

NIMHD grant: Ancillary to HCHS/SOL: Gut Origins of Latino Diabetes Study  
(R01 MD011389; R Kaplan, RD Burk, & RD Knight)



# **HCHS/SOL GOLD Ancillary Study Derived Variable Dictionary**

**February 2019  
INV Version 2.0**

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

Yanping Teng  
Daniela Sotres-Alvarez

# HCHS/SOL GOLD Ancillary Study

## Investigator Use Database INV2 (Feb. 2019)

### Derived Variable Dictionary

#### Table of Contents

|                                    |   |
|------------------------------------|---|
| <b>1. DESIGN</b> .....             | 3 |
| 1.1 WEIGHT_EXPANDED_GOLD .....     | 3 |
| 1.2 WEIGHT_NORM_CENTER_GOLD .....  | 3 |
| 1.3 WEIGHT_NORM_OVERALL_GOLD ..... | 3 |
| <b>2. ADMINISTRATIVE</b> .....     | 4 |
| 2.1 KIT_ID .....                   | 4 |
| 2.2 BURK_ID .....                  | 4 |
| 2.3 CLINDATE_GOLD .....            | 4 |
| 2.4 GOLD_ENROLL .....              | 4 |
| 2.5 SAMLEYN .....                  | 5 |
| 2.6 REC_DATE .....                 | 5 |
| 2.7 SURVEY_DATE .....              | 5 |
| 2.8 S1_FREQ .....                  | 5 |
| 2.9 S2_TYPE .....                  | 5 |
| <b>3. SOCIO-DEMOGRAPHIC</b> .....  | 6 |
| 3.1 AGE_GOLD .....                 | 6 |
| 3.2 AGEGROUP_C6_GOLD.....          | 6 |

## Updates to GOLD Data Release or Documentation

| Version | Date       | Description   | Datasets | Documentation       |
|---------|------------|---|----------|---------------------|
| 1       | 11/30/2018 | 1 <sup>st</sup> data release  | _INV1    | V1.0<br>(Nov. 2018) |
| 2       | 2/1/2019   | <b>2<sup>nd</sup> data release. Updates:</b><br>- 56 participants included which were excluded in INV1 release because missing consent.<br>- GOLD sampling weights were updated to include all 3,489 participants.<br>- Variables SURVEY_DATE and S1_FREQ included in GOLD_PART_DERV. | _INV2    | V2.0<br>(Feb. 2019) |

### 1. DESIGN

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, ver. 4) and Analyses Methods for Visit 2 (March 2018, ver. 1.1) documentation that describes the proper use of sample weights in statistical analyses.

The sampling weights are calculated based on visit 2 non-response adjusted sampling weights and the nonresponse rates by gender and age groups (C6) for GOLD (Frozen as 1/22/2019). The sampling weights for GOLD 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).

#### 1.1. **WEIGHT\_EXPANDED\_GOLD** (GOLD Expanded Census 2010 Calibrated, Trimmed, Nonresponse Adjusted Weights)

The GOLD\_WEIGHT\_EXPANDED sample weight is calibrated, trimmed, non-response adjusted (by gender and age groups) reciprocal of a participant’s probability of selection into the HCHS/SOL GOLD Ancillary Study. It was calibrated using first the Hispanic background distribution and then the 12-category age group/gender distribution using the 2010 Decennial Census.

#### 1.2. **WEIGHT\_NORM\_CENTER\_GOLD** (GOLD Center Normalized, Census 2010 Calibrated, Trimmed, NonResponse Adjusted Weights)

This sampling weight is the expanded sampling weight normalized to GOLD center sample size (frozen as 1/22/2019). The sum of the weights is 3,941.

#### 1.3. **WEIGHT\_NORM\_OVERALL\_GOLD** (GOLD Overall Normalized, Census 2010 Calibrated, Trimmed, NonResponse Adjusted Weights)

This sampling weight is the expanded sampling weight normalized to GOLD overall sample size (frozen as 1/22/2019). The sum of the weights is 3,489.

## 2. ADMINISTRATIVE

### 2.1. **KIT\_ID** **(GOLD Lab ID)**

GOLD Lab ID based on GOL0c and GOL0c1.  
KIT\_ID was created by combining the string text of GOL0c and GOL0c1 variables.

Source variable(s):

GOL0c. Lab ID GOLD Site.  
GOL0c1. Lab ID number.

### 2.2. **BURK\_ID** **(Lab use BURK ID: G####)**

BURK\_ID is a Lab use ID from the month updated Burk Receiving file matched to the GOLD CDART 2 retrieved data by KIT\_ID.

Source variable(s):

BURK\_ID column from the Burk Receiving file.

### 2.3. **CLINDATE\_GOLD** **(GOLD AS Clinic Visit Date)**

GOLD AS Clinic Visit date based on GOL0a.

Source variable(s):

GOL0a. GOLD AS Enrollment and Tracking (GOL) form Completion Date.

### 2.4. **GOLD\_ENROLL** **(GOLD AS Enrolled Participants)**

This is 0/1 variable indicating valid GOLD AS Enrolled Participants (consented to Visit 2, agree to participate and Specimen Kit distributed).

GOLD\_ENROLL=1 if (Consent\_v2=1 and GOL1 =2 or 3 and GOL2 non-missing, and GIC1=1).

Source variable(s):

Consent\_V2. HCHS/SOL Visit 2 study consented.  
GOL1. GOLD AS Enrollment status: 1=Refused  
2=Agreed to participate during V2,  
3=Agreed to participate at contact other than V2.  
GOL2. Specimen kit Distribution date.  
GIC1. Consent to HCHS/SOL GOLD for data and specimen's collection.

**2.5. SAMLEYN  
(GOLD Specimen Sample kit Received by Lab)**

This is 0/1 variable indicating the GOLD AS participants Specimen Sample Kit have been received by the Lab, information from the month updated Burk Receiving file.

Source variable(s):

Specimen sample records based on the Burk Receiving file, matched to the GOLD CDART 2 retrieved data by KIT\_ID.

**2.6. REC\_DATE  
(Rec. Date, from Burk Receiving File)**

This is date variable indicating Rec. Date, from the "Rec. Date" column in Burk Receiving file.

**2.7. SURVEY\_DATE  
(Survey Date, from Burk Receiving File)**

This is date variable indicating Survey Date, from the "Survey Date" column in Burk Receiving file.

**2.8. S1\_FREQ  
(S.1 (Frequency) Response, from Burk Receiving File)**

This variable indicates Bristol Stool Frequencies, from the "S.1 (Frequency) Response" column in Burk Receiving file.

**2.9. S2\_TYPE  
(S.1 (Type) Response, from Burk Receiving File)**

This variable indicates Bristol Stool Characteristics, from the "S.2 (Type) Response" column in Burk Receiving file.

### 3. SOCIO-DEMOGRAPHIC

#### 3.1. AGE\_GOLD (Age in years at GOLD AS clinic visit)

The age of the participant in years at the time of GOLD AS clinic visit is determined from the participant's date of birth (from HCHS Visit 2) and the CASAS clinic visit date.

Source variable(s):

DOB\_V2. Date of Birth from HCHS Visit.

CLINDATE\_GOLD. GOLD AS Clinic Visit date based on GOL0a.

#### 3.2. AGEGROUP\_C6\_GOLD (6-level grouped age of participant)

This is a 6-level ordinal variable determined from the derived variable AGE\_GOLD.

| AGE (yrs) | AGEGROUP_C6_GOLD |
|-----------|------------------|
| 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\_GOLD. Age of participant at the time of GOLD clinic visit