Using exploratory factor analysis of FFQ data to identify dietary patterns among Yup'ik people

Tove K Ryman[1], Melissa A Austin[1,*], Scarlett Hopkins[2], Jacques Philip[2], Diane O'Brien[2], Kenneth Thummel[3] and Bert B Boyer[2]

[1]Department of Epidemiology, School of Public Health, University of Washington, Box 357236, Seattle, WA 98195, USA: [2]Center for Alaska Native Health Research, Irving I, University of Alaska Fairbanks, Fairbanks, AK, USA: [3]Department of Pharmaceutics, University of Washington, Seattle, WA, USA

Abstract

Objective: An FFQ developed by the Center for Alaska Native Health Research for studies in Yup'ik people includes market foods and subsistence foods such as moose, seal, waterfowl and salmon that may be related to disease risk. Because the FFQ contains >100 food items, we sought to characterize dietary patterns more simply for use in ongoing pharmacogenomics studies.
Design: Exploratory factor analysis was used to derive a small number of 'factors' that explain a substantial amount of the variation in the Yup'ik diet. We estimated factor scores and measured associations with demographic characteristics and biomarkers.
Setting: South-west Alaska, USA.
Subjects: Yup'ik people (*n* 358) aged ≥18 years.
Results: We identified three factors that each accounted for ≥10% of the common variance: the first characterized by 'processed foods' (e.g. salty snacks, sweetened cereals); the second by 'fruits and vegetables' (e.g. fresh citrus, potato salad); and the third by 'subsistence foods' (seal or walrus soup, non-oily fish). Participants from coastal communities had higher values for the 'subsistence' factor, whereas participants from inland communities had higher values for the 'fruits and vegetables' factor. A biomarker of marine intake, $\delta^{15}N$, was correlated with the 'subsistence' factor, whereas a biomarker of corn- and sugarcane-based market food intake, $\delta^{13}C$, was correlated with 'processed foods'.
Conclusions: The exploratory factor analysis identified three factors that appeared to reflect dietary patterns among Yup'ik based on associations with participant characteristics and biomarkers. These factors will be useful for chronic disease studies in this population.

Traditionally, the diet of Yup'ik people living in the Yukon–Kuskokwim Delta of south-west Alaska has included a variety of culturally important regional subsistence foods such as moose, seal, waterfowl and salmon, many of which are eaten seasonally, as well as foods purchased in the community market (i.e. small grocery store). In this study population, as in other indigenous populations, there is an ongoing nutritional transition from subsistence foods to more market-based foods[1–5]. Bersamin *et al.*[4] found that an average of 22% of energy intake of Yup'ik people was from traditional foods, with a higher percentage in older people. Increases in consumption of market-based foods among indigenous circumpolar populations may lead to increased prevalence of obesity and a variety of chronic diseases such as CHD and diabetes[6,7]. Conversely, the traditional marine-based diet of Yup'ik people typically includes high levels

of EPA and DHA, and may have strong beneficial effects in preventing chronic disease risk by lowering circulating TAG and inflammatory markers normally associated with obesity[8,9].

FFQ are a useful method for collecting food exposure data in large epidemiological studies because they are relatively easy and economical to administer[10], and they can capture with one interview foods that may only be eaten seasonally. This is particularly beneficial for the use in the Yup'ik population, where it can be difficult to access communities during all seasons of the year. Given that FFQ collect frequency of consumption for a large number of foods in order to characterize usual dietary patterns, methods are needed to reduce the amount of dietary data to a few underlying variables that can be used for both association studies between risk factors and health outcomes and to reduce participant burden during

Corresponding author: Email maustin@u.washington.edu

data collection. Factor analysis methods have been used to confirm hypothesized dietary factors[11] and to describe dietary patterns[12–15]. For example Hu *et al.* identified two major dietary patterns using factor analysis of a 131-item FFQ, which were then used for assessing the relative risk of CHD by differing diets[14].

The objectives of the present study were to identify dietary patterns using exploratory factor analysis and to use these factors to evaluate the differences in dietary patterns by demographic characteristics and dietary biomarkers among a Yup'ik population. Dietary data were collected in this study population of Yup'ik people for research on dietary factors and gene–diet interactions associated with chronic disease. The long-term goals of that research include assessing the association of genetic variants in drug-metabolizing enzymes of warfarin with clotting time and how dietary factors like vitamin K intake from leafy green vegetables and DHA/EPA intake from subsistence foods may modify these associations.

Methods

Study sample

A semi-quantitative FFQ developed by the Center for Alaska Native Health Research was administered as part of a community-based participatory research study in the Yukon–Kuskokwim Delta region of Alaska. Recruitment methods are published elsewhere[16,17]. Briefly, the 435 study participants, representing a convenience sample of Yup'ik people from nine communities – three inland (>5 miles from the coast) and six coastal – were recruited between September 2009 and September 2011. Sampling was longitudinal such that participants could have completed more than one FFQ. In this case, only the most recent FFQ was used for the current analysis. Participants could elect to be interviewed in Yup'ik by a native Yup'ik speaker trained to administer the FFQ. Participants were excluded from analysis if they self-reported their ethnicity as non-native Alaskans or 'other', were <18 years of age or if the FFQ was considered by the interviewer to be of 'poor' or 'unknown' quality.

