Using SAS to Expand the Application of Standard Measures and Guide Statistical Explorations: Creating Healthy Eating Index Scores Using Nutrition Data System for Research Output

David A. Ludwig, David C. Landy, Joy M. Kurtz, and Tracie L. Miller
Division of Pediatric Clinical Research, University of Miami Miller School of Medicine, Miami, FL

Background

Composite measures are frequently developed in healthcare and often involve summarizing several variables to provide a more valid or complete description of a concept such as hospital quality or patient health. For instance, new measures may be created to assess adherence to guidelines or to describe how combinations of variables combine to predict risk. Applying a composite measure can be challenging for several reasons. First, the variables comprising the composite may not be readily available in a single data file, such as from a quality improvement report or patient record, or the variables may not have been measured at the appropriate level, such as per procedure versus total stay expenditures. Second, the calculation method may be complex, requiring the performance of algebraic manipulations within a series of logical statements. However, all of these issues can be overcome by creating a SAS program which can not only combine various data sources and summarize data at different levels but which can also automate algebraic calculations and handle complex logical statements. Once created, several issues can limit the application of composite measures both in practice and in research. Issues such as difficulty in the interpretation of the underlying latent concept, questions concerning reliability and validity, and difficulty applying the new measure to specific situations which can be addressed using SAS programs written to focus attention on interpretation and relevant statistical issues.

We describe the creation of a SAS program to calculate a measure of diet quality, the Healthy Eating Index (HEI) 2005 (http://www.cnpp.usda.gov/HealthyEatingIndex.htm), using output from a commonly employed dietary software package, Nutrition Data System for Research (NDSR, http://www.ncc.umn.edu/products/ndsr.html) . Currently, application of the HEI in research is limited by the challenges posed in calculating the HEI using NDSR output. First, the necessary variables are stored in three different output files, two of which are at the individual-day level and one of which is at the food-individual-day level. Second, the calculation process involves a combination of algebraic manipulations and logical statements. In this paper we describe how SAS can be used to overcome these issues. We also offer suggestions for increasing usability such as with the %INCLUDE statement and creating a simplified version utilizing the ability to run a SAS program in batch mode. Finally, we show how the program can be used to explore related statistical issues. Specifically, because the HEI is a ratio of variables at the person-day level, the mean HEI can be calculated as the mean of the individual-day ratios or as the ratio of the means of the individual-days. Finally, when multiple days are recorded per individual, it is possible to estimate the HEI reliability and formulate projections as to the optimum number of daily food recall determinations.

Measuring and Analyzing Dietary Data (The HEI 2005 and the NDSR)

The HEI 2005 is a composite score calculated from 12 components of food group consumption indexed to the Dietary Guidelines for Americans, 2005.[1] Although there are several methods for obtaining the types, quantity, and frequency of foods consumed by individuals over a given period of time, in general, individuals are asked to recall their food choices, portion sizes, and frequencies and then these reported consumptions are classified into food group components that are scored and summed to provide an overall index of diet quality. In order to obtain an HEI score, reported foods must be analyzed and then classified by either type (e.g., fruits and vegetables) or composition (e.g., oils and sodium). Perhaps the most widely used computer based methodology for this type of classification is the NDSR (see web link above). This MS Windows based program allows for the detailed inputting of the reported foods and produces several data files contain virtual every known constituent of dietary intake. The dietary information generated from this program can then be used to calculate the HEI.[2]

As previously stated, several issues make the calculation of HEI somewhat tedious. First, the HEI scoring algorithm is somewhat complicated. In order to obtain proper values, some components require segmented linear interpolation and/or algebraic solutions. Second, the NDSR generates multiple data files containing a plethora of dietary variables with some of these variables being measured at the level of the individual food though values are needed at the level of the individual per day. These issues make hand calculation of the HEI for other than a few individuals highly impractical.

