The Journal of Nutrition
Nutritional Epidemiology

Harmonization of Food-Frequency Questionnaires and Dietary Pattern Analysis in 4 Ethnically Diverse Birth Cohorts[1–3]

Russell J de Souza,[4,7,9]* Michael A Zulyniak,[5] Dipika Desai,[10] Mateen R Shaikh,[4] Natalie C Campbell,[5] Diana L Lefebvre,[5] Milan Gupta,[5,9] Julie Wilson,[11] Gita Wahi,[6] Stephanie A Atkinson,[6] Koon K Teo,[5] Padmaja Subbarao,[8] Allan B Becker,[12] Piushkumar J Mandhane,[13] Stuart E Turvey,[14] Malcolm R Sears,[5] and Sonia S Anand,[4,5,10] for the NutriGen Alliance Investigators

Departments of [4]Clinical Epidemiology and Biostatistics, [5]Medicine, and [6]Pediatrics, McMaster University, Hamilton, Ontario, Canada; [7]Department of Nutritional Sciences and [8]Hospital for Sick Children and Department of Pediatrics, University of Toronto, Toronto, Ontario, Canada; [9]Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Ontario, Canada; [10]Population Health Research Institute, Hamilton Health Sciences and McMaster University, Hamilton, Ontario, Canada; [11]Six Nations Health Services, Ohsweken, Ontario, Canada; [12]Department of Pediatrics and Child Health, Faculty of Medicine, University of Manitoba, Winnipeg, Manitoba, Canada; [13]Department of Pediatrics, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada; and [14]British Columbia Children's Hospital and Child and Family Research Institute, Department of Pediatrics, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

Abstract

Background: Canada is an ethnically diverse nation, which introduces challenges for health care providers tasked with providing evidence-based dietary advice.

Objectives: We aimed to harmonize food-frequency questionnaires (FFQs) across 4 birth cohorts of ethnically diverse pregnant women to derive robust dietary patterns to investigate maternal and newborn outcomes.

Methods: The NutriGen Alliance comprises 4 prospective birth cohorts and includes 4880 Canadian mother-infant pairs of predominantly white European [CHILD (Canadian Healthy Infant Longitudinal Development) and FAMILY (Family Atherosclerosis Monitoring In earLY life)], South Asian [START (SouTh Asian birth cohoRT)-Canada], or Aboriginal [ABC (Aboriginal Birth Cohort)] origins. CHILD used a multiethnic FFQ based on a previously validated instrument designed by the Fred Hutchinson Cancer Research Center, whereas FAMILY, START, and ABC used questionnaires specifically designed for use in white European, South Asian, and Aboriginal people, respectively. The serving sizes and consumption frequencies of individual food items within the 4 FFQs were harmonized and aggregated into 36 common food groups. Principal components analysis was used to identify dietary patterns that were internally validated against self-reported vegetarian status and externally validated against a modified Alternative Healthy Eating Index (mAHEI).

Results: Three maternal dietary patterns were identified—''plant-based,'' ''Western,'' and ''health-conscious''—which collectively explained 29% of the total variability in eating habits observed in the NutriGen Alliance. These patterns were strongly associated with self-reported vegetarian status (OR: 3.85; 95% CI: 3.47, 4.29; r^2 = 0.30, P < 0.001; for a plant-based diet), and average adherence to the plant-based diet was higher in participants in the fourth quartile of the mAHEI than in the first quartile (mean difference: 46.1%; r^2 = 0.81, P < 0.001).

Conclusion: Dietary data collected by using FFQs from ethnically diverse pregnant women can be harmonized to identify common dietary patterns to investigate associations between maternal dietary intake and health outcomes. *J Nutr* 2016;146:2343–50.

Keywords: FFQ, food frequency questionnaire, harmonization, multiethnic, PCA, prospective cohort, principal component analysis, dietary patterns, ethnicity

Introduction

Methodologic advances in dietary measurement in large epidemiologic studies, such as the development of valid and reproducible semiquantitative FFQs (1, 2), have facilitated the study of associations between dietary intake and health and disease outcomes, such as cancer and cardiovascular disease. This is often approached with a "reductionist" lens by examining associations between specific food items (3–6), single nutrients (5, 7), or sources of nutrients (8, 9) and health outcomes. This approach is reflective of public health approaches to food and nutrient recommendations. It has advanced our understanding and treatment of specific nutrient deficiency syndromes (e.g.,

Manuscript received June 7, 2016. Initial review completed July 14, 2016. Revision accepted September 9, 2016.
First published online October 5, 2016; doi:10.3945/jn.116.236729.

folate fortification to prevent neural tube defects) and facilitated the identification and removal of particularly harmful components of food from the food supply (e.g., removal of partially hydrogenated vegetable oils). However, long-term diet is likely a stronger determinant of diet-related chronic disease risk than the consumption of any single food item or nutrient (10), and thus single-food (e.g., coffee) or single-nutrient studies (e.g., dietary cholesterol) are often misleading (11, 12) because they fail to capture the complex interplay between foods and nutrients consumed as meals over long periods of time. To overcome the limitations of single-nutrient or single-food studies, the empirical derivation of dietary patterns—defined as "the quantities, proportions, variety or combinations of different foods and beverages in diets, and the frequency with which they are habitually consumed" (13)—has been proposed as a method to characterize exposure that more accurately reflects how we consume foods or nutrients, and these patterns can be assessed for their associations with health and disease (14–18).

Canada is an ethnically diverse nation (19), which introduces challenges for health care providers tasked with providing evidence-based dietary advice, because much of what we know about diet and disease is rooted in studies in white European populations. Dietary choice is closely tied to ethnicity (e.g., foods, cooking methods, and eating habits) (20), and the degree to which an individual or community consumes ethnically traditional foods can be influenced by immigration, acculturation, and duration of residency in a host country (21).

In preparation for investigations into the role of maternal nutrition on maternal and newborn outcomes in a multiethnic birth cohort consortium, we developed an approach to derive harmonized dietary patterns in pregnant women. This article describes the methods used to derive and to validate dietary patterns identified at a single time point in the cross-sectional analysis of a prospective birth cohort and outlines the unique challenges faced and the methodologic approaches used to address them.

