HCHS/SOL VISIT 2
DERIVED VARIABLE DICTIONARY

INV3.1 – July 2020

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

Version 3.1 – Visit 2 Investigator Use Data Sets (V2_INV3 data file)

HCHS/SOL Visit 2 Derived Variable Dictionary

Table of Contents

Updates to HCHS/SOL Visit 2 Data Release or Documentation

Version	Date	Description	Documentation
_INV1	12/6/2016	**1st Data Release (interim data; n=9,329)**	V1.0 (Dec 2016)
_INV2	3/26/2018	**2nd Data Release (Final V2 data except for ECHO; n=11,623)** ECHO dataset undergoing final data cleaning and quality assurance by the ECHO Reading Center (Harvard)	V2.0 (March 2018)
_INV3	3/1/2019	**3rd Data Release (Updated final V2 data; n=11,623)** **Summary of updates (details are at the end):** **PART_DERV_V2 (Participant Derived Variables)** - 9 new variables - 17 updated/corrected variables **ECHO (Echocardiography):** Final data received by ECHO Reading Center on 01/15/2019, processed on 1/30/2019. **PCE (Pregnancy Complications)** - 15 records removed because were not valid pregnancy complications **PCE_DERV** - Changed conversion of pregnancy length from months to weeks - Updated 5 variables that use pregnancy length **ALL FORMS. Identical to data release INV2** except that we added variable FORM which has language of administration of that particular form.	V3.0 (March 2019)
_INV3	7/13/2020	**4th Data Release (Update only PART_DERV_V2_INV3)** **Summary of updates (details are at the end):** **PART_DERV_V2 (Participant Derived Variables)** - 27 new variables - 11 updated variables - 4 renamed variables	V3.1 (July 2020)

1. DESIGN

1.1 PSU_ID (Primary Sampling Unit ID Clustering Variable- masked)

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

1.2 STRAT (Sampling Design Stratification Variable - masked)

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

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

1.3 LISTNUM (List Address Record Assigned)

Sample design variable used in statistical analyses that is the following:

Hispanic Surname versus non-Hispanic surname address list identifier from the original Genesys sampling address lists used in recruitment.

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

1.4 PSU_CNT (BG-level Hispanic Household Frame Counts)

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

1.5 STRAT_CNT (Stratum-level Hispanic Household Frame Count)

Sample design stratification household frame variable used in statistical analyses.

1.6 LIST_CNT (Sample List - Level Hispanic)

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

**For the following Sample Weight variable(s): Please see HCHS/SOL Manuscript “Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos” (Annals of Epidemiology, Volume 20, Issue 8, August 2010, Pages 629-641) for information on how the sample weights were created. Also see the HCHS/SOL Analyses Methods at Baseline (Sept. 2016) and Analyses Methods for Visit 2 (March 2018) documentation that describes the proper use of sample weights in statistical analyses.

As in any complex survey design, and as done in HCHS/SOL baseline, sampling weights account for non-response. One important and big difference between non-response at baseline and visit 2 is that at baseline all we knew from non-responders was age and sex (from screening) whereas at visit 2 we know all their baseline data. The calculation of the the final sampling weights (FROZEN AS OF 3/6/2018) for visit 2 is based on the sampling weights for visit 1 and taking into account the participant nonresponse for visit 2. See the

Analyses Methods for Visit 2 (March 2018) document for more information. There were eight variables used to account for non-response: Center, gender, age groups at Visit 1 (AGEGROUP_C6), AFU Refusal (Y/N), distance from clinic from recent Visit 2 address (categeorized) , GFR categories using serum creatinine only (GFRSCR_C3), Hispanic/Latino background (BKGRD1_C7), design stratification variable (STRAT), and then 3-level education attainment (EDUCATION_C3) are used to form finer strata and the nonresponse rate for visit 2 is then calculated for each stratum. The sampling weights are calculated based on visit 1 sampling weights and these nonresponse rates for visit 2. The sampling weights for V2 are then trimmed (to handle extreme values), calibrated (to the US 2010 Census population in the target population census tracks), and normalized (so that the sum of the sampling weights adds to the total sample size).

1.7 Weight_ Expanded_V2 (HCHS V2 Expanded Census 2010 Calibrated, Trimmed, Nonresponse Adjusted Weights (frozen as of 03/06/2018))

The expanded sample weight is calculated from the visit 1 sampling weights and nonresponse rates for visit 2, and then trimmed and calibrated.

1.8 Weight_ Norm_Overall_V2 (HCHS V2 Overall Normalized, Census 2010 Calibrated, Trimmed, Nonresponse Adjusted Weights - (frozen as of 03/06/2018))

The normalized sample weight is normalized to the overall sample size. **This final sample weight variable is used for MOST analyses.**

1.9 Weight_Norm_Center_V2 (HCHS V2 Center Normalized, Census 2010 Calibrated, Trimmed, NonResponse Adjusted Weights (frozen as of 03/06/2018))

ONLY use this sampling weight for analyses for which the population of interest is only ONE field center.

2. ADMINISTRATIVE

2.1 Center (Participant's field center)

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

Source variable(s):
SUBJID. Subject ID

2.2 Clindate_V2 (Visit 2 Clinic Date)

This is a SAS date variable which documents the date of the participant's clinic visit. It is derived from the mode date collected from the following form dates: Informed Consent tracking form (ICT), the CDART auto-populated Demographics form (DEM), Clinic Exam Checklist (CHK), Participant Disability Screener (PDE), or Personal Identifiers (IDE) forms, which are usually collected on the day of the exam.

Source variable(s):
ICT0A. Date of completion of the Informed Consent Form
DEM0A. Date of completion of the CDART auto-populated Demographics form
CHK0A. Date of completion of the Clinic Exam Checklist
PDE0A. Date of completion of the Participant Disability Screener Form
IDE0A. Date of completion of the Personal Identifiers Form

2.3 Consent_V2 (Informed consent status for participation in study)

This is a binary variable that determines whether or not a study participant consented to participate in the Visit 2 examination study.

If ICT1=1 then consent=1;
Else consent=0;

Source variable(s):
ICT1. Agrees to participant in HCHS/SOL study (yes/no)

2.4 Any_Permit_V2 (Any permission for public use of data, based on ICT3 and ICT5)

This is a binary variable that combines the 2 permission variables to determine whether or not a study participant gave any permission for public use of data.

IF ICT3=1 and ICT5=1 → 1
otherwise 0

Response Format: 0 = use not permitted
1 = permitted

Source variable(s):
ICT3. Agrees to allow HCHS/SOL and investigators HCHS/SOL works with to study samples (blood, cells and urine) in current and future research
ICT5. Agrees to allow HCHS/SOL and investigators HCHS/SOL works with to use genetic material (DNA/RNA) in current and future research

2.5 Duration_V2 (Difference between Arrival Time and Exit Exam Time (hrs))

This is a continuous variable that calculates the total exam duration (in hrs) using the start and end time from the Clinic Exam Checklist form (CHK).

DURATION_V2 = (chk30-chk4)/3600

Source variable(s):
CHK4. Arrival time
CHK30. Exit time

2.6 YRS_BTWN_V1V2 (Elapsed time between visits 1 and 2 (yrs))

This is a continuous variable that calculates the total elapsed time (in yrs) between HCHS Visit 1 and HCHS Visit 2 clinic dates.

YRS_BTWN_V1V2 = (clindate_v2-clindate)/365.25

Source variable(s):
CLINDATE. Visit 1 Clinic Date
CLINDATE_V2. Visit 2 Clinic Date

2.7 HomeVisit_V2 (Visit 2 Exam home visit flag based on CHK)

This is a binary variable that indicates whether or not a study participant participated in a home visit for HCHS Visit 2.

IF CHK31a=5 → 1
otherwise 0

Response Format: 0 = Not a home visit
1 = Home visit

Source variable(s):
CHK31a. Summary of Exam Visit 2

2.8 Full_AFU_Eligible (Core Baseline Requirements for AFU Eligibility)

A 0/1 numeric variable that defines a participant as meeting all requirements to determine if eligible for annual follow-up. The minimum subset of required forms and procedures are as follows (permanently missing forms are NOT included to determine completeness):

Questionnaires:

Informed Consent (version A or B)
Personal Information (PIE/PIS)
Personal Identifiers (IDE/IDS)
Medical History (MHE/MHS)

Procedures:
Anthropometry (ANT)
Sitting Blood Pressure (SBP)
Biospecimen Collection (BIO)

Additional criterion to be considered eligible for annual follow-up are as follows:
- Consent status to participate in Baseline Examination: Must have ICTA1=1 or ICTB1=1.
- Consent status to participate in yearly Annual Follow-Up: For those participants with an ICTA form, they must have ICTA3=1.
- Sample Blood Draws: At least one sample blood draw was collected. From the response matrix on the BIOA form, questions BIOA14A1-14A10 refer to samples not drawn for each tube (1-10). All 10 must be checked which indicates that no blood samples were drawn from participant.

Source variable(s):
ICTA1. Agrees to participate in baseline examination.
ICTA3. Agrees to yearly contact to answer questions about health/update contact information.
ICTB1. Agrees to participate in baseline examination and yearly contact and to answer questions about health/update contact information.
BIOA14A1-BIOA14A10. Blood drawing incidents – sample not drawn (tubes 1-10)

3. SOCIO-DEMOGRAPHIC

3.1 Age_V2 (Age at visit 2 (yrs))

This is the age of the participant in years (an integer variable) at the time of the participant's clinic visit. It is determined from the participant's date of birth (confidential variable not in release) and the clinic visit date.

AGE_V2 = INTEGER of (CLINDATE_V2– DOB_V2)/365.25

Source variable(s):
CLINDATE_V2. Visit 2 Clinic Date
DOB_V2. Date of birth

3.2 Agegroup_C6_V2 (6-level grouped age of participant)

This is the categorical (grouped) age of the participant. It is determined from the derived variable AGE_V2.

AGE_V2	AGEGROUP_C6_V2
18 – 24	1
25 – 34	2
35 – 44	3
45 – 54	4
55 – 64	5
65 +	6

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

Source variable(s):
AGE_V2. Age of participant at the time of their clinic visit

3.3 Agegroup_C6_NHANES_V2 (6-Level Age Groups (Visit 2 NHANES standardization))

This is the categorical (grouped) age of the participant. It is determined from the derived variable AGE_V2 for use with analyses requiring NHANES age standardization

AGE	Agegroup_C6_NHANES
18 – 29	1
30 – 39	2
40 – 49	3
50 – 59	4
60 – 69	5
70+	6

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

Source variable(s):
AGE_V2. Age of participant at the time of their clinic visit

3.4 Gender_V2 (Gender (F=Female, M=Male))

This is a categorical character variable which describes the participant's gender, female (F) or male (M).

DEM1	Gender
1	M
2	F

When missing use gender collected at baseline exam (GENDER)

Response Format: M=Male
F=Female

Source variable(s):
DEM1. Gender of person obtained from the CDART auto-populated Demographics form
GENDER. Gender from HCHS baseline

3.5 Income_V2 (Grouped yearly household income at Visit 2)

This derived variable groups each subject to one income category. The income is set to missing if both SEE4 and SEE5 are missing. Income is set to SEE4 if non-missing. Otherwise income is set to SEE5.

Income	SEE4	SEE5
1	1	.
2	2	.
3	3	.
4	4	.
5	5	.
6	.	1
7	.	2
8	.	3
9	.	4
10		5
.	.	.

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

Source variable(s):

SEE4. Low income range
SEE5. High income range

3.6 Income_C3_V2 (3-level grouped yearly household income at Visit 2)

This derived variable groups each subject to one income category based exclusively on SEE3. SEE3 was answered by some participants who refused to answer the follow up questions SEE4 and SEE5. Participants with missing SEE3 grouped together.

SEE3=1 → 1
SEE3=2 → 2
SEE3is missing or zero → 3

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

Source variable(s):
SEE3. Your household income for the year

3.7 Income_C5_V2 (5-level grouped yearly household income at Visit 2)

This derived variable groups each subject to one income category. The income is set to missing if both SEE4 and SEE5 are missing. Income_C5 is set to SEE4 if non-missing. Otherwise income is set to SEE5.

```
if SEE4 in(1) then INCOME_C5_V2=1;
else if SEE4 in(2,3) then INCOME_C5_V2=2;
else if SEE4 in(4,5) | SEE5=1 then INCOME_C5_V2=3;
else if SEE5 in(2,3) then INCOME_C5_V2=4;
else if SEE5 in(4,5) then INCOME_C5_V2=5;
else INCOME_C5_V2=.;
```

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

Source variable(s):
SEE4. Low income range
SEE5. High income range

3.8 Education_C3_V2 (3-level group education level at Visit 2)

The educational status derived variable groups the attainment of a high school diploma or an equivalent degree, not having that high school or equivalent degree, or attainment of education beyond a high school equivalent (i.e. college or vocational). This variable is based on the original Visit 1 response EDUCATION_C3 with updates to educational attainment indicated on questions SEE14 and SEE15, as follows:

```
 if see14=0 then EDUCATION_C3_V2=EDUCATION_C3;
else if see14=1 then do;
      if see15 in(1,2) & (EDUCATION_C3=1 | EDUCATION_C3<=.Z) then
      EDUCATION_C3_V2=1;

      else if see15 in(1,2) & EDUCATION_C3 in(2,3) then
      EDUCATION_C3_V2=EDUCATION_C3;

      else if see15 in(3) & (EDUCATION_C3 in(1,2) | EDUCATION_C3<=.Z) then
      EDUCATION_C3_V2=2;

      else if see15 in(3) & EDUCATION_C3 in(3) then
      EDUCATION_C3_V2=EDUCATION_C3;

      else if see15 in(4,5) then EDUCATION_C3_V2=3;
 end;
```

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

Source variable(s):
EDUCATION_C3. Education Status (3 levels)
SEE14. Have you been involved in any educational or training program since the first SOL center visit
SEE15. What was the highest grade/level of education achieved

Note: Currently the variable EDUCATION_C3_V2 was set as missing when SEE15=6 (“Others”) with additional text information.

3.9 N_HC_V2 (Current Health Insurance Coverage at Visit 2)

This variable groups the health insurance coverage status of the participant. The definition also checks whether or not any reason was giving for stopping coverage.

```
hi_ind=sum(hce32a,hce32b,hce32c,hce32d,hce32e,hce32f,hce32g,hce32h,hce32i,hce32j)
;

stop_ind=sum(hce37a,hce37b,hce37c,hce37d,hce37e,hce37f,hce37g,hce37h,hce37i,hce3
7j,hce37k,hce37l,hce37m);

    if hce31<=.Z and hi_ind>0 then N_HC_V2=1;
    else if hce31<=.Z and stop_ind>0 then N_HC_V2=0;
    else if hce31=1 then N_HC_V2=1;
    else if hce31=0 then N_HC_V2=0;
```

Response format: 0=No current health insurance (No)

1=Currently have health insurance (Yes)

Source variable(s):
HCE31. Do you have health insurance or other health care coverage?
HCE32a-HCE32j. What type of health insurance or health care coverage do you have?
HCE37a –HCE37m. Reasons you stopped being covered by health insurance.

