

Performance of a food-frequency questionnaire in the US NIH–AARP (National Institutes of Health–American Association of Retired Persons) Diet and Health Study

Frances E Thompson^{1,*}, Victor Kipnis¹, Douglas Midthune¹, Laurence S Freedman², Raymond J Carroll³, Amy F Subar¹, Charles C Brown¹, Matthew S Butcher⁴, Traci Mouw¹, Michael Leitzmann¹ and Arthur Schatzkin¹

¹US National Cancer Institute, EPN 4016, 9000 Rockville Pike, Bethesda, MD 20893-7344, USA; ²Gertner Institute for Epidemiology, Tel Hashomer, Israel; ³Department of Statistics, Texas A&M University, College Station, TX, USA; ⁴Information Management Services, Inc., Silver Spring, MD, USA

Submitted 2 August 2006; Accepted 12 March 2007; First published online 5 July 2007

Abstract

Objective: We evaluated the performance of the food-frequency questionnaire (FFQ) administered to participants in the US NIH–AARP (National Institutes of Health–American Association of Retired Persons) Diet and Health Study, a cohort of 566 404 persons living in the USA and aged 50–71 years at baseline in 1995.

Design: The 124-item FFQ was evaluated within a measurement error model using two non-consecutive 24-hour dietary recalls (24HRs) as the reference.

Setting: Participants were from six states (California, Florida, Pennsylvania, New Jersey, North Carolina and Louisiana) and two metropolitan areas (Atlanta, Georgia and Detroit, Michigan).

Subjects: A subgroup of the cohort consisting of 2053 individuals.

Results: For the 26 nutrient constituents examined, estimated correlations with true intake (not energy-adjusted) ranged from 0.22 to 0.67, and attenuation factors ranged from 0.15 to 0.49. When adjusted for reported energy intake, performance improved; estimated correlations with true intake ranged from 0.36 to 0.76, and attenuation factors ranged from 0.24 to 0.68. These results compare favourably with those from other large prospective studies. However, previous biomarker-based studies suggest that, due to correlation of errors in FFQs and self-report reference instruments such as the 24HR, the correlations and attenuation factors observed in most calibration studies, including ours, tend to overestimate FFQ performance.

Conclusion: The performance of the FFQ in the NIH–AARP Diet and Health Study, in conjunction with the study's large sample size and wide range of dietary intake, is likely to allow detection of moderate (≥ 1.8) relative risks between many energy-adjusted nutrients and common cancers.

Keywords

Diet
Epidemiological methods
Validation study
Cancer
Questionnaires

In 1995 the National Institutes of Health (NIH) and the AARP, formerly the American Association of Retired Persons, initiated a large prospective cohort study, the NIH–AARP Diet and Health Study, to study relationships between diet and cancer. The study population consisted of members of AARP in six states (California, Florida, Pennsylvania, New Jersey, North Carolina and Louisiana) and two metropolitan areas (Atlanta, Georgia and Detroit, Michigan) who were 50–71 years of age at baseline.

One criticism of some cohort studies is the lack of heterogeneity in reported diet. In order to address this problem, the original design of the NIH–AARP Diet and Health Study called for a two-stage sampling frame: col-

lection of baseline food intake data for a large pool of respondents, and selection into the cohort based on these data, with over-sampling in extreme categories of dietary intake¹. Ultimately, this two-stage plan was abandoned, as two unanticipated factors emerged. First, the actual response rate to the baseline questionnaire (17.6%) was lower than the anticipated response (33.0%); second, the range of intake observed was greater than expected. Thus, the decision to include as subjects the entire baseline population led to statistical power approaching that expected for the two-stage design¹.

Baseline dietary intake was assessed by a food-frequency questionnaire (FFQ). Measurement error, both

*Corresponding author: Email thompsonf@mail.nih.gov

random and systematic, in the FFQ method may be substantial^{2,3}, generally leading to bias in observed relative risks and loss of power to detect diet–disease relationships^{4,5}. Thus, a calibration sub-study was carried out in approximately 2000 subjects, with two 24-hour dietary recalls (24HRs) as the reference instrument. The main purpose was to evaluate the performance of the FFQ instrument in this study and to correct estimated diet–disease relationships for measurement error. We present here validation results for 24 nutrients and for servings of fruits and vegetables. Evaluation of the FFQ-reported intake for most other food groups requires a different methodology due to a non-trivial probability of zero reported intake on any given 24HR and will be presented in a separate paper.

Methods

Design

The design of the NIH–AARP Diet and Health Study is described in detail elsewhere^{1,6}. Briefly, a baseline questionnaire which included a 124-item FFQ was mailed to 3.5 million members of AARP, in three waves. The first wave of 250 000 questionnaires was mailed in October 1995. The calibration sub-study participants were selected from the 46 970 subjects who had responded to the first wave as of January 1996. For purposes of defining the eligible sample for selection into the calibration sub-study, we excluded individuals for a variety of reasons⁶ leaving a pool of 38 691 eligible for selection into the calibration study.

The calibration study was designed to match the original two-stage design of the entire cohort study. A stratified sampling design using FFQ-derived estimates of intake of percentage energy from fat, fruit and vegetables, fibre and red meat to form strata was implemented⁶. For each of these dietary variables, five categories of intake were specified for each gender, consisting of approximately 10% in each of the lowest and highest categories, with the remaining 80% distributed equally among the middle three categories. The five categories and four dietary variables were used to partition eligible respondents into 20 strata. Two additional strata were added to reflect high alcohol intakes (>90 g per day). A combination of fixed and random sampling within strata was implemented to ensure approximate representation from each gender, enhance minority inclusion and increase dietary heterogeneity for these four dietary variables. Extreme intakes (strata 1 and 5) on any of the four variables were sampled at the highest rate; proportionately fewer were sampled from strata 2 and 4; and even fewer were sampled from stratum 3.

In order to attain a final sample of 2000, a pool of 6150 individuals was chosen, and telephone numbers for 3901 individuals obtained. Recruitment calls were attempted

on 3647 and 2923 were reached; 128 were excluded because they could not be located, had language problems, or were too ill to participate. Of the 2795 individuals invited to participate, 2055 agreed to participate. Because of changes in an FFQ exclusion criterion (exclude if >10 rather than >40 scanning errors), two participants were excluded subsequently. Trained interviewers administered two non-consecutive, unannounced 24HRs by telephone, randomly assigned by day of the week. The two recalls were well separated in time (a median of 21 days apart and 75% separated by 14 or more days) from March to September 1996. Of the 2053 participants who completed the first 24HR, 1986 provided a second 24HR. All 2053 were mailed a second FFQ, similar to the baseline instrument, in October 1996; 1445 were returned. The lack of temporal agreement between the recalls and the FFQs may create error, although there is little evidence to suggest that nutrient intakes vary dramatically from season to season.

For these analyses, we excluded 100 individuals from the calibration sub-study, consistent with exclusions made on the overall baseline cohort, for the following reasons: subsequent drop outs ($n = 78$); pre-baseline registry reports of cancer ($n = 9$); and death-only reports of cancer ($n = 13$). Thus, the analysed calibration sample consisted of 1953 individuals (987 men, 966 women) with at least the baseline FFQ and one 24HR. Data collection for the NIH–AARP Diet and Health Study was approved by the National Cancer Institute (NCI) and the Westat, Inc. institutional review boards.

Study instruments

The FFQ used in the NIH–AARP Diet and Health Study was an early version of the Diet History Questionnaire (DHQ) developed at the NCI⁷. Frequency responses were asked for 124 food items; portion sizes for 116. An additional 21 questions asked about specific food choices and cooking practices. The US Department of Agriculture's (USDA) Continuing Survey of Food Intakes of Individuals (CSFII) survey databases (1989–91 initially, and 1994–96 as it became available) were used to develop a nutrient composition database for the FFQ⁸. Estimates of individual intake were calculated for 29 nutrients and 30 food groups.

In the 24HR interviews, participants were asked to report all foods and beverages consumed on the day before the interview. Interviewers used a Food Probe List containing standardised probes specific to foods in over 100 food categories. Data were coded using the Food Intake Analysis System (FIAS), version 2.3, developed at the University of Texas; the same nutrient composition database is used for both FIAS and USDA's CSFII. Data checks were performed on reports with extremely high values for fat, total energy and total fruit and vegetable intakes, and corrections were made to the data when extreme values were due to coding errors.

