

# Validating an FFQ for intake of episodically consumed foods: application to the National Institutes of Health–AARP Diet and Health Study

Douglas Midthune<sup>1,\*</sup>, Arthur Schatzkin<sup>2</sup>, Amy F Subar<sup>3</sup>, Frances E Thompson<sup>3</sup>, Laurence S Freedman<sup>4</sup>, Raymond J Carroll<sup>5</sup>, Marina A Shumakovich<sup>1</sup> and Victor Kipnis<sup>1</sup>

<sup>1</sup>Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, Executive Plaza North – Suite 3131, 6130 Executive Blvd – MSC 7354, Bethesda, MD 20892-7354, USA; <sup>2</sup>Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA;

<sup>3</sup>Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD, USA; <sup>4</sup>Gertner Institute for Epidemiology and Health Policy Research, Sheba Medical Center, Tel Hashomer, Israel; <sup>5</sup>Department of Statistics, Texas A&M University, College Station, TX, USA

Submitted 10 October 2010: Accepted 20 February 2011: First published online 13 April 2011

## Abstract

**Objective:** To develop a method to validate an FFQ for reported intake of episodically consumed foods when the reference instrument measures short-term intake, and to apply the method in a large prospective cohort.

**Design:** The FFQ was evaluated in a sub-study of cohort participants who, in addition to the questionnaire, were asked to complete two non-consecutive 24 h dietary recalls (24HR). FFQ-reported intakes of twenty-nine food groups were analysed using a two-part measurement error model that allows for non-consumption on a given day, using 24HR as a reference instrument under the assumption that 24HR is unbiased for true intake at the individual level.

**Setting:** The National Institutes of Health–AARP Diet and Health Study, a cohort of 567 169 participants living in the USA and aged 50–71 years at baseline in 1995.

**Subjects:** A sub-study of the cohort consisting of 2055 participants.

**Results:** Estimated correlations of true and FFQ-reported energy-adjusted intakes were 0·5 or greater for most of the twenty-nine food groups evaluated, and estimated attenuation factors (a measure of bias in estimated diet–disease associations) were 0·4 or greater for most food groups.

**Conclusions:** The proposed methodology extends the class of foods and nutrients for which an FFQ can be evaluated in studies with short-term reference instruments. Although violations of the assumption that the 24HR is unbiased could be inflating some of the observed correlations and attenuation factors, results suggest that the FFQ is suitable for testing many, but not all, diet–disease hypotheses in a cohort of this size.

## Keywords

Diet  
Food  
Epidemiological methods  
Questionnaires  
Validation studies

Most large prospective cohorts use an FFQ to measure dietary intake. It is well known that an FFQ has substantial measurement error that can affect the results of such studies, leading to bias and the loss of power to detect diet–disease relationships<sup>(1,2)</sup>. In order to evaluate the measurement error in an FFQ, and to correct observed diet–disease relationships for bias due to measurement error, many cohort studies include calibration sub-studies in which another, less biased, dietary instrument is administered as a reference instrument. The reference instrument is usually a short-term instrument such as a 24 h dietary recall (24HR) or food record.

Methods for evaluating an FFQ's ability to measure foods/nutrients that are consumed daily have been

developed based on measurement error models that explicitly or implicitly assume that true usual intake and reported intake from the FFQ and reference instrument are all continuous variables<sup>(3–5)</sup>. These methods have sometimes been used to evaluate 'episodically consumed' foods, or foods that are not consumed nearly every day by almost everyone in the population<sup>(6–8)</sup>. This can be problematic if the reference instrument covers only a short time period, since short-term instruments may have a substantial proportion of subjects reporting zero intake of an episodically consumed food, violating the assumption that the reported intake is continuous.

Recently, a measurement error model for episodically consumed foods has been developed and used in dietary

\*Corresponding author: Email midthund@mail.nih.gov

surveillance to estimate population distributions of usual intakes of such foods<sup>(9–11)</sup> and to correct for measurement error in diet–health relationships when the 24HR is the main dietary instrument<sup>(12)</sup>. The model allows for non-consumption on a given day by separating the probability to consume from the amount consumed on a consumption day using a two-part model<sup>(9)</sup>. The model has also been extended to a ‘three-part’ model to estimate the joint distribution of intakes of an episodically consumed food and energy<sup>(10)</sup>. In the present paper, we use these models to evaluate an FFQ’s ability to measure intake of episodically consumed foods when the reference instrument measures short-term intake. After fitting a model that describes the relationship between the short-term reference and the FFQ, we use Monte Carlo methods to estimate the relationship between true and FFQ-reported intakes.

In 1995, the National Institutes of Health (NIH) and the AARP, formerly the American Association of Retired Persons, initiated a large prospective cohort study called the NIH–AARP Diet and Health Study, which was designed to study relationships between diet and cancer. The study uses an FFQ to measure diet and includes a calibration sub-study of about 2000 subjects who in addition to the FFQ were administered two 24HR. Thompson *et al.*<sup>(13)</sup> evaluated the ability of the NIH–AARP FFQ to measure nutrient intake. In the present paper we assess the FFQ’s ability to measure intakes of twenty-nine food groups.

## Methods

### Study design

The design of the NIH–AARP Diet and Health Study is described in detail elsewhere<sup>(14)</sup>. Briefly, a baseline questionnaire that included a 124-item FFQ was mailed to 3.5 million members of AARP in 1995–1996. A total of 617 119 men and women returned the questionnaire, and, after excluding some whose questionnaires were deemed to be of poor quality or who declined to participate, a cohort of 567 169 subjects was established. Age at baseline in the cohort ranged from 50 to 71 years.

Calibration sub-study participants were selected from the 46 970 subjects who had returned questionnaires as of January 1996. Subjects in the sub-study were asked to complete two non-consecutive unannounced 24HR administered over the telephone by trained interviewers. Of the 2795 individuals invited to participate in the sub-study, 2055 agreed and completed at least one 24HR (97% completed both). The two 24HR were separated in time, with 50% separated by at least 21 days and 75% separated by at least 14 days. In our analysis, we include 1942 subjects (984 men, 958 women) after excluding 113 subjects who subsequently dropped out of the cohort study, had pre-baseline reports of cancer or death-only reports of cancer.

### Study instruments

The FFQ used in the NIH–AARP study was an early version of the Diet History Questionnaire (DHQ) developed at the National Cancer Institute (NCI)<sup>(15)</sup>. Frequency responses were asked for 124 food items; portion sizes for 116. An additional twenty-one questions asked about specific food choices and cooking practices. Databases from the US Department of Agriculture’s (USDA) Continuing Survey of Food Intakes of Individuals (CSFII) (1989–91, 1994–96) were used to develop a nutrient composition database for the FFQ<sup>(16)</sup>. The MyPyramid Equivalents Database (MPED) version 1.0, developed by USDA<sup>(17)</sup>, was used to obtain food group intakes in MPED servings consistent with 2005 Dietary Guidelines for Americans<sup>(18)</sup>. The MPED disaggregates components of food mixtures into food groups (e.g. pepperoni pizza components are placed into grain, dairy, vegetable and meat 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 standardized 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 when extreme values were due to coding errors.

