NHLBI grant: Ancillary to HCHS/SOL: Sleep Patterns as risk factors for disease in the Hispanic Community Health Study (R01 HL098297, S Patel)

Sueno Ancillary Study
Sleep Patterns as a Risk Factor for Disease in the Hispanic Community Health Study

Investigator Use Database Overview
Version 2.1
July 2015

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

Daniela Sotres-Alvarez
Courtney Page

ACKNOWLEDGMENTS

Sueño Study would not have been possible without the participation of HCHS/SOL Study participants and the support of study member staff. Sueno Study was supported by grant **R01HL098297** from the National Heart, Lung, and Blood Institute (NHLBI).

Special thanks to:

Abreu, Maria de los Angeles, MPA, CCPR
UNC at Chapel Hill - HCHS/SOL CC, NC

Antony, Vidya
UNC at Chapel Hill - HCHS/SOL CC, NC

Balbin, Elizabeth, MA
University of Miami - Miami, FL

Bollinger, Nicole W, BA
Brigham and Women's Hospital

Chambers, Claudia P., BPh
University of Miami - Miami, FL

Espinoza, Angelica, BA
University of Illinois - Chicago, IL

Herbas, Marcos, BS
Northwestern University - Chicago, IL

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

Loredo, Jose S., MD, MS, MPH, FCCP
University of California – San Diego, SD

Luciano, Jose
Einstein College of Medicine - Bronx, NY

Martello-Jaurretche, Maria, MS
Einstein College of Medicine - Bronx, NY

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

Page, Courtney, BS
UNC at Chapel Hill - HCHS/SOL CC, NC

Patel, Sanjay R., MD
Brigham and Women's Hospital

Ramirez, Maricelle, BA
Brigham and Women's Hospital

Ramos, Alberto R., MD
University of Miami - Miami, FL

Reid, Kathryn J., PhD
Northwestern University - Chicago, IL

Rubio, Ileana
University of California – San Diego, SD

Rueschman, Michael, MPH
Brigham and Women’s Hospital

Seiger, Ashley
Brigham and Women's Hospital

Sotres-Alvarez, Daniela
UNC at Chapel Hill - HCHS/SOL CC, NC

Teng, Yanping, MD, MSPH
UNC at Chapel Hill - HCHS/SOL CC, NC

Velazco, Erika
University of California – San Diego, SD

Zee, Phyllis C., MD, PhD
Northwestern University - Chicago, IL

SUEÑO Sleep Ancillary Study Investigator Use Database
Version 2.1, July 2015

Updates To Sueño Data Release or Documentation

Version	Date	Description	Datasets	Documentation
1	3/25/2014	1st data release	_INV1	V1 (March 2014)
1.1	6/24/2014	- Sueño sampling weight - Background eligibility for Bronx change (15 Mexicans) - Correction for one participant which makes him ineligible (AHI>50 in HCHS)	PART_DERV_ SUENO_INV1	V1.1 (June 2014)
2.0	3/13/2015	2nd data release (ONLY to PI Sanjay Patel) - ID changed for ONE participant in ALL datasets as by accident at the site he/she was assigned an ID from another participant. - Sampling weights had to be recalculated for that ONE ID change. Changes are miniscule in that site. The rest are identical. - 15 data entry errors were corrected in APEA4 (measured weight in kg), and hence their derived variables were recalculated (e.g. BMI_sueno).	_INV2	V2.0 (March 2015)
2.1	**7/31/2015**	**2nd data release (to all Sueño PIs)** - Updates done in version 2.0 (3/13/2015) - Added medication derived variables to PART_DERV_SUENO_IU2 - Created BMI using SOL height - Created hypertension and diabetes using SUENO medications	**_INV2**	**V2.1 (July 2015)**

1. INTRODUCTION

This document describes the content and structure of the Investigator Use datasets created for the SUEÑO Ancillary Study. This database contains all the data collected for SUEÑO participants, 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 SUEÑO derived variable file.

2. STUDY OBJECTIVES

Sueño HCHS/SOL Ancillary Study was designed to assess the prevalence of poor sleep habits in Hispanic population living in the US, identify predictors of poor sleep patterns, and define the impact of poor sleep on health consequences including obesity, diabetes, hypertension, and cardiovascular disease by measuring objectively sleep for a week (actigraph device) in a subgroup of 2200 participants of the parent HCHS/SOL study.