Statistical methods

Two parameters are important for evaluating the performance of a dietary assessment instrument (Q), in this case, the FFQ: the Pearson correlation coefficient $\rho_{Q,T}$ between reported and true intake, often referred to as the validity coefficient, and the attenuation factor λ_Q , which is the slope in the regression of true intake on reported intake. If an FFQ-measured exposure is categorised into quantiles, the correlation $\rho_{Q,T}$ determines the extent of misclassification due to measurement error⁹. In a univariate model relating dietary exposure to disease, the observed relative risk for a given change in the exposure is equal to the true relative risk raised to power λ_Q . In dietary studies, λ_Q is usually confined between zero and one, therefore attenuating (biasing toward one) true relative risks. An attenuation factor close to zero indicates severe attenuation and an attenuation factor close to one signals very little attenuation. For example, the true relative risk of 2 for a given change in the exposure would be observed as $2^{0.5} = 1.41$ if $\lambda_Q = 0.5$ and only $2^{0.1} = 1.07$ if $\lambda_Q = 0.1$. In addition, the correlation with true intake is related to statistical power to detect diet–disease relationships. Compared with using true intake, the required sample size to reach the desired statistical power should be increased by the inverse of the squared correlation with true intake. For example, if $\rho_{Q,T} = 0.5$, the required sample size should be increased $1/0.5^2 = 4$ times to achieve the nominal power.

Ideally, one would like to measure true intake in the calibration study to estimate the correlation with true intake and attenuation factor. Unfortunately, diet cannot be measured precisely in free-living populations, either using dietary assessment instruments or biomarkers of intake. Estimation of both parameters in this situation can still be performed using statistical modelling but requires administration of a reference instrument. The reference instrument does not have to measure intake perfectly and may contain error of its own, but this error is required to be (a) independent of true intake and (b) independent of error in the FFQ¹⁰. Throughout this paper we shall refer to these two conditions as defining a correct reference instrument. Under conditions (a) and (b), the correlation with true intake is equal to the correlation between the FFQ and reference instrument adjusted for within-person random error in the reference measure¹¹. Its estimation in the calibration study requires that the reference instrument be administered at least twice. The attenuation factor is equal to the slope of the regression of the reference measure on the FFQ, and its estimation requires only one administration of the reference instrument¹².

In the literature, the correlation with true intake and the attenuation factor are usually estimated by applying the method of moments to those individuals in the calibration study who have fully completed the FFQ and necessary reference measurements^{13–15}. Alternatively, both parameters can be estimated by fitting a measurement error

model for the FFQ and reference measurements to all participants of the calibration study using the method of maximum likelihood¹⁰. Under a multivariate normal distribution, the second approach produces the same results as the method of moments if there are no missing data, but otherwise leads to more efficient estimation. We adopted the latter approach in the present study using the general measurement error model for the FFQ that was suggested by Kipnis *et al.*¹⁰ (see Appendix B for a more detailed explanation).

In the NIH–AARP calibration study, the reference instrument was the 24HR. The distributions of most nutrient intakes derived from the 24HRs and FFQs were typically skewed and contained some extreme values that might overly influence parameter estimation. Therefore, the measurement error model was fit after first removing outliers and transforming intakes to approximate normality using the Box–Cox transformation family¹⁶ as explained in detail in Appendix A. Besides normalising the data, the transformation also aimed to improve properties of the 24HR as a reference instrument by making its within-person variation constant and independent of its between-person variation. For FFQ-reported data, the number of removed outliers ranged from one for protein and cholesterol to 24 for vitamin C. For 24HR data, the number of removed outliers ranged from three for saturated fat and vitamin C to 56 for vitamin E.

We estimated correlations with true intake and attenuation factors for both absolute and energy-adjusted nutrient intakes, using the residual energy-adjustment method¹⁷. Due to substantial day-to-day variation in the 24HR, nutrient residuals calculated from its repeat administrations may have substantially correlated within-person random errors, therefore violating requirements for a correct reference measure, even if those requirements were satisfied for the absolute nutrient and total energy intakes. To address this problem, we estimated correlations and attenuation factors for a nutrient residual by fitting simultaneous measurement error models¹⁸ to the nutrient of interest and total energy intakes reported on the FFQ and 24HR, as outlined in Appendix B. The parameters of the measurement error model were estimated using the SAS Mixed procedure¹⁹.

As described above, in AARP the calibration study subjects were selected using stratified sampling; i.e. subjects from the first wave in the main study were selected into the calibration study with probabilities based on their reported intakes on the baseline FFQ for the five strata-determining dietary variables (percentage energy from fat, dietary fibre, fruit and vegetables, red meat and alcohol). Since the missing mechanism depends only on the value of the baseline FFQ, and not on the unobserved 24HR and second FFQ, the data are missing at random. The maximum likelihood estimator ignores missing data as long as they are missing at random, at the expense of including data for the nutrient(s) of interest as well as five

strata-determining dietary variables from all subjects in the main study (not just those in the calibration study)²⁰. On the other hand, the distributions of dietary intake in the calibration study were similar to those in the final baseline cohort, justifying a simple random sample analysis. We fit the measurement error models using both stratified and simple random sample estimation procedures. Since results were similar, we present the simple random sample estimates only.

Results

Comparison of baseline cohort and calibration sub-sample

Demographic and selected exposure characteristics of the eligible baseline cohort and the calibration sub-sample are shown in Table 1. The calibration sub-sample reflects the full baseline cohort with regard to demographic composition (age, race/ethnicity, educational attainment), environmental exposure status (smoking, physical activity, diet, body mass index, hormone use), self-reported health status and family history of cancer. In addition, despite the stratified sampling into the calibration study, the median values of the dietary factors presented (Table 1) and their distributions (results not shown) were similar for both groups. This indicates that the calibration study was practically a simple random sample from the main cohort.

Comparison among median intakes

Estimates of median intakes reported on the FFQ and 24HR after adjustment for within-person variation are presented in Table 2. We present medians rather than means because the distributions are skewed. The estimates for both instruments were obtained by estimating the means on the transformed scale with approximately normal (symmetric) distribution (see Methods section) and then back-transforming the means to the original scale. Under the assumption of a correct reference instrument, this procedure applied to the 24HR produces estimates of median usual intake on the original scale.

Comparison of the first and second administrations of each instrument – the FFQ and the 24HR – generally showed lower reported intakes on the second administration (Table 2). However, most differences were trivial. Comparisons between FFQ and 24HR generally showed few large differences. When we compared the median of the first FFQ with the median of the first 24HR (on the transformed scale), few differences of 15% or more were observed: intakes of protein (men), thiamin (men) and riboflavin (men) on the 24HR were higher than on the FFQ; reports of vitamin B₆ (women), vitamin B₁₂ (women), and fruits and vegetables were higher on the FFQ than the 24HR. Estimates of other nutrients including

energy, carbohydrate, fat, fibre and all energy-adjusted macronutrients were similar for the two instruments.

Estimated correlations with true intake and attenuation factors

Using 24HR as a reference instrument, estimates of correlation coefficients between FFQ-reported and true nutrient intakes and of attenuation factors for FFQ are presented in Table 3. Correlations for energy were 0.39 and 0.22 for men and women, respectively. For men, correlations for absolute nutrient intakes ranged from 0.37 (vitamin E) to 0.67 (dietary cholesterol). For women, correlations ranged from 0.23 (vitamin E) to 0.56 (dietary cholesterol). Attenuation factors for energy were 0.26 among men and 0.15 among women. For men, attenuation factors for absolute nutrient intakes ranged from 0.24 (protein) to 0.53 (cholesterol); for women, they ranged from 0.16 (protein) to 0.43 (vitamin C). After energy adjustment, estimated correlations with true intake and attenuation factors generally rose, especially for women. Correlation coefficients ranged from 0.40 (vitamin E) to 0.76 (saturated fat) among men, and from 0.36 (vitamin E) to 0.70 (vitamin B₆) among women. Attenuation factors ranged from 0.26 (protein) to 0.68 (saturated fat) among men and from 0.24 (vitamin E) to 0.62 (vitamin B₆) among women.

Discussion

The NIH–AARP Diet and Health Study is among the largest prospective studies of diet and cancer in the world, providing large numbers of incident cases for a variety of cancers. It also provides the largest calibration dataset in the USA, with 24HR data on nearly 2000 individuals.

