Development and evaluation of the US Healthy Food Diversity index

Maya Vadiveloo[1], L. Beth Dixon[2], Tod Mijanovich[3], Brian Elbel[4,5] and Niyati Parekh[4,6]*

[1]*Department of Nutrition, Harvard School of Public Health, 655 Huntington Avenue, Building 2, Room 347, Boston, MA 02115, USA*

[2]*Department of Public Health, Food Studies and Nutrition, David B. Falk College of Sport and Human Dynamics Syracuse University, 426 Ostrom Avenue, Syracuse, NY 13244, USA*

[3]*Department of Humanities and Social Sciences in the Professions, Steinhardt School of Culture, Education, and Human Development, New York University, 411 Lafayette Street, 5th Floor, New York, NY 10003, USA*

[4]*Department of Population Health, NYU School of Medicine, New York University, 227 East 30th Street, 6th Floor, New York, NY 10010, USA*

[5]*NYU Robert F. Wagner Graduate School of Public Service, 295 Lafayette Street, New York, NY 10012, USA*

[6]*Department of Nutrition, Food Studies and Public Health, New York University, 411 Lafayette Street, 5th Floor, New York, NY 10003, USA*

Abstract

Varied diets are diverse with respect to diet quality, and existing dietary variety indices do not capture this heterogeneity. We developed and evaluated the multidimensional US Healthy Food Diversity (HFD) index, which measures dietary variety, dietary quality and proportionality according to the 2010 Dietary Guidelines for Americans (DGA). In the present study, two 24 h dietary recalls from the 2003–6 National Health and Nutrition Examination Survey (NHANES) were used to estimate the intake of twenty-six food groups and health weights for each food group were informed by the 2010 DGA. The US HFD index can range between 0 (poor) and $1 - 1/n$, where n is the number of foods; the score is maximised by consuming a variety of foods in proportions recommended by the 2010 DGA. Energy-adjusted Pearson's correlations were computed between the US HFD index and each food group and the probability of adequacy for fifteen nutrients. Linear regression was run to test whether the index differentiated between subpopulations with differences in dietary quality commonly reported in the literature. The observed mean index score was 0·36, indicating that participants did not consume a variety of healthful foods. The index positively correlated with nutrient-dense foods including whole grains, fruits, orange vegetables and low-fat dairy (r 0·12 to 0·64) and negatively correlated with added sugars and lean meats (r −0·14 to −0·23). The index also positively correlated with the mean probability of nutrient adequacy (r 0·41; $P<0·0001$) and identified non-smokers, women and older adults as subpopulations with better dietary qualities. The US HFD index may be used to inform national dietary guidance and investigate whether healthful dietary variety promotes weight control.

Key words: Dietary variety: Healthy Food Diversity index: Healthy food variety: Dietary quality: Diet index

The recommendation to consume a wide number of foods and food groups was a distinct pillar of the US national dietary guidance until concern arose about potential adverse associations between dietary variety and obesity[1]. Historically, individuals who ate a greater variety of foods were more likely to consume adequate nutrients and reduce their mortality and the risk of chronic disease[2–5]. Because of these benefits, dietary variety recommendations have been present since 1916 in the USA[6,7], and dietary variety was emphasised in the first four enumerations of the Dietary Guidelines for Americans (DGA) (1980–1995)[8].

However, over the past 15 years, including dietary variety as a distinct guideline has become controversial because evidence suggests that consuming a wider assortment of foods may be related to excess energy intake and adiposity[9,10]. This concern became particularly salient during this time period because approximately 12 000 new food products or product extensions were introduced into the marketplace each year, and variety

within many of these foods may have promoted overconsumption and exacerbated the current obesity epidemic[11–14]. In response, the concept of dietary variety became embedded within other dietary recommendations in the DGA rather than being included as a separate recommendation[10].

Greater variety within the current food landscape may encourage people to eat beyond requirements because heterogeneity in food choice can stimulate appetite by enhancing the pleasure associated with eating[15]. However, in theory, a dietary pattern comprising a variety of healthful, low-energy-dense foods may encourage long-term weight control by enhancing the pleasure associated with eating lower-energy, nutrient-dense foods[12]. Moreover, because diets with limited variety often become tedious and unsustainable over longer durations, restricting variety may inadvertently discourage long-term weight control[16].

Diets with greater variety have been shown to promote energy intake and adiposity in short-term experimental studies[15], but are inconclusive in epidemiological studies[10]. Existing evidence suggests that dietary variety within nutrient-dense foods (e.g. fruits, vegetables, whole grains) encourages health[2], but it is unclear whether these benefits persist when diets are also influenced by the diversity of less healthful foods in the marketplace. It is also difficult to study this question comprehensively because of limitations in methods used to measure dietary variety in existing research. Dietary variety has been assessed by summing the number of distinct foods and the number of major and minor food groups[6,17], and by examining variety only within healthful or less healthful food groups[4,18,19]. Existing dietary variety indices insufficiently consider dietary quality (i.e. concordance with dietary guidelines) and proportionality (i.e. the proportional distribution of foods and food groups in the diet), which are vital inputs to energy intake and health in an obesogenic food environment[20,21]. Although dietary variety is a component included in some diet quality indices, these indices often dichotomise measures of dietary variety (i.e. 'met' v. 'not met'), which can obscure the associations between dietary variety and health[1]. Proportionality is also a crucial input because variety within less healthful foods may only be detrimental when consumed in excess, and healthful food variety may only be beneficial when consumed in adequate quantities. As a result, measurement error and bias within incomplete existing methods make it necessary to reconsider the tools that we use to assess dietary variety. Furthermore, the absence of a multidimensional index to simultaneously measure dietary variety, dietary quality and proportionality limits our ability to investigate whether promoting variety among nutrient-dense foods increases consumption of these foods and assists with weight control by making healthier dietary patterns more sustainable.

Recently, Drescher et al.[22] developed the Healthy Food Diversity (HFD) index to measure these three components in relation to German dietary guidelines, which provided an important framework for developing a similar index within a US population. The objective of the present study was to develop the US HFD index to measure dietary variety (defined as the number of foods), dietary quality (defined as

concordance with the 2010 DGA) and proportionality (defined as the distribution or balance of food groups in the diet). The US HFD index was designed to approach its maximum with consumption of a greater (v. fewer) number of foods, with greater consumption of more healthful (v. less healthful) food groups, and with consumption of food groups within the proportions recommended by the 2010 DGA (v. equal distributions of all foods).

Herein, we describe the theory and methodology underlying the development of the US HFD index, and evaluate the validity of the index in a representative sample of the US population using data from the 2003–6 National Health and Nutrition Examination Survey (NHANES). The index that we have developed can ultimately be used to evaluate the associations between dietary variety and health outcomes, especially those related to body weight and obesity.

Methods

Overview

We developed the US HFD index consistent with the validated German algorithm[22] and based on the 2010 DGA[21]. We first describe the process for developing health weights and for computing the index. Next, we discuss how we evaluated the US HFD index using the measures of validity used previously to evaluate dietary pattern indices[23]. Our evaluation included the following procedures: (1) confirming that the dietary quality and proportionality components reflected the most recent dietary guidance; (2) examining the energy-adjusted correlation between each food group and the US HFD index to evaluate the influence of each food group on the total score; (3) examining the energy-adjusted correlation between the US HFD index and the probability of adequacy for fifteen nutrients and with a pre-existing diet quality score (i.e. the Dietary Approaches to Stop Hypertension (DASH) diet score) to assess the validity of the dietary quality component; (4) examining whether the US HFD index varied between subpopulations with differences in dietary quality shown in the literature[23–26] and how it compared with existing variety indices.

Population description

The US HFD index was developed using the continuous NHANES 2003–6. The NHANES is a multi-stage, nationally representative survey of non-institutionalised US civilians over the age of 2 years[27]. Approximately 10 000 individuals were surveyed in each 2-year sampling cycle using a multi-stage, area-cluster design with differential selection probabilities for designated demographic groups[28].