3.10 Employed_V2 (4-level grouped employment status)

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

```
if (see8=0) then hours_per_week=0;
else if (see8>.Z & see9>.Z & see10>.Z and see11>.Z) then do;
   hours_per_week=0;
   if (see10>0) then hours_per_week=hours_per_week+(see10a*see9/12);
   if (see11>0) then hours_per_week=hours_per_week+(see11a*see9/12);
end;

if (see7=1) then EMPLOYED=1;
else if (see8=0) then EMPLOYED=2;
else if (.Z<hours_per_week & hours_per_week<=35) then EMPLOYED=3;
else if (.Z<hours_per_week & hours_per_week>35) then EMPLOYED=4;
else EMPLOYED=.;
```

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

Source variable(s):
SEE7. Are you retired?
SEE8. In the past 12 months, did you have any paid employment?
SEE9. In the past 12 months, how many months did you work?
SEE10. When you were working during the past 12 months, in an average month, how many full-time jobs did you have?
SEE10A. On average, how many hours per week did you work in those full-time jobs?
SEE11. When you were working during the past 12 months, in an average month, how many part-time jobs did you have?
SEE11A. On average, how many hours per week did you work in those part-time jobs?

4. ACCULTURATION

4.1 Lang_Pref_V2 (Language preference at Visit 2 - (1=Span,2=Eng))

This variable determines which language was preferred to be used for the visit 2 examination. First, the Participant Eligibility form (ELE) is used which directly asks the participant which language they prefer to use. If this variable is missing, then three core forms are used: Participant Disability Screener (PDE), Personal Identifiers (IDE), and Medical History (MHE). A majority of those used in one language determines which language was preferred.

Response format: 1=Spanish
2=English

Source variable(s):
ELE1: Do you prefer to communicate in Spanish or English?
Form Variable of PDE
Form Variable of IDE
Form Variable of MHE

4.2 YRSUS_V2 (Number of years lived in the US (50 States) at Visit 2)

This is a numeric variable indicating the number of years lived in the US (50 States). It uses the Visit 1 YRSUS as well as duration between V1 and v2 to update the years in US for Visit 2, as follows:

If US_born_V2=1 then set YRSUS_V2=Age_V2;
Else YRSUS_V2 = YRSUS_V1+YRS_BTWN_V1V2.

Source variable(s):
US_BORN_V2. Born in US updated at Visit 2.
YRSUS_V1. Years lived in the US at Visit 1.
YRS_BTWN_V1V2. Elapsed time between visits 1 and 2 (yrs)

4.3 YRSUS_C2_V2 (2-level grouped years lived in the US (50 States))

This is a 2-level grouped numeric variable that defined less than 10years lived in the US versus 10 or more years lived in the US. It uses Number of years lived in the US (50 States) at Visit 2 , as follows:

if .Z< YRSUS_V2<10 then YRSUS_C2_V2=1;

else if YRSUS_V2>=10 then YRSUS_C2_V2=2;

else YRSUS_C2_V2=.;

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

Source variable(s):
YRSUS_V2. Years lived in the US at Visit 2.

4.4 YRSUS_C3_V2 (3-level grouped years lived in the US at Visit 2 (50 States))

This is a 3-level grouped numeric variable that defined less than 10years lived in the US versus 10 or more years lived in the US, plus US born group. It uses Number of years lived in the US (50 States) at Visit 2 and US born information, as follows:

If US_BORN_V2 = 1 then YRSUS_C3_V2 = 3;
Else if and YRSUS_C2_V2=2 then YRSUS_C3_V2 = 2;
Else if and YRSUS_C2_V2=1 then YRSUS_C3_V2 = 1;

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

Source variable(s):
YRSUS_C2_V2. 2-level grouped years lived in the US at Visit 2.
US_BORN_V2. Born in US updated at Visit 2.

4.5 POB_V2 (Place of Birth - country (corrected))

This variable determines the country of origin of the participant using several different variables collected in HCHS baseline and HCHS Visit 2. These variables are used to determine country of origin using multiple questions on place of birth, citizenship by country of birth, inconsistencies in Visit 2 responses as well as Visit 1 responses, and correcting for likely typographical errors using a combination of Visit 2 data (HCE form questions 39a-e and question 40) as well as participant and maternal/paternal place of birth from HCHS Visit 1 (PIE form).

```
POB_V2=HCE39b;

if POB_V2 in(" ","~") then POB_V2=PIEA4;
if (POB_V2^=PIEA4) and (PIEA4=PIEA11 or PIEA4=PIEA14) then POB_V2=PIEA4;
if HCE39=1 and HCE40=1 and POB_V2 in('Q',' ') then POB_V2='63';
if (HCE40=2 and POB_V2='56') and (PIEA11='57' or piea14='57') then POB_V2='57';
if HCE39a='55' and POB_V2 in('58','63') and HCE40=2 then POB_V2='57';
if POB_V2 in('62','64') and (POB_V2=PIEA4) and HCE40=1 then POB_V2='63';
```

Response format: Country codes 01-99, as found in HCHS Visit 1 Personal Information form (PIEA).

Source variable(s):
HCE39. Where were you born? (In US, outside of US)
HCE39a. Specify State or territory (in US)
HCE39b. Specify country (outside of US)
HCE40. Are you a U.S. citizen?
PIEA4. Born in which Country/territory (from Visit 1)

PIEA11. Mother's country of origin (from Visit 1)
PIEA14. Father's country of origin (from Visit 1)

4.6 POB_C8_V2 (8-level Place of Birth Groups)

This variable categorizes the participants' place of birth for use in analyses.

```
 POB_C8_V2=.;
 if POB_V2 = '63' then POB_C8_V2 = 1; **US Mainland;
 else if POB_V2 in('29', '57', '66') then POB_C8_V2 = 2; **US Territories only;
 else if POB_V2 = '18' then POB_C8_V2 = 3; **Cubans;
 else if POB_V2 = '21' then POB_C8_V2 = 4; **Dominican;
 else if POB_V2 = '46' then POB_C8_V2 = 5; **Mexicans;
 else if POB_V2 in('10', '17', '23', '30', '33', '48', '51') then POB_C8_V2 = 6; **Central
Americans;
 else if POB_V2 in('04', '11', '12', '14', '16', '22', '52', '53', '64', '65' ) then POB_C8_V2 = 7;
**South Americans;
 else if POB_V2^=" " then POB_C8_V2 = 8; **Other;
```

Response format:
1=US Mainland
2=US Territories only
3=Cuba
4=Dominican Republic
5=Mexico
6=Central America
7=South America
8=Other

Source variable(s):
POB_V2. Place of Birth - country (corrected)

4.7 US_Born_V2 (US Born (US States only))

This variable groups the place of birth of the participant to the United States (50 states only) or other place of birth using place of birth from Visit 2 and corrections for instances when place of birth (POB) is missing.

```
 if missing(POB_V2)=0 and POB_V2 not in('Q', '.', ' ') then do;
   if POB_V2 = '63' then US_BORN_V2 = 1;
   else US_BORN_V2 = 0;
end;

 if missing(POB_V2)=1 or POB_V2 in('Q', '.', ' ') then do;
   if HCE39_R =1 or HCE40_R in(1,2) then US_NATIVE_V2 = 1;
   else US_NATIVE_V2=0;
 end;

 if HCE40_R=. and POB_V2=" " then US_BORN_V2 = .;
```

Response format: 0=Not born in 50 US States.
1=Born in 50 US States Only

Source variable(s):
POB_V2. Place of Birth - country (corrected).
HCE39_R. Where were you born? (In US, outside of US, recode)
HCE40_R. Are you a U.S. citizen? (Recode)

4.8 US_Native_V2 (US Native (US States and Territories))

This variable groups the place of birth of the participant to the United States (50 states and US territories which includes Puerto-Rico, Guam, and Virgin Islands) or other place of birth using place of birth from Visit 2 and corrections for instances when place of birth (POB) is missing.

```
 if missing(POB_V2)=0 and POB_V2 not in('Q', '.', ' ') then do;

   if POB_V2 in('29', '57', '63', '66') then US_NATIVE_V2=1;
   else US_NATIVE_V2 = 0;
 end;

 if missing(POB_V2)=1 or POB_V2 in('Q', '.', ' ') then do;
   if HCE39_R =1 or HCE40_R in(1,2) then US_NATIVE_V2 = 1;
   else US_NATIVE_V2=0;
 end;

 if missing(HCE39_R)=1 and missing(hce40_R)=1 and missing(POB_V2)=1 then
US_NATIVE_V2=.;
```

Response format: 0=Not born in US or US territory
1=Born in US or US territory

Source variable(s):
POB_V2. Place of Birth - country (corrected).
HCE39_R. Where were you born? (In US, outside of US, recode)
HCE40_R. Are you a U.S. citizen? (Recode)

4.9 IMGEN_C3_V2 (3-level Immigrant Generation (US Terr = Foreign Born))

This variable determines the participants immigrant generation status based on a combination of the country of origin of the participant (US mainland versus not) and maternal and paternal parents (US mainland versus not), as follows:

First generation is defined as foreign-born with foreign-born parents. When place of birth status is missing (POB_V2) then responses to other questions on citizenship by place of birth are used.

Second generation is defined as US born with **at least one** foreign-born parent. In addition, foreign-born adults with at least one US born parent are classified as second generation. Foreign-born adults with at least one US-born parent are considered US-born citizens (must have both parents with non-missing responses).

The third generation (or greater) is defined as US born respondent and both US born mother and father (must have both parents with non-missing responses).

```
if (PIEA11=" " or PIEA11='Q') and (PIEA14=" " or PIEA14='Q') then miss_parents=1;

  if POB_V2 ^ in('Q',' ') then do;

    if (POB_V2 ^='63') and (PIEA11 ^='63' or PIEA14 ^='63') then IMGEN_C3_V2=1;

    else if (POB_V2='63') and (miss_parents^=1 and (PIEA11^='63' or PIEA14^='63')) then
IMGEN_C3_V2=2;

    else if (POB_V2='63') and (miss_parents^=1 and (PIEA11='63' and PIEA14='63')) then
IMGEN_C3_V2=3;

  end;


 if IMGEN_C3_V2=. then do;

    if HCE40_R in(0,2,3,4) then IMGEN_C3_V2=1;

  end;
```

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

Source variable(s):
POB_V2. Place of Birth - country (corrected).
HCE40_R. Are you a U.S. citizen? (Recode)
PIEA11. Mother's country of origin
PIEA14. Father's country of origin

4.10 IMGEN_A_C3_V2 (3-level Immigrant Generation (US Terr = US Born))

This variable, similar to IMGEN_C3_V2, determines the participants immigrant generation status based on a combination of the country of origin of the participant (US mainland/territories versus not) and maternal and paternal parents (US mainland/territories versus not), as follows:

First generation is defined as foreign-born with foreign-born parents. When place of birth status is missing (POB_V2) then responses to other questions on citizenship by place of birth are used.

Second generation is defined as US born with **at least one** foreign-born parent. In addition, foreign-born adults with at least one US born parent are classified as second generation. Foreign-born adults with at least one US-born parent are considered US-born citizens (must have both parents with non-missing responses).

The third generation (or greater) is defined as US born respondent and both US born mother and father (must have both parents with non-missing responses).

```
  if POB_V2 ^ in('Q',' ') then do;
    if (POB_V2 ^in('29','57','63','66')) and (PIEA11 ^in ('29', '57', '63', '66') or PIEA14
^in('29','57','63','66')) then IMGEN_A_C3_V2=1;
    else if (POB_V2 in('29','57','63','66')) and (miss_parents^=1 and (PIEA11
^in('29','57','63','66') or PIEA14 ^in('29','57','63','66'))) then IMGEN_A_C3_V2=2;
    else if (POB_V2 in('29','57','63','66')) and (miss_parents^=1 and (PIEA11
in('29','57','63','66') and PIEA14 in ('29','57','63','66'))) then IMGEN_A_C3_V2=3;
  end;

  if IMGEN_A_C3_V2=. then do;
    if HCE40_R in(0,4) then IMGEN_A_C3_V2=1;
  end;
```

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

Source variable(s):
POB_V2. Place of Birth - country (corrected).
HCE40_R. Are you a U.S. citizen? (Recode)
PIEA11. Mother's country of origin
PIEA14. Father's country of origin

4.11 SASH_LANG_V2 (Short acculturation scale for Hispanics – language subscale, Visit 2)

This is a numeric variable. A 5-point likert-type scale is assigned to each question with 1 =Only Spanish and 5= Only English. The average of the 6 questions indicates the degree of language acculturation. If more than 1 item is missing, SASH_LANG_V2 is set to missing.

SASH_LANG_V2 = average (of ace1-6).

Source variable(s):
ACE1. In general, what language(s) do you read and speak?

ACE2. What was the language(s) you used as a child?
ACE3. What language(s) do you usually speak at home?
ACE4. In which language(s) do you usually think?
ACE5. What language(s) do you usually speak with your friends?
ACE6. In general, what language(s) are the movies, T.V. and radio programs you prefer to watch and listen to?

4.12 SASH_SOC_V2 (Short acculturation scale for Hispanics – Social subscale, Visit 2)

This is a numeric variable. A 5-point likert-type scale is assigned to each question with 1 = All Hispanic/Latino and 5= All non-Hispanic/non-Latino. The average of the 4 questions indicates the degree of social acculturation. If any item is missing, SASH_SOC_V2 is set to missing.

SASH_SOC_V2 = average (of ace7-10).

Note: this variable is missing too many people because the participants refused item 4 “If you could choose your children’s friends you would want them to be…”.(DSA)

Source variable(s):
ACE7. Your close friends are:
ACE8. You prefer going to social gatherings/parties at which the people are:
ACE9. The persons you visit or who visit you are:
ACE10. If you could choose your children’s friends, you would want them to be:

4.13 Language_Subscore_MESA_V2 (Language Subscore – MESA, Visit 2)

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

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

ACE3	Language_Subscore_MESA
1	0.0
2	0.5
3	1.0
4	1.5
5	2.0
Missing	Missing

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

Source variable(s):
ACE3. What language(s) do you usually speak at home?

4.14 Nativity_Subscore_MESA_V2 (Nativity Subscore – MESA, Visit 2)

This is a nativity subscore of the acculturation score similar to the one used in MESA. It is based on nativity and years of residence in the U.S at Visit 2.

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

If US_Born_V2 is missing or (US_Born_V2=0 and YRSUS_V2 is missing) then Nativity_Subscore_MESA_V2 is missing.
Otherwise:
Nativity_Subscore_MESA_V2 is 3, if US_Born_V2=1;
2, if US_Born_V2=0 and 20<=YRSUS_V2;
1, if US_Born_V2=0 and 10<=YRSUS_V2<20;
0, if US_Born_V2=0 and .z<YRSUS_V2<10.

Source variable(s):
US_Born_V2. US born visit 2 version. Grouped place of birth – 50 US States only
YRSUS_V2. Years lived in the US at Visit 2.

4.15 Accult_MESA_V2 (Acculturation Score – MESA, Visit 2)

This is a summary acculturation score based on nativity, language spoken at home, and years of residence in the U.S. (similar to the acculturation score used in MESA). It is defined as the sum of nativity subscore and language subscore at Visit 2

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

If Nativity_Subscore_MESA_V2 is missing or Language_Subscore_MESA_V2 is missing then Accult_MESA_V2 is missing;

Otherwise
Accult_MESA_v2= Nativity_Subscore_MESA_V2 + Language_Subscore_MESA_V2.