The FFQ used in this study is an early version of the NCI DHQ, which has been evaluated in two other studies^{21–23}. The principal difference between the NIH–AARP Diet and Health Study FFQ and the DHQ is the response format: the AARP FFQ uses a grid while the DHQ does not. In the Eating at America's Table Study (EATS)²¹, four 24HRs were used as a reference instrument to evaluate the DHQ as well as the Willett and Block questionnaires. The estimated correlations between DHQ-reported and true nutrient intakes and the attenuation factors for DHQ in EATS were generally high and similar to those in the current study, especially after energy adjustment (see Table 4 below). A second study, the Observing Protein and Energy Nutrition (OPEN) study^{22,23}, used urinary nitrogen and doubly labelled water as established reference biomarkers for intakes of protein and energy, respectively, to estimate the measurement error structure in the DHQ and 24HR, and to evaluate the performance of the 24HR as a reference instrument for protein, energy and energy-adjusted

Table 1 Cohort and calibration sub-sample characteristics at baseline by gender: NIH–AARP Diet and Health Study, 1995–1996

| Characteristic | Men | | Women | |
|---|----------------------------------|-------------------------------------|----------------------------------|-------------------------------------|
| | Baseline cohort (n = 298 449) | Calibration sub-sample (n = 987) | Baseline cohort (n = 204 844) | Calibration sub-sample (n = 966) |
| Age (years), mean | 62.2 | 62.3 | 61.9 | 62.0 |
| Height (in), mean | 70.1 | 70.1 | 64.3 | 64.4 |
| Weight (lb), mean | 190.6 | 189.2 | 157.0 | 157.8 |
| BMI (kg m ⁻²), mean | 27.2 | 27.0 | 26.7 | 26.7 |
| Race/ethnicity (%) | | | | |
| White, non-Latino | 92.7 | 91.7 | 89.6 | 92.2 |
| Black, Latino, Asian, other | 6.2 | 7.6 | 8.8 | 7.0 |
| Unknown | 1.2 | 0.7 | 1.5 | 0.7 |
| Education (%) | | | | |
| <12 years | 5.7 | 5.7 | 6.0 | 4.5 |
| 12 years/high school | 15.4 | 13.9 | 25.3 | 24.6 |
| Vocational/technical | 9.2 | 12.1 | 10.6 | 10.8 |
| Some college | 22.1 | 20.7 | 24.8 | 27.0 |
| College graduate | 21.8 | 22.1 | 14.9 | 15.8 |
| Postgraduate | 23.2 | 23.7 | 15.2 | 14.3 |
| Unknown | 2.6 | 1.9 | 3.3 | 3.0 |
| Family history of cancer (%) | | | | |
| No family history of cancer | 47.2 | 48.8 | 43.3 | 43.0 |
| Some family history of cancer | 47.3 | 46.1 | 51.6 | 52.2 |
| Unknown | 5.4 | 5.1 | 5.1 | 4.9 |
| Smoking status (%) | | | | |
| Never smoker | 29.7 | 28.7 | 44.3 | 41.3 |
| Current smoker | 58.3 | 57.9 | 39.7 | 40.1 |
| Former smoker | 10.4 | 12.1 | 14.2 | 16.4 |
| Unknown | 1.7 | 1.4 | 1.8 | 2.3 |
| Physical activity (%) | | | | |
| Never/rarely | 14.3 | 13.9 | 21.8 | 20.9 |
| 1–3/month | 13.1 | 11.5 | 14.4 | 14.5 |
| 1–2/week | 22.0 | 21.7 | 21.1 | 20.7 |
| 3–4/week | 28.2 | 30.5 | 25.1 | 26.5 |
| >4/week | 21.5 | 21.8 | 16.3 | 16.5 |
| Unknown | 0.9 | 0.7 | 1.3 | 0.9 |
| Self-rated health status (%) | | | | |
| Excellent | 17.5 | 18.3 | 16.9 | 18.0 |
| Very good | 36.0 | 35.5 | 35.0 | 35.9 |
| Good | 34.8 | 35.4 | 34.9 | 33.0 |
| Fair | 10.5 | 9.3 | 11.4 | 10.8 |
| Poor* | — | — | — | — |
| Unknown | 1.3 | 1.5 | 1.8 | 2.3 |
| Currently using HRT (%) | | | | |
| No | — | — | 54.4 | 55.7 |
| Yes | — | — | 44.5 | 43.2 |
| Unknown | — | — | 1.2 | 1.1 |
| Dietary intake on baseline FFQ | | | | |
| Servings of fruits and vegetables per day, median | 6.2 | 6.3 | 6.0 | 6.0 |
| % energy from fat, median | 30.6 | 30.8 | 29.9 | 29.7 |
| Fibre (g), median | 18.4 | 19.2 | 16.1 | 15.8 |
| Total energy (kcal), median | 1866 | 1896 | 1456 | 1496 |
| Total fat (g), median | 61.4 | 64.1 | 47.0 | 48.5 |
| Carbohydrate (g), median | 235.3 | 242.3 | 195.6 | 199.1 |
| Protein (g), median | 70.4 | 71.8 | 55.6 | 55.8 |

NIH – National Institutes of Health; BMI – body mass index; HRT – hormone replacement therapy; FFQ – food-frequency questionnaire.

* Those reporting poor health were excluded from the baseline cohort.

protein intakes. Implications of the OPEN study, as well as other biomarker studies, on the current analysis are discussed below.

In the NIH–AARP Diet and Health Study, estimated correlations with true absolute nutrient intakes were generally in the range of 0.4–0.6. Energy adjustment generally improved correlation coefficients and attenuation factors, particularly among women, where relatively

low validity coefficients for energy were observed. We compare these correlation coefficients with correlations attained in previously published validation studies (Table 4). Studies included for comparison are those which present gender-specific data with ‘deattenuated’ and/or energy-adjusted correlation coefficients. These studies compared FFQs with multiple days of 24HRs^{21,24–26} or food records²⁷ or both²⁸. Most studies

Table 2 Estimates of median daily intakes of nutrients/dietary constituents reported on the FFQ and 24HR in the calibration sub-sample, after adjustment for within-person variation: NIH-AARP Diet and Health Study, 1995–1996

| Nutrient*/dietary constituent (intake per day) | Men (n = 987) | | | | Women (n = 966) | | | |
|--|---------------|--------|-------|--------|-----------------|--------|-------|--------|
| | FFQ | | 24HR | | FFQ | | 24HR | |
| | First | Second | First | Second | First | Second | First | Second |
| Energy (kcal) | 1953 | 1867 | 2231 | 2192 | 1540 | 1464 | 1592 | 1568 |
| Protein (g) | 74.2† | 71.7 | 89.3 | 89.0 | 58.2 | 55.9 | 62.5 | 63.6 |
| Protein (% energy) | 15.1 | 15.3 | 15.8 | 16.1 | 15.2 | 15.2 | 15.8 | 16.2 |
| Carbohydrate (g) | 249.1 | 238.7 | 280.9 | 272.9 | 206.5 | 194.7 | 208.0 | 202.4 |
| Carbohydrate (% energy) | 51.7 | 51.7 | 51.3 | 50.7 | 54.2 | 53.4 | 53.0 | 52.4 |
| Fat (g) | 64.5 | 62.8 | 73.9 | 73.9 | 49.9 | 49.2 | 51.9 | 51.9 |
| Fat (% energy) | 30.5 | 31.0 | 30.6 | 31.2 | 29.7 | 30.7 | 29.9 | 30.3 |
| Saturated fat (g) | 19.9 | 19.2 | 23.6 | 23.6 | 15.1 | 14.9 | 16.7 | 16.8 |
| Monounsaturated fat (g) | 24.4 | 24.0 | 27.3 | 27.2 | 18.4 | 18.3 | 18.8 | 18.7 |
| Polyunsaturated fat (g) | 14.0 | 13.7 | 14.9 | 14.9 | 11.8 | 11.6 | 11.0 | 11.0 |
| Cholesterol (mg) | 205 | 201 | 248 | 252 | 148 | 147 | 161 | 173 |
| Dietary fibre (g) | 19.5 | 18.7 | 19.6 | 19.1 | 16.6 | 15.7 | 14.5 | 14.3 |
| Vitamin A (µg RE) | 1300 | 1222 | 1057 | 1089 | 1226 | 1137 | 865 | 867 |
| Vitamin E (mg α-TE) | 8.73 | 8.64 | 8.50 | 8.60 | 7.52 | 7.28 | 6.32 | 6.36 |
| Vitamin C (mg) | 138 | 126 | 122 | 120 | 133 | 124 | 105 | 108 |
| Thiamin (mg) | 1.60 | 1.56 | 1.85 | 1.84 | 1.26 | 1.19 | 1.29 | 1.28 |
| Riboflavin (mg) | 1.91† | 1.85 | 2.15 | 2.10 | 1.56 | 1.47 | 1.54 | 1.58 |
| Niacin (mg) | 23.0 | 22.3 | 26.4 | 26.2 | 17.8 | 17.0 | 18.1 | 18.3 |
| Folate (mg) | 306 | 296 | 332 | 323 | 266 | 249 | 248 | 253 |
| Vitamin B ₆ (mg) | 2.02 | 1.98 | 2.18 | 2.13 | 1.66† | 1.56 | 1.53 | 1.57 |
| Calcium (mg) | 761 | 729 | 821 | 809 | 661 | 622 | 631 | 645 |
| Iron (mg) | 15.6 | 15.2 | 17.0 | 16.8 | 12.2 | 11.6 | 12.0 | 11.7 |
| Magnesium (mg) | 338 | 325 | 341 | 334 | 284 | 270 | 255 | 253 |
| Phosphorus (mg) | 1238 | 1193 | 1437 | 1409 | 992 | 946 | 1025 | 1034 |
| Zinc (mg) | 10.98 | 10.67 | 12.00 | 11.81 | 8.30 | 7.89 | 8.30 | 8.52 |
| Potassium (mg) | 3487 | 3343 | 3453 | 3418 | 2031 | 2869 | 2592 | 2641 |
| Vitamin B ₁₂ (mg) | 4.78 | 4.63 | 4.51 | 4.47 | 3.55† | 3.46 | 2.83 | 2.96 |
| Fruit & vegetables (Pyramid servings) | 6.4 | 6.1 | 6.1 | 6.3 | 6.3† | 5.9 | 5.1 | 5.4 |

FFQ – food-frequency questionnaire; 24HR – 24-hour dietary recall; NIH – National Institutes of Health; RE – retinal equivalents; α-TE – α-tocopherol equivalents.