The analytic sample was limited to non-pregnant, non-lactating adults aged 20 years and older with complete dietary data on 2 d (n 7585), as shown in Fig. 1. Pregnant and lactating individuals were excluded (n 587) because of differing nutrient needs[29]. Consistent with existing NHANES studies, the sample was further restricted to adults with energy intakes between 1674 and 29 288 kJ/d (400 and 7000 kcal/d) (n 7470)[30,31]. Self-reported demographic variables such as age, race, sex,
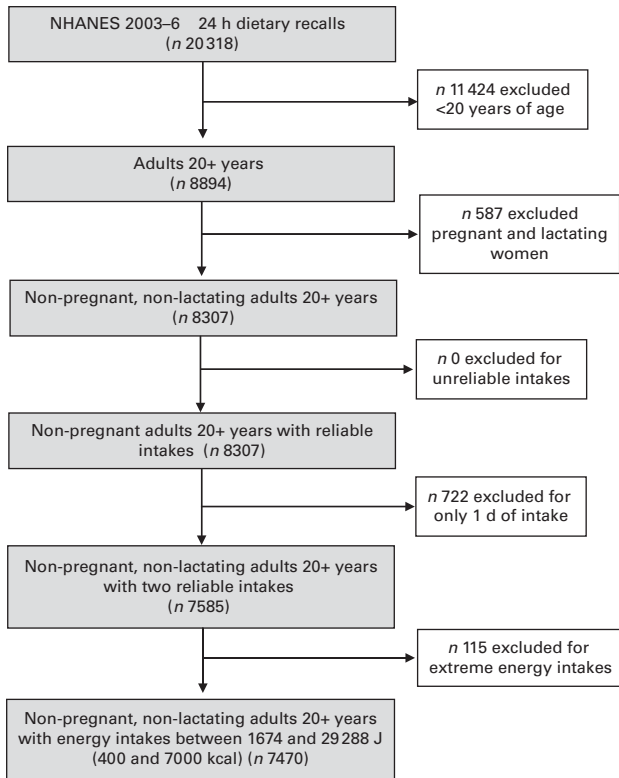
Fig. 1. Analytic dataset of the 2003–6 National Health and Nutrition Examination Survey (NHANES).

smoking status, income and education level were included in the analytic dataset[32,33]. We restricted analyses to the 2003–6 NHANES because FFQ propensity data were only available among a subsample of participants from 2003–6 (*n* 6001).

Assessment of diet

Participants provided 24 h dietary recall data to trained interviewers using the US Department of Agriculture (USDA) Automated Multiple Pass Method, which is a validated and reliable method for measuring food intake[34,35]. Food models and regular household measures were provided to improve the accuracy of portion size estimates. A second 24 h dietary recall by phone was scheduled for 3–10 d later for all participants, and approximately 87–90 % of the participants provided recalls that were considered complete and reliable by survey staff on both recall days[34,35]. The FFQ used in the NHANES was a 124-item questionnaire based on the National Cancer Institute Diet History Questionnaire that can be used to better estimate the probability of food consumption for food groups consumed infrequently[36,37].

Development of the US Healthy Food Diversity index

The US HFD index developed in the present study was based on the German HFD index created by Drescher *et al.*[22]. The equation used to generate the US HFD index is a modification of the Berry Index (BI)[38] used in economics studies, which measures the number and proportion of foods

$(BI = (1 - \sum s_i^2)$, where s is the share or proportion of each individual food i by weight)[22].

To assess dietary quality, a health factor (hf) for each food group was established based on its recommended proportion in the diet in accordance with the US Dietary Guidelines. The health value (hv) of the diet was calculated by multiplying the reported share by the volume of each food by its respective hf and summing it in order to capture both diet quality and proportionality. The overall US HFD index was calculated by multiplying the hv of the diet by the BI. Because the maximum hv that can be achieved is 0·25, the US HFD index was calibrated by dividing it by its maximum.

The index encourages proportionality by penalising consumption of a single high-quality food group or equal consumption of all food groups, which ensures that neither a high hv nor a high BI can independently generate a high US HFD index. The score ranged between 0 (representing a diet with a single food) and nearly 1 (representing a diet with many foods and with a higher proportion of foods from more heavily weighted food groups). The equation used to generate US HFD index scores is given by:

$$US\,HFD\,index = (1 - \sum s_i^2) \times hv,$$

where s_i is the share or proportion of each individual food or food group i based on the volume of the total diet, and

$$hv = \sum hf_i \times s_i.$$

Development of dietary quality and proportionality components

hf were based on both quantitative and qualitative guidelines enumerated in the 2010 DGA as well as the Report of the Dietary Guidelines Advisory Committee (DGAC). Intake recommendations for each of the twenty-six food groups were derived from the 2000-kcal USDA Food Pattern included in the 2010 DGA[21], and qualitative recommendations were based on the committee report[39].

To ensure that food group recommendations were in standardised units, approximate cup measurements were used to convert food groups listed with ounce recommendations to cup recommendations. We used the USDA National Nutrient Database for Standard Reference[40] to determine the best factors for converting ounces, teaspoons and grams into cup measures. There was less variability in energy (J) associated with converting ounces of meat and other protein foods to volume measures than with converting volume to weight. Consequently, we judged that using volumes rather than weights best maintained the proportionality distribution of food groups described in the 2000-kcal USDA Food Pattern from the 2010 DGA.

hf were generated by multiplying the recommended proportion of the twenty-six food groups derived from the 2000-kcal USDA Food Pattern by subjective values quantified by the authors from the DGAC report (Table 1). The DGAC report 'emphasises' the intake of plant-based foods, low-fat dairy products, seafood and oils, 'includes' animal products and refined grains, and 'limits' intake of discretionary

Table 1. Development of health factors (hf) for each food group using qualitative and quantitative recommendations for daily food group intakes based on the 2000-kcal US Department of Agriculture (USDA) Food Pattern in the 2010 Dietary Guidelines for Americans

Food groups	Recommended amount (cups)	Share of food subgroup (cup)	Broad food share × share of food subgroup	hf
Emphasises (0·78)*	9·49/12·15			
Whole grains†	1·5	1·5/9·49 = 0·16	0·78 × 0·16	0·12
Low-fat milk‡	3	3/9·49 = 0·32	0·78 × 0·32	0·25
Vegetables	2·5	2·5/9·49 = 0·26		
Dark green vegetables	0·2	0·2/2·5 = 0·08	0·78 × 0·26 × 0·08	0·02
Red and orange vegetables§	0·8	0·8/2·5 = 0·32	0·78 × 0·26 × 0·32	0·07
Legumes	0·2	0·2/2·5 = 0·08	0·78 × 0·26 × 0·08	0·02
Starchy vegetables‖	0·7	0·7/2·5 = 0·28	0·78 × 0·26 × 0·28	0·06
Other vegetables	0·6	0·6/2·5 = 0·24	0·78 × 0·26 × 0·24	0·05
Fruits¶	2	2/9·49 = 0·21	0·78 × 0·21	0·16
Nuts, seeds and soya products**	0·6 oz/0·08 cup	0·08/9·49 = 0·008	0·78 × 0·008	0·006
Seafood††	1·2 oz/0·30 cup	0·30/9·49 = 0·030	0·78 × 0·03	0·02
Oils‡‡	27 g = 5·4 tsp/0·11 cup	0·11/9·49 = 0·01	0·78 × 0·01	0·008
Includes (0·2)*	2·43/12·15			
Meat (oz)§§	1·8 oz/0·45 cup	0·45/2·43 = 0·19	0·2 × 0·19	0·04
Poultry	1·5 oz/0·38 cup	0·38/2·43 = 0·16	0·2 × 0·16	0·03
Eggs	0·4 oz/0·1 cup	0·1/2·43 = 0·04	0·2 × 0·04	0·01
Refined grains†	1·5 cup	1·5/2·43 = 0·62	0·2 × 0·62	0·12
Limits (0·02)*	0·23/12·15			
Discretionary solid fats‡‡	16 g/0·07 cup	0·07/0·23 = 0·30	0·02 × 0·3	0·006
Added sugar‖‖	32 g/0·16 cups	0·16/0·23 = 0·70	0·02 × 0·7	0·01
Total	12·15			1

tsp, Teaspoon.

* The umbrella terms 'emphasise', 'include' and 'limit' reflect our interpretation of the qualitative recommendations of the 2010 Dietary Guidelines for Americans. The food groups within these categories reflect 78, 20, and 2% of the total food volume, respectively. The hf for each food group was calculated by multiplying the broad-food-share recommendation by the specific intake recommendation in the 2000-kcal USDA Food Pattern. For example, the hf for whole grains is equal to 0·12 (0·78 × 0·16) because whole grains are within the food groups to 'emphasise' (78% of the total food volume) and represent 16% of the volume within the 'emphasised' food groups.