Demographic data were also collected, including location of community (i.e. coastal or inland), participant age, household size, sex and cultural identification (as self-reported adherence to 'Kass'aq' (white) and Yup'ik lifestyle). Answers to cultural identification questions were consolidated into those participants reporting 'not at all' or 'some' in one group and 'a lot' into a second group. The cultural identification questions were not mutually exclusive, for example a participant could report 'a lot' for both the Yup'ik and Kass'aq lifestyle. Age was categorized into 'younger' and 'older' using the mean age (41·6 years) in the sample as the cut-off point.

The study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human subjects were approved by the University of Alaska Fairbanks Institutional Review Board and the Yukon–Kuskokwim Health Corporation Human Studies Committee. Informed consent was obtained prior to participation in the study.

Food frequency data

The FFQ was developed from a list of the most common foods consumed in this population and was derived from 24 h food recalls ($n \approx 2000$) collected in this region. The resulting FFQ was piloted and further reduced in size to decrease participant burden. The FFQ version used for the current analysis included 163 specific food questions (e.g. frequency of fresh citrus) grouped into fifteen categories (e.g. fruits, vegetables, ocean meats). Frequency of alcohol consumption was not included in the questionnaire as possession of alcohol is illegal in these communities. Participants reported how frequently they typically consumed each food during the previous 12 months and, for traditional subsistence foods, if they ate the food seasonally. Four frequency scales were used in the FFQ for the 163 food items. As described below, we used two of these scales that were the most amenable to factor analysis and that were used for most of the foods (132 of the 163 foods, 81 %).

Among the 132 food questions included in our analysis (Supplementary Materials, Supplemental Table 1), frequency of consumption was measured on a 9-point scale which differed for foods and beverages. For both scales the lowest frequency option was 'never or less than once per month', for foods the highest frequency option was '2+ times/d' and for beverages the highest frequency option was '6+ times/d'. After considering possible methodological approaches to quantify food frequency over the course of a year, we elected to convert the frequency of consumption from the 9-point scales to a continuous scale of annual consumption by multiplying the reported frequency of consumption to a 365 d scale (e.g. if the food was reportedly eaten weekly, then the annual consumption was 52·17) for analysis. If the frequency was reported as a range (e.g. eaten every 2–3 weeks), the mean of the range was used to calculate annual consumption. For the thirty-eight traditional subsistence foods typically eaten seasonally (e.g. wild bird eggs, wild greens, seal and walrus soup; see Supplemental Table 1 for complete list), it was elicited from the participant if he/she ate the food seasonally or year round. If the participant reported eating the food seasonally, the annual consumption was calculated as the product of the reported annual consumption and the proportion of the year that food was typically available (determined by local participants and referred to here as the 'seasonal weight'). Weights for all seasonal foods are provided in Supplemental Table 1.

For the factor analysis, annual food consumption values were transformed to the natural log scale to

improve normality of the frequency distributions. We assessed potential outliers as annual consumption values >2 SD from the mean, but retained all potential outlier values as they were judged realistic frequencies of consumption by two investigators (B.B.B., S.H.).

Biomarkers of food intake

We analysed two biomarkers of dietary intake for the Yup'ik study population in relation to the estimated factor scores. Because the nitrogen isotope ratio ($\delta^{15}N$) is elevated in traditional marine foods, red blood cell (RBC) $\delta^{15}N$ is strongly correlated with traditional marine intake[18] and intake of the *n*-3 PUFA EPA and DHA[19]. The carbon isotope ratio ($\delta^{13}C$) is elevated in market foods made from corn or sugarcane[18,20] and meat from corn-fed animals[18,21]. In the present Yup'ik study population, RBC $\delta^{13}C$ is most strongly associated with corn-based market foods[18]. We analysed these biomarkers as a way to assess how well the estimated factors from the FFQ characterize the dietary components measured with biomarkers. RBC samples were prepared and assessed for carbon and nitrogen stable isotope ratios at the Alaska Stable Isotope Facility as previously described[18,19]. By convention and for ease of interpretation, isotope ratios are presented as delta values in 'permil' relative to international standards: $\delta X = [(R_{sample} - R_{standard})/(R_{standard})] \times 1000$ ‰, where X is ^{15}N or ^{13}C, R is the ratio of heavy to light isotope, and the standards are atmospheric nitrogen and Vienna Pee Dee Belemnite for carbon. We concurrently prepared and ran multiple laboratory standards to assess analytical accuracy and precision; these were within 0·1 ‰ and 0·2 ‰, respectively, for both isotopes. The range of isotopic variation in our data set (4·1 for $\delta^{13}C$, 8·0 for $\delta^{15}N$) was large relative to analytical precision (0·2 ‰).

Factor analysis

As our objective was to estimate underlying dietary patterns among the Yup'ik study population, we used exploratory factor analysis. Factor analysis is a statistical approach that uses many correlated dietary components to derive a smaller number of latent variables or 'factors' that explain a substantial amount of the variation in diet. Exploratory factor analysis does not require *a priori* hypotheses about the data; rather, it is used to understand the underlying structure of a large set of variables. All analyses were conducted in the statistical software package SAS version 9·3, with the exploratory factor analysis procedure using the SAS maximum-likelihood parameter estimate and orthogonal varimax rotation. The number of factors retained in the analysis was based on retaining factors that accounted for >10 % of the common variance as well as interpretability. Figures were created using R (http://www.r-project.org/) with the default Gaussian kernel for smoothing. To reduce the 132 foods to a number that could be analysed using factor analysis

with our sample size, we developed a two-stage exploratory factor analysis process to identify relevant factors.

