Latent Class Analysis Is Useful to Classify Pregnant Women into Dietary Patterns[1–3]

Daniela Sotres-Alvarez,[4]* Amy H. Herring,[4,5] and Anna Maria Siega-Riz[5,6]

[4]Department of Biostatistics and [6]Departments of Epidemiology and Nutrition, Gillings School of Global Public Health, and [5]Carolina Population Center, University of North Carolina, Chapel Hill, NC 27599

Abstract

Empirical dietary patterns are derived predominantly using principal components, exploratory factor analysis (EFA), or cluster analysis. Interestingly, latent variable models are less used despite their being more flexible to accommodate important characteristics of dietary data and despite dietary patterns being recognized as latent variables. Latent class analysis (LCA) has been shown empirically to be more appropriate to derive dietary patterns than k-means clustering but has not been compared yet to confirmatory factor analysis (CFA). In this article, we derived dietary patterns using EFA, CFA, and LCA on food items, tested how well the classes from LCA were characterized by the factors from CFA, and compared participants' direct classification from LCA on food items compared with 2 a posteriori classifications from factor scores. Methods were illustrated with the Pregnancy, Infection and Nutrition Study, North Carolina, 2000–2005 ($n = 1285$ women). From EFA and CFA, we found that food items were grouped into 4 factors: Prudent, Prudent with coffee and alcohol, Western, and Southern. From LCA, pregnant women were classified into 3 classes: Prudent, Hard core Western, and Health-conscious Western. There was high agreement between the direct classification from LCA on food items and the classification from the 2-step LCA on factor scores [$\kappa = 0.70$ (95% CI = 0.66, 0.73)] despite factors explaining only 25% of the total variance. We suggest LCA on food items to study the effect for mutually exclusive classes and CFA to understand which foods are eaten in combination. When interested in both benefits, the 2-step classification using LCA on previously derived factor scores seems promising.   J. Nutr. 140: 2253–2259, 2010.

Introduction

Dietary patterns are useful to study the effects of overall diet on health outcomes as opposed to the effects of individual nutrients or foods (1). Because dietary patterns are not directly observed, they are measured with a dietary intake instrument and when assessed a posteriori they are empirically derived using predominantly (2–4) principal components, exploratory factor analysis (EFA),[7] and cluster analysis (5). These are multivariate data reduction methods and do not consider dietary patterns as latent variables (i.e. unobserved random variables). Interestingly, latent variable models (6) have rarely been used despite being recognized as useful in reflecting complex relations between diet

and disease at the 2000 International Workshop on Dietary Patterns (7). Their flexibility offers several advantages (8) for dietary pattern analysis such as modeling observed outcomes with different distributions simultaneously, accounting for correlated errors, adjusting for covariates (e.g. energy intake and age), and multi-group analysis (e.g. testing hypothesis between genders). In particular, confirmatory factor analysis (CFA) attempts to explain the correlations between many observed variables (e.g. food items) by few underlying continuous latent variables (called factors; e.g. dietary patterns) as opposed to EFA, where the relationship between the observed variables and the factors is not specified.

Another type of latent variable model, although less known in epidemiology, are latent class models (9) and to date, only 2 studies (8,10) have used them to derive dietary patterns. In latent class analysis (LCA), individuals are assumed to belong to one of K mutually exclusive classes but for which class membership is unknown, and through a statistical model the latent class explains the associations among the observed variables. LCA relaxes the strict assumptions of conditional independence and same error variance for all outcomes and clusters assumed in K-means clustering, and Fahey et al. (8) showed LCA had a better model fit [measured by the Bayesian information criterion (BIC)] for their data. LCA is useful to study unobserved heterogeneity characterized by several unidentified groups that

behave differently and, in this sense, is similar to nonhierarchical cluster analysis. However, LCA is a model-based clustering approach, not a partition based on a numerical criterion optimization. Technically speaking, in nonhierarchical cluster analysis, the group membership is a parameter, whereas in LCA it is an unobserved random variable. Hence, in LCA, individuals have a predicted probability for belonging to each class, which reflects the uncertainty of class membership. Similar to cluster analysis, the number of groups K is assumed to be known, although this is almost never the case.

Dietary patterning literature has used the terms "dietary patterns", "eating patterns", and "food patterns" interchangeably and regardless of the statistical method used to derive them. However, solutions from these methods are, conceptually and statistically, different. In PCA, EFA, and CFA, food items are grouped according to the degree to which they are correlated to each other and individuals have a score for each dietary pattern. By contrast, in cluster analysis and LCA, individuals are grouped into mutually exclusive dietary patterns such that within groups they have similar food intake. The former methods are useful to understand which foods are consumed in combination and to study associations between dietary patterns and health outcomes, whereas clustering methods are useful to classify individuals and to estimate the risk of an outcome for a group compared with a referent group. Indeed, even when dietary patterns are derived using factor analysis (FA), investigators are still interested in classifying the individuals based on their factor scores. In practice, when there are only 2 factors an easy way to classify them is from the cross-tabulation of the factor scores' quantiles (1,11). However, when there are more than 2 factors, the total number of cells from the cross-tabulation of the factors scores' quantiles might be too large and it could be difficult to collapse into mutually exclusive groups without making any strong subjective decisions. An alternative a posteriori approach is to perform a LCA on factor scores to classify individuals.

