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)

Investigator Use Database Overview
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

ACKNOWLEDGMENTS

SOLNAS Study would not have been possible without the participation of HCHS/SOL Study participants and the support of study member staff. The Study of Latinos: Nutrition & Physical Activity Assessment Study (SOLNAS) was supported by grant #11414319, R01HL095856-01, R01HL095856-02, R01HL095856-03, R01HL095856-04 from the National Heart, Lung, and Blood Institute (NHLBI).

Special thanks to:

Abreu, Maria de los Angeles, MPA, CCPR
UNC at Chapel Hill

Austin, Mary
Nutrition Center-University of Minnesota

Beasley, Jeanette M., PhD
Albert Einstein College of Medicine, Bronx, NY

Butte, Nancy, Ph.D.
Baylor College of Medicine

Chambers, Claudia P., BPh
University of Miami

Denenberg, Julie, MA
UC at San Diego

Gellman, Marc D., PhD
University of Miami

Gernhofer, Niki, Nutritionist
University of Illinois, Chicago

Himes, John H., PhD, MPH
Nutrition Center-University of Minnesota

Lopez-Pino, Daniel, MA, MPH
Einstein College of Medicine, Bronx, NY

Leiendecker Foster, Catherine, MS, MLS (ASCP), MB - University of Minnesota

Martello, Maria, MS
Albert Einstein College of Medicine, Bronx, NY

Matthews, Charles, Ph.D., NCI

Mossavar-Rahmani, Yasmin, PhD, RD, CDN
Albert Einstein College of Medicine, Bronx, NY

Oceguera, Alejandra
UC San Diego/San Diego State Univ.

Prentice, Ross, Ph.D.,
Fred Hutchinson Cancer Research Center, Seattle, WA

Rock, Cheryl, PhD, RD
UC at San Diego

Shaw, Pamela A., PhD
Perelman School of Medicine, U. of Pennsylvania

Siega-Riz, Anna Maria, PhD, RD, LDN
UNC at Chapel Hill

Sotres-Alvarez, Daniela, DrPH
UNC at Chapel Hill

Stevens, Mary
Nutrition Center-University of Minnesota

Stoutenberg, Mark, Ph.D.
University of Miami

Talavera, Ana C., MPH
San Diego State University

Teng, Yanping, MD, MSPH
UNC at Chapel Hill

Thyagarajan, Bharat, PhD
Central Lab-University of Minnesota

Van Horn, Linda, Ph.D., RD
Northwestern Univ., Chicago

William W. Wong, Ph.D., C.N.S., F.A.C.N.
Baylor College of Medicine

Wylie-Rosett, Judith, Ed.D., RD
Albert Einstein College of Medicine, Bronx, NY

Study of Latinos Nutrition Assessment Study (SOLNAS)
Investigator Use Database
Ver. 2.0 (Oct 2014)

Updates to SOLNAS Data Release or Documentation

Version	Date	Description	Datasets	Documentation
1	10/28/2013	1st data release	_INV1	V1 (Oct 2013)
2	10/31/2014	Data released for 1st time: - Multicultural Diet FFQ (only Bronx) - Sugar biomarkers (fructose, sucrose, glucose) from Jeannette Beasley AS -Other biomarkers: carotene, tocopherol, zeaxanthin, cyptoxanthin, lycopene, B12, folate, retinol and Retinyl Palmitate; - PABA results Food Groups Servings/Day (NDSR 9) - QC_LAB and QC_DLWA Data corrections: - Two to VEEA2 (gender) - Two to VTEA2A (measured weight)	_INV2	V2 (Oct 2014)

1. INTRODUCTION

This document describes the content and structure of the Investigator Use datasets created for the Nutrition and Physical Activity Assessment Study (SOLNAS) HCHS/SOL Ancillary Study. This database contains all the data collected for the 485 adults that participated in SOLNAS Visit 1, subject to constraints (described within) to preserve participant confidentiality by de-identifying the data. HCHS/SOL main study data included in this database release is a limited number of socio-demographic and acculturation variables which appear in the SOLNAS derived variable file.