### Statistical analysis

We evaluate the FFQ in terms of its ability to detect diet–disease relationships in observational studies. Two important parameters for characterizing this ability are the correlation of true and FFQ-reported intakes and the attenuation factor. The correlation of true and FFQ-reported intakes is a measure of the statistical power to detect diet–disease relationships, while the attenuation factor for FFQ-reported intake is a measure of the bias in estimated relationships. Both parameters are functions of the joint distribution of true and FFQ-reported intakes. Although one cannot observe true usual intake in free-living populations, one can estimate its distribution and its relationship to the FFQ-reported intake using statistical models and appropriate reference instruments.

#### Statistical model for episodically consumed foods

The model for episodically consumed foods is described in detail in Kipnis *et al.*<sup>(12)</sup>, who use the model to correct for measurement error when 24HR is the main dietary instrument. In the present application, FFQ is the main instrument and 24HR is used as a reference instrument.

For individual  $i$ ,  $i = 1, \dots, n$ , let

$T_{ij}$  be the true intake of an episodically consumed food on day  $j$

$p_i = P(T_{ij} > 0 | i)$  be the true probability to consume on a given day

$A_i = E(T_{ij}|T_{ij} > 0, i)$  be the true average amount consumed on a consumption day

$T_i = E(T_{ij}|i) = p_i \times A_i$  be the true usual intake of the episodically consumed food

$R_{ij}$  be the 24HR-reported intake of the episodically consumed food on day  $j$

$Q_i$  be the FFQ-reported intake of the episodically consumed food.

We assume that an individual's 24HR-reported intake  $R_{ij}$  is an unbiased estimate of true usual intake  $T_i$ . In particular, we assume that the probability to report consumption is equal to the true probability to consume,  $p_b$ , and that the average reported amount on a consumption day is equal to the true average amount consumed on a consumption day,  $A_i$ . Then the mean of  $R_{ij}$  equals  $p_i \times A_i = T_i$ . We note that this is a strong assumption that may not be exactly true, although it is generally believed that a 24HR is less biased than an FFQ (see Discussion section for more on this).

We also assume that, after appropriate transformations, the relationship between the FFQ-reported intake and the probability to consume can be described by a logistic regression model and that the relationship between the FFQ-reported intake and the amount consumed on a consumption day can be described by a linear regression model. The resulting two-part model can be written as:

$$\text{logit}(p_i) = \beta_{10} + \beta_{11} \times Q_i^* + U_{1i} \tag{1}$$

and

$$(R_{ij}^*|R_{ij} > 0) = \beta_{20} + \beta_{21} \times Q_i^* + U_{2i} + \varepsilon_{2ij}, \tag{2}$$

where  $\beta_{k0}$  and  $\beta_{k1}$  are the intercept and slope in the logistic or linear regression;  $U_{1i}$  and  $U_{2i}$  are person-specific random effects that have a bivariate normal distribution with mean zero, variances  $\sigma_{U_1}^2$  and  $\sigma_{U_2}^2$ , and correlation  $\rho_{U_1, U_2}$ ; and  $\varepsilon_{2ij}$  is within-person random error that is normally distributed with mean zero and variance  $\sigma_{\varepsilon_2}^2$ ; and  $\varepsilon_{2ij}$  is independent of  $(U_{1i}, U_{2i})$ . We include random effects  $U_{1i}$  and  $U_{2i}$  to allow for individual variations in probability and amount that are not explained by the FFQ. Variables  $Q_i^*$  and  $R_{ij}^*$  are Box-Cox transformations of  $Q_i$  and  $R_{ij}$  to scales on which they are approximately normal<sup>(19)</sup> (see Appendix 1 for details).

Equations (1) and (2) define a non-linear mixed-effects model that can be fit using the NLMIXED procedure in SAS to obtain maximum likelihood estimates of the model parameters  $\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \sigma_{U_1}^2, \sigma_{U_2}^2, \rho_{U_1, U_2}$  and  $\sigma_{\varepsilon_2}^2$ . For foods that are consumed every day, the model simplifies to:

$$R_{ij}^* = \beta_{30} + \beta_{31} \times Q_i^* + U_{3i} + \varepsilon_{3ij}. \tag{3}$$

Under the model assumptions, true usual intake  $T_i$  can be written as a function of  $Q_i, U_{1i}$  and  $U_{2i}$  (or  $U_{3i}$ ), and one can estimate relationships between true and FFQ-reported intakes by generating a Monte Carlo distribution of  $T_i$  and  $Q_i$  (see Appendix 2 for details).

Note that under the model for episodically consumed foods, intake on a given day ( $T_{ij}$  and  $R_{ij}$ ) can be zero, but usual intake ( $T_i$ ) is assumed to be greater than zero

(although it can be arbitrarily small). There may be foods which some people never consume (e.g. alcohol). Kipnis *et al.*<sup>(12)</sup> describe an extension of the present model that allows  $T_i$  to be zero; with only two 24HR per person, however, it is difficult in practice to distinguish never consumers from infrequent consumers.

A SAS macro that calls the NLMIXED procedure to fit the model for episodically consumed foods (equations (1) and (2)) or foods consumed every day (equation (3)) is available online<sup>(20)</sup>. Prior to fitting the model, we removed outliers of  $Q_i^*$  and positive  $R_{ij}^*$  for each food group, where outliers were defined to be values that fell below the 25th percentile of the distribution of the variable minus two interquartile ranges or above the 75th percentile plus two interquartile ranges. The average number of outliers removed for  $Q_i^*$  was 2 (men) and 4 (women), and the average number removed for  $R_{ij}^*$  was 4 (men) and 3 (women).

*Correlation with true intake and attenuation factor*

The attenuation factor and correlation with true intake are measures of the bias and loss of power in diet-disease studies due to measurement error in the FFQ. We assume that measurement error is non-differential with respect to disease; that is, that reported intake  $Q_i$  contributes no additional information about disease risk beyond that provided by true intake  $T_i$ . Suppose the true diet-disease relationship follows a logistic model:

$$\text{logit}(r_i) = \alpha_0 + \alpha_1 \times T_i^*, \tag{4}$$

where  $r_i$  is the probability of disease given true usual intake  $T_i$ ,  $\alpha_0$  and  $\alpha_1$  are the intercept and slope in the logistic regression, and  $T_i^*$  is a Box-Cox transformation of  $T_i$  to a scale on which it is approximately normal. The logistic regression model does not require covariates to have any particular distribution. In practice, however, covariates with skewed distributions are often transformed to make extreme values less influential.