† For grains 1/2 cup = 1 oz.

‡ Low-fat dairy groups included in the MyPyramid Equivalents Database (MPED) are milk, cheese and yogurt.

§ Red/orange vegetable groups included in the MPED are tomatoes and orange vegetables.

‖ Starchy vegetables included in the MPED are potatoes and other starchy vegetables.

¶ Fruits included in the MPED are citrus fruits, melon and berries, and other fruits and include 100% fruit juices.

** For nuts, seeds and soya products, 1/8 cup is approximately equal to 0·5 oz. The foods included in the MPED for this group included nuts and seeds and soy foods.

†† Fish included in the MPED for this group included high *n*-3 fish and low *n*-3 fish.

‡‡ Converting fat in g to cups was done as follows. Fat contains 38 J/g (9 kcal/g) and 188 J/tsp (45 kcal/tsp) (16 g × 9 = 144 kcal; 144 kcal/45 kcal per tsp = 3·2 tsp); 1 cup = 48 tsp.

§§ For meats, 1/4 cup is approximately equal to 1 oz. The lean meats included in the MPED for this group included lean meats, frankfurters and organ meats.

‖‖ Sugar contains 4·2 g/tsp; 1 cup = 48 tsp.

solid fats and added sugar in its recommendations for a healthful diet. When the twenty-six food groups are positioned within these larger DGAC categories, seven food groups comprise the foods to 'emphasise', four belong to foods to 'include' and two belong to foods to 'limit'. Despite being less nutrient-dense than whole grains, we categorised refined grains as a food group to 'include' rather than 'limit' because the DGA allows for moderate refined grain intake in a healthy diet by only recommending that individuals consume at least half of their grains from whole grains. When calculated as a proportion of the total food volume, food groups to 'emphasise' represented 78% of the total food volume, and food groups to 'include' and 'limit' represented 20 and 2% of the total food volume, respectively.

We calculated the hf in Table 1 by first establishing the relative proportion by volume that each food group represented within its larger 'emphasise', 'include' or 'limit' category (i.e. 1·5 cups of 9·49 cups in the food groups to 'emphasise' should come from whole grains, or 16% by volume). Then, we multiplied the recommended proportion of each food group by the broader DGAC's 'emphasise, 'include' and 'limit' category recommendations (i.e. 78, 20 or 2% of the total food volume). For example, because whole grains are within the food groups to 'emphasise' (78% of the total food volume) and represent 16% of the volume within the 'emphasised' food groups, the hf for whole grains is equal to 0·12 (0·78 × 0·16).

Calculation of the US Healthy Food Diversity index

Dietary data from two 24 h dietary recalls in the 2003–4 and 2005–6 NHANES were downloaded and merged with the Demographics File and MyPyramid Equivalents Database 2.0 and 3.0, respectively. The MyPyramid Equivalents Database 2.0 and 3.0 are databases created by the USDA to translate the individual foods files in the NHANES data into their equivalent food group amounts[41]. In the MyPyramid Equivalents Database, there are a total of thirty-two food groups (*n* 6) and subgroups (*n* 26) including alcohol, which was not included in these analyses. By merging the detailed individual food files in the NHANES with the MyPyramid Equivalents Database, each food code was disaggregated into its component parts. For example, mixed dishes such as pizza were separated into their refined grain, low-fat dairy, tomato and discretionary fat components. Similarly, whole milk was disaggregated into low-fat dairy and discretionary fat.

We calculated the US HFD index for each NHANES participant based on 2 d dietary recall data using all the available food codes (method 1). To compare how differing diet assessment methods may affect the US HFD index, we also computed the US HFD index at the food-group level, i.e. using only the total intake of the twenty-six food groups rather than including all of the foods within the twenty-six food groups (method 2). The key differences between the calculation methods is that the first method allowed for each individual food code to count towards the food count component of the US HFD index whereas the latter technique only allowed for the twenty-six food groups to count towards the food count component of the index. This comparison is meaningful because it helps illustrate the differences that could result from using 24 h recall data v. FFQ data in future studies. An example of these calculations using all foods v. twenty-six food groups and their influence on the distribution of the score are described in the online supplementary material. The US HFD index was also generated with usual

intake estimates of each of the twenty-six food groups by utilising macros developed by the US National Cancer Institute (method 3; Fig. 2) [42]. Utilising two 24 h recalls and relevant covariates (e.g. age, sex, race and education)[43], we fit a one- or two-part model with fixed effects and person-specific random effects to estimate the probability of intake as well as the consumption amount at the individual level[42].

Evaluation of the US Healthy Food Diversity index

Examining the energy-adjusted correlation between each food group and the US Healthy Food Diversity index. We examined energy-adjusted Pearson's correlations between the US HFD index and the intake of each of the twenty-six food groups to compute the US HFD index. The 24 h recall data were used to estimate intake for food groups that were consumed more regularly. Food propensity data reported by a subset of participants (n 6001) were used to estimate daily intake frequency rather than the amount of vegetables,
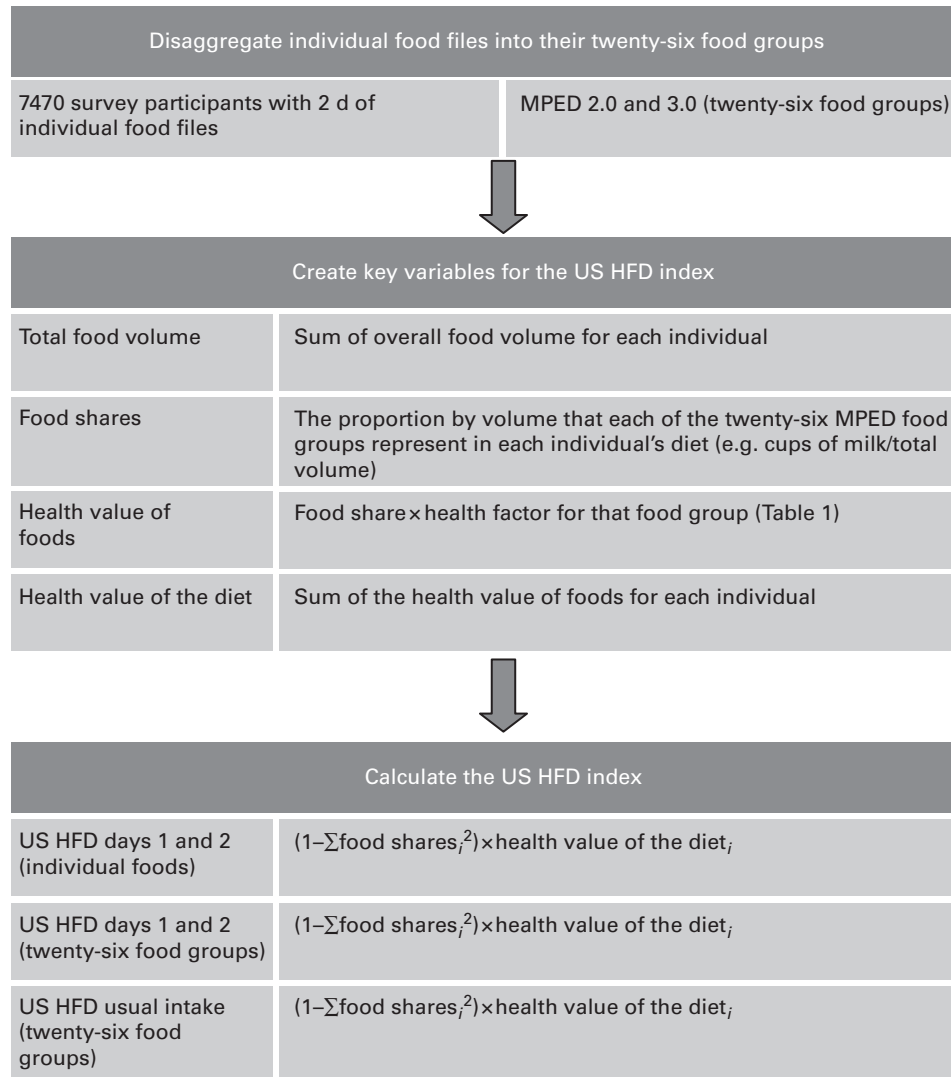