* Values for nutrients are from dietary intakes only and do not include reported supplements.

† The first FFQ value is ≥15% different from the first 24HR value, on the transformed scale.

presented deattenuated energy-adjusted estimates^{21,25–28}. Munger *et al.*²⁴ presented energy-adjusted estimates without first deattenuating. In general, within the range of correlations observed among these studies, those from the NIH-AARP Diet and Health Study are among the higher ones. For example, for energy-adjusted total fat, correlations in other studies range from 0.55 to 0.77, compared with AARP correlations of 0.72 (men) and 0.62 (women). Similarly, for energy-adjusted fibre, correlations in other studies range from 0.24 to 0.80, compared with AARP correlations of 0.72 (men) and 0.66 (women). Although the multitude of studies does give a qualitative sense of how well various nutrients are estimated using FFQs, one should exercise caution in interpreting the results of the direct comparison of validity coefficients from these studies conducted in different populations with different dietary assessment reference instruments²⁹.

Correlations with true intake and attenuation factors in the study were estimated using 24HR as a reference instrument, i.e. assuming that 24HR-reported intake is unbiased at the individual level and includes only within-person random error that is independent of true intake and error in the FFQ. Recent studies using reference biomarkers^{10,23,30} have demonstrated that these assump-

tions may not hold – for protein, total energy and energy-adjusted protein, the 24HR has both intake-related and person-specific biases correlated with their counterparts in the FFQ. Thus, using the 24HR as a reference instrument may lead to biased estimates of the FFQ performance. For example, Kipnis *et al.* derived a formula relating the true attenuation factor to the one estimated using the 24HR as a reference instrument²³. They demonstrate that intake-related bias in 24HR resulting in the flattened slope phenomenon, i.e. when the regression of the 24HR-reported on true intake has a slope smaller than one, tends to underestimate the attenuation factor (overestimate true attenuation). On the other hand, person-specific bias in 24HR as well as its correlation with person-specific bias in the FFQ leads to an overestimated attenuation factor (underestimated true attenuation). The resulting bias in the estimated attenuation factor therefore depends on the relative values of these three parameters that characterise deviations of 24HR from the requirements to a correct reference measure. In the OPEN study, the interplay of the flattened slope and person-specific bias in 24HR as well as correlation between person-specific biases in the 24HR and FFQ led to a minimal overestimation of the FFQ performance (20% or

Table 3 Estimated correlation coefficients and attenuation factors between the FFQ and truth using a measurement error model, unadjusted and adjusted for energy intake: NIH–AARP Diet and Health Study, 1995–1996

| Nutrient/dietary constituent | Correlation coefficient | | Attenuation factor | |
|------------------------------|-------------------------|-----------------|--------------------|-----------------|
| | Men (n = 987) | Women (n = 966) | Men (n = 987) | Women (n = 966) |
| Energy | 0.39 | 0.22 | 0.26 | 0.15 |
| Protein | | | | |
| Unadjusted | 0.41 | 0.30 | 0.24 | 0.16 |
| Adjusted | 0.43 | 0.50 | 0.26 | 0.31 |
| Carbohydrate | | | | |
| Unadjusted | 0.39 | 0.32 | 0.32 | 0.23 |
| Adjusted | 0.71 | 0.64 | 0.62 | 0.50 |
| Fat | | | | |
| Unadjusted | 0.58 | 0.39 | 0.42 | 0.28 |
| Adjusted | 0.72 | 0.62 | 0.57 | 0.45 |
| Saturated fat | | | | |
| Unadjusted | 0.65 | 0.48 | 0.52 | 0.37 |
| Adjusted | 0.76 | 0.69 | 0.68 | 0.58 |
| Monounsaturated fat | | | | |
| Unadjusted | 0.58 | 0.41 | 0.41 | 0.30 |
| Adjusted | 0.71 | 0.62 | 0.56 | 0.46 |
| Polyunsaturated fat | | | | |
| Unadjusted | 0.47 | 0.33 | 0.32 | 0.22 |
| Adjusted | 0.53 | 0.56 | 0.38 | 0.28 |
| Cholesterol | | | | |
| Unadjusted | 0.67 | 0.56 | 0.53 | 0.38 |
| Adjusted | 0.68 | 0.68 | 0.64 | 0.52 |
| Dietary fibre | | | | |
| Unadjusted | 0.50 | 0.42 | 0.39 | 0.31 |
| Adjusted | 0.72 | 0.66 | 0.60 | 0.55 |
| Vitamin A | | | | |
| Unadjusted | 0.46 | 0.42 | 0.35 | 0.31 |
| Adjusted | 0.59 | 0.59 | 0.47 | 0.45 |
| Vitamin E | | | | |
| Unadjusted | 0.37 | 0.23 | 0.27 | 0.17 |
| Adjusted | 0.40 | 0.36 | 0.32 | 0.24 |
| Vitamin C | | | | |
| Unadjusted | 0.60 | 0.54 | 0.47 | 0.43 |
| Adjusted | 0.70 | 0.65 | 0.55 | 0.53 |
| Thiamin | | | | |
| Unadjusted | 0.38 | 0.37 | 0.28 | 0.26 |
| Adjusted | 0.62 | 0.59 | 0.52 | 0.48 |
| Riboflavin | | | | |
| Unadjusted | 0.48 | 0.44 | 0.34 | 0.31 |
| Adjusted | 0.64 | 0.67 | 0.47 | 0.50 |
| Niacin | | | | |
| Unadjusted | 0.40 | 0.32 | 0.27 | 0.18 |
| Adjusted | 0.56 | 0.57 | 0.44 | 0.39 |
| Folate | | | | |
| Unadjusted | 0.48 | 0.49 | 0.42 | 0.39 |
| Adjusted | 0.64 | 0.69 | 0.60 | 0.59 |
| Vitamin B ₆ | | | | |
| Unadjusted | 0.51 | 0.46 | 0.39 | 0.34 |
| Adjusted | 0.69 | 0.70 | 0.61 | 0.62 |
| Calcium | | | | |
| Unadjusted | 0.54 | 0.48 | 0.40 | 0.36 |
| Adjusted | 0.63 | 0.64 | 0.46 | 0.47 |
| Iron | | | | |
| Unadjusted | 0.42 | 0.36 | 0.33 | 0.27 |
| Adjusted | 0.59 | 0.56 | 0.52 | 0.48 |
| Magnesium | | | | |
| Unadjusted | 0.45 | 0.35 | 0.35 | 0.24 |
| Adjusted | 0.64 | 0.66 | 0.58 | 0.53 |
| Phosphorus | | | | |
| Unadjusted | 0.46 | 0.36 | 0.31 | 0.23 |
| Adjusted | 0.61 | 0.63 | 0.42 | 0.44 |
| Zinc | | | | |
| Unadjusted | 0.45 | 0.40 | 0.30 | 0.27 |
| Adjusted | 0.49 | 0.57 | 0.35 | 0.43 |

Table 3 *Continued*

| Nutrient/dietary constituent | Correlation coefficient | | Attenuation factor | |
|------------------------------|-------------------------|-----------------|--------------------|-----------------|
| | Men (n = 987) | Women (n = 966) | Men (n = 987) | Women (n = 966) |
| Potassium | | | | |
| Unadjusted | 0.45 | 0.39 | 0.33 | 0.26 |
| Adjusted | 0.63 | 0.67 | 0.51 | 0.51 |
| Vitamin B ₁₂ | | | | |
| Unadjusted | 0.48 | 0.44 | 0.31 | 0.32 |
| Adjusted | 0.47 | 0.47 | 0.32 | 0.34 |
| Fruit & vegetables | | | | |
| Unadjusted | 0.59 | 0.46 | 0.47 | 0.37 |
| Adjusted | 0.72 | 0.61 | 0.61 | 0.54 |

FFQ – food-frequency questionnaire; NIH – National Institutes of Health.

less for both correlation and attenuation) for absolute protein intake. The resulting overestimation for total energy intake, however, ranged from 120% (correlation with true intake in men) to 230% (attenuation in women)²³. Biases this large would make it difficult to estimate moderate diet–cancer relationships using non-energy-adjusted models, even in a study as large as AARP. For example, if the estimated attenuation factor of 0.28 for absolute fat intake in women were biased upward by 20%, then the true attenuation factor would be about 0.23, and a true fat–breast cancer relative risk of 1.80 would be observed as 1.14. If the estimate were biased upward by 100%, then the true attenuation factor would be about 0.14, and a relative risk of 1.80 would be observed as only 1.09.