CFA and LCA can provide interesting insights into dietary patterning and to date there are no studies that have compared the dietary patterns derived by these 2 methods. We aimed to: *1)* derive dietary patterns using EFA, CFA, and LCA on food items; *2)* test how well the classes from LCA were characterized by the factor scores from CFA; and *3)* compare participants' direct classification from LCA on food items compared with 2 a posteriori classifications from factor scores. We used data from the 3rd cohort of the Pregnancy, Infection and Nutrition (PIN) Study.

Methods

Study population. We used data from the 3rd cohort of the PIN Study (December 2000 to June 2005). The study recruited pregnant women seeking services from prenatal clinics at University of North Carolina Hospitals. Study protocols were reviewed and approved in accordance with the ethical standards of the Institutional Review Board of the University of North Carolina School of Medicine. A total of 1875 women (2006 singleton pregnancies) were enrolled who fulfilled the minimum age of 16 y and <20 wk of gestation, from which 1352 women (1442 pregnancies) had complete dietary data. For this analysis, only 1 pregnancy was randomly selected when a woman had several pregnancies with complete dietary assessments. The mean age was 29.5 ± 5 y (range 16–47 y), 78.5% were married, 17.8% had ≤12 y of education, one-half were nulliparous, 10.3% smoked during months 1–6 of pregnancy, 74.4% were white, and 15.9% black. Based on the categories established by the Institute of Medicine guidelines and using pregravid weight, 14.3% were underweight (BMI < 19.8 kg/m^2), 52.6% were normal (19.8–26.0 kg/m^2), 10.5% were overweight (>26.0–29.0 kg/m^2), and 22.6% were obese (>29.0 kg/m^2).

Dietary intake assessment. Dietary intake was assessed through a self-administered, semiquantitative, 119 food item Block FFQ (12) to measure usual intake in the past 3 mo. It was administered at 26–29 wk of gestation to reflect diet during the second trimester. Dietsys+Plus version 5.6 with an updated food composition table based on nutrient values from the NHANES III and the USDA 1998 nutrient databases was used to calculate daily energy intake in kcal and g/d. We excluded women with daily energy intakes below the 2.5th or above the 97.5th percentiles (1000 and 4765 kcal, respectively) as an attempt to exclude implausible energy intakes, leaving 1285 women for the analysis.

The number of FFQ food items to derive the dietary patterns was reduced from 119 to 105 (**Supplemental Table 1**), because 9 food items were rarely consumed (<10% consumption), and alcoholic drinks (beer, spirits, wine) and low-fat milks (skim, 1%, and 2%) were combined into 2 groups due to very small counts. Given that many food items' distributions were skewed and had a lump at zero due to nonconsumers, the indicators were categorized. Most were categorized into a 3-level variable: nonconsumers (g/d = 0) and below or above the median of consumption among consumers (g/d > 0) to distinguish low and high consumption. Eleven food items were dichotomized as below or above the median, because there were too few nonconsumers and 9 were dichotomized as consumed or not consumed, because there were too few consumers.

Statistical analysis. We derived dietary patterns by EFA, CFA, and LCA. First, we conducted an EFA on 105 ordinal food items using weighted least squares and factors were derived orthogonal using Varimax rotation. We decided the number of factors from a combination of the scree plot and the interpretation of the factor loadings. Dietary patterns' names were given according to the foods with higher loadings and also based on the literature. Second, we performed CFA on the dietary patterns derived by EFA including only food items with loadings in absolute value ≥ 0.25, allowing food items to load on multiple factors. We specified correlated errors between coffee and cream and iced tea and sugar, because the FFQ asked specifically if these condiments were usually added to these drinks. We conducted a CFA with correlated factors to test if, after constraining some of the loadings to zero, factors were still orthogonal. We adjusted for energy intake, parity, smoking status, education, age, and race and assessed goodness-of-fit with the root mean square error of approximation (13).

To determine mutually exclusive groupings, we used LCA to derive dietary patterns including only food items with EFA loadings ≥ 0.25. First, we fit LCA without covariates with 2 to 4 classes to determine the number of classes using the Lo-Mendell-Rubin likelihood ratio test. After selecting the number of classes, the model was adjusted for energy intake and covariates. Because in LCA each individual has a predicted probability for belonging to each class, we classified them into the class with the highest associated probability of class membership. We interpreted and named the classes from the conditional probabilities of consumption. Finally, we compared nutrient intake between classes using the Mann-Whitney test with a Bonferroni correction for multiple comparisons.

