**ReadMe for NHANES 2009-2010 data and FPED**

**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 FPED 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 each individual, using a set of 24HRs (in this example, two 24HRs). Additional code that calculates HEI-2010 component and total scores for a single 24HR 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 2009-2010 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:** This SAS 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 SAS 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 2009-2010 Dietary Interview data: Total Nutrient Intakes, First and Second Day files](http://wwwn.cdc.gov/Nchs/Nhanes/Search/DataPage.aspx?Component=Dietary&CycleBeginYear=2009)
* [NHANES 2009-2010 Demographic Variables and Sample Weights](http://wwwn.cdc.gov/Nchs/Nhanes/Search/DataPage.aspx?Component=Demographics&CycleBeginYear=2009)
* [Food Patterns Equivalents Database (FPED), 2009-2010: FPED\_DR1TOT\_0910\_sas.exe and FPED\_DR2TOT\_0910\_sas.exe](http://www.ars.usda.gov/Services/docs.htm?docid=23869) (these are zipped files that contain the SAS files “fped\_dr1tot\_0910.sas7bdat” and “fped\_dr2tot\_0910.sas7bdat”)

***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 Food Patterns Equivalents Database (FPED) and NHANES output, but others are created as part of the SAS program.

The FPED totals data provide the following components used directly in the calculation of HEI-2010: Total Fruit; Whole Grains; Dairy; 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; 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 FPED, which gives Food Patterns Equivalents per person per day for days 1 and 2 (FPED1 and FPED2, respectively), and rename variables.  
        
      Then, FPED1 and FPED2 are concatenated to create the FPED dataset.
   2. Reads in the NHANES 2009-2010 Total Nutrient Intakes, First Day and Second Day datasets (DR1TOT\_F and DR2TOT\_F, respectively) and rename variables.

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

**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 2009-2010 Demographic dataset (DEMO\_F) 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** **FPED, Nutrient, and Demographic datasets to create the COHORT dataset.**

First, the FPED and Nutrient dataset are merged by SEQN and DAY to create the NUTFDPYR dataset. The NUTFDPYR dataset is then merged with the Demographic dataset by SEQN to create the COHORT dataset.

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

Calculation note for WHOLEFRT: FWHOLEFRT sums together Citrus, Melons and Berries with Other Fruits to generate the value for Whole Fruit (non-juice) consumption. (FWHOLEFRT = DR1T\_F\_CITMLB (cups) + DR1T\_F\_OTHER (cups)).

Calculation note for MONOPOLY: MONOPOLY sums together monounsaturated and polyunsaturated fatty acids (MONOPOLY = DR1TMFAT + DR1TPFAT). 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 soy (ALLMEAT=DR1T\_PF\_MPS\_TOTAL(oz) +DR1T\_PF\_EGGS(oz) +DR1T\_PF\_NUTSDS(oz) +DR1T\_PF\_SOY(oz) ); while SEAPLANT sums together all fish and plant proteins, including fish, nuts, seeds, and soy(SEAPLANT=DR1T\_PF\_SEAFD\_HI(oz) +DR1T\_PF\_SEAFD\_LOW(oz) +DR1T\_PF\_NUTSDS(oz) +DR1T\_PF\_SOY(oz)).

*An additional step is then required to determine how to include LEGUMES in ALLMEAT and SEAPLANT (see Step 5).*

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 (DR1T\_V\_TOTAL and DR1T\_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, named LEGUME\_ADDED\_V\_TOTAL, LEGUME\_ADDED\_BEANGRN, LEGUME\_ADDED\_ALLMEAT, and LEGUME\_ADDED\_SEAPLANT, all used in the HEI scoring macro 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 whether 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).