Performance of a Food Frequency Questionnaire for Estimating Ultraprocessed Food Intake According to the Nova Classification System in the United States NIH-AARP Diet and Health Study

Journal of Nutrition

Loftfield, Erikka; Zhang, Pei; O'Connell, Caitlin P.; Kahle, Lisa L.; Herrick, Kirsten et al

https://doi.org/10.1016/j.tjnut.2025.04.029

The Journal of Nutrition 155 (2025) 2376–2384

THE JOURNAL OF NUTRITION

journal homepage: https://jn.nutrition.org/

Nutritional Epidemiology

Performance of a Food Frequency Questionnaire for Estimating Ultraprocessed Food Intake According to the Nova Classification System in the United States NIH-AARP Diet and Health Study

Erikka Loftfield [1,*,†], Pei Zhang [2,†], Caitlin P O'Connell [1], Lisa L Kahle [3], Kirsten Herrick [4], Leila Abar [1], Neha Khandpur [5,6,7,8], Eurídice Martínez Steele [7], Hyokyoung G Hong [2]

[1] *Metabolic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, United States;* [2] *Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, United States;* [3] *Information Management Services, Rockville, MD, United States;* [4] *Division of Cancer Control and Population Sciences, Risk Factor Assessment Branch, National Cancer Institute, National Institutes of Health, Rockville, MD, United States;* [5] *Division of Human Nutrition and Health, Wageningen University, the Netherlands;* [6] *Department of Nutrition, Harvard T.H. Chan School of Public Health, United States;* [7] *NUPENS, University of São Paulo, Brazil;* [8] *NUPENS, School of Public Health, University of São Paulo, São Paulo, Brazil*

A B S T R A C T

Background: How accurately ultraprocessed food (UPF) intake is measured by food frequency questionnaire (FFQ) is unknown.
Objectives: We evaluated the performance of a 124-item FFQ for estimating UPF intake with Nova.
Methods: In 1311 NIH-AARP Diet and Health Study participants who completed 2 nonconsecutive 24-hour dietary recalls (24HDR) and 2 FFQ, we used 2 methods to estimate UPF intake. First, 4 experts independently matched FFQ items to food descriptions from another FFQ classified previously according to Nova. Discordant descriptions were discussed, and a consensus reached. Second, data from the 1994–1996 Continuing Survey of Food Intake of Individuals were used to match disaggregated FFQ items to Food and Nutrient Database for Dietary Studies (FNDDS) and standard reference (SR) codes. Nova classification of FNDDS and SR codes was based on database linkage. We evaluated FFQ performance using a measurement error model with 2 24HDR as the reference to estimate validity coefficients ($\rho_{Q,T}$) and attenuation factors (λ_Q). We applied parametric bootstrapping to construct 95% confidence intervals, accounting for the repeated measures structure in the data.
Results: For energy-adjusted, kilocalories from UPF (kcal/d), $\rho_{Q,T}$ were 0.50 (0.39–0.61) and 0.44 (0.30–0.59), for males and females, respectively, for the expert consensus method (ECM) and 0.52 (0.40–0.63) and 0.43 (0.29–0.57) for the food code method (FCM). For energy-adjusted, gram weight from UPF (g/d), $\rho_{Q,T}$ were 0.65 (0.60–0.71) and 0.66 (0.60–0.72) for the ECM and 0.66 (0.52–0.65) and 0.66 (0.59–0.72) for the FCM; λ_Q were $\geq$0.50 for both methods. UPF variables, defined using the ECM and FCM but expressed in the same unit, were highly correlated ($r \geq 0.97$).
Conclusions: FCM, performed similarly to ECM but has the potential to standardize UPF exposures across studies because the Nova categorization is assigned at a more granular level through database linkage. UPF intake based on energy-adjusted gram weight outperformed intake based on energy alone.

Keywords: dietary assessment, ultraprocessed food, Nova, measurement error model, epidemiology

Abbreviations: 24HDR, 24-hour dietary recall; AARP, American Association of Retired Persons; CI, confidence interval; CSFII, Continuing Survey of Food Intake of Individuals; ECM, Expert Consensus Method; FCM, food code method; FFQ, food frequency questionnaire; FNDDS, Food and Nutrient Database for Dietary Studies; HEI, Healthy Eating Index; HPFS, Health Professionals Follow-Up Study; MLE, Maximum Likelihood Estimation; MoM, Method of Moments; NHS I and II, Nurses' Health Studies I and II; RR_Q, observed relative risk; RR_T, true relative risk; SR, standard reference; UPF, ultraprocessed food; WWEIA, What We Eat in America.

* Corresponding author. *E-mail address:* erikka.loftfiled@nih.gov (E. Loftfield).

† EL and PZ are cofirst authors.

https://doi.org/10.1016/j.tjnut.2025.04.029
Received 15 January 2025; Received in revised form 9 April 2025; Accepted 29 April 2025; Available online 5 May 2025
0022-3166/Published by Elsevier Inc. on behalf of American Society for Nutrition.

Introduction

Food frequency questionnaires (FFQs) serve as a valuable tool for measuring usual or long-term diet in large epidemiologic studies; however, measurement error, both random and systematic, in the FFQ can be substantial though it differs, in terms of magnitude and direction, by the nutrient, food, food group, or dietary pattern being measured as well as by participant characteristics such as sex [1]. In prospective studies, measurement error is generally assumed to be nondifferential with respect to future disease risk, resulting in relative risk estimates that are biased toward the null and a loss of statistical power to detect diet-disease associations [2]. Evaluating FFQ performance for measuring a dietary exposure of interest using a reference instrument is critical to characterizing measurement error in a study population and correcting it. In addition to enhancing a study's internal validity, accounting for measurement error in an FFQ can increase reproducibility by facilitating comparisons of dietary exposure assessments across studies using different dietary questionnaires.

Ultraprocessed food (UPF), defined according to the Nova classification system as "industrial formulations manufactured from substances derived from foods or synthesized from other organic sources" [3], are abundant and widespread, accounting for a majority of calories consumed at the population level by children [4] and adults [5] in the United States. These ready-to-eat, ready-to-heat foods and beverages tend to be less expensive and more shelf-stable relative to minimally or unprocessed foods, and they often contain high amounts of added sugar, sodium, and saturated fat but little or no whole foods [3]. As global availability and consumption of UPF [3] have increased, so too have rates of obesity [6] and related chronic diseases [7–9]. Higher UPF consumption has been consistently associated with cardiometabolic disease risk and obesity in epidemiological studies [4,10], and obesity is an established risk factor for cardiovascular disease, type 2 diabetes, and more than a dozen cancer types [7–9]. Thus, there is growing concern about the potential adverse impact of UPF consumption on human health [10]. However, assessing intake of UPF in population-based studies is challenging. Dietary questionnaires, including FFQ, query a limited number of food items and typically lack sufficient detail to reliably determine the degree of food processing. A wider range of food and beverage items and more detailed information on food sources and preparation can be obtained from 24-hour dietary recalls (24HDRs). However, few prospective studies with follow-up for mortality as well as incident cancer and other chronic disease outcomes have collected multiple 24HDR to estimate usual or long-term UPF intake [11], which is likely the relevant exposure when considering associations with chronic disease and mortality risk.