Methods

Study population

The NutriGen Alliance is a multiethnic birth cohort consortium composed of 4 ethnically diverse cohorts of pregnant women representing several geographic regions across Canada. These cohorts were assembled to understand the early-life determinants of cardiometabolic risk, allergy, and asthma. Each cohort enrolled pregnant women in their second or third trimester and will follow the mother and infant from pregnancy through delivery and into childhood. The NutriGen Alliance provides a platform to investigate the joint influences of dietary intake, genetics, and the gut microbiome on the development of maternal and infant health outcomes in a Canadian context. As of February 2016, 5000 women with dietary data have been enrolled across the 4 cohort studies. There are 3047 pregnant women from the Canadian Healthy Infant Longitudinal Development (CHILD)[15] Study (22), representing 5 ethnic groups [white European (74%), East or South East Asian (12%), Aboriginal (4%), South Asian (3%), and African or other (12%) origin] recruited from 6 urban and rural Canadian cities (Vancouver, British Columbia; Edmonton, Alberta; Winnipeg, Manitoba; Morden, Manitoba; Winkler, Manitoba; Toronto, Ontario). There are 839 pregnant women included from the Family Atherosclerosis Monitoring In earLY life (FAMILY) Study (23), representing 5 ethnic groups [white European (74%), East or South East Asian (1%), Aboriginal (1%), South Asian (1%), and African or other (4%) origin] recruited from the Greater Hamilton Area, Ontario. There are 1006 South Asian mothers from the SouTh Asian birth cohoRT (START) (24) recruited from the Peel Region, Ontario, and 108 of an anticipated 300 Aboriginal mothers from the Aboriginal Birth Cohort (ABC) (25) recruited from the Six Nations Reserve, Ontario. Comprehensive clinical and dietary data from all pregnant women were collected from all 4 cohorts. Ethical approval was obtained for each study independently, and informed consent was obtained from all individual participants included in the study.

For this analysis, women who did not satisfactorily complete the FFQ [i.e., did not answer ≥10 questions (~6%)] or who reported an implausible energy intake (<500 or >6500 kcal/d) were excluded. One individual reported an implausibly high intake of a single food item (i.e., 64 servings of lettuce/d). The exclusion of this participant's FFQ, or replacement of the implausible reported value with a value equal to the 99th percentile of the "plausible" values (12 servings/d), produced identical dietary patterns; as such, the implausible value was included. The final number of women included in our analysis was 4880 (**Supplemental Table 1**).

Assessment of dietary intake and dietary patterns

FFQs. In the CHILD study, maternal diet was assessed by using a semiquantitative FFQ, adapted from the Fred Hutchinson Cancer Center tool (26). In the FAMILY, START, and ABC cohorts, semiquantitative FFQs developed for the Study of Health and Risk in Ethnic Groups (SHARE) Study (27) were used to assess maternal dietary intake during pregnancy, modified to capture ethnicity-specific foods (SHARE-based FFQs). ABC, FAMILY, and START FFQs were analyzed by using a database linked to the Canadian Nutrient File; the CHILD FFQ was analyzed by using the USDA nutrient database, modified for a Canadian setting (28), allowing estimation of energy intake. The development and validation of these tools have been described previously (29–31).

Frequency of consumption and serving size. The included FFQs used different serving size reference portions and frequency of consumption options. The CHILD FFQ provided respondents with categorical frequency options from which to choose (e.g., never to >2 servings/d), whereas in the SHARE-based FFQs, response categories were open-ended. Thus, we harmonized serving sizes of the SHARE-based FFQs to those in CHILD FFQ (**Supplemental Table 2**) (32, 33). Detailed steps describing the calculations and methods used to harmonize serving sizes across the cohorts are presented in **Supplemental Table 3**.

Food groupings. To create common food groups across the cohorts, individual FFQ items from each study were aggregated into groups of foods of similar nutrient profile and type (e.g., poultry, leafy greens, legumes, etc.). In some cases, food groups contained only a single item that uniquely reflected a particular dietary pattern (e.g., French fries reflect fast- and convenience-food consumption; **Supplemental Table 4**). We grouped foods in a way that has been used in previous dietary pattern analysis studies that examined associations between dietary habits and cardiometabolic conditions, allergies, or common clinical biomarkers

[1] Supported by a Canadian Institute for Health Research (RFA 201301FH6; 2013–2018) grant in Food & Health Population Health Research (RJdS, MG, GW, SAA, KKT, PS, ABB, MR Sears, SSA), a CIHR RCT Fellowship grant (MTP201410, MAZ), and AllerGen NCE, Inc. (MAZ, DLL, PS, ABB, PJM, SET, MR Sears, SSA). The SouTh Asian birth cohoRT (START) data was collected as part of a bilateral ICMR/CIHR funded program.

[2] Author disclosures: RJ de Souza has served as an external resource person to the WHO's Nutrition Guidelines Advisory Group on *trans* fats and saturated fats. The WHO paid for his travel and accommodation to attend meetings from 2012 to 2015 to present and discuss this work. He has also done contract research for the Canadian Institutes of Health Research's Institute of Nutrition, Metabolism, and Diabetes, Health Canada, and the WHO for which he received remuneration. MA Zulyniak, D Desai, MR Shaikh, NC Campbell, DL Lefebvre, M Gupta, J Wilson, G Wahi, SA Atkinson, KK Teo, P Subbarao, AB Becker, PJ Mandhane, SE Turvey, MR Sears, and SS Anand, no conflicts of interest.

[3] Supplemental Tables 1–8 and Supplemental Figure 1 are available from the "Online Supporting Material" link in the online posting of the article and from the same link in the online table of contents at http://jn.nutrition.org.

*To whom correspondence should be addressed. E-mail: rdesouz@mcmaster.ca.

[15] Abbreviations used: ABC, Aboriginal Birth Cohort; CHILD, Canadian Healthy Infant Longitudinal Development; DNBC, Danish National Birth Cohort; FAMILY, Family Atherosclerosis Monitoring In earLY life; mAHEI, modified Alternative Healthy Eating Index; MoBa, Norwegian Mother and Child Cohort Study; PCA, principal components analysis; SHARE, Study of Health and Risk in Ethnic Groups; START, SouTh Asian birth cohoRT.

(e.g., fasting plasma glucose, cholesterol, and TGs) (32–35). For example, bacon, breakfast sausages, low-fat and regular hotdogs, lunchmeats, and canned meats were combined into a single category called "processed meats."