Fortunately, energy adjustment appears to improve the situation substantially. The original rationale for this procedure was to evaluate the nutrient effect independent of overall energy intake³¹. A second rationale for the procedure is to reduce measurement error³². Because the error in reported energy intake is correlated with the error in reported intake for most other nutrients on the same instrument, controlling for energy reduces this correlated error. The published OPEN results²³ used the density method rather than the residual method to energy-adjust protein, but the results for residual protein were similar to those for protein density. In OPEN, energy-adjusted protein on 24HR had substantially reduced person-specific bias and somewhat more pronounced flattened slope compared with absolute reported protein intake. Correlation between person-specific biases in the FFQ and 24HR, however, increased twofold in men and fourfold in women. As a result of the interplay among these factors, using the 24HR as a reference instrument produced an essentially unbiased estimate of the attenuation factor for energy-adjusted protein intake in men, while in women the estimated attenuation factor was biased upward by about 40%. If similar biases are assumed for energy-adjusted fat intake in the AARP calibration study, the estimated attenuation factor of 0.45 for

Table 4 Correlation coefficients for nutrients, unadjusted and adjusted for energy intake, by gender and study

| Nutrient/dietary constituent | Men | | | | | | | Women | | | | | | | |
|------------------------------|------|------------------------|--------------------|--------------------|--------------------|---------------------|---------------------|-------|------------------------|--------------------|----------------------|--------------------|---------------------|---------------------|-------------------------|
| | NCI | | Willett | | Block | Hawaii | ACS | NCI | | Willett | | Block | Hawaii | ACS | WHI |
| | AARP | EATS-DHQ ²¹ | EATS ²¹ | Rimm ²⁷ | EATS ²¹ | Stram ²⁵ | Flagg ²⁶ | AARP | EATS-DHQ ²¹ | EATS ²¹ | Munger ²⁴ | EATS ²¹ | Stram ²⁵ | Flagg ²⁶ | Patterson ²⁸ |
| Energy | 0.39 | 0.49 | 0.20 | 0.43 | 0.45 | 0.48 | 0.27 | 0.22 | 0.48 | 0.18 | 0.30 | 0.45 | 0.28 | 0.40 | 0.37 |
| Protein | | | | | | | | | | | | | | | |
| Unadjusted | 0.41 | 0.47 | 0.30 | | 0.47 | 0.51 | | 0.30 | 0.46 | 0.14 | | 0.43 | 0.38 | | |
| Adjusted | 0.43 | 0.57 | 0.58 | 0.44 | 0.61 | 0.44 | 0.29 | 0.50 | 0.60 | 0.54 | 0.16 | 0.53 | 0.67 | 0.33 | 0.51 |
| Carbohydrate | | | | | | | | | | | | | | | |
| Unadjusted | 0.39 | 0.53 | 0.40 | | 0.46 | 0.49 | | 0.32 | 0.50 | 0.33 | | 0.49 | 0.37 | | |
| Adjusted | 0.71 | 0.63 | 0.67 | 0.73 | 0.64 | 0.68 | 0.73 | 0.64 | 0.69 | 0.63 | 0.45 | 0.66 | 0.80 | 0.51 | 0.67 |
| Total fat | | | | | | | | | | | | | | | |
| Unadjusted | 0.58 | 0.52 | 0.25 | | 0.53 | 0.57 | | 0.39 | 0.55 | 0.30 | | 0.53 | 0.39 | | |
| Adjusted | 0.72 | 0.62 | 0.60 | 0.67 | 0.55 | 0.69 | 0.58 | 0.62 | 0.66 | 0.65 | 0.62 | 0.67 | 0.77 | 0.66 | 0.64 |
| Percentage energy from fat | | 0.66 | 0.65 | | 0.60 | 0.69 | 0.49 | | 0.67 | 0.64 | | 0.66 | 0.76 | 0.65 | |
| Saturated fat | | | | | | | | | | | | | | | |
| Unadjusted | 0.65 | 0.57 | 0.34 | | 0.58 | 0.61 | | 0.48 | 0.60 | 0.41 | | 0.56 | 0.45 | | |
| Adjusted | 0.76 | 0.68 | 0.66 | 0.75 | 0.67 | 0.71 | 0.64 | 0.69 | 0.66 | 0.66 | 0.59 | 0.65 | 0.76 | 0.66 | 0.63 |
| Monounsaturated fat | | | | | | | | | | | | | | | |
| Unadjusted | 0.58 | 0.51 | 0.27 | | 0.56 | | | 0.41 | 0.56 | 0.31 | | 0.55 | | | |
| Adjusted | 0.71 | 0.60 | 0.63 | 0.68 | 0.54 | | | 0.62 | 0.62 | 0.64 | 0.62 | 0.60 | | | 0.64 |
| Polyunsaturated fat | | | | | | | | | | | | | | | |
| Unadjusted | 0.47 | 0.52 | 0.24 | | 0.41 | | | 0.33 | 0.48 | 0.19 | | 0.43 | | | |
| Adjusted | 0.53 | 0.61 | 0.52 | 0.37 | 0.33 | | | 0.56 | 0.64 | 0.47 | 0.43 | 0.48 | | | 0.54 |
| Cholesterol | | | | | | | | | | | | | | | |
| Unadjusted | 0.67 | 0.48 | 0.52 | | 0.57 | | | 0.56 | 0.61 | 0.52 | | 0.55 | | | |
| Adjusted | 0.68 | 0.64 | 0.70 | 0.76 | 0.67 | | 0.58 | 0.68 | 0.66 | 0.68 | 0.21 | 0.57 | | 0.44 | 0.60 |
| Dietary fibre | | | | | | | | | | | | | | | |
| Unadjusted | 0.50 | 0.62 | 0.44 | | 0.62 | 0.59 | | 0.42 | 0.60 | 0.41 | | 0.62 | 0.46 | | |
| Adjusted | 0.72 | 0.80 | 0.73 | 0.68 | 0.77 | 0.71 | 0.64 | 0.66 | 0.77 | 0.68 | 0.24 | 0.80 | 0.73 | 0.61 | 0.70 |
| Vitamin A | | | | | | | | | | | | | | | |
| Unadjusted | 0.46 | 0.60 | 0.43 | | 0.59 | 0.59 | | 0.42 | 0.58 | 0.43 | | 0.47 | 0.58 | | |
| Adjusted | 0.59 | 0.69 | 0.51 | 0.61 | 0.65 | 0.64 | 0.45 | 0.59 | 0.62 | 0.54 | 0.14 | 0.50 | 0.77 | 0.26 | 0.30 |
| Vitamin E | | | | | | | | | | | | | | | |
| Unadjusted | 0.37 | 0.55 | 0.23 | | 0.39 | | | 0.23 | 0.43 | 0.17 | | 0.28 | | | |
| Adjusted | 0.40 | 0.57 | 0.48 | 0.42 | 0.24 | | 0.23 | 0.36 | 0.51 | 0.46 | 0.79 | 0.19 | | 0.27 | 0.29 |
| Vitamin C | | | | | | | | | | | | | | | |
| Unadjusted | 0.60 | 0.66 | 0.61 | | 0.60 | 0.63 | | 0.54 | 0.61 | 0.54 | | 0.65 | 0.66 | | |
| Adjusted | 0.70 | 0.74 | 0.78 | 0.77 | 0.70 | 0.71 | 0.65 | 0.65 | 0.68 | 0.71 | 0.53 | 0.72 | 0.73 | 0.65 | 0.37 |
| Thiamin | | | | | | | | | | | | | | | |
| Unadjusted | 0.38 | 0.61 | 0.54 | | 0.56 | | | 0.37 | 0.54 | 0.33 | | 0.45 | | | |
| Adjusted | 0.62 | 0.83 | 0.76 | | 0.72 | | | 0.59 | 0.65 | 0.57 | 0.68 | 0.57 | | | 0.66 |
| Riboflavin | | | | | | | | | | | | | | | |
| Unadjusted | 0.48 | 0.63 | 0.55 | | 0.60 | | | 0.44 | 0.58 | 0.49 | | 0.53 | | | |
| Adjusted | 0.64 | 0.78 | 0.79 | | 0.68 | | | 0.67 | 0.66 | 0.69 | 0.32 | 0.59 | | | 0.65 |
| Niacin | | | | | | | | | | | | | | | |
| Unadjusted | 0.40 | 0.55 | 0.40 | | 0.56 | | | 0.32 | 0.48 | 0.25 | | 0.34 | | | |
| Adjusted | 0.56 | 0.65 | 0.57 | | 0.59 | | | 0.57 | 0.64 | 0.53 | | 0.46 | | | 0.63 |

