

## Racial/Ethnic Heterogeneity in Diet of Low-income Adult Women in the United States: Results from National Health and Nutrition Examination Surveys 2011-2018

Briana Joy K. Stephenson<sup>1</sup>, Walter C. Willett<sup>2</sup>

### Affiliations

1. Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA (BJKS)
2. Departments of Nutrition and Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA (WCW)

### ABSTRACT

*Background.* Poor diet is a major risk factor of cardiovascular and chronic diseases, particularly for low-income women. However, the pathways by which race/ethnicity plays a role in this risk factor have not been fully explored.

*Objective.* This observational study aims to identify dietary consumption differences by race/ethnicity of US women living at or below the 130% poverty income level from 2011-2018.

*Design.* A total of 3005 adult women aged 20-80 years from the National Health and Nutrition Examination Survey (2011-2018) living at or below the 130% poverty-income level with at least one complete 24-hr dietary recall were classified into 5 self-identified racial/ethnic subgroups (Mexican, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Asian). Dietary consumption patterns were defined by 29 major food groups summarized from the Food Pattern Equivalents Database and derived via a robust profile clustering model which identifies foods that share consumption patterns across all low-income adult women, and foods that differ in consumption patterns based on race/ethnic subgroups.

*Results.* Legumes (protein and vegetable) were the most differentiating foods identified across all racial/ethnic subgroups and were primarily consumed by Mexican and Other Hispanic women. Non-Hispanic Asian women were most likely to favor a high consumption of prudent foods (fruits, vegetables, whole grains). Non-Hispanic White and Black women shared the most similarities in consumption patterns but differed in foods such as milk, poultry, and eggs.

*Conclusions.* Differences among consumption behaviors of low-income women were found along racial/ethnic lines. Efforts to improve nutritional health of low-income adult women should consider racial/ethnic differences in diet to appropriately focus interventions.

*Keywords:* dietary patterns, low-income, robust profile clustering, NHANES, women, race/ethnicity.

**Short Running Title:** Diet Heterogeneity of Low-income US Adult Women

**Abbreviations:** CVD, cardiovascular disease; SES, socioeconomic status; RPC, robust profile clustering; NHANES, National Health and Nutrition Examination Survey; FNDDS, Food and Nutrition Database for Dietary Studies; LDL, low density lipoprotein; BMI, body mass index; MCMC, Markov Chain Monte Carlo; LCA, latent class analysis; SNAP, Supplemental Nutrition Assistance Program; WIC, women, infants, and children.

**Disclaimers:** N/A.

**Sources of Support:** Study supported in part by National Heart Lung and Blood Institute (NHLBI) grant R25 (HL105400) to Victor G. Davila-Roman and DC Rao.

## 1 INTRODUCTION

2 Diet is one of the largest contributing factors to cardiovascular and other chronic  
3 diseases (Mokdad, 2018). The impact of cardiovascular disease (CVD) related outcomes on  
4 women aged 35-54 has increased in the past twenty years, while decreasing among other  
5 subgroups (Kuehn, 2019; Kalinowski, 2019; Thompson, 2017; Van Dyke, 2018). This gender  
6 disparity has been suggested to result from differences in diagnosis, care, and research  
7 specifically focused on women populations (Kuehn, 2019). CVD research that ignores gender  
8 differences create misleading extrapolation and a gap in understanding women-specific risk  
9 factors. Amidst this gender gap, racial and economic disparities have also continued to persist.  
10 Reportedly, 45% of Black women have some type of CVD, compared to 32% of white women  
11 (Williams, 2009). As disparities in CVD health continue to widen, the pathways through which  
12 factors like gender or race/ethnicity contribute to these differences are still not fully understood  
13 (Winkleby, 1999).

14 Dietary intake and behaviors has been widely studied in an attempt to identify  
15 modifiable pathways for improving overall health and preventing chronic diseases. Many  
16 studies have focused on dietary quality, using standardized adherence scores (e.g. Healthy  
17 Eating Index, Mediterranean Diet Score) (Hu, 2020; Onvani, 2017; Pate, 2015). While others  
18 have focused on individual foods/nutrients to examine diet-disease relationships (Basu, 2013;  
19 Drewnowski, 2013).

20 Foods and/or nutrients are commonly consumed together and share an intercorrelation  
21 structure, and biological synergy may exist among dietary components. For these reasons,  
22 dietary patterns can be a useful way to define exposure in studies of diet–disease  
23 relationships, complementing those of specific foods or nutrients (Hu, 2000; Schulze, 2018).

24 Defining patterns by consumption of specific foods can also provide a simplified way to identify  
25 dietary behaviors of subgroups.

26 Latent class models are an easy and interpretable technique to identify dietary patterns,  
27 based on the consumption or nonconsumption of a large set of foods. However, this technique  
28 relies on a population composition where the single differentiating behavior present is diet.  
29 Populations that contain a mixture of demographics (e.g., culture, gender, SES) that may  
30 mediate differences in dietary behaviors pose challenges to the standard structure of latent  
31 class models. If a certain mediating demographic comprises the majority of the population,  
32 behaviors reflected in those dietary patterns will be dominated by the larger demographic.  
33 Smaller-sized demographics may have differing features, but are often overlooked and treated  
34 as noise. The focus of analysis is controlled by the identifiable behaviors present in the larger-  
35 sized demographic. For example, in many national surveys and cohorts of adults in the United  
36 States, the study population is often a majority of non-Hispanic white participants living above  
37 the 130% poverty income level. Modeling dietary behaviors from study populations sharing this  
38 makeup will yield behaviors reflective of this demographic, and prevent us from better  
39 understanding factors impacting the populations at greatest risk for health outcomes  
40 (racial/ethnic minority and low-income) (Stephenson, 2021; Field, 2007; Gavin, 2011)

41 More flexible model approaches have recently been introduced to improve the way we  
42 consider demographic features that may drive dietary behaviors. Robust Profile Clustering  
43 (RPC) is an extension to the latent class model that distinguishes consumption patterns that  
44 may be shared across the study population, and those that may be specific to a defined  
45 demographic. The technique has previously been applied to identify differences in maternal  
46 diet by geography, as well as adult diet of Hispanics/Latinos based on cultural background and

47 US residency (Stephenson, 2019; Stephenson, 2020). With a focus on low-income adult  
48 women, this paper aims to apply the RPC model to identify racial/ethnic differences amongst  
49 an at-risk demographic.

## 50 **METHODS**

### 51 *National Health and Nutrition Examination Survey*

52 National Health and Nutrition Examination Survey (NHANES) is a publicly available  
53 population-based repeated cross-sectional survey. Approximately 9,000 people are sampled  
54 annually from 15 unique counties of varying socioeconomic and racial/ethnic backgrounds.  
55 Survey sampling weights are provided to generate population-based estimates of collected  
56 measures.

57 Our analysis focused on low-income female participants aged 20 and over who self-  
58 identified as Mexican-American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black,  
59 Non-Hispanic Asian. Participants who identified as Mixed/Other or missing race/ethnicity  
60 information were excluded. Low-income was defined as those reporting a ratio of family  
61 income to poverty level at or below the 130%. We pooled four survey cycles for analysis  
62 (2011-2012, 2013-2014, 2015-2016, 2017-2018). Sampling weights were adjusted for pooled  
63 analysis in accordance with NHANES survey methods and guidelines (Johnson, 2014; Chen,  
64 2018; Chen, 2018).