We want to estimate the bias in the estimation of log odds ratio  $\alpha_1$  caused by using reported intake  $Q_i^*$  rather than  $T_i^*$  in equation (4). Since  $T_i^*$  and  $Q_i^*$  are transformed using different Box-Cox transformations, the interpretation of  $\alpha_1$  depends on which variable is in the model. In order to make the interpretations comparable, we first standardize the transformed variables so that a unit change equals the change from the 10th to the 90th percentile of true intake  $T_i$  on that scale. We can then interpret  $\alpha_1$  as the log odds ratio comparing the 90th and 10th percentiles of true intake.

To a close approximation, fitting equation (4) using  $Q_i^*$  rather than  $T_i^*$  leads to estimating not the true risk parameter  $\alpha_1$  but the product  $\tilde{\alpha}_1 = \gamma_1 \alpha_1$ , where  $\gamma_1$  is the slope in the linear regression of  $T_i^*$  v.  $Q_i^{*(21)}$ . The value  $\gamma_1$  is called the attenuation factor and is interpreted as the multiplicative bias in estimating log odds ratio  $\alpha_1$  due to measurement error in  $Q_i$ .

The loss of statistical power due to using  $Q_i^*$  rather than  $T_i^*$  in equation (4) is related to the correlation between  $T_i^*$  and  $Q_i^*$ , which we will call  $\rho_{TQ}$ . If a study would need a sample size of  $n$  to attain a desired power using  $T_i^*$  to measure intake, then the study would need a sample size of  $\tilde{n} = n/\rho_{TQ}^2$  to attain the same power using  $Q_i^*$ <sup>(22)</sup>. For both the correlation with true intake and the attenuation factor, one represents the ideal value. A correlation of one means no loss of power, while an attenuation factor of one means no bias in estimated risk. In a univariate diet–disease model, the attenuation factor is usually between zero and one, indicating that the estimated log odds ratio is biased towards zero, or attenuated.

One can estimate  $\gamma_1$  and  $\rho_{TQ}$  by generating a Monte Carlo distribution of  $T_i^*$  and  $Q_i^*$ , based on the models described in the previous section. Under the model assumptions, the Monte Carlo distribution will be approximately the same as the distribution in the real population, so that estimates based on the Monte Carlo distribution will be approximately unbiased (see Appendix 2 for details).

#### Energy-adjusted intake

Researchers are often interested in ‘energy-adjusted’ diet–disease relationships; that is, relationships between food intake and disease when total energy intake is held constant<sup>(23)</sup>. One popular energy-adjustment method is the ‘residual’ method, in which one first calculates the residual in the regression of food  $v$ . energy intake (after transforming both to approximate normality) and then relates residual intake to disease<sup>(23)</sup>. For simplicity, we refer to residual intake as ‘energy-adjusted’ intake.

To evaluate FFQ-reported energy-adjusted intake, we fit the three-part food and energy model described in Freedman *et al.*<sup>(10)</sup> and generate Monte Carlo distributions of true and FFQ-reported food and energy intakes. We then calculate true and reported residual intakes from the Monte Carlo distributions and use them to estimate the correlation with truth and the attenuation factor for residual intake (see Appendix 3 for details).

## Results

Table 1 shows the percentage of subjects in the calibration sub-study having zero intake on the 24HR or FFQ for thirty-two food groups. The food groups range from those that are rarely consumed to those that are consumed almost every day. For example, 98% of men and women reported zero intake of organ meat on both 24HR, while 99% reported non-zero intake of total grains on both 24HR.

Table 2 presents sample means for reported intakes of the thirty-two food groups. The means include both zero and non-zero amounts. In men, FFQ-reported intake tended to be less than 24HR-reported intake, while in women it tended to be greater. For men, the FFQ mean was at least 20% smaller than the 24HR mean for twelve

food groups, and at least 20% larger for six food groups. For women, the FFQ mean was at least 20% smaller than the 24HR mean for five food groups, and at least 20% larger for ten food groups.

Table 3 presents estimated correlations of true and FFQ-reported intakes and attenuation factors for twenty-nine food groups. Three food groups (yoghurt, organ meat, soya) are not included because they are too rarely consumed to obtain stable estimates. Results are presented for both unadjusted and energy-adjusted (residual) intakes. For the five most commonly consumed food groups (non-whole grains, total grains, total vegetables, added sugars, discretionary fat (solid)), estimates were obtained using the method for foods consumed every day, described in the Methods section. For the rest of the food groups, estimates were obtained using the method for episodically consumed foods. After energy adjustment, most food groups had correlations with true intake greater than 0.5; the food groups with lowest correlations after energy adjustment were legumes (0.34 for women), potatoes (0.35 for women), discretionary fat (oil) (0.38 for women, 0.43 for men) and low omega fish (0.42 for men). Attenuation factors were generally greater than 0.4, although several food groups had lower values; the food groups with lowest attenuation factors after energy adjustment were discretionary fat (oil) (0.18 for women), potatoes (0.23 for women), legumes (0.28 for women) and other starchy vegetables (0.29 for women).

Table 4 shows the number of incident cancers in the NIH–AARP cohort by gender and cancer type during the follow-up period, 1995 to 2003<sup>(24)</sup>. Table 4 also shows for each cancer type the study’s power to detect an odds ratio of 1.5 using FFQ-reported intake if  $\rho_{TQ} = 1$  (no loss of power due to measurement error) and if  $\rho_{TQ} = 0.5$ . The odds ratio compares the 90th to 10th percentile of true intake in a univariate diet–disease model (see Appendix 4 for details). For common cancer types such as prostate, breast, lung and colorectal, the power to detect the association is at least 85% when  $\rho_{TQ} = 0.5$ . For less common types such as myeloid leukaemia, thyroid and liver, the power is less than 30%.

## Discussion

We have proposed a methodology to evaluate an FFQ’s ability to measure intake of episodically consumed foods and used it to evaluate the FFQ in the NIH–AARP Diet and Health study. The methodology uses a two-part model designed for such foods<sup>(9,12)</sup> and Monte Carlo methods to estimate the relationship between true and FFQ-reported intakes. In order to evaluate energy-adjusted intake of such foods, we use a three-part food and energy model<sup>(10)</sup>.