Source variable(s):
Nativity_Subscore_MESA_V2. Nativity Subscore – MESA at Visit 2
Language_Subscore_MESA_V2. Language Subscore – MESA at Visit 2.

5. Sociocultural Scores

5.1 From Visit 2 Social Support (SSE)

Interpersonal Support Evaluation List (ISEL) based on Social Support 12 items (SSE1-SSE12)

Scoring process:

Step 1: convert the 12 item values from 1, 2, 3, 4 to 0, 1, 2, 3 by minus one;

Step 2: reverse Items 1, 2, 7, 8, 11, 12 to 1R, 2R, 7R, 8R, 11R, 12R by using 3 minus the values of Items 1, 2, 7, 8, 11, 12;

Step 3: Subscale scoring:

5.1.1 Appraisal Subscale

ISEL_app_cnt_v2=n(of 2R, 4, 6, 11R);
label ISEL_app_cnt_v2='ISEL-Appraisal Subscale non-missing items count,0-4, V2';

ISEL_app_v2=sum(of 2R, 4, 6, 11R);
if ISEL_app_cnt_v2<**4** then ISEL_app_v2=**.**;
label ISEL_app_v2='Interpersonal Support Evaluation List (ISEL)-Appraisal Subscale, V2';

5.1.2 Belonging Subscale

ISEL_belong_cnt_v2=n(of 1R, 5, 7R, 9);
label ISEL_belong_cnt_v2='ISEL-Belonging Subscale non-missing items count,0-4, V2';

ISEL_belong_v2=sum(of 1R, 5, 7R, 9);
if ISEL_belong_cnt_v2<**4** then ISEL_belong_v2=**.**;
label ISEL_belong_v2='Interpersonal Support Evaluation List (ISEL)-Belonging Subscale, V2';

5.1.3 Tangible Subscale

ISEL_tangible_cnt_v2=n(of 3, 8R, 10, 12R);
label ISEL_tangible_cnt_v2='ISEL-Tangible Subscale non-missing items count,0-4, V2';

ISEL_tangible_v2=sum(of 3, 8R, 10, 12R);
if ISEL_tangible_cnt_v2<**4** then ISEL_tangible_v2=**.**;

label ISEL_tangible_v2='Interpersonal Support Evaluation List (ISEL)-Tangible Subscale, V2';

5.1.4 Overall Support: Sum of 3 Subscales

ISEL_all_V2=sum of items (2R, 4, 6, 11R) , (1R, 5, 7R, 9) and (3, 8R, 10, 12R)
Missing level control at CSCC: the score was calculated only when missing < 20% of total items (allow up to 2 missing out of 12), and adjusted by multiply 12/11 or 12/10 when there is only 1 or 2 missing items.
label ISEL_all_V2='Interpersonal Support Evaluation List (ISEL)-overall, V2';

Primary Reference:

Cohen, S., Memelstein, R., Kamarck, T., & Hoberman, H. (1985). Measuring the functional components of social support. In I.G. Sarason & B. Sarason (Eds.), Social support: Theory, research and application (pp.73-94). The Hague: Martinus Nijhoff.

Type of Measure: Modified. Original scale consists of 40 items (ten items in each of the four sub-cales measuring separate aspects of social support). The modified version includes only three sub-scales (the self-esteem sub-scale was excluded because it overlaps with the self-esteem measure) and only four highest-loading items for each sub-scale.

Purpose: To assess perceived availability of four types of social support (appraisal, belonging, self-esteem, and tangible). Availability of social support has been linked to reduced mortality (Rosenberg, Orth-Gomer, Wedel, & Wilhemsen, 1993) and improved psychological state (Cohen & Wills, 1985).

Description: Respondents indicate the extent to which sentences describing availability of different types of social support in their lives are true or false. No time frame or referent period is used.

Scaling: 0 = Definitely False; 1 = Probably False; 2 = Probably True; 3 = Definitely True

items: 12

Sample items: “If I were sick, I could easily find someone to help me with me daily chores.” (tangible) “I don’t often get invited to do things with others.” (reversed; belonging) “When I need suggestions on how to deal with a personal problem, I know someone I can turn to.” (appraisal)

Psychometrics:

Reliability: undergraduate students, alpha = .77 - .86 general population, alpha = .88 - .90

Validity: Correlates positively with other support scales (Inventory of Socially Supportive Behaviors), with number of close friends, and with the measure of the quality of marital relationships (Partner Adjustment Scale). The sub-scales are also associated in the predicted direction with related trait measures: self-esteem subscales correlates with self-esteem measure whereas appraisal subscale correlates with self-disclosure measure.

Source variable(s):
SSE1 – SSE12: the degree of response to the particular social support statements (from 1 - 4).

Note: the following 3 subscales are not recommended to be used in analysis and had been removed from the part_derv data. Please only use the total score.

5.2 From Visit 2 Family Cohesion (FCE)

5.2.1 Cohesion Subscale non-missing item count

Fmly_Cohesion_cnt_V2=non-missing count of FCE1 FCE2R FCE3 FCE4 FCE6 FCE7R FCE8 FCE9;

→FCE2R, FCE7R revised value 0/1 from FCE2, FCE7.

Label Fmly_Cohesion_cnt_v2='Family Cohesion Subscale Scaled Raw Score non-missing items count,0-8, V2';

5.2.2 Cohesion Subscale Raw Score

Fmly_Cohesion_raw_V2=sum of FCE1 FCE2R FCE3 FCE4 FCE6 FCE7R FCE8 FCE9;

→FCE2R, FCE7R revised value 0/1 from FCE2, FCE7.

Missing level control at CSCC: the score was calculated only when missing < 20% of total items (allow 1 missing out of 8), and adjust the values by multiply 8/7 when there is only 1 missing item.
Label Fmly_Cohesion_raw_V2='Family Cohesion Subscale Scaled Raw Score, V2';

5.2.3 Cohesion Subscale Scaled Score

if Fmly_Cohesion_raw_V2=0 then Fmly_Cohesion_scale_V2= 4;
if Fmly_Cohesion_raw_V2=1 then Fmly_Cohesion_scale_V2=11;
if Fmly_Cohesion_raw_V2=2 then Fmly_Cohesion_scale_V2=18;
if Fmly_Cohesion_raw_V2=3 then Fmly_Cohesion_scale_V2=25;
if Fmly_Cohesion_raw_V2=4 then Fmly_Cohesion_scale_V2=31;
if Fmly_Cohesion_raw_V2=5 then Fmly_Cohesion_scale_V2=38;

if Fmly_Cohesion_raw_V2=6 then Fmly_Cohesion_scale_V2=45;
if Fmly_Cohesion_raw_V2=7 then Fmly_Cohesion_scale_V2=52;
if Fmly_Cohesion_raw_V2=8 then Fmly_Cohesion_scale_V2=59;
if Fmly_Cohesion_raw_V2=9 then Fmly_Cohesion_scale_V2=65;

Label Fmly_Cohesion_scale_V2='Family Cohesion Subscale Scaled Score, V2';

5.3 From Visit 2 Chronic Stress (STE)

5.3.1 Chronic Stress Scores

Based on STE1-STE8, STE1a-STE8a, STE1b-STE8b, STE8c.

According to Patty Gonzalez's email on 11/08/2011, add the Chronic Stress scores for Visit 2 STE data questions 1-8 serise:

Reference: Bromberger, J. T., & Matthews, K. A. (1996). A longitudinal study of the effects of pessimism, trait anxiety, and life stress on depressive symptoms in middle-aged women. Psychol Aging, 11, 207-213.

Items: 1-8 on STE

Description: Each question consists of 3 parts (did participants experience the stressor (0=No, 1=Yes); has the stressor been a problem for six months or more (0=No, 1=Yes); and severity of stressor (1=Not very stressful, 2=Moderately stressful, 3=Very stressful)

Items: 8

Scoring:

Total number of Chronic stressors: Sum each event that participants answered affirmatively (Yes=1): Item 1, Item 2, Item 3, Item 4, Item 5, Item 6, Item 7, Item 8. Scores range from 0 to 8.

Per Linda Gallo's email on June 13, 2012:

Add together the number of chronic stressors, meaning stressors endorsed as having occurred, and indicated to have lasted at least 6 months based on part a of the item. So to score one point, the person would both have to state that 1) yes the stressor occurred and 2) yes it lasted at least 6 months. (Regardless of severity-Patricia Gonzalez, 06/29/2012) Add up these points and that is the total number of "chronic" (stressors lasting at least 6 months) stressors.

Recoded items of STE1 – STE8 for CHR_STR_TOT_V2:
N_CS_Q1_V2 - N_CS_Q8_V2

CHR_STR_TOT_V2= Sum of (recode items of STE1 – STE8)
The score was calculated only if there is no missing item.
Label = Total number of Chronic Stressors, V2.

5.3.2 Moderate to very Chronic Stress Scores

Recode (STE Items 1b, 2b, 3b, 4b, 5b, 6b, 7b, 8c) into 1=0, 2=1, 3=1.

Per Linda Gallo's email on June 13, 2012:

Add together the number of "chronic" stressors rated as moderately or very stressful. To score one point, the person would have to state that 1) yes the stressor occurred and 2) yes it lasted at least 6 months and 3) it was moderately or very stressful. Then add the number of these points and that is the total number of "moderately or very stressful chronic stressors".

Recoded items of STE1 – STE8 for M_V_TOT_V2:
M_V_Q1_V2 – M_V_Q8_V2

M_V_TOT_V2= Sum Recoded Items (STE1b, 2b, 3b, 4b, 5b, 6b, 7b, 8c),
The score was calculated only if there is no missing item.
Label=Total Moderate to Very Chronic Stress Score, V2.

6. Well Being

6.1 CESD10_V2 (10-item CES-D summary score for Visit 2)

We want to create a variable based upon the responses for WBE1-WBE10. This variable is conceptually the sum of WBE1-WBE10. However since we allow for some missing values in the data, for those observations with missing it is defined as 10 times the mean of the non-missing. Another complication is that out of the 10 questions, most of the questions query about negative well-being status, while for questions 5 and 8 they query about positive well-being status. Thus, in order that higher scores always have the same meaning, the responses to questions 5 and 8 are reversed (i.e. 0 changed to 3, 1 changed to 2, 2 changed to 1, and 3 changed to 0).

This is a numeric variable with values ranging from 0 to 30. Visit 2 CESD10_v2 score modified the existing Visit 1 algorithm for CESD10 from WBEA form so that it works with the Visit 2 WBE array. The CESD items are WBE1 to WBE10. The two positive affect items that are reverse scored are WBE5 and WBE8. Missing if more than 2 of the 10 items (WBE1-WBE10) are missing

Sum of (WBE1-WBA10), where positively worded items (WBE5, WBE8) are reverse coded. For participants with 1 or 2 missing items, CESD10 is equal to the weighted average of the nonmissing items * 10

Response format for WBE1-WBE10:
0 = rarely or none of the time
1 = some or a little of the time
2 = occasionally or a moderate amount of time
3 = all of the time

Source variable(s):
WBE1. I was bothered by things that usually don't bother me
WBE2. I had trouble keeping my mind on what I was doing
WBE3. I felt depressed
WBE4. I felt that everything I did was an effort
WBE5. I felt hopeful about the future
WBE6. I felt fearful
WBE7. My sleep was restless
WBE8. I was happy
WBE9. I felt lonely
WBE10. I could not "get going"

6.2 GAD7_V2 (GAD-7 Sum Score at Visit 2)

This is a numeric variable with values ranging from 0 to 21, or missing.

If WBE11 to WBE17 has no more than one missing value then compute sum score
GAD7_V2 = sum of items WBE11 to WBE17
If 2 or more items missing then GAD7_V2= missing.

Response format for WBE11-WBE17: 0 = Not at all
1 = Several days
2 = More than half the days
3 = Nearly every day

Source variable(s):
WBE11. Feeling nervous, anxious or on edge.
WBE12. Not being able to stop or control worrying.
WBE13. Worrying too much about different things.
WBE14. Trouble relaxing.
WBE15. Being so restless that it is hard to sit still.
WBE16. Becoming easily annoyed or irritable.
WBE17. Feeling afraid as if something awful might happen.

6.3 ANXIETY_V2 (Generalized anxiety disorder possible at Visit 2)

This is a dichotomous variable with values 1=Yes/2=No, or missing.

If GAD7_V2 GE 10 then Anxiety_V2 =1
Else if (**0** LE GAD7_V2 LT 10) then Anxiety_V2=0

Otherwise Anxiety_V2=.

Response format for ANXIETY_V2: 0 = No
1 = Yes

Source variable(s):
GAD7_V2. GAD-7 Sum Score at Visit 2.

7. ANTHROPOMETRY

7.1 BMI_V2 (Body Mass Index (kg/m^2) at Visit 2)

This is a numeric variable.

Missing if the height variable, ANT2, is missing.

If ANT2 is not missing then BMI_V2 = ANTA4 / (ANT2/100)^2

Source variable(s):
ANT2. Standing height in centimeters
ANTA4. Weight in kilograms

7.2 BMIGRP_C4_V2 (4-level grouped Body Mass Index – WHO at Visit 2)

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

if 0≤BMI_V2<18.5 then BMIGRP_C4_V2=1;
else if 18.5≤BMI_V2<25 then BMIGRP_C4_V2=2;
else if 25≤BMI_V2<30 then BMIGRP_C4_V2=3;
else if 30≤BMI_V2 then BMIGRP_C4_V2=4;
else if BMI_V2=.Z then BMIGRP_C4_V2=.;

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

Source variable(s):
BMI_V2. Body Mass Index (kg/m^2) at Visit 2

7.3 BMIGRP_C6_V2 (6-level grouped Body Mass Index – WHO at Visit 2)

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

if 0≤BMI_V2<18.5 then BMIGRP_C6_V2=1;

else if 18.5≤BMI_V2<25 then BMIGRP_C6_V2=2;
else if 25≤BMI_V2<30 then BMIGRP_C6_V2=3;
else if 30≤BMI_V2<35 then BMIGRP_C6_V2=4;
else if 35≤BMI_V2<40 then BMIGRP_C6_V2=5;
else if 40≤BMI_V2 then BMIGRP_C6_V2=6;
else if BMI_V2=.Z then BMIGRP_C6_V2=.;

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

Source variable(s):
BMI_V2. Body Mass Index (kg/m^2) at Visit 2

7.4 AFUCOPD_self_byV2 (Self-reported COPD/Emph or CB only in AFU by V2)

This is a 0/1 numeric variable that indicates a self-report of COPD, emphysema or Chronic bronchitis only reported in AFU by V2.

Among participants who did not have self-reported COPD, emphysema or Chronic bronchitis in visit 1:
if any OPE5=1 by Visit 2 visit, then AFUCOPD_SELF_byV2=1;

Response format: 0 = No
1 = Yes

Source variable(s):
OPE5. Self-report COPD/emphysema or CB in AFU visit.

7.5 COPD_EVER_V2 (Self-reported COPD/Emph or CB by V2)

This is a 0/1 numeric variable that indicates a self-report of COPD, emphysema or Chronic bronchitis by V2.

