). Low risk (<5%), Borderline risk (5% to 7.4%), Intermediate risk (7.5% to 19.9%), and High risk (≥20%).

Format: Numeric, 6.2

(0.11 would reflect 11% risk)

Algorithm: Code is based on equations mentioned in the reference paper. To derive variables an R package created by the PREVENT others was used from their github. Special permission is needed to access the github. Base variables needed for the equations include sex, age (30-79), total cholesterol, hdl-c, systolic blood pressure, current smoking status, diabetes status, use of statins, use of blood pressure medications, egfr (CKD-EPI 2021), and BMI.

Source Variables: Dervar4: sex, age, totchol, hdl, sbp, smoker_current, diabetes, bp_meds_sr, egfr_ckdepi2021_creat, meds_statin, bmi

Reference: Khan SS, Matsushita K, Sang Y, Ballew SH, Grams ME, et al. Development and validation of the American Heart Association's PREVENT equations. *Circulation.* 2024; 149(6): 430-449.

Khan SS, Coresh J, Pencina MJ, Ndumele CE, Rangaswami J, et al. Novel prediction equations for absolute risk assessment of total cardiovascular disease incorporating cardiovascular-kidney-metabolic health: a scientific statement from the American Heart Association. *Circulation.* 2024; 148(24): 1982-2004.

20.07. PREVENT™ Equation 10-year Risk Score for Heart Failure: prevent_b_hf10yr

Variable Name: prevent_b_hf10yr

Description:	A score that predicts 10-year risk of heart failure using the PREVENT Equations (using only base variables). Low risk (<5%), Borderline risk (5% to 7.4%), Intermediate risk (7.5% to 19.9%), and High risk (≥20%).
Format:	Numeric, 6.2 (0.11 would reflect 11% risk)
Algorithm:	Code is based on equations mentioned in the reference paper. To derive variables an R package created by the PREVENT others was used from their github. Special permission is needed to access the github. Base variables needed for the equations include sex, age (30-79), total cholesterol, hdl-c, systolic blood pressure, current smoking status, diabetes status, use of statins, use of blood pressure medications, egfr (CKD-EPI 2021), and BMI.
Source Variables:	Dervar4: sex, age, totchol, hdl, sbp, smoker_current, diabetes, bp_meds_sr, egfr_ckdepi2021_creat, meds_statin, bmi
Reference:	Khan SS, Matsushita K, Sang Y, Ballew SH, Grams ME, et al. Development and validation of the American Heart Association's PREVENT equations. *Circulation.* 2024; 149(6): 430-449. Khan SS, Coresh J, Pencina MJ, Ndumele CE, Rangaswami J, et al. Novel prediction equations for absolute risk assessment of total cardiovascular disease incorporating cardiovascular-kidney-metabolic health: a scientific statement from the American Heart Association. *Circulation.* 2024; 148(24): 1982-2004.

21. Life's Simple Seven (LSS)

LS7 Reference: Lloyd-Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, et al. 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; 121: 586-613.

21.01. LS7 AHA Smoking Categorization: ls7_smk3cat

Variable Name: ls7_smk3cat

Description: Poor Health is defined as: (1) Current Smoker

Intermediate Health is defined as: (1) Quit < 12 months ago

Ideal Health is defined as: (1) Never smoked; or (2) Quit ≥ 12 months ago

Note: JHS Exam 4 does not ask how long it's been since the participant quit smoking. Our selection includes 1, Never used; 2, Rarely used in past, but not in past 30 days; 3, Regularly used in past, but not in past 30 days; 4, Used in past 30 days, less than weekly, but at least once per month; 5, Used in past 30 days, less than daily, but at least once a week; 6, Used daily or almost daily in past 30 days

If participant selected 1 or 2 we classified them as 'Ideal Health', if participant selected 3 we classified them as intermediate health. If participant selected 4, 5, or 6 we classified them as 'Poor Health'.

