NHLBI grant: Ancillary to HCHS/SOL: Nutrition & Physical Activity Assessment Study
(R01 HL095856, Y Mossavar-Rahmani)

SOLNAS Ancillary Study
Study of Latinos: Nutrition and Physical Activity Assessment Study (SOLNAS)

Data Dictionary
Version 2.0
October 31, 2014

Prepared by
HCHS/SOL Coordinating Center
Collaborative Studies Coordinating Center
UNC Department of Biostatistics

Daniela Sotres-Alvarez
Maria de los Angeles Abreu
Yanping Teng
Courtney Page

Table of Contents

1. HCHS/SOL Derived variables

1.1 DESIGN

1.1.1 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:

Field Center and selected block group identifier.

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

1.2 ADMINISTRATIVE

1.2.1 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):
ID. Subject ID

1.2.2 Centernum – Numeric (Participant's field center)

This is a numeric variable with four possible values derived from the city of origin: 1= Bronx, 2= Chicago, 3= Miami, 4= San Diego. Center cannot have missing values because each valid participant ID has an affiliated field center.

Source variable(s):

ID.	Subject ID
CENTER.	Center (Participant's field center)

1.2.3 Clindate (Date of the participant's clinic visit)

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

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

Source variable(s):

PIEA0A. Date of completion of the Personal Information Form
ICTA0A or ICTB0A. Date of completion of the Informed Consent Form (version A or B)
SBPA0A. Date of completion of the Sitting Blood Pressure Form
ANTA0A. Date of completion of the Anthropometry Form

1.2.4 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 criterions 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 must are 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)

1.3 SOCIO-DEMOGRAPHIC

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

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

AGE = INTEGER of (PIEA2– PIEA0A)/365.25

When missing use age collected as recruitment: ELEA2
and if this missing, then use HSRA5A4-HSRA5N4 or HSRB4A4-HSRB4J4 based on linkage in ELEA of participant to roster row of household screening form.

Source variable(s):

PIEA0A.	Date of completion of the Personal Information Form
PIEA2.	Date of birth
ELEA2	Age of person screened
HSRA5A4-5N4	Age of person rostered from household (rows A-N)
HSRB4A4-4J4	Age of person rostered from household (rows A-J)

1.3.2 Agegroup_C2 (2-level grouped age of participant)

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

AGE	AGEGROUP_C2
18 – 44	1
45+	2

Response Format: 1=Ages 18-44
2=Ages 45+

Source variable(s):

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

1.3.3 Agegroup_C6 (6-level grouped age of participant)

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

AGE	AGEGROUP_C6
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. age of participant at the time of their clinic visit

1.3.4 Gender (Gender (F=Female, M=Male))

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

PIEA1	Gender
1	M
2	F

When missing use gender collected as recruitment:
HSRA5A3-HSRA5N3 or HSRB4A3-HSRB4J3 based on linkage in ELEA of participant to roster row of household screening form

Source variable(s):

PIEA1. Gender
HSRA5A3-5N3 – Gender of person rostered from household (rows A-N)
HSRB4A3-4J3 – Gender of person rostered from household (rows A-J)

1.3.5 Gendernum (Gender – Numeric (0=Female, 1=Male))

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

Gender	Gendernum
F	0
M	1

Response Format: 0=Female
1=Male

Source variable(s):

Gender. (Gender (F=Female, M=Male))

1.3.6 Education_C2 (2-level group education level)

The educational status derived variable groups the attainment of a high school diploma or an equivalent degree, or not having that degree.

```
if piea19a in(1,2,3) & piea21a in(0,.) then EDUCATION_C2=1;
else if (piea19a in(1,2,3) & piea21a in(1)) |
(piea19a in(4,5)) then EDUCATION_C2=2;
else if piea19a in(6) then EDUCATION_C2=.;
else EDUCATION_C2=.;
```

Response format: 1 = No high school diploma or GED
2 = At least High school diploma or GED

Source variable(s):

PIEA19a. Highest (or closest) grade/level of education achieved (6 levels)
PIEA21a. High school diploma or equivalent (yes/no)

1.3.7 Education_C3 (3-level group education level)

The educational status derived variable groups the attainment of a high school diploma or an equivalent degree, not having that high school or equivalent degree, or attainment of education beyond a high school equivalent (i.e. college or vocational)

```
if piea19a in(1,2,3) & piea21a in(0,.) then EDUCATION_C3=1;
else if piea19a in(1,2,3) & piea21a in(1) then EDUCATION_C3=2;
else if piea19a in(4,5) then EDUCATION_C3=3;
else if piea19a in(6) then EDUCATION_C3=.;
else EDUCATION_C3=.;
```

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

Source variable(s):

PIEA19a. Highest (or closest) grade/level of education achieved (6 levels)
PIEA21a. High school diploma or equivalent (yes/no)

1.3.8 Income (Grouped yearly household income)

This derived variable groups each subject to one income category. The income is set to missing if both ECEA3 and ECEA4 are missing. The income is also set to missing if both ECEA3 and ECEA4 are answered.