If baseline COPE_EVER missing and AFUCOPD_Self_ByV2 missing and MHE12 ne (1 or 0) then set COPD_EVER_V2 to missing;
Otherwise if baseline COPE_EVER=1 OR AFUCOPD_Self_ByV2=1 OR MHE12=1 then COPD_EVER_V2=1;
Otherwise set COPD_EVER_V2=0;

Response format: 0 = No
1 = Yes

Source variable(s):
COPD_EVER. Self-report COPD/emphysema or CB in V1.

AFUCOPD_Self_byV2. Self-report COPD/emphysema or CB in AFU visit by V2.
MHE12. Self-report COPD/emphysema or CB at visit 2.

7.6 Charlson_V2 (Charlson comorbidity index at Visit 2)

This is a numeric variable. This index of comorbidity in HCHS based on the Charlson Index which is cited in the medical literature. The index is a weighted summary measure of the presence of several comorbid conditions. An index can be created based on the baseline HCHS medical history and derived variables, similarly one can be produced from the Visit 2 medical history and its related derived variables. The presence of disease signs and symptoms can change over time so the computed index can vary from V1 to V2. The index is the sum of the weighted (0/1) indicators for the presence of a condition.

Conditions	**Charlson Weight**	**V1 Definition**	**V2 Definition**
1. Myocardial infarct	1	MHEA4=1	MHE1=1
2. Congestive heart failure	1	MHEA5=1	MHE4=1
3. Peripheral vascular disease	1	MHEA14=1	MHE10=1
4. Cerebrovascular disease	1	STROKE_TIA=1	MHE5=1 or MHE6=1
5. Dementia or MCI	1	0 < TOTAL_6item < 3	MHE23A=1
6. Chronic pulmonary disease	1	RSEA43=1 or RSEA44=1	COPD_EVER_V2=1
7. Connective tissue disease	1	MHEA23=1	n/a at Visit 2
8. Ulcer disease or GERD	1	MHEA19=1 or MHEA20=1	n/a at Visit 2
9. Mild liver disease	1	MHEA18=1 AND MHEA18C=0	MHE11=1 AND MHE11C=0
10. Diabetes	1	DIABETES2=3	DIABETES4_V2=3
11. Moderate or severe renal disease	2	CKD in(3,4,5)	CKD_V2 in(3,4,5)
12. Diabetes with end organ damage	2	DIABETES2=3 AND CKD in(4,5)	DIABETES4_V2=3 AND CKD_V2 in(4,5)
13. Any tumor	2	MHEA26=1	MHE17=1
14. Leukemia or Lymphoma	2	MHEA26A4=1	LeuLym_V2=(MHE17A4 =4)
15. Moderate or severe liver disease	3	MHEA18=1 AND MHEA18C=1	MHE11=1 AND MHE11C=1

CHARLSON_V2 = sum of (condition 1* Charlson wt, condition 2*Charlson wt, … condition 15*Charlson wt)

Source variable(s):
For MHE variables: Since the first SOL visit, has a doctor said that you had any of the following medical problems:
MHE1. Heart attack
MHE4. Heart Failure
MHE5. Stroke
MHE6. A mini-stroke or TIA (transient ischemic attack)
MHE10. Peripheral arterial disease (problems with circulation, blocked arteries to the legs)
MHE11. Liver disease
MHE11C. Cirrhosis (type of liver disease)
COPD_Ever_V2. Self-reported non-TB emphysema, chronic bronchitis, or COPD by V2.
MHE17. Cancer or a malignant tumor
MHEA17A4. Blood/lymph glands (type of cancer/tumor)
MHE23A. Dementia
DIABETES4_V2. 3-level grouped Diabetes based on self-reported antidiabetic med use and ADA criteria, V2
CKD_V2. Chronic Kidney Disease using eGFR only (NIDDK) at Visit 2

7.7 WAIST_HIP_V2 (Waist to hip ratio at Visit 2)

We want to create a numerical variable providing the waist to hip ratio. This is obtained by simply dividing the waist girth by the hip girth. If either of these two values is missing then the ratio is considered to be missing.

This is a numeric variable.

Missing if either waist girth or hip girth is missing.

If neither waist girth nor hip girth are missing then WAIST_HIP_V2 = ANT10A / ANT10B.

Source variable(s):
ANT10A. Waist girth measured in centimeters
ANT10B. Hip girth measured in centimeters

8. BLOOD PRESSURE MEASURES

8.1 Hypertension2_V2 (Hypertension using NHANES definition, Visit 2)

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

Response format: 0=Not hypertensive (No)

1=Hypertensive (Yes)

Source variable(s):
SBP5. Systolic blood pressure.
SBP6. Diastolic blood pressure.
MUE26D. Were any of the medications you took during the last 4 weeks for high blood pressure or hypertension?
Note: MUE26D missing had been reset to 0 when MUE2=1 (confirmed not taking any medications).

8.2 Hypertension_C4_V2 (4-level grouped hypertension, Visit 2)

Classifies participants into 4 categories based on hypertension and hypertensive medication use: no hypertension, pre-hypertension, treated hypertension, and untreated hypertension

HYPERTENSION2_V2=0 and ((120<=SPBA5<140) or (80<=SPBA6<90)) → 2
Other HYPERTENSION2_V2=0 → 1
HYPERTENSION2_V2=1 and MUE26D=1 → 3
Other HYPERTENSION2_V2=1 → 4
Missing HYPERTENSION2_V2 → missing

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

Source variable(s):
Hypertension2_V2. Hypertension using NHANES definition.
SBP5. Systolic blood pressure.
SBP6. Diastolic blood pressure.
MUE26D. Were any of the medications you took during the last 4 weeks for high blood pressure or hypertension?
Note: MUE26d missing had been reset to 0 when MUE2=1 (confirmed not taking any medications).

8.3 Hypertension2_AHA_V2 (Hypertension using New ACC/AHA definition, Visit 2)

This is a 0/1 numeric variable. Hypertension is defined here using the following new ACC/AHA definition: if the systolic or diastolic BP is greater than or equal to 130/80 or if the participant self-reported as currently taking antihypertensive mediations. Participants without a blood pressure measurement and no medication use were assumed to be not hypertensive.

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

Source variable(s):
SBP5. Systolic blood pressure.
SBP6. Diastolic blood pressure.
MUE26D. Were any of the medications you took during the last 4 weeks for high blood pressure or hypertension?
Note: MUE26D missing had been reset to 0 when MUE2=1 (confirmed not taking any medications).

8.4 Hypertension_AHA_C5_V2 (5-level grouped hypertension using new ACC/AHA definition, Visit 2)

Classifies participants into 5 categories based on hypertension_AHA_C2 and hypertensive medication use: no hypertension, pre-hypertension, treated hypertension, and untreated hypertension

IF HYPERTENSION2_AHA_V2=0 and ((120<=SBP5<130) AND SBP6<80) → 2
Otherwise HYPERTENSION2_AHA_V2=0 → 1
HYPERTENSION2_AHA_V2=1 and MUE26D=1 → 3
Otherwise if HYPERTENSION2_AHA_V2=1 and ((SBP5>=140) OR (SBP6>=90)) → 5
Otherwise if HYPERTENSION2_AHA_V2=1 → 4
Missing HYPERTENSION2_AHA_V2 → missing

Response format: 1 = No hypertension
2 = Elevated-hypertension
3 = Treated hypertension
4 = Untreated Stage-1 hypertension
5 = Untreated Stage-2 hypertension

Source variable(s):
Hypertension2_V2. Hypertension using New ACC/AHA definition.
SBP5. Systolic blood pressure.
SBP6. Diastolic blood pressure.
MUE26D. Were any of the medications you took during the last 4 weeks for high blood pressure or hypertension?
Note: MUE26D missing had been reset to 0 when MUE2=1 (confirmed not taking any medications).

New ACC/AHA Guideline reference: https://www.acc.org/latest-in-cardiology/articles/2017/11/08/11/47/mon-5pm-bp-guideline-aha-2017

8.5 AFUHypert_self_byV2 (Self-reported hypertension only in AFU by V2)

This is a 0/1 numeric variable that indicates a self-report of hypertension only reported in AFU by V2.

Among participants who did not have Hypertension awareness in visit 1:
if any OPE8=1 by Visit 2 visit, then AFUHypert_SELF_byV2=1;

Response format: 0 = No
1 = Yes

Source variable(s):
OPE8. Self-report hypertension in AFU visit.

8.6 Hypert_Awareness_V2 (Awareness of Hypertension – NHANES, by Visit 2)

This is a 0/1 numeric variable and is defined by hypertension awareness at baseline, or the self-report of a doctor's diagnosis of hypertension in AFU and Visit 2 (MHE15=1) for those who were defined as having hypertension using the NHANES definition (HYPERTENSION2_V2=1) only.

Response format: 0=Not aware (No)
1=Aware (Yes)

Source variable(s):
Hypertension2_V2. Hypertension using NHANES definition at V2.
Hypert_awareness. Hypertension awareness at baseline.
AFUHypert_Self_by V2. Self-reported hypertension only in AFU by Visit 2.
MHE15. Since our last telephone interview with you, has a doctor or health professional told you that you had high blood pressure or hypertension? (at V2)

8.7 Hypert_Treatment_V2 (Treatment of Hypertension – NHANES, Visit 2)

This is a 0/1 numeric variable and is defined by the participant's self-report awareness of hypertension (HYPERT_AWARENESS_V2) and self-report use of antihypertensive medications (MUE26D=1) for those who were defined as having hypertension using the NHANES definition (HYPERTENSION2_V2=1) only. Note that participants who had missing data for hypertensive medication use were set to 0 (no medication use).

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

Source variable(s):
Hypertension2_V2. Hypertension using NHANES definition.
Hypert_Awareness_V2. Awareness of Hypertension using NHANES definition.
MUE26D. Were any of the medications you took during the last 4 weeks for high blood pressure or hypertension?
Note: MUE26d missing had been reset to 0 when MUE2=1 (confirmed not taking any medications).

8.8 Hypert_Control_V2 (Control of Hypertension – NHANES, Visit 2)

This is a 0/1 numeric variable and is defined for participants whose blood pressure are below the cutoff of 140/90 (both systolic and diastolic must be less than the cutoff values) for those who were defined as having hypertension using the NHANES definition (HYPERTENSION2_V2=1) only and also indicated whether/not they were receiving treatment for hypertension (HYPERT_TREATMENT_V2= 0 or 1).

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

Source variable(s):
HYPERTENSION2_V2. Hypertension using NHANES definition
HYPERT_TREATMENT_V2. Treatment of Hypertension - NHANES
SBPA5. Systolic blood pressure. SBPA6. Diastolic blood pressure.

9. LABORATORY MEASURES

9.1 Dyslipidemia_V2 (High LDL, low HDL, or high triglycerides – NIH at Visit 2)

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

If laba69≤.Z & laba68≤.Z & laba67≤.Z then DYSLIPIDEMIA_V2= missing.
Else if laba69 ≥ 160 OR 0≤laba68<40 OR laba67 ≥ 200 then DYSLIPIDEMIA_V2=1
Else DYSLIPIDEMIA_V2=0.

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

Source variable(s):
LABA67. Triglycerides (mg/dL)
LABA68. HDL - cholesterol (mg/dL)
LABA69. LDL - Cholesterol (mg/dL)

9.2 Dyslipidemia_C3_V2 (Dyslipidemia Status at Visit 2)

This is a 3 category variable that defines whether a participant has dyslipidemia, and if so what the treatment status is of the condition.

```
if dyslipidemia_v2<=.Z | reported_chol_meds<=.Z then DYSLIPIDEMIA_C3_V2=.;
else if dyslipidemia_v2=0 then DYSLIPIDEMIA_C3_V2=1;
```

else if dyslipidemia_v2=1 & reported_chol_meds=0 then DYSLIPIDEMIA_C3_V2=2;
else if dyslipidemia_v2=1 & reported_chol_meds=1 then DYSLIPIDEMIA_C3_V2=3;

Response format: 1 = Not dyslipidemic (No)
2 = Dyslipidemic without treatment
3 = Dyslipidemic undergoing treatment

Source variable(s):
Dyslipidemia_v2. Dyslipidemia status at visit 2
MUE26E (through reported_chol_meds). Self-reported high blood cholesterol medication intake
Note: MUE26E missings had been reset to 0 when MUE2=1 (confirmed not taking any medications).

9.3 Cardiac_risk_ratio_V2 (Cardiac Risk Ratio (TC/HDL) – AHA at Visit 2)

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

If laba66 and laba68 are both non-missing, then:
CARDIAC_RISK_RATIO_V2=laba66/laba68.

Source variable(s):
LABA66. Total Cholesterol (mg/dL)
LABA68. HDL - cholesterol (mg/dL)

9.4 Dyslipidemia_TCHDL_V2 (Dyslipidemia based on TCHDL>5 at Visit 2)

This is a 0/1 variable that checks for dyslipidemia as measured by a high total cholesterol to HDL cholesterol ratio.

if cardiac_risk_ratio_v2<=.Z then DYSLIPIDEMIA_TCHDL_V2=.;
else if cardiac_risk_ratio_v2>5 then DYSLIPIDEMIA_TCHDL_V2=1;
else DYSLIPIDEMIA_TCHDL_V2=0;

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

Source variable(s):
Cardiac_risk_ratio_v2. Cardiac Risk Ratio (TC/HDL) – AHA at Visit 2

9.5 Dys_TCHDL_Med_V2 (Dyslipidemia based on TCHDL>5 or lipid reduction drug use at Visit 2)

This is a 0/1 variable that checks for dyslipidemia as measured by either a high total cholesterol to HDL ratio, or by self-reported high blood cholesterol medication intake.

```
   if dyslipidemia_tchdl_v2<=.Z & reported_chol_meds <=.Z then
DYS_TCHDL_MED_V2=.;
   else if dyslipidemia_tchdl_v2=1 | reported_chol_meds =1 then
DYS_TCHDL_MED_V2=1;
   else DYS_TCHDL_MED_V2=0;
```

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

Source variable(s):
Dyslipidemia_tchdl_v2. Dyslipidemia based on TCHDL>5 at Visit 2
MUE26E (through reported_chol_meds). Self-reported high blood cholesterol medication intake
Note: MUE26E missing had been reset to 0 when MUE2=1 (confirmed not taking any medications).

9.6 TOTAL_CHOL_C3_V2 (Total Cholesterol at Visit 2)

This is a 3 category variable defining a participant's cholesterol status.

```
   if .Z<laba66<200 then TOTAL_CHOL_C3_V2_V2=1;
   else if 200<=laba66<=239 then TOTAL_CHOL_C3_V2=2;
   else if laba66>=240 then TOTAL_CHOL_C3_V2=3;
   else TOTAL_CHOL_C3_V2=.;
```

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

Source variable(s):
LABA66. Total Cholesterol (mg/dL)

9.7 Insulin_Fast (Insulin, fasting (converted to mU/L))

This variable is defined using fasting glucose results from the central lab. It is converted from pmol/L to mU/L by dividing the fasting insulin variable by 6.