Disaggregate individual food files into their twenty-six food groups	
7470 survey participants with 2 d of individual food files	MPED 2.0 and 3.0 (twenty-six food groups)

Create key variables for the US HFD index	
Total food volume	Sum of overall food volume for each individual
Food shares	The proportion by volume that each of the twenty-six MPED food groups represent in each individual's diet (e.g. cups of milk/total volume)
Health value of foods	Food share × health factor for that food group (Table 1)
Health value of the diet	Sum of the health value of foods for each individual

Calculate the US HFD index	
US HFD days 1 and 2 (individual foods)	$(1 - \sum \text{food shares}_i^2) \times \text{health value of the diet}_i$
US HFD days 1 and 2 (twenty-six food groups)	$(1 - \sum \text{food shares}_i^2) \times \text{health value of the diet}_i$
US HFD usual intake (twenty-six food groups)	$(1 - \sum \text{food shares}_i^2) \times \text{health value of the diet}_i$

Fig. 2. Stepwise methodology for generating the US Healthy Food Diversity (HFD) index. MPED, MyPyramid Equivalents Database.

legumes, nuts, soya products, organ meats, frankfurters and fish food groups because a high proportion of participants did not consume these food groups on recall days and because 2 d of recall do not reliably estimate usual intake among episodically consumed food groups[44].

Computing the probability of nutrient adequacy. We calculated the probability of adequacy for fifteen key nutrients of interest[45–47], which were selected based on their relevance in assessing dietary quality and because they are nutrients of public health concern in the USA[45]. Only nutrients obtained from foods and beverages rather than from dietary supplements were examined[48,49]. The probability of adequacy for each nutrient was calculated using the Estimated Average Requirement (EAR) and CV for the population, which is possible when the requirement distribution for a nutrient is normal[45,46]. Fe was the only nutrient where the probability of inadequacy at different levels of Fe intake was taken from sex-specific published tables[50].

As described by Barr *et al.*[51], online supplementary Fig. S1 shows the stepwise methodology for calculating the probability of nutrient adequacy. Briefly, z-scores were calculated for each individual to determine the probability that dietary intake met the requirement for selected nutrients.

$$z\text{-score} = \frac{(\text{mean observed intake} - \text{EAR})}{\sqrt{(\text{variance}_{\text{requirement}} + (\text{variance}_{\text{within}}/n))}}.$$

Variance in the requirement and within-individual variance were taken from published tables[50]. We examined both unadjusted and energy-adjusted correlations between the US HFD index and the probability of adequacy for each of the fifteen individual nutrients and the mean probability of adequacy for all fifteen nutrients to ensure that the probability of adequacy was not increasing as a function of consuming more food[45]. We also examined the correlations between the US HFD index and fibre, Na, total fat and cholesterol. For the purposes of comparison, we examined the correlation between two variety indices (the BI and a food count measure of variety) and the probability of adequacy for individual nutrients.

Evaluation of the US Healthy Food Diversity index across subpopulations and with overall diet quality. We evaluated whether the US HFD index differed between smokers and non-smokers, men and women, and by age group because research has suggested that diets of these groups differ by measures of quality[26] and, to an extent, by variety[3]. We computed sample-weighted means for the US HFD index for each subpopulation and used linear regression to determine whether the index differed by smoking status, sex or age category. We also generated a DASH score and assessed the concordance of the US HFD index, BI and food count with the DASH diet by computing energy-adjusted correlations. The DASH score measures dietary quality using the measures of total and whole grains, added sugars, Na, fruits, vegetables, dairy products, meat, total and saturated fat, and nuts and legumes. Methods for calculating the score have been previously published[52]. Although the US HFD index is a dietary variety index that includes measures of quality and

proportionality, higher scores should capture a more nutrient-dense dietary pattern, making comparison with the DASH diet score relevant.

Statistical analyses

SAS version 9.3 (SAS Institute, Inc.) was used to conduct all statistical analyses. Appropriate strata, cluster and sample weights were selected based on NHANES analytic guidelines[28]. Research conducted with de-identified data from the publicly available NHANES does not constitute human subjects research per the Institutional Review Board at New York University.

Results

Mean intake of the twenty-six food groups used in calculating US Healthy Food Diversity index scores in men and women

On average, both men and women exceeded intake for food groups that should be consumed in moderation and consumed less than what is recommended within food groups that are encouraged. The mean 2 d intake of whole grains was 0·40 for men and 0·34 cups for women, which is below the recommendation of ≥1·5 cup by more than 1 cup. Total fruit intake was also approximately 1 cup below the recommendation (2 cups) for both sexes, and total milk intake was approximately 2 cups below the recommendation (3 cups). Conversely, intake of added sugars (0·38 cups/18·2 teaspoons) and intake of discretionary solid fats (0·19 cups/9·1 teaspoons) were more than twice the amount recommended (0·16 and 0·07 cups, respectively) (see online supplementary Table S1).

Calculation of US Healthy Food Diversity index scores using the three methods

Table 2 provides summary information for the US HFD index calculated using the three different methods (method 1: 2 d mean reported intakes, all individual foods; method 2: 2 d mean reported intakes, twenty-six food groups; method 3: usual intakes, twenty-six food groups) to demonstrate how different calculation techniques influence the total mean score. Mean US HFD index scores were similar regardless of the calculation method used; mean scores calculated using method 2 or method 3 differed by 0–8% from the estimates generated using method 1. Overall, the food-group level data (method 2) slightly reduced the score relative to the calculation based on all the reported individual foods (method 1). For example, the mean US HFD index value calculated using only twenty-six food groups (0·34) was approximately 6% lower than the mean US HFD index value calculated using all individual foods (0·36). Additionally, as expected, the range of scores was wider when using all individual foods (0·03–0·68) than when using the twenty-six food groups (0·01–0·56). US HFD index scores were higher in the overall population (0·39) and in men (0·38) and women (0·39) when usual intake estimates of the twenty-six food groups were used (method 3) and would increase further if more foods or food groups were included.

Table 2. Mean US Healthy Food Diversity (HFD) index scores among adults aged 20 years and older from the 2003–6 National Health and Nutrition Examination Survey computed using the three methods*†

(Mean values with their standard errors)

	Overall population (n 7470)		Men (n 3786)		Women (n 3684)	
	Mean	SE	Mean	SE	Mean	SE
2 d means (all individual foods) (method 1)	0·36	0·002	0·36	0·003	0·37	0·002
2 d means (twenty-six food groups) (method 2)	0·34	0·002	0·34	0·002	0·35	0·002
Usual intake‡ (twenty-six food groups) (method 3)	0·39	0·001	0·38	0·002	0·39	0·001

* The US HFD index was calculated using the following equation: US HFD index $= (1 - \Sigma s_i^2) \times$ hv, where s_i is the share or proportion of each individual food or food group i based on the volume of the total diet; hv $= \Sigma hf_i \times s_i$, where hv is the health value and hf is the health factors of food groups.

† Note that a diet comprised of the twenty-six food groups in the proportions recommended by the US Department of Agriculture has a US HFD index score of 0·49.

‡ Usual intake of the twenty-six food groups was estimated using methodology developed by the US National Cancer Institute. Using two 24 h recalls and relevant covariates, a one- or two-part model with fixed effects and person-specific random effects was fitted to estimate the probability of intake as well as the consumption amount at the individual level based on the specified covariates[42]. Consistent with existing studies, the covariates used to predict usual intake included race, age, sex and education level[43]. Propensity data from the FFQ were also used in exploratory models to examine whether it improved usual intake estimates for episodically consumed foods[43].

Correlations between the US Healthy Food Diversity index and components

Table 3 shows energy-adjusted Pearson's correlations between the US HFD index and each of the twenty-six food groups to compute the score for the total population and by sex, demonstrating the linear association between each food group and the overall score. For both men and women, the US HFD index was positively correlated with intake of whole grains, milk, citrus fruits, melon and berries, other fruits and orange vegetables (r 0·12 to 0·63, $P < 0.0001$). Conversely, the US HFD index was negatively correlated with intakes of poultry, lean meats and added sugars (r −0·14 to −0·23, $P < 0.0001$). Index scores were positively correlated with fibre intake (r 0·19, $P < 0.0001$) and negatively correlated with Na, total fat, and cholesterol intake (r −0·08 to −0·12, $P < 0.0001$). Some small differences were observed between men and women. For example, there was a negative correlation between the US HFD index and intake of oils in women (r −0·14) but not in men. Similarly, the US HFD index was positively correlated with discretionary fat intake in men (r 0·11), but not in women. In both sexes, the US HFD index was positively correlated with overall dietary quality as measured in accordance with the dietary recommendations of the DASH (r 0·32 in men; r 0·38 in women). The correlation with DASH recommendations was stronger with the US HFD index than with the BI (r 0·17 in men, r 0·21 in women) and total food count (r 0·25, both sexes).