Format: 0="Poor Health"

1="Intermediate Health"

2="Ideal Health"

Algorithm: After removing participants who refused to answer smoking questions we took the max of the 6 options asked above for cigarettes, small cigars, large cigars, pipe, or hookah.

smoke=max(tob103a, tob103b, tob103c, tob103d, tob103e);

if smoke in (**1**,**2**) then ls7_smk3cat=**2**; *2 includes those who rarely used in past;

if smoke=**3** then ls7_smk3cat=**1**; *3 is those who regularly used in past;

if smoke in (**4**,**5**,**6**) then ls7_smk3cat=**0**;

Source Variables: TOB1: tob103a-tob103e

21.02. LS7 Indicator for Ideal Health via Smoking Status: ls7_smk_ideal

Variable Name: ls7_smk_ideal

Description: Ideal health indicator based on AHA smoking status classification.

Format: 0=”No”

1=”Yes”

Algorithm: if ls7_smk3cat in (**0**,**1**) then ls7_smk_ideal=**0**;

else if ls7_smk3cat=**2** then ls7_smk_ideal=**1**;

else ls7_smk_ideal=**.**;

Source Variables: Dervar4: ls7_smk3cat

21.03. LS7 AHA BMI Categorization: ls7_bmi3cat

Variable Name: ls7_bmi3cat

Description: Poor Health: BMI ≥ 30 (Obese)

Intermediate Health: 30 > BMI ≥ 25(Overweight)

Ideal Health: 25>BMI (Normal)

Format: 0=”Poor Health”

1=”Intermediate Health”

2=”Ideal Health”

Algorithm: if **0**<bmi<**25** then ls7_bmi3cat=**2**;

else if **25**<=bmi<**30** then ls7_bmi3cat=**1**;

else if **30**<=bmi then ls7_bmi3cat=**0**;

else ls7_bmi3cat=**.**;

Source Variables: Dervar4: bmi

21.04. LS7 Indicator for Ideal Health via BMI: ls7_bmi_ideal

Variable Name: ls7_bmi_ideal

Description: Ideal health indicator based on AHA BMI classification.

Format: 0=”No”

1=”Yes”

Algorithm: if ls7_bmi3cat in (**0**,**1**) then ls7_bmi_ideal=**0**;

else if ls7_bmi3cat=**2** then ls7_bmi_ideal=**1**;

else ls7_bmi_ideal=**.**;

Source Variables: Dervar4: ls7_bmi3cat

21.05. LS7 AHA Physical Activity Categorization: ls7_pa3cat

Variable Name: ls7_pa3cat

Description: Poor Health is defined as: (1) 0 mins of moderate physical activity; and (2) 0 mins of vigorous physical activity

Intermediate Health: (1) 0 < mins of moderate physical activity < 150; or (2) 0 < mins of vigorous physical activity < 75; or (3) 0 < mins of combined moderate and vigorous physical activity < 150

Ideal Health: (1) mins of moderate physical activity ≥ 150; or (2) mins of vigorous physical activity ≥ 75; or (3) mins of combined moderate and vigorous physical activity ≥ 150

Format: 0="Poor Health"

1="Intermediate Health"

2="Ideal Health"

Algorithm: if moderatePA >= **150** or vigorousPA >= **75** or mvlpa >= **150** then ls7_pa3cat=**2**;

else if **0**<moderatepa<**150** or **0**<vigorouspa<**75** or **0**<mvlpa<**150** then ls7_pa3cat=**1**;

else if mvlpa=**0** then ls7_pa3cat=**0**;

else ls7_pa3cat=**.**;

Source Variables: Dervar4: mvlpa

Supplement: moderapa, vigorouspa

21.06. LS7 Indicator for Ideal Health via Physical Activity: ls7_pa_ideal

Variable Name: ls7_pa_ideal

Description: Ideal health indicator based on AHA Physical Activity classification.