The objective of the first stage of the analysis was to reduce the number of food variables to less than thirty-six, i.e. less than ten per participant in our sample. We first assessed correlations between foods and conducted exploratory factor analyses within each of the fifteen food categories. In each of these categories, we then evaluated the three foods that had the highest loading for each factor. One food from among the top three foods loading on each factor was selected to be used in the final exploratory factor analysis, by two of the investigators (B.B.B., S.H.). Additionally, although it was not in the top three foods for the vegetable factor, the food 'wild greens' was selected to be included in the final exploratory factor analysis as it was particularly important to our vitamin K intake research questions. In total, twenty-two foods were selected for inclusion in our final factor analysis. The twenty-two food variables selected for the final factor analysis are noted in the last column of Supplemental Table 1 (see Supplementary Materials).

In the second stage, exploratory factor analysis was conducted using the twenty-two selected food variables. We qualitatively considered foods with a factor loading coefficient >0·40 to load highly on that factor, and these foods were used to provide the factor with a descriptive name. Using the factor loading coefficients, we calculated estimated factor scores for each factor for all participants by summing the product of the factor loading coefficient and the participants' reported annual consumption frequency for each food item. For a participant, a higher estimated factor score for a particular factor indicated a higher correlation with foods that had high loading coefficients on that factor. The estimated factor scores were assessed for normality using descriptive statistics and frequency histograms.

To measure associations between the estimated factors and demographic characteristics (i.e. community location, age, gender, cultural identification) for each of the estimated factors, we conducted *t* tests assuming unequal variance. We also assessed the associations using the non-parametric Wilcoxon–Mann–Whitney test for estimated factor scores which were not normally distributed, and the results were similar to the *t* test. Additionally, we conducted a multivariate linear regression analysis for each factor including all of the study participant characteristics (community location, age, sex, Kass'aq lifestyle and Yup'ik lifestyle) to assess independent associations. We measured associations between the dietary biomarkers and the estimated factor scores using Pearson correlations. To assess if the estimated factors were associated with the biomarkers independent of other participant characteristics, we conducted a multivariate linear regression analysis with the participant characteristics (listed above) and the estimated factor score as the independent variables. *P* values ≤ 0·05 were considered statistically significant for all analyses.

Results

Sample characteristics

FFQ data were available from a total of 435 participants, of whom sixty-eight were excluded based on age, ethnicity and quality of the interview. Of the remaining 367 participants, 358 had complete data for the selected foods used in the final factor analysis (Table 1). Among this sample, more participants were from coastal communities (58%) than inland communities (42%). Participants ranged in age from 18 to 91 years, with a mean age of 41·6 years. A larger proportion (58%) of the sample was female than male. Only 16% of the participants reported that they adhere 'a lot' (v. 'not at all' or 'some') to a Kass'aq lifestyle. Nearly half of participants reported 'not at all/some' adherence to Yup'ik lifestyle and the other half reported 'a lot'; trends were similar in coastal and inland communities. 'A lot' of self-reported adherence to Yup'ik lifestyle was associated with being in the older age group ($\chi^2 = 46·4$, $P < 0·001$).

Exploratory factor analysis

The estimated factor analysis within the fifteen food categories resulted in one factor per food category for eleven of the categories. For beverages, desserts and snacks, and fruits, two factors loaded, and for vegetables four factors loaded. Thus, the selection of one food from each factor plus the addition of wild greens resulted in twenty-two foods used in the factor analysis.

The distributions of annual consumption of the twenty-two foods selected for the final exploratory factor analysis were generally highly right-skewed, due to a large number of participants reporting consumption frequencies of 'never or less than once per month'. Transformation of values to the natural log scale improved normality considerably, as illustrated by skewness and kurtosis values (Table 2). That is, prior to the log transformation, skewness and kurtosis values were >1 for all foods; compared with after, when the absolute skewness or kurtosis value was ≤1 for fifteen and five foods, respectively. Pearson correlations between the twenty-two selected food variables were typically positive, with the strongest correlations between salty snacks and sweetened drinks ($r = 0·43$) and salty snacks and pizza ($r = 0·40$; Supplementary Materials, Supplemental Table 2).

The final exploratory factor analysis of the twenty-two selected foods resulted in three factors (Table 3). The three factors had eigenvalues >1·0 and each accounted for at least 10% of the common variance, whereas the nineteen factors not retained had eigenvalues <1·0. The first factor, with an eigenvalue of 5·93 and high loading coefficients (>0·40) for salty snacks, sweetened cereals, pizza, sweetened drinks and hot dogs and lunch meat, was labelled 'processed foods'. The second factor with an eigenvalue of 3·15 had high loading coefficients for fresh citrus, potato salad, citrus juice, corn, green beans and green salad, and was labelled 'fruits and

Table 1 Characteristics of participants with complete FFQ data for all items included in factor analysis: Yup'ik people (n 358) aged ≥18 years, south-west Alaska, USA, September 2009–2011