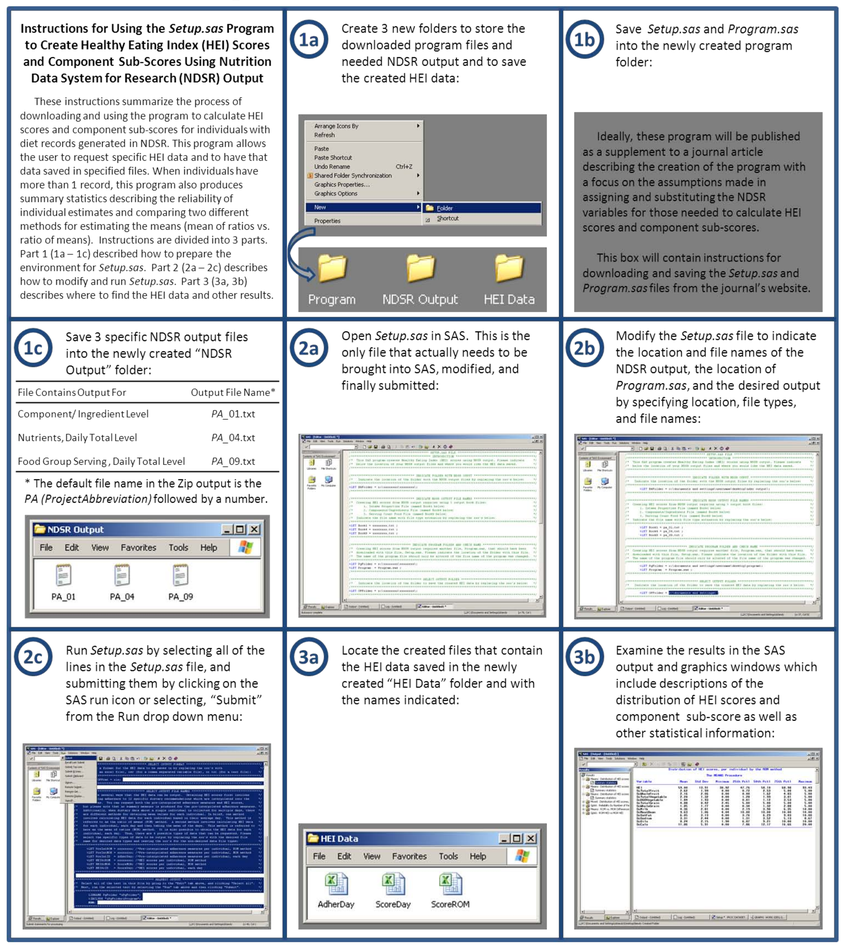
Although simple generic HEI scoring programs have previously been written (http://riskfactor.cancer.gov/tools/hei/tools.html), these programs require a very high degree of dietary data preprocessing. This limits the portability and in turn the utility of these programs by third party users. The following SAS program was constructed to solve this problem. It was written so that the raw data files from the NDSR program can be used without any type of editing or manipulation. All that is required is a basic knowledge of MS Windows (for setting up folders and files), three output files which are generated from the NDSR program (i.e., Book1, Book4, and Book9), and basic SAS software (Cary, NC).

Overview of Using the SAS Program

We created a program in SAS that would automate the process of calculating HEI scores and component sub-scores using the unedited NDSR output. To increase usability, we created two SAS program files. The first program (Setup.sas) is meant to be opened and edited by the user (Appendix). This file allows the user to indicate where the NDSR output is located and indicate preferences regarding the calculated HEI data. The second program (Program.sas) contains the code to calculate the HEI component scores and total based on the user specified information. This second file does not need to be edited or even opened by the user.

Using these programs simply requires following a series of basic steps which can be divided into 3 parts. First, users prepare the environment for Setup.sas. Second, users modify and run Setup.sas. Third, users open the designated files containing the HEI composite score data and examine the other results. The steps involved in conducting these parts are summarized in an instruction page.

Instructions Page



Modify and Run *Setup.sas* (Part 2)

First, the user will need to open the file *Setup.sas* in SAS. This file is used to indicate the location of the NDSR output needed to calculate the HEI scores, where the HEI result (i.e., data) should be saved, as well as specify amongst output options. This *Setup.sas* file is the only file that actually needs to be brought into SAS, edited and finally submitted.