Format: 0="No"

1="Yes"

Algorithm: if ls7_pa3cat in (**0**,**1**) then ls7_pa_ideal=**0**;

else if ls7_pa3cat=**2** then ls7_pa_ideal=**1**;

else ls7_pa_ideal=**.**;

Source Variables: Dervar4: ls7_pa3cat

21.07. LS7 AHA Nutrition Categorization: ls7_diet3cat

Variable Name: ls7_diet3cat

Description: Components (based on 2000-kcal diet):

- Fruits and vegetables: ≥ 4.5 cups/day
- Fish: > 3.5 ounces, twice per week

- Sodium: < 1500 mg/day

- Sugary beverages: < 450 kcal/wk

- Whole grains: ≥ 3 servings/day

Poor Health: 0-1 components

Intermediate Health: 2-3 components

Ideal Health: 4-5 components

Format: 0=”Poor Health”

1=”Intermediate Health”

2=”Ideal Health”

Algorithm: *Derived directly by the FFQ reading center

Source Variables: lifesimple7_clean: ls7_score

21.08. LS7 Indicator for Ideal Health via Nutrition: ls7_diet_ideal

Variable Name: ls7_diet_ideal

Description: Ideal health indicator based on AHA Nutrition classification.

Format: 0=”No”

1=”Yes”

Algorithm: if ls7_diet3cat in (**0**,**1**) then ls7_diet_ideal=**0**;

else if ls7_diet3cat=**2** then ls7_diet_ideal=**1**;

else ls7_diet_ideal=**.**;

Source Variables: Dervar4: ls7_diet3cat

21.09. LS7 AHA Cholesterol Categorization: ls7_chol3cat

Variable Name: ls7_chol3cat

Description: Poor Health (High) is defined as: (1) Total Chol. ≥ 240 mg/dL

Intermediate Health (Borderline High) is defined as: (1) 240> Total Chol ≥ 200; or (2) 200> Total Chol (if treated)

Ideal Health (Desirable/Normal) is defined as: (1) 200> Total Chol (if untreated)

Format: 0=”Poor Health”

1=”Intermediate Health”

2=”Ideal Health”

Algorithm: if totchol>=**240** or totchol=**.H** then ls7_chol3cat=**0**;

else if **240**>totchol>=**200** or (meds_cholesterol=**1** and (**200**>totchol>=**0** or totchol=**.L**)) then ls7_chol3cat=**1**;

else if **200**>totchol>=**0** and meds_cholesterol=**0** then ls7_chol3cat=**2**;

else ls7_chol3cat=**.**;

Source Variables: Dervar4: totchol, meds_cholesterol

21.10. LS7 Indicator for Ideal Health via Blood Cholesterol: ls7_chol_ideal

Variable Name: ls7_chol_ideal

Description: Ideal health indicator based on AHA Cholesterol classification.

Format: 0=”No”

1=”Yes”

Algorithm: if ls7_chol3cat in (**0**,**1**) then ls7_chol_ideal=**0**;

else if ls7_chol3cat=**2** then ls7_chol_ideal=**1**;

else ls7_chol_ideal=**.**;

Source Variables: Dervar4: ls7_chol3cat

21.11. LS7 AHA Blood Pressure Categorization: ls7_bp3cat

Variable Name: ls7_bp3cat

Description: Poor Health defined as: (1) SBP ≥ 140; or (2) DBP ≥ 90

Intermediate Health defined as: (1) 140> SBP ≥120; or (2) 90 > DBP ≥ 80; or (3) 120>SBP and 80 > DBP if treated

Ideal Health is defined as: (1) 120 > SBP (if untreated); and (2) 80 > DBP (if untreated)

Note: We used self-reported BP medication for ‘treated’ vs ‘untreated’ due to many BP medications being used for non-HTN related reasons.

Format: 0=”Poor Health”

1=”Intermediate Health”

2=”Ideal Health”

Algorithm: if sbp>=**140** or dbp>=**90** then ls7_bp3cat=**0**;

else if **140**>sbp>=**120** or **90**>dbp>=**80** then ls7_bp3cat=**1**;

else if **120**>sbp>**0** and **80**>dbp>**0** and meds_bp_sr=**1** then ls7_bp3cat=**1**;

else if **120**>sbp>**0** and **80**>dbp>**0** and meds_bp_sr=**0** then ls7_bp3cat=**2**;

else ls7_bp3cat=**.**;

Source Variables: Dervar4: sbp, dbp, meds_bp_sr

21.12. LS7 Indicator for Ideal Health via Blood Pressure: ls7_bp_ideal

Variable Name: ls7_bp_ideal

Description: Ideal health indicator based on AHA Blood Pressure classification.