	All	Coastal communities	Inland communities
Sample size	358	208	150
Mean age (years)	41·6	41·9	41·1
Range	18–91	18–91	18–79
Average household size	6·0	6·0	5·9
Range	1–21	1–21	1–14
Sex (% female)	58·4	59·1	57·3
Kass'aq (white) lifestyle (%)*			
Missing	0·8	1·0	0·7
Not at all/some	83·5	79·3	89·3
A lot	15·6	19·7	10·0
Yup'ik lifestyle (%)*			
Missing	0·6	1·0	0·0
Not at all/some	48·3	50·0	46·0
A lot	51·1	49·0	54·0

*Self-reported.

vegetables'. Finally, the third factor, with an eigenvalue of 1·06 and high loading factors for seasonal foods, seal or walrus soup and non-oily fish (not dried), was labelled 'subsistence foods'. Estimated factor weights were approximately normally distributed for 'processed foods' and 'fruits and vegetables', while 'subsistence foods' had a bimodal distribution.

Estimated factor score associations

Participants residing in inland communities had a greater mean estimated score for 'fruits and vegetables' (Factor 2; $P < 0·001$), indicating higher frequency of consumption of foods with greater loading coefficients on these factors (Table 4 and Fig. 1(a)). In contrast, participants residing in coastal communities had a significantly greater mean score for 'subsistence foods' (Factor 3; $P < 0·001$; Fig. 1(c)). Younger participants (age <41·6 years) had a significantly greater estimated mean factor score for 'processed foods' (Factor 1; $P < 0·001$) and a significantly lower mean factor score for 'subsistence foods' (Factor 3; $P = 0·035$) compared with older participants (aged ≥41·6 years; Table 4). There were no significant differences in mean factor score between males and females for any of the factors. Reporting 'a lot' for Kass'aq lifestyle was significantly associated with a greater estimated mean factor score for 'processed foods' (Factor 1; $P = 0·001$). Conversely, participants reporting 'a lot' for Yup'ik lifestyle had a significantly lower estimated mean factor score for 'processed foods' (Factor 1; $P < 0·001$) and a significantly greater estimated mean factor score for 'subsistence foods' (Factor 3; $P = 0·025$; Table 4). All associations remained significant in multivariate models except the association between age and the 'subsistence foods' factor ($P = 0·374$). Completion of high school was not associated with any of the three factors (data not presented).

RBC δ^{15}N was negatively correlated with the 'processed foods' (Factor 1; $r = -0·492$, $P < 0·001$) and 'fruits and

Table 2 Characteristics of the annual food consumption distribution for twenty-two foods included in the factor analysis: Yup'ik people (*n* 358) aged ≥18 years, south-west Alaska, USA, September 2009–2011

	Untransformed values					Natural log-transformed values				
Foods	Mean	Median	SD	Skewness	Kurtosis	Mean	Median	SD	Skewness	Kurtosis
Salty snacks	55·38	30·00	88·85	3·01	12·32	1·63	3·40	3·71	−0·95	−0·77
Sweetened cereals	51·04	12·00	96·40	3·37	15·33	0·40	2·48	4·21	−0·26	−1·75
Pizza	31·22	12·00	72·56	6·04	49·40	0·05	2·48	4·02	−0·22	−1·80
Sweetened drinks	229·75	78·27	389·52	2·43	6·17	2·25	4·36	4·43	−0·76	−1·12
Hot dogs and lunch meat	29·91	12·00	62·12	5·86	50·79	0·40	2·48	3·92	−0·43	−1·67
Fried chicken	22·83	12·00	36·94	3·82	18·94	0·59	2·48	3·72	−0·62	−1·48
Canned tuna	8·69	0·01	28·03	7·79	81·93	−2·53	−4·61	3·43	1·09	−0·74
Fresh citrus	31·84	12·00	64·61	5·61	44·87	0·90	2·48	3·72	−0·72	−1·28
Potato salad	11·40	0·01	33·07	6·26	44·90	−1·75	−4·61	3·70	0·56	−1·60
Citrus juice	67·44	24·00	170·07	3·99	16·46	−0·06	3·18	4·40	0·01	−1·84
Corn	56·14	30·00	89·94	3·67	19·08	2·08	3·40	3·39	−1·29	0·12
Green beans	42·32	12·00	70·31	2·76	8·25	1·08	2·48	3·83	−0·70	−1·27
Green salad	8·88	0·01	27·55	6·17	45·96	−2·44	−4·61	3·46	1·01	−0·89
Pudding and jello	10·82	0·01	23·31	6·58	64·21	−1·30	−4·61	3·76	0·29	−1·86
Dried salmon	116·93	52·18	128·64	1·72	4·30	3·26	3·95	3·13	−1·82	2·09
Wild game soup*	89·48	41·09	106·76	2·48	9·88	3·39	3·68	2·36	−2·13	4·84
Pancakes	59·64	30·00	107·94	4·03	19·90	2·11	3·40	3·36	−1·31	0·23
Market berries in akutaq	19·44	0·01	54·45	7·95	87·67	−1·01	−4·61	3·94	0·23	−1·84
Seal or walrus soup*	23·86	12·00	40·49	3·49	15·06	0·60	2·48	3·66	−0·62	−1·40
Non-oily fish (not dried)*	20·63	3·96	42·45	4·38	24·88	−0·32	1·38	3·87	−0·12	−1·83
Wild greens*	10·03	0·01	30·64	7·48	71·68	−1·71	−4·61	3·54	0·51	−1·55
Bird soup*	40·96	30·00	51·53	2·81	9·93	2·47	3·40	2·70	−1·94	2·65

Values represent the annual frequency of consumption.
*Seasonal food items.