Second, the user should follow the instructions in the *Setup.sas* file. At this step, the user will specify the location of the folder containing the NDSR output files, specify the names of the 3 specific NDSR output files needed, and specify the location of the folder in which the *Program.sas* file is located. The user will then indicate the location of the folder in which the HEI data should be saved, indicate the format for the HEI data to be saved and finally, indicate which types of possible HEI data are wanted by replacing the xxx's with a file name selected by the user. Besides the standard HEI component scoring and scaling, users can request the raw equivalents per 1000 kcal. These are pre-interpolated values on each of the 12 HEI component sub-scores. Also, when multiple days are evaluated per individual, the summary statistic which relates the ratio of the component to the energy intake can be calculated in two different ways. One way is to separately average both the component and the energy across days and then form the ratios of these two averages (i.e., ratio of means, ROM). Alternatively, for each subject, the ratio of the component to energy can be calculated separately for each day and these ratios averaged (i.e., mean of ratios, MOR).

Third, the user should select all lines in the *Setup.sas* file and submit.

Setup.SAS

```
/*********************************************** SETUP.SAS FILE ***********************************************/
/*********************************************** INTRODUCTION *************************************************/
/*  This SAS program creates Healthy Eating Index (HEI) scores using NDSR output. Please indicate             */
/*  below the location of your NDSR output files and where you would like the HEI data saved.                 */
/**************************************************************************************************************/

/********************************** INDICATE FOLDER WITH NDSR OUTPUT **********************************/
/*  Indicate the location of the folder with the NDSR output files by replacing the xxx's below:      */
/******************************************************************************************************/
     %LET BkFolder = x:\xxxxxxxx\xxxxxxxx\;
/******************************************************************************************************/

/********************************** INDICATE NDSR OUTPUT FILE NAMES ***********************************/
/*  Creating HEI scores from NDSR output requires using 3 output files:                               */
/*     1. Intake Properties File (named Book1 below)                                                  */
/*     2. Components/Ingredients File (named Book4 below)                                             */
/*     3. Serving Count Food File (named Book9 below)                                                 */
/*  Indicate the file name with file type extension by replacing the xxx's below:                     */
/******************************************************************************************************/
     %LET Book1 = xxxxxxxx.txt ;
     %LET Book4 = xxxxxxx.txt ;
     %LET Book9 = xxxxxxxx.txt ;
/******************************************************************************************************/

/********************************** INDICATE PROGRAM FOLDER AND CHECK NAME ****************************/
/*  Creating HEI scores from NDSR output requires another file, Program.sas, that should have been    */
/*  downloaded with this file, Setup.sas. Please indicate the location of the folder with this file.  */
/*  The name of the program file should only be altered if the file name of the program was changed.   */
/******************************************************************************************************/
     %LET PgFolder = x:\xxxxxxxx\xxxxxxxx\;
     %LET Program  = Program.sas ;
/******************************************************************************************************/

/********************************** SELECT OUTPUT FOLDER **********************************************/
/*  Indicate the location of the folder to save the created HEI data by replacing the xxx's below:    */
/******************************************************************************************************/
     %LET OPFolder = x:\xxxxxxxx\xxxxxxxx\;
/******************************************************************************************************/

/********************************** SELECT OUTPUT FORMAT **********************************************/
/*  Indicate a format for the HEI data to be saved in by replacing the xxx's with                     */
/*  xls (for an excel file), csv (for a comma separated variable file), or txt (for a text file):     */
/******************************************************************************************************/
     %LET OPFrmt = xxx;
/******************************************************************************************************/

/********************************** SELECT OUTPUT FILE NAMES ******************************************/
/*  There are several ways that the HEI data can be output.  Obtaining HEI scores first involves      */
/*  calculating adherence to 12 specific dietary recommendations one which are interpolated into the */
/*  HEI scores.  You can request both the pre-interpolated adherence measures and HEI scores,        */
/*  but please note that no summary measure is produced for the pre-interpolated adherence measures. */
/*  Additionally, when dietary data about a single individual is collected for multiple days, there  */
/*  are different methods for obtaining mean values for each individual. In brief, one method       */
/*  involves calculating HEI data for each individual based on their average day.  This method is    */
/*  refered to as the ratio of means (ROM) method.  A second method involves calculating HEI data    */
/*  for each individual, each day and then taking the mean of the days.  This method is referred to  */
/*  here as the mean of ratios (MOR) method.  It is also possible to obtain the HEI data for each    */
/*  individual, each day.  Thus, there are 6 possible types of data that can be requested. Please     */
/*  select the specific types of data to be output by replacing the xxx's with the desired file      */
/*  name for desired data types and leaving the xxx's for the non-desired data file types:           */
/******************************************************************************************************/
     %LET PreIntROM = xxxxxxxx; /*Pre-interpolated adherence measures per individual, ROM method     */
     %LET PreIntMOR = xxxxxxxx; /*Pre-interpolated adherence measures per individual, MOR method     */
     %LET PreIntID  = xxxxxxxx; /*Pre-interpolated adherence measures per individual, each day       */
     %LET HEIScROM  = xxxxxxxx; /*HEI scores per individual, ROM method                              */
     %LET HEIScMOR  = xxxxxxxx; /*HEI scores per individual, MOR method                              */
     %LET HEIScID   = xxxxxxxx; /*HEI scores per individual, each day                                */
/******************************************************************************************************/

/********************************** REQUEST OUTPUT ****************************************************/
/*  Select all of the text in this file by going to the "Edit" tab above, and clicking "Select All".  */
/*  Next, run the selected text by selecting the "Run" tab above and then clicking "Submit".          */
/******************************************************************************************************/
     LIBNAME PgFolder "&PgFolder";
     %INCLUDE "&PgFolder&Program";
     RUN;
/******************************************************************************************************/
```

