# A New, Obesity-specific Healthy Eating Index (OS-HEI)

Radwan, Amr[1]; Gil, Jose Maria[2]; Variyam, J.N.[3]

[1]CREDA – UPC – IRTA, Barcelona, Spain, amr.radwan@upc.edu
And Department of Agricultural Economics, Cairo University, Cairo, Egypt.
[2] CREDA – UPC – IRTA, Barcelona, Spain, chema.gil@upc.edu
[3] Food Economics Division - Economic Research Service, USDA, Washington D.C., USA, JVARIYAM@ers.usda.gov

Disclaimer "The views expressed in this paper are those of the authors and do not reflect the views of the Economic Research Service or the USDA."

**Paper prepared for presentation at the EAAE-AAEA Joint Seminar**
**'Consumer Behavior in a Changing World: Food, Culture, Society"**

March 25 to 27, 2015
Naples, Italy

## Abstract

Because of the strong relationship between health status and diet quality, many indices have been developed to measure diet quality. Although most of these indices are intended to show a relationship with health outcomes, in all indices this association has been proved to be moderate, especially in the case of obesity. Moreover, all suggested indices are based on a high level of subjectivity in relation to choosing the components to build up the index and the cut-off points for each component. Even in the case of less subjective indices, such as the Healthy Eating Index (HEI), it depends on a review of scientific literature on nutrition and health effects done by a distinguished panel of scientists and the HEI components and their weighting are objectively based on this review. Both issues have reinforced the need to develop a disease- and obesity-specific diet quality index, which is the main objective of this paper. To reduce subjectivity, we propose the use of a new, non-parametric approach – the multivariate adaptive regression splines (MARS) – to develop our new, Obesity-specific Healthy Eating Index (OS-HEI), which is a data-driven, non-parametric tool and allows for interaction between the different items. The data used comes from the 2007–08 and 2009–10 National Health and Nutrition Examination Surveys (NHANESs). While the data from 2007–08 has been used to develop the new index, the data from 2009–10 has been used to validate the results. The traditional HEI-2010 index has been used as a benchmark. Results indicate that the OS-HEI notably outperforms the HEI-2010 in predicting obesity prevalence.



**Key words:** Obesity, Overweight, Healthy Eating Index, Obesity-Specific Healthy Eating Index (OS-HEI), Diet Quality indices.

**Topic:** Consumer protection in the new scenarios: is there a role for policies?

## Introduction

Obesity is a growing, worldwide epidemic. Overweight and obesity are defined as 'abnormal or excessive fat accumulation that may impair health'.[1] Many different studies worldwide have shown association between food intake and obesity in both children and adults.[2] Promoting a healthy diet could be considered a top priority for many countries. A key step to doing so is to measure the diet quality. Although diet quality is universally recognized as a key determinant of overweight and obesity prevalence, there is still a lack of consensus on how to measure it.

To achieve this objective of being able to assess diet quality, many diet quality indices have been developed during the last decades and some of them used to predict the health outcomes (such as obesity) related to food quality.

In fact, most existing indices are able to predict health outcomes to some extent, but the associations are generally modest for all dietary scores, casting doubts on their validity. This may be explained by the many arbitrary choices in the development of an index, and the lack of insight into the consequences of these choices. The main choices relate to the components to include in the score, the cut-off values to compare intake with, and the exact method of scoring. In addition, diet quality scores may still not adequately deal with the main reasons for a holistic approach, which is the correlation between intakes of various dietary groups. [3] This weak association could also be due to a wide variety of health outcomes, as some nutrients or foods could be considered determinant only for specific health outcomes, while irrelevant for other health outcomes. This casts a shadow of doubt on the validity of these indices and raises the need to develop disease- and obesity-specific diet quality indices. Even in the case of less subjective indices, such as the HEI, they depend on a review of scientific literature on nutrition and health effects done by a distinguished panel of scientists, and the HEI components and their weighting are objectively based on this review. To avoid subjectivity, we used the MARS in developing our new OS-HEI. The main advantage of using the MARS is that it is a data-driven, non-parametric tool which allows you to avoid subjectivity in choosing index items and their cut-off points. Moreover, the MARS allows for interaction between the different items, permitting a holistic approach which takes into account the interaction between different nutrients and food groups. In this paper, the NHANES data set for the year 2007–08 has been used for the development of our new index, while the NHANES data set for the year 2009–10 has been used to validate the index and evaluate its ability to predict the effect of diet quality on obesity prevalence. An association was found between the new OS-HEI and obesity prevalence, and this association increased significantly by controlling for some socio-demographics, such as age and gender. Moreover, our new OS-HEI notably outperformed the HEI-2010 in predicting obesity prevalence.

The rest of this paper is organized as follows. Section 2 provides a review of diet quality indices and their association with health outcomes, in particular obesity, with special emphasis on the HEI. A brief description of the data used (NHANES) is given in Section 3. The methodological approach applied in the development and validation of our new OS-HEI is explained in Section 4. The main results are discussed in Section 5. Finally, Section 6 provides some concluding remarks.

## Diet quality indices

Diet quality appears to have no official definition in related literature. Definitions vary widely, depending on the tools used to measure it. Traditionally, a common perception has been that dietary quality reflects nutrient adequacy. Nutrient adequacy refers to a diet that meets requirements for energy and all essential nutrients.[4] The more recent worldwide concern regarding overweight and obesity prevalence is that this is mainly caused by excess intake of certain nutrients and foods, shifting the definition of dietary quality to include both concepts of nutrient deficiency and over-nutrition.[6]

According to Drewnowski[7], in many cases, the only criteria for defining healthy foods is being free of problematic ingredients, such as fat, sugar and sodium, and not by the beneficial nutrients they might contain.

Nutrient-dense foods lack a common definition.[7 8] In 1977, Guthrie[7] asserted that there were only limited efforts to define the concept of a nutritious food, finding only some general statements that did not depend on clear standards or criteria. In 2004, Lackey and Kolasa[8] affirmed that there was still no agreement on the definition of a nutritious or healthy food and beverage. Past attempts to quantify the nutrient density of foods have been based on a variety of calories-to-nutrient scores, nutrients-per-calorie indexes and nutrient-to-nutrient ratios.[6]

Assessment of diet quality is concerned with both the quality and variety of the diet as a whole, rather than individual nutrients, and allows for the evaluation of how closely eating patterns align with dietary recommendations. Panagiotakos[9] (pp. 1) defined indices as ‘composite tools aiming to measure and quantify a variety of clinical conditions, behaviors, attitudes and beliefs that are difficult to be measured quantitatively and accurately’.

Diet quality indices add an important dimension to dietary assessment, as a composite measure of diet has been noted as preferable to an index of a single nutrient or food in the area of dietary assessment.[10 11] Diet quality indices, or scores, are tools that provide an overall rating, on a numeric scale, of an individual’s dietary intake with regards to nutrient and/or dietary recommendations. [12] In addition, Hu[13] emphasized the importance of using composite indices to avoid the problems which could result from entering a large number of highly correlated explanatory variables (i.e., the components of the index) in a model. This could generate multi-collinearity problems, resulting in less robust estimations of the coefficients and less accurate predictions.

Many indices of overall diet quality were developed during the last decades. Two types of measurement can be identified – predefined diet quality indices that assess compliance with prevailing dietary guidance as dietary patterns and empirically derived diet patterns. [14]

Predefined, or theoretically defined, indices consist of nutritional variables essentially, nutrients and foods or food groups that are assumed to be either healthful or detrimental. The index variables are quantified and summed to provide an overall measure of dietary quality. The definition of diet quality depends on attributes selected by the investigator. It is built upon current nutrition knowledge or theory, or based on a diet that has been proven to be healthy, such as the Mediterranean diet.