The model for episodically consumed foods is designed for studies in which the reference instrument covers only a short time period and the probability of zero intake is substantial. In the NIH–AARP study, the reference

**Table 1** Percentage of subjects having zero intakes of MPED food groups on 24HR or FFQ; NIH–AARP Diet and Health Study

MPED food group	Men ( <i>n</i> 984*)			Women ( <i>n</i> 958*)		
	% with zero intake on both 24HR	% with non-zero intake on both 24HR	% with zero intake on FFQ	% with zero intake on both 24HR	% with non-zero intake on both 24HR	% with zero intake on FFQ
Milk Group						
Cheese	24.2	34.8	0.6	25.1	30.5	0.8
Milk	4.7	79.7	0.0	5.6	76.6	0.0
Yoghurt	91.2	2.3	62.1	85.9	4.2	40.5
Total dairy	1.3	90.6	0.0	2.4	88.6	0.0
Grain Group						
Non-whole grains	0.2	98.8	0.0	0.0	99.1	0.0
Whole grains	15.1	53.7	0.1	16.0	52.3	0.1
Total grains	0.1	99.4	0.0	0.0	99.4	0.0
Fruit Group						
Citrus, melon, berry	10.2	65.3	0.0	5.9	71.4	0.2
Other fruit	14.6	58.4	0.2	14.7	57.3	0.1
Total fruit	2.8	84.1	0.0	1.9	86.1	0.0
Vegetable Group						
Dark green vegetables	60.7	7.5	2.0	55.7	9.2	0.9
Orange vegetables	33.8	22.0	0.1	32.2	25.4	0.2
Potatoes	28.3	27.1	0.1	32.8	20.7	0.0
Other starchy vegetables	52.8	7.5	0.3	56.4	8.2	0.4
Tomatoes	11.1	48.4	0.0	14.7	45.6	0.3
Other vegetables	1.5	87.4	0.0	1.1	83.9	0.0
Total vegetables	0.4	95.7	0.0	0.2	93.4	0.0
Legumes	72.7	4.5	3.3	76.9	2.1	7.8
Meat Group						
Red meat	22.1	40.8	0.0	27.1	30.5	0.2
Poultry	38.3	19.7	0.2	34.0	21.5	0.0
Fish (high omega)	74.0	4.1	3.2	76.0	3.3	2.5
Fish (low omega)	64.1	6.7	0.2	69.2	4.8	0.4
Franks, luncheon meat	51.4	14.3	0.4	61.0	7.1	1.4
Organ meat	98.0	0.0	51.6	98.3	0.0	59.4
Meat, poultry & fish	1.5	90.8	0.0	1.8	85.9	0.0
Eggs	20.0	42.7	0.2	20.8	35.6	0.1
Nuts & seeds	49.3	18.7	0.4	53.1	12.9	0.5
Soya	59.1	8.3	94.3	57.8	6.9	95.5
Alcoholic beverages	54.7	23.0	23.6	65.1	16.8	30.4
Added sugars	0.1	99.4	0.0	0.0	99.1	0.0
Discretionary fat (oil)	2.0	86.0	0.0	1.6	81.6	0.0
Discretionary fat (solid)	0.0	100.0	0.0	0.0	100.0	0.0

MPED, MyPyramid Equivalents Database; 24HR, 24 h dietary recall; NIH, National Institutes of Health.

\*Percentages for the 24HR are based on the 953 men and 926 women who completed two 24HR.

instrument is the repeat application of a single 24HR. Some other studies use as reference the average of many (up to 28) days of 24HR or food records<sup>(25–27)</sup>. In such studies, simpler measurement error models may be used. Such studies tend to be small (fewer than 200 subjects), however, and it is generally considered that study designs with more subjects and fewer days per subject are more efficient<sup>(28,29)</sup>.

In epidemiological studies, the most important characteristics in determining the utility of an FFQ are the correlation of true and FFQ-reported intakes and the attenuation factor. We estimated these characteristics for

twenty-nine food groups in the NIH–AARP calibration sub-study. After energy adjustment, correlations of true and FFQ-reported intakes were estimated to be 0.5 or greater, and attenuation factors 0.4 or greater, for most of the food groups, including some that are of particular interest to nutritional epidemiologists, such as whole grains, total fruit, total vegetables, red meat and alcoholic beverages.

A limitation of our analysis (and of most FFQ validation studies) is our reliance on 24HR (or similar self-report instrument) as a reference instrument. We have assumed that the 24HR provides unbiased estimates of food group intake. Recent studies using biomarkers as references,

**Table 2** Mean reported MPED food group intakes on 24HR and FFQ, with standard errors; NIH–AARP Diet and Health Study

MPED food group (unit)	Men (n 984*)				Women (n 958*)			
	24HR		FFQ		24HR		FFQ	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Milk Group (cup equivalents)								
Cheese	0.51	0.02	0.25†	0.01	0.38	0.02	0.18†	0.01
Milk	1.04	0.03	1.17	0.04	0.85	0.03	1.02	0.04
Yoghurt	0.04	0.01	0.05	0.01	0.07	0.01	0.09‡	0.01
Total dairy	1.60	0.04	1.47	0.04	1.30	0.03	1.30	0.04
Grain Group (oz. equivalents)								
Non-whole grains	6.56	0.11	4.74†	0.07	4.57	0.07	3.76	0.06
Whole grains	1.13	0.04	1.18	0.03	0.83	0.03	0.89	0.02
Total grains	7.69	0.12	5.92†	0.09	5.41	0.08	4.65	0.08
Fruit Group (cup equivalents)								
Citrus, melon, berry	0.77	0.03	0.91	0.03	0.72	0.02	0.86	0.03
Other fruit	0.90	0.04	1.19‡	0.04	0.68	0.02	1.14‡	0.03
Total fruit	1.67	0.05	2.10‡	0.06	1.40	0.03	2.00‡	0.05
Vegetable Group (cup equivalents)								
Dark green vegetables	0.15	0.01	0.22‡	0.01	0.15	0.01	0.28‡	0.01
Orange vegetables	0.14	0.01	0.17	0.01	0.13	0.01	0.18‡	0.01
Potatoes	0.51	0.02	0.42	0.01	0.33	0.01	0.34	0.01
Other starchy vegetables	0.14	0.01	0.18‡	0.01	0.10	0.01	0.15‡	0.00
Tomatoes	0.34	0.01	0.38	0.01	0.26	0.01	0.33‡	0.01
Other vegetables	1.02	0.03	0.62†	0.02	0.85	0.02	0.66†	0.02
Total vegetables	2.30	0.05	1.99	0.04	1.82	0.04	1.94	0.04
Legumes (cup equivalents)	0.10	0.01	0.13‡	0.01	0.06	0.01	0.08‡	0.00
Meat Group (oz. lean meat equivalents)								
Red meat	2.33	0.08	1.92	0.05	1.41	0.05	1.21	0.03
Poultry	1.41	0.06	1.01†	0.03	1.24	0.05	0.95†	0.03
Fish (high omega)	0.32	0.03	0.19†	0.01	0.18	0.02	0.15	0.01
Fish (low omega)	0.81	0.06	0.53†	0.02	0.45	0.03	0.43	0.02
Franks, luncheon meat	0.66	0.03	0.72	0.02	0.36	0.02	0.40	0.02
Organ meat	0.04	0.01	0.03†	0.00	0.03	0.01	0.02	0.00
Meat, poultry & fish	5.57	0.10	4.39†	0.09	3.67	0.07	3.15†	0.07
Eggs	0.46	0.02	0.35†	0.01	0.30	0.01	0.25	0.01
Nuts & seeds	0.63	0.05	0.60	0.03	0.34	0.03	0.32	0.02
Soya	0.05	0.01	0.001†	0.00	0.03	0.01	0.001†	0.00
Alcoholic beverages (drinks)	0.82	0.05	1.10‡	0.09	0.45	0.03	0.56‡	0.06
Added sugars (teaspoons)	16.66	0.41	12.75†	0.37	12.08	0.28	9.83	0.31
Discretionary fat (oil) (g)	17.70	0.57	17.72	0.39	12.69	0.40	15.82‡	0.37
Discretionary fat (solid) (g)	45.71	0.84	37.08	0.73	32.00	0.59	27.06	0.53

MPED, MyPyramid Equivalents Database; 24HR, 24 h dietary recall; NIH, National Institutes of Health.