65 Dietary intake was measured using two 24-hour dietary recalls for each participant,  
66 collected initially in person, and the second recall was collected via telephone three to ten days  
67 later. Participants with at least one complete dietary recall were included for analysis.  
68 Participants with two complete recalls available, were averaged over the two days. Recalls  
69 were collected as part of the What We Eat in America survey component of NHANES (Bodner-

70 Montville, 2006). Nutrient composition for all foods and beverages reported were calculated  
71 using the Food and Nutrition Database for Dietary Studies (FNDDS). These nutrients were  
72 then converted into food pattern equivalents in accordance with the Dietary Guidelines for  
73 Americans (Bowman, 2014; Bowman, 2017; Bowman, 2018; Bowman, 2020). The food pattern  
74 equivalents database summarized the diet data into 29 major food groups. Levels of  
75 consumption for each food group were categorized into four levels, no consumption, low  
76 consumption (lower tertile of positive consumption), medium consumption (middle tertile of  
77 positive consumption), high consumption (upper tertile of positive consumption), where tertiles  
78 were calculated based on the overall population (Liu, 2019; Sotres-Alvarez, 2010).

79 CVD risk factors were treated as binary outcomes. High cholesterol was defined as  
80 participants with more than 200 total cholesterol or LDL cholesterol > 150 mg/dL or self-report  
81 taking cholesterol lowering medication. Obesity was defined as participants with a BMI > 30  
82 kg/m<sup>2</sup>. Hypertension was defined as participants with an average blood pressure reading of  
83 systolic blood pressure greater than 140 mg/dL, diastolic blood pressure greater than 90  
84 mg/dL, or self-reported medication use of blood pressure lowering medication. Diabetes is  
85 defined as participants with >126 mg/dL or self-reported diabetes medication use. Current  
86 smoker is defined as those who responded yes to the question “SMQ040: Do you now smoke  
87 cigarettes?”. Participants with complete diet data but missing one or more CVD risk factor data  
88 were still included for analysis but not included in the descriptive reporting for that missing risk  
89 factor.

### 90 *Robust Profile Clustering*

91 Robust Profile Clustering (RPC) is a flexible extension of the latent class model. First  
92 introduced in Stephenson et al., the model breaks apart the assumption that all participants

93 assigned to the same latent class (diet pattern) will share the same consumption behaviors for  
94 all food items observed (Stephenson, 2019). Alternatively, the RPC allows participants who  
95 share overall consumption patterns of a subset of foods to assume a global pattern, and  
96 participants who share unique consumption patterns within a predefined subgroup of the  
97 population assume a localized pattern.

98 The model assumes a probabilistic framework comprised of three components:

- 99 1) *Global pattern assignment*. This is consistent with the standard latent class model,  
100 where a global pattern describes the latent dietary behaviors shared amongst the  
101 overall population. The assignment of the global pattern is determined through a  
102 probability vector that describes the probability of a person being assigned to one global  
103 pattern over another.
- 104 2) *Local pattern assignment*. Within each subpopulation there exists another set of latent  
105 dietary behaviors that share attributes specific to their subpopulation, referred here as  
106 local patterns. The assignment of the local pattern is determined through a probability  
107 vector that describes the probability of a person from this subpopulation being assigned  
108 to one local pattern over another.
- 109 3) *Global/Local indicator*. Each food item used to describe each respective pattern has a  
110 probability of assuming the pattern detailed at the global level or the local level. This  
111 indicator, which is unique for each participant in the study population, is determined  
112 through a probability describing the likelihood of this food item to assume a global  
113 versus local pattern for that individual. The lower the probability the more likely it is to  
114 assume a local pattern. The higher the probability the more likely it is to assume a  
115 global pattern.

116 Each dietary pattern (global or local) is described with a probability matrix of  $p$  food item rows,  
 117 where each row describes the probability of consumption at each of the  $d$  possible  
 118 consumption levels. The modal pattern for each profile can be identified by the consumption  
 119 level category with the highest probability value in that row.

120 We can describe the model mathematically, using the following notation. Let  $y_i =$   
 121  $(y_{i1}, \dots, y_{ip})$  denote the self-reported levels of consumption of  $p$  unique food/beverage items by  
 122 individual  $i \in (1, \dots, n)$ , where  $n$  is the total number of participants in the study population. Let  
 123  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_{K_0})$  denote the probability vector for global pattern assignment, and  $\boldsymbol{\lambda}^{(s_i)} =$   
 124  $(\lambda_1^{(s_i)}, \dots, \lambda_{K_s}^{(s_i)})$  denote the probability vector of local pattern assignment where  $s_i \in (1, \dots, S)$   
 125 indicates the subpopulation index of individual  $i \in (1, \dots, n)$ . Let  $\boldsymbol{\theta}_{0j \cdot | h} = (\theta_{(0j1|h)}, \dots, \theta_{(0jd|h)})$   
 126 denote the consumption  $d$ -length probability vector for food item  $j \in (1, \dots, p)$ , given  
 127 assignment to global pattern  $h$ . Let  $\boldsymbol{\theta}_{1j \cdot | l}^{s_i} = (\theta_{1j1|l}^{s_i}, \dots, \theta_{1jd|l}^{s_i})$  denote the consumption probability  
 128 vector for food item  $j \in (1, \dots, p)$ , given assignment to local pattern  $l$  within subpopulation index  
 129  $s_i \in (1, \dots, S)$ . Let  $G_{ij}$  denote the global/local indicator variable for individual  $i \in (1, \dots, n)$  and  
 130 food item  $j \in (1, \dots, p)$ . Finally, let  $K_0$  and  $K_s$  denote the number of global and local patterns,  
 131 respectively. The subject-specific likelihood of the RPC model can be described as

$$132 \quad f(y_i | s_i, \boldsymbol{\pi}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_1^{(s_i)}, \boldsymbol{\lambda}^{(s_i)}, G_{ij}) = \sum_{h=1}^{K_0} \pi_h \prod_{j=1}^p \theta_{(0jr|h)}^{1(y_{ij}=r)} \prod_{j=1}^p \sum_{l=1}^{K_s} \lambda_l^{(s_i)} \prod_{r=1}^d (\theta_{1jr|l}^{s_i})^{1(y_{ij}=r)}$$

133  
 134 Estimation of the parameters of this model can be performed using a Bayesian approach. We  
 135 fit the model using a Gibbs sampling algorithm, with conditional posterior distributions  
 136 described previously (Stephenson, 2020). Noninformative, flat priors were selected to let the  
 137 observed data drive model estimation. The number of global and local pattern were unknown a  
 138 *priori*. As a result, we overfit the RPC model with 30 global and local patterns each so that



139 when run using Markov chain monte carlo (MCMC), an interpretable set of nonempty global  
140 and local patterns would remain (Van Havre, 2015). Posterior computation, MCMC  
141 diagnostics, prior sensitivities, and convergence were performed as described previously  
142 (Stephenson, 2019; Stephenson, 2020). Data was preprocessed using SAS 9.4 (Cary Institute,  
143 2013). RPC model was analyzed in Matlab 2021a. Survey sampling weights accounting for  
144 NHANES study design were incorporated *post hoc* and summarized using the survey package  
145 in R 4.0. All data and code to replicate analysis has been made publicly available on GitHub  
146 repository (<https://github.com/bjks10/RPC/tree/master/NHANES>).