2. SOLNAS STUDY AIMS

The study was designed to enroll 476 participants from HCHS/SOL. The primary aims are:

Aim 1. To compare energy and protein data from the 24-hr dietary recall to the gold standard biomarkers Doubly Labeled Water (DLW) for energy and urinary nitrogen for protein in the HCHS/SOL study; to compare physical activity energy expenditure data from study questionnaires to the Actical (an accelerometer for measuring physical activity), DLW and indirect calorimetry.

Aim 2. To contrast measurement error properties of: (i) the 24-hr dietary recall; (ii) the 24-hr dietary recall with the addition of the Food Propensity Questionnaire (FPQ) (iii) HCHS/SOL Physical Activity Questionnaire; (iv) the Multi-Cultural Food Frequency Questionnaire (formerly known as Tufts University Food Frequency Questionnaire.) The latter was only administered at the Bronx site which consists primarily of Puerto Rican & Dominican participants.

Aim 3. To use the fitted measurement error model to produce calibrated intake and physical activity measures on the full HCHS/SOL cohort for use in analyses of clinical outcomes. For this population, clinical outcomes of interest include cardiovascular disease and self-reported diabetes.

3. STUDY DESIGN

To address SOLNAS study aims each of the four HCHS/SOL field centers recruited approximately 120 persons from HCHS/SOL cohort. The age range is 18-79, and recruitment was monitored to ensure: 1) an equal distribution of participants at each field center, 2) a minimum of 60 participants in each Hispanic/Latino subgroup; 3) an age, sex and BMI group (under/normal weight, overweight, obese) distribution that is representative of the parent study. For a subset of 96 participants (the reliability subsample), the full protocol was repeated approximately 6 months later. **Figure 3.3 provides an overview of SOLNAS;** the four-letter mnemonic refers to the dataset name described in section 5**.** Electronic copies of the **study protocol and manuals of operation** are also included for reference with this data release.

3.1. Participants

All study participants were invited to participate within seven months of completing HCHS/SOL study baseline visit provided they have completed the second dietary recall. SOLNAS staff obtained potentially eligible participants from HCHS/SOL Coordinating Center. Language barrier was not a reason for exclusion. All contact with participants that are Spanish speakers not proficient in English was done using the appropriate language. Before scheduling SOLNAS Visit 1, interested participants received a telephone call for further screening (form TSEA) such as participating in a weight loss program, experienced more than 15 lb weight loss in the four weeks preceding their visit, started taking diabetic medications or had bladder control problems. During the phone screening, they were also asked about temporary conditions.