Many predefined diet quality scores have been proposed. Four of them have gained most attention – the Healthy Eating Index[15,] the Diet Quality[16], the Healthy Diet Indicator[17], and the Mediterranean Diet Score[18]. Many others have been developed by carrying out several modifications to these indices.

Indices differ in several aspects, such as the items included, the cut-off values used and the exact method of scoring, indicating that many arbitrary choices were made.[19] This means that development of such indices involves a high level of subjectivity.

Indices' components range from nutrients only, to adherence to recommended food group servings, to variety within healthful food groups. [12] Kant[10] mentioned that the construction of diet quality indices is mainly based on food groups, or specific foods, or nutrient intakes, or derived from combinations of both foods and nutrient intakes. Waijers *et al.* [3] added that scores used national dietary guidelines or a Mediterranean pattern as a reference. Most indices, including the HEI, are based on both food groups and nutrients, while some, such as the Healthy Food Index (HFI), are based on foods and food groups.

Almost all indices included five food groups, which are: vegetables and fruits, cereals or grain, dairy, and meat and meat products. Regarding nutrients, there is a consensus on including fat in the index – fat introduced in different forms as a total fat, and/or saturated fatty acid (SFA), or the ratio of monounsaturated fatty acids (MUFA) to SFA. Cholesterol and alcohol are also included in many indices. Moreover, the number of dietary variables used in each dietary index varies. It can be observed that although the vast majority of indices have been constructed using nine or 10 components, there are some indices developed using fewer or more components. [3]

Once the variables have been selected to be included in an index, they need to be quantified. The simplest method is to use a cut-off value for each index item and assign a score of '0' if consumption is lower than this value (or higher) and '1' if consumption is higher (or lower) than this cut-off point, respectively.

Although very important, the relative contribution of the various index items to the total score has rarely been addressed. In most existing indices, all items contribute equally to the total score, in other words, all components have the same weighting. Only a few indices have been constructed assigning specific weightings to some components (e.g. HEI-2010). However, Kourlaba *et al.*[20] mentioned that the weightings were attributed arbitrarily. The authors did not explain how these weightings were calculated; they only reported that higher weightings were assigned to components considered to be more important for diet quality, based on specific dietary guidelines. This could be considered as a main contribution of this paper, as we used the MARS model in weighting our new OS-HEI. This indicates that weighting index components is totally objective and depends only upon the role of each component in overweight and obesity prevalence.

Development of diet quality indices includes a high level of subjectivity, as many arbitrary choices related to the index items, or components, that should be included in the index, the cut-off values that should be used for each component and the weightings that should be assigned to each component, take part in the composition of an index.

Empirically derived dietary patterns represent another way of examining dietary patterns, using an a posteriori approach, in which statistical methods like factor and cluster analysis are used to generate patterns from collected dietary data.

In factor analysis of dietary patterns, the so-called factors are discerned based on correlations between variables – generally foods or food groups. Correlated variables are grouped together, distinct from groups of variables with which they are not correlated. Individuals are given a score for each factor. In contrast to factor analysis, cluster analysis does not aggregate intake variables but individuals into relatively homogeneous subgroups (clusters) with similar diets. A summary score for each pattern can be derived and used in either correlation or regression analysis to examine relationships between various eating patterns and the outcome of interest, such as nutrient intake, cardiovascular risk factors and other biochemical indicators of health. [21]

Although factor and cluster analysis could be considered as data-driven techniques, a degree of subjectivity exists, as choices have to be made in each of the consecutive steps in the analytical process. These steps are somewhat alike for factor and cluster analysis. First, the foods or food groups for entry into the analysis need to be selected, foods need to be assigned to food groups, and input variables can (or cannot) be adjusted, for example, for energy intake. The analysis itself and the identification of the dietary patterns or clusters are not straightforward either and also involve choices. In the majority of factor analysis studies in nutritional epidemiology, principal component analysis has been applied, using orthogonal rotation and eigenvalues >1. In cluster analysis studies, K-means method [22] was most often used, but Ward's [23] method was also regularly applied. Both the parameters of the resulting factor and cluster solutions and the interpretability, as decided on by the researcher, determine which solution is finally reported. The number of derived factors reported generally ranged from two to 25 and for most studies the total of variance explained by all factors was limited, in general, between 15% and 40%. The number of resulting clusters varied from two to eight. The researcher also gives names to the factors or clusters, and although the factor or cluster loadings are generally reported in the published results, labeling does play a critical role in the interpretation. Up to present, there is not yet enough insight with regard to what extent outcomes are influenced by choices, including treatment of the input variables and the factor or clustering method used.[21 24]

*The Healthy Eating Index (HEI)*

The HEI is considered one of the most prevalent diet quality indices. It is a summary measure of the overall quality of people's diets. The HEI[15] is an index developed by the US Department of Agriculture (USDA) and the US Department of Health and Human Services (USDHHS), based on the *Dietary Guidelines for Americans 1995,* [25] as a tool to measure compliance with dietary guidance. The HEI consisted of 10 components – grains, vegetables, fruits, milk and meat intakes, total fat and saturated fat intakes as a percentage of total energy intake, and cholesterol and sodium intakes as milligrams and variety in a person's diet. Scores between 0 and 10 were assigned to each component – where 0 indicates non-compliance with recommended amounts or ranges, while 10 indicates intakes close to recommended amounts or ranges. The scores in-between these limits were computed proportionally. Total scores were calculated simply by adding the scores assigned to each component (giving equal weight for each component), having an index with values ranged between 0 and 100. A total score greater than 80 was considered 'good', scores of 51–80 indicated 'needs improvement', and scores less than 51 were considered 'poor'.[26]

From 2000 onwards, the HEI was slightly modified to reflect changes in dietary guidelines.[27] HEI-2005 was developed following the release of the *Dietary Guidelines for Americans 2005* [28] and in response to the increased emphasis on important aspects of diet quality, such as whole grains, various types of vegetables, specific types of fat and the introduction of the new concept of 'discretionary calories'.[29] This updated version of the HEI consisted of 12 components – total fruit, whole fruit, total vegetables, dark green and orange vegetables and legumes, total grains, whole grains, milk, meat and beans, oils, saturated fat intake as a percentage of total energy intake, sodium intake as g/1,000 kcal and the calories from solid fat, alcohol and added sugar (SoFAAS) as a percentage of total energy intake. Individual nutrient intakes are first transformed into a base of 1,000 calories for diet groups one to nine and 11. For the nutrient intakes of the first nine diet groups, the intake of each group is compared with the corresponding recommended intake of that group. If the nutrient intake from a diet group, say total fruit, meets the recommended quantity, it will receive the maximum HEI score for that group. If the nutrient intake for the diet group is zero, that group gets a zero HEI score. Intakes between zero and the recommended quantity (maximum level) are scored proportionately. For the 11th diet group, sodium, the maximum score will be given if the sodium intake is less than the recommended amount. For diet groups such as saturated fat and SoFAAS, the HEI scores are based on the percentage of energy obtained from those groups relative to the total energy gained from food consumption. If the energy from saturated fat (the 10th diet group) is less than or equal to 7% of the total energy from food consumption, the saturated fat diet group receives the highest HEI score. If the energy from SoFAAS (the 12th diet group) is less than or equal to 20% of the total energy, the SoFAAS diet group gets the maximum HEI score (for more details, see Guenther *et al.*[29]). While the minimum HEI score for all diet groups is zero, the maximum HEI scores of different diet groups vary. The first six food groups receive the maximum HEI scores of five, the SoFAAS group receives a maximum score of 20, and the rest of the five diet groups receive maximum scores of 10. The total score (ranging from 0 to 100), similar to the original HEI, is simply the sum of all HEI individual scores, and can be used to assess the overall diet quality of the food.