With over 500,000 participants and >20 years of follow-up for cancer incidence and mortality, the NIH-AARP Diet and Health Study is among the largest studies in the world for studying potential health risks associated with UPF intake. Additionally, it includes the largest calibration dataset with 24HDR data in a United States cohort, providing a unique opportunity to apply the Nova classification to FFQ data and assess the performance. Prospective cohort studies, including NIH-AARP, have generally used a validated FFQ or other diet history instrument to assess dietary intake [1,12]. Although the Nova classification system has been applied to several FFQ and other diet history instruments to study the associations of UPF intake with chronic disease risk [10,13,14], validation studies of Nova-defined, FFQ-derived UPF intake are lacking. This raises concerns about exposure misclassification as well as study validity and reproducibility because the Nova classification relies on ingredient-level data and food source for accurate classification [15].

Consequently, it is critical to evaluate FFQ performance for evaluating UPF specifically, and the Nova classification more broadly, and to, whenever possible, account for measurement error in observed diet-disease association studies. Herein, our objectives were 2-fold as follows: *1)* Estimate FFQ-derived UPF intake by expert consensus, which is the approach often used in FFQ-based cohort studies [16], and then by disaggregation of FFQ items into standard recipes, Food and Nutrient Database for Dietary Studies (FNDDS) codes and standard reference (SR) codes followed by linkage to a Nova database developed in the United States NHANES [15]. *2)* Validate FFQ-derived UPF intake using a calibration sub-study of >1300 participants who completed 2 nonconsecutive 24HDRs and 2 FFQ in the NIH-AARP Diet and Health Study. A secondary objective was to estimate and validate FFQ intake from Nova groups 1 (unprocessed or minimally processed foods), 2 (processed culinary ingredients), and 3 (processed foods).

Methods

Design

The NIH-AARP Diet and Health Study design, including the calibration sub-study, has been described in detail elsewhere [1,17,18]. In brief, a baseline questionnaire, including a 124-item FFQ, was mailed to ~ 3.5 million AARP members in 3 waves. In October 1995, the first wave of 250,000 questionnaires was mailed, and the calibration sub-study participants were selected from a pool of ~38,000 eligible participants who responded by January 1996. Of the 2795 individuals who were invited to participate in the calibration sub-study, 2055 agreed to participate, and 1988 completed 2 nonconsecutive, unannounced 24HDRs by telephone on a random day of the week [1]. The study was approved by the Special Studies Institutional Review Board of the United States National Cancer Institute. Our analytic sample included 1311 individuals (n = 661 males and n = 650 females) who met the original study eligibility criteria [1] as well as those who completed a second valid 24HDR and a second valid 124-item FFQ in October 1996 (Supplemental Figure 1).

Nova assessment

Demographic and lifestyle variables were queried about on the self-administered baseline questionnaire, which also assessed dietary intake via an FFQ that included 124 food and beverage items as well as portion size for 116 items and 21 questions on intake of low-fat, high-fiber foods and food preparation, resulting in 204-line items. Briefly, nationally representative survey data from the USDA's 1994–1996 Continuing Survey of Food Intake of Individuals (CSFII), which sampled noninstitutionalized individuals who resided in all 50 states with oversampling of the low-income populations, was used to develop a nutrient composition database [18,19] for the FFQ, and estimates of

individual intake were calculated in gram weight and energy. Prior research has shown that using nationally representative United States dietary data in a systematic way to create an FFQ nutrient database yields less bias in nutrient estimates in FFQ-based studies of United States adults [18]. To assign gram weight and calorie intake to a Nova classification (i.e., group 1: unprocessed or minimally processed foods; group 2: processed culinary ingredients; group 3: processed foods; group 4: ultra-processed foods), defined according to Martinez-Steele et al. [20], we used 2 approaches.

First, we used an expert consensus method (ECM), similar to what has been employed in other FFQ-based cohort studies, to assign Nova categorization [16]. Four experts (EL, LA, NK, EMS) independently matched NIH-AARP food and beverage descriptions for 204-FFQ line items to Nurses' Health Studies (NHS) I and II and the Health Professionals Follow-Up Study (HPFS) food and beverage descriptions [16]. In the published food list by Khandpur et al. [16], combined dishes that were presented jointly in the FFQ were not disaggregated into their components. However, when NIH-AARP FFQ line items matched multiple descriptions (e.g., "Pies, fruit" in NIH-AARP matched to "Pie, home-baked," AND "Pie, home-baked or ready-made," AND "Pie, ready-made" in NHS I/II and HPFS); calorie and gram intake were split equally across the matching descriptions (e.g., 1/3 for each "Pie" description; Supplemental Table 1). Overall, 4-way agreement for assignment of food and beverage descriptions was 63%, and 3-way agreement was 72%. Multiple potential matches (e.g., assigning "Pie, home-baked" only compared with all 3 matching descriptors) were the primary cause of discordant assignments. Discordant items were discussed, and descriptions were assigned based on group consensus. The original NIH-AARP assigned description was retained for items where no proxy description existed. Next, using the published food list [16], food and beverage descriptions were linked to *1)* a primary Nova categorization and *2)* a sensitivity analysis Nova categorization, whereby food items with insufficient information to support a single classification were assigned to a non–ultraprocessed Nova group (i.e., Nova group 1 or Nova group 3) as their primary categorization, as outlined by Khandpur et al [16]. Line items without a proxy description were discussed and assigned a primary and sensitivity Nova categorization based on expert consensus.

Next, we used a food code method (FCM) to disaggregate FFQ items and assign Nova categorization at a more granular level. To accomplish this, we translated FFQ line items into individual food codes using the 1994–1996 USDA CSFII database, which serves as the basis for the NIH-AARP FFQ and nutrient database [18]. These individual food codes were then translated into 3513 USDA FNDDS codes, which were disaggregated further to component FNDDS and SR codes using the approach described by Steele et al. [15] for classifying foods according to Nova. For example, a single NIH-AARP FFQ item "fruit pies" corresponded to 50 FNDDS food codes, which could each be broken down further into SR codes (e.g., "pie, cherry, one crust" decomposed to 9 SR codes, including 3 in Nova group 1, 3 in Nova group 2, and 3 in Nova group 4); when an FFQ item comprised 2 or more food codes, the proportion of gram and energy intake was estimated based on the sample weight of the person who reported that food code and how frequently that food code was reported in CSFII [18]. Similar to the FFQ FCM, dietary intake data from 24HDR was translated into FNDDS food codes, which were disaggregated using the approach described by Steele et al. [15]. Nova classification for both the FFQ FCM and 24HDR relied on linkage of disaggregated FNDDS and SR codes to the What We Eat in America (WWEIA) and the NHANES Nova database created by Steele et al. Following the Nova classification, gram weights and energy values were recalculated so that the sum of all Nova groups and subgroups equaled the original estimates in NIH-AARP for each participant for each dietary assessment tool and administration.

Statistical analysis

To evaluate the comparability of the baseline cohort and calibration sub-sample, we generated descriptive statistics for key demographic, lifestyle, and health-related factors. To evaluate differences in kilocalorie and gram weight intake estimation from UPF and other Nova groups, we estimated median daily intake in kilocalories (kcal), percentage energy (% energy), gram weight (g), and percentage gram weight (% g) for Nova groups 1 through 4 for the FFQ and the 24HDR at each administration and for each Nova classification method (i.e., ECM and FCM). We then calculated the percentage difference for the FFQ-derived measure and the first 24HDR value following square root transformation and noted differences that when differences were $\geq$15% [1]. To explore correlations between Nova variables, derived from the first FFQ administration using the ECM and FCM, we calculated Pearson correlations, which were visualized as a correlation matrix.