## 147 **RESULTS**

148 A total of 3005 female adult participants living at or below the 130% family poverty  
149 income level were included for analysis of this study (**Figure 1**). Demographic information of  
150 these participants are provided in **Table 1**. Participants were mostly between the ages of 20-34  
151 years of age (32.8%), self-identified as Non-Hispanic White (49.9%), married or living with a  
152 partner (45.9%), at least some college (43.6%). Most participants reported having high  
153 cholesterol.

154 Local patterns were defined by five racial/ethnic subgroups. The demographic  
155 characteristics of these racial/ethnic subgroups are described in Table 1. Non-Hispanic Black  
156 women were more likely to be never married, whereas all other racial/ethnic subgroups were  
157 more likely to be married or living with a partner. Mexican and Other Hispanic women were  
158 more likely to have less than a high school education. The other three racial/ethnic subgroups  
159 were more likely to have at least some college education.

160 *Global Dietary Profiles*

161 Five global patterns were derived. **Figure 2** illustrates the modal pattern of consumption  
162 for each profile. Global profiles 1 and 2 have the most similarities, sharing consumption modes  
163 for 22 of 29 food items. Both profile patterns favor a high intake of cheese, oils, fats, and  
164 sugars, as well as grains, other vegetables, and legumes. They favor no consumption of  
165 potatoes, other starchy vegetables, soybean products and alcohol. Where they differ is in the  
166 consumptions of meats and fruits. Global profile 1 favored a higher consumption of fruits,  
167 juices, high n-3 seafood, poultry, and green vegetables. Global profile 2 favored a higher  
168 consumption of organ meat and other red/orange vegetables. Global profiles 3 and 4 shared  
169 14 consumption modes. These two profile patterns both favored a medium intake of other  
170 fruits, dark green vegetables, soybean products, oils, and sugars. However, global profile 3  
171 favored a higher consumption of tomatoes, refined grains, eggs, and cheese. Global profile 4  
172 comparatively favored a high consumption of organ meat. Global profile 5 had the least diet  
173 diversity favoring consumption of only 7 foods. Of these, poultry and other starchy vegetables  
174 favored a high level of consumption, while the remaining five foods (refined grains, soybean,  
175 oils, solid fats, and added sugar) favored a low level of consumption.

176 The distribution of these global patterns amongst the study population is provided in  
177 **Table 2**. Of immediate note is that global profile 5 contains about 70 percent of the population.  
178 This is likely due to the more populous racial/ethnic subgroups, Non-Hispanic White and Black  
179 women being assigned to this group. Mexican women were more likely to be assigned to  
180 global profile 2. Other Hispanic women were more likely to be assigned to global profile 3.  
181 Non-Hispanic Asian women were more likely to be assigned to global profile 4. Women in  
182 global profile 1 were more likely to be married or living with a partner and have less than a high  
183 school degree and had the largest average household size. Women assigned to global profile

184 2 averaged the highest healthy eating index score, whereas those assigned to global profile 5  
185 averaged the lowest.

### 186 *Racial/Ethnic-specific Local Patterns*

187 As mentioned before, under the RPC model, women assigned to the global patterns are  
188 not beholden to all the consumption patterns described at the global level. To determine which  
189 food items assume which pattern, we must consider **Figure 3**. This figure provides a  
190 probability heatmap of a food item (y-axis) assuming the global pattern for each race/ethnicity  
191 subgroup (x-axis). Two food items had a higher probability for assuming a global pattern,  
192 legumes (protein) and legumes (vegetable). Mexican and Other Hispanic women had over  
193 94% probability of assuming a global pattern. Recalling that the Mexican and Other Hispanic  
194 groups were more likely to be assigned to global profiles 2 and 3 we note that the mode of  
195 consumption for legumes in these profiles was a high and medium consumption level,  
196 respectively. Similarly, Non-Hispanic Asian women had a 74% probability of assuming the  
197 global patterns. These women had the largest representation in global profile 4, which had a  
198 high proportion of a low consumption of these food items. Both Non-Hispanic White and Black  
199 women were split in whether they would assume the global or local pattern. Focusing on their  
200 most representative global profile 5 consumption mode, and that for each of these subgroups  
201 at the local level, we see that they were most likely to not consume these food items at all.

202 Given that the remainder of the food items were most strongly associated with a local  
203 pattern, we examine the locally stratified patterns for each of these racial/ethnic subgroups.  
204 **Figure 4** provides a full distribution plot of the pattern of each of the observed foods given  
205 assignment to each local profile. Each of the racial/ethnic subgroups had a single local profile

206 to explain the different probabilities of consumption. Consumption modes are identifiable by  
207 the vertical bar of relative height.

208 Added sugars, solid fats, refined grains, and oils were consumed by all racial/ethnic  
209 subgroups. However, Non-Hispanic Asian women were the only subgroup that had some  
210 women who did not consume these foods at all during the two recall days of record. Mexican  
211 women were more likely to favor a high consumption of other fruit, refined grains, eggs, and  
212 milk, with a medium consumption of other vegetables and tomatoes. Other Hispanic women  
213 showed similar patterns to Mexican women, but favored a lower consumption of tomatoes,  
214 other vegetables, refined grains, and eggs. Non-Hispanic Black and Non-Hispanic White  
215 women showed the most similarities in their patterns. These two subgroups favored a low  
216 consumption of tomatoes, other vegetables, refined grains, cheese, and oils. Comparatively,  
217 Non-Hispanic Black women favored a high consumption of eggs (medium vs low), whereas  
218 Non-Hispanic White women favored a higher consumption of milk (high vs low) and solid fats  
219 (medium vs low). Non-Hispanic Asian had the most unique dietary pattern compared to the  
220 other four, with a high consumption of other fruit, other vegetables. While meats appear to  
221 have a mode of non-consumption for all groups, we see that Non-Hispanic Asian women were  
222 least likely to consume cured meats across subgroups, and most likely to consume non-  
223 specified meat and seafood. At the high consumption level, poultry and seafood (low-n3) were  
224 most likely to be consumed by Non-Hispanic Black women.

225 These local dietary patterns are consistent with the healthy eating index scores  
226 provided in Table 1. Average energy intake was highest amongst Mexican women and lowest  
227 amongst non-Hispanic Asian women. Non-Hispanic Asian women, whom had the highest  
228 probability of high consumption of foods like whole grains, other red/orange vegetables, and no

229 consumption of cheese, milk, oils, solid fats, and added sugars, had the highest HEI score  
230 (57.4 +/- 1.5), and the lowest proportion of all five CVD risk factors. Both Non-Hispanic White  
231 and Black women had HEI scores lower than 50 with favored high consumption of starchy  
232 vegetables, refined grains, oils, and added sugars, and shared the highest proportion of  
233 women with high cholesterol. Non-Hispanic Black women also had the highest proportion of  
234 women with hypertension and obesity. Diabetes was most prevalent among Other Hispanic  
235 women.