Probability of nutrient adequacy and correlations with the US Healthy Food Diversity index

We examined mean nutrient intakes compared with the respective EAR and the mean probability of adequacy for fifteen nutrients (see online supplementary Table S2). Although mean intake for most nutrients exceeded the EAR, the mean probability of adequacy ranged from 0·27 to 0·98 in men and from 0·13 to 0·88 in women for individual nutrients. The mean probability of adequacy for the fifteen

nutrients was 0·68 in men and 0·60 in women, indicating that more than half of the participants met their intake requirements for the selected nutrients.

Table 4 shows unadjusted and adjusted correlations for energy intake between the US HFD index and the mean probability of adequacy for all the fifteen nutrients and the probability of adequacy for each nutrient. The mean probability of nutrient adequacy was correlated with 2 d energy intake in both men (r 0·72) and women (r 0·68). US HFD index scores were also correlated with the mean probability of adequacy for the fifteen nutrients (r 0·38–0·40, $P < 0.0001$). After controlling for energy intake, these correlations were not attenuated (r 0·41, $P < 0.0001$).

Similar to the German HFD index[22], the US HFD index was correlated with the probability of adequacy for most individual nutrients. In men, the US HFD index significantly correlated with the probability of adequacy for all nutrients with the exception of vitamin E and niacin after adjustment for energy. Correlations ranged from 0·09 to 0·59 without adjustment for energy and from 0·01 to 0·62 after adjustment for energy. In women, the US HFD index was correlated with the probability of adequacy for all nutrients except vitamin E. Correlations ranged from 0·04 to 0·55 without adjustment for energy and 0·01 to 0·58 after adjustment for energy.

We also examined the correlation between the BI and food count measures of variety with the probability of nutrient adequacy. These indices were more correlated with energy intake (r 0·14–0·33) than the US HFD index (r 0·10–0·14). Both the BI and food count measures correlated with the probability of adequacy for all nutrients with unadjusted correlations that ranged from 0·18–0·50 and from 0·09–0·41 after adjustment for energy.

US Healthy Food Diversity index scores between subpopulations

Survey-weighted mean US HFD index scores by smoking status, sex and age groups are shown in Table 5. The mean

Table 3. Energy-adjusted Pearson's correlations between the US Healthy Food Diversity (HFD) index and 2 d mean intakes of twenty-six MyPyramid food groups in adults aged 20 years or older in the 2003–6 National Health and Nutrition Examination Survey†

	US HFD		
Food groups	Overall population (*n* 7470)	Men (*n* 3786)	Women (*n* 3684)
Total grains			
Whole grains	0·22**	0·24**	0·21**
Refined grains	−0·01	0·01	−0·02
Total vegetables			
Dark green vegetables	0·05*	0·07 (*P*=0·05)	0·02
Orange vegetables	0·12**	0·11**	0·11*
Tomatoes	−0·002	0·01	−0·01
Potatoes	−0·07*	−0·05*	−0·09**
Starchy vegetables	0·02	0·04	−0·01
Other vegetables	0·01	0·03	−0·02
Total fruits			
Citrus fruits, melon and berries	0·24**	0·25**	0·23**
Other fruits	0·34**	0·34**	0·34**
Total milk			
Milk	0·63**	0·64**	0·64**
Yogurt	0·17**	0·15**	0·16**
Cheese	0·22**	0·23**	0·20**
Meats, poultry, fish			
Legumes	0·02	0·03	0·01
Nuts	0·09**	0·12**	0·07*
Soya	0·02	0·02	0·01
Lean meats	−0·23**	−0·24**	−0·20**
Frankfurters	−0·05*	−0·04	−0·04
Organ meats	−0·01	0·005	−0·02
Poultry	−0·16**	−0·15**	−0·17**
Overall fish intake	−0·04*	0·02	−0·10*
Eggs	−0·03	−0·01	−0·04
Oils			
Oils	−0·09**	−0·07	−0·14**
Extras			
Added sugars	−0·14**	−0·12**	−0·19**
Discretionary fats	0·07**	0·11**	0·002
Other nutrients			
Fibre	0·19**	0·18**	0·19**
Na	−0·09**	−0·06*	−0·12**
Total fat	−0·08**	−0·03	−0·17**
Cholesterol	−0·12**	−0·10**	−0·12**
Total sugars‡	−0·004	0·001	−0·02
DASH score	0·35**	0·32**	0·38**

DASH, Dietary Approaches to Stop Hypertension.
* $P < 0.05$, ** $P < 0.0001$.
† For the analysis of the vegetable groups, legumes, nuts, soya, organ meats, frankfurters, eggs and fish food groups, correlations were based on the frequency of intake reported in the FFQ by a subset of participants (*n* 3062 women and *n* 2939 men) because of the high proportion of zero intakes for these food subgroups.
‡ The added sugars food group was negatively correlated with the energy-adjusted US HFD index. To estimate added sugars from nutrient intake, we examined the energy-adjusted correlation between the US HFD index and total sugars. Because total sugars include sugars from dairy products, fruit and vegetables, we also adjusted for intake of these three food groups.

US HFD index score was 0·36 for the overall population. Mean scores were approximately 27 % lower than the score achieved by consuming a dietary pattern congruent with the 2000-kcal USDA Food Pattern (US HFD: 0·49). The US HFD index was lower among smokers than among non-smokers (0·34 *v.* 0·38, $P < 0.0001$), among men than among women (0·36 *v.* 0·37, $P < 0.005$), and among younger than among older individuals (0·35–0·38 *v.* 0·41, $P < 0.0001$).

Non-smokers had higher energy-adjusted intakes of some food groups that were positively correlated with the US HFD index including whole grains, other fruits, orange vegetables, and citrus fruits, melon and berries ($P < 0.0001$), as well as low-fat milk ($P < 0.05$) (see online supplementary Table S3). Non-smokers also consumed less lean meat, frankfurters, discretionary fat and added sugar ($P < 0.05$) after adjustment for energy. The small difference in score between men and women may have been due to greater consumption of negatively weighted food groups including potatoes, lean meat and frankfurters ($P < 0.05$) in men. Women also had greater consumption of milk ($P < 0.01$) and intake of citrus fruits, melon and berries ($P < 0.05$) than men. However, women had higher energy-adjusted intakes of sugar and discretionary fats ($P < 0.05$). No differences were observed in the intake of whole grains, other fruits or orange vegetables between men and women.

M. Vadiveloo *et al.*

Table 4. Unadjusted and energy-adjusted Pearson's correlations between the US Healthy Food Diversity (HFD) index and the probability of nutrient adequacy in adults aged 20 years or older in the 2003−6 National Health and Nutrition Examination Survey†