We used 2 approaches to compare the dietary patterns derived by FA and LCA. The goal for the first approach was to describe how well the classes from LCA could be characterized by the factor scores from CFA; we compared factor scores' means among classes. The second approach examined whether participants' direct classification into dietary patterns using LCA on food items agreed with their classification using factor scores. Because FA does not classify participants directly, we classified them a posteriori fitting LCA on the 4 continuous factor scores. We assumed conditional independence given the class, and different factor means and variances by class. For this ad hoc 2-step procedure, we determined the same number of classes obtained directly from the LCA on food items. For comparison purposes, we also classified women by cross-tabulating the 4 factor scores' tertiles. Because dietary patterns' membership is unknown, we could not test which classification was best but only whether the direct classification using LCA on food items and the 2 a posteriori classifications from factors agreed or not. Agreement was assessed with the weighted kappa statistic. For all tests, $P < 0.05$ was considered significant unless a Bonferroni correction was made to

the significance level to account for multiple comparisons, in which case it was explicitly stated. Statistical analyses were performed using SAS/STAT software, version 9.1 of the SAS System for Windows (14), the procedure PROC LCA (15), and Mplus Version 5.1 (16) to fit EFA, CFA, and the ad hoc 2-step procedure. Supplemental Table 2 compares selected software to fit LCA.

Results

FA. According to the scree plot from EFA, after the 4th factor, factors did not contribute much to explain the variance of the data (the first 6 Eigenvalues were 10.22, 8.66, 4.36, 3.19, 2.62, and 2.59). One factor loaded high (>0.25) in many fruits and vegetables, whole grains, yogurt, vegetable soup, and beans; it was called FA-Prudent (**Table 1**). A second factor loaded high on processed meat, hamburger, French fries, soft drinks, and Southern foods (coleslaw, corn, collards, green beans, fried chicken and fish, pork, corn bread, and iced tea); it was called FA-Southern. A 3rd factor loaded on green salad and dressing, tomatoes, broccoli, spinach, fish not fried, whole grains, coffee, and alcohol; it was called FA-Prudent with coffee and alcohol. The 4th factor loaded high in fast food, salty snacks, and

TABLE 1 Selected exploratory and confirmatory factor loadings for 4-factor model (PIN Study, 2000–2005)[1–3]

Food item	FA-Prudent		FA-Southern		FA-Western		FA-Prudent coffee and alcohol		R^2
	EFA	CFA	EFA	CFA	EFA	CFA	EFA	CFA	
Oranges or tangerines	0.47	0.49							0.23
Apples or pears	0.45	0.51							0.24
Coleslaw or cabbage			0.50	0.46					0.20
Greens (e.g. collards)			0.51	0.41	−0.31	−0.19			0.14
Raw tomatoes							0.54	0.65	0.37
Spinach	0.36	0.25					0.50	0.47	0.35
Carrots	0.41	0.40					0.37	0.26	0.30
Green salad							0.66	0.70	0.42
Salad dressing							0.53	0.52	0.25
Yogurt	0.36	0.36	−0.31	−0.25			0.38	0.29	0.36
Low fat milk			−0.41	−0.27					0.07
Baked beans	0.40	0.43							0.18
Vegetable stew	0.50	0.51							0.24
Beef (e.g. roast, steak)			0.46	0.61					0.33
Pork (e.g. chops, roasts, dinner ham)			0.50	0.63					0.34
Ribs or spareribs			0.58	0.62					0.33
Fried chicken			0.63	0.73					0.44
Fried fish			0.48	0.46					0.20
Chicken not fried							0.36	0.29	0.08
Fish not fried							0.56	0.67	0.39
Hot dogs or dinner sausage			0.55	0.66					0.37
Bacon			0.53	0.69					0.40
Breakfast sausage			0.56	0.65					0.36
Meat substitutes (not just soy)	0.40	0.56	−0.52	−0.54					0.53
White bread			0.40	0.48					0.21
Bagels or muffins					0.44	0.35			0.12
Whole wheat bread (e.g. dark, rye)	0.33	0.35	−0.42	−0.36			0.43	0.29	0.40
High-fiber cereals	0.30	0.40							0.15
Salty snacks (e.g. chips, popcorn)					0.44	0.35			0.12
Ice cream					0.41	0.37			0.13
Doughnuts or pastries					0.43	0.60			0.31
Cake					0.39	0.54			0.26
Coffee							0.36	0.30	0.09
Alcohol (beer, spirits, and wine)							0.32	0.20	0.04
Vitamin C-rich drinks (e.g. Kool-Aid, Hi-C)			0.53	0.40			−0.38	−0.19	0.19
Drinks with some juice (e.g. Sunny D)			0.47	0.46					0.20
French fries or fried potatoes			0.40	0.40	0.34	0.12			0.20
Hamburger or cheeseburger			0.50	0.67	0.34	0.03			0.39
Pizza					0.48	0.38			0.14
Cheese dish (e.g. macaroni and cheese)					0.41	0.44			0.18
Tacos or burritos					0.49	0.52			0.25

[1] The full table with all 105 food items included in EFA is available as Supplemental Table 1.
[2] The confirmatory 4-factor model was adjusted for energy intake, nulliparous, smoker, white, education, and age. It included correlated errors between coffee and cream, and iced tea and sugar or honey. Some factors were correlated; $r = 0.49$ between FA-Southern and FA-Western, $r = 0.38$ between FA-Prudent and FA-Prudent with coffee and alcohol, and $r = 0.17$ between FA-Prudent and FA-Western.
[3] Sample size was 1285 women for EFA and 1219 women for CFA due to missing values in some covariates.

pastries; it was called FA-Western. Most food items loaded on only 1 factor, 12 loaded on 2 factors, and 3 loaded on 3 factors. Seven food items (cheese, eggs, nonfortified cereal, pudding, orange juice, diet soft drinks, and butter) with EFA loadings < 0.25 for all factors were excluded from CFA and LCA.