## 236 **DISCUSSION**

237 The application of this model allowed a deeper look into the consumption behaviors of  
238 adult women in the United States that live at or below the 130% poverty income level, and how  
239 those behaviors may differ by race/ethnicity. The standard latent class model relies on a  
240 homogeneous population where consumption habits may differ. Non-Hispanic White women  
241 are the larger racial/ethnic demographic. Consequently, this demographic and their  
242 consumption behaviors overpower the patterns reflected across all five global diet patterns.  
243 The RPC generated a more accurate identification of racial/ethnic dietary differences amongst  
244 these women.

245 Legumes (protein and vegetable) were the two most differentiating food features that  
246 were reflected across the five global patterns. Consumption of legumes (beans) has been  
247 associated with a myriad of nutritional benefits on lowering cardiovascular disease risk  
248 (Winham, 2009; Mattei, 2014; Winham, 2007; Finley, 2007; Bazzano, 2011; Thompson, 2012;  
249 Mitchell, 2009). However, the consumption patterns of these foods, continue to fall along  
250 cultural lines. Mexican and Other Hispanic women, where beans are a cultural staple of the  
251 home, were likely to have a high and medium consumption of legumes (Global patterns 1-3),

252 respectively. Non-Hispanic Asian women, who had a higher representation in global pattern 4,  
253 were likely to have a low consumption. Non-Hispanic Black and White women, who made up  
254 the largest representation of the participants assigned to global pattern 5, were likely to not  
255 consume legumes at all.

256         Locally, we were able to see the consumption distribution of the remaining 27 food  
257 items by race/ethnicity. Unhealthy food items (refined grains, solid fats, added sugars) shared  
258 some level of consumption across all racial/ethnic subgroups, except for Non-Hispanic Asian  
259 women. Despite poorer expected outcomes and risk factors expected amongst low-income  
260 adults, Non-Hispanic Asian women had the healthiest of outcomes and lower prevalence of all  
261 CVD risk factors except cholesterol. The diets reflected in this subgroup had a higher  
262 proportion of high consumers of fruits, vegetables, and whole grains. Mexican and Other  
263 Hispanic women shared several similarities in consumption patterns across most foods, with a  
264 higher consumption by Mexican women of refined grains, tomatoes, eggs, and cheese. Non-  
265 Hispanic White and Black women also shared several similarities in consumption patterns, with  
266 higher consumption of poultry and eggs by Non-Hispanic Black women and lower consumption  
267 of dairy (milk, cheese).

268         Previous studies who examined the diets of low-income populations, have focused  
269 analysis on adherence scores to examine diet quality, emphasizing consumption of fruits,  
270 vegetables, and whole grains (Lin, 2005; Leung, 2012). Adjustments for intake or survey  
271 weighted means are adjusted using regression analysis focusing on specific foods or nutrients  
272 on independent models. Our analysis utilized the RPC model to examine the combination of all  
273 foods summarized by the dietary intake tool, under a latent class approach. Additionally,  
274 findings of prior studies indicated an overall lower intake of fruits and vegetables, whereas our

275 joint-stratified model was able to identify certain racial/ethnic groups where healthier eating  
276 habits are present.

277 This analysis highlights one of the methodological strengths of the RPC model: 1) the  
278 ability to reduce the number of models to better understand the heterogeneity; 2) generating a  
279 joint-stratified latent class structure with a global LCA identifying patterns shared across  
280 multiple subgroups, and a stratified LCA to capture local patterns for each subgroup; 3)  
281 implementing an overfitted latent class model to determine the appropriate number of diet  
282 patterns, globally and locally in one single model.

283 With expansion of NHANES 2011 data collection, starting in 2011, our analysis was  
284 able to leverage additional racial/ethnic details (e.g. Non-Hispanic Asian, Mexican vs Other  
285 Hispanic) not previously available in prior NHANES diet research. This allowed a more in-  
286 depth analysis of dietary patterns shared across all low-income women, and across multiple  
287 racial/ethnic subgroups.

288 As with most models, the output is only as good as the input. Dietary patterns reflected  
289 in this model are based on the consumption amounts reported on two 24-hour dietary recalls  
290 by NHANES participants. Measurement error was not accounted for in this model, so cases of  
291 underreporting and overreporting are possible and patterns should be considered with  
292 potential misclassification in mind. As a result, the associations between race/ethnicity and diet  
293 patterns reported are likely to have been understated.

294 Food choices made by low-income women are often influenced by cost (Wiig, 2009).  
295 Federal assistance programs, such as Supplemental Nutrition Assistance Program (SNAP)  
296 and special SNAP for Women, Infants, and Children (WIC), have been created to assist in  
297 providing affordable options, but the selection of foods purchased and consumed may not be

298 motivated with nutritional health in mind (Wardle, 2000). Wang et al. examined diet quality of  
299 adults from NHANES 1999-2010 and found that while income disparities accounted for  
300 differences in Non-Hispanic Black and White populations, it did not for other racial/ethnic  
301 subgroups (e.g. Mexican-Americans) (Wang, 2014). In other racial/ethnic subgroups, cultural  
302 influence can play an even more significant role in both diet and participation in federal  
303 assistance programs. For example, our study found that Non-Hispanic Asian women had the  
304 highest adherence to healthy habits and the lowest participation in SNAP or WIC federal  
305 assistance programs. The reasons behind these dietary differences and choices remain  
306 complex and open for further research.

307         This approach effectively identified heterogeneity in diet quality and foods consumed  
308 among subgroups of low-income women, defined by race/ethnicity, that have previously not  
309 been fully appreciated in overall population diet analysis. With a more recent focus on how to  
310 improve nutrition amongst federal assistance programs for low-income populations,  
311 approaches such as the RPC, may be of value for targeting efforts to improve diet quality  
312 within these subgroups and for a better understanding of health disparities overall.

313



314 **ACKNOWLEDGEMENTS**

315 The authors thank the study participants of NHANES for their important contributions. The  
316 authors responsibilities were as follows: BJKS designed the research, analyzed, interpreted  
317 the data, wrote the manuscript, and had primary responsibility of content. WCW assisted with  
318 data interpretation, and critical revisions of the manuscript. All authors read and approved the  
319 final manuscript. This study was supported in part by NHLBI grant R25 HL105400 to DC Rao  
320 and Victor G. Davila-Roman.