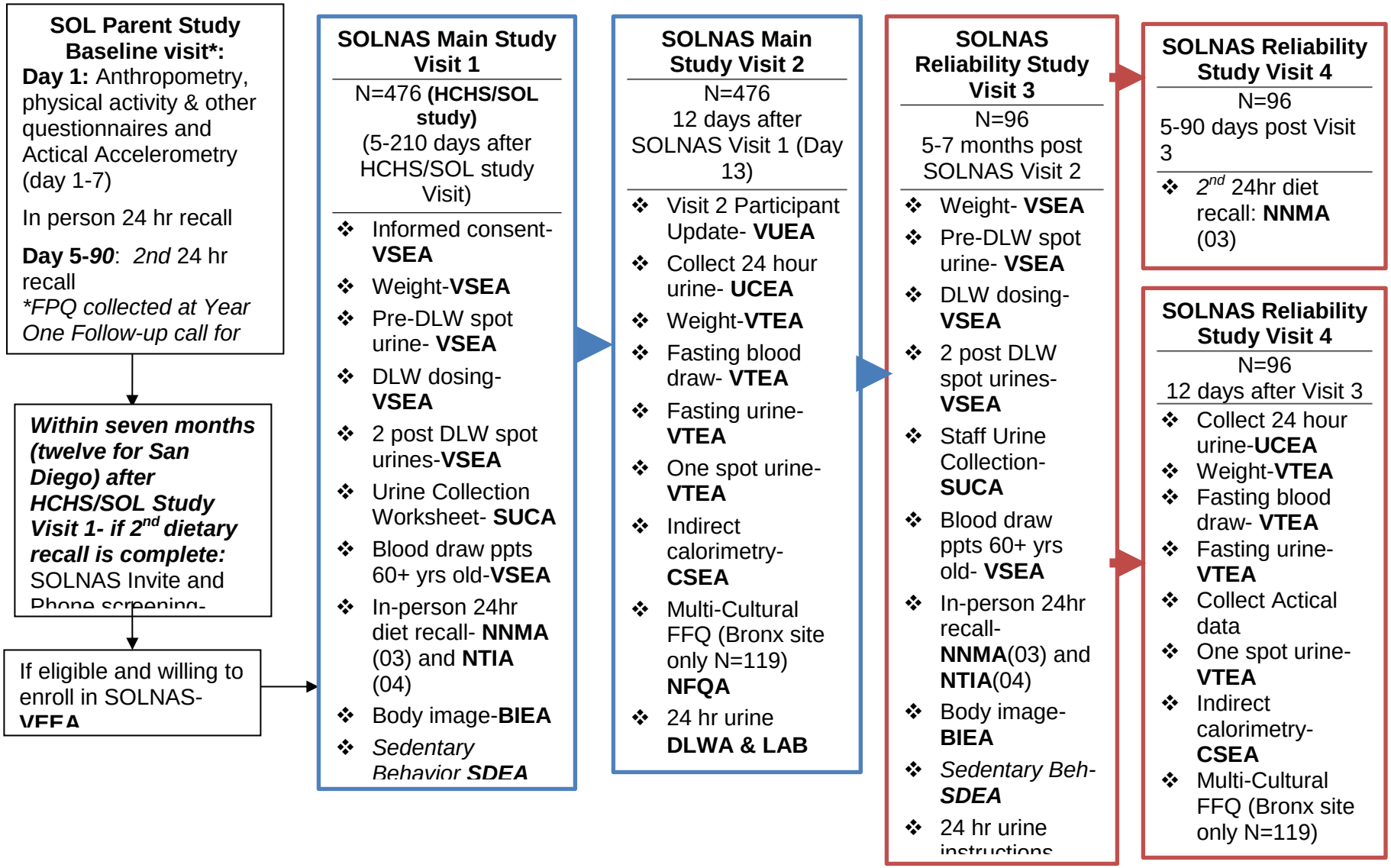
3.2. Questionnaires and Procedures

Table 1. Key fields and number of records per dataset

Dataset	Description	Key fields	Number of records		
			Total	Main	Reliability
Screening					
TSEA	Telephone Screening Interview	ID + SOLNAS_STUDY	1,479	1358	121
Visit 1 & 3					
VEEA	Visit 1 & 3 Eligibility Form	ID + SOLNAS_STUDY	585	487	98
VSEA	THE SOLNAS Visit 1 & 3 Form	ID + SOLNAS_STUDY	583	485	98
SUCA	Urine Collection Worksheet for Staff	ID + SOLNAS_STUDY	577	480	97
BIEA	Body Image Questionnaire	ID + SOLNAS_STUDY	583	485	98
SDEA	Sedentary Behavior Questionnaire	ID + SOLNAS_STUDY	583	485	98
Visit 2 & 4					
VUEA	Visit 2 & 4 Update Worksheet	ID + SOLNAS_STUDY	577	482	95
UCEA	Record Sheet 24hr Urine Collection	ID + SOLNAS_STUDY	578	481	97
VTEA	THE SOLNAS Visit 2 & 4 Form	ID + SOLNAS_STUDY	574	478	96
CSEA	Calorimetry Summary Form	ID + SOLNAS_STUDY	572	477	95
BIOMARKERS					
DLWA[1]	Doubly Labeled Water	ID+SOLNAS_STUDY+ SAMPLESEQ	3,017	2,504	513
QC_DLWA	Blinded duplicates for DLW lab	ID+SOLNAS_STUDY+ SAMPLESEQ	135	135	0
LAB_ SOLNAS	Other (e.g. protein, sodium, potassium,etc)	ID+SOLNAS_STUDY	573	477	96
QC_LAB_ SOLNAS	Blinded duplicates for Central Lab	ID+SOLNAS_STUDY	49	49	0
24HR DIETARY RECALLS (NDSR Files)					
NNMA[2]	NDSR03 Nutrients at the Meal Level	ID + SOLNAS_VISIT + MEALID	3,499	2,549	950
NTIA[2]	NDSR04 Nutrients at the Daily Level	ID + SOLNAS_VISIT	673	485	188
N09A[2]	NDSR09 Food group servings at the Daily Level	ID + SOLNAS_VISIT	673	485	188
Multicultural Diet FFQ (Bronx only)					
NFQA	Daily Total Nutrient from Multicultural FFQ	ID + SOLNAS_VISIT	143	119	24
Physical Activity (PA) Accelerometer data					
PA_CNTS	Daily Counts per Minute (day level)	ID + DAY	539	NA	539
PA_DERV	PA Derived Variables (participant level)	ID	92	NA	92
Derived Variables					
PART_ DERV_ SOLNAS	Participant Derived Variables	ID	485	485	NA