Format: 0=”No”

1=”Yes”

Algorithm: if ls7_bp3cat in (**0**,**1**) then ls7_bp_ideal=**0**;

else if ls7_bp3cat=**2** then ls7_bp_ideal=**1**;

else ls7_bp_ideal=**.**;

Source Variables: Dervar4: ls7_bp3cat

21.13. LS7 AHA Diabetes Categorization: ls7_dm3cat

Variable Name: ls7_dm3cat

Description: Poor Health defined as: (1) fasting glucose ≥ 126 mg/dL; or (2) HbA1c ≥ 6.5%; or (3) report of taking diabetes medications

Intermediate Health defined as: (1) 6.5% > HbA1c ≥ 5.7%; or (2) 126 mg/dL > FBG ≥ 100 mg/dL; and (3) no report of taking diabetes medications

Ideal Health defined as: (1) 5.7% > HbA1c; (2) 100 mg/dL > FBG; and (3) no report of taking diabetes medications

Format: 0=”Poor Health”

1=”Intermediate Health”

2=”Ideal Health”

Algorithm: if fbg_s>=**126** or hba1c>=**6.5** or meds_dm_sr=**1** then ls7_dm3cat=**0**;

else if (**126**>fbg_s>=**100** or **6.5**>hba1c>=**5.7**) and meds_dm_sr=**0** then ls7_dm3cat=**1**;

else if **5.7**>hba1c>**0** and **100**>fbg_s>**0** and meds_dm_sr=**0** then ls7_dm3cat=**2**;

else ls7_dm3cat=**.**;

Source Variables: Dervar4: fbg_s, hba1c, meds_dm_sr

21.14. LS7 Indicator for Ideal Health via Diabetes: ls7_dm_ideal

Variable Name: ls7_dm_ideal

Description: Ideal health indicator based on AHA Diabetes classification.

Format: 0=”No”

1=”Yes”

Algorithm: if ls7_dm3cat in (**0**,**1**) then ls7_dm_ideal=**0**;

else if ls7_dm3cat=**2** then ls7_dm_ideal=**1**;

else ls7_dm_ideal=**.**;

Source Variables: Dervar4: ls7_dm3cat

21.15. LS7 Score: ls7_score

Variable Name: ls7_score

Description: LS7’s score is derived by adding up the 3-way categorizations of all 7 components (smoking, BMI, physical activity, diet, cholesterol, blood pressure, and diabetes)

Format: Scale (0-14)

Algorithm: if cmiss(ls7_smk3cat, ls7_bmi3cat, ls7_pa3cat, ls7_diet3cat, ls7_chol3cat, ls7_bp3cat, ls7_dm3cat)>0 then ls7_score=.;

else ls7_score=ls7_smk3cat + ls7_bmi3cat + ls7_pa3cat + ls7_diet3cat + ls7_chol3cat + ls7_bp3cat + ls7_dm3cat;

Source Variables: Dervar4: ls7_smk3cat, ls7_bmi3cat, ls7_pa3cat, ls7_diet3cat, ls7_chol3cat, ls7_bp3cat, ls7_dm3cat

22. Life's Essential Eight (LE8)

LE8 Reference: Lloyd-Jones DM, Allen NB, Anderson CA, Black T, Brewer LC, et al. 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; 146(5): e18-e43.

22.01. LE8 Smoking Score: le8_smk

Variable Name: le8_smk

Description: Metric: Combustible tobacco; or secondhand smoke exposure

Scoring: 100 = Never smoker; 75 = Former smoker, quit >=5 y; 50 = Former smoker, quit 1–<5 y; 25 = Former smoker, quit <1 y; 0 = Current smoker
Subtract 20 points (unless score is 0) for any close contact with people where they were smoking.