Source variable(s):
LABA96. Fasting Insulin (pmol/L)

9.8 Fasttime_V2 (Fasting time (hrs.))

This variable combines fasting time variables collected from the Biospecimen Collection (BIO) form, as follows (for non-missing BIO7):

```
if bio7 > .Z then do;
   if (bio7=1) | (bio8='0:00't) then t=0;
   else if bio7=2 then t=1;
   else if bio7=3 then t=2;
   FASTTIME_V2=((((bio10+(t*'24:00't))-bio8)/60)/60);
End;
```

Source variable(s):
BIO7. On which day did you last eat or drink anything except water: today, yesterday, or the day before yesterday?
BIO8. And at what time was that?
BIO10. Collection time

9.9 OGTT_Time_V2 (Elapsed time before OGTT blood draw)

This variable calculates the elapsed time (in hrs) between fasting blood draw and OGTT blood draw time variables collected from the Biospecimen Collection (BIO) form, as follows:

```
OGTT_TIME_V2 = (bio26-bio25)/(60);
```

Source variable(s):
BIO25. Time glucola given
BIO26. Time of collection of post-glucola samples

9.10 GFR_V2 (Insulin, fasting (calibrated, converted to mU/L))

This is a numeric variable that calculates estimates eGFR. Units for estimated GRF are 60mL/min/1.73m2. It will only calculate GFR if age, gender race and Creatinine (LABA76) are numeric, as follows:

```
 IGRFFactor=1;
 If gender_v2 = 'M' then IGRFFactor = 1;
 If gender_v2 = 'F' Then
   IGRFFactor = IGRFFactor * 0.742;

 If race = 4 Then
   IGRFFactor = IGRFFactor * 1.210;

 GFR_V2 = 175 * (laba76**-1.154)*(Age_V2**-0.203) * (IGRFFactor);
```

Source variable(s):
LABA76. Creatinine (mg/dL) from Visit 2
GENDER_V2. Gender at Visit 2
AGE_V2. Age at Visit 2
RACE. Race Group (self-report) from Visit 1

9.11 CKD_V2 (Chronic Kidney Disease using eGFR only – NIDDK at Visit 2)

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

if GFR_V2≥90 then CKD_V2=1 ('Normal')
else if 60≤GFR_V2<90 then CKD_V2=2 ('Mild')
else if 30≤GFR_V2<60 then CKD_V2=3 ('Moderate')
else if 15≤GFR_V2<30 then CKD_V2=4 ('Severe')
else if .Z<GFR_V2<15 then CKD_V2=5 ('End-Stage')

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

Source variable(s):
GFR_V2. Glomerular Filtration Rate, estimated

9.12 GFRscys_V2 (eGFR based on serum cystatin C w/o demographics at Visit 2)

Estimated Glomerular filtration rate at visit 2, based on serum cystatin C and no demographic factors.

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

$GFRscys\_V2 = 76.7 * LABA101^{-1.19}$

Source variable(s):
LABA101. Cystatin C (mg/L) from visit 2

9.13 GFRscys_scr_V2 (eGFR based on serum cystatin C, serum creatinine, gender, age and race at Visit 2)

Estimated Glomerular filtration rate at visit 2, based on serum cystatin C, serum creatinine, and demographics gender, age and race (black vs. nonblack).

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

If (GENDER_V2=F and RACE=4) then $135 * \min(LABA76/0.7, 1)^{-0.248} * \max(LABA76/0.7, 1)^{-0.601} * \min(LABA101/0.8, 1)^{-0.375} * \max(LABA101/0.8, 1)^{-0.711} * 0.995^{AGE\_V2} * 0.969 * 1.08$

If (GENDER_V2=F and RACE≠4) then $135 * \min(LABA76/0.7, 1)^{-0.248} * \max(LABA76/0.7, 1)^{-0.601} * \min(LABA101/0.8, 1)^{-0.375} * \max(LABA101/0.8, 1)^{-0.711} * 0.995^{AGE\_V2} * 0.969$

If (GENDER_V2=M and RACE=4) then 135 * min(LABA76/0.9, 1)$^{-0.207}$ *
max(LABA76/0.9, 1)$^{-0.601}$ * min(LABA101/0.8, 1)$^{-0.375}$ * max(LABA101/0.8, 1)$^{-0.711}$ *
0.995^{AGE_V2} *1.08
If (GENDER_V2=M and RACE≠4) then 135 * min(LABA76/0.9, 1)$^{-0.207}$ *
max(LABA76/0.9, 1)$^{-0.601}$ * min(LABA101/0.8, 1)$^{-0.375}$ * max(LABA101/0.8, 1)$^{-0.711}$ *
0.995^{AGE_V2}

Source variable(s):
GENDER_V2. Gender at Visit 2
RACE. Race Group (self-report) from Visit 1
AGE_V2. Age at Visit 2
LABA76. Creatinine (mg/dL) from Visit 2
LABA101. Cystatin C (mg/L) from Visit 2

9.14 GFRscr_V2 (eGFR based on serum creatinine, gender, age, and race)

Estimated Glomerular filtration rate at Visit 2, based on serum creatinine, and demographics gender, age and race (black vs. nonblack).

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

If (GENDER_V2=F and RACE=4) then 141 * min(LABA76/0.7, 1)$^{-0.329}$ *
max(LABA76/0.7, 1)$^{-1.209}$ * 0.993^{AGE_V2} *1.018 *1.159
If (GENDER_V2=F and RACE≠4) then 141 * min(LABA76/0.7, 1)$^{-0.329}$ *
max(LABA76/0.7, 1)$^{-1.209}$ * 0.993^{AGE_V2} *1.018
If (GENDER_V2=M and RACE=4) then 141 * min(LABA76/0.9, 1)$^{-0.411}$ *
max(LABA76/0.9, 1)$^{-1.209}$ * 0.993^{AGE_V2} *1.159
If (GENDER_V2=M and RACE≠4) then 141 * min(LABA76/0.9, 1)$^{-0.411}$ *
max(LABA76/0.9, 1)$^{-1.209}$ * 0.993^{AGE_V2}

Source variable(s):
GENDER_V2. Gender at Visit 2
RACE. Race Group (self-report) from Visit 1
AGE_V2. Age at Visit 2
LABA76. Creatinine (mg/dL) from Visit 2

9.15 GFRscr_A_V2 (eGFR based on serum creatinine, gender, age and not race at Visit 2)

Estimated Glomerular filtration rate at Visit 2, based on serum creatinine, and demographics gender and age only, and ignores race.

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

If (GENDER_V2=F) then $141 * \min(LABA76/0.7, 1)^{-0.329} * \max(LABA76/0.7, 1)^{-1.209} * 0.993^{AGE\_V2} * 1.018$
If (GENDER_V2=M) then $141 * \min(LABA76/0.9, 1)^{-0.411} * \max(LABA76/0.9, 1)^{-1.209} * 0.993^{AGE\_V2}$

Source variable(s):
GENDER_V2. Gender at Visit 2
AGE_V2. Age at Visit 2
LABA76. Creatinine (mg/dL) from visit 2

9.16 EGFR_LT60_V2 (2-level grouped eGFR based on serum creatinine, gender, age and not race at Visit 2 (GFRSCR_a<60))

This binary variable determines if the estimated glomerular filtration rate at visit 2, based on serum creatinine, and demographics gender and age only, and ignores race, is less than 60 mg/dL (=1).

if 0<=GFRscr_a_V2<60 then EGFR_LT60_V2=1;
else if GFRscr_a_V2>=60 then EGFR_LT60_V2=0;

Source variable(s):
GFRscr_A_V2. eGFR based on serum creatinine, gender and age only at visit 2

9.17 AC_Ratio_GT30_V2 (2-level grouped serum albumin-creatinine ratio at Visit 2 (laba81>=30))

This binary variable determines if the serum albumin-creatinine ratio at Visit 2 is greater than 30 mg/dL (=1).

if laba81>=30 then AC_RATIO_GT30_V2=1;
else if 0<=laba81<30 then AC_RATIO_GT30_V2=0;

Source variable(s):
LABA81. Urine Albumin/creatinine ratio from Visit 2

9.18 CKD2_V2 (Chronic Kidney Disease using eGFR only - NIDDK)

This variable is estimated based on the estimated GFR value and serum albumin-creatinine ratio at Visit 2, as follows:

if EGFR_LT60_V2=1 | AC_RATIO_GT30_V2=1 then CKD2_V2=1;
else if EGFR_LT60_V2<=.Z & AC_RATIO_GT30_V2<=.Z then CKD2_V2=.;
else CKD2_V2=0;

Response format: 0=No CKD
1=CKD

Source variable(s):
EGFR_LT60_V2. 2-level grouped eGFR based on serum creatinine, gender, age and not race at Visit 2 (GFRSCR_a<60)
AC_RATIO_GT30_V2. 2-level grouped serum albumin-creatinine ratio at Visit 2 (laba81>=30)

9.19 Incident_GFR_V1V2 (Incident eGFR (>=60 in V1,<60 in V2))

This is a binary variable that determines the incidence of low eGFR between Visit 1 and visit 2, as follows:

if EGFR_LT60=0 & EGFR_LT60_V2=1 then INCIDENT_GFR_V1V2=1;
else if EGFR_LT60<=.Z | EGFR_LT60_V2<=.Z then INCIDENT_GFR_V1V2=.;
else INCIDENT_GFR_V1V2=0;

Response format: 0=No
1=Yes

Source variable(s):
EGFR_LT60. 2-level grouped eGFR based on serum creatinine, gender, age and not race at Visit 1 (GFRSCR_a<60)
EGFR_LT60_V2. 2-level grouped eGFR based on serum creatinine, gender, age and not race at Visit 2 (GFRSCR_a<60)

9.20 Incident_ACR_V1V2 (Incident Albumin-creatinine ratio (<30 in V1,>=30 in V2))

This is a binary variable that determines the incidence of high serum albumin-creatinine ratio between Visit 1 and visit 2, as follows:

if AC_RATIO_GT30=0 & AC_RATIO_GT30_V2=1 then INCIDENT_ACR_V1V2=1;
else if AC_RATIO_GT30<=.Z | AC_RATIO_GT30_V2<=.Z then INCIDENT_ACR_V1V2=.;
else INCIDENT_ACR_V1V2=0;

Response format: 0=No
1=Yes

Source variable(s):
AC_RATIO_GT30. 2-level grouped serum albumin-creatinine ratio at Visit 1 (laba81>=30)
AC_RATIO_GT30_V2. 2-level grouped serum albumin-creatinine ratio at Visit 2 (laba81>=30)

9.21 Incident_CKD_V1V2 (Chronic Kidney Disease using eGFR only - NIDDK)

This is a binary variable that determines the incidence of low eGFR and/or high serum albumin-creatinine ratio between Visit 1 and Visit 2 as well as incorporating a GFR decreasing rate of 1+ per year among those without chronic kidney disease at baseline (CKD2=0), as follows:

GFRSCR_DIFF=GFRSCR_a_V2-GFRSCR_a;

```
GFRSCR_YEAR=GFRSCR_DIFF/YRS_BTWN_V1V2;
if GFRSCR_YEAR>.Z then do;
  if GFRSCR_YEAR<=-1 then GFR_PLUS1=1;
  else GFR_PLUS1=0;
end;

if CKD2=0 then do;
   if INCIDENT_GFR_V1V2=1 & GFR_PLUS1=1 then INCIDENT_CKD_V1V2=1;
   else if INCIDENT_ACR_V1V2=1 then INCIDENT_CKD_V1V2=1;
   else if INCIDENT_GFR_V1V2<=.Z & INCIDENT_ACR_V1V2<=.Z then
INCIDENT_CKD_V1V2=.;
   else INCIDENT_CKD_V1V2=0;
end;
```

Response format: 0=No
1=Yes

Source variable(s):
GFRSCR_a. eGFR based on serum creatinine, gender, age and not race at Visit 1
GFRSCR_a_V2. eGFR based on serum creatinine, gender, age and not race at Visit 2
YRS_BTWN_V1V2. Elapsed time between visits 1 and 2 (yrs))
CKD2. Chronic Kidney Disease using eGFR (creatinine only, no race) and albumin-creatinine ratio (NIDDK) at Visit 1
INCIDENT_GFR_V1V2. Incident eGFR (>=60 in V1,<60 in V2
INCIDENT_ACR_V1V2. Incident Albumin-creatinine ratio (<30 in V1,>=30 in V2)

9.22 GFR_PLUS1_V1V2 (eGFR decline by average>=1/year between V1 and V2)

Estimated Glomerular filtration decline by average>=1/year between V1 and V2.

Define GFRSCR_DIFF = GFRSCR_a_V2 - GFRSCR_a
Define GFRSCR_YEAR = GFRSCR_DIFF/YRS_BTWN_V1V2

Then define GFR_PLUS1_V1V2 as follows:

If missing(GFRSCR_YEAR) = 0 then do (for all non-missing GFRSCR_YEAR):
IF GFRSCR_YEAR <= -1 then set GFR_PLUS1_V1V2 = 1
Otherwise then set GFR_PLUS1_V1V2 = 0

Source variable(s):
GFRSCR_A_V2. eGFR based on serum creatinine, gender, age and not race at Visit 2.
GFRSCR_A. eGFR based on serum creatinine, gender, age and not race at baseline
YRS_BTWN_V1V2. Years between baseline to Visit 2 visit.

10. CLINICAL CHARACTERISTICS

10.1 Diabetes_Lab_V2 (3-level grouped Diabetes - Lab, V2)

This is a numeric variable that describes the stages of diabetes at visit 2. This definition takes into account serum glucose levels adjusted for fasting time and, if available, post-OGTT glucose levels, and A1C percentages. This variable does NOT include any medication use variables.

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

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

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

Source variable(s):
LABA70. Fasting glucose (mg/dL)
LABA71. Glucose, post OGGT (mg/dL)
LABA72. % Glycosylated Hemoglobin (A1C)
FASTTIME_V2. Elapsed time (hrs) between the time the participant last consumed anything and the blood draw

10.2 SelfMed_Antidiab_V2 (Anti-Diabetics Medication used – Self-reported, V2)

This is a 0/1 numeric variable and is defined by the participant's self-report use of anti-diabetics medications (Yes if MUE26C=1, No if MUE26c=0, after reset missing to zero when confirmed not taking any medications, MUE2=1). Note that participants who had missing data or unsure (MUE26c=9) for antidiabetics medication use were set to missing.

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

Source variable(s):
MUE26C. Were any of the medications you took during the last 4 weeks for high blood sugar or diabetes?
MUE2. Not bringing any of the medications you took during the last 4 weeks, because you forgot/unable to, or took no medication?

10.3 GESTATIONAL_DM_V2 (Date-validated Gestational Diabetes flag, by Visit 2)

This is a 0/1 numeric intermediate variable that flags the participants reported gestational diabetes (GDM) in PCE and without any Diabetes reported after the last GDM. NOTE: it is NOT an indicator of whether the women had a pregnancy with gestational diabetes in between visits; if interested in gestational diabetes look at the PCE dataset.

if any PCE1g=1 and PCE1g2=0 then GESTATIONAL_DM_V2=1;
else set= 0.

When a self-reported diabetes happened in AFU or V2 after the last gestational diabetes report date, reset Gestational_DM_V2 = 0.

Response format: 0 = No
1 = Yes

Source variable(s):
PCE1g. Had diabetes during pregnancy.
PCE1g2. Had diabetes before this pregnancy.
PCE1B. BOD or end of pregnancy date.
DMSelf_LL_DATE. Self-reported diabetes by V2 interval lower limit date.