Dietary pattern analysis

To identify dietary patterns within the FFQ data, we used the "psych" package (version 1.5.6) within R (version 3.1.2) to perform a principal components analysis (PCA) with an orthogonal varimax rotation (16). The statistical details of PCA as a means to reduce the dimensionality of the FFQ are beyond the scope of this article, but we refer interested readers to several excellent reviews (10, 33, 36–39). The number of dietary patterns retained was determined by visual inspection of scree plots in conjunction with eigenvalues (>1.0) and principal component interpretability (15, 40, 41). Three sensitivity analyses of dietary patterns were conducted (by using the same PCA approach as described) but limiting the sample to: *1*) women diagnosed with type 2 diabetes before their current pregnancy (n = 107; with or without hypertension), *2*) women diagnosed with hypertension before their current pregnancy (n = 190; with or without type 2 diabetes), and *3*) those without type 2 diabetes (n = 4720) or hypertension (n = 4632) before their current pregnancy.

We labeled each dietary pattern (i.e., groups of foods with similarly high factor loadings) with a descriptor that reflected the highly loaded food groups (e.g., "Western" compared with "prudent" patterns). The PCA scores for each pattern obtained for each individual represented how closely their food choices reflected each of the empirically derived dietary patterns, with a higher score reflecting a greater degree of adherence to that dietary pattern. Dietary pattern scores were adjusted to the mean total population energy intake by using the residual method (42, 43).

Dietary pattern adherence score

We created a dietary pattern adherence score that would more intuitively represent an individual's degree of adherence to each of the identified dietary patterns. To do this, "cardinal food groups" that characterized each dietary pattern were defined as those food groups with an absolute factor (dietary pattern) loading score ≥0.30 (**Supplemental Table 5**) (44, 45). Daily servings of each of the cardinal food groups were converted into quintiles, by using the distribution of servings within the study population, and assigned "quintile scores" from 1 (<20th percentile) to 5 (≥80th percentile) (**Supplemental Table 6**). These quintile scores for each of the food groups were summed to derive a numerical indicator of how closely an individual's diet reflected a given pattern. For example, "processed foods" had an absolute loading score ≥0.30 (0.55) for the "Western" diet but not for "plant-based" (−0.22) or "health-conscious" (0.13) dietary patterns. In this case, the quintile score for "processed foods" contributed to the total score for the Western dietary pattern but not to the plant-based or health-conscious dietary patterns. An individual's score for that specific diet was divided by the maximum score possible for the diet and multiplied by 100 to quantify the degree to which an individual adheres to each of the given dietary patterns (on a scale of 1–100) (**Table 1**).

Internal and external validation of dietary pattern scores

PCA summary scores were internally validated against self-reported vegetarian practice by using a logistic regression model. It was hypothesized that higher plant-based diet scores would be associated with higher odds of self-reported vegetarian status. PCA summary scores were externally validated against the modified Alternative Healthy Eating Index (mAHEI) (46) by comparing differences in mean scores between extreme quartile groups for PCA dietary patterns. An mAHEI diet score was calculated for each participant. Participants received 10 points for each of the following food items they consumed above (healthful foods) or below (less-healthful foods) a threshold: ≥5 servings vegetables, ≥4 servings fruit, ≥1 serving nuts or soy proteins, ≥3 servings whole grains, a ratio of ≥4 servings fish to 1 serving meat and eggs, and ≤0.5 servings of less-healthy foods (i.e., fried foods and processed meats); intermediate intakes were scored proportionally between 0 and 10. The maximum mAHEI score was 60. For this analysis, "processed meats" was included in the mAHEI "fried foods" category to capture *trans*-fat consumption. The mAHEI category for "alcohol consumption" was not included in this analysis of pregnant women. A design feature of the mAHEI [and other indexes, such as the Healthy Eating Index (47)] is that it rewards the consumption of "healthy" foods (5 items contribute to the score) more heavily than the avoidance of "unhealthy" foods (1 item contributes to the score); however, this feature does not preclude its usefulness as a valuable external validation tool for our derived dietary patterns. To do this, we compared mean plant-based, health-conscious, and Western diet scores between individuals in the lowest mAHEI-points quartile (i.e., <15 points, "least healthy") and those in the fourth mAHEI quartile (i.e., ≥45 points, "most healthy"). Differences in mean scores between diet groups were used to assess validity (e.g., higher plant-based scores were expected in those in the fourth mAHEI quartile than in those in the first quartile and higher Western scores were expected in those in the first mAHEI quartile than in those in the fourth quartile).

Results

PCA-derived patterns

Overall, 4880 valid FFQs were harmonized across 4 cohorts (Supplemental Table 1). The dimensionality of the food-group matrix was reduced from the 152–167 items queried within each individual study FFQ to 36 harmonized food groups (Supplemental Table 4), and 93 food items were common to all 4 instruments. A total of 59 and 70 foods were unique to CHILD and START FFQs, respectively; 64 were unique to the FAMILY FFQ; and 6 were unique to the ABC FFQ (**Figure 1**). The PCA identified 3 primary dietary patterns within the NutriGen Alliance with eigenvalues of 4.02, 3.20, and 3.05, which collectively explained 29% of the diet variability within the harmonized FFQ data set. The dietary patterns were classified as "plant-based," "Western," and "health-conscious" to emphasize the prominent food groups that defined each pattern. These categorizations reflect previously described dietary patterns in large cohort studies (Supplemental Table 5) (32–35, 48). In sensitivity analyses, the PCA-derived dietary patterns within subgroups of mothers who reported prepregnancy diabetes (n = 107) or hypertension (n = 190) were similar (e.g., plant-based, Western, and health-conscious) to those derived with the entire sample population or to those groups without hypertension (n = 4632) or type 2 diabetes (n = 4720).

The number of food groups with a loading factor ≥|0.30| were 10 for the plant-based, 13 for the Western, and 14 for the health-conscious patterns. The plant-based pattern was characterized by vegetables, legumes, fermented dairy, whole grains, nonmeat dishes, and a lack of red meat; the Western pattern had a high loading of sweets, red and processed meats, French fries, starchy vegetables, condiments, and sweet drinks; and the health-conscious pattern was characterized by seafood, poultry, and red meats; eggs; cruciferous vegetables; leafy greens; fruit; refined grains; stir-fried dishes; and condiments.