Note: JHS Exam 4 does not ask how long it's been since the participant quit smoking. Our selection includes 1, Never used; 2, Rarely used in past, but not in past 30 days; 3, Regularly used in past, but not in past 30 days; 4, Used in past 30 days, less than weekly, but at least once per month; 5, Used in past 30 days, less than daily, but at least once a week; 6, Used daily or almost daily in past 30 days

If participant selected 1 their score was set to 100. If 2 was selected their score was set to 62.5 (midpoint of 75 and 50 since we can't say for sure when they quit smoking). If the participant selected 3 then their score was set to 25 (This coincides with LS7's definition so we wanted to make them match). If the participant selected 4, 5, or 6 then their score was set to 0.

Format: Scale (0-100)

Algorithm: *After removing participants who refused to answer smoking questions we took the max of the 6 options asked above for cigarettes, small cigars, large cigars, pipe, or hookah.

smoke=max(tob103a, tob103b, tob103c, tob103d, tob103e);

*The question on second-hand smoking asked the average number of hours per week the participant were they in close contact with someone who was smoking;

if tob101a=**0** then environ_exp=**0**;

else if tob101a>**0** then environ_exp=**1**;

else environ_exp=**.**;

*Scoring;

if smoke=**1** then le8_smk=**100**;

if smoke=**2** then le8_smk=**62.5**; *We need to leave a special note of this one, since it was rarely used in past, but not

in past 30 days we took midpoint of 50 and 75;

if smoke=**3** then le8_smk=**25**;

if smoke in (**4**,**5**,**6**) then le8_smk=**0**;

if environ_exp=**1** then le8_smk=le8_smk-**20**; *Subtract 20 for environmental exposure;

if le8_smk<**0** and le8_smk^=**.** then le8_smk=**0**; *0 minimum;

Source Variables: TOB1: tob101a, tob103a-tob103e

22.02. LE8 BMI Score: le8_bmi

Variable Name: le8_bmi

Description: Scoring for Life's Essential 8 BMI

BMI (kg/m2)

Scoring:

100 = <25;

70 = 25.0–29.9;

30 = 30.0–34.9;

15 = 35.0–39.9;

0 >=40.0

Format: Scale (0-100)

Algorithm: if **25**>bmi>**0** then le8_bmi=**100**;

else if **30**>bmi>=**25** then le8_bmi=**70**;

else if **35**>bmi>=**30** then le8_bmi=**30**;

else if **40**>bmi>=**35** then le8_bmi=**15**;

else if bmi>=**40** then le8_bmi=**0**;

else le8_bmi=**.**;

Source Variables: Dervar4: bmi

22.03. LE8 Physical Activity Score: le8_pa

Variable Name: le8_pa

Description: Scoring for Life's Essential 8 Physical Activity Component

Metric: Minutes of moderate or vigorous leisure physical activity (MVLPA) per week

Scoring:

100 = >= 150;

90 = 120–149;

80 = 90–119;

60 = 60–89;

40 = 30–59;

20 = 1–29;

0 = 0

Format: Scale (0-100)

Algorithm: if mvlpa >= **150** then le8_pa=**100**;

else if **120**<=mvlpa<**150** then le8_pa=**90**;

else if **90**<=mvlpa<**120** then le8_pa=**80**;

else if **60**<=mvlpa<**90** then le8_pa=**60**;

else if **30**<=mvlpa<**60** then le8_pa=**40**;

else if **1**<=mvlpa<**30** then le8_pa=**20**;

else if **0**<=mvlpa<**1** then le8_pa=**0**;

else le8_pa=**.**;

Source Variables: Dervar4: mvlpa

22.04. LE8 Diet (DASH) Score: le8_diet_dash

Variable Name: le8_diet_dash

Description: Scoring for Life's Essential 8 Diet Component using DASH Diet Score

Metric: DASH diet score

Scoring (population)

100 = >= 95%ile;

80 = 75-94%ile;

50 = 50-74%ile;