Table 3 Standardized factor loadings obtained from exploratory factor analysis, using varimax rotation, of annual consumption of FFQ items: Yup'ik people (*n* 358) aged ≥18 years, south-west Alaska, USA, September 2009–2011

Foods	Processed foods (Factor 1)	Fruits and vegetables (Factor 2)	Subsistence foods (Factor 3)
Salty snacks	**0·60**	0·25	−0·03
Sweetened cereals	**0·55**	0·06	−0·02
Pizza	**0·53**	0·21	−0·09
Sweetened drinks	**0·52**	0·13	0·04
Hot dogs and lunch meat	**0·46**	0·34	0·11
Fried chicken	0·36	0·29	0·07
Canned tuna	0·35	0·24	0·08
Fresh citrus	0·26	**0·54**	0·15
Potato salad	0·25	**0·47**	−0·01
Citrus juice	0·31	**0·47**	0·12
Corn	0·37	**0·46**	−0·04
Green beans	0·20	**0·45**	0·12
Green salad	0·22	**0·45**	0·04
Pudding and jello	0·25	0·35	0·07
Dried salmon	0·11	0·29	0·06
Wild game soup*	0·17	0·23	0·15
Pancakes	−0·01	0·21	0·10
Market berries in akutaq	0·17	0·20	0·03
Seal or walrus soup*	0·24	−0·13	**0·82**
Non-oily fish (not dried)*	−0·06	0·09	**0·47**
Wild greens*	−0·08	0·32	0·39
Bird soup*	0·00	0·16	0·35

Bold font indicates factor loadings >0·40 for interpretation.
*Seasonal food items.

vegetables' (Factor 2; $r = -0.144$, $P = 0.007$) factors and positively correlated with the 'subsistence foods' factor (Factor 3; $r = 0.252$, $P < 0.001$; Table 5). In comparison, RBC δ^{13}C was positively correlated with 'processed foods'

(Factor 1; $r = 0.364$, $P < 0.001$). δ^{13}C was not associated with either 'fruits and vegetables' (Factor 2) or 'subsistence foods' (Factor 3). Adjusting δ^{13}C for δ^{15}N did not improve the association between δ^{13}C and the 'processed foods' factor (Factor 1). With the exception of the association between RBC δ^{15}N and 'fruits and vegetables' (Factor 2; $P = 0.258$), all significant associations remained significant in the multivariate analysis assessing the associations adjusted for demographic characteristics.

Discussion

In the current exploratory factor analysis we identified three factors that characterized dietary patterns among Yup'ik people: 'processed foods', 'fruits and vegetables' and 'subsistence foods'. These three factors were easily interpretable and the associations were consistent with other studies of Native Alaskan diet[1,4,18,22]. Moreover, the factors were associated with participant characteristics and biomarkers, as hypothesized based on other studies. Specifically, we observed differences in factor scores between community locations, participant age, self-reported cultural identification and δ^{15}N and δ^{13}C biomarkers.

Greater mean factor scores for 'processed foods' were found in younger participants, those with self-reported greater adherence to Kass'aq lifestyle and those reporting lower adherence to a Yup'ik lifestyle in the bivariate analysis and additionally in inland communities in the multivariate analysis. A previously described marker of corn- and sugarcane-based market foods, δ^{13}C[18], was

Table 4 Mean natural log-transformed factor scores for each factor by characteristics of study participants: Yup'ik people (*n* 358) aged ≥18 years, south-west Alaska, USA, September 2009–2011

		Processed foods (Factor 1)			Fruits and vegetables (Factor 2)			Subsistence foods (Factor 3)		
	n	Mean	*P**	*P*†	Mean	*P**	*P*†	Mean	*P**	*P*†
Community location										
Coastal	208	−0·064	0·081	0·032	−0·227	<0·001	<0·001	0·245	<0·001	<0·001
Inland	150	0·089			0·315			−0·340		
Age										
Younger (<41·6 years)	186	0·412	<0·001	<0·001	−0·007	0·870	0·897	−0·089	0·044	0·374
Older (≥41·6 years)	172	−0·445			0·007			0·097		
Sex										
Female	209	−0·052	0·157	0·727	0·018	0·632	0·495	−0·033	0·400	0·286
Male	149	0·074			−0·025			0·046		
Kass'aq (white) lifestyle‡										
Not at all/some	299	−0·064	0·001	<0·001	−0·011	0·488	0·059	0·009	0·702	0·507
A lot	56	0·324			0·083			−0·037		
Yup'ik lifestyle‡										
Not at all/some	173	0·301	<0·001	<0·001	−0·042	0·326	0·419	−0·103	0·025	0·042
A lot	183	−0·286			0·044			0·104		