10.4 AFUDM_self_byV2 (Self-reported Diabetes flag in AFU by V2, excluded gestational only diabetes)

This is a 0/1 numeric intermediate variable that indicates a self-report of diabetes only reported from annual follow-up 1 to 8 before V2.

Among participants who did not have self-reported diabetes in visit 1:
if any OPE7=1 by Visit 2 visit, then AFUDM_SELF_byV2=1;

Among the participant pool with gestational diabetes reported in PCE data, AFU self-reported diabetes happened before or same time as the last PCE confirmed no diabetes before pregnancy date had been reset to 0.

Response format: 0 = No
1 = Yes

Source variable(s):
OPE7. Self-report diabetes in AFU visit.
PCE1G2. No diabetes before pregnancy.
PCE1B. BOD or end of pregnancy date.
Gestational_DM_V2: the confirmed gestational diabetes.

10.5 Diabetes_self_V2 (Diabetes - self report prevalent cases by visit 2)

This is a 0/1 numeric variable that indicates all the prevalent self-reported cases of diabetes by Visit 2. In other words, it includes those who self-reported as diabetics at

baseline, or during annual follow-up calls (AFU 1 to 8 before V2), or at Visit 2. It removes gestational diabetes identified from PCE (Pregnancy Complications questionnaire) when the date of the pregnancy (PCE1B) with gestational diabetes (PCE1G=1) is between the interval of the diabetes reported from AFU or Visit 2. For example, if someone reported diabetes at AFU4 and the PCE1B is in between AFU3 and AFU4 then it is no longer consider diabetes. Similarly, we further excluded those self-reported diabetes cases when in the last pregnancy of the PCE form they indicated that before that last pregnancy they have never had diabetes before (i.e. PCE1G2=0).

if (DIABETES_SELF=1 | AFUDM_SELF_BYV2 =1 | MHE14=1) AND
GESTATIONAL_DM_V2 NE 1 then DIABETES_SELF_V2=1;
else if INMHE=0 then DIABETES_SELF_V2=.;
else DIABETES_SELF_V2=0;

Response format: 0 = No
1 = Yes

Source variable(s):
DIABETES_SELF. Diabetes - self report only from Visit 1
AFUDM_SELF_BYV2. Self-reported Diabetes flag only in AFU by V2.
GESTATIONAL_DM_V2. Gestational Diabetes flag in Visit 2.
MHE14. Self-report diabetes or high blood sugar – from Visit 2

10.6 Diabetes3_V2 (3-level grouped Diabetes based on self-reported diabetes and ADA lab criteria, by V2)

This is a numeric variable that describes the stages of diabetes based on American Diabetes Association definition (*Diabetes Care* 2010;33:S62-69) with an additional self-report criterion. This definition takes into account serum glucose levels adjusted for fasting time and, if available, post-OGTT glucose levels, A1C percentages, and self-report of diabetes by Visit 2.

if DIABETES_SELF_V2=1 | DIABETES_LAB_V2=3 then DIABETES3_V2=3;
else if DIABETES_LAB_V2=2 then DIABETES3_V2=2;
else if DIABETES_SELF_V2=. & DIABETES_LAB_V2 in (1,.) then DIABETES3_V2=.;
else DIABETES3_V2=1;
Response format: 1 = Non-diabetic
2 = Pre-diabetic
3 = Diabetic

Note: this diabetes variable was defined by the algorithm strictly. Per Diabetes SIG decision and memo, Diabetes definition 3 classification at baseline is not part of definition specs and the HCHS/SOL study lab classification at baseline was not considered as a formal diagnosis.

Source variable(s):
DIABETES_LAB_V2. 3-level grouped Diabetes – Lab criterion only, V2
DIABETES_SELF_V2. Diabetes – self-report only V2

10.7 Diabetes4_V2 (3-level grouped Diabetes based on self-reported antidiabetics med and ADA lab criteria, V2)

This is a numeric variable that describes the stages of diabetes based on American Diabetes Association definition with an additional self-report antidiabetic medication information. This definition takes into account serum glucose levels adjusted for fasting time and, if available, post-OGTT glucose levels, A1C percentages, and self-report of antidiabetics med at Visit 2.

if SelfMed_Antidiab_V2=1 | DIABETES_LAB_V2=3 then DIABETES4_V2=3;
else if DIABETES_LAB_V2=2 then DIABETES4_V2=2;
else if SelfMed_Antidiab_V2 & DIABETES_LAB_V2 both missing then DIABETES4_V2=.;
else DIABETES4_V2=1;
Response format: 1 = Non-diabetic
2 = Pre-diabetic
3 = Diabetic

Note: this diabetes variable was defined by the algorithm strictly. Per Diabetes SIG decision and memo, Diabetes definition 4 classification at baseline is not part of definition specs and the HCHS/SOL study lab classification at baseline was not considered as a formal diagnosis.

Source variable(s):
DIABETES_LAB_V2. 3-level grouped Diabetes – Lab criterion only, V2
SelfMed_Antidiab_V2. Self-reported Antidiabetic Medication use at Visit 2.

10.8 Diabetes5_V2 (3-level grouped Diabetes based on Self-reported diabetes, self-reported antidiabetics med and ADA lab criteria, by V2)

This is a numeric variable that describes the stages of diabetes based on American Diabetes Association definition with an additional self-reported diabetes diagnosis and self-report antidiabetic medication information by Visit 2.

if DIABETES_SELF_V2=1 | SelfMed_Antidiab_V2=1 | DIABETES_LAB_V2=3 then DIABETES5_V2=3;
else if DIABETES_LAB_V2=2 then DIABETES5_V2=2;
else if DIABETES_SELF_V2 & SelfMed_Antidiab_V2 & DIABETES_LAB_V2 all missing then DIABETES5_V2=.;
else DIABETES5_V2=1;
Response format: 1 = Non-diabetic
2 = Pre-diabetic
3 = Diabetic

Note: this diabetes variable was defined by the algorithm strictly. Per Diabetes SIG decision and memo, Diabetes definition 5 classification at baseline is not part of definition specs and the HCHS/SOL study lab classification at baseline was not considered as a formal diagnosis.

Source variable(s):
DIABETES_LAB_V2. 3-level grouped Diabetes – Lab criterion only, V2
DIABETES_SELF_V2. Diabetes – self-report only V2
SelfMed_Antidiab_V2. Self-reported Antidiabetic Medication use at Visit 2.

10.9 Diabetes3_C4_V2 (4-level grouped Diabetes based on Diabetes Definition 3, by Visit 2)

This is a numeric variable that assesses diabetes using Diabetes Definition 3 classification by visit 2, and medication use based on MUE26C (self-reported medication for high blood sugar or diabetes).

DIABETES3_V2=1 → 1 (normal)
DIABETES3_V2=2 → 2 (pre-diabetics)
DIABETES3_V2=3 and (MUE26C =1) → 3 (diabetes with treatment)
DIABETES3_V2=3 and (MUE26C NE 1) → 4 (diabetes without treatment)
Missing DIABETES3 → missing

Source variable(s):
DIABETES3_V2. 3-level grouped Diabetes definition 3 classification by Visit 2.
MUE26C. Self-reported anti-diabetic medication use.
Note: MUE26C missing had been reset to 0 when MUE2=1 (confirmed not taking any medications).

10.10 Diabetes4_C4_V2 (4-level grouped Diabetes based on Diabetes Definition 4, at Visit 2)

This is a numeric variable that assesses diabetes using Diabetes Definition 4 classification at visit 2, and medication use based on MUE26C (self-reported medication for high blood sugar or diabetes).

DIABETES4_V2=1 → 1 (normal)
DIABETES4_V2=2 → 2 (pre-diabetics)
DIABETES4_V2=3 and (MUE26C =1) → 3 (diabetes with treatment)
DIABETES4_V2=3 and (MUE26C NE 1) → 4 (diabetes without treatment)
Missing DIABETES4_V2 → missing

Source variable(s):
DIABETES4_V2. 3-level grouped Diabetes definition 4 classification at Visit 2.
MUE26C. Self-reported anti-diabetic medication use.
Note: MUE26C missing had been reset to 0 when MUE2=1 (confirmed not taking any medications).

10.11 Diabetes5_C4_V2 (4-level grouped Diabetes based on Diabetes Definition 5, by Visit 2)

This is a numeric variable that assesses diabetes using Diabetes Definition 5 classification by visit 2, and medication use based on MUE26C (self-reported medication for high blood sugar or diabetes).

DIABETES5_V2=1 → 1 (normal)
DIABETES5_V2=2 → 2 (pre-diabetics)
DIABETES5_V2=3 and (MUE26C =1) → 3 (diabetes with treatment)
DIABETES5_V2=3 and (MUE26C NE 1) → 4 (diabetes without treatment)
Missing DIABETES5_V2 → missing

Source variable(s):
DIABETES5_V2. 3-level grouped Diabetes definition 5 classification by Visit 2.
MUE26C. Self-reported anti-diabetic medication use.
Note: MUE26C missing had been reset to 0 when MUE2=1 (confirmed not taking any medications).

10.12 METS_IDF3_V2 (Metabolic Syndrome w/o Scanned Meds for Trig&HDL - IDF definition 3)

This is a 0/1 numeric variable with 1=metabolic syndrome present and 0=no metabolic syndrome in Visit 2.

The International Diabetes Federation (IDF) provides the following definition of metabolic syndrome. The IDF definition emphasizes central obesity because it is "independently" associated with each of the other metabolic syndrome components, as well as insulin resistance. Another difference between the IDF and NCEP-ATP III definitions is the recommended use of ethnic-specific waist girth thresholds (thresholds for Asian populations are lower than those for European and North American groups). Further, the ATP III definition includes 110 mg/dL cutoff for fasting plasma glucose (although the American Diabetes Association recently decided to lower the threshold for impaired fasting glucose to 100 mg/dL).

The IDF diagnosis of metabolic syndrome is made:

a) **Abdominal_obesity_IDF_V2**: ANT10A ≥ 94 (men), ANT10A≥ 80 cm (women)

b) Plus any 2 of the following:

- **High_Trig_V2**: LABA67≥150 mg/dL
- **Low_HDL_V2**: LABA68< 40 mg/dL (men), LABA68< 50 mg/dL (women)
- **Elevated_BP_selfMeds_V2**: SBP5 ≥ 130 or SBP6 ≥ 85 mm Hg or use of anti-hypertension medications (MUE26d=1) (note: requires non-missing SPB, DBP, or medication use).

- **IFG_IDF_selfmeds_V2**: LABA70 ≥ 100 mg/dl, or previous diagnosis of diabetes (Diabetes_self_V2=1), or use of anti-diabetic medications (MUE26c=1).

First, each of the components is evaluated (0/1, or missing).

If abdominal obesity is missing or 2 or more or the other components are missing components, then METS_IDF3_V2 should be missing.

Otherwise, METS_IDF3_V2 = 0 if abdominal obesity is 0 or abdominal obesity is 1 and sum of four remaining components is 0 or 1.
= 1 if abdominal obesity is 1 and the sum of the four remaining components is 2, 3, or 4.

Source variable(s):
GENDER_V2. Gender
Diabetes_Self_V2. Self-reported diabetes diagnosis by Visit 2.
ANT10A. Girth – waist
LABA67. Triglycerides (mg/dL)
LABA68. HDL cholesterol (mg/dL)
LABA70. Fasting glucose (mg/dL)
SBP5. Average Systolic blood pressure
SBP6. Average Diastolic blood pressure
MUE26c. Self-reported anti-diabetic medication use
MUE26d. Self-reported anti-hypertensive medication use
Note: MUE26c/MUE26d missing had been reset to 0 when MUE2=1 (confirmed not taking any medications).

10.13 METS_NCEP2_V2 (Metabolic Syndrome w/ self-reported meds – NCEP-ATP III)

This is a 0/1 numeric variable with 1=metabolic syndrome present and 0=no metabolic syndrome in Visit 2. This variable incorporates the original METS_NCEP definition with changes to include self-reported medication use instead of scanned/transcribed medication use.

The diagnosis of metabolic syndrome is made when three or more of the following risk factors are present (*indicates variable substitutions from original definition):

***Elevated_BP_SelfMeds**: SBP5 ≥ 130 or SBP6 ≥ 85 mm Hg or self-report use of anti-hypertension medications (MUE26d=1) (note: requires non-missing SPB, DBP, or self-report medication use).
***High_Trig**: LABA67 ≥150 mg/dL
***Low_HDL**: LABA68< 40 mg/dl (men), LABA68< 50 mg/dl (women)
***IFG_NCEP_SelfMeds**: LABA70 ≥ 100 mg/dl or self-report use of anti-diabetic medications (MUE26c=1)
***Abdominal_obesity_NCEP**: ANT10A≥102 cm (men), ANT10A≥88 cm (women)
First, each of the components is evaluated (0/1, or missing).

If 3 or more components are missing components, then METS_NCEP2_V2 should be missing.

Otherwise, METS_NCEP2_V2 = 0 if sum of five components is 0, 1, 2
= 1 if sum of five components is 3, 4, 5

Source variable(s):
GENDER_V2. Gender at Visit 2
ANT10A. Girth - waist
LABA67. Triglycerides (mg/dL)
LABA68. HDL cholesterol (mg/dL)
LABA70. Fasting glucose (mg/dL)
SBP5. Average systolic blood pressure
SBP6. Average diastolic blood pressure
MUE26c. Self-report anti-diabetic medication use
MUE26d. Self-reported anti-hypertensive medication use
Note: MUE26c/MUE26d missing had been reset to 0 when MUE2=1 (confirmed not taking any medications).

10.14 DM3_Aware_V2 (Diabetes Awareness among Diabetes Definition 3 classification by Visit 2)

This is a 0/1 numeric variable that describes the awareness of diabetes among the diabetes definition 3 classified diabetic group by Visit 2.

If (diabetes_self_v2 missing and Selfmed_antidiab_V2 missing) OR Diabetes3_V2 ne 3 then set DM3_Aware_v2 as missing;
Otherwise If diabetes_self_v2=**1** or Selfmed_antidiab_V2=**1** then DM3_Aware_v2=**1**;
Otherwise DM3_Aware_v2=**0**;

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

Source variable(s):
DIABETES_SELF_V2. Diabetes – self-report only
Selfmed_antidiab_V2. Self-reported antidiabetic med used in the past 4 week prior Visit 2.
Diabetes3_V2. 3-level grouped Diabetes – includes self-report diabetes diagnoses by Visit 2.

10.15 DM4_Aware_V2 (Diabetes Awareness among Diabetes Definition 4 classification by Visit 2)

This is a 0/1 numeric variable that describes the awareness of diabetes among the diabetes definition 4 classified diabetic group by Visit 2.

If (diabetes_self_v2 missing and Selfmed_antidiab_V2 missing) OR Diabetes4_V2 ne 3 then set DM4_Aware_v2 as missing;
Otherwise If diabetes_self_v2=**1** or Selfmed_antidiab_V2=**1** then DM4_Aware_v2=**1**;

Otherwise DM4_Aware_v2=**0**;

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

Source variable(s):
DIABETES_SELF_V2. Diabetes – self-report only
Selfmed_antidiab_V2. Self-reported antidiabetic med used in the past 4 week prior Visit 2.
Diabetes4_V2. 3-level grouped Diabetes – includes self-report antidiabetics medication use at Visit 2.