The factor loadings from EFA and CFA were similar (Table 1, **Supplemental Table** 3) except for French fries and hamburger for FA-Western and real fruit juice excluding orange juice for FA-Southern. Hence, for the dietary patterns assessed by CFA, we kept the names given from EFA. However, the overall test for the correlations between the 4 factors being zero was significant ($P < 0.001$) and this model had a slightly better fit than the one with uncorrelated factors. The highest correlation was between FA-Southern and FA-Western ($r = 0.49$; $P < 0.001$), and 0.38 ($P < 0.001$) between FA-Prudent and FA-Prudent with coffee and alcohol. Although significant, the correlation between FA-Prudent and FA-Western was much smaller ($r = 0.17$; $P = 0.035$). The correlated errors between coffee and cream, and iced tea and sugar were significant. The food items for which their variance was better accounted for by the factors ($R^2 > 0.4$) were: green salad, fried chicken, bacon, whole wheat bread, and meat substitutes. Supplemental Table 4 presents regression coefficients for CFA and supplemental Table 5 presents correlations between factor scores and daily dietary nutrient intakes.

LCA. We chose 3 classes because the model was different from the one with 2 classes ($P = 0.011$) but not different from the one with 4 classes ($P = 0.748$). One class had higher probabilities of consuming more fruits and vegetables, whole grains, baked beans, nuts, fish and chicken (not fried), yogurt, water, and low-fat milk; we called it LCA-Prudent (**Fig. 1** includes 9 food items with marked differences between classes; **Supplemental Fig. 1** shows all 98 food items). Women in this class had higher consumption of fiber, folate, and vitamins (**Table 2**). The second class had high probabilities for consuming higher amounts of fast food, salty snacks, and sweets but also for fruits and vegetables. It was called LCA-Health Conscious Western and had significantly higher median percent of energy from fat and sweets compared with the LCA-Prudent class, but the micronutrient intake was similar. A 3rd class was less likely to eat fruits, vegetables, yogurt, low-fat milk, coffee, alcohol, nuts, and beans and more likely to consume fried fish and chicken, sausages, white bread, and soft drinks. It was called LCA-Hard Core Western and had significantly lower micronutrient intake compared with the other 2 classes but fat intake similar to the LCA-Health Conscious Western class. With respect to Southern foods, the LCA-Prudent class had higher percentages of nonconsumers and there were no differences between the 2 LCA-Western classes. Overall, there were 32.8% women in the LCA-Prudent, 34.6% in the LCA-Health Conscious Western, and 32.6% in the LCA-Hard Core Western. However, the prevalence depends on parity, smoking status, race, and education. White, nulliparous, older, and more educated women were more likely to be in the LCA-Prudent class than in LCA-Hard Core Western (**Table 3**). Heavier women were significantly less likely to be in the LCA-Prudent class. Women with higher energy intake were 2 to 3 times more likely to be in the LCA-Health Conscious Western class than in LCA-Hard Core Western class.

Comparison between factor scores and classes. The LCA-Prudent and LCA-Health Conscious Western classes had significantly higher means for FA-Prudent and FA-Prudent with coffee and alcohol factors compared with the LCA-Hard Core Western class (**Fig. 2A**). The LCA-Prudent class had significantly lower
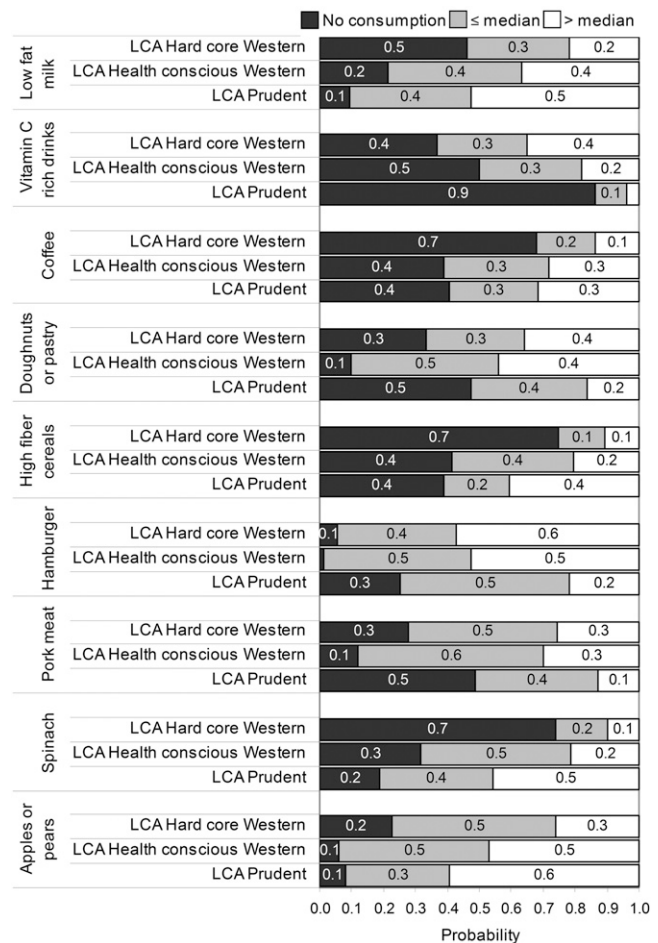

