

Health and Retirement Study

Health Care and Nutrition Study (HCNS) 2013

Version 5.0, November 2018

Data Description

2013 Health Care and Nutrition Study V5.0

Nutrient Totals Data Description

1. Introduction

This nutrient totals dataset is a supplement to the 2013 Health Care and Nutrition Study (HCNS), which contained questions about health care access, food purchases, food consumption and nutrition (including vitamins and other supplements). This dataset uses the responses from Section C – Food and Nutrition, to calculate calorie and nutrient totals for each respondent across 215 nutrient variables. The HRS HCNS was based on the Harvard food frequency questionnaire originally proposed by Willett and colleagues, and these estimates of nutrient intake utilize the nutrient tables provided by the Harvard School of Public Health (<https://regepi.bwh.harvard.edu/health/nutrition.html>).

For more information about the 2013 Health Care and Nutrition Study (HCNS), see the HRS website, documentation, data description and “off-year” studies: <http://hrsonline.isr.umich.edu>.

The HRS is funded under a cooperative agreement between the National Institute on Aging (NIA) and the Survey Research Center at the University of Michigan. The HRS is designed to study labor force, health, and family transitions of the U.S. population aged 51 and older, and the impact of those transitions on economic resources, claims on structured programs such as Social Security, Medicare, and Medicaid, and informal assistance and transfers to and from family members.

The National Institute on Aging (NIA) provided funding (U01 AG009740) for the 2013 HCNS, which was conducted by the Survey Research Center (SRC), at the Institute for Social Research (ISR), at the University of Michigan.

By receiving the dataset, you agree to use it for research and statistical purposes only, and make no effort to identify respondents. In addition, you agree to send us a copy of publications you produce based on the data. (See Obtaining the Data at the end of this document for additional details).

The data release history for HCNS is:

Release Number	HCNS Data	HCNS Nutrient Totals Data
V1.0	First data release, weights not included	
V2.0	Weights added	
V3.0	No change	First release of nutrient totals
V4.0	No change	Correction to RACE variable
V5.0	No change	Correction to the nutrient totals associated with C9AG Salt Added

1.5. Correction to Nutrients Associated with C9AG, HCNS V5.0

An error was discovered in the nutrient calculations associated with C9AG, Salt added. The nutrients associated with C9AG Salt added, and the associated nutrient totals affected by this update are:

Nutrient	Nutrient total
Calcium	CALC_SUM
Iron	IRON_SUM
Magnesium	MAGN_SUM
Potassium	K_SUM
Sodium	SODIUM_SUM
Zinc	ZN_SUM
Copper	CU_SUM
Manganese	MN_SUM

2. The Sample Interviewed in the 2013 HRS HCNS

In November 2013, questionnaires were mailed to a subsample of HRS respondents (n= 12,418). The sample for the 2013 HCNS consists of all living HRS respondents, and their spouse/partners, who were not included in the 2013 Consumption and Activities Mail Survey (CAMS). The field period for the 2013 HCNS was late November 2013 through early May 2014.

The main data file for the HCNS, released in August 2014, contains data for 8,073 respondents, with a simple response rate of 65% percent.

For the nutrient totals dataset, we removed 37 respondents who answered less than 3% of the food consumption questions in Section C, and 1 respondent reported as using a feeding tube. The data file for the nutrient totals contains data for 8,035 respondents.

For Section C, 97% of respondents answered 90% or more of the food consumption questions. Missing data was imputed using the steps described below (4-2).

3. The HCNS Nutrient Totals Data

Data in the 2013 HCNS Nutrient Totals release is divided into the following sections:

Section A	Identifiers and Demographics
Section B	Nutrient Totals
Section C	Food Frequencies

Section A contains respondent identifiers and the five demographic variables from the core HRS used to impute missing data. Section B contains the totals for each nutrient variable for each respondent. Section C contains the daily frequencies for each food, including imputed data.

4. Nutrient Totals: Calculations

The source files used for deriving the nutrient totals were obtained from the Harvard University School of Public Health's download site:

<https://regepi.bwh.harvard.edu/health/nutrition.html>

The totals were calculated as follows:

1. Convert categorical responses to numeric responses (daily portion sizes).
2. Impute missing data for 6 food items with the least amount of missing data, 1 from each food category, using 5 respondent descriptors pulled from the core Health and Retirement Study (age, race, gender, years of education, and BMI).
3. Impute the remaining missing data for consumption frequency in Section C using these 11 predictors (5 respondent predictors and 6 food items).
4. Impute missing data for the 5 categorical variables describing fat content for particular food items (yogurt, cheese, margarine, salad dressing).
5. Assign brand/type of cereal and margarine for missing open ended questions.
6. Map foods mentioned in *C16 Other foods eaten at least once a week* to the nutrient data set.
7. Calculate the totals for each nutrient.