1 DLWA is multi-record dataset (key fields are: ID+ SOLNAS_STUDY+SAMPLESEQ) with 7 or 5 records per participant depending whether they had blood drawn (participants 60 or older) or not.
2 The reliability study had two 24hr dietary recalls.

3.3. Diagram of SOLNAS Study Design and Protocol Overview



4. DATABASE STRUCTURE

4.1. Data Set Organization

There is one table (SAS data set) in the database for each type of data collection form (provided as PDFs). The data values from one completed paper form are stored in one record in the corresponding table (observation in the SAS data set). Each data item on a paper form is stored as one or more columns (variables) in the data set. Collection of direct measurements during examination procedures can also result in the creation of a data file.

A special derived variable dataset (PART_DERV_SOLNAS_INV2) has been created with SOLNAS specific variables (e.g. BMI using weight and height from SOLNAS VSEA form) and with some HCHS/SOL main study variables such as socio-demographic and acculturation variables. These variables are defined and described in a separate document called **"SOLNAS Derived Variable Dictionary INV2".**

A codebook has been produced for each data set by SOLNAS Main Study or SOLNAS Reliability Study. A careful review of the codebooks, in conjunction with the forms, is critical to interpreting the data. The codebook provides a description of every variable in the data set as well as the frequency and meaning of variables' values. Analysts are *strongly* encouraged to use the codebooks, paying attention to the data user notes contained in this document.

4.2. Form and Data Set Naming Conventions

Each SOLNAS data collection instrument (PDF form) has a unique four-letter mnemonic associated with it (e.g., TSEA is the mnemonic for the SOLNAS Telephone Screening Interview form, Version A). In SOLNAS, all forms have only one version (A). Corresponding data sets begin with the same first four letters of the mnemonic and a character string "INV2" for Investigator Use Version 2. For example, the data set for Telephone Screening Interview investigators-version1 is "TSEA_INV2". Note, since the questionnaire battery for the ancillary study has both English and Spanish language versions of the forms each has been merged into one common data record format which follows the main HCHS/SOL study conventions. For example, the VTE and the VTS both map to the VTEA_ INV2 in this data release. The variable FORM in each dataset has the unique four-letter mnemonic and allows distinguishing which forms were completed in English and which ones in Spanish.

4.3. Key Fields for Data Records

The unique identification of a participant data record within a file is determined by two primary key fields for forms that are collected once per visit (see SOLNAS MOP). These key fields are:

1) **ID**: A random 8-digit identification code unique to each participant. This ID is the same identifier from HCHS/SOL Main Study. Therefore, SOLNAS datasets can be merged with HCHS/SOL datasets by ID.

2) **SOLNAS_STUDY:** Identifies whether the form was administered at SOLNAS Main Study ("M") or at SOLNAS Reliability Study ("R").

