**ReadMe file for NHANES 2001-2002 and MPED, version 1.0**

**HEI-2005 Population Ratio Method**

This ReadMe file describes the methodology implemented in the provided SAS program, HEI2005\_NHANES0102\_MC\_PopulationScore.sas. The principles underlying this approach are described in Freedman et al., J Nutr, 2008, 138(9):1725-9 ([see abstract in PubMed](http://www.ncbi.nlm.nih.gov/pubmed/18716176)). The SAS program, along with the required macros, calculate mean Healthy Eating Index-2005 (HEI-2005) component and total scores and corresponding standard errors and confidence intervals for a population, subpopulation, or group, from one-day 24-hour recall data collected in NHANES 2001-2002. The specific instructions on how to successfully run the HEI procedure are within the provided SAS program. Because NHANES dietary data are weighted to equalize the days of the week and are collected year-round, these mean scores can be considered estimated scores of the population’s long-term or “usual” intake.

The program can be adapted to calculate HEI-2005 scores for other cycles of NHANES or for other data sources that use a complex sampling strategy.

**NOTE:** Another version of this SAS program is also available for non-survey datasets. It uses the analysis of data from the Automated Self-Administered 24-hour Recall (ASA24) system as an example. Also, additional code that calculates HEI-2005 component and total scores at the individual level, for each day or multiple days of 24-hour recalls or food records are available on the [NCI Healthy Eating Index website](https://epi.grants.cancer.gov/hei/sas-code.html).

This code is an update to that previously available for calculating HEI-2005 component and total scores and standard errors. In contrast to the previous code, this version does not require the use of SUDAAN; rather, SAS survey procedures are used to account for the survey design. This version also does not involve the use of an Excel spreadsheet to calculate confidence intervals for HEI-2005 components and total scores as the previous version did.

Because the HEI-2005 is a multi-dimensional construct involving 12 densities (amounts of food groups or nutrients per 1,000 calories), a simple method for estimating standard errors is not available. In this code, a Monte Carlo simulation step is included for the calculation of standard errors; this step simulates the densities for 10,000 samples to allow stable estimation of the standard errors.

The SAS program includes four main sections:

1. Calculations at the individual participant level to obtain variables needed to calculate HEI-2005 scores.
2. Calculation of weighted means and a variance-covariance matrix and generation of a Monte Carlo dataset, enabling standard errors to be calculated.
3. Allocation of legumes (beans and peas) and application of the HEI-2005 scoring algorithm.
4. Calculation of HEI-2005 component and total scores and their standard errors and confidence intervals.

***Required datasets:***

* [NHANES 01-02 dietary intake and demographics data](https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/wweia-documentation-and-data-sets/)
* [MyPyramid Equivalents Database (MPED), version 1.0](https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/mypyramid-equivalents-product-downloads/)
* CNPP MyPyramid Equivalents Databases for Whole Fruit and Fruit Juice for NHANES 2001-02 – ([cnppmypyrequivdb\_v1\_wjfrt.zip](https://origin.www.cnpp.usda.gov/HealthyEatingIndex-2005report.htm))
* CNPP Addendum to the MyPyramid Equivalents Database (MPED), version 2.0B - included in this folder

***Required macros:****available on* *[NCI Healthy Eating Index website](https://epi.grants.cancer.gov/hei/sas-code.html).*

* [hei2005.beanspeas.allocation.macro.sas](https://epi.grants.cancer.gov/hei/sas-code.html)
* [hei2005.score.macro.sas](https://epi.grants.cancer.gov/hei/sas-code.html)

The 12 components of the HEI-2005 calculated by the macros are:

* Total Fruit;
* Whole Fruit;
* Total Vegetables;
* Dark Green and Orange Vegetables and Legumes;
* Total Grains;
* Whole Grains;
* Milk;
* Meat and Beans;
* Oils;
* Saturated Fat;
* Sodium;
* Calories from Solid Fats, Alcoholic beverages, and Added Sugars (SoFAAS).