FIGURE 1 Probabilities of consumption for selected food items by dietary patterns derived using LCA. The latent class model was adjusted for energy intake, nulliparous, smoker, white, education, and age. It included correlated errors between coffee and cream, and iced tea and sugar and honey. Values are estimated probabilities and the legend indicates the ordinal level of consumption, $n = 1219$.

means for FA-Southern and FA-Western factor scores than the LCA-Health Conscious Western class. The LCA-Health Conscious Western class had a significantly higher FA-Western mean than the LCA-Hard Core Western class, and the FA-Southern means were not significantly different.

The second approach compared the direct classification into 3 dietary patterns using LCA on the food items compared with the a posteriori classification using LCA on the 4 factor scores. These latter classes were interpreted by comparing the means of the factors as before (Fig. 2B). We called one class 2-Step Prudent/Anti-Southern, because it had means significantly higher than zero for FA-Prudent and FA-Prudent with coffee and alcohol factors and a negative mean for the FA-Southern factor. A second class had the highest FA-Western mean but also had means significantly higher than zero for FA-Prudent and FA-Prudent with coffee and alcohol; it was called 2-Step Western/Prudent. Finally, the 3rd class had lower means for FA-Prudent and FA-Prudent with coffee and alcohol and a higher mean for the FA-Western factor; it was called 2-Step Western. To compare patterns derived from LCA to those derived from the 2-step method, we mapped 2-Step Western to LCA-Hard Core Western, 2-Step Western/Prudent to LCA-Health Conscious Western, and 2-Step Prudent/Anti-Southern to LCA-Prudent. There was high agreement between the 2 classifications [$\kappa = 0.70$ (95% CI =

TABLE 2 Daily dietary nutrient intakes by latent class (PIN Study, 2000–2005)[1,2]

| | Latent class | | | | | |
| | LCA-Prudent | | LCA-Health Conscious Western | | LCA-Hard Core Western | |
	Median	IQR	Median	IQR	Median	IQR
Total energy,[3] *kcal*	1870[b]	688	2190[a]	857	2010[b]	927
Fat, *g*	65.4[c]	29.2	79.6[a]	38.3	72.1[b]	43.7
Saturated fat, *g*	21.8[b]	10.2	26.3[a]	13.5	24.5[a]	15.4
Cholesterol, *g*	0.18[b]	0.12	0.23[a]	0.14	0.22[a]	0.15
(n-3) fatty acids, *g*	1.8[b]	1.1	2.0[a]	1.2	1.7[b]	1.2
Fiber, *g*	19.6[a]	10.9	18.2[b]	8.6	13.5[c]	8.5
Iron, *mg*	14.9[a]	6.7	15.4[a]	6.4	13.4[b]	8.3
Folate, *mg*	0.42[a]	0.18	0.41[a]	0.16	0.34[b]	0.18
Calcium, *g*	1.06[a]	0.51	0.98[a]	0.51	0.85[b]	0.53
Vitamin D,[4] *IU*	199[a]	188	189[ab]	178	156[b]	189
Vitamin A,[5] *μg RE*	1500[a]	966	1340[b]	715	919[c]	671
Vitamin E, *mg α-TE*	9.9[a]	5.9	10.3[a]	5.3	7.6[b]	4.9
Zinc, *mg*	11.1[a]	5.4	11.7[a]	5.5	9.1[b]	5.5
α-carotene, *μg RE*	661[a]	943	595[a]	690	323[b]	469
β-carotene, *μg RE*	3680[a]	3380	3340[a]	2780	1920[b]	2100
Fat, *% energy*	31.6[b]	7.3	33.6[a]	6.5	33.3[a]	8.7
Protein, *% energy*	15.2[a]	3.0	14.2[b]	2.8	13.0[c]	3.5
Carbohydrates, *% energy*	55.8[a]	9.3	54.3[b]	8.4	55.2[b]	10.5
Sweets, *% energy*	8.5[b]	8.4	11.3[a]	7.6	11.3[a]	11.0
Foods consumed, *n*	68[b]	11	81[a]	9	64[c]	13

[1] Data are presented as median and interquartile range (IQR), $n = 1219$ women due to missing values in some covariates. Medians in a row with superscripts without a common letter differ, $P < 0.003$ (Bonferroni correction for 20 multiple comparisons within class).