3. STUDY DESIGN

Original Sueño Study protocol requested each of the four field centers to recruit approximately 550 HCHS/SOL participants aged 18-64 years and within 24 months form HCHS/SOL baseline examination from the Hispanic/Latino backgrounds of interest as depicted in Table 3.1. Each site needed to ensure a minimum goal of 330 participants per each Hispanic/Latino background in order to adequately power analyses of these groups. One year after recruitment started, the window of eligibility was increased to 30 months. The recruitment period was from October 2010 to December 2013 and recruited 2,252 participants across all four HCHS/SOL four field centers (Bronx, Chicago, Miami, and San Diego).

The eligibility criteria to participate in Sueno are:

- To be an HCHS/SOL main study cohort participant (n=16,415)
- Allow to be contacted about future ancillary studies
- Do not self-reported narcolepsy or being treated for obstructive sleep apnea at HCHS/SOL baseline
- Do not had an urgent AHI referral (AHI 3% desaturation >50) from HCHS/SOL baseline sleep monitor (SLPA16)
- Age 18 to 64 years at Sueño visit
- HCHS/SOL baseline exam less than 24 months (30 months) from Sueño visit

Table 3.1 Sueño Protocol Breakdown of Hispanic/Latino background by field center

	BRONX	CHICAGO	MIAMI	SAN DIEGO	TOTAL
Cuban	0	0	330	0	330
Mexican	0	110	0	550	660
Puerto Rican	220	220	0	0	440
Dominican	330	0	0	0	330
Central/South American	0	220	220	0	440
TOTAL	**550**	**550**	**550**	**550**	**2200**

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_SUENO_INV1) has been created with SUENO specific variables (e.g. BMI using weight and height from SUENO APE 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 **"SUENO Derived Variable Dictionary INV1".**

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

Table 4.1. Sueño datasets

Administrative	Form Code (SAS dataset)	Number of observations
Pre-visit screening (eligibility, safety)	ANE	5,010
Procedures		
Anthropometry	APE	2,252
Sleep Data (Actigraph device)	SAW	2,218
Questionnaires		
Acculturation Stress	ATE	2,251
Medication & Supplement use	MDE	2,250
Medical History	MQE	2,244
Neighborhood Stress	NSE	2,252
Sleep Questionnaire	SPE	2,251
Sleep Questionnaire II	SQE	2,250
Sleep Attitudes	SSE	2,251
Well being	WLE	2,252
Work Schedule	WSE	2,252
Derived Variables		
Participant Derived	PART_DERV_SUENO	2,252

4.2. Form and Data Set Naming Conventions

Each SUENO data collection instrument (PDF form) has a unique four-letter mnemonic associated with it. Corresponding data sets begin with the same four letters of the mnemonic, followed by the character string "_INV1" for Investigator Use Version 1. For example, the data set for "Pre-visit Screening-ANE" investigators-version1 is "ANEA_INV1". The naming convention serves both to identify the originating form and provide version control when subsequent generations of datasets are produced. 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 ANEA and the ANSA both map to the ANEA_ INV1 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 key field ID is a random 8-digit identification code unique to each HCHS/SOL participant and is the same in Sueno Study.

4.4. Common Variables Across Data Sets

Two variables appear in all datasets that are part of the DMS database but are not relevant for Sueno: LINENUMBER, and VISIT. Only FORM is useful.

VERSION: Version of the data collection form. 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. All SUENO forms have only 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 (ANEA vs. ANSA Screening Forms).

4.5. Variable Naming Conventions

While the key field (ID) have the same name on each SAS dataset, 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, question 3 (Has participant been diagnosed by physician with narcolepsy?) on the Screening form ANE is named ANEA3 on the corresponding SAS file ANEA_INV1.

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 SUENO) was re-derived for use in all data sets as a random identifier code for participants.

2) CENTER, is a real code to distinguish among participating field centers was created for the Investigator Use database and is included in the Participant derived variable set PART_DERV_SUENO_INV1.

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

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

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.

5. DESCRIPTION OF DATA COLLECTION FORMS AND ACTIGRAPHY DATA

5.1. Administrative

5.1.1. Telephone Screening Interview (ANE)

This 10 item instrument performs an initial screening that collects exclusion criteria information from participants included in the eligibility lists provided by the CC. Additional exclusions criteria reviewed with this form are: age, pregnancy, narcolepsy, and sleep apnea. The form is also used to inquire about special needs, such as any medical conditions that would affect the examination or the appointment time or impediments in hearing or reading.