4-1. Convert categorical responses to numeric responses (daily portion sizes).

The data in Section C was first converted from a categorical response to a numeric value reflecting servings per day using Harvard University's food serving conversion guides. For example, 1 serving per week is equivalent to 0.14 servings per day (1/7), and 5-6 servings per week is equivalent to 0.8 servings per day (5.5 / 7).

4-2. Impute missing data for 6 food items with the least amount of missing data, 1 from each food category, using 5 respondent descriptors pulled from the core HRS (age, race, gender, years of education, BMI).

For the first level of imputation, we chose a variable with a low percentage of missing data (less than 2%) from each of the 6 food groups:

- C3N cheese (American, cheddar)
- C4D bananas
- C5G broccoli
- C6L hamburger, lean
- C7K white rice
- C8C soda with caffeine and sugar

Missing data for these six food variables was imputed by regressing each food on 5 predictor variables from the core HRS study.

- AGE
- GENDER
- RACE
- YEARS OF EDUCATION
- BMI (body mass index, calculated using the formula $703 * (\text{height} / \text{weight}^2)$)

We used that equation to predict values for all respondents in the dataset, including those with missing data. We then sorted by the predicted value, and each case with missing data was filled in using the observed value of the case nearest in predicted value.

(Note: There are 15 respondents in the HCNS data who do not have a core HRS interview. For those respondents, missing data in the 5 predictor variables were filled in using information from the spouse/partner's core interview, or with mean values.)

4-3. Impute the remaining missing data for consumption frequency in Section C using the 11 predictors described above (5 respondent predictors and 6 food items).

The 5 predictor variables from the core study along with the 6 predictor foods (including the imputed values), were then used to impute the remaining continuous food variables in Section C.

Missing data for a given food item in this section ranged from 0.8% to 16%, with an average of 2%.

For C9AH, an open ended question which asked how many teaspoons of sugar the respondent added to beverages each day, the top 1% of values were set to missing (because they were extremely high) and were imputed.

4-4. Impute missing data for the 5 categorical variables describing fat content for particular food items (yogurt, cheese, margarine, salad dressing).

Missing data for these 5 categorical variables ranged from 4 % to 16%.

HNC3K TYPE YOGURT was imputed when either of the yogurt questions HNC3I or HNC3J contained a value greater than 0, including imputed values. About 5.6 % of the responses were imputed for HNC3K TYPE YOGURT.

HNC3O TYPE CHEESE was imputed when any of the cheese questions HNC3L, HNC3M, or HNC3N contained a value greater than 0, including imputed values. About 4% of the responses were imputed for HNC3O TYPE CHEESE.

HNC3S FORM MARGARINE and HNC3T TYPE MARGARINE were imputed when either of HNC3Q SPREADABLE BUTTER or HNC3R MARGARINE contained a value greater than 0, including imputed values. About 12% of the values for HNC3S FORM MARGARINE, and about 16% of the values for HNC3T TYPE MARGARINE were imputed.

HNC9AO TYPE SALAD DRESSING was imputed when HNC9AN SALAD DRESSING contained a value greater than 0, including imputed values. About 9% of the values for HNC9AO were imputed.

To impute these five categorical variables, we did an ordered logistic regression, using the 5 predictor variables from the core study (age, gender, race, years of education and BMI), along with related predictor variables.

Variable	Predictors (in addition to 5 predictors from core study)
C3K type yogurt (regular, low fat, nonfat)	C3I sweetened yogurt C3J low carb yogurt C3A skim milk C3B 2% milk C3C whole milk
C3O type cheese (regular, low fat, nonfat)	C3L cottage cheese C3M cream cheese C3L Amer/cheddar C3A skim milk C3B 2% milk C3C whole milk
C3S form margarine (stick, tub, spray/squeeze) C3T type margarine (regular, light, nonfat)	C3Q spreadable butter C3R margarine C3A skim milk C3B 2% milk C3C whole milk
C9AO type salad dressing (nonfat, low fat, olive oil dressing, other vegetable oil dressing)	C9AL low fat mayonnaise C9AM regular mayonnaise C9AN salad dressing C9AP olive oil

The data were then sorted by the probability of NONFAT, and missing values were filled in by taking the observed value from the record nearest to the missing value.

4-5. Assign brand/type of cereal and margarine for missing open ended questions.

The open ended questions HNC3U1 and HNC3U2 asked respondents to specify the brand and type of margarine or spreadable butter they consumed most often. This information was necessary to link to the nutrient dataset for margarine.