Income	ECEA3	ECEA4
1	1	.
2	2	.
3	3	.
4	4	.
5	5	.
6	.	1
7	.	2
8	.	3
9	.	4
10		5
.	.	.
.	Not missing	Not missing

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

Source variable(s):

ECEA3. Low income range
ECEA4. High income range

1.3.9 Income_C5 (5-level grouped yearly household income)

This derived variable groups each subject to one income category. The income is set to missing if both ECEA3 and ECEA4 are missing. The income is also set to missing if both ECEA3 and ECEA4 are answered.

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

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

Source variable(s):

ECEA3. Low income range
ECEA4. High income range

1.3.10 Marital_Status (3-level grouped marital status)

This derived variable groups each subject's marital status.

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

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

Source variable(s):

PIEA3. Marital Status

1.3.11 Employed (4-level grouped employment status)

This variable groups the employment status of the participant. This variable includes retirees who are or are not employed.

```
if (ocea1=1 & (ocea5=3 | ocea5≤.Z)) then EMPLOYED=1;
else if (ocea1^=1 & ocea5=3) then EMPLOYED=2;
else if (ocea5=2) then EMPLOYED=3;
else if (ocea5=1) then EMPLOYED=4;
else EMPLOYED=.;
```

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

Source variable(s):

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

1.3.12 Occupation_Curr (Grouped current occupation for employed participants)

For participants who are currently employed (part or full time), this is the type of occupation the participant had at the time of the baseline examination.

```
if (ocea16a=7) then OCCUPATION_CURR=1;
else if (ocea16a=11) then OCCUPATION_CURR=2;
else if (ocea16a=6) then OCCUPATION_CURR=3;
else if (1≤ocea16a≤4) then OCCUPATION_CURR=4;
else if (ocea16a not in(.,1,2,3,4,6,7,11)) then OCCUPATION_CURR=5;
else OCCUPATION_CURR=.;
```

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

Source variable(s):

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

1.3.13 Occupation_Long (Grouped longest held occupation)

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

if (ocea28a=7) then OCCUPATION_CURR=1;
else if (ocea28a=11) then OCCUPATION_CURR=2;
else if (ocea28a=6) then OCCUPATION_CURR=3;
else if (1≤ocea28a≤4) then OCCUPATION_CURR=4;
else if (ocea28a not in(.,1,2,3,4,6,7,11)) then OCCUPATION_CURR=5;
else OCCUPATION_CURR=.;

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

Source variable(s):

OCEA28a. Type of job held the longest.

1.4 ACCULTURATION

1.4.1 Bkgrd1_C7 (7-level re-classification of Hispanic/Latino Background)

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

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

Source variable(s):

PIEA5A. Hispanic/Latino background
PIEA5B. Text response for other Hispanic/Latino background

1.4.2 US_Born (Grouped place of birth – 50 US States only)

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

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

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

Source variable(s):

PIEA4. Born in which Country/territory.

1.4.3 US_Native (Grouped place of birth – US states and territories)

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

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

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

Source variable(s):

PIEA4. Born in which Country/territory.

1.4.4 Lang_Pref (Language Preference for Baseline Examination)

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

Response format: 1=Spanish
2=English

Source variable(s):

Form Version of PIEA/PISA
Form Version of IDEA/IDSA
Form Version of MHEA/MHSA

1.4.5 IMGEN_C2 (2-level Immigrant Generation)

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

First generation is defined as foreign-born with foreign-born parents (same as IMGEN_C3 and IMGEN_C4).

The second generation and beyond definition:

If US_BORN in (0, 1) **with** both IMGEN_11 and IMGEN_14 are missing, then set IMGEN_C2 as missing.
Else if US_BORN=1 then classified as 2nd + generation.
Else if US_BORN =0 **with** at least one US born parent then classified as 2nd + generation
Else set IMGEN_C2 as missing.