Publication of the *Dietary Guidelines for Americans 2010* [30] provoked a second major update of the HEI.[31] The HEI-2010 holds several characteristics of the HEI-2005:

a) it has 12 components, most of them unchanged, comprising nine adequacy and three moderation components;

b) it uses a density approach to set standards, per 1,000 calories, or as a percentage of calories; and

c) it employs least restrictive standards.

On the other hand, the main changes to the index include:

a) the greens and beans component replaces dark green and orange vegetables and legumes;

b) specific choices from the protein group, reflected through adding seafood and plant proteins;

c) fatty acids, a ratio of polyunsaturated and monounsaturated to saturated fatty acids, replaces oils and saturated fat, to acknowledge the recommendation to replace saturated fat with monounsaturated and polyunsaturated fatty acids; and

d) a moderation component, refined grains, replaces the adequacy component, total grains, to assess overconsumption.

We used the same components of food groups and nutrients and cut-off points as the HEI-2010 (as the last available version of the HEI) in developing our new OS-HEI.

*Low association between diet quality indices and health outcomes*

There is a continuing need to examine the relationship between diet quality and health in the population. The traditional approach of investigating the association between single nutrients or food and the risk of related health outcomes is fraught with problems because of the complexity of people's diets, the possible correlations in nutrient intake and the possible interactions in the effect of several foods/nutrients. [32] It is widely accepted that individuals do not consume isolated nutrients or foods but complex combinations of foods, consisting of several nutrients and non-nutrients.[32] In response to such need, diet quality indices are progressively being used to examine epidemiological associations between dietary intake and nutrition-related health outcomes. [12]

An inverse association of healthful dietary patterns with all-cause mortality (the annual number of deaths in a given age group, per the population in that age group, usually expressed per 100,000) and cardiovascular disease (CVD) risk was reported in most studies. However, the magnitude of risk reduction was modest and was attenuated after control for confounders.[14] Diet quality scores were weakly associated with lowered risk of CVD in men[33] but were not associated with a reduced chronic disease risk in women.[34] Measures of overall diet quality have been associated with biomarkers of chronic disease risk and health outcomes.[35 36] However, large cohort studies have often shown conflicting results between diet quality scores and chronic disease.

A systematic literature review of 30 observational studies found that the significant association between a diet index score and Body Mass Index (BMI) and obesity were consistently negative.[37] However, some studies have failed to find similar relationships between diet quality and weight measures.[38 39] Guo *et al.*[40] examined the association between the HEI and obesity, using a cross-sectional analysis of data from 10,930 adults who participated in the third NHANES, finding that a low HEI score was associated with overweight and obesity. However, the overall effectiveness of these guidelines in disease prevention needs to be studied further. In a recent study, Marshall *et al.*[41] examined associations between diet quality and weight status in populations at risk of over-nutrition, by reviewing 26 studies, and found a significant relationship in only eight.[42 - 49] The relationship between diet quality and weight status seems to be inconsistent with studies finding negative and positive associations. In a study of French adults, higher diet quality scores were weakly associated with lower BMI and lower blood pressure for men only, but were not associated with plasma lipid profiles.[36] Asghari *et al.*[50] investigated the performances of the priori dietary pattern, including the Mediterranean Diet Score (MDS), the HEI-2005 and the Diet Quality Index-International (DQI-I), which were compared as simple indices for predicting overweight or obesity. No significant relationship between diet quality indices, obesity and abdominal obesity were found, indicating that the ability of diet quality indices to predict obesity and abdominal obesity depends on how well these indices correlate with changes in energy balance as the primary focus in obesity.

Several studies have examined the association between the HEI and morbidity. A weak association was detected between HEI scores and the risk of chronic disease, with the exception of cancer risk.[33 34] Moreover, while Kant *et al.*[51] reported that the HEI score was associated with obesity and biomarkers of CVD and diabetes, Fung *et al.*[52] at the same time, published a study indicating that the HEI score is not significantly correlated with any of the biomarkers for CVD. Reedy *et al.*[53] revealed that the HEI-2005 is also inversely correlated with colorectal cancer risk.

From the abovementioned literature, it can be observed that the already proposed indices are adequate tools concerning the evaluation of diet quality, but they have shown moderate predictive ability in relation to chronic diseases and health determinants such as obesity. [21] For most of the indices, association with disease or mortality were generally moderate, casting doubts on the validity of these indices. This therefore emphasizes the need for the new OS-HEI.

## Data: The National Health and Nutrition Examination Survey (NHANES)

The NHANES is a program of studies designed to assess the health and nutritional status of adults and children in the US. The survey is characterized by combining interviews and physical examinations. The NHANES program began in the early 1960s and has been conducted as a series of surveys focusing on different population groups or health topics. In 1999, the NHANES became a continuous program that has a changing focus on a variety of health and nutrition measurements to meet emerging needs. The survey examines a nationally representative sample of about 10,000 people each year.

The NHANES interview includes demographic-, socioeconomic-, dietary-, and health-related questions. The examination component consists of medical, dental, and physiological measurements, as well as laboratory tests administered by highly trained medical personnel.

Findings from this survey are used to determine the prevalence of major diseases and risk factors for diseases. Information is also used to assess nutritional status and its association with health promotion and disease prevention. NHANES findings are also the basis for national standards for such measurements as height, weight, and blood pressure. Data from this survey is used in epidemiological studies and health sciences research, which help develop sound public-health policy, direct and design health programs and services, and expand health knowledge for the nation.

The data on two-day food consumption and nutrient intakes for 2007–08 (for the development of OS-HEI) and 2009–10 (for the validation of OS-HEI) are obtained from the NHANES databases, including data from the Dietary Interview of Individual Foods (DIIF) and the Dietary Interview of Total Nutrient Intakes (DITN). The DIIF provides detailed information on the types (corresponding to USDA food codes) and amount (in grams) of food and beverages consumed by NHANES participants in two days. The

DITN provides information on individual nutrient intakes based on data from the DIIF and *USDA Food and Nutrient Database for Dietary Studies* (FNDDS) (USDA Agricultural Research Service.[54] The FNDDS provides information on nutrient information for each food listed in USDA food codes. The nutrient information helps transform individual food intake into nutrient intake. The total calorie intake from food consumption from the DITN is used to transform the nutrient intake from absolute amount into intake per 1,000 calories. The MyPyramid Equivalents Database[55] is used to transform individual food and nutrient intakes into cup or ounce equivalents of diet groups, corresponding to those in the *Dietary Guidelines for Americans 2010* [30], which helps calculate the HEIs of different diet groups.

## Development and validation of the Obesity-specific Healthy Eating Index (OS-HEI)

We adopted a semi-empirical approach in developing the novel OS-HEI, where we used the same food and nutrient groups as in HEI-2010, following the same cut-off points as the MyPyramid Equivalents Database guidelines, and then used the MARS model to determine which food groups should build up our index and the weighting of each group.