In addition, datasets include variable **SOLNAS_VISIT** which is the SOLNAS consecutive visit number (1 and 2 for SOLNAS Main Study, and 3 and 4 for Reliability Study). **SOLNAS_STUDY** and **SOLNAS_VISIT** are equivalent and in combination with ID uniquely identify a record (except for 24hr recalls in the Reliability Study where there are two 24hr dietary recalls and SOLNAS_VISIT identifies the recall). For example an individual that participated in the Reliability Study has two records in VSEA: one record with SOLNAS_STUDY="M" (SOLNAS_VISIT=1) and another record with SOLNAS_STUDY="R" (SOLNAS_VISIT=3). This same participant has 3 records in dietary dataset NTIA (nutrients at the daily level NDSR file 03) one for each SOLNAS visit: 1, 3 and 4.

NOTE: Other datasets (PA_CNTS, NNMA, and DLWA) have other key fields (see Table 1 in section 3.2.

4.4. Common Variables Across Data Sets

VERSION: Version of the data collection form. A one character variable indicating which version of the paper form was used to collect the data. Possible values for VERSION are "A", "B", and "C", representing the first, second, and third versions, respectively. Version remained a constant in the ancillary study. In SOLNAS all forms are version "A".

FORM: The original 3-letter form code that appears on the paper-based forms or on the form code selection menu in the DMS uses the convention of having the third letter designate the language version in use. Use this variable to detect changes in language of administration. The standard taken from the main study uses "E" for English language forms versus "S" for the Spanish language version (VTE vs. VTS for the Visit 1 or 3 Form).

4.5. Variable Naming Conventions

While the key field and sort variables (ID, SOLNAS_VISIT, SOLNAS_STUDY) have the same name on each SAS file (see Sections 4.3 and 4.4), other SAS variables are unique to a specific form. To predictably and uniquely link data items to forms, these form-specific variable names begin with the same three characters as the data set name, followed by the form version letter, and then the question number as indicated on the form. For example,

variable for question 3 (gender) of the Telephone Screening form is named TSEA3 on the corresponding SAS file TSEA_ INV2.

4.6. Changes to Variables to Preserve Confidentiality

As part of the study commitment to complying with HIPAA regulations for participant confidentiality and in following guidelines from NHLBI/NIH the Coordinating Center has made explicit modifications and/or deletions to variables that were common across all forms. All participant ID values were transformed from the original ID to random values to produce Investigator Use data files that protect the confidentiality of the individual. However, the authorized user will need to actively attend to the security and confidentiality of these Investigator Use files as part of the end user agreement.

1) HCHS/SOL ID (same ID used in SOLNAS) was re-derived for use in all data sets as a random identifier code for participants.

2) Addresses, phone numbers, and SSN of the participants were omitted from these files.

3) CENTER is a real code to distinguish among participating field centers created in the Participant derived variable set PART_DERV_SOLNAS_INV2.

4) STAFF ID codes were deleted across all forms and not substituted.

5) DATES were kept unaltered.

4.7. Missing Values

The study database employs a standard set of special missing value codes (see study codebook) that have contextual meaning. Since SAS allows numeric variables to assume up to 27 unique missing values, ".A to .Z, and ." the Coordinating Center uses several of these special missing codes to convey additional meaning to the analyst. Here is a table that describes that usage of missing values in HCHS/SOL.

Missing value	Meaning
. or blank	Empty field, missing
.Q	Don't know / refused
.S	Skipped field
.L	Below lower limit of analysis
.H	Above higher limit of analysis

Selective recodes may need to be made to make use of known refusals, or to account for skip patterns in coding derived variables based on multiple items in a form. Laboratory variables with results reported as "< number", or "> number" for values below or above the assay limits are set to the special values of ".L" or ".H".

5. DESCRIPTION OF DATA COLLECTION FORMS / DATABASE TABLES

5.1. Telephone Screening Interview (TSE)

This 19 item instrument performs an initial screening that collects exclusion criteria information from participants included in the eligibility lists provided by the CC. Reasons for exclusion include: if a participant is taking insulin or medications for diabetes, is currently participating in a weight loss program or has experienced more than 15 lb weight loss in the four weeks preceding their visit, has recently started taking diabetic medications or has bladder control problems, or has other temporary conditions.