To evaluate the performance of Nova variables, we used a measurement error model to estimate the following: *1)* the validity coefficient (ρ), which is the Pearson correlation coefficient between true and reported intake, and *2)* the attenuation factor (λ), defined as the slope of the regression of true intake on reported intake [1]. An attenuation factor approaching zero implies significant bias in relative risk estimation (i.e., severe attenuation toward the null), whereas a value approaching 1 suggests minimal bias. The estimation of both parameters requires the administration of a reference instrument, or 24HDR, at least twice per person to ensure that within-person variation can be accounted for, and we assume that 24HDR intake is unbiased at the individual level and that within-person random error is independent of true intake and error in the FFQ. We define N, E, and R to represent the Nova intake variable in either kilocalories or grams, total energy intake, and the residual model, respectively. The residual model R refers to the adjustment of N for E by regression analysis [21]. We followed the framework of Thompson et al. [1] in using the 24HDR as the reference instrument. The formulas for ρ_N, λ_N, ρ_E, λ_E, ρ_R, and λ_R are derived in Appendix B of Thompson et al. [1]. Although Thompson et al. [1] employed the Maximum Likelihood Estimation (MLE) method [1], we estimated these parameters using the Method of Moments (MoM) [22] for 1311 participants who had fully completed both the 2 FFQs and the 2 24HDRs by sex. MoM is straightforward and does not require iterative optimization, which can be advantageous when estimating a large number of parameters. In our study, where the number of parameters diverges, the MLE was unstable and more sensitive to violations of distributional assumptions. We present our model and assumptions in the Appendix A Supplementary data.

Before parameter estimation, we applied a square root transformation to each variable to address the significant number of

zeros in some nutrient variables and to manage the extensive range observed. We also attempted a logarithmic transformation, after adding a small constant number; however, these variables still exhibited a bimodal distribution after logarithmic transformation. This deviates from the normality assumption required for our estimation process. We also considered Box-Cox transformation, but this would result in different transformations for each *N* and for *E*, making interpretation difficult.

We estimated validity coefficients and attenuation factors stratified by sex. To construct the 95% CIs for ρ_N, ρ_E, and ρ_R, as well as λ_N, λ_E, and λ_R, we employed a parametric bootstrapping technique with 1000 replications. Given that observations between different time points are correlated, which originates from the same subjects, we specifically used this approach to adequately account for the repeated measures structure in the data. The R codes for executing the analysis are available at https://github.com/BettyzhangPei.

There is no standard of interpretation for validity coefficients and attenuation factors. However, in nutritional epidemiology, correlation coefficients between 2 dietary assessment methods (e.g., self-reported dietary intake and dietary biomarker) >0.5 have been considered strong [23]. Similarly, given an attenuation factor of 0.5 for a dietary exposure and a true relative risk (RR_T) of 1.5, we would expect an observed relative risk (RR_Q) of 1.2 (i.e., if $RR_T = 1.5$ & $\lambda_Q = 0.5$ then $RR_Q = 1.5^{0.5} = 1.2$) [1], suggesting that it is feasible to detect a modest association between a dietary exposure with an attenuation factor of ≥0.5 and disease risk in large cohort studies.

Results

The calibration sub-sample reflects the baseline cohort with regard to baseline demographic factors, including age, race/ethnicity, and educational attainment, as well as modifiable risk factors, including BMI, smoking status, physical activity, and diet quality assessed via the Healthy Eating Index 2015 (HEI-2015) [24] (Table 1). As noted previously by Thompson et al. [1], similar distributions in baseline characteristics indicate that the calibration sub-sample was a random sample of the cohort.

TABLE 1
Baseline characteristics in the NIH-American Association of Retired Persons (AARP) Diet and Health Study cohort and calibration sub-study analytic sample by sex.

Baseline characteristic	Males		Females	
	Baseline cohort (n = 319,026)	Calibration sub-sample (n = 661)	Baseline cohort (n = 221,074)	Calibration sub-sample (n = 650)
Age (y), mean (SD)	62.3 (5.3)	62.5 (5.2)	62.0 (5.4)	62.3 (5.4)
BMI (kg/m^2), mean (SD)	27.2 (4.3)	27.0 (4.4)	26.8 (6.0)	26.8 (5.8)
Total HEI-2015 score, mean (SD)	66.9 (9.7)	66.5 (10.2)	68.8 (9.4)	68.3 (9.7)
Race/ethnicity (n, %)				
Non-Hispanic White	295,145 (92.5%)	618 (93.5%)	197,833 (89.5%)	609 (93.7%)
Non-Hispanic Black	8747 (2.7%)	23 (3.5%)	12,248 (5.5%)	27 (4.2%)
Hispanic	5949 (1.9%)	10 (1.5%)	4057 (1.8%)	8 (1.2%)
Asian	4125 (1.3%)	5 (0.8%)	2421 (1.1%)	1 (0.2%)
Pacific Islander	360 (0.1%)	0 (0%)	280 (0.1%)	0 (0%)
American Indian/Alaskan Native	785 (0.2%)	1 (0.2%)	718 (0.3%)	0 (0%)
Unknown	3915 (1.2%)	4 (0.6%)	3517 (1.6%)	5 (0.8%)
Education (n, %)				
<12 y	18,371 (5.8%)	29 (4.4%)	13,398 (6.1%)	19 (2.9%)
12 y/high school graduate	49,514 (15.5%)	89 (13.5%)	55,726 (25.2%)	169 (26.0%)
Vocational/technical training	29,396 (9.2%)	83 (12.6%)	23,333 (10.6%)	72 (11.1%)
Some college	70,433 (22.1%)	137 (20.7%)	54,782 (24.8%)	169 (26.0%)
College graduate	142,657 (44.7%)	312 (47.2%)	66,363 (30.0%)	203 (31.2%)
Unknown	8655 (2.7%)	11 (1.7%)	7472 (3.4%)	18 (2.8%)
Smoking status (n, %)				
Never smoker	92,625 (29.0%)	198 (30.0%)	96,577 (43.7%)	264 (40.6%)
Former smoker	180,981 (56.7%)	373 (56.4%)	85,355 (38.6%)	261 (40.2%)
Current smoker	32,779 (10.3%)	71 (10.7%)	31,194 (14.1%)	105 (16.2%)
Unknown	12,641 (4.0%)	19 (2.9%)	7948 (3.6%)	20 (3.1%)
Physical activity (n, %)				
Never/rarely	46,455 (14.6%)	91 (13.8%)	48,833 (22.1%)	139 (21.4%)
Low (few times a month)	41,566 (13.0%)	78 (11.8%)	31,518 (14.3%)	94 (14.5%)
Moderate (few times per week)	69,935 (21.9%)	144 (21.8%)	46,454 (21.0%)	138 (21.2%)
High (many times a week)	157,963 (49.5%)	345 (52.2%)	91,270 (41.3%)	274 (42.2%)
Unknown	3107 (1.0%)	3 (0.5%)	2999 (1.4%)	5 (0.8%)
Self-rated health status (n, %)				
Excellent	54,197 (17.0%)	125 (18.9%)	36,019 (16.3%)	127 (19.5%)
Very good	112,959 (35.4%)	235 (35.6%)	76,172 (34.5%)	237 (36.5%)
Good	111,908 (35.1%)	237 (35.9%)	78,054 (35.3%)	209 (32.2%)
Fair	35,468 (11.1%)	54 (8.2%)	26,719 (12.1%)	61 (9.4%)
Poor[1]	—	0 (0%)	—	0 (0%)
Unknown	4494 (1.4%)	10 (1.5%)	4110 (1.9%)	16 (2.5%)

Abbreviations: HEI, Healthy Eating Index.
[1] Participants with poor self-rated health were excluded from the baseline cohort.