	Correlation with the probability of nutrient adequacy‡											
	Men (*n* 3786)						Women (*n* 3684)					
	US HFD index		BI		Food count		US HFD index		BI		Food count	
	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
Energy intake	0·14**		0·14**		0·31**		0·10**		0·20**		0·33**	
Mean probability of nutrient adequacy§	0·40**	0·41**	0·37**	0·39**	0·51**	0·43**	0·38**	0·41**	0·43**	0·41**	0·53**	0·44**
Fe	0·12**	0·09*	0·22**	0·19**	0·21**	0·14**	0·22**	0·19**	0·34**	0·30**	0·38**	0·29**
Vitamin A	0·44**	0·42**	0·33**	0·30**	0·38**	0·30**	0·37**	0·35**	0·35**	0·30**	0·43**	0·35**
Vitamin E	0·09*	0·01	0·20**	0·15**	0·38**	0·28**	0·04	0·01	0·20**	0·14**	0·32**	0·21**
Vitamin C	0·26**	0·23**	0·30**	0·28**	0·32**	0·28**	0·22**	0·20**	0·32**	0·29**	0·34**	0·30**
Thiamin	0·33**	0·30**	0·27**	0·23**	0·37**	0·26**	0·29**	0·28**	0·29**	0·23**	0·38**	0·25**
Riboflavin	0·42**	0·40**	0·30**	0·27**	0·40**	0·30**	0·41**	0·40**	0·33**	0·27**	0·39**	0·28**
Niacin	0·09**	0·01	0·30**	0·18**	0·34**	0·22**	0·11*	0·05*	0·30**	0·24**	0·38**	0·25**
Vitamin B_6	0·19**	0·12*	0·30**	0·27**	0·34**	0·22**	0·22**	0·18**	0·36**	0·31**	0·38**	0·26**
Vitamin B_{12}	0·32**	0·28**	0·20**	0·15**	0·24**	0·12**	0·30**	0·27**	0·25**	0·20**	0·27**	0·16**
Folate	0·14**	0·06*	0·25**	0·21**	0·38**	0·27**	0·08**	0·04*	0·22**	0·16**	0·33**	0·24**
P	0·31**	0·28**	0·31**	0·28**	0·35**	0·23**	0·34**	0·33**	0·35**	0·29**	0·40**	0·27**
Mg	0·28**	0·24**	0·29**	0·26**	0·44**	0·33**	0·32**	0·31**	0·36**	0·31**	0·49**	0·39**
Cu	0·23**	0·17**	0·33**	0·31**	0·47**	0·37**	0·22**	0·19**	0·39**	0·35**	0·50**	0·41**
Zn	0·22**	0·16**	0·21**	0·16**	0·31**	0·16**	0·24**	0·21**	0·30**	0·23**	0·33**	0·17**
Ca	0·59**	0·62**	0·18**	0·13**	0·31**	0·16**	0·55**	0·58**	0·18**	0·09**	0·27**	0·11**

BI, Berry Index.

* $P<0.05$, ** $P<0.0001$.

† All analyses were adjusted using appropriate sample weights for the complex survey design.

‡ Mean 2 d nutrient intakes were correlated with the 2 d mean US HFD index. We computed both unadjusted Pearson's correlations and energy- adjusted correlations.

§ Mean probability of adequacy was calculated by averaging the probability of adequacy for the fifteen individual nutrients. The correlation between the mean probability of nutrient adequacy and energy intake was 0·72 in men and 0·68 for in women.

Table 5. US Healthy Food Diversity index scores for adults aged 20 years or older by smoking status, sex and age group in the 2003−6 National Health and Nutrition Examination Survey*

(Mean values and standard errors)

	2 d mean	SE
Overall population (n 7470)	0·36	0·002
Smoking status		
Non-smokers (n 5812)	0·37**	0·002
Smokers†‡ (n 1652)	0·34	0·003
Sex		
Men (n 3786)	0·36**	0·003
Women† (n 3684)	0·37	0·002
Age groups (years)		
20−35 (n 1853)	0·35**	0·003
36−50 (n 1932)	0·36**	0·003
51−65 (n 1713)	0·36**	0·003
66−80 (n 1431)	0·38**	0·003
>80† (n 541)	0·41	0·004

*All analyses were adjusted using appropriate cluster, stratum, and sample weights for the complex survey design.
** $P < 0.0001$.
† Denotes the referent group.
‡ Participants were coded as smokers if they reported currently smoking cigarettes 'every day' or 'some days', and as non-smokers if they did not report smoking at least 100 cigarettes in their lifetime or if they reported not currently smoking cigarettes.

Discussion

The potential for dietary variety to increase energy intake and thus obesity has led to the removal of the distinct US guideline to 'eat a variety of foods', even though healthful, varied diets may be optimal for wellness. However, the role of dietary variety in obesity is equivocal in part because existing dietary variety indices do not incorporate measures of dietary quality and proportionality in their assessments[10]. The primary objective of these analyses was to develop a dietary variety score using US dietary guidance that also measured dietary quality and proportionality.

Because dietary variety, quality and proportionality are included in the US HFD index, a number of dietary patterns can improve scores depending on whether individual foods or food groups are used in the calculation. With either dietary input, a dietary pattern with between-group variety and proportionality (i.e. consumption of the twenty-six food groups within the recommended proportions of the USDA) but limited within-group variety among the twenty-six food groups (i.e. consuming one food within each of these groups) would score 0·49. Scores would increase by consuming a higher proportion of foods from more healthful food groups, whereas scores would decrease when less healthful food groups are consumed in higher proportions. When individual foods (v. food groups) are used, increasing within-group variety among more heavily weighted food groups also improves scores. From the perspective of public health, the latter technique is useful for clarifying whether a wider assortment of healthful foods increases consumption and promotes a dietary pattern favourably associated with health.

Based on these analyses, mean US HFD index scores within the US population were 27 % lower than what would be achieved by consuming a diet within the recommended proportions of the USDA but minimal within-group variety, suggesting that the typical adult is not selecting a diet composed of a variety of healthful foods. On average, US adults overconsumed food groups requiring moderation and inadequately consumed 'emphasised' food groups. Shifting intake from 'limit' to 'emphasised' food groups would improve the US HFD index.

Our evaluation of the US HFD index suggests that it measures dietary variety in conjunction with dietary quality. The index was most strongly correlated with low-fat milk intake because it was the most heavily weighted food group within the 2000-kcal USDA Food Pattern. However, the index also correlated with many other healthful food groups and nutrients, while negative correlations were observed for foods and nutrients that should be limited. Taken together, these findings support the previous assertion that adults must shift intake from less heavily weighted food groups to 'emphasised' food groups to improve US HFD index scores. This is further reinforced by the higher scores observed in non-smokers who consumed more dairy products, whole grains and fruits, and less added sugar and lean meats than smokers. Similar patterns, albeit to a lesser extent, were observed among women v. men, potentially accounting for the slightly higher scores. Although observed differences between these groups are small, the US HFD index is relatively sensitive to slight changes in the diet such that increasing fruit intake by <1 serving while also reducing meat intake will improve the index by nearly 10 %.

The US HFD index also positively correlated with overall nutrient adequacy, demonstrating that individuals who consumed a diet in accordance with the US guidelines were more likely to meet recommended nutrient intakes. Because energy intake and nutrient adequacy were strongly correlated in the present study and in the literature[45], it is possible that higher US HFD index scores were a marker for a higher-energy dietary pattern. However, the US HFD index was less correlated with energy than both the BI and food count, and the correlation between the US HFD index and the mean probability of adequacy did not attenuate with adjustment for energy. Hence, it is apparent that higher index scores were not achieved by consuming more food but rather by consuming a higher proportion of nutrient-dense foods.

The US HFD index correlated most with the probability of adequacy for vitamin A, riboflavin and Ca, probably reflecting positive correlations between the index and sources of low-fat dairy products, orange vegetables and fruits, which are key food sources of vitamin A in the USA[53,54]. Milk is also the primary source of both riboflavin and Ca in the USA[53]. Conversely, the US HFD index was not correlated with niacin adequacy in men or with vitamin E adequacy in both sexes. Key sources of vitamin E include ready-to-eat-cereals, oils, margarine and baked goods[53], and high intake of these food groups negatively correlated with the US HFD index. Correlations were also weaker for vitamin B_6, Fe, niacin (in women) and folate. Ready-to-eat cereals are important sources of all of these nutrients, while poultry and beef are the major sources of niacin, vitamin B_6 and Fe[53]. The US HFD index negatively correlated with lean meats,

poultry and added sugars, potentially accounting for the weaker correlation between the index and these nutrients. Conversely, the BI and food count positively correlated with these nutrients, highlighting how these indices captured different dietary patterns.

The US HFD index appeared to capture overall diet quality better than existing variety scores because it more strongly correlated with the DASH diet score (0·38 in women and 0·32 in men) than either the BI or a food count measure. Similar to existing diet quality scores, the US HFD index also differentiated between populations with established differences in dietary quality[23]. In the present study, the US HFD index was higher among non-smokers *v.* smokers, women *v.* men, and older *v.* younger adults, all of whom have been previously shown to consume more healthful diets[26]. Taken together, our findings suggest that the US HFD index has the content and construct validity and performs comparably with the German HFD, the only other index that has simultaneously evaluated dietary variety, dietary quality and proportionality[22,55].