5.2. Visit 1 & 3 FORMS

5.2.1. Visit 1 & 3 Eligibility Form (VEE)

This 11 item instrument evaluates if the participant continues to be eligible at the time of the first visit. Information is collected on changes in the information previously provided on the Telephone Screening questionnaire that would make the participant ineligible at time of Visit 1 (e.g. began taking insulin or medications for diabetes, pregnancy, etc.)

5.2.2. THE SOLNAS Visit 1 & 3 FORM (VSE)

This 16-item instrument records height, weight, DLW dosage information, and bio-specimen data. It collects lab ID #'s for regular samples and QC samples including times of collection. It also contains the Specimen Inventory Checklist for samples.

5.2.3. Urine Collection Worksheet for Staff (SUC)

This 3 item instrument collects data on the weight of the 24-hour urine collection specimen completed by participants between visits.

5.2.4. Body Image Questionnaire (BIE)

This 15-item questionnaire assesses a participant's perception of their body image by having them select their body image from a card containing images. It provides participants assessment in self-perception of his/her body size, his/her desired size and the healthiest size and provides participant's satisfaction measurement with his/her body size.

5.2.5. Sedentary Behavior Weekday and Weekend Questionnaire (SDE)

The 18-item questionnaire measures a participant's sedentary activities during the week and weekend.

5.3. Visit 2 & 4 FORMS

5.3.1. Visit 2 or 4 Participant Update Worksheet (VUE)

This 8 item instrument evaluates if a participant is still eligible at the time of the second visit. If participants are not eligible, they may be rescheduled.

5.3.2. SOLNAS Visit 2 or Visit 4 Form (VTE)

This 16-item instrument records height, weight, urine collection and blood collection data. It collects lab ID #'s for regular samples and QC samples including times of collection. It also contains the number of PABA tablets taken by participants and the Specimen Inventory Checklist for samples.

5.3.3. Record Sheet for 24-hour Urine Collection (UCE)

This 12-item questionnaire collects data on the 24-hour urine collection completed by participants. This includes the time collected, the consumption times of the PABA-B-Vitamin tablets, and if participants missed any collections.

5.3.4. Calorimetry Summary Form (CSE)

This 5-item instrument collects participant calorimetry summary data.

5.4. RECOVERY BIOMARKERS

5.4.1. Energy Biomarker: Doubly Labeled Water (DLW)

This dataset from the Reading Center at Baylor College of Medicine Children's Nutrition Research Center, has the total energy expenditure in kcal/day over the 2 weeks of the Study (variable DLWA33) among other variables. **This is a multi-record dataset with key fields: ID, SOLNAS_STUDY (M or R) and SAMPLESEQ (consecutive number from 1 to 7 that maps to DLWA4).** Values were calculated using SOLNAS participants' data collected from DLSA, SPFA, VSEA, VTEA, CSEA, TSEA and VEEA forms (see appendix A for a full description of the dataset).

SampleSeq	DLWA4	Description
1	U30	Vial 30 (Visit 1 spot urine time 0 - fasting)
2	U33	Vial 33 (Visit 1 spot urine time 3 – 3 hrs after DLW)
3	U34	Vial 34 (Visit 1 spot urine time 4 – 4 hrs after DLW
4	U50	Vial 50 (Visit 2 spot urine time 0 - fasting)
5	U51	Vial 51 (Visit 2 spot urine time 1 – 1 hr after fasting urine)
6	P40	Vial 40 (Visit 1 plasma aliquot for 60+)
7	P20	Vial 20 (Visit 2 plasma aliquot for 60+)

- Data provided by the DLW Lab are from variable DLWA20 to DLWA38.
- Only variables DLWA20 (Number of days sample was collected since dosing), DLWA22 (MEAN ^{2}H (^{o}/oo)) and DLWA22 (MEAN ^{18}O (^{o}/oo)) will have data for all sample collections (i.e. key field SampleSeq 1 to 5 corresponding to U30 to U51, and when available, for SampleSeq = 6 and 7 for P40 & P20, respectively)
- Only SampleSeq=1 (DLWA4='U30') has data for ALL variables (DLWA20 to DLWA37), and DLWA38 (Ratio of TEE by plasma over TEE by urine; only older subjects w/plasma) is only present if SampleSeq = 6 and 7 (P20 & P40) are available (i.e. blood drawn because participant is 60 or older)
- Only SampleSeq=6 (DLWA4='P40'), when available, has data from variable DLWA23 to DLWA38. Note that, participants 60 or older will have DLWA33 (TEE0-12 days (kcal/d) Total Energy Expenditure) calculated for both urine and plasma.
- Variables DLWA9 to DLWA10 and DLWA16 to DLWA19 were purposely dropped from data release since these variables are available directly from SOLNAS forms.