321

322 REFERENCES

323

- 324 1. Mokdad AH, Ballestros K, Echko M, et al. The state of US health, 1990-2016: burden of  
325 diseases, injuries, and risk factors among US states. *JAMA*. 2018;319(14):1444-1472.  
326 PMID: PMC5933332
- 327 2. Kuehn, BM; State of the Heart for Women: Leaders Highlight Advances in Women's Heart  
328 Care at the American Heart Association, 2019.
- 329 3. Kalinowski J, Taylro JY, Spruill TM. Why are young black women at high risk for  
330 cardiovascular disease? *Circulation*, 2019;139(8):1003-1004.
- 331 4. Thompson LE, Daugherty SL. Gender disparities in cardiovascular disease prevention.  
332 *Heart*, 2017;103(7):479-480.
- 333 5. Van Dyke M, Greer S, Odom E, Schieb L, Vaughan A, Kramer M, Casper M. Heart disease  
334 death rates among blacks and whites aged  $\geq 35$  years — United States, 1968–2015.  
335 *MMWR Surveill Summ*. 2018;67(5):1-11. PMID: PMC5877350
- 336 6. Williams RA. Cardiovascular Disease in African American women: A Health Care  
337 Disparities Issue. *J Natl Med Assoc*, 2009;101(6):536-540. PMID: 19585921.
- 338 7. Winkleby MA, Cubbin C, Ahn DK, Kraemer HC. Pathways by which SES and ethnicity  
339 influence cardiovascular disease risk factors. *Ann N Y Acad Sci*. 1999;896(1):191-209.  
340 PMID: 10681898
- 341 8. Hu EA, Steffen LM, Coresh J, Appel LJ, Rebholz CM. Adherence to the healthy eating  
342 index–2015 and other dietary patterns may reduce risk of cardiovascular disease,  
343 cardiovascular mortality, and all-cause mortality. *J Nutr*. 2020 Feb 1;150(2):312-21.
- 344 9. Onvani S, Haghghatdoost F, Surkan PJ, Larijani B, Azadbakht L. Adherence to the Healthy  
345 Eating Index and Alternative Healthy Eating Index dietary patterns and mortality from all  
346 causes, cardiovascular disease and cancer: a meta-analysis of observational studies. *J of*  
347 *Hum Nutr and Diet*. 2017 Apr;30(2):216-26.
- 348 10. Stephenson BJK, Herring A, Olshan A. Robust clustering with subpopulation-specific  
349 deviations. *J Am Stat Assoc*. 2020;115(530):521-537. PMID: PMC7500490
- 350 11. Stephenson BJK, Sotres-Alvarez D, Siega-Riz AM, Mossavar-Rahmani Y, Daviglius ML,  
351 Van Horn L, Herring AH, Cai J. Empirically derived dietary patterns using robust profile  
352 clustering in the Hispanic Community Health Study/Study of Latinos. *J Nutr*.  
353 2020;150(10):2825-2834. PMID: PMC7549309.
- 354 12. Stephenson BJ, Dominici F. Identifying Dietary Consumption Patterns from Survey Data: a  
355 Bayesian Nonparametric Latent Class Model. medRxiv. 2021 Jan 1.
- 356 13. Field AE, Aneja P, Austin SB, Shrier LA, De Moor C, Gordon-Larsen P. Race and gender  
357 differences in the association of dieting and gains in BMI among young adults. *Obesity*.  
358 2007 Feb;15(2):456-64.
- 359 14. Gavin JR, Fox KM, Grandy S. Race/Ethnicity and gender differences in health intentions  
360 and behaviors regarding exercise and diet for adults with type 2 diabetes: A cross-sectional  
361 analysis. *BMC public health*. 2011 Dec;11(1):1-8.
- 362 15. Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. National Health and Nutrition  
363 Examination Survey: Sample design, 2011–2014. National Center for Health Statistics.  
364 *Vital Health Stat* 2014; 2(162).
- 365 16. Chen TC, Parker JD, Clark J, Shin HC, Rammon JR, Burt VL. National Health and Nutrition  
366 Examination Survey: Estimation procedures, 2011–2014. National Center for Health  
367 Statistics. *Vital Health Stat* 2018; 2(177).

- 368 17. Chen TC, Clark J, Riddles MK, Mohadjer LK, Fakhouri THI. National Health and Nutrition  
369 Examination Survey, 2015–2018: Sample design and estimation procedures. National  
370 Center for Health Statistics. *Vital Health Stat* 2020; 2(184).
- 371 18. Bodner-Montville, J., Ahuja, J., Ingwersen, L., Haggerty, E., Enns, C., & Perloff, B. USDA  
372 Food and Nutrient Database for Dietary Studies: Released on the web. *J Food Comp Anal*  
373 2006; 19, S100. Doi: [10.1016/j.jfca.2006.02.002](https://doi.org/10.1016/j.jfca.2006.02.002)
- 374 19. Bowman SA, Clemens JC, Friday JE, Thoerig RC, and Moshfegh AJ. Food Patterns  
375 Equivalents Database 2011-12: Methodology and User Guide [Online]. Food Surveys  
376 Research Group, Beltsville Human Nutrition Research Center, Agricultural Research  
377 Service, U.S. Department of Agriculture, Beltsville, Maryland. Available at:  
378 <http://www.ars.usda.gov/nea/bhnrc/fsrg>
- 379 20. Bowman SA, Clemens JC, Friday JE, Lynch KL, and Moshfegh AJ. Food Patterns  
380 Equivalents Database 2013-14: Methodology and User Guide [Online]. Food Surveys  
381 Research Group, Beltsville Human Nutrition Research Center, Agricultural Research  
382 Service, U.S. Department of Agriculture, Beltsville, Maryland. May 2017. Available at:  
383 <http://www.ars.usda.gov/nea/bhnrc/fsrg>
- 384 21. Bowman SA, Clemens JC, Shimizu M, Friday JE, and Moshfegh AJ. Food Patterns  
385 Equivalents Database 2015-2016: Methodology and User Guide [Online]. Food Surveys  
386 Research Group, Beltsville Human Nutrition Research Center, Agricultural Research  
387 Service, U.S. Department of Agriculture, Beltsville, Maryland. September 2018. Available  
388 at: <http://www.ars.usda.gov/nea/bhnrc/fsrg>
- 389 22. Bowman SA, Clemens JC, Friday JE, and Moshfegh AJ. Food Patterns Equivalents  
390 Database 2017-2018: Methodology and User Guide [Online]. Food Surveys Research  
391 Group, Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S.  
392 Department of Agriculture, Beltsville, Maryland. October 2020. Available at:  
393 <http://www.ars.usda.gov/nea/bhnrc/fsrg>
- 394 23. Liu L, Shih YCT, Strawderman RL, Zhang D, Johnson BA, Chai H. Statistical analysis of  
395 zero-inflated nonnegative continuous data: a review. *Statist Sci.* 2019;34(2):253-279. Doi:  
396 10.1214/18-STS681
- 397 24. Sotres-Alvarez D, Herring AH, Siega-Riz AM. Latent class analysis is useful to classify  
398 pregnant women into dietary patterns. *J Nutr.* 2010;140(12):2253-2259. PMID:  
399 PMC2981007
- 400 25. Van Havre Z, White N, Rousseau J, Mengersen K. Overfitting Bayesian mixture models  
401 with an unknown number of components. *PloS one.* 2015 Jul 15;10(7):e0131739
- 402 26. SAS Institute Inc. 2013. *SAS® 9.4 Statements: Reference*. Cary, NC: SAS Institute Inc.
- 403 27. Winham DM, Webb D, Barr A. Beans and good health. *Nutrition Today*, 2008;5:201-209.
- 404 28. Mattei J, Campos H. Perceptions and behaviors of legume consumption among Puerto  
405 Rican adults. *Health Behav Policy Rev.* 2014;1:38-49.
- 406 29. Winham DM, Hutchins AM, Johnston CS. Pinto bean consumption reduces biomarkers for  
407 heart disease risk. *J Am Coll Nutr.* 2007; 26:243-249. PMID:17634169.
- 408 30. Finley JW, Burrell JB, Reeves PG. Pinto bean consumption changes SCFA profiles in fecal  
409 fermentations, bacterial populations of the lower bowel, and lipid profiles in blood of  
410 humans. *J Nutr.* 2007;137(11):2391-2398. PMID:17951475.
- 411 31. Bazzano LA, Thompson AM, Tees MT, Nguyen CH, Winham DM. Non-soy legume  
412 consumption lowers cholesterol levels: a meta-analysis of randomized controlled trials. *Nutr*  
413 *Metab Cardiovasc Dis.* 2011; 21(2):94-103. PMID:19939654.