\*Means for the 24HR are based on the 953 men and 926 women who completed two 24HR.

†FFQ mean at least 20% smaller than 24HR mean.

‡FFQ mean at least 20% larger than 24HR mean.

however, have shown that the 24HR is biased for energy, protein and energy-adjusted protein intake, and that these biases sometimes, but not always, lead to overestimation of correlations with true intake and attenuation factors when the 24HR is used as a reference instrument<sup>(21,30)</sup>. While no such biomarkers are presently known for any food groups, it is not unreasonable to expect similar biases for at least some food groups. To the extent that this is so, our estimates of the correlations with true intake and attenuation factors could be biased and may overestimate the true parameters.

The two-part model used in the current analysis has been validated by computer simulations<sup>(9)</sup>. In addition, graphical methods have been developed to assess the model's goodness-of-fit to specific data<sup>(12)</sup>. A comparison of Tables 1 and 3 indicates that the precision of the estimated correlations and attenuation factors is related to the frequency with which a food is consumed. The standard

errors of the estimated correlations and attenuation factors for less frequently consumed food groups, such as legumes, fish and other starchy vegetables, tend to be larger than those for more frequently consumed food groups such as milk, whole grains and red meat, and, as we saw with other starchy vegetables in men, there is a possibility that the measurement error model will fail to converge if the food group is infrequently consumed. This is because there is less information about the amount consumed on consumption days when there are fewer consumption days in the data. In particular, if there are only a few subjects who have non-zero consumption on multiple days, then it is difficult to separate between- and within-person error (i.e. difficult to estimate the variances of  $U_{2i}$  and  $\varepsilon_{2ij}$ ). To estimate infrequently consumed foods with more precision, it would be necessary to have a larger calibration sub-study.

**Table 3** Estimates of the correlation of true and FFQ-reported food intakes ( $\rho_{QT}$ ) and the attenuation factor ( $\lambda$ ) for FFQ-reported food intake, with standard errors; NIH–AARP Diet and Health Study

MPED food group	Model	Men (n 984)				Women (n 958)			
		$\rho_{QT}$	SE	$\lambda$	SE	$\rho_{QT}$	SE	$\lambda$	SE
Cheese	Unadjusted	0.59	0.06	0.55	0.05	0.42	0.08	0.35	0.06
	Energy-adjusted	0.63	0.08	0.58	0.05	0.48	0.10	0.42	0.06
Milk	Unadjusted	0.68	0.03	0.53	0.03	0.67	0.03	0.50	0.03
	Energy-adjusted	0.70	0.03	0.53	0.03	0.73	0.03	0.52	0.03
Total dairy	Unadjusted	0.61	0.04	0.44	0.03	0.58	0.03	0.42	0.03
	Energy-adjusted	0.63	0.04	0.44	0.03	0.73	0.03	0.52	0.03
Whole grains	Unadjusted	0.59	0.04	0.55	0.04	0.48	0.04	0.48	0.05
	Energy-adjusted	0.65	0.04	0.61	0.04	0.53	0.05	0.52	0.05
Non-whole grains	Unadjusted	0.34	0.04	0.24	0.03	0.39	0.05	0.25	0.03
	Energy-adjusted	0.50	0.05	0.37	0.04	0.47	0.06	0.31	0.04
Total grains	Unadjusted	0.35	0.04	0.24	0.03	0.39	0.05	0.23	0.03
	Energy-adjusted	0.54	0.05	0.39	0.03	0.53	0.05	0.31	0.03
Citrus, melon, berry	Unadjusted	0.64	0.03	0.54	0.03	0.57	0.03	0.43	0.03
	Energy-adjusted	0.70	0.03	0.59	0.03	0.62	0.03	0.46	0.03
Other fruit	Unadjusted	0.70	0.03	0.68	0.04	0.60	0.03	0.52	0.03
	Energy-adjusted	0.74	0.03	0.71	0.04	0.64	0.04	0.55	0.03
Total fruit	Unadjusted	0.70	0.02	0.61	0.03	0.58	0.03	0.46	0.03
	Energy-adjusted	0.76	0.03	0.66	0.03	0.65	0.04	0.51	0.03
Dark green vegetables	Unadjusted	0.75	0.10	0.57	0.06	0.52	0.07	0.50	0.05
	Energy-adjusted	0.78	0.10	0.59	0.06	0.58	0.08	0.56	0.06
Orange vegetables	Unadjusted	0.62	0.10	0.41	0.05	0.57	0.07	0.48	0.05
	Energy-adjusted	0.71	0.10	0.50	0.05	0.62	0.07	0.54	0.05
Potatoes	Unadjusted	0.58	0.12	0.37	0.05	0.38	0.07	0.27	0.05
	Energy-adjusted	0.60	0.14	0.40	0.06	0.35	0.11	0.23	0.05
Other starchy vegetables	Unadjusted	*		*		0.50	0.19	0.26	0.07
	Energy-adjusted	*		*		0.56	0.18	0.29	0.07
Tomatoes	Unadjusted	0.44	0.11	0.29	0.05	0.60	0.09	0.42	0.05
	Energy-adjusted	0.54	0.11	0.39	0.06	0.64	0.10	0.45	0.05
Other vegetables	Unadjusted	0.46	0.05	0.37	0.04	0.44	0.05	0.34	0.05
	Energy-adjusted	0.50	0.06	0.43	0.05	0.54	0.06	0.44	0.05
Total vegetables	Unadjusted	0.46	0.04	0.32	0.03	0.42	0.05	0.32	0.04
	Energy-adjusted	0.55	0.05	0.43	0.04	0.52	0.05	0.44	0.04
Legumes	Unadjusted	0.44	0.08	0.44	0.08	0.40	0.23	0.27	0.07
	Energy-adjusted	0.50	0.09	0.48	0.08	0.34	0.20	0.28	0.07
Fish high omega	Unadjusted	0.48	0.11	0.59	0.10	0.46	0.17	0.45	0.12
	Energy-adjusted	0.55	0.13	0.66	0.11	0.60	0.15	0.56	0.12
Fish low omega	Unadjusted	0.39	0.09	0.42	0.09	0.74	0.14	0.47	0.08
	Energy-adjusted	0.42	0.09	0.47	0.09	0.71	0.13	0.50	0.09
Red meat	Unadjusted	0.56	0.05	0.50	0.04	0.83	0.10	0.47	0.04
	Energy-adjusted	0.54	0.05	0.55	0.05	0.84	0.09	0.52	0.05
Poultry	Unadjusted	0.47	0.11	0.34	0.05	0.39	0.09	0.25	0.04
	Energy-adjusted	0.53	0.10	0.42	0.05	0.46	0.09	0.33	0.05
Franks, luncheon meat	Unadjusted	0.61	0.07	0.54	0.05	0.55	0.13	0.36	0.05
	Energy-adjusted	0.64	0.07	0.60	0.05	0.67	0.15	0.39	0.06
Meat, poultry & fish	Unadjusted	0.44	0.04	0.27	0.03	0.45	0.06	0.23	0.03
	Energy-adjusted	0.44	0.06	0.31	0.05	0.53	0.05	0.33	0.03
Eggs	Unadjusted	0.70	0.06	0.78	0.05	0.54	0.11	0.53	0.07
	Energy-adjusted	0.69	0.05	0.81	0.05	0.55	0.11	0.56	0.07
Nuts & seeds	Unadjusted	0.54	0.04	0.58	0.05	0.48	0.07	0.48	0.06
	Energy-adjusted	0.54	0.06	0.56	0.06	0.60	0.10	0.50	0.07
Alcoholic beverages	Unadjusted	0.80	0.03	0.67	0.04	0.81	0.03	0.68	0.04
	Energy-adjusted	0.82	0.03	0.70	0.03	0.81	0.03	0.65	0.04
Added sugars	Unadjusted	0.55	0.03	0.41	0.03	0.46	0.04	0.39	0.03
	Energy-adjusted	0.63	0.03	0.44	0.03	0.58	0.03	0.43	0.03
Discretionary fat (oil)	Unadjusted	0.46	0.05	0.34	0.04	0.30	0.07	0.19	0.04
	Energy-adjusted	0.43	0.05	0.34	0.04	0.38	0.15	0.18	0.04
Discretionary fat (solid)	Unadjusted	0.65	0.04	0.47	0.03	0.50	0.04	0.36	0.03
	Energy-adjusted	0.76	0.03	0.61	0.03	0.64	0.04	0.49	0.03