25 = 25-49%ile;

0 = 1-24%ile

Format: Scale (0-100)

Algorithm: Percentiles are gathered through Proc Univariate in SAS where max=100th percentile, p95=95th percentile, p75=75th percentile, p50=50th percentile, p25=25th percentile, and min=0th percentile. Coding for le8_diet_dash is then completed by

if max>=dash_score>=p95 then le8_diet_dash=**100**;

else if p95>dash_score>=p75 then le8_diet_dash=**80**;

else if p75>dash_score>=p50 then le8_diet_dash=**50**;

else if p50>dash_score>=p25 then le8_diet_dash=**25**;

else if p25>dash_score>=min then le8_diet_dash=**0**;

else le8_diet_dash=**.**;

Source Variables: Dervar4: dash_score

22.05. LE8 Cholesterol Score: le8_chol

Variable Name: le8_chol

Description: Scoring for Life's Essential 8 Cholesterol Component

Blood lipid: Non-HDL cholesterol (mg/dl)

Scoring:

100 = <130;

60 = 130-159;

40 = 160-189;

20 = 190-219;

0 >=220

Subtract 20 points (unless score is 0) If drug-treated level.

Format: Scale (0-100)

Algorithm: if hdl=**.H** then hdl=**200***sqrt(**2**); *Set value for out of range high for categorization purposes;

non_hdl=totchol-hdl;

if non_hdl>=**220** then le8_chol=**0**;

else if non_hdl>=**190** then le8_chol=**20**;

else if non_hdl>=**160** then le8_chol=**40**;

else if non_hdl>=**130** then le8_chol=**60**;

else if **130**>non_hdl>**0** then le8_chol=**100**;

else le8_chol=**.**;

*if drug-treated substract 20 points;

if meds_cholesterol=**1** then le8_chol=le8_chol-**20**;

if le8_chol<**0** and le8_chol^=**.** then le8_chol=**0**;

if meds_cholesterol=**.** and le8_chol>=**20** then le8_chol=**.**; *if no valid medication entry then we cannot get an accurate score;

Source Variables: Dervar4: totchol, hdl, meds_cholesterol

22.06. LE8 Blood Pressure Score: le8_bp

Variable Name: le8_bp

Description: Scoring for Life's Essential 8 Blood Pressure Component

Blood pressure: Systolic and diastolic BPs (mm Hg)

Scoring:

100 = <120 and <80 (optimal);

75 = 120–129 and <80 (elevated);

50 = 130–139 or 80–89 (stage 1 hypertension);

25 = 140–159 or 90–99;

0 >=160 or >=100

Subtract 20 points (unless score is 0) if drug-treated level.

Note: For drug-treated level self-reported use of BP medications was used due to BP medications often being used for non-hypertensive related reasons

Format: Scale (0-100)

Algorithm: if sbp>=**160** or dbp>=**100** then le8_bp=**0**;

else if **160**>sbp>=**140** or **100**>dbp>=**90** then le8_bp=**25**;

else if **140**>sbp>=**130** or **90**>dbp>=**80** then le8_bp=**50**;

else if **130**>sbp>=**120** and **80**>dbp>**0** then le8_bp=**75**;

else if **120**>sbp>**0** and **80**>dbp>**0** then le8_bp=**100**;

else le8_bp=**.**;

*Subtract 20 if drug treated;

if meds_bp_sr=**1** then le8_bp=le8_bp-**20**;

if le8_bp<**0** and le8_bp^=**.** then le8_bp=**0**;

if meds_bp_sr=**.** and le8_bp>=**20** then le8_bp=**.**; *If no valid SR BP Med then we can't get a reliable LE8 score;

Source Variables: Dervar4: sbp, dbp, meds_bp_sr

22.07. LE8 Diabetes Score: le8_dm

Variable Name: le8_dm

Description: Scoring for Life's Essential 8 Diabetes Component

Blood glucose: FBG (mg/dl) or HbA1c (%)

Scoring:

100 = No history of diabetes and FBG <100 (or HbA1c <5.7)