Our understanding of the association between dietary variety and adiposity in US populations is equivocal[10] because we have relied on tools that exclusively measure dietary variety[17] or variety only within certain food groups[18]. Even though food count measures of variety positively correlate with nutrient adequacy and inversely correlate with added sugars and saturated fats[2], heterogeneity in diet quality may still exist because low- and high-intake consumers of a particular food cannot be differentiated. This is especially problematic when many nutrient-dense foods are consumed in small quantities because energy-adjusted correlations between the variety index and the food group may be statistically significant, but not practically meaningful.

In contrast to existing diet quality indices that also measure variety[56], the US HFD index is not constrained by only evaluating a predetermined number of foods or by dichotomising dietary intakes into categories of 'met'/'not met' with regard to specific quantitative recommendations[1]. The US HFD index reduces misclassification of individuals by distinguishing between individuals who consume a highly *v.* adequately varied diet, by including all foods and by considering consumption amount. Because these forms of bias may incorrectly inform our understanding of the associations between dietary variety, adiposity and chronic disease, it is important that future studies utilise a measurement tool that is more strongly associated with dietary quality. We found that the US HFD index correlated with the energy-adjusted probability of adequacy for most nutrients, and that it more strongly correlated with the DASH diet score than either the BI or food count, suggesting that it may be a better tool.

Some limitations of these analyses should be mentioned. By selecting the USDA Food Pattern to measure dietary quality, low-fat dairy food groups may be more weighted than necessary to achieve a high-quality dietary pattern[1]. However, by incorporating recommendations to consume a plant-based diet into the health weights, varied non-dairy or vegetarian food patterns still scored well, as evidenced by the positive correlations between the US HFD index and intake of plant foods and the negative correlations with meat and poultry.

Future research is needed to develop alternative health weights for other high-quality diet patterns. Additionally, self-reported dietary data are subject to some systematic error[57], though the influence of this error was minimised to the extent possible by using 24 h dietary recall data that has less bias than other dietary assessment methods[58] and by utilising propensity data from the FFQ to better capture episodically consumed food groups. Importantly, the conclusions drawn from the present study were similar regardless of whether usual intakes or 2 d intakes were used to calculate the US HFD index. Finally, although there are no 'gold' standard dietary variety scores against which the US HFD index can be calibrated, the use of other measures of validity along with comparison against other variety indices increased our confidence in the robustness of this tool.

There are a number of unique features of the US HFD index that are worthy of mention. To our knowledge, the US HFD index is the first tool developed using the US national dietary guidance and a nationally representative population that jointly examines dietary variety, dietary quality and proportionality. Moreover, the theoretical validity of this investigation was enhanced because we used the validated German index[22] to inform the development of the US HFD index. Methodologically, we were able to utilise relatively precise 24 h dietary recall data[58], while also correcting for the inability of 2 d of recall data to adequately reflect the usual intake. Although it did not change our conclusions, developing the index using both 2 d means and usual intake estimates ensured that using individual food-level data did not introduce bias.

In conclusion, the present study shows that the US HFD index is an efficacious tool for the simultaneous measurement of dietary variety, quality and proportionality. From an epidemiological perspective, it may be used to better understand whether promoting variety within healthful foods can be an effective weight-control strategy and reduce the risk of chronic disease within diverse US populations. The US HFD index may also be used to help make the dietary guidelines more clear and actionable with regard to the benefits of dietary variety.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0007114514002049

Acknowledgements

writing, and critically reviewed the manuscript for important intellectual content. All authors read and approved the final manuscript and took full responsibility for the final content and have no conflicts of interest to report.

References

1. Dixon LB, Cronin FJ & Krebs-Smith SM (2001) Let the pyramid guide your food choices: capturing the total diet concept. *J Nutr* **131**, 461S–472S.
2. Murphy SP, Foote JA, Wilkens LR, *et al.* (2006) Simple measures of dietary variety are associated with improved dietary quality. *J Am Diet Assoc* **106**, 425–429.
3. Kant AK, Schatzkin A, Harris TB, *et al.* (1993) Dietary diversity and subsequent mortality in the First National Health and Nutrition Examination Survey Epidemiologic Follow-up Study. *Am J Clin Nutr* **57**, 434–440.
4. Kant AK & Graubard BI (2005) A comparison of three dietary pattern indexes for predicting biomarkers of diet and disease. *J Am Coll Nutr* **24**, 294–303.
5. Mirmiran P, Azadbakht L, Esmaillzadeh A, *et al.* (2004) Dietary diversity score in adolescents – a good indicator of the nutritional adequacy of diets: Tehran lipid and glucose study. *Asia Pac J Clin Nutr* **13**, 56–60.
6. Kant AK, Block G, Schatzkin A, *et al.* (1991) Dietary diversity in the US population, NHANES II, 1976–1980. *J Am Diet Assoc* **91**, 1526–1531.
7. Kennedy E (2004) Dietary diversity, diet quality, and body weight regulation. *Nutr Rev* **62**, S78–S81.
8. De Lorenzo A, Deurenberg P, Pietrantuono M, *et al.* (2003) How fat is obese? *Acta Diabetol* **40**, Suppl. 1, S254–S257.
9. Azadbakht L, Mirmiran P, Esmaillzadeh A, *et al.* (2006) Dietary diversity score and cardiovascular risk factors in Tehranian adults. *Public Health Nutr* **9**, 728–736.
10. Vadiveloo M, Dixon LB & Parekh N (2013) Associations between dietary variety and measures of body adiposity: a systematic review of epidemiological studies. *Br J Nutr* **109**, 1557–1572.
11. Putnam J & Allshouse J (1999) Food consumption, prices, and expenditures, 1970–97. In *US Department of Agriculture Statistical Bulletin* no. 965. Washington, DC: Government Printing Office.
12. Remick AK, Polivy J & Pliner P (2009) Internal and external moderators of the effect of variety on food intake. *Psychol Bull* **135**, 434–451.
13. Jekanowski MD & Binkley JK (2000) Food purchase diversity across U.S. markets. *Agribusiness* **16**, 417–433.
14. Moss M (2013) *Salt Sugar Fat: How the Food Giants Hooked us United States*. New York: Random House.
15. Raynor HA & Epstein LH (2001) Dietary variety, energy regulation, and obesity. *Psychol Bull* **127**, 325–341.
16. Hill AJ (2007) The psychology of food craving. *Proc Nutr Soc* **66**, 277–285.
17. Drewnowski A, Henderson SA, Driscoll A, *et al.* (1997) The Dietary Variety Score: assessing diet quality in healthy young and older adults. *J Am Diet Assoc* **97**, 266–271.
18. McCrory MA, Fuss PJ, McCallum JE, *et al.* (1999) Dietary variety within food groups: association with energy intake and body fatness in men and women. *Am J Clin Nutr* **69**, 440–447.
19. Roberts SB, Hajduk CL, Howarth NC, *et al.* (2005) Dietary variety predicts low body mass index and inadequate macronutrient and micronutrient intakes in community-dwelling older adults. *J Gerontol A Biol Sci Med Sci* **60**, 613–621.
20. Ruel MT (2003) Operationalizing dietary diversity: a review of measurement issues and research priorities. *J Nutr* **133**, 3911S–3926S.
21. U.S. Department of Agriculture & U.S. Department of Health and Human Services (2010) *Dietary Guidelines for Americans, 2010*, 7th ed. Washington, DC: U.S. Government Printing Office.
22. Drescher LS, Thiele S & Mensink GB (2007) A new index to measure healthy food diversity better reflects a healthy diet than traditional measures. *J Nutr* **137**, 647–651.
23. Guenther PM, Reedy J, Krebs-Smith SM, *et al.* (2008) Evaluation of the Healthy Eating Index-2005. *J Am Diet Assoc* **108**, 1854–1864.
24. Subar AF, Harlan LC & Mattson ME (1990) Food and nutrient intake differences between smokers and non-smokers in the US. *Am J Public Health* **80**, 1323–1329.
25. Dallongeville J, Marécaux N, Fruchart J-C, *et al.* (1998) Cigarette smoking is associated with unhealthy patterns of nutrient intake: a meta-analysis. *J Nutr* **128**, 1450–1457.
26. Hiza HA, Casavale KO, Guenther PM, *et al.* (2013) Diet quality of Americans differs by age, sex, race/ethnicity, income, and education level. *J Acad Nutr Diet* **113**, 297–306.
27. Dwyer J, Picciano MF & Raiten DJ (2003) Collection of food and dietary supplement intake data: What We Eat in America – NHANES. *J Nutr* **133**, 590S–600S.
28. National Center for Health Statistics & Centers for Disease Control and Prevention (2006) *Analytic and Reporting Guidelines: The National Health and Nutrition Examination Survey (NHANES)*. Hyattsville, MD: National Center for Health Statistics, Centers for Disease Control and Prevention.
29. Breslow RA, Guenther PM, Juan W, *et al.* (2010) Alcoholic beverage consumption, nutrient intakes, and diet quality in the US adult population, 1999–2006. *J Am Diet Assoc* **110**, 551–562.
30. Dixon LB, Sundquist J & Winkleby M (2000) Differences in energy, nutrient, and food intakes in a US sample of Mexican-American women and men: findings from the Third National Health and Nutrition Examination Survey, 1988–1994. *Am J Epidemiol* **152**, 548–557.
31. Vernarelli JA & Lambert JD (2013) Tea consumption is inversely associated with weight status and other markers for metabolic syndrome in US adults. *Eur J Nutr* **52**, 1039–1048.
32. Centers for Disease Control and Prevention & National Center for Health Statistics (2006) *NHANES 2003–2004 Data Documentation, Codebook, and Frequencies: Smoking and Tobacco Use*. Hyattsville, MD: Centers for Disease Control and Prevention, National Center for Health Statistics. http://www.cdc.gov/nchs/nhanes/nhanes2003-2004/SMQ_C.htm (accessed 6 February 2014).
33. Centers for Disease Control and Prevention & National Center for Health Statistics (2008) *NHANES 2005–2006 Data Documentation, Codebook, and Frequencies*. Hyattsville, MD: Centers for Disease Control and Prevention, National Center for Health Statistics. http://www.cdc.gov/nchs/nhanes/nhanes2005-2006/SMQ_D.htm (accessed 6 February 2014).
34. Centers for Disease Control and Prevention & National Center for Health Statistics (2006) *NHANES 2003–2004 Data Documentation, Codebook, and Frequencies: Dietary Interview-Individual Foods*. Hyattsville, MD: Centers for Disease Control and Prevention, National Center for Health Statistics. http://www.cdc.gov/nchs/nhanes/nhanes2003-2004/DR1IFF_C.htm (updated November 2007 and accessed 6 February 2014).