Some of these components come directly from the MyPyramid Equivalents Database (MPED) and NHANES 2001-02, but others are created as part of the SAS program.

The MPED and NHANES data provide the following components used directly in the calculation of HEI-2005: Total Fruit; Total Grains; Whole Grains; Oils; Saturated Fat; and Sodium.

*Additional steps are necessary to create the remaining components that are a combination of variables:* *Total Vegetables; Dark Green and Orange Vegetables and Legumes; Whole Fruit; Dairy; Meat and Beans; and Calories from SoFAAS.*

**This program carries out 9 steps:**

1. **Reads in and combines the required datasets and variables, makes necessary exclusions, and edits to the datasets.**
   1. Reads in the MPED, which gives MyPyramid equivalents per 100 grams of food.

The list below includes the variables from MPED used in the calculation of HEI-2005 components.

Total Fruit (F\_TOTAL) in cup equivalents

Total Vegetables (V\_TOTAL) in cup equivalents

Dark Green Vegetables (V\_DRKGR) in cup equivalents

Orange Vegetables (V\_ORANGE) in cup equivalents

Legumes (LEGUMES) in cup equivalents

Total Grains (G\_TOTAL) in ounce equivalents

Whole Grains (G\_WHL) in ounce equivalents

Total Milk (D\_TOTAL) in cup equivalents

Meat, Poultry, Fish (M\_MPF) in ounce equivalents

Eggs (M\_EGG) in ounce equivalents

Nut and Seeds (M\_NUTSD) in ounce equivalents

Soybean products (M\_SOY) in ounce equivalents

Discretionary oil (DISCFAT\_OIL) in grams

Discretionary solid fat (DISCFAT\_SOL) in grams

Added sugars (ADD\_SUG) in teaspoon equivalents

MPED for USDA Survey Food Codes, Version 1.0 (MyPyrEquivDB\_v1) can be extracted from the downloadable MyPyramid equivalents food data file, [Myfddata.exe](https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/mypyramid-equivalents-product-downloads/), on the Agricultural Research Service web site. All foods in the database are uniquely coded with an 8-digit USDA food code. This database contains the number of equivalents per 100 grams of food for the following food groups, which are used to calculate component and total scores for the HEI-2005.

MPED, version 2.0B database contains 942 food codes that were reported in NHANES 2005-06 and 2007-08 data cycles for individuals age 2 years and older (excluding breastfed children) who had complete recalls. The data for two food codes found in the Addendum should be used in place of those found in the MPED Version 1.0. These food codes are 11710800 and 11710801. This addendum is included in this folder.

* 1. CNPP MPED for Whole Fruit and Fruit Juice for NHANES 2001-2002 (cnppmypyrequivdb\_v1\_wjfrt.zip):

Calculation note for Whole Fruit:This database was created by CNPP from the MPED for USDA Survey Food Codes, Version 1.0. A SAS- formatted database can be downloaded from the CNPP web site [cnppmypyrequivdb\_v1\_wjfrt.zip](https://origin.www.cnpp.usda.gov/HealthyEatingIndex-2005report.htm). Foods containing fruit were assigned to either Whole Fruit (WHOLEFRT) or Fruit Juice (FRTJUICE) in cup equivalents per 100 grams of foods. In the case of foods that contain both whole fruit and fruit juice, e.g., fruit canned in fruit juice, the Total Fruit equivalents were assigned to either Whole Fruit or Fruit Juice, whichever was the majority ingredient according to its description or recipe in the Food and Nutrient Database for Dietary Studies (FNDDS), version 1.0, developed by the Agricultural Research Service. MPED, version 1.0, and CNPP MPED for Whole Fruit and Fruit Juice for NHNAES 2001-02 are first combined and further merged with NHANES 01-02 individual food intake data to calculate individuals’ 1-day food-group intake, expressed in MyPyramid equivalents (i.e., cup-, ounce-, and teaspoon-equivalents).