The process for developing the new OS-HEI can be summarized in the following steps:

- **Step 1**: Calculating individual diet scores for each participant in the NHANES 2007–08, using the HEI and OS-HEI – We calculated the scores for each component, and while these scores in the HEI-2010 ranged from 0 to 5, 10 or 20, our scores ranged from 0 to 100 for all components. We also followed the same methodology of calculating scores in the HEI-2010 (for details see Guenther *et al.*, 2010).

  **Step 2**: We ran the MARS model, using the OS-HEI individual scores as independent variables and BMI and waist circumference as dependent variables to calculate the relative importance of each component.

- **Step 3**: The total OS-HEI is calculated by multiplying each score by its relative importance (calculated in step 2) and the sum of all the resulting weighted scores have a total score ranging from 0 to 100.

In the following section we will explain the MARS model.

*Multivariate adaptive regression splines (MARS)*

Our paper is the first attempt to use the MARS model to develop diet quality indices. The MARS model is capable of overcoming the high level of subjectivity involved in the development of predefined diet quality indices. MARS also outperform traditional techniques, such as factor analysis and cluster analysis, which are normally used in developing empirical diet quality indices. Moreover, the MARS model is capable of taking into account the interactions between the different components of diet quality indices.

The MARS model, first introduced as a data mining tool,[56] is able to address the above limitations of factor analysis, cluster analysis and other classical methods. The MARS model is a nonparametric method; hence, it is expected to perform as well as, or even better than, the classical regression techniques when distributional assumptions are not satisfied. It also allows for local models and, thus, for a more accurate function approximation. The MARS model is not affected by any volume of missing data, since it automatically introduces indicator functions for every variable that contains missing values. Furthermore, this method is designed to capture higher-order interactions, even in high-dimensional settings. However, unlike other available nonparametric methods that can capture complex relationships among the variables, such as the classification and regression tree (CART) or artificial neural networks (ANNs), MARS produce very simple and easy to interpret models.

MARS performance depends on data structure, [57] but is generally known for predictive accuracy, computational speed and simplicity of interpretation. Leathwick *et al.*[58] compared general additive models (GAM) and MARS models and highlighted the advantages of MARS in cases involving large data sets. MARS models are also parsimonious and provide more extensive predictions. Muñoz and Felicísimo[59] used two different ecological data sets to compare MARS over other modeling techniques, such as multiple linear regression (MLR), principal component regression (PCR) and CART, observing that MARS performed consistently well. Using motor vehicle injury data consisting of 59 cases and 689 controls and with up to 3% missing values for some of the variables, Kuhnert *et al.*[60] showed that MARS outperformed CART and MLR, in terms of accuracy and flexibility as a modeling tool. Haughton and Loan[61] compared different statistical techniques to model vulnerability from a panel of 4,272 households. They showed that MARS, together with CART, were the most parsimonious models and were able to capture nonlinearities and interaction effects.

The main advantage of MARS, compared with other regressions, such as the logistic regression, is that the MARS model is a data-driven technique. Instead of fitting a single regression equation for the model, MARS can generate many piecewise regression equations, which allow the researcher to obtain more consistent and unbiased estimates of the covariates.

The main principle of MARS is based on searching for every point where linearity breaks. These cut-off points of the covariate, where the slope of the line changes, are called knots. The knot defines the end of one domain and the beginning of another. Between two knots, a linear (or cubic) regression line is fitted to that range of data. When the slope is not changing along the entire range, no knots are detected and a single linear regression is defined between the covariate and the dependent variable, as in the parametric approach. As mentioned, in MARS the data is left to reveal the variable knot locations, while the user need not input any specification into the model.

Based on knots detected in the process, basis functions are defined to re-express the relations between the dependent variable and its covariates. Basis functions in MARS, which serve as independent variables, are truncated linear functions, which address the problem of discontinuity of recursive partitioning algorithms. To model basis functions, MARS uses the so-called hinge functions or hockey-stick functions, which take the following expression:

$$(X - t_k)_+ = \begin{cases} X - t_k, & \text{if } \mathrm{X} \geq \mathrm{t_k}, \\ 0, & \text{else} \end{cases} \tag{1}$$

where $\mathrm{t_k}$ is a constant called knot.

In contrast to additive models, MARS allows interactions up to an order specified by the user, and trades off the interaction order and complexity of the additive functions and interactions.[62] [63] Piecewise linear functions can not only be formed from hinge functions, but they can be multiplied among them to form nonlinear functions.

The MARS model can be written as:

$$y = \sum_{i=1}^{M} \beta_i B_i(X) \tag{2}$$

where, $B_i$ (i = 1,2,…,M) are the basis functions and $\beta_i$ are the coefficients to be estimated.

MARS is a stepwise process that uses both forward and backward progressions for robust and unbiased parameter estimations. It starts by maximizing all possible effects of explanatory variables in the forward model and then removes the least effective functions in the backward model, using the ordinary least squares method, in order to minimize the so-called generalized cross validation (GCV) indicator,[64] given by:

$$\mathrm{GCV} = \frac{\frac{1}{\mathrm{N}} \sum_{i=1}^{N} \left[ y_i - \hat{f}_M(X_i) \right]^2}{\left[ 1 - \frac{d - M}{N} \right]^2} \tag{3}$$

where N is the number of observations, M is the number of basis functions in the model and $\hat{f}$ denotes the fitted values of the current MARS model. The numerator refers to the common residual sum of squares (RSS), that is penalized by the denominator, which accounts for the increasing variance as the model complexity increases. The penalizing parameter 'd' is chosen by the user, although the conventional value is d = 4. A lower (higher) value of d generates a model with more (fewer) basis functions. Thus, the GCV can be considered as a form of regularization by trading off goodness of fit against model complexity. In MARS models, the RSS cannot be used for comparing models, as the RSS always increases as MARS terms are dropped, which means that if the RSS were used to compare models, the backward step of model construction would always choose the largest model.

The main disadvantage of MARS is the low prediction power with insufficient sample size. This is not the case in our analysis, as we have a quite big data set which consists of 9,000 observations. Moreover, Briand *et al.*[65] mentioned that the model might suffer from multi-collinearities, as MARS picks-up interactions between predictive variables involved in the model. The MARS methodology also has a risk of over fitting, because of the very exhaustive search that is conducted to identify nonlinearities and interactions. This drawback could be controlled by choosing the appropriate penalty term of the model.

The importance of each explanatory variable is calculated as the square root of the GCV value of a sub-model from which all basis functions that involve this variable have been removed, minus the square root of the GCV value of the selected model.

**Results**

Table 1 summarizes the descriptive statistics for the individual scores of both the HEI-2010 and OS-HEI. For both indices the minimum for all scores is zero. The maximum for all components in the case of the OS-HEI is constant and equals 100, while in the case of HEI-2010, the maximum varies and ranges from 5, in the case of total vegetables, greens and beans, total fruit, whole fruit, total protein food, and seafood and plant protein; 10 for whole grain, dairy, fatty acids, refined grains, and sodium; the highest maximum of 20 assigned only for the empty calories component.

The lowest mean score is observed for greens and beans for both indices, with a value of 0.94 and 18.76 for the HEI-2010 and OS-HEI respectively. While the highest mean score is detected for empty calories, with a value of 10.97 in the case of the HEI-2010, the highest mean score of 76.31 corresponds to total protein food for the OS-HEI.