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

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

Source variable(s):

PIEA4. Participant's country of origin
PIEA11. Mother's country of origin
PIEA12. Maternal grandmother's country of origin

PIEA13. Maternal grandfather's country of origin
PIEA14. Father's country of origin
PIEA15. Paternal grandmother's country of origin
PIEA16. Paternal grandfather's country of origin

1.4.6 YRSUS (Years lived in the US)

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

1. If participant was NOT born in the US, then the years lived in the US is equal to the response of PIEA7.

2. If participant was born in the US, then the years lived in the US is equal to the participant's age (calculated from participant's date of birth).

Response format: 0=Not a US citizen
1=US citizen

Source variable(s):

Age. Age of participant in years.
PIEA4. Participant's country of origin
PIEA7. Years lived in mainland US.

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

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

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

Source variable(s):

YRSUS. Years lived in the US.

1.4.8 SASH_ALL (Short acculturation scale for Hispanic)

This is a numeric variable. The original acculturation questionnaire ("Development of a Short Acculturation Scale for Hispanics", Marin et. Al., *Hispanic Journal of Behavioral Sciences* 1987; 9; 18) contains 12 items but the following 2 questions were dropped from use in HCHS/SOL:

1. In what language(s) are the T.V. programs you usually watch?
2. In what language(s) are the radio programs you usually listen to?

A 5-point likert-type scale is assigned to each question with 1 =Only Spanish / All Hispanic/Latino and 5= Only English / All non-Hispanic/non-Latino. The average of the 10 questions indicates the degree of acculturation.

Missing if more than 2 of the 10 items (SCEA1-10) are missing
Otherwise,
SASH = average(of scea1-10);

Source variable(s):

SCEA1. In general, what language(s) do you read and speak?
SCEA2. What was the language(s) you used as a child?
SCEA3. What language(s) do you usually speak at home?
SCEA4. In which language(s) do you usually think?
SCEA5. What language(s) do you usually speak with your friends?
SCEA6. In general, in what language(s) are the movies, T.V. and radio programs you prefer to watch and listen to?
SCEA7. Your close friends are:
SCEA8. You prefer going to social gatherings/parties at which the people are:
SCEA9. The persons you visit or who visit you are:
SCEA10. If you could choose your children's friends, you would want them to be:

1.5 ANTHROPOMETRY

1.5.1 BMI (Body Mass Index kg/m2) for HCHS/SOL Baseline Examination

This is a numeric variable for HCHS/SOL baseline Body Mass Index.

Missing if the variable, HEIGHT, is missing.
If HEIGHT is not missing then BMI = ANTA4 / $(HEIGHT/100)^2$

Source variable(s):

HEIGHT. Corrected height in centimeters
ANTA4. Weight in kilograms

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

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

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

Response format: 1=Underweight
2=Normal
3=Overweight
4=Obese

Source variable(s):

BMI. Body Mass Index in kg/m^2 in HCHS/SOL baseline.

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

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

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

Response format: 1=Underweight
2=Normal
3=Overweight
4=Obese I
5=Obese II
6=Obese III

Source variable(s):

BMI. Body Mass Index in kg/m^2 in HCHS/SOL baseline.

2. SOLNAS Derived Variables

2.1 Administrative

2.1.1 DATEV1 (Date for SOLNAS visit 1 Main Study)

DateV1=VSEA0a when visit=1 in VSE.

Source variable(s):

VSEA0a. Completion date for VSE visit 1.

2.1.2 DATEV2 (Date for SOLNAS visit 2 Main Study)

DateV2=VTEA0a when visit=2 in VTE.

Source variable(s):

VTEA0a. Completion date for VTE visit 2.

2.1.3 DATEV3 (Date for SOLNAS visit 3 Reliability Study)

DateV3=VSEA0a when visit=3 in VSE.

Source variable(s):

VSEA0a. Completion date for VSE visit 3.

2.1.4 DATEV4 (Date for SOLNAS visit 4 Reliability Study)

DateV4=VTEA0a when visit=4 in VTE.

Source variable(s):

VTEA0a. Completion date for VTE visit 4.

2.1.5 DAYS_V1M1 (Days between HCHS clinic visit and SOLNAS visit 1)

Days_V1M1=Clindate – VSEA0a (where visit =1).

Source variable(s):

Clindate. HCHS/SOL baseline clinic visit date
VSEA0a. Completion date for VSE visit 1.