Table 2 shows estimated median daily intake in kilocalories (kcal), percentage energy (% energy), gram weight (g), and percentage gram weight (% g) for Nova groups 1 through 4 for the FFQ and the 24HDR at each administration and for each Nova classification method (i.e., ECM and FCM). For absolute energy (kcal/d) and gram (g/d) intake, median reported FFQ intake was consistently lower on the second administration for males and females and for intake of unprocessed/minimally processed food (Nova group 1) and UPF (Nova group 4) using both the ECM and the FCM. In contrast, relative measures for energy and grams were more similar between the first and second FFQ administrations for male and females and across all Nova groups, particularly for the FCM. Based on population median intakes, FFQ tended to overestimate gram weight intake (g/d and % g/d) for group 1 and underestimate gram weight intake for group 4 as compared with 24HDR. For energy intake, the ECM and the FCM generally overestimated group 1 intake for both male (except for kcal/d using FCM) and females, but the ECM resulted in greater overestimation. Group 4 energy intake was consistently underestimated by FFQ using the ECM and FCM for both males and females, though the ECM resulted in greater underestimation. Comparisons between FFQ and 24HDR revealed a few substantial differences. For example, when we compared the square root transformed median of the first FFQ and 24HDR for each absolute and relative energy and gram weight intake by the Nova group, absolute energy intake from UPF derived using the ECM, but not the FCM, exceeded ≥15% in both males and females.

For each Nova group, variables derived using the different approaches to Nova classification, with the same unit of measure, were highly correlated. For example, Pearson correlation coefficients were >0.99 for daily absolute gram weight from UPF (group 4) derived using the ECM, the ECM sensitivity classification, and the FCM (Figure 1 and Supplemental Table 2). Additionally, these variables were moderately positively correlated with those for daily absolute energy from UPF (all r ≥ 0.52), percentage energy from UPF (all r ≥ 0.32), and percentage gram weight from UPF (all r ≥ 0.75). Absolute gram weight intake from group 1 food was not correlated with absolute gram weight intake from UPF (all r ≤ 0.08) but was positively correlated with absolute energy intake from UPF (all r ≥ 0.17) and negatively correlated with percentage energy from UPF (all r ≤ −0.25) and percentage gram weight from UPF (all r ≤ −0.43; Figure 1 and Supplemental Table 2).

Using 2 24HDR as the reference instrument, we estimated validity coefficients (ρ) and attenuation factors (λ) with 95% CI for true and FFQ-reported Nova intakes classified using 2 different methods (Table 3). The validity coefficients for absolute daily energy and gram weight intakes in our analytic sample were 0.39 (0.31, 0.46) and 0.55 (0.49, 0.60) for males and 0.24 (0.15, 0.32) and 0.45 (0.39, 0.52) for females, respectively. For males, validity coefficients for daily energy intake from UPF using the FCM, without and with adjustment for total energy intake, were 0.43 (0.36, 0.51) and 0.52 (0.40. 0.63), respectively. For females, validity coefficients for daily energy from UPF using the FCM, without and with adjustment for total energy intake, were slightly lower than for males at 0.33 (0.24, 0.41) and 0.43 (0.29. 0.57), respectively. Validity coefficients for daily energy intake from UPF using the ECM main or sensitivity Nova classifications were similar. Validity coefficients for daily gram weight intake from UPF using the FCM were higher than those for energy intake. Using the FCM, validity coefficients for daily gram weight intake from UPF were 0.59 (0.53, 0.65) and 0.66 (0.42. 0.65), without and with adjustment for total energy, for males and 0.59 (0.53, 0.66) and 0.66 (0.59. 0.72) for females. Again, validity coefficients for daily gram weight intake from UPF using the ECM main or sensitivity Nova classifications were similar.

Discussion

Validation of an error-prone dietary assessment (or closely related) instrument, using a reference instrument, is critical to

TABLE 2
Median daily energy and gram weight intakes for Nova groups measured by FFQ, using either a food code or expert consensus method for Nova classification, and 24HDR, using a FCM, in the calibration sub-sample.

Nova group (unit)	Males (n = 661)						Females (n = 650)					
	FFQ ECM (main)		FFQ FCM		24HDR		FFQ ECM (main)		FFQ FCM		24HDR	
	First	Second	First	Second	First	Second	First	Second	First	Second	First	Second
Group 1, (kcal/d)	887	850	783	758	801	806	709	679	646	617	579	584
Group 1, (% energy/d)	46.3	47.3	41.6	41.8	36.6	37.7	48.8[1]	49.0	43.6	43.7	36.3	37.8
Group 1, (g/d)	1960	1900	1930	1870	1590	1620	1840[1]	1720	1810[1]	1690	1340	1320
Group 1, (% g/d)	71.5	70.8	70.2	70.1	63.2	62.6	76.0	77.0	74.7	75.9	65.8	65.6
Group 2, (kcal/d)	76.1[1]	73.7	68.8[1]	66.8	46.7	51.7	70.2[1]	69.8	51.6	46.7	39.2	32.3
Group 2, (% energy/d)	4.00[1]	4.15	3.50[1]	3.49	2.16	2.60	4.29[1]	4.71	3.32[1]	3.19	2.35	2.08
Group 2, (g/d)	18.0[1]	18.2	13.4	13.2	13.4	13.4	17.0[1]	16.7	10.4	10.2	10.4	9.31
Group 2, (% g/d)	0.63	0.68	0.46	0.47	0.52	0.56	0.73[1]	0.72	0.43	0.44	0.50	0.44
Group 3, (kcal/d)	134	136	137	137	157	149	91.9	87.7	93.0	89.3	101	87.3
Group 3, (% energy/d)	6.96	7.09	7.07	7.19	6.81	7.22	6.02	6.15	6.06	6.19	6.40	5.68
Group 3, (g/d)	102	107	125[1]	126	88.0	95.5	70.7	68.3	88.0[1]	85.1	58.2	49.8
Group 3, (% g/d)	3.63	4.02	4.58	4.62	3.53	3.67	2.92	2.93	3.59	3.70	2.86	2.54
Group 4, (kcal/d)	713[1]	697	825	794	1030	1020	548[1]	527	640	625	768	773
Group 4, (% energy/d)	38.5	38.2	44.1	43.7	48.9	46.9	37.1	36.9	43.5	43.7	48.8	48.9
Group 4, (g/d)	592	580	599	598	693	660	432	392	460	415	520	541
Group 4, (% g/d)	21.2	20.7	21.8	21.2	27.9	27.4	18.2[1]	17.2	18.9	18.3	25.7	26.2

Abbreviations: 24HDR, 24-hour dietary recall; EMC, expert consensus method, main classification; FCM, food code method; FFQ, food-frequency questionnaire; Group 1, unprocessed/minimally processed; Group 2, processed culinary ingredient; Group 3, processed; Group 4, ultraprocessed.
[1] The first FFQ value is ≥15% different from the first 24HR value following square root transformation.