Table 4 Continued

| Nutrient/dietary constituent | Men | | | | | | | Women | | | | | | | |
|------------------------------|------|------------------------|--------------------|--------------------|--------------------|---------------------|---------------------|-------|------------------------|--------------------|----------------------|--------------------|---------------------|---------------------|-------------------------|
| | NCI | | Willett | | Block | Hawaii | ACS | NCI | | Willett | Block | Hawaii | ACS | WHI | |
| | AARP | EATS-DHQ ²¹ | EATS ²¹ | Rimm ²⁷ | EATS ²¹ | Stram ²⁵ | Flagg ²⁶ | AARP | EATS-DHQ ²¹ | EATS ²¹ | Munger ²⁴ | EATS ²¹ | Stram ²⁵ | Flagg ²⁶ | Patterson ²⁸ |
| Folate | | | | | | | | | | | | | | | |
| Unadjusted | 0.48 | | | | | | | 0.49 | | | | | | | |
| Adjusted | 0.64 | | | 0.70 | | | 0.51 | 0.69 | | | 0.26 | | | 0.43 | 0.57 |
| Vitamin B ₆ | | | | | | | | | | | | | | | |
| Unadjusted | 0.51 | 0.65 | 0.52 | | 0.63 | | | 0.46 | 0.54 | 0.33 | | 0.48 | | | |
| Adjusted | 0.69 | 0.79 | 0.70 | 0.73 | 0.64 | | | 0.70 | 0.65 | 0.61 | | 0.57 | | | 0.66 |
| Calcium | | | | | | | | | | | | | | | |
| Unadjusted | 0.54 | 0.69 | 0.57 | | 0.62 | 0.52 | | 0.48 | 0.66 | 0.55 | | 0.62 | 0.49 | | |
| Adjusted | 0.63 | 0.81 | 0.79 | 0.60 | 0.72 | 0.64 | 0.57 | 0.64 | 0.73 | 0.67 | 0.49 | 0.66 | 0.69 | 0.66 | 0.73 |
| Iron | | | | | | | | | | | | | | | |
| Unadjusted | 0.42 | 0.59 | 0.46 | | 0.55 | | | 0.36 | 0.49 | 0.30 | | 0.42 | | | |
| Adjusted | 0.59 | 0.71 | 0.67 | 0.32 | 0.71 | | 0.28 | 0.56 | 0.59 | 0.52 | -0.09 | 0.57 | | 0.35 | 0.66 |
| Magnesium | | | | | | | | | | | | | | | |
| Unadjusted | 0.45 | 0.61 | 0.43 | | 0.64 | | | 0.35 | 0.57 | 0.44 | | 0.60 | | | |
| Adjusted | 0.64 | 0.79 | 0.82 | 0.71 | 0.76 | | | 0.66 | 0.78 | 0.83 | | 0.81 | | | 0.72 |
| Phosphorus | | | | | | | | | | | | | | | |
| Unadjusted | 0.46 | 0.58 | 0.39 | | 0.56 | | | 0.36 | 0.55 | 0.33 | | 0.55 | | | |
| Adjusted | 0.61 | 0.73 | 0.77 | 0.63 | 0.68 | | | 0.63 | 0.69 | 0.65 | 0.52 | 0.62 | | | |
| Zinc | | | | | | | | | | | | | | | |
| Unadjusted | 0.45 | 0.42 | 0.37 | | 0.44 | | | 0.40 | 0.46 | 0.28 | | 0.48 | | | |
| Adjusted | 0.49 | 0.49 | 0.66 | 0.71 | 0.42 | | 0.14 | 0.57 | 0.51 | 0.43 | | 0.58 | | 0.18 | 0.47 |
| Potassium | | | | | | | | | | | | | | | |
| Unadjusted | 0.45 | 0.58 | 0.43 | | 0.60 | | | 0.39 | 0.59 | 0.38 | | 0.59 | | | |
| Adjusted | 0.63 | 0.76 | 0.79 | 0.73 | 0.72 | | | 0.67 | 0.76 | 0.75 | 0.65 | 0.75 | | | 0.67 |
| Vitamin B ₁₂ | | | | | | | | | | | | | | | |
| Unadjusted | 0.48 | | | | | | | 0.44 | | | | | | | |
| Adjusted | 0.47 | | | 0.52 | | | | 0.47 | | | 0.47 | | | | 0.21 |

NCI – National Cancer Institute; ACS – American Cancer Society; WHI – Women's Health Initiative; EATS – Eating at America's Table Study; DHQ – Diet History Questionnaire.

Stram *et al.*²⁵ used energy from non-alcohol sources only. Estimates in this table are those for whites only.

Flagg *et al.*²⁶ used retinol for vitamin A.

Patterson *et al.*²⁸ used retinol for vitamin A.

residual fat intake in women would correspond to a true attenuation factor of about 0.32. Without residual confounding from energy, a true relative risk of 1.8 between the medians of the first and fifth quintiles of the residual fat distribution would be observed as 1.21. With a projected 3700 cases of breast cancer after 5 years of follow-up¹, the statistical power to detect this relative risk using a two-sided test with a 5% significance level would be 98%.

While we do not have reference biomarkers for nutrients other than protein and energy, it seems likely that the 24HR-based estimates of the FFQ performance for some other nutrients are biased. Thus, we think it best to consider the estimated correlations with true intakes and attenuation factors in Table 3 as relative rather than absolute measures of the FFQ performance, and emphasise comparison with corresponding estimates in other studies. However, similar to protein in the OPEN study, the estimated correlation coefficients and attenuation factors improved substantially for most nutrients in Table 3 after energy adjustment. This improvement might be expected to hold for true correlations and attenuation factors.

While we do not know the precise direction and magnitude of biases in evaluating the FFQ using 24HR as a reference instrument, it is reasonable to conclude that reported energy intake may be most useful as an adjusting covariate for estimating intake of energy-adjusted nutrients. The FFQ's ability to capture energy-adjusted intake appears to be superior to its ability to capture absolute intake. Thus, energy adjustment not only allows one to assess the presence of an isoenergetic nutrient effect, but may also increase the statistical power to detect diet–disease relationships. Based on the estimated correlations with true intake and attenuation factors in Table 3, and even taking into account that some of them may overestimate the actual values, for energy-adjusted nutrients, the sample size of the NIH–AARP cohort study appears to be large enough and its range of intake wide enough to compensate for the loss of power due to measurement error in the FFQ and to detect moderate associations (≥ 1.8) between diet and common cancers.

Acknowledgements

Sources of funding: This research was supported (in part) by the Intramural Research Program of the NIH, NCI.

Conflict of interest declaration: None.

Authorship responsibilities: Study design – A.S., C.C.B., L.S.F., R.J.C., A.F.S., F.E.T.; study management – T.M., A.S., M.L.; analysis – V.K., D.M., C.C.B., F.E.T., M.S.B., L.S.F., R.J.C.; writing – F.E.T., V.K., D.M., A.F.S., A.S., L.S.F., R.J.C., M.L.

Acknowledgements: We are indebted to the participants in the NIH–AARP Diet and Health Study for their outstanding cooperation. In addition, the authors gratefully acknowledge the work of Westat, Inc. in conducting the NIH–AARP Diet and Health Study.