NIH, National Institutes of Health; MPED, MyPyramid Equivalents Database.

\*For Other starchy vegetables in men, the measurement error model failed to converge.

A number of studies have validated FFQ for intakes of foods or food groups in American adults, including those described by Salvini *et al.*<sup>(25)</sup>, Flagg *et al.*<sup>(7)</sup> and Millen *et al.*<sup>(8)</sup>. Direct comparison with these studies is complicated

by the fact that the food groups validated were generally not the same as in the present study and were not measured in MPED servings. Further, some studies, such as Salvini *et al.*<sup>(25)</sup>, used food records rather than 24HR as reference

**Table 4** Number of incident cancers in the NIH–AARP Diet and Health Study 1995–2003; power to detect an odds ratio of 1.5\* using FFQ-reported intake if the correlation of true and FFQ-reported intakes  $\rho_{TQ} = 1$  and if  $\rho_{TQ} = 0.5$ 

Type of cancer	Men (n 262 642)			Women (n 183 535)		
	No. of cases	Power if $\rho_{QT} = 1$	Power if $\rho_{QT} = 0.5$	No. of cases	Power if $\rho_{QT} = 1$	Power if $\rho_{QT} = 0.5$
Prostate	15,949	1.00	1.00	–	–	–
Breast	–	–	–	5478	1.00	1.00
Endometrial	–	–	–	1041	1.00	0.72
Ovarian	–	–	–	475	0.93	0.41
Lung	3769	1.00	1.00	2288	1.00	0.97
Colorectal	3031	1.00	0.99	1457	1.00	0.86
Melanoma	1485	1.00	0.86	543	0.96	0.45
Bladder	1246	1.00	0.80	235	0.68	0.23
Non-Hodgkin's lymphoma	1114	1.00	0.75	605	0.97	0.50
Head and neck	939	1.00	0.68	300	0.78	0.28
Kidney	857	1.00	0.64	322	0.81	0.29
Pancreatic	601	0.97	0.49	348	0.84	0.31
Stomach	440	0.91	0.38	127	0.43	0.14
Oesophagus	425	0.90	0.37	76	0.28	0.10
Brain	356	0.85	0.32	146	0.48	0.16
Myeloma	331	0.82	0.30	157	0.51	0.17
Myeloid leukaemia	288	0.77	0.27	119	0.41	0.14
Liver	238	0.69	0.23	72	0.27	0.10
Thyroid	153	0.50	0.16	176	0.56	0.18

\*Odds ratio comparing the 90th to 10th percentile of true intake in a univariate diet–disease model.

instrument. To the extent that comparisons can be made, results of the present study are generally similar to the earlier studies. For example, Salvini *et al.*<sup>(25)</sup> reported energy-adjusted correlations for intake of fish, eggs (men) and tomatoes that were similar to those in Table 3, although the correlation for egg intake in women was somewhat higher in their study (0.77 compared to 0.55). Flagg *et al.*<sup>(7)</sup> reported energy-adjusted correlations for total grains, total vegetables and red meat that were similar to those in the present study.

The study most comparable to ours is an analysis of the Eating at America's Table Study (EATS) reported by Millen *et al.*<sup>(8)</sup>. In that analysis, the NCI's DHQ was validated for food groups derived from the USDA Pyramid Servings Database<sup>(31)</sup>, a database that is similar to MPED but based on earlier dietary guidelines. The DHQ is a later version of the FFQ used in the NIH–AARP study. In general, energy-adjusted correlations in EATS and the present study are similar, although there are some differences. For example, energy-adjusted correlations for total vegetables were 0.63 (men) and 0.66 (women) in EATS, compared with 0.55 (men) and 0.52 (women) in the present study. Possible explanations for these differences include the facts that the EATS sample was comprised of subjects aged 20–70 years, while the NIH–AARP sample was older (50–71) years, and the EATS analysis did not use methods designed for episodically consumed foods.

As shown in Table 4, when the correlation of true and FFQ-reported intakes is at least 0.5, the NIH–AARP study will have at least 85% power to detect moderate diet–disease associations (odds ratios 1.5 or greater) for common cancer types such as prostate, breast, lung and colorectal. For less common types such as thyroid or liver, however, the power to detect such associations will be much lower. Similarly, when the attenuation factor is at least 0.4, moderate

diet–disease associations may be substantially underestimated, but not to the point where they disappear altogether. For example, if the true odds ratio is 1.5 ( $\alpha_1 = \log(1.5)$  in equation (4)) and the attenuation factor is 0.4, then the estimated odds ratio will have mean equal to about  $1.5^{0.4} = 1.18$ . Moreover, when the attenuation factor is small, say less than 0.2, attempting to 'deattenuate' estimates will give unreliable results and is not advised. When the attenuation factor is at least 0.4, however, it is possible to deattenuate an estimated log odds ratio by dividing it by the attenuation factor, giving an approximately unbiased estimate<sup>(4)</sup>. In summary, the levels of correlation and attenuation factor that we have estimated indicate that the NIH–AARP FFQ is suitable for estimating and testing many, but not all, diet–disease relationships in the NIH–AARP cohort.