[2] The latent class model was adjusted for energy intake, nulliparous, smoker, white, education, and age. It included correlated errors between coffee and cream, and iced tea and sugar or honey.

[3] 1 kcal = 4.1868 kJ.

[4] 40 IU = 1 μg.

[5] RE, retinol equivalent.

0.66, 0.73)]. From the 1219 women, the number of women classified in the same dietary pattern with both classifications (diagonal of the contingency table represents agreement) was 287, 356, and 307 for the Health Conscious Western, Hard Core Western, and Prudent classes.

To illustrate what has been done previously in the literature to classify participants into dietary patterns derived by FA, we categorized the 4 factor scores into tertiles. Hence, their cross-tabulation yielded 81 patterns that were subjectively collapsed into the 3 groups obtained by the direct classification from LCA: Prudent, Health Conscious Western, and Hard Core Western. We classified as Prudent those with high or medium tertiles for FA-Prudent and FA-Prudent with coffee and alcohol and low tertiles for both FA-Southern and FA-Western. The group Hard Core Western was defined as those with high or medium tertiles for FA-Western and low tertiles for both FA-Prudent and FA-Prudent with coffee and alcohol. The remaining 71 patterns were considered Health Conscious Western. The agreement between the 3-LCA on 98 food items and this particular classification was $\kappa = 0.29$ (95% CI = 0.26, 0.33). From the 1219 women, the number of women classified in the same dietary pattern with both classifications was 73, 418, and 86 for the Health Conscious Western, Hard Core Western, and Prudent classes.

Discussion

We found that food items were grouped into 3 distinctive factors among pregnant women from PIN: Prudent, Western, and Southern. In addition, a 4th factor grouped coffee and alcohol with food items also considered in a Prudent pattern. Using LCA

to derive dietary patterns, women were grouped into 3 classes: Prudent and 2 types of Western diets, Hard core Western, and Health conscious Western. It seems there may be a group of women commonly in the Western pattern who, due to their

TABLE 3 Odds ratios for covariates of 3-LCA on 98 food items (PIN Study, 2000–2005)[1]

| | LCA-Prudent | | LCA-Health Conscious Western | |
Covariate	Odds ratio	P-value	Odds ratio	P-value
Nulliparous	1.7	0.011	1.2	0.287
Smoker	0.4	0.053	0.8	0.356
White	3.3	<0.001	2.4	0.001
Age, *y*				
25–29	2.7	0.016	2.3	0.005
30–34	8.7	<0.001	4.7	<0.001
35–47	9.1	<0.001	5.7	<0.001
Education				
Grades 13-16	3.9	0.003	1.8	0.027
≥Grade 17	11.6	<0.001	3.2	0.002
Pregravid BMI				
Underweight	2.1	0.013	1.6	0.117
Overweight	0.4	0.011	0.9	0.706
Obese	0.2	<0.001	0.7	0.198
Energy intake				
2nd quartile	1.1	0.687	2.1	0.013
3rd quartile	1.2	0.500	3.4	<0.001
4th quartile	0.7	0.386	3.4	0.003

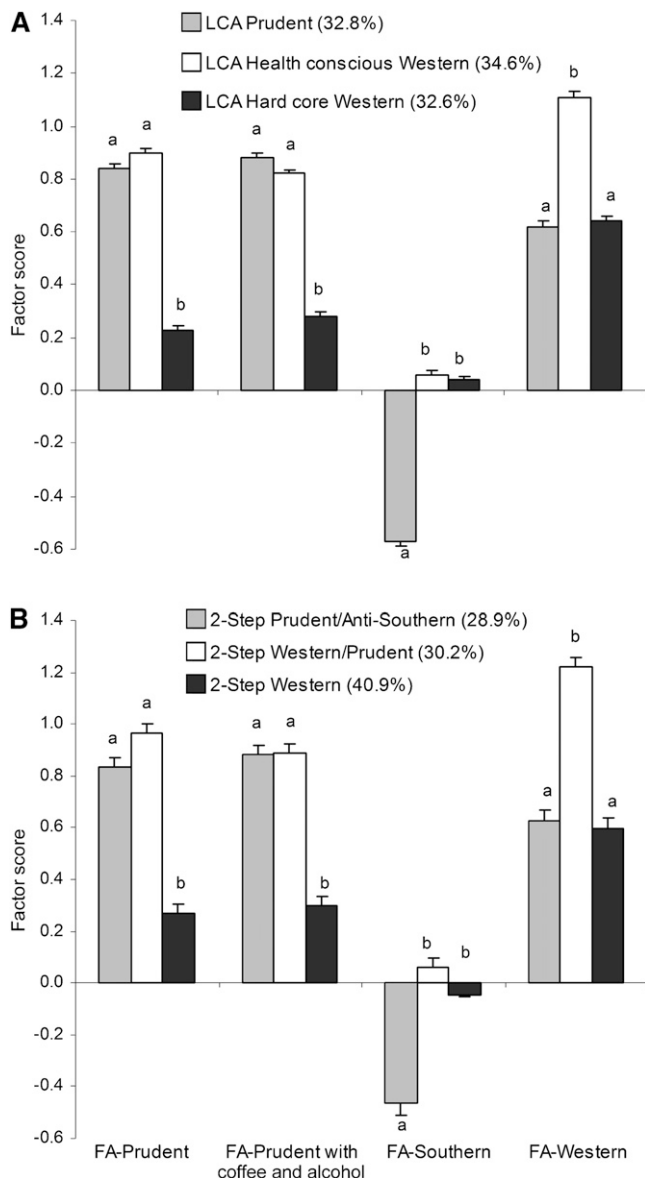[1] The reference class is LCA-Hard Core Western.