The dietary PCA scores for each individual were as follows: −1.8 to 6.1 (plant-based), −3.7 to 6.6 (Western), and −2.8 to 9.1 (health-conscious). When adjusted for total energy intake by using the residual method (49) to a mean total energy intake of 2000 kcal/d (equal to the mean energy intake of mothers in the NutriGen Alliance), the range of loading scores for dietary patterns were −2.2 to 5.5 (plant-based), −5.4 to 4.7 (Western), and −4.0 to 7.8 (health-conscious). Negative values indicate that an individual's dietary pattern is not generally reflective of the specific PCA-derived pattern (i.e., plant-based, Western, or health-conscious), and positive values indicate that an individual's dietary pattern is generally reflective of the specific PCA-derived pattern.

TABLE 1 Quantification of quintile dietary scores for each individual within the NutriGen Alliance cohort

Step	Description
1. Identify characteristic food groups for each diet	Identify the food groups in each dietary pattern that load most strongly (i.e., ≥|0.30|) and characterize them (e.g., ``processed meat´´ for the Western diet; see Supplemental Table 5).
2. Assign quintile scores for consumption frequency	Convert the serving frequencies for each characteristic food group to quintiles, from 1 to 5. This will give individuals in the lowest (<20%) and highest (≥80%) consumption frequencies for any food group a score of 1 and 5, respectively.
3. Calculate the participant quintile diet score for each diet	For each diet, sum the quintile scores of the foods that characterize the diet (identified in step 1). For foods that are inversely associated with a diet (e.g., ``meat´´ in the ``prudent´´ diet), individuals with a quintile score of 1, 2, 3, 4, or 5 would receive 5, 4, 3, or 2 points or 1 point, respectively, for that food group for that diet. When complete, each participant will have a total quintile score for each of the diets identified (e.g., plant-based, Western, and ``health-conscious´´).
4. Calculate the maximum quintile score for each diet	Multiply the total number of characteristic foods for each diet by 5. This is the maximum score for that diet. For example, the plant-based diet has 10 characteristic food groups; multiplied by 5 this gives a maximum score of 50 [e.g., 10 (food items) × 5 (maximum points for each food item) = 50 (maximum possible score)].
5. Determine relative adherence to dietary patterns	Divide each person's diet scores (step 3) by the maximum scores for each diet (step 4). This will reflect how closely each person's reported dietary patterns match each of the identified dietary patterns on a scale from 0% to 100%. For example, a person with scores of 34% plant-based, 75% Western, and 47% health-conscious would suggest that their diet is most similar to the Western pattern, with foods common to the prudent and health-conscious patterns consumed less frequently.

In a second PCA, indicators for each ethnicity were included in the PCA to estimate the effect of ethnicity on the derived dietary patterns (**Supplemental Table 7**). "Other vegetables" no longer loaded at ≥0.30 within the health-conscious dietary pattern, but the dietary patterns were equivalent to those observed in the original PCA reported in Supplemental Table 5. Univariate regression showed that the summary scores from the PCA that did not include ethnicity correlated strongly with the summary scores when ethnicity was included: plant-based (r^2 = 0.97, P < 0.001), Western (r^2 = 0.94, P < 0.001), and health-conscious (r^2 = 0.96, P < 0.001).

Diet scores

The maximum adherence diet scores for the plant-based, Western, and health-conscious diets were 50, 65, and 70 total quintile points, respectively. Energy-adjusted PCA scores were well correlated with the energy-adjusted, quintile-based diet scores (r^2 values: plant-based = 0.75, P < 0.001; Western = 0.47, P < 0.001; health-conscious = 0.51, P < 0.001).

By using this scoring method, the plant-based diet had a mean adherence of 57.1%, the Western diet had a mean adherence of 58.6%, and the health-conscious diet had a mean adherence of 59.2% (**Supplemental Table 8**). There were clear differences between the 4 major ethnic groups (i.e., those with ≥200 participants) with respect to average dietary pattern scores. South Asians most closely adhered to the plant-based diet (mean ± SD score: 77.9% ± 12.5%), whereas East and South East Asians were least adherent (47.7% ± 10.3%). The Western diet was most strongly adhered to by Aboriginal people (63.3% ± 9.2%) and least strongly by South Asians (47.6% ± 9.5%). The health-conscious diet was strongly followed by East or South East Asians (66.9% ± 9.2%) and least strongly adhered to by South Asians (51.5% ± 10.1%).

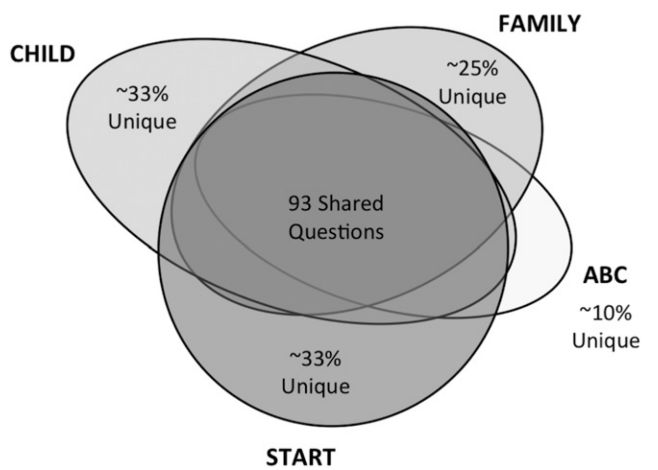

FIGURE 1 Venn diagram of the similarity and differences between the food items queried within individual study cohorts (i.e., ABC, CHILD, FAMILY, and START) that comprise the NutriGen Alliance cohort (n = 4880). Unlisted similarities of foods questioned between studies are ≤10% similar. ABC, Aboriginal Birth Cohort Study; CHILD, Canadian Healthy Infant Longitudinal Development Study; FAMILY, Family Atherosclerosis Monitoring In earLY life Study; START, SouTh Asian birth cohoRT Study.

Validation assessments

Internal validity. To assess the internal validity and robustness of the harmonized NutriGen dietary patterns, we also derived the patterns within each of the individual cohorts separately (ABC, CHILD, FAMILY, and START) and found that the cohort-specific dietary patterns reflected those of the harmonized NutriGen cohort. CHILD presented 2 primary diets: ovopescetarian (plant-based with fish and eggs) and Western; FAMILY presented 2 primary diets: health-conscious and Western; START presented 3 primary diets: plant-based, Western, and health-conscious; and ABC presented 2 primary diets: health-conscious and Western.