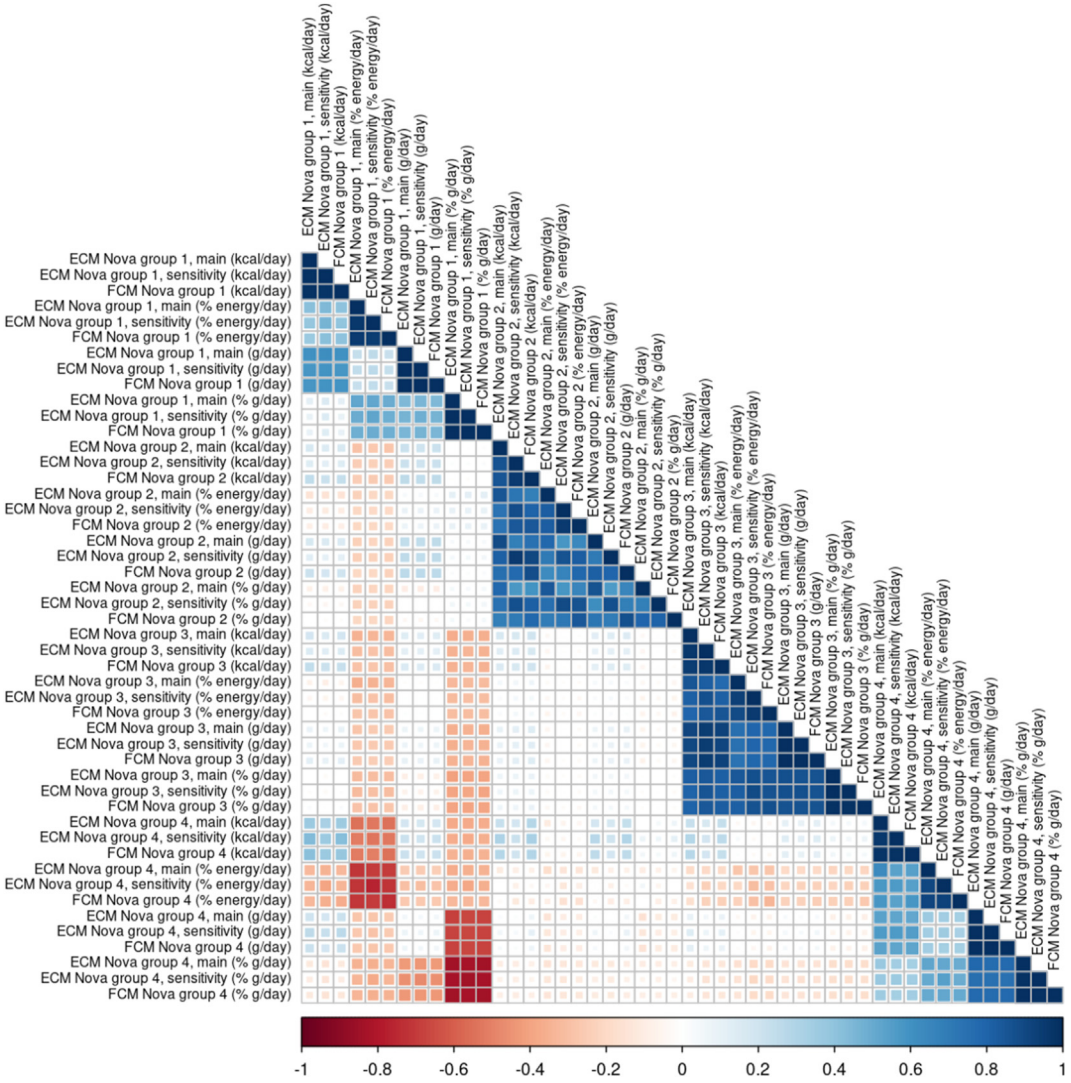

FIGURE 1. Pearson correlation matrix of dietary intake by degree of food processing, defined by Nova group, unit of measure, defined as absolute or relative daily energy or gram weight, and FFQ classification method, defined as the ECM (main or sensitivity) or FCM, using the first administered FFQ in the NIH-American Association of Retired Persons (AARP) Diet and Health Study calibration sub-sample (n = 1311); ECM, expert consensus method; FCM, food code method; FFQ, food frequency questionnaire.

quantifying the degree of measurement error in dietary data, but it is most informative when the performance of the dietary exposure of interest (e.g., UPF intake) is quantified. Additionally, within a study, validation data for an exposure of interest, can be used to calibrate relative risk estimates from Cox regression models [12,25,26]. With over 500,000 participants and >20 years of follow-up for cancer incidence and mortality, the NIH-AARP Diet and Health Study is among the largest studies in the world for studying potential health risks associated with UPF intake. Therefore, it is critical to evaluate how well UPF intake is measured by FFQ in the full cohort and to account for measurement error in the FFQ in future analyses of UPF exposures with mortality and cancer risk. Moreover, our study has the potential to improve study rigor and reproducibility of UPF research more broadly by serving as a benchmark for comparative validation of other FFQ that have been used to assess dietary intake in large prospective cohort studies.

In our study, the validity coefficients, which if close to 0 indicate severe misclassification and if close to 1 indicate little misclassification, for absolute energy and gram weight intake from UPF, defined using the FCM, were 0.43 and 0.59 for males and 0.33 and 0.59 for females, respectively, before adjustment for total energy intake. Adjustment for total energy intake generally improved validity coefficients, particularly for energy intake from UPF. Validity coefficients tended to be higher for males than females, which is in line results for other nutrient components in the NIH-AARP Study [1]. The attenuation factors, which if close to 0 indicate severe attenuation of the effect estimate and if close to 1 indicate little attenuation, followed a similar pattern as the validity coefficients. For the FCM, for absolute energy and gram weight intake from UPF, attenuation factors were 0.34 to 0.47 for males and 0.25 to 0.45 for females, respectively, before adjustment for total energy intake. Adjustment for total energy generally moved estimates toward 1, including for other Nova groups, and attenuation factors tended to be higher in males than females. Validity coefficients and attenuation factors were similar for the FCM and ECM (both main and sensitivity variables), and UPF intake defined using the

TABLE 3
Estimated validity coefficients (ρ) and attenuation factors (λ) between FFQ-reported Nova intakes, derived using 2 different methods for classification, and true Nova intakes, estimated using a measurement error model and 24HDR as the reference instrument, for absolute daily energy and gram weight intakes, unadjusted and adjusted for total energy intake, in the NIH-American Association of Retired Persons (AARP) Diet and Health Study calibration sub-sample ($n = 1311$).