*Unadjusted *t* test assuming unequal variance; for Factor 3 the Wilcoxon–Mann–Whitney test *P* values were similar to *t* test *P* values.
†Adjusted for the other study participant characteristics using a multivariate linear regression model, *n* 354 due to missing values.
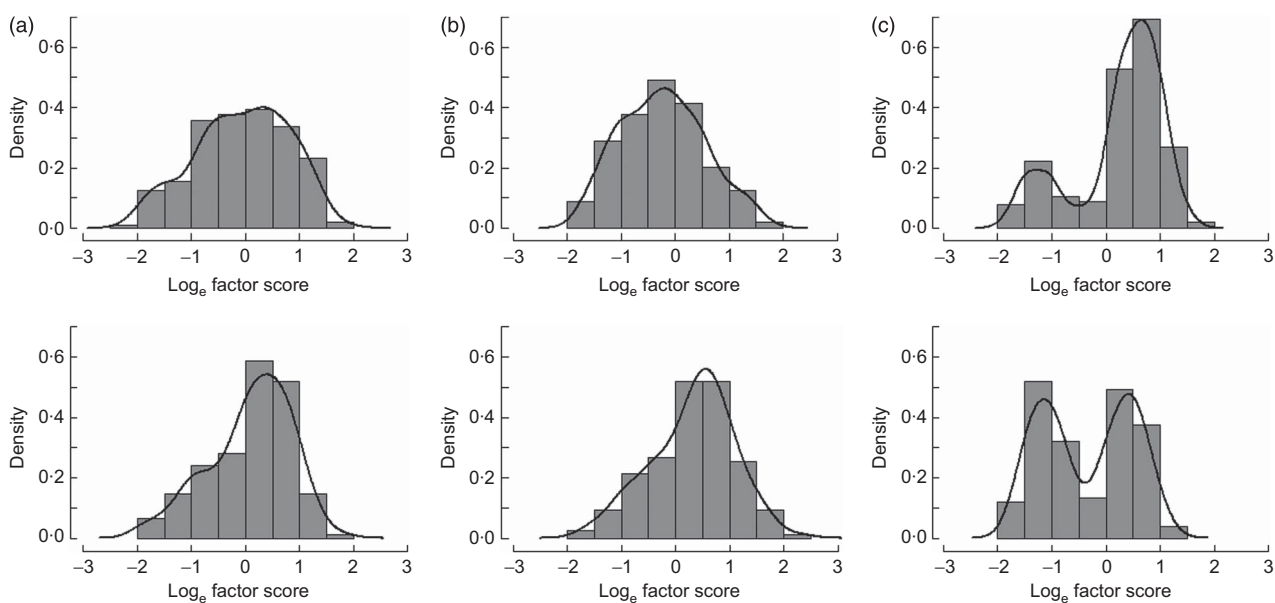‡Sample sizes vary due to missing values.

Fig. 1 Distribution of estimated factor scores by community location (top row; coastal; bottom row, inland): (a) 'processed foods' (Factor 1); (b) 'fruits and vegetables' (Factor 2): and (c) 'subsistence foods' (Factor 3). ── represents the Gaussian kernel density line; *x*-axis refers to the natural log-transformed estimated factor scores (▧). Yup'ik people (*n* 358) aged ≥18 years, south-west Alaska, USA, September 2009–2011

positively correlated with this factor, and δ^{15}N, a marker of traditional marine foods[18], had a strong negative correlation with this factor. Because these biomarkers were previously validated for this Yup'ik study population using 24 h food recalls and 3 d food records[18], the correlations between the biomarkers and the identified dietary pattern strengthen our confidence that the factor analysis approach is accurately characterizing the dietary patterns in this population.

The 'fruit and vegetable' factor was significantly associated only with residing in an inland community. This could partially be explained by the easier and more regular access to a regional hub of ~4000 people that has grocery stores closely resembling those found in a large city (e.g. having a full fresh produce and dairy section). Some of the foods included in the 'fruits and vegetables' factor are fresh (e.g. fresh citrus and green salad) and thus would not be easily available in more remote communities.

Table 5 Pearson correlations between RBC δ^{15}N and δ^{13}C and the estimated factor scores: Yup'ik people (*n* 356)* aged ≥18 years, southwest Alaska, USA, September 2009–2011

	Processed foods (Factor 1)			Fruits and vegetables (Factor 2)			Subsistence foods (Factor 3)		
	r	*P*	*P*†	*r*	*P*	*P*†	*r*	*P*	*P*†
RBC δ^{15}N	−0·492	<0·001	<0·001	−0·144	0·007	0·258	0·252	<0·001	0·027
RBC δ^{13}C	0·364	<0·001	<0·001	0·000	0·996	0·211	0·058	0·276	0·173

RBC, red blood cell.
*Biomarker data missing on two participants.
†Adjusted for the other study participant characteristics (community location, age, sex, Kass'aq lifestyle and Yup'ik lifestyle) using a multivariate linear regression model, *n* 352 due to missing values.

As expected, this factor was not highly correlated with either δ^{15}N or δ^{13}C.

The mean 'subsistence foods' factor score was significantly higher in coastal communities, in older participants and among those participants with greater adherence to a Yup'ik lifestyle. Other studies of diet in indigenous populations have reported similar trends of older participants consuming more subsistence foods[1,4,22]. However, older age was no longer significantly associated with the 'subsistence foods' factor after adjusting for other characteristics in the multivariate analysis, potentially due to the strong association between being older and adhering to a Yup'ik lifestyle. A biomarker of traditional marine intake (δ^{15}N) was correlated with this factor, as expected from previous studies[18]. Taken together, these three factors will be useful for characterizing Yup'ik diet for future studies in this population.