There were 6027 respondents with an actual or imputed non 0 value in one or both of the margarine/spreadable butter questions (HNC3Q or HNC3R). Of these, 3429 (57%) had missing or incomplete data in HNC3U1 (brand), and 5569 (92%) had missing or incomplete data in HNC3U2 (type).

In the cases with missing or incomplete brand data, the link was made to the nutrient file for margarine (or oil) using the data in HNC3S and HNC3T (form and type). In cases where the form and type mapped to more than one possible brand, the most representative brand was chosen.

For example, there are 4 different brands of light stick margarine in the margarine file (Blue Bonnet, Brummel & Brown, I Can't Believe It's Not Butter, and Imperial). Respondents with light stick values in HNC3S and HNC3T who are missing a brand in HNC3U1, are linked to Blue Bonnet.

For HNC7A Cold breakfast cereal, there were 6788 respondents (of 8035) in the dataset who provided an answer greater than 0 (Never), or for whom a response greater than 0 was imputed.

Of these, 5243 respondents did not answer the follow up open ended question HNC7D, "What brand and type of cold breakfast cereal do you usually eat?" These respondents were all assigned General Mills Cheerios for the purpose of computing nutrient totals.

4-6. Map foods mentioned in "C16 Other foods eaten at least once a week" to the nutrient data set.

In this step, we added 63 additional foods mentioned in the open ended question C16, which were not included in the questionnaire. (Foods mentioned in C16 which were already included in the questionnaire, were only included in nutrient totals if the respondent had skipped that item in the main questionnaire. This was done prior to imputing missing data.)

4-7. Calculate the totals for each nutrient

Once all missing data was imputed, the totals for each variable in the nutrient data set were calculated. Nutrient data for margarine and cereal was pulled from MARGARINE.XLSX, CEREAL.XLSX, and OIL.XLSX. The remaining nutrient data was pulled from FOOD.2011.XLSX. All of these files were obtained from the Harvard University School of Public Health download site.

Foods in the questionnaire that did not map directly to the FOOD.2011.XLSX data set were mapped using averages of similar foods. For example, HNC3B asks about 1% *or* 2% milk, which are separate lines in the nutrient data set. Responses to that question are mapped to a line averaging those 2 items.

The nutrient values provided in FOOD.2011 are per 100 grams of each food and beverage, so prior to calculating the totals, we scaled the nutrient data to the serving sizes described in the questionnaire.

Some of these gram conversions were obtained from Harvard University. The remaining conversions were obtained from:

<http://nutritiondata.self.com/>

5. Distribution Files

The following extensions are used for the six different types of distribution files:

- .DA for data files,
- .SAS for SAS program statements,
- .SPS for SPSS program statements,
- .DO for Stata DO statements,
- .DCT for Stata dictionary statements, and
- .TXT for codebook files.

The file naming conventions mirror that of the original HCNS data release, with _NT added (for “nutrient totals”). For example,

HCNS13_R_NT_V4.DA contains ASCII data
HCNS13_R_NT_V4.SAS contains corresponding SAS program statements,
HCNS13_R_NT_V4.SPS contains corresponding SPSS program statements,
HCNS13_R_NT_V4.DO contains corresponding Stata DO statements,
HCNS13_R_NT_V4.DCT contains corresponding Stata dictionary statements, and
HCNS13_R_NT_V4.TXT contains the ASCII codebook.

The 2013 HRS HCNS Nutrient Totals Final Release data are provided in ASCII format, with fixed-length records. Use the associated SAS, SPSS or Stata program statements to read the data into the analysis package of your choice. In addition, you will probably want to download the codebook file (HCNS13_R_NT.TXT) and the data description (this document).

6. Program Statements

6A. Using the Files with SAS

To create a SAS system file for a particular dataset, two file types must be present for that dataset -- .SAS program statement files and .DA data files.

To create a SAS system file, load the *.SAS file into the SAS Program Editor.

If the *.SAS file is located in "c:\hcns2013\sas_nt_v4" and the data file is located in "c:\hcns2013\data_nt_v4", you can run the file as is. A SAS system file (*.SD2 or *.SAS7BDAT) will be saved to directory "c:\hcns2013\sas_nt_v4".

If the files are not located in the specified directories, you will need to edit the *.SAS file to reflect the proper path names prior to running the file.

6B. Using the Files with SPSS

To create an SPSS system file for a particular dataset, two file types must be present for that dataset -- .SPS program statement files and .DA data files.

To create an SPSS system file, open the *.SPS file in SPSS as an SPSS Syntax File.