Details and Selected Aspects of Program.sas

Program.sas contains the code to perform the actual manipulations of the NDSR output to produce HEI scores and component sub-scores. *Program.sas* is called upon in the *Setup.sas* program using the %INCLUDE statement which minimizes the number of files the user interacts with and allows the user to interface with the series of programs in a more controlled atmosphere in which navigation around lengthy code, parts of which may be unfamiliar to some users, is avoided. Below, selected aspects of *Program.sas* are described in more detail for those interested in the details of the actual SAS programming.

References

1. Guenther PM, Reedy J, Krebs-Smith SM. Development of the Healthy Eating Index-2005. J Am Diet Assoc 2008;108:1896-1901.
2. Miller PE, Mitchell DC, Harala PL, Pettit JM, Smiciklas-Wright H, Hartman TJ. Development and evaluation of a method for calculating the Healthy Eating Index-2005 using Nutrition Data System for Research. Public Health Nutr 2011;14:306-13.
3. Qiao CG, Wood GR, Lai CD, Luo DW. Comparison of two common estimators of the ratio of the means of independent normal variables in agricultural research. J Appl Math Decis Sci 2006;2006:1-14.

AUTOMATICALLY GENERATED PROGRAM OUTPUT

Exploring Different Estimates of the Mean HEI Score for an Individual:

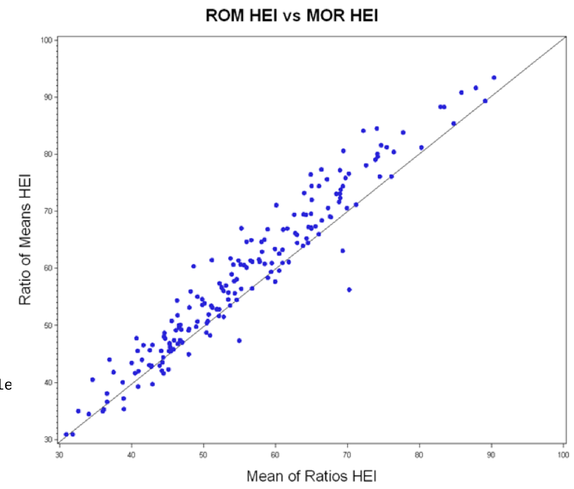
The HEI score is the summation of several component sub-scores, each of which is based on the ratio of two variables, a measure of the amount of a specific food or nutrient consumed in a day and a measure of total consumption for that day. When multiple records are available for an individual, the ratio can be calculated as the average of both variables which is called the ratio of means (ROM) approach or as the average of the ratios of the variables for each day which is called the mean of ratios approach (MOR).[3] These different estimates are merged into a single file and the difference between them for each component sub-score and the HEI score is calculated. The code and an example table and graph are presented below:

```
%MACRO Name(d = , p =);
DATA OPfolder.HEIScn&d;
      SET OPFolder.HEIS&d;
%LET k = %SYSFUNC(COUNTW(&HEI_Sc_Names));
%DO i = 1 %TO &k;
      %LET y  = %SCAN(&HEI_Sc_Names, &i);
      RENAME &y=&p&y;
%END;
RUN;
%MEND;
%Name(d=a, p =ROM); %Name(d=b, p =MOR);

DATA OPFolder.compare;
      MERGE OPFolder.HEIScnb OPFolder.HEIScna; BY Participant_ID;
RUN;

%MACRO Diff;
DATA OPFolder.compare;
      SET OPFolder.compare;
            %LET k = %SYSFUNC(COUNTW(&HEI_Sc_Names));
            %DO i = 1 %TO &k;
                  %LET y  = %SCAN(&HEI_Sc_Names, &i);
                        Diff&y = ROM&y - MOR&y;
            %END;
RUN;
%MEND;
%Diff;

PROC MEANS DATA=OPFolder.compare;
   TITLE 'ROM vs. MOR Differences for HEI scores';
   VAR DiffHEI DiffScTotalFruit DiffScWholeFruit DiffScTotalVegetable
   DiffScDGOVegetable DiffScTotalGrain DiffScWholeGrain DiffScMilk
   DiffScMeatBean DiffScSatFat DiffScSodium DiffScOils DiffScCals;
RUN;

AXIS1 LABEL=(ANGLE=90 ROTATE=0 HEIGHT=2 "Ratio of Means HEI");
AXIS2 LABEL=(HEIGHT=2 "Mean of Ratios HEI") MINOR=NONE;
TITLE 'ROM HEI vs MOR HEI'; SYMBOL value=dot interpol=none;
DATA anno;
      function='move'; xsys='1'; ysys='1'; x=0; y=0; output;
      function='draw'; xsys='1'; ysys='1'; x=100; y=100; output;
RUN;
PROC GPLOT DATA=OPFolder.compare;
      PLOT ROMHEI*MORHEI / VAXIS=AXIS1 HAXIS=AXIS2 ANNO=anno;
      RUN;
```

ROM vs. MOR Differences for HEI scores and component sub-scores

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
DiffHEI	124	2.4426589	4.0756843	-21.4499738	12.3582007
DiffScTotalFruit	124	0.2256818	0.5275919	-0.7591776	1.7417500
DiffScWholeFruit	124	0.1970122	0.4976800	-0.7342397	1.9110712
DiffScTotalVegetable	124	0.1045641	0.3278275	-0.6100893	1.5542596
DiffScDGOVegetable	124	0.2930498	0.7612785	-0.8807825	3.3333333
DiffScTotalGrain	124	0.2524940	0.4871636	-0.7087304	3.0429997
DiffScWholeGrain	124	0.2914218	0.6726610	-0.5495573	3.3333333
DiffScMilk	124	0.2650865	0.5985580	-0.7082368	2.6159600
DiffScMeatBean	124	0.7138303	0.9584041	-0.8099236	5.4208973
DiffScSatFat	124	-0.1216408	1.2963917	-6.4731270	2.4141296
DiffScSodium	124	-0.3177233	0.7041723	-2.9561221	1.5398615
DiffScOils	124	0.8943167	1.1491947	-1.0673972	5.3151389
DiffScCals	124	-0.1900248	1.4453838	-9.1478127	4.2057614