Table 1 summary statistics for individual scores of HEI-2010 and OS-HEI using NHANES database (2007-2008)

| Index Components | HEI-2010 | | | | OS-HEI | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Std. Dev. | Min | Max | Mean | Std. Dev. | Min | Max |
| Total vegetable | 2.46 | 1.71 | 0 | 5 | 52.85 | 34.21 | 0 | 100 |
| Greens and Beans | 0.94 | 1.81 | 0 | 5 | 18.76 | 36.12 | 0 | 100 |
| Total fruit | 2.49 | 2.12 | 0 | 5 | 49.83 | 42.47 | 0 | 100 |
| Whole fruit | 2.14 | 2.29 | 0 | 5 | 42.96 | 45.81 | 0 | 100 |
| Whole grain | 1.93 | 2.86 | 0 | 10 | 19.32 | 28.60 | 0 | 100 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Dairy | 5.29 | 3.61 | 0 | 10 | 52.88 | 36.15 | 0 | 100 |
| Total protein food | 3.82 | 1.57 | 0 | 5 | 76.31 | 31.50 | 0 | 100 |
| Seafood and plant protein | 1.55 | 2.05 | 0 | 5 | 30.95 | 40.98 | 0 | 100 |
| Fatty acids | 4.57 | 3.56 | 0 | 10 | 45.67 | 35.60 | 0 | 100 |
| Refined grains | 5.25 | 3.64 | 0 | 10 | 52.52 | 36.39 | 0 | 100 |
| Sodium | 5.59 | 3.69 | 0 | 10 | 55.93 | 36.94 | 0 | 100 |
| Empty Calories | 10.97 | 6.37 | 0 | 20 | 54.84 | 31.86 | 0 | 100 |

Table 2 shows the basis functions estimates of the MARS model, with BMI as a dependent variable and the 12 individual scores of the OS-HEI as explanatory variables. In order to keep the model as simple as possible, interactions of order 2 and a penalty term that equals 4 have been chosen.[1] As explained earlier, in the backward step, the best model is reached by minimizing the GCV. The optimal model consisted of 14 basis functions beside the intercept (adjusted $R^2 = 0.09$).

Results of the MARS model seem consistent with coefficients having the expected sign. For instance, the first base function indicates that for those with a high consumption of dairy, with a very high score greater than 99.7, increasing the consumption of dairy products also increases the BMI index, which indicates that the cut-off point for dairy consumption is quite suitable for analyzing obesity prevalence. While for most of the participants with a dairy score of less than 99.7, an increase of the individual quality score of dairy consumption by 10% resulted in 0.3 reduction in the BMI. The total protein quality does not affect the BMI for those with a score less than 83, while it has a positive effect on BMI for those with a higher total protein score. In the case of fruit, increasing the individual fruit quality index has a negative effect on BMI. It worth mention that while, protein consumption affects BMI in the case of high levels of protein consumption, fruit consumption affect BMI negatively even with low quantities.

[1] Higher interaction orders (3 and 4) and different penalty terms (2, 3, 5 and 6) were considered. No significant differences were found in terms of basis functions, knots and variable importance.

It was observed that all components have a combined effect, resulting from their interaction with other diet components as well as their sole effect on obesity prevalence. This emphasizes the importance of taking into account the interaction between the different components of the index.

It is worth mentioning that we re-estimated the same model using waist circumference as a dependent variable instead of BMI, with the aim of checking the consistency of our results using different measures of obesity. This model resulted in quite similar results to those of the BMI model, indicating consistency of the results, regardless of the way of measuring weight status.[2]

Table 2 Parameter estimates from the MARS model

| Bases functions | Coefficients | Variable Involved | Knot Value | Variable Involved | Knot Value |
|---|---|---|---|---|---|
| 0 | 22.5758 | | | | |
| 1 | -2.5663 | Total Dairy | 99.72 | | |
| 2 | 0.0327 | Total Dairy | 99.72 | | |
| 3 | 0.1033 | Total Protein | 83.08 | | |
| 4 | -0.0326 | Total Protein | 83.08 | | |
| 5 | -0.0256 | Total Fruit | 3.51 | | |
| 6 | 0.5278 | Total Fruit | 3.51 | | |
| 7 | 0.0002 | Total Vegetable | 10.93 | | |
| 8 | 0.0020 | Total Vegetable | 10.93 | Total Fruit | 3.51 |
| 9 | -0.0042 | Whole Fruit | 5.13 | Total Fruit | 3.51 |
| 10 | 0.0207 | Whole Grain | 9.73 | Total Dairy | 99.72 |
| 11 | 0.0918 | Whole Grain | 9.73 | | |
| 12 | 0.0152 | Total Vegetable | 0.00 | | |
| 13 | 0.0007 | SOFAAS | 31.23 | Total Protein | 83.08 |
| 14 | 0.0009 | SOFAAS | 31.23 | Total Protein | 83.08 |

Underlined cells indicate Basis functions of type max (0, independent-knot), otherwise max (0, knot-independent)

[2] Results of this model are available upon request from the authors.

The importance of each explanatory variable is calculated as the square root of the GCV value of a sub-model from which all basis functions that involve this variable have been removed, minus the square root of the GCV value of the selected model. Individual scores for dairy, total fruit, total protein and total vegetables are the most important in predicting the prevalence of obesity and overweight. The individual scores from the whole fruit, whole grains and empty calories groups also play a role in predicting obesity (Table 3). These results seem to be consistent with the related literature. Dairy has been shown to reduce risk for several chronic diseases, including osteoporosis, hypertension, obesity and Type 2 diabetes.[66 - 68] Authors and experts used to suggest that those who are obese, or at risk of obesity, should eat more fruit and vegetables.[69] The association of meat consumption with health can be described as U-shaped – in moderate quantities it is assumed to be beneficial, but its intake should not be too high, as high consumption levels are considered unfavorable.[3] Cereal products are not the only foods that provide dietary fiber. Moreover, the health effect of whole grains is not attributed to fiber alone, but also to the micronutrients, antioxidants and non-nutritive dietary constituents, such as phytoestrogens in wheat bran and beta-glucans in oats[70] – because of this, whole-grain products should be distinguished from refined grains in diet quality indices.

The other five individual scores seem to be irrelevant in predicting obesity. Some of these scores, such as sodium consumption, have no importance in predicting obesity because no direct link can be identified between sodium consumption and obesity prevalence. Other components have no importance in predicting obesity, although their consumption could affect obesity prevalence – such as refined grains, because their effects are overlooked by the greater effect of similar components such as whole grains. This highlights the fact that, in many cases, healthy foods are defined by the absence of problematic ingredients, such as fats, sugar, and sodium, rather than the presence of any beneficial nutrients they might contain.[6] Our results suggest that more attention should be given to beneficial nutrients.