The unadjusted and energy-adjusted PCA summary scores were validated against the self-reported dichotomous variable "vegetarian status" (which included self-reports of lactovegetarians, ovovegetarians, vegetarians, and vegans). For the unadjusted PCA scores, a single-unit increase in the plant-based diet PCA score was associated with 3-fold greater odds of self-reporting as a "vegetarian" or being a nonconsumer of meat (OR: 3.35; 95% CI: 3.03, 3.68; $r^2 = 0.26$, $P < 0.001$), whereas a single-unit increase in either the Western (OR: 0.36; 95% CI: 0.31, 0.42; $r^2 = 0.08$, $P < 0.001$) or health-conscious (OR: 0.60; 95% CI: 0.53, 0.68; $r^2 = 0.02$; $P < 0.001$) diets were negatively associated with self-reported vegetarian status. For energy-adjusted PCA scores, the plant-based diet was similarly positively associated with self-reported vegetarian status (OR: 3.85; 95% CI: 3.47, 4.29; $r^2 = 0.30$, $P < 0.001$), and both the Western (OR: 0.29; 95% CI: 0.24, 0.34; $r^2 = 0.08$, $P < 0.001$) and health-conscious (OR: 0.67; 95% CI: 0.59, 0.75; $r^2 = 0.01$, $P < 0.001$) diets were negatively associated with self-reported vegetarian status.

External validity. Individuals in the lowest (least healthy) mAHEI quartile had a lower adherence to the plant-based (mean diet score: 35.8% ± 7.9% in quartile 1 compared with 81.8% ± 11.2% in quartile 4; $r^2 = 0.81$, $P < 0.001$) and health-conscious (41.8% ± 8.7% in quartile 1 compared with 56.0% ± 13.6% in quartile 4; $r^2 = 0.23$, $P < 0.001$) dietary patterns than those in the highest (most healthy) mAHEI quartiles (**Supplemental Figure 1**). Individuals in the lowest mAHEI quartile adhered more strongly to the Western dietary pattern (57.7% ± 12.9% in quartile 1 compared with 52.9% ± 15.0% in quartile 4; $r^2 = 0.02$, $P < 0.001$) than those in the highest mAHEI quartile.

Discussion

This study describes, to our knowledge, a novel application of a methodologic approach to harmonize dietary data collected with cohort-specific, independently validated FFQs across 4 ethnically diverse birth cohorts. This effort represents an exemplar readily extensible to settings outside of Canada. Such harmonization efforts are increasingly common (50) for other types of data, and directed criteria and guidelines have been developed (i.e., PhenX Toolkit, available from: https://www.phenxtoolkit.org) to facilitate the pooling of maternal and infant data across birth cohorts (51).

We identified 3 unique dietary patterns, which we named "plant-based," "Western," and "health-conscious," that closely resemble previously documented patterns in a cohort of the Toronto Nutrigenomics and Health Study, a multiethnic cohort of young Canadian men and women residing in the Greater Toronto area ($n = 1153$) (52). In this study, 3 patterns—prudent, Western, and Eastern—were identified by using a single semiquantitative FFQ and explained 16% of the dietary variance, which is less than the 29% that our harmonized analysis explained. Although dietary pattern studies typically identify 2 major dietary patterns (14, 15, 53), the similarity of the NutriGen and Toronto Nutrigenomics and Health Study dietary patterns likely reflects a similar ethnic composition of the cohorts.

In the present study, we faced the challenge of post hoc harmonization. An excellent example of forward thinking on harmonization is provided by the merger of FFQ data collected from 2 birth cohorts: the Danish National Birth Cohort (DNBC; $n = 70{,}183$) and the Norwegian Mother and Child Cohort Study (MoBa; $n = 87{,}000$) (54). Despite some unique regional items within each FFQ, food items were comparable and aggregated into common higher-order food groups (e.g., fruit, legumes, etc.). The harmonization was aided by a high degree of ethnic homogeneity and cooperation between the DNBC and MoBa study teams during MoBa's development, which facilitated the development of an FFQ that was very similar to the DNBC FFQ. Nevertheless, we show that retrospective harmonization across diverse ethnic cohorts is possible (27). Furthermore, we were well powered to detect small differences (i.e., 3–4%) in dietary pattern adherence, even within ethnic groups in which one may expect homogeneity of dietary intakes (Supplemental Table 8).

The NutriGen Alliance dietary patterns showed good internal and external validity. The plant-based score was strongly associated with self-reported vegetarian status, although even this association is likely diluted because "vegetarian" was inconsistently defined across the cohorts: for example, in the CHILD cohort, pregnant women "reported abstinence from meats," whereas in the FAMILY, START, and ABC cohorts a "vegetarian" status question was asked. A single-unit increase in the plant-based score increased the odds of being a vegetarian (i.e., non–meat eater) by >3-fold; conversely, a unit increase in the Western diet reduced these odds by ~70%. The health-conscious diet score was less useful at predicting vegetarian status: a single-unit increase reduced the likelihood of vegetarian status by ~40%. Our analyses suggest that 3 dietary patterns can accurately discriminate between groups who consume distinct dietary patterns (i.e., vegetarians from non-vegetarians).

Our external validation against the mAHEI (46), which has been used previously to assess diet quality in pregnant women (55), found that mAHEI score was associated with greater adherence to the plant-based and health-conscious dietary patterns and lower adherence to the Western diet, which confirms alignment of our dietary patterns with external methods for assessing diet quality. Total energy was adjusted for in the analysis to reduce confounding and random error due to differences in food intake resulting from differences in body size, metabolic efficiency, and physical activity. In some studies, it may be desirable to not account for energy if excess food energy is causally implicated in the relation between certain foods or diets and specific outcomes (e.g., when modeling the association between high-energy sugar-sweetened beverages and obesity). However, it is often desirable to isolate the effect of a specific food item or nutrient from its unspecific contribution to total energy intake when assessing diet-disease associations (e.g., the unique contribution of *trans* fat from other energy-containing nutrients of the foods in which it is contained). In a comparison of dietary patterns derived with and without energy adjustment, Northstone et al. (43) found that "white bread" was positively loaded on the "processed diet" in an unadjusted model but, after energy-adjustment, was negatively loaded for the "health-conscious diet." Balder et al. (56) proposed that, in an energy-adjusted model, the avoidance of high-energy foods in favor of low-energy healthy alternatives (i.e., choosing lower-energy-dense brown bread rather than high-energy-dense white bread) is a salient feature of "health-conscious" diets; therefore, energy-unadjusted and -adjusted models characterize similar dietary patterns and are therefore comparable. In the present study, the likelihood of vegetarian status according to participant plant-based, Western, and health-conscious dietary pattern scores were comparable in unadjusted and energy-adjusted models. It has been recommended that energy adjustment be performed post-PCA (43, 56) to simplify the interpretation of the results.