* 1. Reads in the NHANES 2001-02 Total Nutrient Intakes, First Day dataset (NUTRIENT).

NHANES 2001-02 datasets for food intake (drxiff\_b), nutrient intake (drxtot\_b) provide the following variables:

Dietary day one sample weight (WTDRD1)

Foodcode (DRDIFDCD)

Energy (DRXTKCAL) in kcal

Saturated fat (DRXTSFAT) in grams

Sodium (DRDTSODI) in grams

Alcohol (DRXIALCO) in grams

Carbohydrates (DRXICARB) in grams

NHANES 2001-02 datasets for food item intakes (drxiff\_b), total daily nutrients intake (drxtot\_b), and the demographic data (demo\_b) can be downloaded from [The Centers for Disease Control and Prevention, National Center for Health Statistics website](http://wwwn.cdc.gov/nchs/nhanes/search/nhanes01_02.aspx). The NHANES 01-02 food intake dataset, drxiff\_b, includes the list of foods eaten by each individual on Day 1 in gram amounts (DRXIGRMS). The NHANES 01-02 nutrient intake dataset, drxtot\_b, includes each individual’s Day 1 dietary sample weight (WTDRD1) and total nutrient intakes for Day 1.

**NOTE**: This step imports data only from those participants with reliable dietary recalls; this restricts the dataset to those with reliable dietary recalls. To read more about how NHANES determines if a dietary recall is reliable, see [The Centers for Disease Prevention and Control’s - DR1DRSTZ - Dietary recall status information](http://wwwn.cdc.gov/nchs/nhanes/2009-2010/DR1TOT_F.htm#DR1DRSTZ).

* 1. Reads in the NHANES 2001-02 Demographic dataset (DEMO) and keep only variables needed for analysis. This example keeps the following variables.
* SEQN: Respondent sequence number
* RIDAGEYR: Age in years of the participant at the time of screening.
* RIAGENDR: Gender of the participant.
* SDDSRVYR: Data release cycle
* SDMVPSU: Masked variance unit pseudo-PSU variable for variance estimation
* SDMVSTRA: Masked variance unit pseudo-stratum variable for variance estimation

**NOTE:** The variables kept here are one example; researchers should consider which variables they will need in their particular analysis.

**NOTE:**  In this step only participants ages 2 and older are included in the analysis. Keeping this inclusion criteria is recommended as the HEI-2005 is based on the 2005 Dietary Guidelines for Americans, which are intended as guidelines for the US population ages 2 and over.

1. **Moves soy beverages out of Soybean Products (M\_SOY) and into Dairy (D\_TOTAL). This uses the MyPyrEquivDB\_v1 output file. The adjusted totals are generated for each day of dietary intake and saved in a temporary SAS dataset.**

*Calculation note for soy beverages (Milk and Meat and Beans components):* Soy beverages are counted as part of the Milk component of the HEI-2005. This differs from the MPED, which groups them with other soy products in the Meats and Beans group. The reassignment process is completed in this step. Soy beverages (food codes 11310000, 11320000, 11321000, and 11330000) are converted from soybean products (M\_SOY) in ounce equivalents to milk (D\_TOTAL) in cup equivalents, based on the weight in grams of 1 cup.

1. **Combines the required datasets, make the necessary exclusions, and make necessary adjustments at the level of individual foods.**

MPED, version 1.0, and CNPP MPED for Whole Fruit and Fruit Juice for NHNAES 2001-02 are first combined and further merged with NHANES 01-02 individual food intake data to calculate individuals’ 1-day food-group intake, expressed in MyPyramid equivalents (i.e., cup-, ounce-, and teaspoon-equivalents).

1. **Creates additional required variables:** **ALLMEAT, V\_DOL, ALC, CARB, SoFAAS.**

Calculation note for ALLMEAT:ALLMEAT sums together all animal and plant proteins, including meat, poultry, fish, eggs, nuts, seeds, and the soy variable (ALLMEAT = M\_MPF (oz) + M\_EGG (oz) + M\_NUTSD (oz) + M\_SOY (oz)).