- 414 32. Thompson SV, Winham DM, Hutchins AM, Traditional bean and rice meals reduce  
415 postprandial glycemia in adults with type 2 diabetes: a cross-over study. *Nutr J* 2012;11:23.
- 416 33. Mitchell DC, Lawrence FR, Hartman TJ, Curran JM. Consumption of dry beans, peas, and  
417 lentils could improve diet quality in the US population. *J Am Diet Assoc.* 2009; 109(5):909-  
418 913. PMID:19394480.
- 419 34. Lin, B., & United States. Department of Agriculture. Economic Research Service, issuing  
420 body. Nutrition and health characteristics of low income populations. Healthy eating index  
421 (Agriculture information bulletin; no. 796-1). Washington, D.C.]: United States Department  
422 of Agriculture, Economic Research Service 2005.
- 423 35. Leung, C., Ding, E., Catalano, P., Villamor, E., Rimm, E., & Willett, W. (2012). Dietary  
424 intake and dietary quality of low-income adults in the Supplemental Nutrition Assistance  
425 Program. *Am J of Clin Nutr* 2012; 96(5), 977-988.
- 426 36. Wiig Dammann, Kristen, BS, & Smith, Chery, PhD, MPH, RD. Factors Affecting Low-  
427 income Women's Food Choices and the Perceived Impact of Dietary Intake and  
428 Socioeconomic Status on Their Health and Weight. *J Nutr Educ Behav* 2009; 41(4), 242-  
429 253.
- 430 37. Institute of Medicine (US) Committee on Dietary Risk Assessment in the WIC Program.  
431 Dietary Risk Assessment in the WIC Program. Washington (DC): National Academies  
432 Press (US); 2002. 3, Using the Dietary Guidelines as the Basis of Dietary Risk  
433 Criteria. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK220563/>
- 434 38. Wardle J Parmenter K, Waller J. Nutrition knowledge and food intake. *Appetite* 2000;  
435 34(3):269-275.
- 436 39. Wang DD, Leung CW, Li Y, Ding EL, Chiuve SE, Hu FB, Willett WC. Trends in dietary  
437 quality among adults in the United States, 1999 through 2010. *JAMA Intern Med* 2014;  
438 174(10):1587-1595.
- 439  
440  
441

442  
443

**Table 1.** Demographic information of low-income adult women participants pooled from 2011-2018. Percentages account for survey sampling weights. Counts are unweighted.

	<b>Overall</b>	<b>Mexican</b>	<b>Other Hispanic</b>	<b>NH White</b>	<b>NH Black</b>	<b>NH Asian</b>
<i>N</i>	3005	574	414	981	809	227
<i>% (SE)</i>	--	15.9 (2.2)	9.4 (1.3)	49.9 (3.3)	19.2 (2.1)	5.6 (1.0)
<i>Age Group, %</i>						
20-34 (n=777)	32.8 (1.5)	30.6 (4.6)	34.4 (4.6)	29.3 (3.1)	30.7 (3.6)	38.5 (5.7)
35-49 (n=729)	25.1 (1.2)	42.1 (4.0)	24.8 (3.8)	24.9 (2.3)	29.2 (3.9)	22.1 (4.6)
50-64 (n=740)	23.0 (1.0)	17.2 (3.2)	23.1 (3.3)	26.5 (2.6)	24.9 (2.1)	22.2 (5.1)
65+ (n=702)	19.2 (1.1)	10.1 (1.9)	17.7 (2.7)	19.4 (2.3)	15.2 (2.3)	17.3 (5.5)
<i>Marital Status, %</i>						
Married/Living with Partner (n=1292)	45.9 (2.0)	57.7 (3.1)	44.5 (5.9)	43.2 (3.3)	24.5 (2.9)	58.7 (7.1)
Separated/Widowed/Divorced (n=1038)	30.4 (1.3)	26.7 (2.9)	31.0 (3.3)	40.6 (3.1)	35.9 (2.7)	17.9 (5.9)
Never married (n=674)	23.7 (2.1)	15.7 (2.3)	24.5 (5.7)	16.2 (2.4)	39.6 (3.0)	24.0 (5.8)
<i>Education, %</i>						
At least some college (n=1184)	43.6 (1.9)	23.9 (3.3)	25.4 (3.4)	49.4 (3.2)	42.7 (2.9)	37.5 (7.1)
HS/GED (n=745)	26.5 (1.0)	21.2 (3.6)	27.4 (4.8)	28.0 (2.2)	32.3 (2.7)	35.4 (5.3)
Less than HS (n=1076)	29.9 (1.8)	54.9 (4.0)	47.2 (5.6)	22.6 (2.7)	25.0 (2.2)	27.1 (6.0)
<i>Federal Assistance (SNAP/WIC), %</i> (n=3005)	49.2 (1.8)	53.2 (4.5)	58.4 (4.2)	41.1 (2.6)	68.8 (2.3)	26.9 (5.0)
<i>Household Size</i> (n=3005)	3.4 (0.1)	4.2 (0.2)	3.7 (0.2)	3.0 (0.1)	3.4 (0.2)	3.7 (0.3)
<i>HEI-2015 Total Score</i> (n=3005)	49.5 (0.4)	51.7 (0.9)	52.7 (1.1)	49.0 (0.7)	47.6 (0.6)	57.4 (1.5)
<i>Energy intake (kcal)</i> (n=3005)	1751 (21)	1811 (51)	1622 (77)	1762 (43)	1764 (74)	1554 (67)
<i>Body Mass Index (kg/m<sup>2</sup>)</i> (n=2969)	30.1 (0.2)	31.3 (0.6)	31.2 (0.7)	30.4 (0.6)	32.2 (0.5)	24.6 (0.7)
<i>CVD Risk factors, %</i>						
High Blood Pressure or meds (n=2925)	34.1 (1.7)	21.9 (3.7)	26.5 (3.8)	34.7 (2.7)	45.4 (3.2)	24.2 (6.7)
Diabetes or medication (n=3005)	10.3 (1.0)	13.1 (2.4)	14.1 (2.9)	10.5 (1.5)	12.9 (2.4)	7.5 (3.2)

Obese (n=2969)	46.0 (1.8)	47.8 (4.9)	49.1 (3.7)	45.3 (3.3)	58.2 (3.1)	18.1 (5.0)
High cholesterol (n=1911)	54.3 (1.6)	52.0 (2.8)	58.1 (4.1)	55.3 (2.6)	50.8 (3.4)	61.3 (5.1)
Current Smoker (n=3005)	23.3 (1.7)	8.1 (2.0)	9.4 (2.5)	34.3 (2.9)	28.4 (3.4)	4.8 (2.8)