A salient feature of our cohort was its ethnic diversity. Downstream dietary pattern analyses within diverse cohorts often require adjustment for ethnicity (16, 57), which is most often accomplished by including ethnicity as a covariate in multivariable models. An alternative approach is to include ethnicity in the PCA when deriving dietary patterns, which would help account for the tight conceptual linking of diet and culture. In the present study, including ethnicity in the PCA only marginally affected the dietary patterns (Supplemental Table 7), and these dietary pattern scores derived with ethnicity correlated strongly with those derived without including ethnicity in the PCA ($r^2 \geq 0.94$). However, adjusting for ethnicity in the PCA makes it impossible to assess whether the associations between dietary patterns and health outcomes are modified by ethnicity. Thus, leaving ethnicity out of the PCA derivation of dietary patterns gives maximum flexibility to the researcher in future analyses of dietary patterns and health outcomes.

We developed a diet score to simplify the interpretation of the dietary patterns. Individual summary scores for each principal component reflect how closely each person follows a given dietary pattern (e.g., prudent, Western, and health-conscious), but factor loading scores are difficult to interpret because the score and the range of scores vary across dietary patterns. However, by only focusing on foods that contribute strongly to each dietary pattern (i.e., "cardinal features" with loading scores $\geq|0.30|$) and by calculating a diet score ranging from 1% (minimal adherence) to 100% (full adherence) for each of the diets, the dietary patterns scores have the straightforward interpretation of how closely dietary habits reflects each of the empirically derived plant-based, Western, and health-conscious diets. Because this intuitive approach loses little information, and there is a strong correlation between diet scores and PCA scores, the derived dietary scores can be used in place of the summary scores for regression analyses for easier interpretability and presentation of results.

Our study has some limitations. Maternal diet was collected by using a single administration of a self-completed FFQ. Although these instruments have been validated, recall bias and measurement error are acknowledged limitations of these tools. However, given the prospective nature of our planned analyses (i.e., the association between maternal food choices and future maternal and infant health) and the large number of individuals involved, we anticipate that random error will be comparable to other prospective cohort studies that have used this method of assessment. This problem can be attenuated if multiple measures of diet are available (58). In addition, scree plots identified 3 patterns—with eigenvalues >3.0 each that collectively explained 29% of the dietary variability—of several possible patterns detected by the PCA. Minor patterns, which explain a smaller degree of variation, were not retained. Future studies may need to increase the number of dietary patterns to characterize less-common dietary patterns in their study population of interest. We addressed the issue of reverse confounding such that a pre-existing medical condition, such as prepregnancy diabetes or hypertension, may influence dietary intake in pregnancy by conducting a sensitivity analysis in those women with type 2 diabetes or hypertension. Our analyses showed that, within each subgroup, the PCA-derived dietary patterns did not differ substantially from each other or from our patterns derived by using the complete sample. In addition, although nutrients were not the focus of the present study, future analyses in these 4 harmonized birth cohorts that focus on macro- and micro-nutrient analyses will require harmonization of the nutrient data when different nutrient databases were used.

In conclusion, this study addressed a novel challenge, the merging and harmonization of multiple FFQ data sets collected from pregnant women of diverse ethnicities with the use of an established methodology for dietary pattern analysis. We showed a valid approach to merge both similar and distinct FFQ data sets to investigate how maternal diet during pregnancy contributes to maternal and infant health and disease.

Acknowledgments

Additional NutriGen Alliance study investigators are as follows: Jospeh Beyene, Judah A Denburg, Sarah McDonald, Andrew Mente, David Meyre, Katherine Morrison, Guillaume Pare, and Michael Surette. The ABC study investigators are Sonia S Anand, Gita Wahi, and Julie Wilson; the ABC research staff members are A Darlene Davis, Trista Hill, Dorothy Green, and Lori Davis Hill. The CHILD study investigators are as follows: Sonia S Anand, Allan B Becker, Dean Befus, Jeff Brook, Wen-Yi Lou, Piush J Mandhane, Gregory Miller, Andrew Sandford, Malcolm R Sears, Padmaja Subbarao, and Stuart E Turvey; and the CHILD research staff members are Diana L Lefebvre and Joanne M Duncan. The START Canada investigators are as follows: Sonia S Anand, Rebecca Anglin, Joseph Beyene, Milan Gupta, Guillaume Pare, and Katherine Morrison.

RJdS, MG, GW, SAA, KKT, PS, ABB, PJM, SET, MR Sears, and SSA developed the overall research plan; all listed authors and cohort investigators contributed to data collection; RJdS, MAZ, DD, MR Shaikh, NCC, and DLL contributed to analysis of the data for this manuscript; and RJdS, MAZ, and SSA prepared the final manuscript. All authors provided input into data interpretation and read and approved the final manuscript.

References

1. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol 1985;122:51–65.
2. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. Am J Epidemiol 1992;135:1114–26; discussion: 27–36.
3. Schernhammer ES, Hu FB, Giovannucci E, Michaud DS, Colditz GA, Stampfer MJ, Fuchs CS. Sugar-sweetened soft drink consumption and risk of pancreatic cancer in two prospective cohorts. Cancer Epidemiol Biomarkers Prev 2005;14:2098–105.
4. Fung TT, Malik V, Rexrode KM, Manson JE, Willett WC, Hu FB. Sweetened beverage consumption and risk of coronary heart disease in women. Am J Clin Nutr 2009;89:1037–42.
5. Yudin J, Morland J. Sugar intake and myocardial infarction. Am J Clin Nutr 1967;20:503–6.
6. Giovannucci E, Rimm EB, Stampfer MJ, Colditz GA, Ascherio A, Willett WC. Intake of fat, meat, and fiber in relation to risk of colon cancer in men. Cancer Res 1994;54:2390–7.
7. Willett WC, Hunter DJ, Stampfer MJ, Colditz G, Manson JE, Spiegelman D, Rosner B, Hennekens CH, Speizer FE. Dietary fat and fiber in relation to risk of breast cancer: an 8-year follow-up. JAMA 1992;268:2037–44.