10.16 DM5_Aware_V2 (Diabetes Awareness among Diabetes Definition 5 classification by Visit 2)

This is a 0/1 numeric variable that describes the awareness of diabetes among the diabetes definition 5 classified diabetic group by Visit 2.

If (diabetes_self_v2 missing and Selfmed_antidiab_V2 missing) OR Diabetes5_V2 ne 3 then set DM5_Aware_v2 as missing;
Otherwise If diabetes_self_v2=**1** or Selfmed_antidiab_V2=**1** then DM5_Aware_v2=**1**;
Otherwise DM5_Aware_v2=**0**;

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

Source variable(s):
DIABETES_SELF_V2. Diabetes – self-report only
Selfmed_antidiab_V2. Self-reported antidiabetic med used in the past 4 week prior Visit 2.
Diabetes5_V2. 3-level grouped Diabetes – includes self-report diabetes diagnoses by Visit 2 and self-reported antidiabetics medication use at Visit 2.

10.17 DM3_Control_V2 (Diabetes controlled: Diabetes Definition 3 Classified at Visit 2 and A1C < 7%)

This is a 0/1 numeric variable that indicates controlled diabetes at Visit 2. Defined only for diabetics group classified by diabetes definition 3.

If both LABA72 and Diabetes3_v2 missing, then set DM3_control_v2 as missing;
Otherwise If .Z < LABA72 < **7** and Diabetes3_v2=**3** then DM3_control_v2=**1**;
Otherwise DM3_control_v2=**0**;

Response format: 0=Uncontrolled diabetes
1=Controlled diabetes

Source variable(s):
Diabetes3_V2. 3-level grouped Diabetes – includes self-reported diabetes diagnosis by Visit 2.

LABA72. HbA1C

10.18 DM4_Control_V2 (Diabetes controlled: Diabetes Definition 4 Classified at Visit 2 and A1C < 7%)

This is a 0/1 numeric variable that indicates controlled diabetes at Visit 2. Defined only for diabetics group classified by diabetes definition 4.

If both LABA72 and Diabetes4_v2 missing, then set DM4_control_v2 as missing;
Otherwise If .Z < LABA72 < **7** and Diabetes3_v2=**3** then DM4_control_v2=**1**;
Otherwise DM4_control_v2=**0**;

Response format: 0=Uncontrolled diabetes
1=Controlled diabetes

Source variable(s):
Diabetes4_V2. 3-level grouped Diabetes – includes self-report antidiabetics medication use at Visit 2.
LABA72. HbA1C

10.19 DM5_Control_V2 (Diabetes controlled: Diabetes Definition 5 Classified at Visit 2 and A1C < 7%)

This is a 0/1 numeric variable that indicates controlled diabetes at Visit 2. Defined only for diabetics group classified by diabetes definition 5.

If both LABA72 and Diabetes5_v2 missing, then set DM5_control_v2 as missing;
Otherwise If .Z < LABA72 < **7** and Diabetes3_v2=**3** then DM5_control_v2=**1**;
Otherwise DM5_control_v2=**0**;

Response format: 0=Uncontrolled diabetes
1=Controlled diabetes

Source variable(s):
Diabetes5_V2. 3-level grouped Diabetes – includes self-report diabetes diagnoses by Visit 2 and self-reported antidiabetics medication use at Visit 2.
LABA72. HbA1C

10.20 DM_TRT_V2 (Diabetes Treatment at Visit 2)

This is a 0/1 numeric variable that describes the treatment of diabetes at Visit 2.

If (mue26c missing or =9) then set DM_trt_v2 missing;
Otherwise set DM_trt_v2=MUE26c (0/1).

Response format: 0=No treatment of diabetes
1=has treatment of diabetes

Source variable(s):
MUE26c. High blood sugar or diabetes – self-reported med use
Note: MUE26c missing had been reset to 0 when MUE2=1 (confirmed not taking any medications).

10.21 PreCVD_V2 (Prevalent Cardiovascular Disease at Visit 2)

This is a 0/1 variable that combines self-report of heart attack, or procedure (angioplasty, stent, bypass) or stroke.

If PreCVD, MHE1, MHE2 and MHE5 all missing then set preCVD_v2 missing.
Otherwise if any of (PreCVD, MHE1, MHE2 or MHE5) =1 then set PreCVD_v2=1;
Otherwise set PreCVD_v2=0.

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

Source variable(s):

WHE1. Since baseline visit, has a doctor said that you had Heart attack?
WHE2. Since baseline visit, has a doctor said that you had A balloon angioplasty, a stent, or bypass surgery to the arteries in your heart to improve the blood flow to your heart?
MHE5. Since baseline visit, has a doctor said that you had Stroke?
PreCVD. Prevalent Cardiovascular Disease at baseline.

10.22 INCIDENT_CVD_V1V2 (incident Cardiovascular Disease from visit 1 to visit 2)

This is a 0/1 variable that denotes the incident cardiovascular disease cases from visit 1 to visit 2.

If preCVD missing AND PreCVD_v2 missing, then set as missing;
Otherwise if preCVD=0 and PreCVD_v2=1 then set incident_CVD_v1v2=1;
Otherwise incident_CVD_v1v2=0.

Response format: 0=No Incident CVD (No)
1=Incident CVD (Yes)

Source variable(s):
PreCVD. Prevalent CVD at baseline
PreCVD_V2. Prevalent CVD at Visit 2.

10.23 Diabetes3_LL_DATE (Diabetes censored interval lower limit date using diabetes definition 3)

If baseline Diabetes3=3 then Diabetes3_LL_date= Missing;

Otherwise if AFUDM_Self_ByV2=1 happened in AFU1 or the first AFU visit, then Diabetes3_LL_date= baseline Clindate;
Otherwise if AFUDM_Self_ByV2=1 then Diabetes3_LL_date= the previous AFU visit GHE0a;
Otherwise if Diabetes3_V2=3 then Diabetes3_LL_date= the last AFU GHE0a before V2;
Otherwise if Diabetes3_V2 =1 or 2 then Diabetes3_LL_date = Clindate_V2;

Source variable(s):
Diabetes3. 3-level prevalent diabetes classified by diabetes definition 3 at baseline.
Diabetes3_V2. 3-level prevalent diabetes classified by diabetes definition 3 by Visit 2.
AFUDM_Self_byV2. Self-reported diabetes only at AFU by Visit 2.
GHE0a. AFU interview date.
Clindate. Baseline visit date.
Clindate_V2. Visit 2 visit date.

10.24 Diabetes3_LL (Time in days between baseline to Diabetes censored interval lower limit using diabetes definition 3)

If Diabetes3_LL_date=Missing then Diabetes3_LL= (Diabetes3_LL_DATE);
Otherwise Diabetes3_LL= (Diabetes3_LL_DATE – baseline Clindate);

Source variable(s):
Clindate. Baseline visit date.
Diabetes3_LL_DATE. Diabetes censored interval lower limit date using diabetes definition 3.

10.25 Diabetes3_UL_DATE (Diabetes censored interval upper limit date using diabetes definition 3)

If baseline Diabetes3=3 then Diabetes3_UL_date=baseline Clindate;
Otherwise if AFUDM_Self_ByV2=1 then Diabetes3_UL_date= the same AFU visit GHE0a;
Otherwise if Diabetes3_V2=3 then Diabetes3_UL_date= Clindate_V2;
Otherwise if Diabetes3_V2 =1 or 2 then Diabetes3_UL_date= Missing;
Source variable(s):
Diabetes3. 3-level prevalent diabetes classified by diabetes definition 3 at baseline.
Diabetes3_V2. 3-level prevalent diabetes classified by diabetes definition 3 by Visit 2.
AFUDM_Self_byV2. Self-reported diabetes only at AFU by Visit 2.
GHE0a. AFU interview date.
Clindate. Baseline visit date.
Clindate_V2. Visit 2 visit date.

10.26 Diabetes3_UL (Time in days between baseline to Diabetes censored interval upper limit using diabetes definition 3)

If Diabetes3_UL_date=Missing then Diabetes3_UL= (Diabetes3_UL_DATE);
Otherwise Diabetes3_UL= (Diabetes3_UL_DATE – baseline Clindate);

Source variable(s):
Clindate. Baseline visit date.

Diabetes3_UL_DATE. Diabetes censored interval upper limit date using diabetes definition 3.

10.27 Diabetes5_LL_DATE (Diabetes censored interval lower limit date using diabetes definition 5)

If baseline Diabetes5=3 then Diabetes5_LL_date= Missing;
Otherwise if AFUDM_Self_ByV2=1 happened in AFU1 or the first AFU visit, then Diabetes5_LL_date= baseline Clindate;
Otherwise if AFUDM_Self_ByV2=1 then Diabetes5_LL_date= the previous AFU visit GHE0a;
Otherwise if Diabetes5_V2=3 then Diabetes5_LL_date= the last AFU GHE0a before V2;
Otherwise if Diabetes5_V2 =1 or 2 then Diabetes5_LL_date = Clindate_V2;

Source variable(s):
Diabetes5. 3-level prevalent diabetes classified by diabetes definition 5 at baseline.
Diabetes5_V2. 3-level prevalent diabetes classified by diabetes definition 5 by Visit 2.
AFUDM_Self_byV2. Self-reported diabetes only at AFU by Visit 2.
GHE0a. AFU interview date.
Clindate. Baseline visit date.
Clindate_V2. Visit 2 visit date.

10.28 Diabetes5_LL (Time in days between baseline to Diabetes censored interval lower limit using diabetes definition 5)

If Diabetes5_LL_date=Missing then Diabetes5_LL= (Diabetes5_LL_DATE);
Otherwise Diabetes5_LL= (Diabetes5_LL_DATE – baseline Clindate);

Source variable(s):
Clindate. Baseline visit date.
Diabetes5_LL_DATE. Diabetes censored interval lower limit date using diabetes definition 5.

10.29 Diabetes5_UL_DATE (Diabetes censored interval upper limit date using diabetes definition 5)

If baseline Diabetes5=3 then Diabetes5_UL_date=baseline Clindate;
Otherwise if AFUDM_Self_ByV2=1 then Diabetes5_UL_date= the same AFU visit GHE0a;
Otherwise if Diabetes5_V2=3 then Diabetes5_UL_date= Clindate_V2;
Otherwise if Diabetes5_V2 =1 or 2 then Diabetes5_UL_date= Missing;
Source variable(s):
Diabetes5. 3-level prevalent diabetes classified by diabetes definition 5 at baseline.
Diabetes5_V2. 3-level prevalent diabetes classified by diabetes definition 5 by Visit 2.
AFUDM_Self_byV2. Self-reported diabetes only at AFU by Visit 2.
GHE0a. AFU interview date.
Clindate. Baseline visit date.
Clindate_V2. Visit 2 visit date.

10.30 Diabetes5_UL (Time in days between baseline to Diabetes censored interval upper limit using diabetes definition 5)

If Diabetes5_UL_date=Missing then Diabetes5_UL= (Diabetes5_UL_DATE);
Otherwise Diabetes5_UL= (Diabetes5_UL_DATE – baseline Clindate);

Source variable(s):
Clindate. Baseline visit date.
Diabetes5_UL_DATE. Diabetes censored interval upper limit date using diabetes definition 5.

10.31 Diabetes3_INDICATOR_V2 (indicator of diabetic status using diabetes definition 3 by visit 2, 1=Yes 0=No)

This is a 0/1 indicator variable diabetic cases classified by diabetes definition 3 by visit 2. This indicator variable could also be used as incident diabetes case indicator within non-diabetic group classified by diabetes definition 3 at baseline.

If Diabetes3_v2 missing, then set as missing;
Otherwise if Diabetes3_v2=3 then set Diabetes3_INDICATOR_V2=1;
Otherwise set Diabetes3_INDICATOR_V2 as 0.

Response format: 0=Not diabetic (No)
1=Diabetes (Yes)

Source variable(s):
Diabetes3_V2. Prevalent diabetes based on self-reported diagnoses and lab criteria by Visit 2.

10.32 Diabetes3_TIME_V2 (Observed time for diabetes incident event using diabetes definition 3 by visit 2)

This variable will be the upper bound approximation of the existing interval time (Diabetes3_LL, Diabetes3_UL) for participants who developed diabetes between visit 1 and visit 2.

Diabetes3_Time_V2 is defined as:
= missing: if the subject has missing v1 (diabetes3=.) or v2 diabetes info (Diabetes3_v2 = .).
Otherwise set = Zero: if the subject has diabetes at v1 (diabetes3 =3).

For incidence analysis, the subset (subpopulation) of participants who do not have diabetes at v1 (diabetes3 in (1,2)) should be analyzed;

Otherwise, set = Diabetes3_LL: if diabetes3_indicator_v2 = 0; right-censored cases, use the lower bound of interval time, which equals to v2 time;
Otherwise, set = Diabees3_UL: if the subject developed diabetes between v1 and v2; incident cases, use the upper bound of interval time.

Source variable(s):
Diabetes3. Prevalent diabetes based on self-reported diagnosis and lab criteria at Visit 1.
Diabetes3_V2. Prevalent diabetes based on self-reported diagnosis and lab criteria by V2.
Diabetes3_LL. Time in days between baseline to Diabetes definition 3 censored interval lower limit.
Diabetes3_UL. Time in days between baseline to Diabetes definition 3 censored interval upper limit.

10.33 Diabetes4_INDICATOR_V2 (indicator of diabetic status using diabetes definition 4 at visit 2, 1=Yes 0=No)

This is a 0/1 indicator variable diabetic cases classified by diabetes definition 4 at visit 2. This indicator variable could also be used as incident diabetes case indicator within non-diabetic group classified by diabetes definition 4 at baseline.

If Diabetes4_v2 missing, then set as missing;
Otherwise if Diabetes4_v2=3 then set Diabetes4_INDICATOR_V2=1;
Otherwise set Diabetes4_INDICATOR_V2 as 0.

Response format: 0=Not diabetic (No)
1=Diabetes (Yes)

Source variable(s):
Diabetes4_V2. Prevalent diabetes based on self-reported med and lab criteria at Visit 2.

10.34 Diabetes5_INDICATOR_V2 (indicator of diabetic status using diabetes definition 5 by visit 2, 1=Yes 0=No)

This is a 0/1 indicator variable diabetic cases classified by diabetes definition 5 by visit 2. This indicator variable could also be used as incident diabetes case indicator within non-diabetic group classified by diabetes definition 5 at baseline.

If Diabetes5_v2 missing, then set as missing;
Otherwise if Diabetes5_v2=3 then set Diabetes5_INDICATOR_V2=1;
Otherwise set Diabetes5_INDICATOR_V2 as 0.

Response format: 0=Not diabetic (No)
1=Diabetes (Yes)

Source variable(s):
Diabetes5_V2. Prevalent diabetes based on self-reported diagnoses, self-reported med and lab criteria by Visit 2.

10.35 Diabetes5_TIME_V2 (Observed time for diabetes incident event using diabetes definition 5 by visit 2)

This variable will be the upper bound approximation of the existing interval time (Diabetes5_LL, Diabetes5_UL) for participants who developed diabetes between visit 1 and visit 2.

Diabetes5_Time_V2 is defined as:
= missing: if the subject has missing v1 (diabetes5=.) or v2 diabetes info (Diabetes5_v2 = .).
Otherwise set = Zero: if the subject has diabetes at v1 (diabetes5 =3).