Calculation note for V\_DOL:V\_DOL sums together Dark Green and Orange Vegetables (V\_DOL = V\_DPYEL (cups) + V\_DRKGR (cups)).

Calculation note for Calories from SoFAAS (EXFAAS): EXFAAS sumps up three variables, ADDSUGC, SOLFATC, and BWCARBC described below, to calculate SoFAAS. The SoFAAS component of the HEI is calculated as a percentage of calories, so all variables that comprise empty calories need to be converted to units of calories.

ADDSUGC: Teaspoons of added sugar (ADD\_ SUG) are converted to calories using the conversion factor 1tsp=16kcal.

SOLDFATC: Grams of solid fat (DISCFAT\_SOL) are converted to calories using the conversion factor 1g=9kcal.

EXALCCAL: Alcoholic beverage intake is defined as beer, wine, or distilled spirits consumed as a beverage. To calculate the calories from alcoholic beverages, beer, wine, and distilled spirits reported or coded separately are used. Cooking wine and alcoholic beverages otherwise used as ingredients are excluded. The first 3 digits of the 8-digit USDA food code are used to indicate alcoholic beverages (931 through 935), and the food code for cooking wine (93401300). Calories from alcoholic beverages are calculated from the amounts of ethanol and carbohydrate contained in the beverages. To prevent double-counting calories from added sugars, any calories from added sugars are subtracted from the calories from alcoholic beverages because they are accounted for in the Added Sugars part of the Calories from SoFAAS component.

*Note for legumes (beans and peas):* In a previous version of the code for calculating HEI-2005 population scores, legume allocation occurred as part of this step. In this updated approach, allocation of legumes now occurs as part of step 6, after the Monte Carlo simulation step. This is intended to better reflect usual intake of legumes.

1. **Calculates NHANES 2001-02 individual food and nutrient intakes for each individual.**

In this step, computations are performed to create HEI-2005 components that are a combination of food groups from MPED.

**NOTE:** Legumes are allocated to the Meat and Beans component and/or to the Total Vegetables and the Dark Green and Orange Vegetables components in step 7.

# Calculates weighted means and a variance-covariance matrix and generates a Monte Carlo dataset, enabling standard errors to be calculated.

This section of the code performs computations necessary for the calculation of standard errors, accounting for the complex sampling design of NHANES 01-02. SAS survey procedures are used.

1. Weighted means and the variance/covariance matrix for the dietary variables of interest are calculated.

ii. A Monte Carlo data set with 10,001 rows is generated using the weighted means and variance/covariance matrix from step i.

# Runs the macro to properly allocate the intakes of LEGUMES in cup equivalents to either Total Protein Foods (ALLMEAT) or Total Vegetables and Dark Green and Orange Vegetables.

This section of the program calls the macro, [hei2005.beanspeas.allocation.macro.sas](https://epi.grants.cancer.gov/hei/sas-code.html).

This step results in 3 additional variables, named LEGUME\_ADDED\_V\_TOTAL, LEGUME\_ADDED\_V\_DOL, and LEGUME\_ADDED\_ALLMEAT are all used in step 8.

Calculation notes for Legumes: Intake of legumes counts toward meeting the standard for the Meat and Beans component first. Once the Meat and Beans standard is met, any additional amount of legumes counts toward the Total Vegetables and the Dark Green and Orange Vegetables and Legumes components.

**NOTE:** Legume amounts in the MPED are in cup equivalents; therefore, the cup equivalents are first converted to ounce equivalents of meat when they are counted for the Meat and Beans component, and are then converted back to cup equivalents when counted as vegetables.

Calculation note regarding conversion of cup equivalents to ounce equivalents One-fourth cup of legumes is equal to 1-ounce equivalent of meat. Thus, the number of cup equivalents of legumes is multiplied by 4 to convert to ounce equivalents of meat.