Dietary constituent	ρ (95% CI)				λ (95% CI)			
	Males ($n = 661$)		Female ($n = 650$)		Males ($n = 661$)		Females ($n = 650$)	
Total daily energy, kcal	0.39	(0.31, 0.46)	0.24	(0.15, 0.32)	0.26	(0.21, 0.32)	0.16	(0.10, 0.22)
Total daily gram weight, g	0.55	(0.49, 0.60)	0.45	(0.39, 0.52)	0.38	(0.33, 0.42)	0.27	(0.23, 0.32)
Food code method								
Group 1, kcal, unadjusted	0.40	(0.32, 0.49)	0.37	(0.27, 0.46)	0.29	(0.23, 0.36)	0.24	(0.18, 0.31)
Group 1, kcal, adjusted	0.46	(0.35, 0.57)	0.46	(0.34, 0.58)	0.42	(0.32, 0.52)	0.39	(0.29, 0.49)
Group 2, kcal, unadjusted	0.52	(0.41, 0.64)	0.61	(0.49, 0.73)	0.42	(0.33, 0.51)	0.45	(0.37, 0.53)
Group 2, kcal, adjusted	0.52	(0.39, 0.65)	0.59	(0.46, 0.72)	0.43	(0.33, 0.52)	0.44	(0.35, 0.53)
Group 3, kcal, unadjusted	0.59	(0.51, 0.67)	0.46	(0.36, 0.55)	0.55	(0.47, 0.63)	0.46	(0.36, 0.56)
Group 3, kcal, adjusted	0.60	(0.51, 0.69)	0.57	(0.45, 0.68)	0.63	(0.53, 0.73)	0.55	(0.44, 0.66)
Group 4, kcal, unadjusted	0.43	(0.36, 0.51)	0.33	(0.24, 0.41)	0.34	(0.27, 0.40)	0.25	(0.18, 0.31)
Group 4, kcal, adjusted	0.52	(0.40, 0.63)	0.43	(0.29, 0.57)	0.50	(0.39, 0.61)	0.35	(0.23, 0.47)
Group 1, g, unadjusted	0.62	(0.57, 0.67)	0.58	(0.53, 0.64)	0.49	(0.44, 0.55)	0.39	(0.35, 0.44)
Group 1, g, adjusted	0.69	(0.60, 0.77)	0.67	(0.57, 0.77)	0.54	(0.47, 0.62)	0.50	(0.43, 0.58)
Group 2, g, unadjusted	0.56	(0.45, 0.67)	0.60	(0.50, 0.70)	0.43	(0.35, 0.51)	0.46	(0.38, 0.53)
Group 2, g, adjusted	0.58	(0.47, 0.69)	0.62	(0.51, 0.72)	0.44	(0.36, 0.53)	0.48	(0.40, 0.56)
Group 3, g, unadjusted	0.74	(0.68, 0.79)	0.64	(0.56, 0.73)	0.57	(0.51, 0.62)	0.51	(0.44, 0.58)
Group 3, g, adjusted	0.74	(0.68, 0.80)	0.70	(0.62, 0.78)	0.64	(0.57, 0.70)	0.55	(0.48, 0.62)
Group 4, g, unadjusted	0.59	(0.53, 0.65)	0.59	(0.53, 0.66)	0.47	(0.41, 0.52)	0.45	(0.39, 0.50)
Group 4, g, adjusted	0.66	(0.52, 0.65)	0.66	(0.59, 0.72)	0.57	(0.50, 0.63)	0.51	(0.45, 0.58)
Expert consensus method (main classification)								
Group 1, kcal, unadjusted	0.39	(0.30, 0.48)	0.36	(0.26, 0.45)	0.27	(0.20, 0.33)	0.22	(0.16, 0.28)
Group 1, kcal, adjusted	0.44	(0.33, 0.55)	0.44	(0.33, 0.56)	0.39	(0.29, 0.49)	0.35	(0.26, 0.45)
Group 2, kcal, unadjusted	0.53	(0.42, 0.65)	0.56	(0.44, 0.68)	0.38	(0.30, 0.46)	0.37	(0.30, 0.45)
Group 2, kcal, adjusted	0.53	(0.40, 0.65)	0.53	(0.40, 0.65)	0.36	(0.28, 0.44)	0.34	(0.26, 0.42)
Group 3, kcal, unadjusted	0.56	(0.48, 0.64)	0.46	(0.37, 0.56)	0.49	(0.42, 0.57)	0.42	(0.33, 0.51)
Group 3, kcal, adjusted	0.55	(0.46, 0.64)	0.53	(0.42, 0.65)	0.53	(0.44, 0.62)	0.46	(0.36, 0.56)
Group 4, kcal, unadjusted	0.43	(0.35, 0.51)	0.32	(0.24, 0.4)	0.33	(0.27, 0.40)	0.24	(0.18, 0.31)
Group 4, kcal, adjusted	0.50	(0.39, 0.61)	0.44	(0.30, 0.59)	0.45	(0.34, 0.55)	0.32	(0.22, 0.42)
Group 1, g, unadjusted	0.62	(0.57, 0.67)	0.58	(0.53, 0.64)	0.49	(0.44, 0.54)	0.39	(0.35, 0.44)
Group 1, g, adjusted	0.68	(0.59, 0.77)	0.67	(0.56, 0.77)	0.54	(0.46, 0.61)	0.50	(0.42, 0.58)
Group 2, g, unadjusted	0.49	(0.38, 0.60)	0.50	(0.40, 0.60)	0.32	(0.25, 0.39)	0.35	(0.28, 0.42)
Group 2, g, adjusted	0.49	(0.37, 0.60)	0.50	(0.40, 0.61)	0.32	(0.25, 0.39)	0.36	(0.28, 0.43)
Group 3, g, unadjusted	0.75	(0.69, 0.81)	0.68	(0.60, 0.76)	0.56	(0.50, 0.61)	0.51	(0.45, 0.57)
Group 3, g, adjusted	0.75	(0.69, 0.81)	0.72	(0.64, 0.80)	0.61	(0.55, 0.67)	0.53	(0.47, 0.60)
Group 4, g, unadjusted	0.59	(0.53, 0.65)	0.59	(0.53, 0.65)	0.46	(0.40, 0.52)	0.44	(0.39, 0.49)
Group 4, g, adjusted	0.65	(0.60, 0.71)	0.66	(0.60, 0.72)	0.55	(0.49, 0.61)	0.50	(0.44, 0.56)
Expert consensus method (sensitivity classification)								
Group 1, kcal, unadjusted	0.42	(0.34, 0.51)	0.39	(0.29, 0.49)	0.30	(0.24, 0.37)	0.25	(0.18, 0.31)
Group 1, kcal, adjusted	0.48	(0.37, 0.59)	0.49	(0.38, 0.61)	0.41	(0.32, 0.50)	0.38	(0.29, 0.47)
Group 2, kcal, unadjusted	0.54	(0.42, 0.65)	0.63	(0.51, 0.74)	0.38	(0.30, 0.45)	0.41	(0.34, 0.49)
Group 2, kcal, adjusted	0.53	(0.41, 0.66)	0.60	(0.47, 0.73)	0.35	(0.27, 0.43)	0.37	(0.29, 0.45)
Group 3, kcal, unadjusted	0.60	(0.52, 0.67)	0.52	(0.43, 0.62)	0.52	(0.45, 0.60)	0.49	(0.40, 0.59)
Group 3, kcal, adjusted	0.59	(0.50, 0.68)	0.62	(0.51, 0.73)	0.54	(0.45, 0.63)	0.52	(0.43, 0.62)
Group 4, kcal, unadjusted	0.43	(0.35, 0.50)	0.34	(0.26, 0.42)	0.32	(0.25, 0.38)	0.25	(0.19, 0.31)
Group 4, kcal, adjusted	0.52	(0.41, 0.63)	0.47	(0.33, 0.61)	0.50	(0.39, 0.62)	0.36	(0.25, 0.47)
Group 1, g, unadjusted	0.62	(0.57, 0.67)	0.59	(0.53, 0.64)	0.49	(0.44, 0.54)	0.39	(0.34, 0.44)
Group 1, g, adjusted	0.58	(0.51, 0.66)	0.57	(0.48, 0.65)	0.36	(0.30, 0.42)	0.30	(0.25, 0.36)
Group 2, g, unadjusted	0.54	(0.43, 0.65)	0.52	(0.42, 0.63)	0.40	(0.32, 0.48)	0.41	(0.33, 0.49)
Group 2, g, adjusted	0.53	(0.42, 0.64)	0.52	(0.41, 0.63)	0.39	(0.31, 0.47)	0.41	(0.33, 0.49)
Group 3, g, unadjusted	0.76	(0.71, 0.81)	0.70	(0.62, 0.78)	0.55	(0.50, 0.60)	0.51	(0.45, 0.57)
Group 3, g, adjusted	0.74	(0.68, 0.79)	0.72	(0.64, 0.80)	0.57	(0.52, 0.63)	0.51	(0.45, 0.57)
Group 4, g, unadjusted	0.58	(0.52, 0.64)	0.59	(0.52, 0.65)	0.46	(0.40, 0.52)	0.44	(0.39, 0.50)
Group 4, g, adjusted	0.62	(0.56, 0.68)	0.63	(0.57, 0.69)	0.53	(0.47, 0.60)	0.51	(0.45, 0.57)

Abbreviations: 24HDR, 24-hour dietary recall; EMC, expert consensus method, main classification; FCM, food code method; FFQ, food-frequency questionnaire; Group 1, unprocessed/minimally processed; Group 2, processed culinary ingredient; Group 3, processed; Group 4, ultraprocessed.