Our findings have potential implications for research on individualized medicine, where diet may be an important modulator of drug disposition or response. For example, we are currently investigating the impact of diet on anticoagulation response during warfarin therapy among Alaskan Native people. Both vitamin K and PUFA intakes in the diet may affect the function of the vitamin K cycle and warfarin response[23,24]. If quantitative information about overall dietary intake of DHA, EPA and vitamin K could be captured with a simplified FFQ and reliable factors representing dietary patterns identified, this could greatly enhance our ability to conduct a large-scale warfarin pharmacogenetic study in the future that includes an examination of gene–diet interactions. Indeed, data are already available supporting the use of FFQ for assessing vitamin K status in population-based studies[25]. A similar approach based on the present work supports use of an FFQ for evaluating DHA/EPA status as well.

The Yup'ik study population consumes many foods seasonally, particularly the foods included in the 'subsistence factor'. The FFQ was administered during different seasons and although the questionnaire asked about diet over the previous year, there may have been recall bias for seasonal foods not eaten during the season when the questionnaire was administered[26,27].

Moreover, for the foods eaten seasonally, it was challenging to identify the most appropriate way to account for this variability in diet for the analysis. We elected to use annual frequency weighted based on seasonal availability. However, this aggregation across seasons could modify some of the findings. For example, Receveur *et al.* found differences in dietary intake between sexes and age groups were more pronounced during seasons of peak traditional food consumption in sixteen indigenous communities in the Canadian Northwest Territories[22]. We might anticipate similar findings, particularly among the older participants who have greater mean levels of seasonal foods included in the 'subsistence' factor.

In the Yup'ik study population there may also be seasonality in consumption of non-subsistence foods due to other geographical and economic influences. In inland communities, travel to regional hubs can be restricted due to weather and ice conditions affecting access to market-based foods. In all communities, market-based food consumption (i.e. those foods in the 'processed foods' and 'fruits and vegetables' factors) is also associated with income and there is seasonality to the availability of income (e.g. Alaskan Permanent Fund payments) and support for purchasing of foods (e.g. food stamps, coupons for the Special Supplemental Nutrition Program for Women, Infants, and Children). These variations may result in seasonal fluctuation in consumption of fruits and vegetables, and to a lesser degree processed foods, and may have important implications for vitamin K levels and clotting time, as well as other diet and chronic disease associations. Additional work is needed to better capture seasonal variation in diet.

In addition to the challenge of accurately measuring seasonal variation in diet, the present study was subject to other limitations related to data collection and analysis. A convenience sampling method was used, possibly limiting the generalizability. Additionally, participants from the same household were not excluded, and thus diet data from all participants may not be independent. The questionnaire collected the frequency that foods were eaten, not actual consumption. For participants who ate a particular food very frequently, the FFQ was not designed to capture a frequency greater than two or more

times daily, and for beverages greater than six or more times daily. Because the 'subsistence foods' factor had a bimodal distribution, we conducted non-parametric association tests for this factor, and the results were found to be comparable to those using parametric methods. The bimodal distribution is the result of a high proportion of participants who reported eating the subsistence foods 'never or less than once per month', possibly reflecting a dietary transition away from subsistence foods to more Western foods. However, given the importance of describing this factor, we elected to retain foods even if they were not frequently eaten if they were in one of the top three loading factors for the food category. The FFQ used has not undergone a rigorous validation exercise. However, the correlations with biomarkers that have been validated in this population[18] strengthen our confidence in the validity, as the biomarkers have been rigorously evaluated and compared with actual fatty acid composition in RBC[19]. Although the specific dietary patterns identified in the current analysis are likely unique to this Yup'ik population, this methodology used could be generalized to other difficult-to-access or minority populations.

Based on these findings, the next steps in this research will be to collect additional FFQ data from a new set of participants and conduct a confirmatory factor analysis. If these results are found to be consistent, factor analysis will be used to reduce the number of FFQ questions asked, decreasing the expense of the interview and the time burden on participants, and thereafter the conducting of a rigorous validation study. In the present analysis the identified dietary factors were not adjusted for energy intake as our interest was in dietary patterns, not intakes of specific nutrients. However, for future analyses it may become important to adjust for energy intake.

Conclusions

We found that exploratory factor analysis identified three dietary factors – 'processed foods', 'fruits and vegetables' and 'subsistence foods' – which, based on the observed associations with participant characteristics and biomarkers, appear to be accurate reflections of dietary pattern among Yup'ik people. The factors identified through these analyses can be used on an ongoing basis to better understand the role of dietary factors in health disparities associated with chronic disease health outcomes in this study population and facilitate future pharmacogenetic studies.

Acknowledgements

Supplementary Materials

For Supplementary Materials for this article, please visit http://dx.doi.org/10.1017/S1368980012005411