If the *.SPS file is located in "c:\hcns2013\spss_nt_v4" and the data file is located in "c:\hcns2013\data_nt_v4", you can run the file as is. An SPSS system file (*.SAV) will be saved to directory "c:\hcns2013\spss_nt_v4".

If the files are not located in the specified directories, you will need to edit the *.SPS file to reflect the proper path names prior to running the file.

6C. Using the Files with Stata

To use Stata with a particular dataset, the following three file types must be present for that dataset -- .DCT files, .DO files, and .DA data files.

Files with the suffix .DA contain the raw data for Stata to read. Files with the suffix .DCT are Stata dictionaries used by Stata to describe the data. Files with the suffix .DO are short Stata programs ("do files") which you may use to read in the data. Load the .DO file into Stata and then submit it.

If the *.DO and *.DCT files are located in "c:\hcns2013\stata_nt_v4" and the data file is located in "c:\hcns2013\data_nt_v4", you can run the .DO file as is.

If the files are not located in these directories, you must edit the *.DO and *.DCT files to reflect the proper path names before you run the files.

Note that the variable names provided in the .DCT files are uppercase. If you prefer lower case variable names, you may wish to convert the .DCT files to lower case prior to use. You may do this by reading the .DCT file into a text or word processing program and changing the case. For instance in Microsoft Word, Edit, Select All, Format, Change Case, lowercase.

7. Linking Respondents across Time

Respondent records in the 2013 HCNS Nutrient Totals Final Release can be linked to respondent records HHID and PN. The sub-household identifiers can be used to link household data with the cross-sectional respondent level data. By wave, these are:

<u>Wave</u>	<u>Subhousehold ID</u>
1992	ASUBHH
1993	BSUBHH
1994	CSUBHH
1995	DSUBHH
1996	ESUBHH
1998	FSUBHH
2000	GSUBHH,
2002	HSUBHH
2004	JSUBHH
2006	KSUBHH
2008	LSUBHH
2010	MSUBHH
2012	NSUBHH
2014	OSUBHH
2016	PSUBHH

8. Registration and Downloading the Data

8A. Registration

HRS data are available for free to researchers and analysts at the HRS Web site. In order to obtain public release data, you must first register at our Web site. Once you have completed the registration process, your username and password will be sent to you via e-mail. Your username and password are required to download any data files.

By registering all users, we are able to document for our sponsors the size and diversity of our user community allowing us to continue to collect these important data. Registered users receive user support, information related to errors in the data, future releases, workshops, and publication lists. The information you provide will not be used for any commercial use, and will not be redistributed to third parties.

8B. Conditions of Use

By registering, you agree to the Conditions of Use governing access to the Health and Retirement Study's public release data. You must agree to

- o not attempt to identify respondents
- o not transfer data to third parties except as specified
- o not share your username and password
- o include specified citations in work based on HRS data
- o provide information to us about publications based on HRS data
- o report apparent errors in the HRS data or documentation files
- o notify us of changes in your contact information

For more information concerning privacy issues and conditions of use, please read "Conditions of Use for Public Data Files" and "Privacy and Security Notice" at the Public File Download Area of the HRS Web site.

8C. Publications Based on Data

As part of the data registration process, you agree to include specified citations and to inform HRS of any papers, publications, or presentations based on HRS data. Please send a copy of any publications you produce based on HRS data, with a bibliographical reference, if appropriate, to the address below.

Health and Retirement Study
Attn: Papers and Publications
The Institute for Social Research, Room 3410
P.O. Box 1248
Ann Arbor, MI (USA) 48106-1248

Alternately, you may contact us by e-mail at hqsquestions@umich.edu with "Attn: Papers and Publications" in the subject line.

9. If You Need to Know More

This document is intended to serve as a brief overview and to provide guidelines to using the 2013 HCNS Final Release Nutrient Totals (Version 4.0) data. Additional information about the HRS can be obtained from the HRS Web site. If you have questions or concerns that are not adequately covered here or on our Web site, please contact us. We will do our best to provide answers.

9A. HRS Internet Site

Health and Retirement Study public release data and additional information about the study are available on the Internet. To access the data and other relevant information, point your Web browser to the HRS Web site.

<http://hrsonline.isr.umich.edu/>

9B. Contact Information

If you need to contact us, you may do so by one of the methods listed below.

Internet: Help Desk at our Web site

E-mail: hqsquestions@umich.edu

Postal service:

Health and Retirement Study
The Institute for Social Research, Room 3050
The University of Michigan
P.O. Box 1248
Ann Arbor, MI 48106-1248

FAX: (734) 647-1186