For incidence analysis, the subset (subpopulation) of participants who do not have diabetes at v1 (diabetes5 in (1,2)) should be analyzed;

Otherwise, set = Diabetes5_LL: if diabetes5_indicator_v2 = 0; right-censored cases, use the lower bound of interval time, which equals to v2 time;
Otherwise, set = Diabees5_UL: if the subject developed diabetes between v1 and v2; incident cases, use the upper bound of interval time.

Source variable(s):
Diabetes5. Prevalent diabetes based on self-reported diagnosis, self-reported medication use and lab criteria at baseline
Diabetes5_V2. Prevalent diabetes based on self-reported diagnosis, self-reported medication use and lab criteria by Visit 2.
Diabetes5_LL. Time in days between baseline to Diabetes definition 5 censored interval lower limit.
Diabetes5_UL. Time in days between baseline to Diabetes definition 5 censored interval upper limit.

10.36 HOMA_IR_V2 (Visit 2 HOMA index of Insulin Resistance)

This is a numeric variable that defined the Homeostasis model assessment of insulin resistance as follows (LABA70 and INSULIN_FAST_V2 must be non-missing):

HOMA_IR_V2 = LABA70*INSULIN_FAST_V2 / 405.

Source variable(s):
LABA70. Fasting Glucose in mg/dL
INSULIN_FAST_V2. Insulin, fasting (converted to mU/L) at visit 2.

10.37 HOMA_B_V2 (Visit 2 HOMA index of Beta Cell Function)

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

If LABA70 ≤ 63 then set HOMA_B as missing, otherwise:

HOMA_B_V2 = 360 * INSULIN_FAST_V2 / (LABA70 - 63).

Source variable(s):
LABA70. Fasting Glucose in mg/dL
INSULIN_FAST_V2. Insulin, fasting (converted to mU/L)

10.38 High_total_Chol2_V2 (Visit 2 Hypercholesterolemia (from LABA66, LABA68, LABA69, MUE26E))

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

If missing(laba66) & missing(laba68) & missing(laba69) & (missing(mue26e) or mue26e=9) then high_total_chol2=missing
Else if (laba66>=240) OR (laba69>=160) OR (0<=laba68<40) OR (mue26e=1) then high_total_chol2=1;
Else high_total_chol2=0;

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

Source variable(s):
LABA66. Total Cholesterol (mg/dL)
LABA68. HDL - cholesterol (mg/dL)
LABA69. LDL - Cholesterol (mg/dL)
MUE26E. Self-reported use Lipid lowering drugs/Antihyperlipidemics medication.
Note: MUE26E missing had been reset to 0 when MUE2=1 (confirmed not taking any medications).

10.39 Current_Smoker_V2 (Current Smoker status at visit 2)

This is a 0/1 indicator variable to denote current smoker status at visit 2 using the variables TBE1 and TBE4.

If TBE1=1 & TBE4 in (1,2) then CURRENT_SMOKER_V2=1
If (TBE1=1 & TBE4=3) or TBE1=0 then CURRENT_SMOKER_V2=0

Response format: 0 = No (i.e., never or former smoker)
1 = Yes (i.e., current smoker)

Source variable(s):
TBE1. Have you ever smoked at least 100 cigarettes in your entire life?
TBE4. Do you NOW smoke daily, some days or not at all?

10.40 STROKE_V2 (Prevalent Stroke, self-report, Visit 2)

STROKE_V2 is a 0/1 variable indicating the self-reported stroke:

If (MHE5=1 | (Stroke = 1) then Stroke_V2=1;
Otherwise if (MHE5<=.Z|MHE5 = 9) then Stroke_V2=.;
Otherwise Stroke_V2=0;

Source variable(s):
STROKE. Self reported stroke from Visit 1

Since the first SOL visit, has a doctor said that you had any of the following medical problems?

MHE5. Stroke?

10.41 STROKE_TIA_V2 (Prevalent Stroke or TIA, self-report, Visit 2)

STROKE_TIA_V2 is a 0/1 variable indicating the self-reported stroke, mini-stroke or TIA:

If (MHE5=1 | MHE6=1) | (Stroke_TIA = 1) then Stroke_TIA_V2=1;
Otherwise if (MHE5<=.Z|MHE5 = 9) & (MHE6<=.Z|MHE6 = 9) then Stroke_TIA_V2=.;
Otherwise Stroke_TIA_V2=0;

Source variable(s):
STROKE_TIA. Self reported stroke or TIA from Visit 1

Since the first SOL visit, has a doctor said that you had any of the following medical problems?

MHE5. Stroke?
MHE6. A mini-stroke or TIA (transient ischemic attack)?

10.42 CDCR_ V2 (Cerebrovascular Disease and Carotid Revascularization, self-report)

CDCR_V2 is a 0/1 variable that checks for medical history of stroke, mini-stroke or TIA, or balloon angioplasty or surgery to the arteries in the neck as follows:

if (MHE5=1 | MHE6=1 | MHE7=1) | (cdcr = 1) then CDCR_V2=1;
else if (MHE5<=.Z|MHE5 = 9) & (MHE6<=.Z|MHE6 = 9) & (MHE7<=.Z|MHE7 = 9)
then CDCR_V2=.;
else CDCR_V2=0;;

This measure is cumulative from the time before Visit 1 until Visit 2.

Response format: 0 = No CDCR (No)
1 = CDCR (Yes)

Source variable(s):
CDCR. Self reported CDCR from Visit 1

Since the first SOL visit, has a doctor said that you had any of the following medical problems?

MHE5. Stroke?
MHE6. A mini-stroke or TIA (transient ischemic attack)?
MHE7. A balloon angioplasty or surgery to the arteries of your neck to prevent or correct a stroke?

10.43 CHD_SELF_V2 (Self-reported CHD, Visit 2)

CHD_SELF_V2 is a 0/1 variable that combines the self-reports of coronary revascularization or heart attack since Visit 1.

If MHE1=1 | MHE2=1 | CHD_SELF = 1 then CHD_SELF=1;
Else if MHE1 in(.,.Q) & MHE2 in(.,.Q) then CHD_SELF=.;
Else CHD_SELF=0;

Source variable(s):
CHD_SELF. Self reported CHD from Visit 1.

Since the first SOL visit, has a doctor said that you had any of the following medical problems?
MHE1: Heart Attack?
MHE2: A balloon angioplasty, a stent, or bypass surgery to the arteries in your heart to improve the blood flow to your heart.

10.44 HBA1C_SI_V2 (Glycosylated Hemoglobin in SI units (mmol/mol))

HBA1C_SI = (LABA72 *10.93) – 23.50, rounded to the nearest integer

Beginning with manuscripts submitted after Jan. 1, 2013, Diabetes requires authors to report
HbA1c levels in both traditional, DCCT-derived units (as %) and SI, IFCC-recommended units (as mmol/mol).

Change of HbA1c reporting to the new SI units, Jones G, Barker G, et al. MJA 2011; 195: 45–46

Source variable(s):
LABA72. % Glycosylated Hemoglobin

10.45 DNA_AVAILABLE_V2 (DNA AVAILABLE at Visit 2)

DNA_AVAILABLE_V2 is a binary variable that determines whether or not a study participant had blood draws successfully conducted without incidents or problems for tubes 4 and 5 to make DNA available to investigators associated or not associated to HCHS/SOL and specialized laboratories (Note: does NOT reflect consent for use on DNA samples, i.e. genotyping).

((bio14a4^=1 & bio14b4^=1 & bio14c4^=1 & bio14d4^=1 & bio14e4^=1 & bio14f4^=1 & bio22a4^=1 & bio22b4^=1 & bio22c4^=1 & bio22d4^=1 & bio22e4^=1) |
(bio14a5^=1 & bio14b5^=1 & bio14c5^=1 & bio14d5^=1 & bio14e5^=1 & bio14f5^=1 & bio22a5^=1 & bio22b5^=1 & bio22c5^=1 & bio22d5^=1 & bio22e5^=1))
then let DNA_Available_V2=1;
Otherwise DNA_Available_V2=0;

Source variable(s):
BIO9. Date - Blood Collection
BIO14A4. Tube 4 – Sample not drawn
BIO14B4. Tube 4 – Partial sample drawn
BIO14C4. Tube 4 – Tourniquet reapplied
BIO14D4. Tube 4 – Fist clenching
BIO14E4. Tube 4 – Needle movement
BIO14F4. Tube 4 – Participant reclining
BIO14A5. Tube 5 – Sample not drawn
BIO14B5. Tube 5 – Partial sample drawn
BIO14C5. Tube 5 – Tourniquet reapplied
BIO14D5. Tube 5 – Fist clenching
BIO14E5. Tube 5 – Needle movement
BIO14F5. Tube 5 – Participant reclining
BIO22A4. Tube 4 – Broken Tube
BIO22B4. Tube 4 – Sample re-centrifuged
BIO22C4. Tube 4 – Clotted
BIO22D4. Tube 4 – Hemolyzed
BIO22E4. Tube 4 – Lipemic
BIO22A5. Tube 4 – Broken Tube
BIO22B5. Tube 4 – Sample re-centrifuged
BIO22C5. Tube 4 – Clotted
BIO22D5. Tube 4 – Hemolyzed

11. MEDICATIONS – OTHER

11.1 HYPERTMED_SELF_V2 (Hypertension Med Use - self-report only at Visit 2)

This is a 0/1 numeric variable. This variable combines self-report antihypertensive medication use along with any indication of taking meds in last 4 weeks.

if (mue2=1 & mue26d<=.Z) | mue26d=0 then HYPERTMED_SELF_V2=0;
else if mue26d=1 then HYPERTMED_SELF_V2=1;

Response format: 0 = No antihypertensive medication use
1 = Antihypertensive medication use

Source Variable(s):
MUE2. Took no medication in last 4 weeks.
MUE26d. Self-reported High blood pressure or hypertension medication use
Note: MUE26d missing had been reset to 0 when MUE2=1 (confirmed not taking any medications).

Variable Modifications

Date release	Variable	Number of records (participants) changed	Notes
3/2019	YRSUS_C2_V2	1,620 missing values were recoded	Corrected calculation. NV3 has only 53 missing values.
3/2019			**Visit 2 derived variables updates after the self-reported med variables (MUE26x in V2) missing values had been reset to zero when MUE2=1 (confirmed not taking any medication in MUE).**
3/2019	DIABETES4_V2	20	Definition identical to DIABETES2_V2 distributed in INV2 data release. In INV3 data release we renamed this variable because it is not directly comparable to DIABETES2 at baseline. Baseline DIABETES2 at baseline uses scanned medications whereas DIABETES4_V2 uses updated self-reported medications MUE26c.
3/2019	HYPERTMED_SELF_V2	0	
3/2019	ELEVATED_BP_SELFMEDS_V2	1	
3/2019	IFG_NCEP_SELFMEDS_V2	29	
3/2019	IFG_IDF_SELFMEDS_V2	29	
3/2019	METS_NCEP2_V2	15	
3/2019	METS_IDF3_V2	0	
3/2019	DM_TRT_V2	3645 -> # of mue26 switched missing to 0, given MUE2=1	Clone of MUE26c, normally used within the category of Diabetics and no changes found in this case.
3/2019	HIGH_TOTAL_CHOL2_V2	20	
3/2019	HYPERTENSION2_V2	1	
3/2019	HYPERTENSION_C4_V2	1	
3/2019	HYPERT_TREATMENT_V2	0	
3/2019	DYSLIPIDEMIA_C3_V2	1282 from missing to 2 (dyslipidemia w/o treatment)	
3/2019	DYS_TCHDL_MED_V2	20	
3/2019	HYPERTENSION2_AHA_V2	1	
3/2019	HYPERTENSION_AHA_C5_V2	1	
7/2020	YRSUS_V2	1800	Algorithm update. 1800 values slightly changed because they are no longer reported as integer values when US_BORN_V2=0.
7/2020	YRSUS_C2_V2	7	Corresponding change with YRSUS_V2
7/2020	YRSUS_C3_V2		NEW VARIABLE
7/2020	DIABETES_SELF_V2	898	Algorithm update by implement AFU self-reported DM and Gestational DM.

Date release	Variable	Number of records (participants) changed	Notes
7/2020	DIABETES3_V2	566	Corresponding change with Diabetes_self_V2
7/2020	GFR_PLUS1_V1V2		RENAMED from GFR_PLUS1
7/2020	SELFMED_ANTIDIAB_V2		NEW VARIABLE
7/2020	AFUDM_SELF_BYV2		NEW VARIABLE (intermediate)
7/2020	GESTATIONAL_DM_V2		NEW VARIABLE (intermediate)
7/2020	DIABETES5_V2		NEW VARIABLE
7/2020	DM3_AWARE_V2		RENAMED from DM_AWARE_V2
7/2020	DM4_AWARE_V2		NEW VARIABLE
7/2020	DM5_AWARE_V2		NEW VARIABLE
7/2020	DM3_CONTROL_V2		RENAMED from DM_CONTROL_V2
7/2020	DM4_CONTROL_V2		NEW VARIABLE
7/2020	DM5_CONTROL_V2		NEW VARIABLE
7/2020	DIABETES3_C4_V2		RENAMED from DIABETES_C4_V2
7/2020	DIABETES4_C4_V2		NEW VARIABLE
7/2020	DIABETES5_C4_V2		NEW VARIABLE
7/2020	DIABETES3_INDICATOR_V2		NEW VARIABLE
7/2020	DIABETES4_INDICATOR_V2		NEW VARIABLE
7/2020	DIABETES5_INDICATOR_V2		NEW VARIABLE
7/2020	DIABETES3_TIME_V2		NEW VARIABLE
7/2020	DIABETES5_TIME_V2		NEW VARIABLE
7/2020	DIABETES3_LL_DATE		NEW VARIABLE
7/2020	DIABETES3_UL_DATE		NEW VARIABLE
7/2020	DIABETES3_LL		NEW VARIABLE
7/2020	DIABETES3_UL		NEW VARIABLE
7/2020	DIABETES5_LL_DATE		NEW VARIABLE
7/2020	DIABETES5_UL_DATE		NEW VARIABLE
7/2020	DIABETES5_LL		NEW VARIABLE
7/2020	DIABETES5_UL		NEW VARIABLE
7/2020	AFUHYPERT_SELF_BYV2		NEW VARIABLE
7/2020	AFUCOPD_SELF_BYV2		NEW VARIABLE
7/2020	COPD_EVER_V2		NEW VARIABLE
7/2020	CHARLSON_V2	1165	COPD_EVER_V2 replaced MHE12 to include AFU and baseline self-reported COPD cases
7/2020	IFG_IDF_SELFMEDS_V2	300	Diabetes_Self_V2 replaced MHE14 to include AFU and baseline self-reported diabetes cases
7/2020	METS_IDF3_V2	115	Corresponding to IFG_IDF_SELFMEDS_V2 update
7/2020	HYPERT_AWARENESS_V2	1987	Include the AFU and baseline self-reported cases
7/2020	HYPERT_TREATMENT_V2	1663	Due to Hypert_awareness_V2 updates
7/2020	HYPERT_CONTROL_V2	1081	Due to Hypert_Treatment_V2 updates
7/2020	DURATION_V2	32	Set to missing when less than 30 minutes or greater than 8 hours