References

1. Bersamin A, Luick BR, Ruppert E *et al.* (2006) Diet quality among Yup'ik Eskimos living in rural communities is low: the Center for Alaska Native Health Research Pilot Study. *J Am Diet Assoc* **106**, 1055–1063.
2. Kuhnlein HV, Receveur O, Soueida R *et al.* (2004) Arctic indigenous peoples experience the nutrition transition with changing dietary patterns and obesity. *J Nutr* **134**, 1447–1453.
3. Bersamin A, Luick BR, King IB *et al.* (2008) Westernizing diets influence fat intake, red blood cell fatty acid composition, and health in remote Alaskan native communities in the Center for Alaska Native Health Study. *J Am Diet Assoc* **108**, 266–273.
4. Bersamin A, Zidenberg-Cherr S, Stern JS *et al.* (2007) Nutrient intakes are associated with adherence to a traditional diet among Yup'ik Eskimos living in remote Alaska native communities: the CANHR Study. *Int J Circumpolar Health* **66**, 62–70.
5. Bjerregaard P, Young TK, Dewailly E *et al.* (2004) Indigenous health in the Arctic: an overview of the circumpolar Inuit population. *Scand J Public Health* **32**, 390–395.
6. McLaughlin JB, Middaugh JP, Utermohle CJ *et al.* (2004) Changing patterns of risk factors and mortality for coronary heart disease among Alaska natives, 1979–2002. *JAMA* **291**, 2545–2546.
7. Haman F, Fontaine-Bisson B, Batal M *et al.* (2010) Obesity and type 2 diabetes in Northern Canada's remote First Nations communities: the dietary dilemma. *Int J Obes (Lond)* **34**, Suppl. 2, S24–S31.
8. Makhoul Z, Kristal AR, Gulati R *et al.* (2011) Associations of obesity with triglycerides and C-reactive protein are attenuated in adults with high red blood cell eicosapentaenoic and docosahexaenoic acids. *Eur J Clin Nutr* **65**, 808–817.

9. Makhoul Z, Kristal AR, Gulati R *et al.* (2010) Associations of very high intakes of eicosapentaenoic and docosahexaenoic acids with biomarkers of chronic disease risk among Yup'ik Eskimos. *Am J Clin Nutr* **91**, 777–785.

10. Willett W (1990) *Nutritional Epidemiology*. New York: Oxford University Press.

11. Pierce BL, Austin MA, Crane PK *et al.* (2007) Measuring dietary acculturation in Japanese Americans with the use of confirmatory factor analysis of food-frequency data. *Am J Clin Nutr* **86**, 496–503.

12. van Dam RM, Rimm EB, Willett WC *et al.* (2002) Dietary patterns and risk for type 2 diabetes mellitus in US men. *Ann Intern Med* **136**, 201–209.

13. Hu FB, Rimm E, Smith-Warner SA *et al.* (1999) Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *Am J Clin Nutr* **69**, 243–249.

14. Hu FB, Rimm EB, Stampfer MJ *et al.* (2000) Prospective study of major dietary patterns and risk of coronary heart disease in men. *Am J Clin Nutr* **72**, 912–921.

15. Khani BR, Ye W, Terry P *et al.* (2004) Reproducibility and validity of major dietary patterns among Swedish women assessed with a food-frequency questionnaire. *J Nutr* **134**, 1541–1545.

16. Mohatt GV, Plaetke R, Klejka J *et al.* (2007) The Center for Alaska Native Health Research Study: a community-based participatory research study of obesity and chronic disease-related protective and risk factors. *Int J Circumpolar Health* **66**, 8–18.

17. Boyer BB, Mohatt GV, Lardon C *et al.* (2005) Building a community-based participatory research center to investigate obesity and diabetes in Alaska natives. *Int J Circumpolar Health* **64**, 281–290.

18. Nash SH, Bersamin A, Kristal AR *et al.* (2012) Stable nitrogen and carbon isotope ratios indicate traditional and market food intake in an indigenous circumpolar population. *J Nutr* **142**, 84–90.

19. O'Brien DM, Kristal AR, Jeannet MA *et al.* (2009) Red blood cell delta 15N: a novel biomarker of dietary eicosapentaenoic acid and docosahexaenoic acid intake. *Am J Clin Nutr* **89**, 913–919.

20. Jahren AH, Saudek C, Yeung EH *et al.* (2006) An isotopic method for quantifying sweeteners derived from corn and sugar cane. *Am J Clin Nutr* **84**, 1380–1384.

21. Jahren AH & Kraft RA (2008) Carbon and nitrogen stable isotopes in fast food: signatures of corn and confinement. *Proc Natl Acad Sci USA* **105**, 17855–17860.

22. Receveur O, Boulay M & Kuhnlein HV (1997) Decreasing traditional food use affects diet quality for adult Dene/Métis in 16 Communities of the Canadian Northwest territories. *J Nutr* **127**, 2179–2186.

23. Leray C, Wiesel M-L, Freund M *et al.* (2001) Long-chain *n*-3 fatty acids specifically affect rat coagulation factors dependent on vitamin K: relation to peroxidative stress. *Arterioscler Thromb Vasc Biol* **21**, 459–465.

24. Kim KH, Choi WS, Lee JH *et al.* (2010) Relationship between dietary vitamin K intake and the stability of anticoagulation effect in patients taking long-term warfarin. *Thromb Haemost* **104**, 755–759.

25. McKeown NM, Jacques PF, Gundberg CM *et al.* (2002) Dietary and nondietary determinants of vitamin K biochemical measures in men and women. *J Nutr* **132**, 1329–1334.

26. Fowke JH, Schlundt D, Gong Y *et al.* (2004) Impact of season of food frequency questionnaire administration on dietary reporting. *Ann Epidemiol* **14**, 778–785.

27. Subar AF, Frey CM, Harlan LC *et al.* (1994) Differences in reported food frequency by season of questionnaire administration: the 1987 National Health Interview Survey. *Epidemiology* **5**, 226–233.