2.1.6 DAYS_V2V1 (Days between SOLNAS visits 1 and 2- Main Study)

Days_V2V1= VTEA0a (where visit=2) - VSEA0a (where visit =1).

Source variable(s):

VSEA0a. Completion date for VSE visit 1.
VTEA0a. Completion date for VTE visit 2.

2.1.7 DAYS_V3V1 (Days between SOLNAS visits 1 and 3)

Days_V3V1= VSEA0a (where visit=3) - VSEA0a (where visit =1).

Source variable(s):

VSEA0a. Completion date for VSE visit 1 and visit 3.

2.1.8 DAYS_V4V3 (Days between SOLNAS visits 1 and 3- Reliability Study)

Days_V4V3= VTEA0a (where visit=4) - VSEA0a (where visit =3).

Source variable(s):

VSEA0a. Completion date for VSE visit 3.
VTEA0a. Completion date for VTE visit 4.

2.1.9 IN_Main (Completed records in SOLNAS Main Study, in both visits 1 and 2)

It's binary variable showing the status that whether the participants complete both visits 1 and 2 in SOLNAS Main study).

When visit=1 in VSE and visit=2 in VTE, then set IN_Main=1.
Otherwise, set it as 0;

Source variable(s):
VISIT variables in VSE and VTE forms.

2.1.10 IN_Reliab (Completed records in SOLNAS reliability Study, in both visits 3 and 4)

It's binary variable showing the status that whether the participants complete both visits 3 and 4 in SOLNAS Reliability study).

When visit=3 in VSE and visit=4 in VTE, then set IN_Reliab=1.
Otherwise, set it as 0;

Source variable(s):
VISIT variables in VSE and VTE forms.

2.1.11 VISIT1 (In SOLNAS Main Study visit 1 based on VSEA)

Set visit1=1 when Visit=1 in VSEA data. Otherwise, set it as 0;

Source variable(s):
VISIT variable in VSEA form.

2.1.12 VISIT2 (In SOLNAS Main Study visit 2 based on VTEA)

Set visit1=1 when Visit=2 in VTEA data. Otherwise, set it as 0;

Source variable(s):
VISIT variable in VTEA form.

2.1.13 VISIT3 (In SOLNAS Reliability Study visit 3 based on VSEA)

Set visit1=1 when Visit=3 in VSEA data. Otherwise, set it as 0;

Source variable(s):
VISIT variable in VSEA form.

2.1.14 VISIT4 (In SOLNAS Reliability Study visit 4 based on VTEA)

Set visit1=1 when Visit=4 in VTEA data. Otherwise, set it as 0;

Source variable(s):
VISIT variable in VTEA form.

2.1.15 VSEA2 (SOLNAS Consent Signed in SOLNAS 1st visit, Informed consent status for participation in SOLNAS study)

This is a binary variable that confirms whether or not a study participant consented to participate in the SOLNAS study during the first SOLNAS visit.

Source variable(s):
VSEA2. SOLNAS Consent Signed (yes/no)

2.2 SOLNAS ANTHROPOMETRY

2.2.1 BMIV1 (Body Mass Index updated in SOLNAS study visit 1)

Missing if the variable, VSEA3a (HEIGHT), is missing.
If VSEA3a is not missing then BMI = VSEA3b/ (VSEA3a /100)^2

Source variable(s):
VSEA3a. Corrected height in centimeters
VSEA3b. Weight in kilograms

2.2.2 BMIGRPV1_C4 (4-level grouped Body Mass Index – in SOLNAS visit 1)

This variable is created using the calculated body mass index updated in SOLNAS 1st visit to define categories based on current WHO classifications.

if 0≤BMIV1<18.5 then BMIGRPV1_C4=1;
else if 18.5≤BMIV1<25 then BMIGRPV1_C4=2;
else if 25≤BMIV1<30 then BMIGRPV1_C4=3;
else if 30≤BMIV1 then BMIGRPV1_C4=4;
else if BMIV1=.Z then BMIGRPV1_C4=.;

Response format: 1=Underweight
2=Normal
3=Overweight
4=Obese

Source variable(s):

BMIV1. Body Mass Index in kg/m^2, updated in SOLNAS 1st visit.