FIGURE 2 Factor scores' means by latent class from 3-LCA on 98 ordinal food items (*A*) or from 3-LCA on 4 factor scores (*B*). Bars represent means ± SEM, *n* = 1219. For each factor, labeled means without a common letter differ, *P* < 0.004 due to Bonferroni's correction.

pregnancies, are making an extra effort to eat fruits and vegetables. Though these women have a similar micronutrient intake compared with the Prudent class, they have a high-energy diet with a high percent of energy from fat and sweets. Prudent and Western patterns have been consistently derived in other populations (2,3) and the Southern pattern has been reported using the NHANES Survey (17). Because pregnancy is a life event characterized by cravings and aversions, it is possible to find unusual dietary patterns even after excluding women with extreme energy intake. However, because food items were categorized, the influence of extreme values is less of a concern. One possible reason for obtaining a Prudent with coffee and alcohol pattern among pregnant women is that other dietary patterns may have underreported coffee and alcohol consumption due to social desirability bias. Another reason is that women in the Prudent pattern were highly educated and they could be more aware than women in Southern and Western patterns that

occasional and very low consumption of coffee and alcohol has not been shown to be harmful to the fetus. The dietary patterns derived by LCA in our study are not comparable to the dietary patterns of other studies that have also used LCA, due to the fact that even after adjusting for covariates, dietary patterns are still population specific. Our study considered only pregnant women in central North Carolina, whereas the other studies included pregnant and nonpregnant women who were Indian (10) and British (8). However, it is interesting to note that the 3 studies found 1 group with considerably higher mean consumption of meat and our LCA Hard core Western class resembles the Convenience cluster identified by Padmadas et al. (10), which had a high preference of refined cereals, whole milk, snacks, and fast food.

In this population, the dietary patterns derived from grouping women into latent classes were well characterized by the dietary patterns derived from FA. Further, results from each method complemented our understanding about the dietary patterns. For example, using FA we identified typical foods from an American Southern cuisine, but when using LCA, we did not identify a "pure" Southern class. However, the LCA-Prudent class was characterized not only by a high FA-Prudent mean but also by a low FA-Southern mean, and by using the 2-step a posteriori classification we were able to better characterize this class as the Prudent/Anti-Southern class. On the other hand, the LCA-Hard core Western and LCA-Health conscious Western classes had different FA-Prudent and FA-Western means, but the FA-Southern means were not different. Having some classes that differ on score's means for some factors and not for others highlights the importance of considering all factors simultaneously and agrees with findings from the only 2 other studies (18,19) that have compared dietary patterns derived from factor and cluster analysis. For example, Costacou et al. (18) derived 4 principal components and 3 clusters and found 2 principal components' means (Mediterranean and Vegetarian) higher in cluster A (Mediterranean) than in their combined clusters BC (Low Mediterranean) but no mean differences between clusters A and BC for the principal components Sweets and Western.

In contrast to EFA and cluster analysis, CFA and LCA consider dietary patterns as latent variables and allow adjusting for covariates, modeling dietary data with different distributions jointly, specifying correlated errors, assessing goodness-of-fit, and conducting multi-group analysis. From CFA we found moderate correlations between Southern and Western, and Prudent and Prudent with coffee and alcohol and a low correlation between Prudent and Western, even though factors were initially derived to be uncorrelated using EFA. Factors can be correlated, because many of the factor loadings in the model were restricted to zero. Testing if they are correlated is important when factors will be used in subsequent analyses to characterize dietary patterns or when they will be derived and used in the same population repeatedly over time. When the factors are to be jointly categorized to derive mutually exclusive dietary patterns before further analysis, lack of independence is less of a concern.

The main advantage for using LCA over CFA is classifying participants into mutually exclusive groups directly as opposed to from the joint classification of the factors. When there are only 2 factors, an easy way to classify participants is from the cross-tabulation of the factor scores' quantiles. However, when there are more factors, LCA avoids making strong subjective decisions for collapsing all possible patterns. We found that there was high agreement between the direct classification from LCA on all 98 food items and the a posteriori one from the 2-step LCA on the 4 factors scores, despite the fact that only 25%

of the total variance was explained by the factors. On the other hand, there was a poor agreement with the subjective classification due to the LCA-Health Conscious Western group, which collapsed all the nonextreme patterns. Our experience suggests that with more than 2 factors, a subsequent LCA may be superior to "eyeballing" the cross-tabulation, which may be very time consuming and may not identify the best classification.