References

- Schatzkin A, Subar AF, Thompson FE, Harlan LC, Tangrea J, Hollenbeck AR, *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. *American Journal of Epidemiology* 2001; **154**: 1119–25.
- Plummer M, Clayton D. Measurement error in dietary assessment: an investigation using covariance structure models. Part I. *Statistics in Medicine* 1993; **12**: 925–35.
- Plummer M, Clayton D. Measurement error in dietary assessment: an investigation using covariance structure models. Part II. *Statistics in Medicine* 1993; **12**: 937–48.
- Freudenheim JL, Marshall JR. The problem of profound mismeasurement and the power of epidemiologic studies of diet and cancer. *Nutrition and Cancer* 1988; **11**: 243–50.
- Freedman LS, Schatzkin A, Wax Y. The impact of dietary measurement on planning sample size required in a cohort study. *American Journal of Epidemiology* 1990; **132**: 1185–95.
- US National Cancer Institute, Division of Cancer Epidemiology and Genetics. NIH–AARP Diet and Health Study [homepage]. <http://www.dietandhealth.cancer.gov>.
- US National Cancer Institute, Division of Cancer Control and Population Sciences. *Risk Factor Monitoring and Methods, Diet History Questionnaire* [online], 8 February 2007. Available at <http://riskfactor.cancer.gov/DHQ/>. Accessed June 2007.
- Subar AF, Midthune D, Kulldorff M, Brown CC, Thompson FE, Kipnis V, *et al.* Evaluation of alternative approaches to assign nutrient values to food groups in food frequency questionnaires. *American Journal of Epidemiology* 2000; **152**: 279–86.
- Walker AM, Blettner M. Comparing imperfect measures of exposure. *American Journal of Epidemiology* 1985; **121**: 783–90.
- Kipnis V, Midthune D, Freedman LS, Bingham S, Schatzkin A, Subar A, *et al.* Empirical evidence of correlated biases in dietary assessment instruments and its implications. *American Journal of Epidemiology* 2001; **153**: 394–403.
- Rosner B, Willett WC. Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. *American Journal of Epidemiology* 1988; **127**: 377–86.
- Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine* 1989; **8**: 1051–69.
- Spiegelman D, Schneeweiss S, McDermott A. Measurement error correction for logistic regression models with an 'alloyed gold standard'. *American Journal of Epidemiology* 1997; **145**: 184–96.
- Wong MY, Day NE, Bashir SA, Duffy SW. Measurement error in epidemiology: the design of validation studies I: univariate situation. *Statistics in Medicine* 1999; **18**: 2815–29.
- Wong MY, Day NE, Wareham NJ. Measurement error in epidemiology: the design of validation studies II: bivariate situation. *Statistics in Medicine* 1999; **18**: 2831–45.
- Box G, Cox D. An analysis of transformations. *Journal of the Royal Statistical Society* 1964; **B26**: 211–52.
- Willett WC. *Nutritional Epidemiology*, 2nd ed. New York: Oxford University Press, 1998.
- Carroll RJ, Midthune D, Freedman LS, Kipnis V. Seemingly unrelated measurement error models with application to nutritional epidemiology. *Biometrics* 2006; **62**: 75–84.
- SAS Institute, Inc. *SAS/STAT User's Guide*. Cary, NC: SAS Institute, Inc., 2000.
- Carroll RJ, Freedman L, Pee D. Design aspects of calibration studies in nutrition, with analysis of missing data in linear measurement error models. *Biometrics* 1997; **53**: 1440–57.

- 21 Subar AF, Thompson FE, Kipnis V, Midthune D, Hurwitz P, McNutt S, *et al.* Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: the Eating at America's Table Study. *American Journal of Epidemiology* 2001; **154**: 1089–99.
- 22 Subar AF, Kipnis V, Troiano RP, Midthune D, Schoeller DA, Bingham S, *et al.* Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN Study. *American Journal of Epidemiology* 2003; **158**: 1–13.
- 23 Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano RP, *et al.* Structure of dietary measurement error: results of the OPEN Biomarker Study. *American Journal of Epidemiology* 2003; **158**: 14–21.
- 24 Munger RG, Folsom AR, Kushi LH, Kaye SA, Sellers TA. Dietary assessment of older Iowa women with a food frequency questionnaire: nutrient intake, reproducibility, and comparison with 24-hour dietary recall interviews. *American Journal of Epidemiology* 1992; **136**: 192–200.
- 25 Stram DO, Hankin JH, Wilkens LR, Pike MC, Monroe KR, Park S, *et al.* Calibration of the dietary questionnaire for a multiethnic cohort in Hawaii and Los Angeles. *American Journal of Epidemiology* 2000; **151**: 358–70.
- 26 Flagg EW, Coates RJ, Calle EE, Potischman N, Thun MJ. Validation of the American Cancer Society Cancer Prevention Study II Nutrition Survey Cohort Food Frequency Questionnaire. *Epidemiology* 2000; **11**: 462–8.
- 27 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. *American Journal of Epidemiology* 1992; **135**: 1114–26.
- 28 Patterson RE, Kristal AR, Tinker LF, Carter RA, Bolton MP, Agurs-Collins T. Measurement characteristics of the Women's Health Initiative food frequency questionnaire. *Annals of Epidemiology* 1999; **9**: 178–87.
- 29 Plummer M, Kaaks R. Commentary: An OPEN assessment of dietary measurement error. *International Journal of Epidemiology* 2003; **32**: 1062–3.
- 30 Kipnis V, Midthune D, Freedman L, Bingham S, Day NE, Riboli E, *et al.* Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public Health Nutrition* 2002; **5**: 915–23.
- 31 Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. *American Journal of Epidemiology* 1986; **124**: 17–27.
- 32 Willett W. Dietary diaries versus food frequency questionnaires – a case of undigestible data. *International Journal of Epidemiology* 2001; **30**: 317–19.
- 33 Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika* 1965; **52**: 591–611.
- 34 Shapiro SS, Francia RS. An approximate analysis of variance test for normality. *Journal of the American Statistical Association* 1972; **67**: 215–16.

Appendix A

Prior to modelling, we performed the following steps for each nutrient and for each gender, in order to appropriately transform the data and remove outliers:

1. For each of the two 24HRs, temporarily remove values that fall below the 25th percentile of the distribution of reported intake minus three interquartile ranges or above the 75th percentile plus three interquartile ranges.
2. Find the Box–Cox power transformation¹⁶ of the remaining 24HR values that maximises the Shapiro–Wilk test statistic for normality^{33,34} for the average of the two transformed 24HRs.
3. Apply this transformation to the values of each 24HR and each FFQ, after adding back the values removed in step (1).
4. For each 24HR and each FFQ, remove as outliers the transformed values that fall below the 25th percentile of the distribution of transformed reported intake minus two interquartile ranges or above the 75th percentile plus two interquartile ranges.

The correlations with true intake and attenuation factors were estimated for each nutrient on their respective transformed scales. Because the Box–Cox transformation is monotonic, adjustment for energy does not depend on the scale to which energy is transformed. Therefore, the correlations with true intake and attenuation factors for energy-adjusted nutrients should be inter-

preted on the scale to which each particular nutrient was transformed.

Appendix B – Measurement error model

For a given person, let T_N and T_E denote true intakes of nutrient N of interest and total energy E , respectively. Let Q_{Lj} and F_{Lj} denote the FFQ- and 24HR-reported intakes of nutrient $L = N, E$ on occasion $j = 1, 2$, respectively. On appropriately transformed scales where all random variables are (approximately) normally distributed, the simultaneous measurement error model¹⁸ for nutrients N and E is given as

$$\begin{aligned}
 Q_{Nj} &= \beta_{N0j} + \beta_{N1} T_N + r_N + e_{Nj}; & j = 1, 2 \\
 Q_{Ej} &= \beta_{E0j} + \beta_{E1} T_E + r_E + e_{Ej}; & j = 1, 2 \\
 F_{Nj} &= \mu_{N0j} + T_N + u_{Nj}; & j = 1, 2 \\
 F_{Ej} &= \mu_{E0j} + T_E + u_{Ej}; & j = 1, 2.
 \end{aligned} \tag{B1}$$

Among the random variables, T_L , $L = N, E$, has mean μ_{T_L} and variance $\sigma_{T_L}^2$. Parameters β_{L0j} and μ_{L0j} , $L = N, E$ and $j = 1, 2$, denote overall biases independent of true intake for the FFQ and 24HR, respectively; for the same nutrient, they are allowed to have different values to reflect the tendency to report lower mean intakes on repeat administrations of the same instrument. It is assumed that the first 24HR is unbiased, so that $\mu_{N0_1} = \mu_{E0_1} = 0$. Slopes β_{N1} and β_{E1} reflect intake-related bias

in the FFQ; their values smaller than 1 lead to the well-known flattened slope phenomenon, when people with lower than average intake overreport it and people with higher than average intake underreport it. Random variables r_N and r_E with means of zero and variances $\sigma_{r_N}^2$ and $\sigma_{r_E}^2$, respectively, denote person-specific biases in the FFQ that are independent of true intake (but not of each other) and represent the differences between total within-person biases and their intake-related components. For nutrient $L = N, E$ and repeat administration $j = 1, 2$, random variables $\{e_{Lj}\}$ and $\{u_{Lj}\}$ denote within-person random errors in the FFQ and 24HR, respectively, which are assumed to be independent of true intakes and person-specific biases. Because the FFQs and 24HRs were administered well separated in time from each other and between their respective repeats, within-person random errors for two different instruments and two different administrations for the same instrument are assumed to be independent. However, for the same administration of each instrument, within-person random errors for nutrients N and E are allowed to be correlated. It is critical that, under this model, the 24HR is assumed a correct reference instrument with errors that are independent of true intakes, errors in the FFQ and, for repeat administrations, of each other.