Table 3 Variable importance for MARS model

| Index components | Number of Basis functions | Variable importance (%) |
|---|---|---|
| Total Dairy | 3 | 22 |
| Total Fruit | 4 | 21 |
| Total Protein | 4 | 20 |
| Total vegetables | 3 | 15 |
| Whole Fruit | 1 | 10 |
| Whole Grains | 2 | 7 |
| Empty Calories | 2 | 5 |

Each individual score was multiplied by the variable importance, to get the OS-HEI total score, which ranged from 0 to 100. Table 4 represents summary statistics for the total scores of the HEI-2010 and OS-HEI, BMI and waist circumference, using the NHANES database (2007–08). It can be observed that the mean value of the OS-HEI is 10 points greater than the HEI-2010, which indicates that the HEI-2010 is underestimating the diet quality in terms of obesity. In addition, a higher standard deviation was observed for the OS-HEI compared with the HEI-2010, which emphasizes that the OS-HEI was more capable of capturing variability in intakes of food and nutrients regarding obesity prevalence.[3]

Table 4 summary statistics for total scores of HEI-2010 and OS-HEI, BMI and waist circumference using NHANES database (2007-2008)

| Variable | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Total score (HEI-2010) | 47.18 | 13.90 | 0 | 100 |
| Total score (OS-HEI) | 53.68 | 18.46 | 0 | 100 |
| BMI | 25.72 | 7.58 | 12.50 | 73.43 |
| Waist circumference | 87.64 | 22.24 | 37.80 | 178.20 |

[3] Same differences in terms of both mean and standard deviation were detected using the 2009–10 NHANES database.

Table 5 presents weights for the different individual scores of the HEI-2010 and OS-HEI. It can be observed that while only four groups, which are dairy, total fruit, total protein and total vegetables, represent around 80% of the important components in the OS-HEI, these represent only 20% of the weight of the HEI-2010. On the other hand, the empty calories component has 20% importance in the HEI-2010 and only 5% in the OS-HEI. This emphasizes the importance of using MARS as an objective tool in weighting the different components.

Table 5 weights for individual scores of HEI-2010 and OS-HEI

| Index components | | HEI-2010 (%) | OS-HEI (%) |
|---|---|---|---|
| 1 | Total vegetable | 5 | 15 |
| 2 | Greens and Beans | 5 | 0 |
| 3 | Total fruit | 5 | 21 |
| 4 | Whole fruit | 5 | 10 |
| 5 | Whole grain | 10 | 7 |
| 6 | Dairy | 10 | 22 |
| 7 | Total protein food | 5 | 20 |
| 8 | Seafood and plant protein | 5 | 0 |
| 9 | Fatty acids | 10 | 0 |
| 10 | Refined grains | 10 | 0 |
| 11 | Sodium | 10 | 0 |
| 12 | Empty Calories | 20 | 5 |

*Validation of the OS-HEI*

We validated the OS-HEI using the 2009–10 NHANES database. We then scored each individual's diet using the NHANES (2009–10), to determine their HEI-2010 and OS-HEI scores. After that, we estimated the correlation between the BMI and waist circumference, as measures of obesity prevalence, and the HEI-2010 and OS-HEI as indices of diet quality. While no significant correlation was detected between the HEI-2010 and the two obesity prevalence measures, a significant negative correlation of -0.053 and - 0.066 between the OS-HEI and BMI and waist circumference, respectively, indicated that increasing diet quality, as measured by the OS-HEI, by 10% could decrease obesity prevalence by approximately 0.5%. This low magnitude of the correlation coefficient is to be expected, as many other factors besides diet quality play a role in obesity prevalence.

Table 6 correlation and significance estimates between obesity prevalence (BMI and waist circumference) and diet quality measured by HEI-2010 and OS-HEI

| Variables | BMI | | Waist Circumference | |
|---|---|---|---|---|
| | Correlation | significance | Correlation | significance |
| HEI-2010 | -0.003 | 0.748 | 0.005 | 0.640 |
| OS-HEI | -0.053 | 0.000 | -0.066 | 0.000 |

We then estimated a MARS model, using the HEI or OS-HEI as the predictor of BMI or waist circumference, to determine the correlation between the HEI or OS-HEI and obesity prevalence. The HEI-2010 failed in predicting the obesity prevalence, with an adjusted $R^2$ equal to 0.000, while the OS-HEI was capable of predicting obesity prevalence in both cases, with an adjusted $R^2$ equal to 0.01. This low value of the adjusted $R^2$ is to be expected, as many other factors besides diet quality play a vital role in obesity prevalence. For instance controlling for age and gender, which are considered the most important determinants in obesity prevalence,[71] increases adjusted $R^2$ to around 0.40. This result emphasizes the importance of designing gender-specific and age-group-specific diet quality indices, which is one of our future research lines.

Table 7 Goodness of fit estimates for the different MARS models

| Explanatory variable | BMI | | Waist Circumference | |
|---|---|---|---|---|
| | GCV | Adjusted $R^2$ | F value | Adjusted $R^2$ |
| HEI-2010 | 59.53 | 0.00 | 492.53 | 0.00 |
| OS-HEI | 59.10 | 0.01 | 487.35 | 0.01 |
| OS-HEI controlling for Age and Gender | 36.12 | 0.39 | 200.65 | 0.59 |

Although it can be argued that total calorie intake could be used instead of the OS-HEI, single nutrient or food group analysis omits the synergistic nature of whole diet. Knowledge was shown to be a stronger predictor of overall diet quality than of any single nutrient or diet quality – keeping with the doctrine that there are no good or bad foods, only bad diets.[72] So, measures of nutritional quality should focus on total diets only not on single foods or nutrients. A holistic approach would be more realistic, as people have diets, they do not just consume nutrients but combinations of foods.

Furthermore, calorie labeling has been found to be inefficient in reducing food consumption and enhancing diet quality in many studies.[73 74]

Moreover, we estimated the MARS model, with the 12 individual scores and total calorie intake, with the aim of constructing an energy-adjusted, diet quality index. Surprisingly, our results indicated that total calorie intake has a significant effect on obesity prevalence, while empty calories become insignificant. In addition, the importance of total calories in this model and that of the empty calorie in the former are quite similar, which indicates that empty calorie component is capturing the total calorie intake of individuals. To keep our OS-HEI comparable with the HEI-2010, we decided not to include the total calorie intake in our model, as its effect is already reflected by the empty calories component.[4]

[4] Results of this model are available from the author upon request.

## Concluding remarks

In this paper, we have described how we have developed a new OS-HEI. Data from the NHANES data set for the year 2007–08 was used for the development of our new index, and the NHANES data set for the year 2009–10 was used to validate the index and evaluate its ability to predict the effect of diet quality on obesity prevalence. In order to avoid the shortcomings of previous diet quality indices, we followed a semi-empirical approach, using the MARS to develop the OS-HEI and avoid subjectivity in choosing food groups included in the index, their weightings and cut-off points. A high association was found between the new OS-HEI and obesity prevalence. Moreover, our new OS-HEI notably outperformed the HEI-2010 and total calories consumed in predicting obesity prevalence. While the HEI-2010 includes 12 components, the OS-HEI includes only seven components, as seafood and plant protein, fatty acids, refined grains and sodium were assumed not to have an effect on obesity prevalence. While the weighting of food groups in the HEI-2010 equals 5, 10 or 20, food group weightings in the OS-HEI ranged from 5% for empty calories to 22% for dairy. This study has provided an initial look at the development and validation of disease- and obesity-specific diet quality indices, offering an OS-HEI capable of predicting obesity prevalence. Moreover, it provides a flexible methodological framework to develop other disease-specific diet quality indices, avoiding subjectivity in doing so. The capability of the OS-HEI in predicting obesity increased significantly by also considering age and gender – this raised the need to develop gender- and age-group-specific healthy-eating indices, especially in the case of obesity.

## References


1- WHO (2013). Fact sheet No 311. [WWW document]. URL http://www.who.int/mediacentre/factsheets/fs311/en/ [accessed 14 October 2013].

2- Bandini, L. G., Vu, D., Must, A., Cyr, H., Goldberg, A., Dietz, W.H. Comparison of high-calorie, low-nutrient-dense food consumption among obese and non-obese adolescents. Obesity Research 1999; 7: 438–443.