444

445

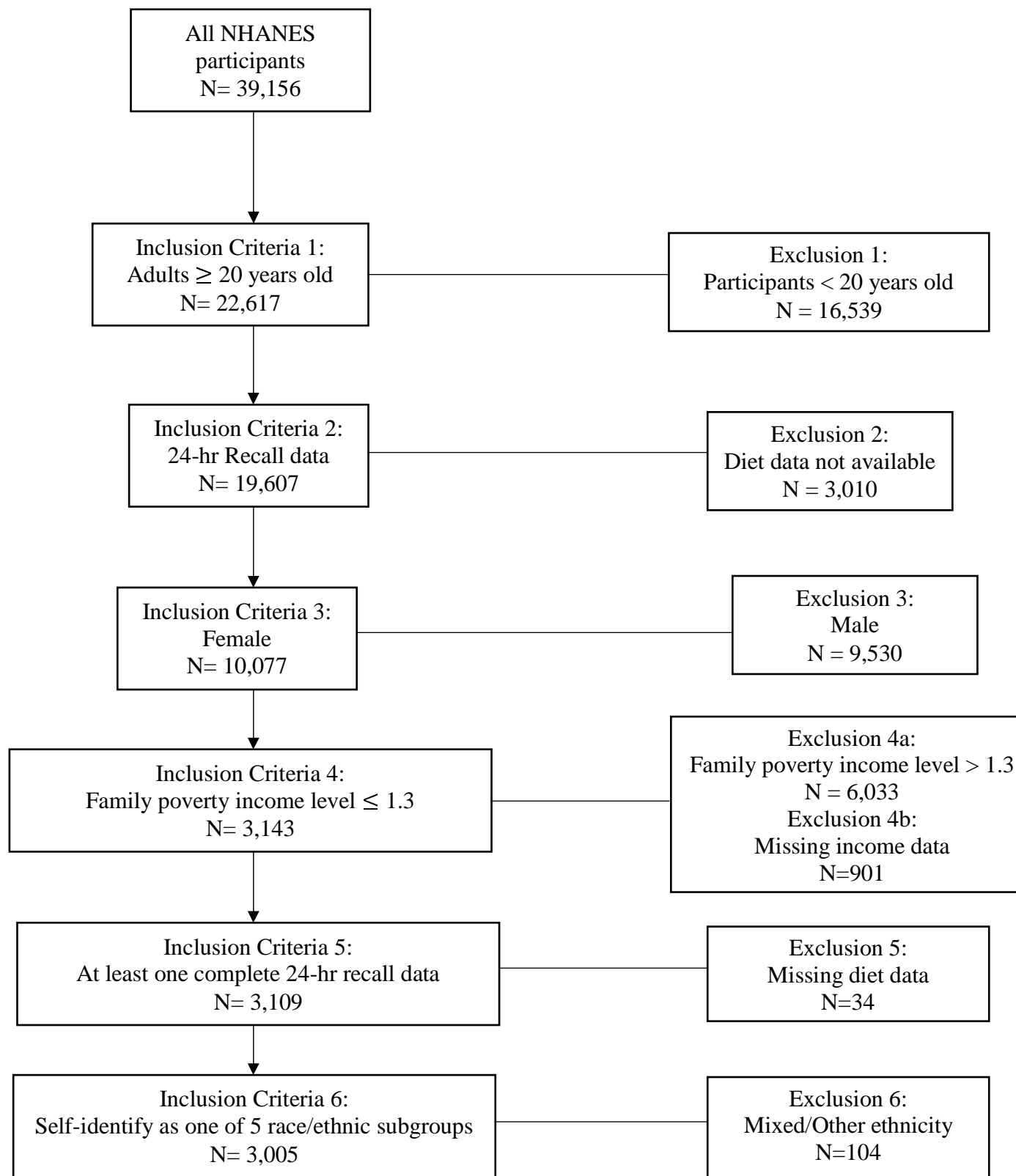
446  
447

**Table 2.** Weighted means with standard errors in parentheses are provided to describe demographic distribution across RPC-derived global profiles.

	Global 1 8.9 (0.7)	Global 2 2.2 (0.4)	Global 3 8.9 (0.8)	Global 4 9.4 (0.7)	Global 5 70.5 (1.2)
<i>Age Group</i>					
20-34	33.1 (4.0)	24.4 (6.9)	33.0 (4.7)	37.0 (4.3)	32.4 (1.7)
35-49	34.3 (4.6)	40.9 (9.5)	28.5 (4.2)	20.2 (3.0)	23.6 (1.3)
50-64	19.3 (3.1)	25.0 (6.7)	22.0 (2.9)	19.6 (3.0)	24.0 (1.3)
65+	13.4 (2.3)	9.6 (3.7)	16.5 (3.6)	23.2 (3.4)	20.0 (1.5)
<i>Race/Ethnicity</i>					
Mexican	29.5 (5.5)	31.7 (7.5)	23.8 (5.9)	19.0 (5.2)	11.2 (2.1)
Other Hispanic	15.2 (4.5)	3.0 (1.7)	23.6 (6.6)	10.6 (4.0)	7.2 (1.3)
Non-Hispanic White	43.7 (7.0)	50.6 (11.1)	40.2 (7.4)	50.9 (7.4)	55.5 (4.2)
Non-Hispanic Black	7.7 (2.9)	7.8 (4.4)	9.5 (3.4)	11.2 (3.0)	21.1 (2.7)
Non-Hispanic Asian	4.0 (1.6)	6.9 (4.7)	2.9 (1.3)	8.3 (4.1)	5.0 (1.1)
<i>Marital Status</i>					
Married/Living with Partner	67.8 (5.5)	26.1 (9.1)	45.9 (6.1)	47.1 (5.9)	38.3 (2.6)
Separated/Divorced/Widowed	17.5 (4.7)	51.3 (11.0)	32.0 (5.8)	32.3 (4.9)	37.4 (2.1)
Never married	14.7 (4.5)	22.5 (9.2)	22.0 (6.7)	20.6 (5.3)	24.4 (2.5)
<i>Education</i>					
At least some college	39.5 (5.8)	52.2 (9.9)	33.8 (4.9)	45.4 (6.9)	43.1 (2.4)
High School/GED	17.9 (4.5)	19.9 (7.9)	27.9 (4.5)	30.6 (6.4)	29.1 (1.7)
Less than High school	42.6 (5.9)	27.9 (8.2)	38.3 (5.7)	23.9 (5.0)	27.9 (2.0)
Federal Assistance Program (SNAP/WIC)	53.2 (5.1)	38.6 (9.3)	58.5 (4.1)	43.4 (4.2)	48.6 (1.8)
<i>Continuous measures</i>					
Household size	4.0 (0.3)	3.4 (0.4)	3.8 (0.2)	3.3 (0.3)	3.2 (0.1)
HEI2015 Total Score	55.3 (1.1)	59.6 (1.5)	51.2 (1.2)	52.6 (1.4)	48.2 (0.5)
Energy (kcal)	2144 (74)	2075 (142)	1831 (87)	1738 (68)	1680 (31)
Body Mass index	30.5 (0.7)	33.5 (2.5)	29.5 (1.1)	28.1 (1.0)	31.0 (0.4)
<i>CVD risk factors</i>					
High blood pressure + meds	27.8 (5.2)	33.4 (9.5)	30.2 (4.3)	33.7 (5.0)	33.2 (2.2)
Diabetes + meds	12.0 (4.2)	19.9 (7.5)	6.2 (1.9)	10.8 (3.5)	11.6 (1.2)
Obese (BMI > 30)	44.0 (5.9)	48.4 (9.5)	41.8 (5.2)	30.3 (6.3)	49.3 (2.2)
High Cholesterol	76.6 (4.1)	77.5 (7.1)	69.8 (5.5)	68.5 (6.3)	72.5 (1.8)
Current smoker	16.3 (5.0)	26.6 (13.0)	16.3 (4.2)	18.7 (6.0)	28.9 (2.2)