5.4.2. Nutrient Recovery Biomarkers from Urine and Serum (LAB_SOLNAS)

The central clinical chemistries laboratory at University of Minnesota Fairview Hospital Clinical Laboratory analyzed most of the biomarkers (cholesterol, triglycerides, urine potassium, nitrogen, urine sodium, vitamin B12, folate, zeaxanthin, cryptoxanthin, lycopene, alpha-carotene, beta-carotene, alpha-tocopherol, gamma-tocophero, retinol, and retinyl palmitate. The urine and serum samples were collected on Visit 2-for main study, and Visit 4-for reliability study. **The sugar biomarkers** (fructose, glucose and sucrose) were analyzed in University of Hawaii by Adrian Franke as part of an ancillary study to SOLNAS awarded to Dr Jeannette Beasley. **PABA results** were provided by the Dr JW Lampe from the Cancer Prevention Program at the Fred Hutchinson Cancer Research Center.

5.5. 24HR DIETARY RECALLS (NDSR FILES)

These files based on the 24 hour dietary recalls contain information on nutrients from foods consumed by the participants. Files report nutrient intake at different levels. **Please see HCHS/SOL Dietary Data Overview, Methods and guidelines for details.**

5.5.1. NDSR File 03 – Nutrients at the meal level (NNM)

This file, based on the dietary 24 hour recalls, contains nutrient totals for each meal or eating occasion within each dietary recall. **It contains multiple records per participant; key fields are ID + SOLNAS_VISIT + MEALID.** Details are documented in HCHS/SOL Dietary Data Overview, Methods and guidelines for details.

5.5.2. NDSR File 04 – Nutrients at the daily totals level (NTI)

This file, based on the dietary 24 hour recalls, contains nutrient totals at the daily level for each dietary recall. It also contains general information about each recall (e.g. day of intake, self-report intake amount, intake reliability assessed by the interviewer). **It contains multiple records per participant; key fields are ID + SOLNAS_VISIT.** Details are documented in HCHS/SOL Dietary Data Overview, Methods and guidelines for details.

5.5.3. NDSR File 09 – Nutrients at the daily totals level (N09)

This file, based on the dietary 24 hour recalls, contains serving counts at the daily level for each dietary recall. There is one serving count for each of the 168 NDSR food groups; see Appendix 10 of the 2011 NDSR Manual for serving counts. **It contains multiple records per participant; key fields are ID + SOLNAS_VISIT.** Details are documented in HCHS/SOL Dietary Data Overview, Methods and guidelines for details.

5.6. Multicultural Food Frequency Questionnaire- Bronx only (NFQ)

This dataset has the daily total nutrient from the Multicultural FFQ. This instrument was only administered by the Bronx Field Center and both on Visit 2-for SOLNAS Main study and Visit 4-for SOLNAS Reliability study.

5.7. Physical Activity from Accelerometer

Please see HCHS/SOL PA data Overview, Methods and guidelines for details.

5.7.1. Physical Activity from Actical (daily counts/minute) (PA_CNTS_SOLNAS)

PA_CNTS is a dataset with multiple records per participant (key fields are ID and DAY); one record per participant ID per worn DAY for the monitor. Thus each ID has between 1 and 6 records, depending on the number of days they wore the Actical. Day 1 is the day after the clinic visit. The file has the original counts (variables CNT1 to CNT1440) for each minute (epoch length) of a calendar day (24*60 = 1440) where CNT1 corresponds to counts starting at midnight (24:00), CNT720 corresponds to counts before noon (11:59), and CNT1440 corresponds to counts at 23:59 hours. Details are documented in HCHS/SOL Physical Activity Data Overview, Methods and Guidelines, and HCHS/SOL Physical Activity Data Dictionary.