Since we are mainly interested in estimating the FFQ correlations with true intake and attenuation factors, it is convenient to re-parameterise measurement error model (B1) in terms of the first and second moments of the joint distribution of the intakes reported on two FFQs and two 24HRs. Define the observed data vector by $\mathbf{D} = (Q_{N1}, Q_{N2}, Q_{E1}, Q_{E2}, F_{N1}, F_{N2}, F_{E1}, F_{E2})^T$. Denote its mean vector by $\boldsymbol{\mu} = (\mu_1, \dots, \mu_8)$ and its variance covariance matrix by $\boldsymbol{\Sigma}$. Note that, according to the model assumptions, $\mu_5 = E(F_{N1}) = \mu_{T_N}$ and $\mu_7 = E(F_{E1}) = \mu_{T_E}$. Define

$$\begin{aligned} \theta_1 &= \text{var}(Q_{Nj}); \quad j = 1, 2 \\ \theta_2 &= \text{cov}(Q_{N1}, Q_{N2}) \\ \theta_3 &= \text{cov}(Q_{N1}, Q_{E1}) = \text{cov}(Q_{N2}, Q_{E2}) \\ \theta_4 &= \text{cov}(Q_{N1}, Q_{E2}) = \text{cov}(Q_{N2}, Q_{E1}) \\ \theta_5 &= \text{cov}(Q_{Nj}, F_{Nj}) = \text{cov}(Q_N, T_N); \quad l, j = 1, 2 \\ \theta_6 &= \text{cov}(Q_{Nj}, F_{Ej}) = \text{cov}(Q_N, T_E); \quad l, j = 1, 2 \\ \theta_7 &= \text{var}(Q_{Ej}); \quad j = 1, 2 \\ \theta_8 &= \text{cov}(Q_{E1}, Q_{E2}) \\ \theta_9 &= \text{cov}(Q_{Ej}, F_{Nj}) = \text{cov}(Q_E, T_N); \quad l, j = 1, 2 \\ \theta_{10} &= \text{cov}(Q_{Ej}, F_{Ej}) = \text{cov}(Q_E, T_E); \quad l, j = 1, 2 \\ \theta_{11} &= \text{var}(F_{Nj}); \quad j = 1, 2 \\ \theta_{12} &= \text{cov}(F_{N1}, F_{N2}) = \sigma_{T_N}^2 \\ \theta_{13} &= \text{cov}(F_{Nj}, F_{Ej}) = \text{cov}(T_N, T_E); \quad l, j = 1, 2 \\ \theta_{14} &= \text{var}(F_{Ej}); \quad j = 1, 2 \\ \theta_{15} &= \text{cov}(F_{E1}, F_{E2}) = \sigma_{T_E}^2. \end{aligned} \quad (\text{B2})$$

From model (B1), it follows that the variance-covariance matrix of vector \mathbf{D} is then given by

$$\boldsymbol{\Sigma} = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 & \theta_5 & \theta_6 & \theta_6 \\ \theta_2 & \theta_1 & \theta_4 & \theta_3 & \theta_5 & \theta_5 & \theta_6 & \theta_6 \\ \theta_3 & \theta_4 & \theta_7 & \theta_8 & \theta_9 & \theta_9 & \theta_{10} & \theta_{10} \\ \theta_4 & \theta_3 & \theta_8 & \theta_7 & \theta_9 & \theta_9 & \theta_{10} & \theta_{10} \\ \theta_5 & \theta_5 & \theta_9 & \theta_9 & \theta_{11} & \theta_{12} & \theta_{13} & \theta_{13} \\ \theta_5 & \theta_5 & \theta_9 & \theta_9 & \theta_{12} & \theta_{11} & \theta_{13} & \theta_{13} \\ \theta_6 & \theta_6 & \theta_{10} & \theta_{10} & \theta_{13} & \theta_{13} & \theta_{14} & \theta_{15} \\ \theta_6 & \theta_6 & \theta_{10} & \theta_{10} & \theta_{13} & \theta_{13} & \theta_{15} & \theta_{14} \end{pmatrix},$$

so that the observed data follow the model

$$\mathbf{D} \sim \text{Normal}[\boldsymbol{\mu}; \boldsymbol{\Sigma}]. \quad (\text{B3})$$

The maximum likelihood estimates for the unknown first and second moments are computed by fitting model (B3) using the SAS Mixed procedure.

For a given nutrient L , $L = N, E$, the correlation with true intake and attenuation factor are given by

$$\rho_{Q,T} = \frac{\text{cov}(Q_L, T_L)}{\sigma_{Q_L} \sigma_{T_L}}$$

and

$$\lambda_{Q_L} = \frac{\text{cov}(Q_L, T_L)}{\sigma_{Q_L}^2},$$

respectively. Under our assumptions, it follows from (B2) that

$$\text{var}(Q_N) = \theta_1; \quad \text{var}(Q_E) = \theta_7; \quad \sigma_{T_N}^2 = \theta_{12}; \quad \sigma_{T_E}^2 = \theta_{15}$$

so that the correlation with true intake and attenuation factor for nutrients N and E are estimated as

$$\hat{\rho}_{Q_N, T_N} = \frac{\hat{\theta}_5}{\sqrt{\hat{\theta}_1 \hat{\theta}_{12}}}; \quad \hat{\lambda}_{Q_N} = \frac{\hat{\theta}_5}{\hat{\theta}_1}$$

and

$$\hat{\rho}_{Q_E, T_E} = \frac{\hat{\theta}_{10}}{\sqrt{\hat{\theta}_7 \hat{\theta}_{15}}}; \quad \hat{\lambda}_{Q_E} = \frac{\hat{\theta}_{10}}{\hat{\theta}_7},$$

respectively.

For nutrient residual R , we have the following respective expressions for the true residual and the residual calculated using the FFQ:

$$T_R = (T_N - \mu_{T_N}) - \frac{\text{cov}(T_N, T_E)}{\sigma_{T_E}^2} (T_E - \mu_{T_E})$$

and

$$Q_{Rj} = (Q_{Nj} - \mu_{Q_{Nj}}) - \frac{\text{cov}(Q_{Nj}, Q_{Ej})}{\text{var}(Q_E)} (Q_{Ej} - \mu_{Q_{Ej}}),$$

with

$$\text{var}(T_R) = \sigma_{T_N}^2 - \frac{\text{cov}^2(T_N, T_E)}{\sigma_{T_E}^2},$$

$$\text{var}(Q_{Rj}) = \sigma_{Q_N}^2 - \frac{\text{cov}^2(Q_{Nj}, Q_{Ej})}{\text{var}(Q_{Ej})}$$

and

$$\begin{aligned} \text{cov}(Q_{Rj}, T_R) = & \text{cov}(Q_{Nj}, T_N) - \frac{\text{cov}(Q_{Nj}, Q_{Ej})\text{cov}(Q_{Ej}, T_N)}{\text{var}(Q_{Ej})} \\ & - \frac{\text{cov}(T_N, T_E)\text{cov}(Q_{Nj}, T_E)}{\sigma_{T_E}^2} \\ & + \frac{\text{cov}(Q_{Nj}, Q_{Ej})\text{cov}(T_N, T_E)\text{cov}(Q_{Ej}, T_E)}{\text{var}(Q_{Ej})\sigma_{T_E}^2}; \\ & j = 1, 2. \end{aligned}$$

From formulas (B2) it follows that, for the nutrient residual, the correlation with true intake and attenuation

factor are estimated as

$$\hat{\rho}_{Q_R, T_R} = \frac{\hat{\theta}_5\hat{\theta}_7\hat{\theta}_{15} - \hat{\theta}_3\hat{\theta}_9\hat{\theta}_{15} - \hat{\theta}_6\hat{\theta}_7\hat{\theta}_{13} + \hat{\theta}_1\hat{\theta}_3\hat{\theta}_{10}}{\sqrt{\hat{\theta}_7\hat{\theta}_{15}(\hat{\theta}_{12}\hat{\theta}_{15} - \hat{\theta}_{13}^2)}}$$

and

$$\hat{\lambda}_{Q_R} = \frac{\hat{\theta}_5\hat{\theta}_7\hat{\theta}_{15} - \hat{\theta}_3\hat{\theta}_9\hat{\theta}_{15} - \hat{\theta}_6\hat{\theta}_7\hat{\theta}_{13} + \hat{\theta}_1\hat{\theta}_3\hat{\theta}_{10}}{\hat{\theta}_{12}\hat{\theta}_{15} - \hat{\theta}_{13}^2},$$

respectively.