448  
449  
450  
451  
452

453 Figure 1. CONSORT diagram of NHANES 2011-2018 participants included for study analysis.  
454 A total of 3005 participants were included after all exclusion/inclusion criteria was applied.  
455  
456  
457

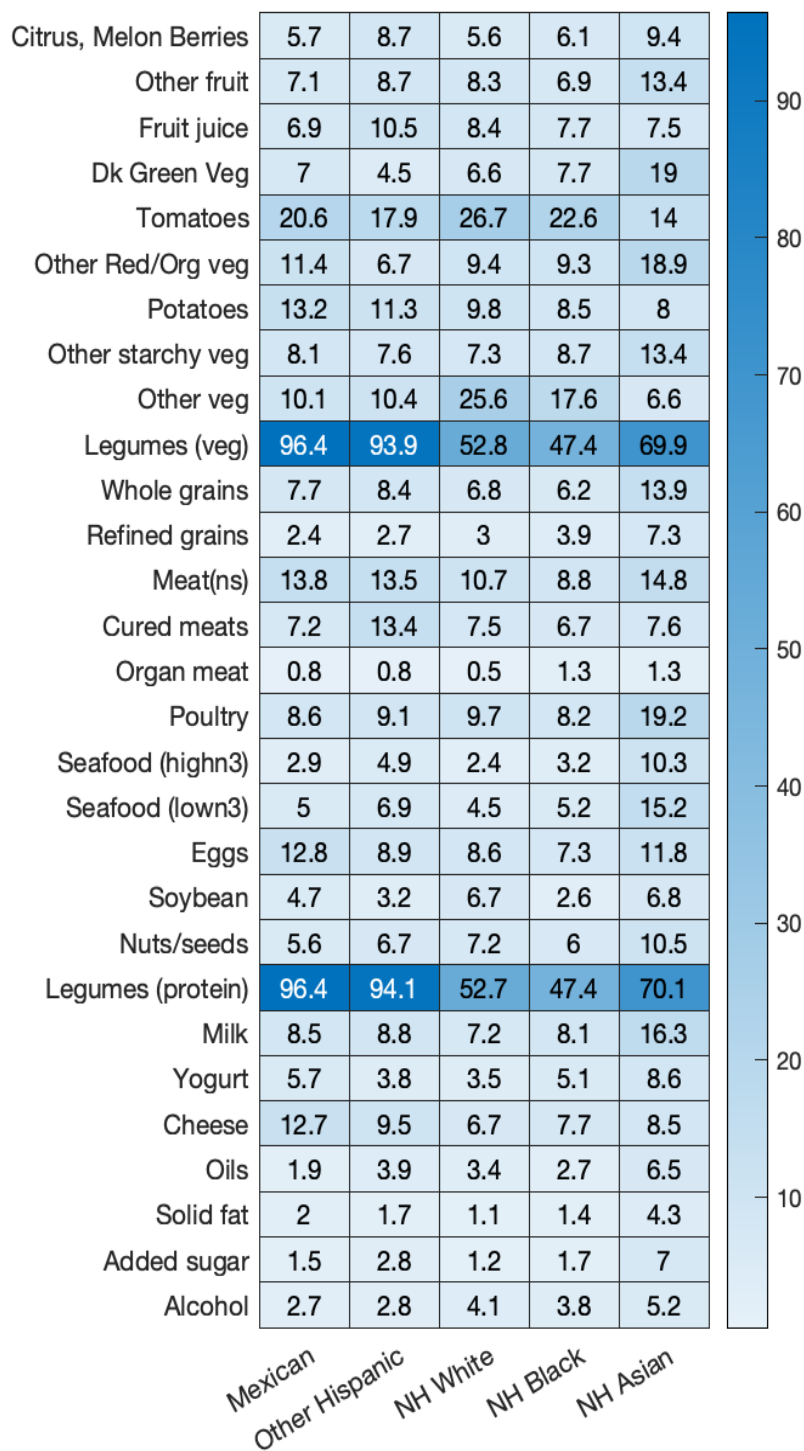




458 **Figure 2.** Consumption modal pattern for five RPC-derived global dietary profiles. Levels of  
 459 consumption are highlighted as follows: 1-Dark Blue (no consumption), 2-Light Blue (low  
 460 consumption), 3-Green (medium consumption), 4-yellow (high consumption). Key features of  
 461 each global dietary profile: Global 1, high legumes, medium poultry; Global 2, high legumes,  
 462 high organ meat; Global 3, medium legumes, medium seafood; Global 4, high organ meat, low  
 463 fruit/veg; Global 5, high starchy vegetable, high poultry.

464						
465						
466	Citrus, Melon Berries	3	3	2	2	1
467	Other fruit	4	4	3	3	1
468	Fruit juice	4	2	2	2	1
469	Dk Green Veg	4	3	3	3	1
470	Tomatoes	4	4	4	2	1
471	Other Red/Org veg	1	2	2	2	1
472	Potatoes	1	1	2	2	1
473	Other starchy veg	1	1	3	2	4
474	Other veg	4	4	3	3	1
475	Legumes (veg)	4	4	3	2	1
476	Whole grains	4	4	2	3	1
477	Refined grains	4	4	4	3	2
478	Meat(ns)	4	3	3	2	1
479	Cured meats	3	3	2	2	1
480	Organ meat	1	4	1	4	1
481	Poultry	3	2	2	2	4
482	Seafood (highn3)	3	1	1	3	1
	Seafood (lown3)	1	1	3	2	1
	Eggs	4	4	4	3	1
	Soybean	1	1	3	3	2
	Nuts/seeds	3	3	3	2	1
	Legumes (protein)	4	4	3	2	1
	Milk	4	4	2	3	1
	Yogurt	3	3	3	2	1
	Cheese	4	4	4	2	1
	Oils	4	4	3	3	2
	Solid fat	4	4	3	2	2
	Added sugar	4	4	3	3	2
	Alcohol	1	1	1	3	1
		1	2	3	4	5
		Global Dietary Profile				

483 **Figure 3.** Heatmap illustrating probability of a given food item assuming a pattern at the global  
 484 level. Foods likely to assume a global pattern are darker in blue hue. Foods likely to assume a  
 485 localized pattern by the racial/ethnic subgroup are lighter in blue hue.



**Figure 4.** Pattern distributions for RPC-derived local profiles for each racial/ethnic subgroups. M=Mexican, H=Other Hispanic, W=Non-Hispanic White, B=Non-Hispanic Black, A=Non-Hispanic Asian.