60 = No diabetes and FBG 100–125 (or HbA1c 5.7–6.4) (prediabetes)

40 = Diabetes with HbA1c <7.0

30 = Diabetes with HbA1c 7.0–7.9

20 = Diabetes with HbA1c 8.0–8.9

10 = Diabetes with HbA1c 9.0–9.9

0 = Diabetes with HbA1c >=10.0

Format: Scale (0-100)

Algorithm: if diabetes=**1** and hba1c>=**10** then le8_dm=**0**;

else if diabetes=**1** and **10**>hba1c>=**9** then le8_dm=**10**;

else if diabetes=**1** and **9**>hba1c>=**8** then le8_dm=**20**;

else if diabetes=**1** and **8**>hba1c>=**7** then le8_dm=**30**;

else if diabetes=**1** and **7**>hba1c>**0** then le8_dm=**40**;

else if diabetes=**0** and (**126**>fbg_s>=**100** or **6.5**>hba1c>=**5.7**) then le8_dm=**60**;

else if diabetes=**0** and (**100**>fbg_s>**0** or **5.7**>hba1c>**0**) then le8_dm=**100**;

else le8_dm=**.**;

Source Variables: Dervar4: fbg_s, hba1c, diabetes

22.08. LE8 Sleep Score: le8_sleep

Variable Name: le8_sleep

Description: Scoring for Life's Essential 8 Sleep Component

Metric: Average hours of sleep per night

Scoring:

100 = 7–<9;

90 = 9–<10;

70 = 6–<7;

40 = 5–<6 or >=10;

20 = 4–<5;

0 = <4

Format: Scale (0-100)

Algorithm: if **9**>sle201a>=**7** then le8_sleep=**100**;

else if **10**>sle201a>=**9** then le8_sleep=**90**;

else if **7**>sle201a>=**6** then le8_sleep=**70**;

else if **6**>sle201a>=**5** or 14>sle201a>=**10** then le8_sleep=**40**;

else if **5**>sle201a>=**4** then le8_sleep=**20**;

else if **4**>sle201a>=**0** then le8_sleep=**0**;

Source Variables: SLE2: sle201a

22.09. Number of missing LE8 Components: le8_nmiss

Variable Name: le8_nmiss

Description: Number of missing LE8 Components. This is a sum of number of missing LE8 items including smoking, BMI, physical activity, diet (DASH), cholesterol, Blood Pressure, diabetes, and sleep

Format: Integer

Algorithm: le8_nmiss=cmiss(le8_smk, le8_bmi, le8_pa, le8_diet_dash, le8_chol, le8_bp, le8_dm, le8_sleep);

Source Variables: Dervar4: le8_smk, le8_bmi, le8_pa, le8_diet_dash, le8_chol, le8_bp, le8_dm, le8_sleep

22.10. LE8 Total Score (Using Dash for the dietary component): le8_total

Variable Name: le8_total

Description: LE8 score with dash diet. This is the average of all 8 components.

Format: Numeric, 0-100

Algorithm: if le8_nmiss=**0** then le8_total=mean(le8_smk, le8_bmi, le8_pa, le8_diet_dash, le8_chol, le8_bp, le8_dm, le8_sleep);

else le8_total=**.**;

Source Variables: Dervar4: le8_smk, le8_bmi, le8_pa, le8_diet_dash, le8_chol, le8_bp, le8_dm, le8_sleep

22.11. LE8 Cardiovascular Health Category (Using Dash for the dietary component): le8_cvh3cat_dash

Variable Name: le8_cvh3cat

Description: LE8 Cardiovascular Health Category with dash diet. This is categorized as high health (1), moderate health (2), or low health (3)

Format: 1=”High”

2=”Moderate”

3=”Low”

Algorithm: if le8_total>=**80** then le8_cvh3cat=**1**;

else if **50**<=le8_total<**80** then le8_cvh3cat=**2**;

else if **0**<=le8_total<**50** then le8_cvh3cat=**3**;

else le8_cvh3cat=**.**;

Source Variables: Dervar4: le8_total