# Runs the HEI-2005 scoring macro which calculates intake density amounts and calculates HEI-2005 component and total scores and their standard errors and confidence intervals.

This section of the program calls the macro, [hei2005.score.macro.sas](https://epi.grants.cancer.gov/hei/sas-code.html).

1. This step uses the Monte Carlo dataset with the beans and peas allocated from step 6 and calls the HEI-2005 scoring macro which calculates intake density amounts and HEI scores.
2. Univariate and means procedures are used to compute one HEI-2005 total score and one set of HEI-2005 component scores and their standard errors and confidence intervals for the group, subgroup, or population.

Below are the HEI-2005 Component Scoring Standards. For more information on HEI components, see [Comparing Versions of the HEI](https://epi.grants.cancer.gov/hei/comparing.html) on the NCI website.

**HEI–2005**[**1**](http://epi.grants.cancer.gov/hei/developing.html#f1) **Components & Scoring Standards**

|  |  |  |  |
| --- | --- | --- | --- |
| **Component** | **Max points** | **Standard for maximum score** | **Standard for minimum score of zero** |
| **Adequacy:** | | | |
| **Total Fruit** | 5 | ≥0.8 cup equiv., per 1,000 kcal | No fruit |
| **Whole Fruit** | 5 | ≥0.4 cup equiv., per 1,000 kcal | No whole fruit |
| **Total Vegetables** | 5 | ≥1.1 cup equiv., per 1,000 kcal | No vegetables |
| **Dark Green, Orange Vegetables and Legumes2** | 5 | ≥0.4 cup equiv., per 1,000 kcal | No dark green/orange vegetables or legumes |
| **Total Grains** | 5 | ≥3.0 oz equiv., per 1,000 kcal | No grains |
| [**Whole Grains**](https://epi.grants.cancer.gov/hei/comparing.html#f3) | 5 | ≥1.5 oz equiv., per 1,000 kcal | No whole grains |
| **Milk3** | 10 | ≥1.3 c equiv., per 1,000 kcal | No dairy |
| **Meat & Beans** | 10 | ≥2.5 oz equiv., per 1,000 kcal | No meat or beans |
| **Oils4** | 10 | ≥12 grams, per 1,000 kcal | No Oil |
| **Moderation:** | | | |
| **Saturated Fats** | 10 | ≤7% of energy | ≥15% of energy |
| **Sodium** | 10 | ≤0.7 gram, per 1,000 kcal | ≥2.0 grams per 1,000 kcal |
| **Calories from Solid Fats, Alcoholic beverages, and Added Sugar (S0FAAS)** | 20 | ≤20% of energy, per 1,000 kcal | ≥ 50% of energy |

1: Intakes between the minimum and maximum levels are scored proportionately, except for Saturated Fat and Sodium

2: Legumes counted as vegetables only after Meat and Beans standard is met.

3: Includes all milk products, such as fluid milk, yogurt, and cheese, and soy beverages.

4: Includes nonhydrogenated vegetable oils and oils in fish, nuts, and seeds.

5: Saturated Fat and Sodium get a score of 8 for the intake levels that reflect the 2005 Dietary Guidelines, <10% of calories from saturated fat and 1.1 grams of sodium/1,000 kcal, respectively.

1. **Displays and saves the** **results** **in the specified output folder** **(see notes in the provided SAS program).**
2. The program saves total score and set of component scores for the population/group of interest, together with minimum and maximum values, standard errors and confidence intervals. An option is provided to export the results into a CSV file that can be opened in Excel.
3. Print

This step is included as a data check. The min and max can be compared to the bounds of HEI-2005 scores – if any scores <0 or >100, this is a red flag.

This code was written by **Lisa Kahle** and **Dennis Buckman** of Information Management Services, Inc., with assistance from **Kevin Dodd** of the National Cancer Institute. Please send any comments and questions regarding this code to [RFAB@mail.nih.gov](mailto:RFAB@mail.nih.gov).