**ReadMe for NHANES 2003-2004 data and MPED 2.0**

**HEI-2010 By Person**

This ReadMe file describes the methodology implemented in the SAS program that demonstrates how to calculate the Healthy Eating Index-2010 (HEI-2010) component and total scores, using MPED and NHANES data. The provided SAS program, along with the required macros, can be used to calculate Healthy Eating Index (HEI)-2010 total and component scores from 24-hour recall data collected in NHANES. The specific instructions on how to successfully run the HEI procedure are within the provided SAS program.

The program calculates HEI-2010 component and total scores for a set of 24HRs (in this example, two 24HRs). Additional code that calculates HEI-2010 component and total scores for a single 24-hour recall is available on the [NCI Healthy Eating Index website](https://epi.grants.cancer.gov/hei/sas-code.html).

The program has been tested using SAS, version 9.2 and uses 2003-2004 NHANES data. This code can be adapted to calculate HEI-2010 scores for other cycles of NHANES or for other data sources that use a complex sampling strategy.

**NOTE:** The program calculates HEI-2010 scores using a simple method that does not account for measurement error. Therefore, HEI scores obtained using this method should be interpreted with appropriate caveats noting the limitations of this method. More information about measurement error in dietary assessment is available in the [Dietary Assessment Primer](http://dietassessmentprimer.cancer.gov/concepts/error/). When multiple days of 24HR data are available, it is possible to adjust for some of the measurement error inherent in dietary data in many cases. Please see the [Research Uses-Choosing a Method](https://epi.grants.cancer.gov/hei/tools.html) webpage for information about other methods of calculating HEI scores which do account for some of the measurement error inherent in self-report dietary data.

The program creates unweighted HEI-2010 scores. However, the demographic file is included in the example below because it will likely be used in any analysis that uses these scores. Make sure to read in the required variables from the demographic dataset needed for your analysis.

***Required datasets:***

* [NHANES 2003-2004 Dietary Interview data: Individual Foods, First Day and Total Nutrient Intakes, First and Second Day files](http://wwwn.cdc.gov/nchs/nhanes/search/DataPage.aspx?Component=Dietary&CycleBeginYear=2003)
* [NHANES 2003-2004 Demographic Variables and Sample Weights](http://wwwn.cdc.gov/nchs/nhanes/search/DataPage.aspx?Component=Demographics&CycleBeginYear=2003)
* [MyPyramid Equivalents Database (MPED), version 2.0](http://www.ars.usda.gov/Services/docs.htm?docid=17565)
* CNPP MyPyramid Equivalents Databases for Whole Fruit and Fruit Juice for the NHANES cycle 2003-2004 ([cnppmyp\_v1nhanes0304\_wjfrt.sas7bdat](https://www.cnpp.usda.gov/healthy-eating-index-support-files-03-04))

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

* [hei2010.beanspeas.allocation.macro.sas](https://epi.grants.cancer.gov/hei/sas-code.html)
* [hei2010.score.macro.sas](https://epi.grants.cancer.gov/hei/sas-code.html)

The 12 components of the HEI-2010 calculated by the macros are:

* Total Vegetables (HEIX1\_TOTALVEG)
* Greens and Beans (HEIX2\_GREEN\_AND\_BEAN)
* Total Fruit (HEIX3\_TOTALFRUIT)
* Whole Fruit (HEIX4\_WHOLEFRUIT)
* Whole Grains (HEIX5\_WHOLEGRAIN)
* Dairy (HEIX6\_TOTALDAIRY)
* Total Protein Foods (HEIX7\_TOTPROT)
* Seafood and Plant Proteins (HEIX8\_SEAPLANT\_PROT)
* Fatty Acids (HEIX9\_FATTYACID)
* Refined Grains (HEIX11\_REFINEDGRAIN)
* Sodium (HEIX10\_SODIUM)
* Empty Calories (HEIX12\_SOFAAS)

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

The MPED data provide the following components used directly in the calculation of HEI-2010: Total Fruit; Whole Grains; and Refined Grains. The NHANES Total Nutrient Intake data provide Sodium.