5.2. Procedures

5.2.1. Anthropometry (APE)

The direct measurements obtained at the anthropometry station are recorded on these entry screens. The APE form also screens for presence of an electronic implantable device (EID), and the ability to stand.

3/2015 UPDATE at _INV2: 15 changes in APEA4 due to data cleaning (body weight incorrectly entered in lb instead of kg).

5.2.2. Actigraphy Data (SAW)

Participants were given a wrist actigraph device (Respironics Spectrum); this device records movement and is used to infer sleep and wake times. The participant was asked to wear the device for 7 consecutive days and nights. At the Field Center, actigraphy data was downloaded and transmitted to the Sleep Reading Center. Data released to

investigators ONLY include sleep studies with valid status (SAWA4=1), and the variable with the invalid reason (SAWA5) was removed.

5.3. Questionnaires

5.3.1. Acculturation Stress (ATE)

This 17 item instrument provides information on stress associated with the process of integrating and existing in a non-native culture.

5.3.2. Medication Use (MDE)

The medication use questionnaire captures the self-report medication use such as prescription, and over-the-counter medications, including cold and allergy medications, vitamins, herbals or supplements. Since participants may or may not know the actual indication for a specific medicine, there is embedded 10 item list of conditions for which medications could be prescribed.

5.3.3. Medical History (MQE)

This 15 item instrument provides information of doctor-diagnosed medical conditions such as hypertension, other cardiovascular diseases and sleep disorders.

5.3.4. Neighborhood Stress (NSE)

This 15 item instrument provides information on stress associated with the conditions in one's external living environment which includes factors such as neighborhood support, safety, cleanliness, noise level, and accessibility to resources.

5.3.5. Sleep Attitudes (SSE)

This 20 item instrument provides information on the participant perception of the importance of sleep and ideas about sleep hygiene.

5.3.6. Sleep Questionnaire (SPE)

This 18 item instrument provides information about sleep patterns and symptoms of sleep disturbances.

Please note that the Spanish translation for question 18 was updated in SUENO on 09-17-2011 to match the English translation; it does not match the Spanish translation in the HCHS main study Sleep Questionnaire (SLSA).

SUENO SPSA18 original translation and HCHS/SOL main study SLSA18 translation: ¿Son estos síntomas peores en el transcurso del día o durante la noche?

SUENO SPSA18 updated translation: ¿Son estos síntomas peores mas tarde en el día o durante la noche?

5.3.7. Sleep Questionnaire II (SQE)

This 25 item instrument provides information about the participant sleep patterns. It also evaluates circadian rhythms, insomnia, and sleep hygiene.

5.3.8. Well-being Questionnaire (WLE)

This 20 item instrument is intended to evaluate the relationships between depression and anxiety with sleep. Also, it provides an assessment of the prevalence of depression and anxiety symptoms among Latinos that have sleep problems.

5.3.9. Work Schedule (WSE)

This 17 item instrument provides information about the participants' current work schedule.

Please note that the Spanish translation for question 3 was updated on 01-12-2011 in SUENO to match the English translation

SUENO WSSA 3 original translation: En una semana típica, ¿Cuántas horas de trabajo dura su día normal de trabajo en su(s) empleo(s)?

SUENO WSSA 3 updated translation: En una semana típica, ¿Cuántas horas trabaja en su(s) empleo(s)?

5.4. Derived Variables (PART_DERV_SUENO)

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 (2,252 observations) at baseline PART_DERV_SUENO_INV1 including 63 who are not eligible due to being 65 or older, interviewed more than 30 months from HCHS/SOL baseline, have AHI>50 at baseline, or not eligible Hispanic/Latino Background for the site). Indicator variable SUENO_ELIGIBLE flags those 2,189 who meet Sueño eligibility criteria.** 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, "SUEÑO Derived Variable Dictionary" for the definitions of the variables included in this special purpose file.

The participant derived file includes several demographic variables from the HCHS/SOL main study, and all the remaining variables are derived from SUENO data. In some cases the variable name ends with _SUENO to explicitly state that this variable is derived from Sueno data. For example variable BMI_SUENO was derived from height and weight measured at Sueno visit (SUENO APEA form).

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 SUEÑO Coordinating Center and their resolutions are included in the derived variable files.