## Acknowledgements

R.J.C.'s research was supported by a grant from the National Cancer Institute (CA57030). None of the authors have any conflict of interest. A.S. directed the study. A.S., A.F.S., F.E.T., L.S.F. and R.J.C. participated in the design of the study. D.M., L.S.F., R.J.C. and V.K. developed the statistical methodology and designed and reviewed the analysis. D.M. and M.A.S. carried out the analysis. D.M. had primary responsibility for writing the manuscript, but all authors contributed to its writing and revision.

## References

1. Freudenheim JL & Marshall JR (1988) The problem of profound mismeasurement and the power of epidemiologic studies of diet and cancer. *Nutr Cancer* **11**, 243–250.

2. Freedman LS, Schatzkin A & Wax Y (1990) The impact of dietary measurement error on planning a sample size required in a cohort study. *Am J Epidemiol* **132**, 1185–1195.
3. Beaton GH, Milner J, Corey P *et al.* (1979) Sources of variance in 24-hour dietary recall data; implications for nutritional study design and interpretation. *Am J Clin Nutr* **32**, 2546–2559.
4. Rosner B, Spiegelman D & Willett WC (1990) Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *Am J Epidemiol* **132**, 734–744.
5. Freedman LS, Carroll RJ & Wax Y (1991) Estimating the relation between dietary intake obtained from a food frequency questionnaire and true average intake. *Am J Epidemiol* **134**, 310–320.
6. Feskanich D, Rimm EB, Giovannucci EL *et al.* (1993) Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. *J Am Diet Assoc* **93**, 790–796.
7. Flagg EW, Coates RJ, Calle EE *et al.* (2000) Validation of the American Cancer Society Cancer Prevention Study II Nutritional Survey Cohort food frequency questionnaire. *Epidemiology* **11**, 462–468.
8. Millen AE, Midthune D, Thompson FE *et al.* (2006) The National Cancer Institute Diet History Questionnaire: validation of pyramid food servings. *Am J Epidemiol* **163**, 279–288.
9. Toozé JA, Midthune D, Dodd KW *et al.* (2006) A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *J Am Diet Assoc* **106**, 1575–1587.
10. Freedman LS, Guenther PM, Krebs-Smith SM *et al.* (2010) A population's distribution of Health Eating Index-2005 component scores can be estimated when more than one 24-hour recall is available. *J Nutr* **140**, 1529–1534.
11. Marriott BP, Olsho L, Hadden L *et al.* (2010) Intake of added sugars and selected nutrients in the United States, National Health and Nutrition Examination Survey (NHANES) 2003–2006. *Clin Rev Food Sci Nutr* **50**, 228–258.
12. Kipnis V, Midthune D, Buckman DW *et al.* (2009) Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics* **65**, 1003–1010.
13. Thompson FE, Kipnis V, Midthune D *et al.* (2008) 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**, 183–195.
14. Schatzkin A, Subar AF, Thompson FE *et al.* (2001) 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**, 1119–1125.
15. US National Cancer Institute, Division of Cancer Control and Population Sciences, Applied Research Program (2010) Risk Factor Monitoring and Methods homepage. <http://www.riskfactor.cancer.gov>
16. Subar AF, Midthune D, Kulldorff M *et al.* (2000) Evaluation of alternative approaches to assign nutrient values to food groups in food frequency questionnaires. *Am J Epidemiol* **152**, 279–286.
17. Friday JE & Bowman SA (2006) *MyPyramid Equivalents Database for USDA Survey Food Codes, 1994–2002 Version 1.0*. Beltsville, MD: USDA, Agricultural Research Service, Community Nutrition Research Group; available at <http://www.ars.usda.gov/ba/bhnrc/fsrg>
18. US Department of Health and Human Services & US Department of Agriculture (2005) *Dietary Guidelines for Americans*. Washington, DC: US Government Printing Office.
19. Box GEP & Cox DR (1964) An analysis of transformations. *J R Stat Soc B* **26**, 211–252.
20. US National Cancer Institute, Division of Cancer Control and Population Sciences, Applied Research Program (2010) Risk Factor Monitoring and Methods. Usual Dietary Intakes: Background. <http://www.riskfactor.cancer.gov/diet/usualintakes>
21. Kipnis V, Midthune D, Freedman LS *et al.* (2001) Empirical evidence of correlated biases in dietary assessment instruments and its implications. *Am J Epidemiol* **153**, 394–403.
22. Kaaks R, Riboli E & van Staveren W (1995) Calibration of dietary intake measurements in prospective cohort studies. *Am J Epidemiol* **142**, 548–556.
23. Willett WC, Howe GR & Kushi LH (1997) Adjustment for total energy intake in epidemiologic studies. *Am J Clin Nutr* **65**, Suppl. 4, 1220S–1228S.
24. George SM, Mayne ST, Leitzmann MF *et al.* (2009) Dietary glycemic index, glycemic load, and risk of cancer: a prospective cohort study. *Am J Epidemiol* **169**, 462–472.
25. Salvini S, Hunter DJ, Sampson L *et al.* (1989) Food-based validation of a dietary questionnaire: the effects of week-to-week variation in food consumption. *Int J Epidemiol* **18**, 858–867.
26. Bohlscheid-Thomas S, Hoting I, Boeing H *et al.* (1997) Reproducibility and relative validity of food group intake in a food frequency questionnaire developed for the German part of the EPIC project. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* **26**, Suppl. 1, S59–S69.
27. Shu XO, Yang G, Jin F *et al.* (2004) Validity and reproducibility of the food frequency questionnaire used in the Shanghai Women's Health Study. *Eur J Clin Nutr* **58**, 17–23.
28. Rosner B & Willett WC (1988) Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. *Am J Epidemiol* **127**, 377–386.
29. Kaaks R, Riboli E & van Staveren W (1995) Sample size requirements for calibration studies of dietary intake measurements in prospective cohort investigations. *Am J Epidemiol* **142**, 557–565.
30. Kipnis V, Subar AF, Midthune D *et al.* (2003) Structure of dietary measurement error: results of the OPEN Biomarker Study. *Am J Epidemiol* **158**, 14–21.
31. Cook A & Friday JE (2004) *Pyramid Servings Database for USDA Survey Food Codes, Version 2.0*. Washington, DC: Agricultural Research Service, US Department of Agriculture.

## Appendix 1

### Box–Cox transformations

The Box–Cox transformation is defined as

$$g(x; \lambda) = \begin{cases} (x^\lambda - 1)/\lambda & \text{if } \lambda > 0 \\ \log(x) & \text{if } \lambda = 0 \end{cases} \quad (5)$$

for some transformation parameter  $\lambda$ . We use Box–Cox transformations to transform  $Q_i$  and positive  $R_{ij}$  to approximate normality, defining  $Q_i^* = g(Q_i; \lambda_Q)$  and  $R_{ij}^* = g(R_{ij}; \lambda_R)$ , and choosing  $\lambda_Q$  and  $\lambda_R$  so as to maximize the Shapiro–Wilk test statistic for normality for  $Q_i$  and positive  $R_{ij}$ . We also define  $T_i^* = g(T_i; \lambda_T)$ , choosing  $\lambda_T$  so as to minimize the Kolmogorov test statistic

for normality for  $T_i$  in the Monte Carlo distribution (see Appendix 2).