*Additional steps are necessary to create the remaining components that are a combination of variables: Total Vegetables; Greens and Beans; Whole Fruit; Dairy; Total Protein Foods; Seafood and Plant Proteins, Fatty Acids and Empty Calories.*

**The SAS program carries out 7 steps:**

1. **Reads in the required datasets and variables and makes necessary edits to the datasets.**
   1. Reads in the MPED, which gives My Pyramid Equivalents per 100 grams of food, and perform edits and corrections, including:

Updates the nutrient and food group values for the following three pizza food codes to those in FPED 2011-2012 to correct for previously identified errors in the MPED 2003-2004 database:

* 58106210, PIZZA, CHEESE, NS AS TO TYPE OF CRUST
* 58106220, PIZZA, CHEESE, THIN CRUST
* 58106230, PIZZA, CHEESE, THICK CRUST

**NOTE**:This correction is specific to 2003-2004 NHANES data. This change should be commented out for any other year of NHANES data.

Moves soy beverages out of Soybean Products (M\_SOY) and into Dairy (D\_TOTAL). The adjusted totals are generated for each day of dietary intake and saved in a temporary SAS dataset.

Calculation note for soy beverages: The calculation of soy beverages affects the Dairy and Total Protein Foods components*.* Soy beverages are counted as part of the Dairy component of the HEI-2010. This differs from the MyPyramid Equivalents Database (MPED), which groups them with other Soybean Products (M\_SOY). Soy beverages (food codes 11310000, 11320000, 11321000, and 11330000) are moved from Soybean Products (M\_SOY), in ounce equivalents, to Total Milk (D\_TOTAL), in cup equivalents, based on the weight in grams of 1 cup. Below are the conversion factors for the four affected food codes:

* 11310000, MILK, IMITATION, FLUID, SOY BASED (1 cup=244 g)
* 11320000, MILK, SOY, READY-TO-DRINK, NOT BABY (1 cup=245 g)
* 11321000, MILK, SOY, READY-TO-DRINK, NOT BABY'S, CHOCOLATE (1 cup=240 g)
* 11330000, MILK, SOY, DRY, RECONSTITUTED, NOT BABY (1 cup=245 g)
  1. Reads in the CNPP MyPyramid Equivalents Databases for Whole Fruit and Fruit Juice for NHANES 2003-04 ([cnppmyp\_v1nhanes0304\_wjfrt.sas7bdat](https://www.cnpp.usda.gov/healthy-eating-index-support-files-03-04)), which gives fruit juice data per 100 grams of food.
  2. Reads in the NHANES 2003-2004 Individual Foods, First Day and Second day datasets (FOOD1 and FOOD2, respectively).  
       
     These two datasets are then merged to create the FOOD dataset, which includes all the food items from both recalls.

**NOTE**: This step imports data only from those participants with reliable dietary recalls; this restricts the dataset to those with reliable dietary recalls.

* 1. Reads in the NHANES 2003-2004 Total Nutrient Intakes, First Day and Second Day datasets (NUTRIENT1 and NUTRIENT2, respectively). (Again, this step imports data only from those participants with reliable dietary recalls.)

These two datasets are then merged to create the NUTRIENT dataset, which includes nutrient information for both recalls combined.

* 1. Reads in the NHANES 2003-2004 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-2010 is based on the 2010 Dietary Guidelines for Americans, which are intended as guidelines for the US population ages 2 and over.

1. **Combines the MPED, Nutrient, and Demographic datasets to create the BOTH dataset.**

Combines the MPED and CNPP Whole Fruit and Fruit Juice database on a food level to create the NEWMPED dataset.

Combines the NEWMPED with the Individual Foods (FOOD) data on a food level to create the FDPYR dataset.

Converts individuals’ food intake amounts from grams to My Pyramid Equivalents.

**NOTE:** MyPyramid equivalent values for total vegetable intake (V\_TOTAL) in the HEI-2010 may be different from V\_TOTAL in the MPED because legumes may be counted as vegetables or meat in the HEI-2010; and total dairy intake (D\_TOTAL) in the HEI-2010 may be different from D\_TOTAL in the MPED because soy beverages are counted as milk in the HEI-2010.

Calculates individual food intake amounts for MyPyramid food groups for one day to create the PYRCALC dataset.

Combines the Nutrient and Demographic datasets on an individual level to create the COHORT dataset.

Combines all data on an individual level to create the BOTH dataset.

**NOTE:** Some participants may be in the COHORT dataset, but not the PYRCALC dataset because the PYRCALC dataset only includes values for those persons with valid food data (so, it does not include participants with valid dietary recalls who ate no foods). In this merge all participants in the COHORT dataset are included so as not to exclude those participants who ate no foods.