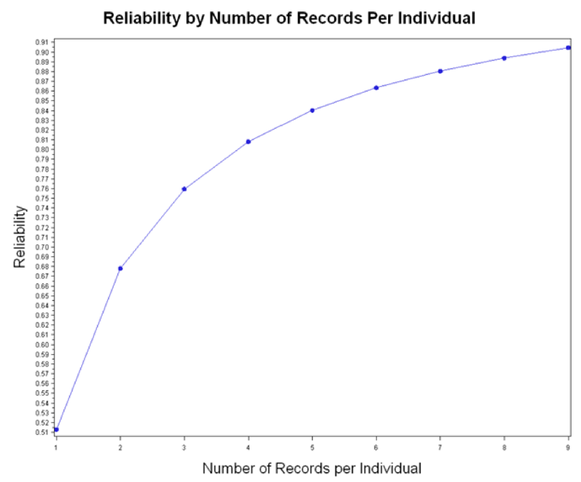


These results show that the ROM approach produces consistently higher estimates of the HEI score compared to the MOR approach. This is consistent with the idea that individuals can consume large amounts of a specific food group one day, the value of which is truncated under the MOR approach but not the ROM approach.

Exploring Study-Specific Reliability:

A major limitation in the study of diet, especially in developed countries with significant heterogeneity in food options, is the large amount of variation across diets which exists both between individuals of a specific group and within individuals. With respect to variation within individuals, it is possible to assess the average reliability of the estimates for individuals and to use this information to estimate the expected average reliability for these individuals given a specific number of daily food recalls per individual.[4] The following code uses the MIXED procedure to estimate the variation between individuals in the NDSR output, sigmasq, and within individuals, tau00. These estimates are used to calculate the number of individual records needed to obtain a reliability of 90%. Then, the estimates are used to calculate the expected reliability associated with each specific number of records between 1 and the number of records needed to obtain a reliability of 90%. These results are then plotted so that the user can understand the reliability of the estimates for specific individuals. This offers important information in deciding how to move forward using the calculated HEI scores. The code and an example graph are presented below:

```
PROC MIXED DATA=OPFolder.HEISb0;
      CLASS Participant_ID;
      MODEL HEI=;
      RANDOM Intercept / SUBJECT=Participant_ID;
      ODS OUTPUT CovParms=OPFolder.CovParms;
RUN;

DATA OPFolder.CovParms;
      SET OPFolder.CovParms;
      IF CovParm='Residual'  THEN CALL SYMPUT('sigmasq', Estimate);
      IF CovParm='Intercept' THEN CALL SYMPUT('tau00'  , Estimate);
RUN;

%MACRO Reliability;
%LET n = %SYSEVALF((&sigmasq/((&tau00/.90)-&tau00)+1), INT);
DATA OPFolder.Reliability;
      %DO i=1 %TO &n;
            records=&i;
            reliability=(&tau00/(&tau00+(&sigmasq/&i)));
            OUTPUT;
      %END;
RUN;
%MEND;
%Reliability;

TITLE 'Reliability by Number of Records Per Individual';
AXIS1 LABEL=(ANGLE=90 ROTATE=0 HEIGHT=2 "Reliability");
AXIS2 LABEL=(HEIGHT=2 "Number of Records per Individual") MINOR=NONE;
SYMBOL value=dot interpol=join;
PROC GPLOT DATA=OPFolder.reliability;
      PLOT reliability*records / VAXIS=AXIS1 HAXIS=AXIS2;
RUN;
```



In this case, the results provided show that with just three days of diet records per individual, the reliability of the estimate for the individual's mean HEI score is over 75% suggesting that these estimates may provide sufficient reliability to use these HEI scores in analyses at the individual level.

Ludwig, Landy, Kurtz, and Miller

UHealth Pediatrics
UNIVERSITY OF MIAMI HEALTH SYSTEM
UNIVERSITY OF MIAMI
MILLER SCHOOL of MEDICINE

SAS® GLOBAL FORUM