8. Howe GR, Jain M, Miller AB. Dietary factors and risk of pancreatic cancer: results of a Canadian population-based case-control study. Int J Cancer 1990;45:604–8.
9. Hu FB, Stampfer MJ, Manson JE, Rimm E, Colditz GA, Speizer FE, Hennekens CH, Willett WC. Dietary protein and risk of ischemic heart disease in women. Am J Clin Nutr 1999;70:221–7.
10. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol 2002;13:3–9.
11. Keys A. Diet and blood cholesterol in population surveys–lessons from analysis of the data from a major survey in Israel. Am J Clin Nutr 1988;48:1161–5.
12. Sacks FM, Castelli WP, Donner A, Kass EH. Plasma lipids and lipoproteins in vegetarians and controls. N Engl J Med 1975;292:1148–51.
13. US Department of Health and Human Services; US Department of Agriculture. 2015–2020 Dietary guidelines for Americans. 8th ed. 2015. Available from: https://health.gov/dietaryguidelines/2015/guidelines/.
14. Heidemann C, Schulze MB, Franco OH, van Dam RM, Mantzoros CS, Hu FB. Dietary patterns and risk of mortality from cardiovascular disease, cancer, and all causes in a prospective cohort of women. Circulation 2008;118:230–7.
15. Hu FB, Rimm EB, Stampfer MJ, Ascherio A, Spiegelman D, Willett WC. Prospective study of major dietary patterns and risk of coronary heart disease in men. Am J Clin Nutr 2000;72:912–21.
16. Iqbal R, Anand S, Ounpuu S, Islam S, Zhang X, Rangarajan S, Chifamba J, Al-Hinai A, Keltai M, Yusuf S, et al. Dietary patterns and the risk of acute myocardial infarction in 52 countries: results of the INTERHEART study. Circulation 2008;118:1929–37.
17. Kontogianni MD, Panagiotakos DB. Dietary patterns and stroke: a systematic review and re-meta-analysis. Maturitas 2014;79:41–7.
18. Okubo H, Miyake Y, Sasaki S, Tanaka K, Murakami K, Hirota Y, Osaka Maternal and Child Health Study Group, Kanzaki H, Kitada M, Horikoshi Y, et al.; Child Health Study Group. Maternal dietary patterns in pregnancy and fetal growth in Japan: the Osaka Maternal and Child Health Study. Br J Nutr 2012;107:1526–33.
19. Fearon JD. Ethnic and cultural diversity by country*. J Econ Growth 2003;8:195–222.
20. Anand SS, Hawkes C, de Souza RJ, Mente A, Dehghan M, Nugent R, Zulyniak MA, Weis T, Bernstein AM, Krauss RM, et al. Food consumption and its impact on cardiovascular disease: importance of solutions focused on the globalized food system: a report from the workshop convened by the World Heart Federation. J Am Coll Cardiol 2015;66:1590–614.
21. Lesser IA, Gasevic D, Lear SA. The association between acculturation and dietary patterns of South Asian immigrants. PLoS One 2014;9:e88495.
22. Subbarao P, Anand SS, Becker AB, Befus AD, Brauer M, Brook JR, Denburg JA, HayGlass KT, Kobor MS, Kollmann TR, et al. The Canadian Healthy Infant Longitudinal Development (CHILD) Study: examining developmental origins of allergy and asthma. Thorax 2015;70:998–1000.
23. Morrison KM, Atkinson SA, Yusuf S, Bourgeois J, McDonald S, McQueen MJ, Persadie R, Hunter B, Pogue J, Teo K. The Family Atherosclerosis Monitoring In earLY life (FAMILY) Study: rationale, design, and baseline data of a study examining the early determinants of atherosclerosis. Am Heart J 2009;158:533–9.
24. Anand SS, Vasudevan A, Gupta M, Morrison K, Kurpad A, Teo KK, Srinivasan K. Rationale and design of South Asian birth cohort (START): a Canada-India collaborative study. BMC Public Health 2013;13:79.
25. Wahi G, Wilson J, Miller R, Anglin R, McDonald S, Morrison KM, Teo KK, Anand SS. Aboriginal Birth Cohort (ABC): a prospective cohort study of early life determinants of adiposity and associated risk factors among Aboriginal people in Canada. BMC Public Health 2013;13:608.
26. Fred Hutchinson Cancer Research Center. Food frequency questionnaires (FFQ) [Internet] [updated 2014 Nov 3; cited 2016 Jun 7]. Available from: http://sharedresources.fhcrc.org/services/food-frequency-questionnaires-ffq.
27. Anand SS, Yusuf S, Vuksan V, Devanesen S, Teo KK, Montague PA, Kelemen L, Yi C, Lonn E, Gerstein H, et al. Differences in risk factors, atherosclerosis and cardiovascular disease between ethnic groups in Canada: the Study of Health Assessment and Risk in Ethnic Groups (SHARE). Indian Heart J 2000;52(7, Suppl):S35–43.
28. Nutrition Research Division (Health Canada). Canadian Nutrient File (CNF) [Internet]. 2010. [cited 2016 Sep 20] Available from: http://www.hc-sc.gc.ca/fn-an/nutrition/fiche-nutri-data/cnf_aboutus-aproposdenous_fcen-eng.php.
29. Korpi-Hyövälti E, Schwab U, Laaksonen DE, Linjama H, Heinonen S, Niskanen L. Effect of intensive counselling on the quality of dietary fats in pregnant women at high risk of gestational diabetes mellitus. Br J Nutr 2012;108:910–7.
30. Kelemen LE, Anand SS, Vuksan V, Yi Q, Teo KK, Devanesen S, Yusuf S. Development and evaluation of cultural food frequency questionnaires for South Asians, Chinese, and Europeans in North America. J Am Diet Assoc 2003;103:1178–84.
31. Patterson RE, Kristal AR, Tinker LF, Carter RA, Bolton MP, Agurs-Collins T. Measurement characteristics of the Women's Health Initiative food frequency questionnaire. Ann Epidemiol 1999;9:178–87.
32. Nettleton JA, Polak JF, Tracy R, Burke GL, Jacobs DR. Dietary patterns and incident cardiovascular disease in the Multi-Ethnic Study of Atherosclerosis. Am J Clin Nutr 2009;90:647–54.
33. Fung TT, Rimm EB, Spiegelman D, Rifai N, Tofler GH, Willett WC, Hu FB. Association between dietary patterns and plasma biomarkers of obesity and cardiovascular disease risk. Am J Clin Nutr 2001;73:61–7.
34. Hu FB, Rimm E, Smith-Warner SA, Feskanich D, Stampfer MJ, Ascherio A, Sampson L, Willett WC. Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. Am J Clin Nutr 1999;69:243–9.
35. Gadgil MD, Anderson CAM, Kandula NR, Kanaya AM. Dietary patterns in Asian Indians in the United States: an analysis of the metabolic syndrome and atherosclerosis in South Asians Living in America Study. J Acad Nutr Diet 2014;114:238–43.
36. Huijbregts P, Feskens E, Räsänen L, Fidanza F, Nissinen A, Menotti A, Kromhout D. Dietary pattern and 20 year mortality in elderly men in Finland, Italy, and The Netherlands: longitudinal cohort study. BMJ 1997;315:13–7.
37. Huijbregts PP, Feskens EJM, Kromhout D. Dietary patterns and cardiovascular risk factors in elderly men: the Zutphen Elderly Study. Int J Epidemiol 1995;24:313–20.
38. Kant AK. Indexes of overall diet quality: a review. J Am Diet Assoc 1996;96:785–91.
39. Millen BE, Quatromoni PA, Gagnon DR, Cupples LA, Franz MM, D'Agostino RB. Dietary patterns of men and women suggest targets for health promotion: the Framingham Nutrition Studies. Am J Health Promot 1996;11:42–52.
40. Kim J-O, Mueller CW. Factor analysis: statistical methods and practical issues. Newbury Park (CA): Sage; 1978.
41. Satija A, Hu FB, Bowen L, Bharathi AV, Vaz M, Prabhakaran D, Reddy KS, Ben-Shlomo Y, Davey Smith G, Kinra S, et al. Dietary patterns in India and their association with obesity and central obesity. Public Health Nutr 2015;18:3031–41.
42. Willett W, Stampfer M. Implications of total energy intake for epidemiologic analysis. In: Willett WW, editor. Monographs in epidemiology and biostatistics: nutritional epidemiology. 2nd ed. New York: Oxford University Press; 1998. p. 273–301.
43. Northstone K, Ness AR, Emmett PM, Rogers IS. Adjusting for energy intake in dietary pattern investigations using principal components analysis. Eur J Clin Nutr 2008;62:931–8.
44. Catsburg C, Kim RS, Kirsh VA, Soskolne CL, Kreiger N, Rohan TE. Dietary patterns and breast cancer risk: a study in 2 cohorts. Am J Clin Nutr 2015;101:817–23.
45. Chocano-Bedoya PO, O'Reilly EJ, Lucas M, Mirzaei F, Okereke OI, Fung TT, Hu FB, Ascherio A. Prospective study on long-term dietary patterns and incident depression in middle-aged and older women. Am J Clin Nutr 2013;98:813–20.
46. Dehghan M, Mente A, Teo KK, Gao P, Sleight P, Dagenais G, Avezum A, Probstfield JL, Dans T, Yusuf S, et al. Relationship between healthy diet and risk of cardiovascular disease among patients on drug therapies for secondary prevention: a prospective cohort study of 31 546 high-risk individuals from 40 countries. Circulation 2012;126:2705–12.
47. Guenther PM, Casavale KO, Reedy J, Kirkpatrick SI, Hiza HA, Kuczynski KJ, Kahle LL, Krebs-Smith SM. Update of the Healthy Eating Index: HEI-2010. J Acad Nutr Diet 2013;113:569–80.
48. Mozaffarian D, Hao T, Rimm EB, Willett WC, Hu FB. Changes in diet and lifestyle and long-term weight gain in women and men. N Engl J Med 2011;364:2392–404.
49. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. Am J Clin Nutr 1997;65(Suppl):1220S–8S.
50. Thompson A. Thinking big: large-scale collaborative research in observational epidemiology. Eur J Epidemiol 2009;24:727–31.