British Journal of Nutrition

35. Centers for Disease Control and Prevention & National Center for Health Statistics (2008) *NHANES 2005–2006 Data Documentation, Codebook, and Frequencies: Dietary Interview*. Hyattsville, MD: Centers for Disease Control and Prevention, National Center for Health Statistics. http://www.cdc.gov/nchs/nhanes/nhanes2005-2006/DRXDOC_D.htm (accessed 6 February 2014).

36. National Center for Health Statistics and Centers for Disease Control (2008) *NHANES 2003–2004 Data Documentation, Codebook, and Frequencies: Food Frequency Questionnaire – Output from DietCalc Software (FFQDC_C)*. Hyattsville, MD: Centers for Disease Control and Prevention, National Center for Health Statistics. http://www.cdc.gov/nchs/nhanes/nhanes2003-2004/FFQDC_C.htm (accessed 6 February 2014).

37. National Center for Health Statistics and Centers for Disease Control (2008) *NHANES 2005–2006 Data Documentation, Codebook, and Frequencies: Food Frequency Questionnaire – Output from DietCalc Software (FFQDC_D)*. Hyattsville, MD: Centers for Disease Control and Prevention, National Center for Health Statistics. http://www.cdc.gov/nchs/nhanes/nhanes2005-2006/FFQDC_D.htm (accessed 6 February 2014).

38. Berry CH (1971) Corporate growth and diversification. *J Law Econ* **14**, 371–383.

39. U.S. Department of Agriculture & Dietary Guidelines Advisory Committee (2010) *Report of the Dietary Guidelines Advisory Committee on the Dietary Guidelines for Americans, 2010, to the Secretary of Agriculture and the Secretary of Health and Human Services*. Washington, DC: U.S. Department of Agriculture, Agricultural Research Service.

40. United States Department of Agriculture & Agricultural Research Service (2012) *USDA National Nutrient Database for Standard Reference, Release 25*. Washington, DC: U.S. Department of Agriculture, Agricultural Research Service. http://ndb.nal.usda.gov/ (updated 7 December 2011 and accessed 6 February 2014).

41. Bowman SA, Friday JE & Moshfegh A (2008) *MyPyramid Equivalents Database, 2.0 for USDA Survey Foods, 2003–2004 [Online]*. Beltsville, MD: Food Surveys Research Group, Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture.

42. National Cancer Institute (2008) *Usual Dietary Intakes: Details of the Method*. Bethesda, MD: National Cancer Institute, National Institutes of Health, and Department of Health and Human Services. http://riskfactor.cancer.gov/diet/usualintakes/details.html (updated February 2008 and accessed 6 February 2014).

43. Kipnis V, Midthune D, Buckman DW, *et al.* (2009) Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics* **65**, 1003–1010.

44. Palaniappan U, Cue RI, Payette H, *et al.* (2003) Implications of day-to-day variability on measurements of usual food and nutrient intakes. *J Nutr* **133**, 232–235.

45. Foote JA, Murphy SP, Wilkens LR, *et al.* (2004) Dietary variety increases the probability of nutrient adequacy among adults. *J Nutr* **134**, 1779–1785.

46. Mirmiran P, Azadbakht L & Azizi F (2006) Dietary diversity within food groups: an indicator of specific nutrient adequacy in Tehranian women. *J Am Coll Nutr* **25**, 354–361.

47. Verger EO, Mariotti F, Holmes BA, *et al.* (2012) Evaluation of a diet quality index based on the probability of adequate nutrient intake (PANDiet) using national French and US dietary surveys. *PLOS ONE* **7**, e42155.

48. Centers for Disease Control and Prevention & National Center for Health Statistics (2007) *2003–2004 Data Documentation, Codebook, and Frequencies: Dietary Interview – Total Nutrients, First Day*. Hyattsville, MD: Centers for Disease Control and Prevention, National Center for Health Statistics. http://www.cdc.gov/nchs/nhanes/nhanes2003-2004/DR1TOT_C.htm (accessed 6 February 2014).

49. Centers for Disease Control and Prevention, National Center for Health Statistics (2008) *2005–2006 Data Documentation, Codebook, and Frequencies: Dietary Interview – Total Nutrient Intakes, First Day*. Hyattsville, MD: Centers for Disease Control and Prevention, National Center for Health Statistics. http://www.cdc.gov/nchs/nhanes/nhanes2005-2006/DR1TOT_D.htm (accessed 6 February 2014).

50. National Academy of Sciences (2006) *Dietary Reference Intakes: The Essential Guide to Nutrient Requirements*. 560 p. [JJ Otten, JP Hellwig, LD Meyers, editors]. Washington, DC: National Academies Press.

51. Barr SI, Murphy SP & Poos MI (2002) Interpreting and using the dietary references intakes in dietary assessment of individuals and groups. *J Am Diet Assoc* **102**, 780–788.

52. Folsom AR, Parker ED & Harnack LJ (2007) Degree of concordance with DASH diet guidelines and incidence of hypertension and fatal cardiovascular disease. *Am J Hypertens* **20**, 225–232.

53. Cotton PA, Subar AF, Friday JE, *et al.* (2004) Dietary sources of nutrients among US adults, 1994 to 1996. *J Am Diet Assoc* **104**, 921–930.

54. Sommer A & Vyas KS (2012) A global clinical view on vitamin A and carotenoids. *Am J Clin Nutr* **96**, 1204S–1206S.

55. Truthmann J, Richter A, Thiele S, *et al.* (2012) Associations of dietary indices with biomarkers of dietary exposure and cardiovascular status among adolescents in Germany. *Nutr Metab* **9**, 92.

56. Kennedy ET, Ohls J, Carlson S, *et al.* (1995) The Healthy Eating Index: design and applications. *J Am Diet Assoc* **95**, 1103–1108.

57. Subar AF, Kipnis V, Troiano RP, *et al.* (2003) Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. *Am J Epidemiol* **158**, 1–13.

58. Dodd KW, Guenther PM, Freedman LS, *et al.* (2006) Statistical methods for estimating usual intake of nutrients and foods: a review of the theory. *J Am Diet Assoc* **106**, 1640–1650.