2.2.3 BMIGRPV1_C6 (6-level grouped Body Mass Index - WHO)

This variable is created using the calculated body mass index updated in SOLNAS 1st visit to define categories based on current WHO classifications.

if 0≤BMIV1<18.5 then BMIGRPV1_C6=1;
else if 18.5≤BMIV1<25 then BMIGRPV1_C6=2;
else if 25≤BMIV1<30 then BMIGRPV1_C6=3;
else if 30≤BMIV1<35 then BMIGRPV1_C6=4;
else if 35≤BMIV1<40 then BMIGRPV1_C6=5;

else if 40≤BMIV1 then BMIGRPV1_C6=6;
else if BMIV1=.Z then BMIGRPV1_C6=.;

Response format: 1=Underweight
2=Normal
3=Overweight
4=Obese I
5=Obese II
6=Obese III

Source variable(s):

BMIV1. Body Mass Index in kg/m^2, updated in SOLNAS 1st visit.

2.3 SOLNAS AS LABORATORY MEASURES

2.3.1 CNTLAB_QC (Flag variable for Central Lab QC completion)

IF VTEA4=1 and (VTEA11a-11g,VTEA11j,11k & 11l ALL=1) then set CNTLAB_QC=1;
Otherwise set it as 0;

Source variable(s):

VTEA4. Is participant in QC Sample (VTE visit=2)?
VTEA11a-11g, VTEA11j, VTEA11k & VTEA11l Quality control specimen inventory Checklist items for Serum, Plasma and Urine specimen.

2.3.2 DLW_QC (Flag variable for Doubly Labeled Water QC completion)

IF (VSEA5=1 & VSEA16a=1 & VSEA16b=1 & VSEA16c=1)
and (VTEA11h=1 & VTEA11i=1) then set DLW_QC=1;
Otherwise set it as 0;

Source variable(s):
VSEA5. Is participant in QC Sample (VSE visit=1)?
VSEA16a – VSEA16c. Quality control inventory Checklist for Times 0, 3 & 4 Urine specimens in VSE.
VTEA11h & VTEA11i. Quality control specimen inventory Checklist items for Times 0 & 1 Urine specimens in VTE.

2.3.3 LABID (SOLNAS Main Study Lab ID assigned in SOLNAS Visits 1 or 2)

VSEA4 in visit 1and VTEA3 in visit 2 make up each other for missing.

Source variable(s):
VSEA4. Lab ID in VSE when visit=1
VTEA3. Lab ID in VTE when visit=2.

2.3.4 LABID_QC (SOLNAS QC Lab ID assigned in SOLNAS Visits 1 or 2)

QC ID were only assigned in SOLNAS Main study. VSEA5a in visit 1and VTEA4a in visit 2 make up each other for missing cases.

Source variable(s):
VSEA5a. QC Lab ID in VSE when visit=1
VTEA4a. QC Lab ID in VTE when visit=2.

2.3.5 LABID_Reliab (SOLNAS Reliability Study Lab ID assigned in SOLNAS Visit 3)

VSEA4 in visit 3 has more completed ID assignments so use it only for reliability study Lab ID.

Source variable(s):
VSEA4. Lab ID in VSE when visit=3

2.4 SOLNAS Flags for Exclusions

2.4.1 Excl_urineLE500_M

This is a 0/1 variable indicating that in SOLNAS Main study the total 24hr urine collection <= 500.

If 0 <= VTEA7A <= 500 then excl_urineLE500 = 1; else excl_urineLE500 = 0;
Response format: 1=Yes, 0=No

Source variable:
VTEA7A: 24-HOUR URINE COLLECTION total weight.

2.4.2 Excl_MissUrineGE2_M

This is a 0/1 variable indicating participant reported two or more missed urine collections (UCEA7) in SOLNAS Main study.

If UCEA7 >= 2 then excl_MissUrineGE2 = 1; Else excl_MissUrineGE2 = 0;
Response format: 1=Yes, 0=No

Source variable:
UCEA7: If you missed urine collections, how many did you miss?

2.4.3 Excl_urineLE500_R

This is a 0/1 variable indicating that in SOLNAS reliability study the total 24hr urine collection <= 500.

If 0 <= VTEA7A <= 500 then excl_urineLE500 = 1; else excl_urineLE500 = 0;
Response format: 1=Yes, 0=No

Source variable:
VTEA7A: 24-HOUR URINE COLLECTION total weight.

2.4.4 Excl_MissUrineGE2_R

This is a 0/1 variable indicating participant reported two or more missed urine collections (UCEA7) in SOLNAS reliability study.

If UCEA7 >= 2 then excl_MissUrineGE2 = 1; Else excl_MissUrineGE2 = 0;
Response format: 1=Yes, 0=No

Source variable:
UCEA7: If you missed urine collections, how many did you miss?