Using the array PYRVAR, sets all the MPED servings variables to zero if the MPED servings variables are missing at this point.

**NOTE**: Missing MPED servings variables indicates that a subject had a reliable dietary recall, but consumed no foods.  If no foods were consumed then the calculated amount is missing and needs to be replaced with zeros.

1. **Creates additional required variables: MONOPOLY, ALLMEAT, SEAPLANT, ADDSUGC, SOLFATC, EXALCCAL and EMPTYCAL10.**

Calculation note for MONOPOLY: MONOPOLY sums together monounsaturated and polyunsaturated fatty acids (MFAT + PFAT = MONOPOLY). To estimate the fatty acid ratio of unsaturated fatty acids to saturated fatty acids, the scoring macro will divide this summed value by saturated fatty acids (MONOPOLY/Total saturated fatty acids (g)).

Calculation note for ALLMEAT and SEAPLANT*:* ALLMEAT sums together all animal and plant proteins, including meat, poultry, fish, eggs, nuts, seeds, and the adjusted soy variable (ALLMEAT = M\_MPF (oz) + M\_EGG (oz) + M\_NUTSD (oz) + M\_SOY (oz)); while SEAPLANT sums together all fish and plant proteins, including fish, the adjusted soy variable, nuts, and seeds (SEAPLANT = M\_FISH\_HI (oz) + M\_FISH\_LO (oz) + M\_SOY (oz) + M\_NUTSD (oz)). An additional step is then required to determine how to include LEGUMES in ALLMEAT and SEAPLANT (see Step 4).

Calculation note for Empty Calories (EMPTYCAL10)*:* EMPTYCAL10 sums up three variables, ADDEDSUGC, SOLIDFATC, and EXALCCAL described below, to calculate Empty calories. The Empty Calories component of the HEI is calculated as a percentage of calories, so all of the variables that comprise empty calories need to be converted to units of calories.

ADDSUGC: The nutrient file includes teaspoons of added sugar, which are converted to calories using the conversion factor 1tsp=16kcal.

SOLDFATC: The nutrient file includes grams of added sugar, which are converted to calories using the conversion factor 1g=9kcal.

EXALCCAL: In the HEI-2010, energy from alcohol is considered to be empty calories, but only when alcohol is consumed beyond moderate amounts. The least restrictive of the two levels defined as moderate drinking in the Dietary Guidelines, 2 drinks per day (converted to 28 grams of ethanol), was used to set the threshold for counting alcohol as empty calories. A value of 2150 calories was used to energy-adjust the alcohol threshold, based on the estimated median energy intake of adults. Because 28 grams ethanol/2150 calories equals 13 grams ethanol/1000 calories, only amounts greater than 13 grams of ethanol/1000 calories are counted towards Empty Calories.

1. **Calculates total food group and nutrient intake over all possible days reported per individual.**
2. **Runs the HEI-2010 scoring macros to properly allocate the intakes of LEGUMES** **in cup equivalents to either Total Protein Foods and Seafood and Plant Proteins (ALLMEAT and SEAPLANT) or Total Vegetables and Greens and Beans (V\_TOTAL and V\_DRKGR).**

This section of the code calls the macro, [hei2010.beanspeas.allocation.macro.sas](https://epi.grants.cancer.gov/hei/sas-code.html).

This step results in 4 additional variables, LEGUME\_ADDED\_V\_TOTAL, LEGUME\_ADDED\_BEANGRN, LEGUME\_ADDED\_ALLMEAT, and LEGUME\_ADDED\_SEAPLANT are all used in Step 6.

Calculation notes for Legumes: Intake of LEGUMES counts toward meeting the standard for the Total Protein Foods (and Seafood and Plant Proteins) components first. Once the Total Protein Foods standard is met, regardless of if the Seafood and Plant Protein standard is met, any additional amount of Legumes counts only towards Total Vegetables (and Greens and Beans).

**NOTE:** Units for V\_LEGUMES, as well as for the HEI components Total Vegetables and Greens and Beans are in cup equivalents, therefore, the cup equivalents are first converted to ounce equivalents of meat when they are counted for the Total Proteins and Seafood and Plant Proteins components, and are then converted back to cup equivalents when counted as vegetables.