2 methods was highly correlated when the unit of measure (i.e., gram weight) was the same.

Overall, the FFQ performed similarly or better for estimating dietary intake according to Nova as it did for estimating intake of 26 dietary components in the original validation study by Thompson et al. [1] in which validity coefficients and attenuation factors ranged from 0.22 to 0.67 and 0.16 to 0.43 before adjustment for total energy intake and improved after adjustment for energy intake. Finally, after adjustment for total energy intake, FFQ performance for estimating UPF intake in gram

weight, defined using FCM, was strong. Validity coefficients were 0.66 (95% CI: 0.52, 0.75) for males and 0.66 (0.59, 0.72) for females, and attenuation factors were all >0.50. Again, the FCM and ECM performance metrics were similar. Comparable validity coefficients and attenuation factors for the FCM and ECM suggest that FFQ-item assignment based on the food descriptions and Nova assignments published by Khandpur et al. result in a similar degree of exposure misclassification and would yield similar relative risk estimates in prospective analyses. However, it is worth noting the assignment of multiple food descriptions to a single NIH-AARP FFQ item, which mirrors how multiple food codes could be assigned to a single FFQ item using on standard recipes, may have impacted the degree of similarity between the 2 methods. In addition, population median UPF intake estimates based on the FCM, as compared with the ECM, were generally closer to those estimated using 24HDR data (either first or second administration), which like the FCM relies on food and SR codes to disaggregate mixed dishes.

As previously reported, the NIH-AARP FFQ is an early version the NCI Dietary History Questionnaire (DHQ) [1]. The NIH-AARP FFQ was evaluated on the basis of 24 nutrients and servings of fruits and vegetables using a measurement error model and data from 2 24HDR collected in the same calibration sub-study used herein [1]. These studies have shown that the NIH-AARP FFQ and the DHQ tend to perform similarly to other instruments, such as the Block and Willett FFQ, after energy adjustment [1,27], suggesting that results from our study may generalize to other studies using these FFQ and a similar approach to Nova classification. Still additional studies using multiple 24HDR or multi-day food records to evaluate FFQ performance for estimating energy and gram weight intake from UPF are warranted. Moreover, cohort studies, whenever possible, should use validation data to perform regression calibration and account for measurement error in FFQ-derived UPF intake as this has the potential to improve study comparability by accounting for systematic errors due to differences in the FFQ used across cohort studies [12] and to strengthen risk estimates where an association exists but is attenuated toward the null due to nondifferential exposure misclassification [26].

Our study is among the first to evaluate the quality of Nova-based dietary intake estimates from the FFQ by modeling their relationship to the unobserved true intake using 24HDR data as a reference instrument. A key strength of our work is that we estimated 2 widely accepted measures of instrument validity, the validity coefficient (ρ) and the attenuation factor (λ), and we also provided confidence intervals for these estimates, which involved addressing several statistical challenges. However, our approach is based on the measurement error model proposed by Thompson et al. [1], and therefore it shares some of the same limitations. Specifically, the model assumes that that 24HDR intake is unbiased at the individual level and that within-person random error is independent of true intake and error in the FFQ. However, studies using unbiased biomarkers of energy and protein intakes have found that these assumptions may not hold [27–30]. Because both the FFQ and 24HDR values may deviate from true intake due to within-person variability and reporting error (i.e., classical-type error), using 24HDR data as the reference instrument, may introduce bias in our estimates of FFQ performance [1]. Finally, Nova group classification of a given food item or component may vary by the time and place of data collection as well as by expert opinion or choice of database linkage. Therefore, it is critical that studies provide sufficient detail on how UPF intake was operationalized. Future research should also consider Nova subgroups, including sugar- and artificially sweetened beverages that might impact energy compared with gram weight estimation of UPF intake differently.

Overall, we found that 2 different methods for applying Nova to FFQ data, one based on expert consensus (i.e., ECM) and one based on standard recipes translated to food codes and then disaggregated to FNDDS and SR codes (i.e., FCM), performed similarly in terms of validity coefficients and attenuation factors. However, the ECM resulted in greater underestimation median absolute energy and gram intake from UPF compared with the FCM, owing in part to the conservative approach of assigning items with insufficient information for Nova classification to a non–ultraprocessed Nova group as their primary or main categorization [16]. In addition, the FCM has the potential to improve study reliability because Nova categorization is assigned at a more granular level through linkage to existing databases developed using FNDDS codes in a nationally representative United States Study [5,15,31]. Finally, UPF intake based on energy-adjusted gram weight outperformed intake based on energy alone, which may explain, in part, why some cohort studies have reported associations between UPF intake and disease risk using absolute or relative gram weight but not relative energy intake from UPF [13]. In conclusion, validation studies of FFQ-based UPF intake variables in other cohorts are warranted to better understand the degree of measurement error in UPF exposures, compare UPF-disease associations across studies, and improve scientific rigor and reproducibility.

Author contributions

EL, NK, EMS designed research; EL, PZ, CPO, LA, NK, EMS, HGH conducted research; PZ, CPO, HGH analyzed data; PZ, LLK, KAH, NK provided essential materials; and EL, PZ wrote the paper. EL, HGH had primary responsibility for final content; and all authors: read and approved the final manuscript. The opinions expressed by the authors are their own and this material should not be interpreted as representing the official viewpoint of the United States Department of Health and Human Services, the NIH, or the National Cancer Institute.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors did not use AI-assisted technologies and take full responsibility for the content of the publication.

Conflicts of interest

The authors report no conflicts of interest.

Funding

This research was supported by the Intramural Research Program at the National Cancer Institute.

Data availability

Data described in the manuscript, code book, and analytic code will be made available upon request pending application and approval, payment, and Data Transfer Agreement. For

additional information please see the NIH-AARP Diet and Health Study website at: https://www.nihaarpstars.com/

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tjnut.2025.04.029.

References

[1] F.E. Thompson, V. Kipnis, D. Midthune, L.S. Freedman, R.J. Carroll, A.F. Subar, Performance of a food-frequency questionnaire in the US NIH-AARP (National Institutes of Health-American Association of Retired Persons) Diet and Health Study, Public Health Nutr 11 (2) (2008) 183–195.

[2] L.S. Freedman, A. Schatzkin, Y. Wax, The impact of dietary measurement error on planning sample size required in a cohort study, Am. J. Epidemiol. 132 (6) (1990) 1185–1195.

[3] C.A. Monteiro, G. Cannon, J.C. Moubarac, R.B. Levy, M.L.C. Louzada, P.C. Jaime, The UN Decade of Nutrition, the NOVA food classification and the trouble with ultra-processing, Public Health Nutr 21 (1) (2018) 5–17.

[4] L. Wang, E. Martinez Steele, M. Du, J.L. Pomeranz, L.E. O'Connor, K.A. Herrick, et al., Trends in consumption of ultraprocessed foods among US Youths Aged 2–19 Years, 1999–2018, JAMA 326 (6) (2021) 519–530.