3- Waijers , P., Feskens, E., Ocke, M. A critical review of predefined diet quality scores. Br J Nutr. 2007; 97: 219–231.

4- Ruel, M.T., 2002. Is dietary diversity an indicator of food security or dietary quality? A review of measurement issues and research needs. FCND Discussion Paper 140. Washington DC: International Food Policy Research Institute.

5- WHO/FAO (World Health Organization/Food and Agriculture Organization of the United Nations). 1996. Preparation and use of food-based dietary guidelines. Geneva: Nutrition Programme, World Health Organization.

6- Drewnowski, A. Concept of a nutritious food: toward a nutrient density score. Am J Clin Nutr. 2005 ; 82(4) :721–32.

7- Guthrie, H. Concept of a nutritious food. Jam Diet Assoc. 1977; 71: 14–19.

8- Lackey, C.J., Kolasa, K.M. Healthy eating: defining the nutrient quality of foods. Nutr Today 2004; 39: 26 –9.

9- Panagiotakos, D. Health measurement scales: methodological issues. Open Cardiovasc Med J. 2009; 3: 160–165

10- Kant, A. Indexes of overall diet quality. J Am Diet Assoc. 1996; 96: 785–791.

11- Gerber, M. The comprehensive approach to diet: a critical review. J. Nutr. 2001; 131(suppl. 11): 3051S–3055S.

12- Wirt, A., Collins, C.E. Diet quality—What is it and does it matter? . Public Health Nutr. 2009; 12: 2473–2492.

13- Hu, F.B. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol. 2002; 13: 3–9.

14- Kant, A. Dietary patterns and health outcomes. J Am Diet Assoc. 2004; 104: 615–635.

15- Kennedy, E.T., Ohls, J., Carlson, S., Fleming, K. The Healthy Eating Index: design and applications. J Am Diet Assoc. 1995; 95: 1103–8.

16- Seymour, J., Calle, E., Flagg, E., Coates, R., Ford, E., Thun, M. Diet quality index as a predictor of short-term mortality in the American Cancer Society cancer prevention study II nutrition cohort. Am J Epidemiol. 2003; 157: 980–988.

17- Huijbregts, P., Feskens, E., Rasanen, L., Fidanza, F., Nissinen, A., Menotti ,A. and Kromhout, D. Dietary pattern and 20 year mortality in elderly men in Finland, Italy, and the Netherlands: longitudinal cohort study. BMJ. 1997; 315: 13–17.

18- Trichopoulou, A., Kouris-Blazos, A., Wahlqvist, M., Gnardellis, C., Lagiou ,P., Polychronopoulos, E., Vassilakou, T., Lipworth, L., Trichopoulos, D. Diet and survival in elderly people. BMJ. 1995; 311: 1457–1460.

19- Nicklas, J.M., Huskey, K.W., Davis, R.B., Wee, C.C. Successful weight loss among obese U.S. adults. Am J Prev Med. 2012; 42(5): 481-5.

20- Kourlaba, G., Kondaki, K., Grammatikaki, E., Roma-Giannikou, E., Manios, Y. Diet quality of preschool children and maternal perceptions/misperceptions: the GENESIS study. Public Health 2009; 123: 738–742.

21- Waijers PMCM , Feskens EJM. Indexes of overall diet quality - A review of the literature. 2005 RIVM Report 350010003.

22- MacQueen, J. B. (1967). 'Some methods for classification and analysis of multivariate observations'. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability 1. University of California Press. 281–97.

23- Ward, J. H., Jr. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 1963; 58: 236–44.

24- Chen H, Ward MH, Graubard BI, Heineman EF, Markin RM, Potischman NA, Russell RM, Weisenburger DD, Tucker KL. Dietary patterns and adenocarcinoma of the esophagus and distal stomach. Am J Clin Nutr. 2005; 75: 137-144.

25- USDA & USDHHS (1995) *Dietary Guidelines for Americans 1995*

26- U.S. Department of Agriculture (USDA), Center for Nutrition Policy and Promotion (CNPP). (1996 sl. rev.). The Food Guide Pyramid (Home and Garden Bulletin Number 252).

27- Basiotis, PP., Carlson, S., Gerrior, S.A., Juan, W.Y., Lino, M., 2002. The Healthy Eating Index: 1999–2000.

28- USDA & USDHHS. (2005). *Dietary Guidelines for Americans 2005*

29- Guenther, P.M., Reedy, J., Krebs-Smith, S.M., Reeve, B.B., Basiotis, P.P., 2007. Development and Evaluation of the Healthy Eating Index-2005: Technical Report. Center for Nutrition Policy and Promotion, U.S. Department of Agriculture. Available at http://www.cnpp.usda.gov/HealthyEatingIndex.htm

30- USDA & USDHHS. (2010). *Dietary Guidelines for Americans 2010*

31- Guenther, P.M., Casavale, K.O., Reedy, J., Kirkpatrick, S.I., Hiza, H.A., Kuczynski, K.J., Kahle, L.L., Krebs-Smith, S.M. Update of the Healthy Eating Index: HEI-2010. J Acad Nutr Diet 2013;113(4): 569–580.

32- Mertz, W. Foods and nutrients. J Am Diet Assoc. 1984; 84: 769–70.

33- McCullough, M.L., Feskanich, D., Stampfer, M.J., et al. Adherence to the Dietary Guidelines for Americans and risk of major chronic disease in women. Am J Clin Nutr. 2000; 72: 1214–22.

34- McCullough, M.L., Feskanich, D., Rimm, E.B., et al. Adherence to the Dietary Guidelines for Americans and risk of major chronic disease in men. Am J Clin Nutr. 2000; 72: 1223–31.

35- Zamora, D., Gordon-Larsen, P., He, K., et al. Are the 2005 Dietary Guidelines for Americans Associated With reduced risk of type 2 diabetes and cardiometabolic risk factors? Twenty-year findings from the CARDIA study. Diabetes Care 2011; 34(5): 1183-1185.

36- Drewnowski , A., Fiddler, E.C., Dauchet, L., et al. Diet quality measures and cardiovascular risk factors in France: applying the Healthy Eating Index to the SU.VI.MAX study. J Am Coll Nutr. 2009; 28(1): 22-29.

37- Togo, P., Osler, M., Sorensen, T.I., et al. Food intake patterns and body mass index in observational studies. Int J Obes Relat Metab Disord. 2001; 25(12): 1741- 1751.

38- Quatromoni, P.A., Copenhafer, D.L., D’Agostino, R.B., et al. Dietary patterns predict the development of overweight in women: The Framingham Nutrition Studies. J Am Diet Assoc. 2002; 102(9): 1239-1246.

39- Villegas, R., Salim, A., Collins, M.M., et al. Dietary patterns in middle-aged Irish men and women defined by cluster analysis. Public Health Nutr. 2004; 7(8): 1017-1024.

40- Guo, S. S., Huang, C., Maynard, L. M., Demerath, E., Towne, B., and Chumlea, W. C. Body Mass Index during Childhood, Adolescence and Young Adulthood in Relation to Adult Overweight and Adiposity: The Fels LongitudinalStudy. International Journal of Obesity and Related Metabolic Disorder 2000; 24: 1628-1635.

41- Marshall S.J., Biddle S.J.H., Gorely T., Cameron N. and Murdey, I. Relationships between media use, body fatness and physical activity in children and youth: a meta-analysis. International Journal of Obesity 2004; 28: 1238–1246.