The benefit of the 2-step a posteriori procedure to classify participants into dietary patterns over LCA directly on food items is estimating the factor loadings. So, first FA helps explain which foods are eaten in combination and a subsequent LCA helps classify the individuals. However, the 2-step a posteriori classification procedure uses predicted factor scores as outcomes and not fixed variables. This could bias the estimates and the efficiency of standard errors by not taking into account the error in prediction. Potentially, we could fit a latent class mixture model (20–22) to simultaneously estimate the factor scores and latent classes. This approach also would allow within-class heterogeneity. However, modeling is computational intensive and hence not yet useful in practice. In summary, we recommend LCA on food items when the main interest is classification to study the effect of mutually exclusive classes and FA when the interest is to understand which foods are eaten in combination and to study associations between food patterns and outcomes. The proposed ad hoc 2-step classification using LCA on previously predicted factor scores from FA combines both benefits and seems promising.

Literature Cited

1. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol. 2002;13:3–9.

2. Newby PK, Tucker KL. Empirically derived eating patterns using factor or cluster analysis: a review. Nutr Rev. 2004;62:177–203.

3. Kant AK. Dietary patterns and health outcomes. J Am Diet Assoc. 2004;104:615–35.

4. Moeller SM, Reedy J, Millen AE, Dixon LB, Newby PK, Tucker KL, Krebs-Smith SM, Guenther PM. Dietary Patterns: Challenges and Opportunities in Dietary Patterns Research: an Experimental Biology Workshop, April 1, 2006. J Am Diet Assoc. 2007;107:1233–39.

5. Everitt B, Landau S, Leese M. Cluster analysis. Leese M, editor. London, New York: Arnold; Oxford University Press; 2001.

6. Rabe-Hesketh S, Skrondal A. Classical latent variable models for medical research. Stat Methods Med Res. 2007;17:5–32.

7. Hoffman K, Schulze MB, Boeing H, Altenburg HP. Dietary patterns: report of an international workshop. Public Health Nutr. 2002;5:89–90.

8. Fahey MT, Thane CW, Bramwell GD, Coward WA. Conditional Gaussian mixture modelling for dietary pattern analysis. J R Stat Soc Ser A Stat Soc. 2007;170:149–66.

9. Hagenaars JA, McCutcheon AL. Applied latent class analysis. Hagenaars JA, McCutcheon AL, editors. Cambridge, New York: Cambridge University Press; 2002.

10. Padmadas SS, Dias JG, Willekens FJ. Disentangling women's responses on complex dietary intake patterns from an indian cross-sectional survey: a latent class analysis. Public Health Nutr. 2006;9:204–11.

11. Knudsen VK, Orozova-Bekkevold IM, Mikkelsen TB, Wolff S, Olsen SF. Major dietary patterns in pregnancy and fetal growth. Eur J Clin Nutr. 2008;62:463–70.

12. Block G, Thompson FE, Hartman AM, Larkin FA, Guire KE. Comparison of two dietary questionnaires validated against multiple dietary records collected during a 1-year period. J Am Diet Assoc. 1992;92:686–93.

13. Skrondal A, Rabe-Hesketh S. Generalized latent variable modeling: multilevel, longitudinal, and structural equation models. Boca Raton (FL): Chapman & Hall/CRC; 2004.

14. SAS Institute Inc. SAS system for windows. Cary (NC): SAS Institute; 2002-2003;9.1.

15. Lanza ST, Collins LM, Lemmon DR, Schafer JL. PROC LCA: A SAS procedure for latent class analysis. Struct Equ Modeling. 2007;14:671.

16. Muthén LK, Muthén BO. Mplus user's guide. Fifth Edition. Los Angeles (CA): Muthén & Muthén. 1998–2006;5.1.

17. Tseng M, Breslow RA, DeVellis RF, Ziegler RG. Dietary patterns and prostate cancer risk in the national health and nutrition examination survey epidemiological follow-up study cohort. Cancer Epidemiol Biomarkers Prev. 2004;13:71–7.

18. Costacou T, Bamia C, Ferrari P, Riboli E, Trichopoulos D, Trichopoulou A. Tracing the Mediterranean diet through principal components and cluster analyses in the greek population. Eur J Clin Nutr. 2003;57:1378–85.

19. Newby PK, Muller D, Tucker KL. Associations of empirically derived eating patterns with plasma lipid biomarkers: a comparison of factor and cluster analysis methods. Am J Clin Nutr. 2004;80:759–67.

20. Muthen B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics. 1999;55:463–9.

21. Muthen B. Beyond SEM: general latent variable modeling. Behaviormetrika. 2002;29:81–117.

22. Muthen B. Should substance use disorders be considered as categorical or dimensional? Addiction. 2006;101 Suppl 1:6–16.