5.7.2. Actical derived variables at the participant level (PA_DERV_SOLNAS)

This dataset has one record per participant (key field is ID) with averaged physical activity (counts/day) for those who have at least three adherent days (≥10 hrs). Those participants who do not have at least three adherent days have missing values for the

averaged values and other derived variables from their studies. These data are derived from PA_CNTS (objectively measured physical activity at the DAY level). An indicator variable (ADHERENTYN) identifies participants with at least three adherent days. Details are documented in HCHS/SOL Physical Activity Data Overview, Methods and Guidelines, and HCHS/SOL Physical Activity Data Dictionary.

5.8. SOLNAS PARTICIPANT DERIVED VARIABLES (PART_DERV_SOLNAS)

The participant derived variable data sets are not associated solely with any particular form because they contain variables from many forms. There is one record per enrolled participant (485 observations). This file is a cross-section of "derived variables" whose values are defined based on combinations of data items (e.g. age from date of birth, or body mass index from height and weight, etc.). See the separate document, "SOLNAS Derived Variable Dictionary" for the definitions of the variables. **There are no sampling weights calculated for this ancillary study. This validation study does not require using sampling weights in their statistical analyses.**

IMPORTANT ANALYSIS NOTE: In a few cases, inconsistencies or omissions in the information required to define these variables could not be corrected on the original data forms (and corresponding files in this database). These idiosyncratic cases were adjudicated by the SOLNAS Coordinating Center and their resolutions are included in the derived variable files.

Appendix. Doubly Labeled Water (DLWA) Data structure

ID	DLWA 1	DLWA 2	SOLNAS _STUDY	SampleSeq	DLWA4	DLWA5	DLWA6	DLWA 7	DLWA8 TO DLWA10	DLWA 11	DLWA 12	DLWA 13	DLWA 14	DLWA 15	DLWA16 TO DLWA19	DLWA20 to DLWA22	DLWA23 TO DLWA37	DLWA 38
Participant masked ID	Site	Lab ID	Main or reliability	Sample collection sequence (consecutive	Sample Collection Sequence	Sample Date (MM/DD/YYYY)	Sample Time (24 hr)	Baseline Y/N	Deleted in Data Release	Dose/Bottle ID	Dose amount (g)	FQ	Dosing time (24h)	Dosing Date (MM/DD/YYYY)	Deleted in Data Release			Ratio TEE plasma/ urine
ID	Site #	Lab ID	M or R	1	U30	VSEA0a	VSEA6a VSEA6a1	Y		VSEA7b	VSEA7a	0.86	VSEA7c VSEA7c1	VSEA0a		Data from DLW Lab	Data from DLW Lab	Only if blood drawn for 60+
ID	Site #	Lab ID	M or R	2	U33	VSEA0a	VSEA10a VSEA10a1	N		.	.	.	.	.		.	.	.
ID	Site #	Lab ID	M or R	3	U34	VSEA0a	VSEA13a VSEA13a1	N		.	.	.	.	.		.	.	.
ID	Site #	Lab ID	M or R	4	U50	VTEA5c	VTEA5a VTEA5b	N		.	.	.	.	.		.	.	.
ID	Site #	Lab ID	M or R	5	U51	VTEA9b	VTEA9a VTEA9a1	N		.	.	.	.	.		.	.	.
ID	Site #	Lab ID	M or R	6	P40	VSEA0a	VSEA11e VSEA11e1	Y		.	.	.	.	.		Data from DLW Lab	Data from DLW Lab	.
ID	Site #	Lab ID	M or R	7	P20	VTEA6e	VTEA6d VTEA6d1	N		.	.	.	.	.		.	.	.

DLWA has data only where specified, otherwise there is no data provided by the DLW Lab (represented here with a dot).
Note: For participants 60 or older there is TEE0-12 days (kcal/d) Total Energy Expenditure calculated both from urine and from plasma.
KEY VARIABLES (i.e. uniquely identify a record) are: ID + SOLNAS_STUDY + SAMPLESEQ