Calculation note regarding conversion of cup equivalents to ounce equivalents*:* A one-fourth cup equivalent of Legumes is equal to a 1-ounce equivalent of Total Protein Foods and Seafood and Plant Proteins. Thus, the number of cup equivalents of Legumes is multiplied by 4 to convert to ounce equivalents of Total Protein Foods and Seafood and Plant Proteins.

1. **Runs the HEI-2010 scoring macro which calculates intake density amounts and HEI total and component scores.**

The HEI-2010 scoring macro, [hei2010.score.macro.sas](https://epi.grants.cancer.gov/hei/sas-code.html), is called to calculate densities for each HEI-2010 component and then apply the scoring algorithm.

Below are the HEI-2010 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–2010**[**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**[**2**](http://epi.grants.cancer.gov/hei/developing.html#f2) | 5 | ≥0.8 cup equiv. per 1,000 kcal | No Fruit |
| **Whole Fruit**[**3**](http://epi.grants.cancer.gov/hei/developing.html#f3) | 5 | ≥0.4 cup equiv. per 1,000 kcal | No Whole Fruit |
| **Total Vegetables**[**4**](http://epi.grants.cancer.gov/hei/developing.html#f4) | 5 | ≥1.1 cup equiv. per 1,000 kcal | No Vegetables |
| **Greens and Beans**[**4**](http://epi.grants.cancer.gov/hei/developing.html#f4) | 5 | ≥0.2 cup equiv. per 1,000 kcal | No Greens and Beans |
| **Whole Grains** | 10 | ≥1.5 oz equiv. per 1,000 kcal | No Whole Grains |
| **Dairy**[**5**](http://epi.grants.cancer.gov/hei/developing.html#f5) | 10 | ≥1.3 cup equiv. per 1,000 kcal | No Dairy |
| **Total Protein Foods**[**6**](http://epi.grants.cancer.gov/hei/developing.html#f6) | 5 | ≥2.5 oz equiv. per 1,000 kcal | No Protein Foods |
| **Seafood and Plant Proteins**[**6**](http://epi.grants.cancer.gov/hei/developing.html#f6)**,**[**7**](http://epi.grants.cancer.gov/hei/developing.html#f7) | 5 | ≥0.8 oz equiv. per 1,000 kcal | No Seafood or Plant Proteins |
| **Fatty Acids**[**8**](http://epi.grants.cancer.gov/hei/developing.html#f8) | 10 | (PUFAs + MUFAs)/SFAs ≥2.5 | (PUFAs + MUFAs)/SFAs ≤1.2 |
| **Moderation:** |  |  |  |
| **Refined Grains** | 10 | ≤1.8 oz equiv. per 1,000 kcal | ≥4.3 oz equiv. per 1,000 kcal |
| **Sodium** | 10 | ≤1.1 gram per 1,000 kcal | ≥2.0 grams per 1,000 kcal |
| **Empty Calories**[**9**](http://epi.grants.cancer.gov/hei/developing.html#f9) | 20 | ≤19% of energy | ≥50% of energy |

**1:** Intakes between the minimum and maximum standards are scored proportionately.

**2:** Includes fruit juice.

**3:** Includes all forms except juice.

**4:** Includes any beans and peas not counted as Total Protein Foods.

**5:** Includes all milk products, such as fluid milk, yogurt, and cheese, and fortified soy beverages.

**6:** Beans and peas are included here (and not with vegetables) when the Total Protein Foods standard is otherwise not met.

**7:** Includes seafood, nuts, seeds, soy products (other than beverages) as well as beans and peas counted as Total Protein Foods.

**8:** Ratio of poly- and monounsaturated fatty acids to saturated fatty acids.

**9:** Calories from solid fats, alcohol, and added sugars; threshold for counting alcohol is >13 grams/1000 kcal.

1. **Displays and saves the results** **in the specified output folder** **(see notes in the provided SAS program).** 
   1. This program saves one HEI-2010 total score and set of component scores for each individual, based on one 24HR. The results are saved as a CSV file.
   2. Calculates an unweighted mean for all individuals in the group.

This step is included as a data check. The min and max can be compared to the bounds of HEI-2010 scores – if any scores <0 or >100, this is a red flag.

The program was written by **Lisa Kahle**, Information Management Services, Inc. Please send any comments and questions regarding this code to [RFAB@mail.nih.gov](mailto:RFAB@mail.nih.gov).