42- Feskanich, D., Ma, J., Fuchs, C.S., et al. Plasma vitamin D metabolites and Risk of colorectal cancer in women. Cancer Epidemiol Biomarkers Prev. 2004; 13: 1502– 8.

43- Mirmiran, P., Azadbakht, L., Esmaillzadeh, A. and Azizi, F. Dietary diversity score in adolescents- a good indicator of the nutritional adequacy of diets: Tehran lipid and glucose study. Asia Pacific Journal of Clinical Nutrition 2004; 13(1): 56-60.

44- Kranz, S., Findeis, J.L., Sundar, S., Shrestha, S. Use of the Revised Children’s Diet Quality Index to assess preschooler’s diet quality, its sociodemographic predictors, and its association with body weight status. J Pediatr (Rio J). 2008; 84(1): 26-34.

45- Chiplonkar, S.A., Tupe, R. Development of a diet quality index with special reference to micronutrient adequacy for adolescent girls consuming a lacto-vegetarian diet. J Am Diet Assoc. 2010; 110(6): 926-31.

46- Kontogianni, M.D., Farmaki, A.E., Vidra, N., Sofrona, S., Magkanari, F., Yannakoulia, M. Associations between lifestyle patterns and body mass index in a sample of Greek children and adolescents. J Am Diet Assoc. 2010; 110(2): 215–221.

47- Manios, Y., Kourlaba, G., Grammatikaki, E., Androutsos, O., et al. Comparison of two methods for identifying dietary patterns associated with obesity in preschool children: The GENESIS study. European Journal of Clinical Nutrition 2010; 64(12 ): 1407-1414.

48- Golley, R.K., Hendrie, G.A., McNaughton, S.A. Scores on the dietary guideline index for children and adolescents are associated with nutrient intake and socioeconomic position but not adiposity. J Nutr. 2011; 141(7): 1340-1347.

49- Jennings, C.L., Lambert, E.V., Collins, M., Joffe, Y., Levitt, N.S., Goedecke, J.H. Determinants of insulin-resistant phenotypes in normalweight and obese Black African women. Obesity 2011; 16: 1602–1609.

50- Asghari G, Mirmiran P, Rashidkhani B, Asghari-Jafarabadi M, Mehran M, Azizi F. The association between diet quality indices and obesity: Tehran Lipid and Glucose Study. Archives of Iranian Medicine. 2012; 15: 599 – 605.

51- Kant, A. K. and Graubard, B. I. Energy density of diets reported by American adults: association with food group intake, nutrient intake, and body weight. International Journal of Obesity 2005; 29: 950-956.

52- Fung, T.T., Rexrode, K.M., Mantzoros, C.S., Manson, J.E., Willett, W.C., Hu, F.B. Mediterranean diet and incidence of and mortality from coronary heart disease and stroke in women. Circulation 2009; 119*:* 1093 –1100.

53- Reedy, J., Mitrou, P.N., Krebs-Smith, S.M., et al. Index-based dietary patterns and risk of colorectal cancer: the NIH-AARP Diet and Health Study. Am J Epidemiol. 2008; 168: 38–48.

54- USDA-ARS (2006) USDA Food and Nutrient Database for Dietary Studies 2005–2006. Available at:

http://www.ars.usda.gov/News/docs.htm?docid=12068 [accessed December 5, 2014].

55- Bowman, S.A., Gortmaker, S.L., Ebbeling, C.B., Pereira, M.A., Ludwig, D.S. Effects of fast-food consumption on energy intake and diet quality among children in a national household survey. Pediatrics 2004; 113: 112–118.

56- Frieadman, J.H. Multivariate additive regression splines. Ann. Of Stat. 1991; 19: 1-67.

57- Ture, M., Kurt, I., Kurum, A. T., Ozdamar, K. Comparing Classification Techniques for Predicting Essential Hypertension. Expert Systems with App. 2005; 29: 583-588.

58- Leathwick, J. R., Elith, J., Hastie, T. Comparative Performance of Generalized Additive Models and Multivariate Adaptive Regression Splines for Statistical Modeling of Species Distributions. Ecol. Mod. 2006; 199: 188-196.

59- Muñoz, J., Felicísimo, A. M. Comparison of Statistical Methods Commonly Used in Predictive Modeling. J. of Vegetation Science 2004; 15: 285-292.

60- Kuhnert, P. M., Do, K., McClure, R. Combining Non-Parametric Models with Logistic Regression: An Application to Motor Vehicle Injury Data. Comp. Stat. and Data Analysis 2000; 34: 371-386.

61- Haughton, D., Loan, L.T.T., 2004. Vulnerability of Vietnamese Households, 1992-1998. Working Paper (June 2004). Ford Foundation/General Statistics Office Project, Vietnam.

62- Frank, I. E. Tutorial: Modern Non-linear Regression Methods. Chemometrics and Intelligent Lab. Sys. 1995; 27: 1-19.

63- De Veaux, R. D., Psichogios, D. C., Ungar, L. H. A Comparison of Two Nonparametric Estimation Schemes: MARS and Neural Networks. Computers and Chem. Engineering 1993; 17(8): 819-837.

64- Kayri, M. Two-step clustering analysis in researches: A case study. Eurasian Journal of Educational Research 2007; 7(28): 89-99.

65-Briand, L.C., Freimut, B., Vollei, F., 2007. Using multiple adaptive regression splines to understand trends in inspection data and identify optimal inspection rates. ISERN TR 00-07 2007. Avaliable from: http://www.salfordsystems.com.

66-Pereira, M.A., Jacobs, D.R. , Van Horn, L., Slattery, M.L., Kartashov, A.I., Ludwig, D.S. Dairy consumption, obesity, and the insulin resistance syndrome in young adults: the CARDIA Study. JAMA. 2002 ; 287(16) : 2081-9.

67-Zemel, M.B., Miller, S.L. Dietary calcium and dairy mudulation of adiposity and obesity risk. Nutrition Reviews 2004; 62: 125-131.

68-Choi ,H.K., Atkinson, K., Karlson, E.W., Curhan, G. Obesity, weight change, hypertension, diuretic use, and risk of gout in men: the health professionals follow-up study. Arch Intern Med. 2005; 165(7): 742-8.

69-Khan, L.K., Sobush, K., Keener, D., Goodman, K., Lowry, A., Kakietek, J., et al. Centers for Disease Control and Prevention. Recommended community strategies and measurements to prevent obesity in the United States. MMWR Recomm Rep. 2009; 58: 1 – 26.

70-King, L. The role of health promotion: between global thinking and local action. Health Promotion Journal of Australia 2006; 17(3): 196-9.

71-RADWAN, A. and GIL, J.M. 2011. Parametric and Non-Parametric Analysis of the Role of Economic Factors on Obesity Prevalence in Spain. The XIIIth EAAE Congress, Zurich, Switzerland, August 30-September 2.

72-Guthrie, H.A. There’s no such thing as “junk food,” but there are junk diets. Health line 1986; 5: 11–12.

73-Ellison, B., Jayson L., and David, D. Looking at the Label and Beyond: The Effects of Calorie Labels, Health Consciousness, and Demographics on Caloric Intake in Restaurants. International Journal of Behavioral Nutrition and Physical Activity 2013; 10-21.

74-Elbel, B., Kersh, R., Brescoll, V.L., Dixon, L.B. Calorie labeling and food choices: a first look at the effects on low-income people in New York City. Health Affairs 2009; 28: 1110 – 1121.