51. Whitehead NS, Hammond JA, Williams MA, Huggins W, Hoover S, Hamilton CM, Ramos EM, Junkins HA, Harlan WR, Hogue CJ. The PhenX Toolkit pregnancy and birth collections. Ann Epidemiol 2012;22:753–8.
52. Brenner DR, Boucher BA, Kreiger N, Jenkins D, El-Sohemy A. Dietary patterns in an ethnoculturally diverse population: of young Canadian adults. Can J Diet Pract Res 2011;72:e161–8.
53. Li F, Hou L-n, Chen W, Chen P-l, Lei C-y, Wei Q, Tan W-l, Zheng S-b. Associations of dietary patterns with the risk of all-cause, CVD and stroke mortality: a meta-analysis of prospective cohort studies. Br J Nutr 2015;113:16–24.
54. Olsen SF, Eva Birgisdottir B, Halldorsson TI, Brantsaeter AL, Haugen M, Torjusen H, Petersen SB, Strøm M, Meltzer HM. Possibilities and considerations when merging dietary data from the world's two largest pregnancy cohorts: the Danish National Birth Cohort and the Norwegian Mother and Child Cohort Study. Acta Obstet Gynecol Scand 2014;93:1131–40.
55. Han CY, Colega M, Quah EPL, Chan YH, Godfrey KM, Kwek K, Saw S-M, Gluckman PD, Chong Y-S, Chong MF-F. A healthy eating index to measure diet quality in pregnant women in Singapore: a cross-sectional study. BMC Nutr 2015;1:1–11.
56. Balder HF, Virtanen M, Brants HAM, Krogh V, Dixon LB, Tan F, Mannisto S, Bellocco R, Pietinen P, Wolk A. Common and country-specific dietary patterns in four European cohort studies. J Nutr 2003;133:4246–51.
57. Thompson JM, Wall C, Becroft DM, Robinson E, Wild CJ, Mitchell EA. Maternal dietary patterns in pregnancy and the association with small-for-gestational-age infants. Br J Nutr 2010;103:1665–73.
58. Paeratakul S, Popkin BM, Kohlmeier L, Hertz-Picciotto I, Guo X, Edwards LJ. Measurement error in dietary data: implications for the epidemiologic study of the diet-disease relationship. Eur J Clin Nutr 1998;52:722–7.