[5] F. Juul, N. Parekh, E. Martinez-Steele, C.A. Monteiro, V.W. Chang, Ultra-processed food consumption among US adults from 2001 to 2018, Am. J. Clin. Nutr. 115 (1) (2022) 211–221.

[6] NCD Risk Factor Collaboration, Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults, Lancet 390 (10113) (2017) 2627–2642.

[7] H. Dai, T.A. Alsalhe, N. Chalghaf, M. Ricco, N.L. Bragazzi, J. Wu, The global burden of disease attributable to high body mass index in 195 countries and territories, 1990–2017: an analysis of the Global Burden of Disease Study, PLoS Med 17 (7) (2020) e1003198.

[8] P.L. Valenzuela, P. Carrera-Bastos, A. Castillo-Garcia, D.E. Lieberman, A. Santos-Lozano, A. Lucia, Obesity and the risk of cardiometabolic diseases, Nat. Rev. Cardiol. 20 (7) (2023) 475–494.

[9] B. Lauby-Secretan, C. Scoccianti, D. Loomis, Y. Grosse, F. Bianchini, K. Straif, et al., Body fatness and cancer—viewpoint of the IARC working group, N. Engl. J. Med. 375 (8) (2016) 794–798.

[10] M.M. Lane, E. Gamage, S. Du, D.N. Ashtree, A.J. McGuinness, S. Gauci, et al., Ultra-processed food exposure and adverse health outcomes: umbrella review of epidemiological meta-analyses, BMJ 384 (2024) e077310.

[11] K.W. Dodd, P.M. Guenther, L.S. Freedman, A.F. Subar, V. Kipnis, D. Midthune, et al., Statistical methods for estimating usual intake of nutrients and foods: a review of the theory, J. Am. Diet. Assoc. 106 (10) (2006) 1640–1650.

[12] S.A. Smith-Warner, D. Spiegelman, J. Ritz, D. Albanes, W\.L. Beeson, L. Bernstein, et al., Methods for pooling results of epidemiologic studies: the Pooling Project of Prospective Studies of Diet and Cancer, Am. J. Epidemiol. 163 (11) (2006) 1053–1064.

[13] R. Cordova, V. Viallon, E. Fontvieille, L. Peruchet-Noray, A. Jansana, K.H. Wagner, et al., Consumption of ultra-processed foods and risk of multimorbidity of cancer and cardiometabolic diseases: a multinational cohort study, Lancet Reg. Health Eur. 35 (2023) 100771.

[14] L. Wang, M. Du, K. Wang, N. Khandpur, S.L. Rossato, J.P. Drouin-Chartier, et al., Association of ultra-processed food consumption with colorectal cancer risk among men and women: results from three prospective US cohort studies, BMJ 378 (2022) e068921.

[15] E.M. Steele, L.E. O'Connor, F. Juul, N. Khandpur, L.G. Baraldi, C.A. Monteiro, et al., Identifying and estimating ultraprocessed food intake in the US NHANES According to the Nova classification system of food processing, J. Nutr. 153 (1) (2023) 225–241.

[16] N. Khandpur, S. Rossato, J.P. Drouin-Chartier, M. Du, E.M. Steele, L. Sampson, et al., Categorising ultra-processed foods in large-scale cohort studies: evidence from the Nurses' Health Studies, the Health Professionals Follow-up Study, and the Growing Up Today Study, J. Nutr. Sci. 10 (2021) e77.

[17] A. Schatzkin, A.F. Subar, F.E. Thompson, L.C. Harlan, J. Tangrea, A.R. Hollenbeck, et al., Design and serendipity in establishing a large cohort with wide dietary intake distributions: the National Institutes of Health-American Association of Retired Persons Diet and Health Study, Am. J. Epidemiol. 154 (12) (2001) 1119–1125.

[18] A.F. Subar, D. Midthune, M. Kulldorff, C.C. Brown, F.E. Thompson, V. Kipnis, et al., Evaluation of alternative approaches to assign nutrient values to food groups in food frequency questionnaires, Am. J. Epidemiol. 152 (3) (2000) 279–286.

[19] K.S. Tippett, C.Y. Cypel (Eds.), Design and operation: the continuing survey of food intakes by individuals and the diet and health knowledge survey, 1994–96, US Dep. Agric. Agric. Res. Serv. (1997).

[20] E. Martinez-Steele, N. Khandpur, C. Batis, M. Bes-Rastrollo, M. Bonaccio, G. Cediel, et al., Best practices for applying the Nova food classification system, Nat. Food 4 (6) (2023) 445–448.

[21] W.C. Willett, G.R. Howe, L.H. Kushi, Adjustment for total energy intake in epidemiologic studies, Am. J. Clin. Nutr. 65 (4 Suppl) (1997) 1220S–1228S. ; discussion 1229S–1231S.

[22] G. Casella, R.L. Berger, Statistical Inference, Second Edition, Duxbury Thomson Learning, 2002.

[23] R. Landberg, P. Karra, R. Hoobler, E. Loftfield, I. Huybrechts, J.I. Rattner, et al., Dietary biomarkers—an update on their validity and applicability in epidemiological studies, Nutr. Rev. 82 (9) (2024) 1260–1280.

[24] S. Julian-Serrano, J. Reedy, K. Robien, R. Stolzenberg-Solomon, Adherence to 5 diet quality indices and pancreatic cancer risk in a large US prospective cohort, Am. J. Epidemiol. 191 (9) (2022) 1584–1600.

[25] D. Spiegelman, A. McDermott, B. Rosner, Regression calibration method for correcting measurement-error bias in nutritional epidemiology, Am. J. Clin. Nutr. 65 (4 Suppl) (1997) 1179S–1186S.

[26] A.G. Hullings, R. Sinha, L.M. Liao, N.D. Freedman, B.I. Graubard, E. Loftfield, Whole grain and dietary fiber intake and risk of colorectal cancer in the NIH-AARP Diet and Health Study cohort, Am. J. Clin. Nutr. 112 (3) (2020) 603–612.

[27] A.F. Subar, F.E. Thompson, V. Kipnis, D. Midthune, P. Hurwitz, S. McNutt, et al., Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: the Eating at America's Table Study, Am. J. Epidemiol. 154 (12) (2001) 1089–1099.

[28] A.F. Subar, V. Kipnis, R.P. Troiano, D. Midthune, D.A. Schoeller, S. Bingham, et al., Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study, Am. J. Epidemiol. 158 (1) (2003) 1–13.

[29] N. Tasevska, D. Midthune, N. Potischman, A.F. Subar, A.J. Cross, S.A. Bingham, et al., Use of the predictive sugars biomarker to evaluate self-reported total sugars intake in the Observing Protein and Energy Nutrition (OPEN) study, Cancer. Epidemiol. Biomarkers Prev. 20 (3) (2011) 490–500.

[30] Y. Park, K.W. Dodd, V. Kipnis, F.E. Thompson, N. Potischman, D.A. Schoeller, et al., Comparison of self-reported dietary intakes from the automated self-administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers, Am. J. Clin. Nutr. 107 (1) (2018) 80–93.

[31] L.E. O'Connor, K.D. Hall, K.A. Herrick, J. Reedy, S.T. Chung, M. Stagliano, et al., Metabolomic profiling of an ultraprocessed dietary pattern in a domiciled randomized controlled crossover feeding trial, J. Nutr. 153 (8) (2023) 2181–2192.