**Appendix 2**

**Monte Carlo distribution of true and FFQ-reported intakes**

Under the assumptions that  $R_{ij}$  is an unbiased estimate of  $T_b$ , and that the model defined by equations (1) and (2) is correct, one can write  $T_i$  as a function of  $Q_i$  and random effects ( $U_{1b}$ ,  $U_{2i}$ ) as:

$$T_i = H(\beta_{10} + \beta_{11}Q_i^* + U_{1i}) \times E\{g^{-1}(\beta_{20} + \beta_{21}Q_i^* + U_{2i} + \varepsilon_{2ij}; \lambda_R) \mid Q_i, U_{2i}\} \approx H(\beta_{10} + \beta_{11}Q_i^* + U_{1i}) \times g^*(\beta_{20} + \beta_{21}Q_i^* + U_{2i}; \lambda_R), \tag{6}$$

where  $H(x)$  is the logistic function and  $g^*(v; \lambda_R)$  is a Taylor-series approximation of the expectation  $E\{g^{-1}(v + \varepsilon_{2ij}; \lambda_R) \mid v\}$ ,

$$g^*(v; \lambda_R) = g^{-1}(v; \lambda_R) + \frac{1}{2} \sigma_{\varepsilon_2}^2 \frac{\partial^2 \{g^{-1}(v; \lambda_R)\}}{\partial v^2}. \tag{7}$$

One can estimate relationships between true and FFQ-reported intakes by generating a Monte Carlo distribution of ( $T_b$ ,  $Q$ ). For each individual  $i$ , generate random effects ( $U_{1b}$ ,  $U_{2i}$ ) having a joint normal distribution with variances ( $\hat{\sigma}_{U_1}^2, \hat{\sigma}_{U_2}^2$ ) and correlation  $\hat{\rho}_{U_1, U_2}$ , and calculate  $T_i$  as in equation (6). Repeat this process  $m = 100$  times for each individual, so that the resulting Monte Carlo distribution has  $n \times m$  pseudo-individuals. Under the assumptions listed above, the Monte Carlo distribution will be approximately the same as the real distribution of ( $T_b$ ,  $Q$ ), and one can use it to estimate the attenuation factor  $\gamma_1$  and correlation with true intake  $\rho_{TQ}$ , described in the main text. We estimate  $\rho_{TQ}$  as the sample correlation of Box–Cox transformed variables  $T_i^*$  and  $Q_i^*$  in the Monte Carlo distribution. Similarly, we estimate  $\gamma_1$  as the slope in the regression of  $T_i^*$  v.  $Q_i^*$ , after standardizing both variables so that a unit change on the transformed scale is equal to the change from the 10th to 90th percentile of true intake on that scale. Standard errors are estimated using a bootstrap method.

The Monte Carlo method for energy-adjusted foods is similar to the method for unadjusted foods, except that we use the parameter estimates from the three-part food and energy model (see Appendix 3) to generate random effects ( $U_{1b}$ ,  $U_{2b}$ ,  $U_{3i}$ ) and create a Monte Carlo distribution of ( $T_{Fi}$ ,  $T_{Ei}$ ,  $Q_{Fi}$ ,  $Q_{Ei}$ ). We then calculate  $T_{Ri}$  as the residual in the regression of  $T_{Fi}^*$  v.  $T_{Ei}^*$ , and  $Q_{Ri}$  as the residual in the regression of on  $Q_{Fi}^*$  v.  $Q_{Ei}^*$ . Finally, we estimate  $\rho_{TQ}$  as the sample correlation of  $T_{Ri}$  and  $Q_{Ri}$ , and  $\gamma_1$  as the slope in the regression of  $T_{Ri}$  v.  $Q_{Ri}$ .

**Appendix 3**

**Three-part food and energy model**

For individual  $i$ ,  $i = 1, \dots, n$ , let

- $T_{Fi}$  be the true usual intake of an episodically consumed food
- $T_{Ei}$  be the true usual intake of energy
- $R_{Fij}$  be the 24HR-reported intake of the episodically consumed food on day  $j$
- $R_{Eij}$  be the 24HR-reported intake of energy on day  $j$
- $Q_{Fi}$  be the FFQ-reported intake of the episodically consumed food
- $Q_{Ei}$  be the FFQ-reported intake of energy.

Under the food and energy model, we assume that the 24HR is unbiased for true intake:

$$E(R_{Fij} \mid i) = T_{Fi} \tag{8}$$

and

$$E(R_{Eij} \mid i) = T_{Ei}, \tag{9}$$

and that, after appropriate transformations, the relationship between 24HR and FFQ can be described by the following three-part non-linear mixed-effects model:

$$\begin{aligned} \text{logit}(p_i) &= \beta_{10} + \beta_{11} \times Q_{Fi}^* + \beta_{12} \times Q_{Ei}^* + U_{1i}, \tag{10} \\ (R_{Fij}^* \mid R_{Fij} > 0) &= \beta_{20} + \beta_{21} \times Q_{Fi}^* + \beta_{22} \times Q_{Ei}^* + U_{2i} + \varepsilon_{2ij} \tag{11} \end{aligned}$$

and

$$R_{Eij}^* = \beta_{30} + \beta_{31} \times Q_{Fi}^* + \beta_{32} \times Q_{Ei}^* + U_{3i} + \varepsilon_{3ij}, \tag{12}$$

where  $p_i$  is the probability that  $R_{Fij} > 0$ ,  $\text{logit}(p) = \log\{p/(1-p)\}$  is the inverse of the logistic distribution function, random effects ( $U_{1b}$ ,  $U_{2b}$ ,  $U_{3i}$ ) have a joint normal distribution with mean zero, within-person random errors ( $\varepsilon_{2ij}$ ,  $\varepsilon_{3ij}$ ) have a joint normal distribution with mean zero, and within-person errors ( $\varepsilon_{2ij}$ ,  $\varepsilon_{3ij}$ ) are independent of random effects ( $U_{1b}$ ,  $U_{2b}$ ,  $U_{3i}$ ). Variables  $Q_{Fi}^*$ ,  $Q_{Ei}^*$ ,  $R_{Fij}^*$  and  $R_{Eij}^*$  are Box–Cox transformations of  $Q_{Fi}$ ,  $Q_{Ei}$ ,  $R_{Fij}$  and  $R_{Eij}$  to scales on which they are approximately normal (see Appendix 1).

**Appendix 4**

**Estimating power in a univariate diet–disease model**

Suppose we are fitting diet–disease model, equation (4), using  $Q_i$  to measure intake and there are  $D$  cases of disease in the cohort. Kaaks *et al.*<sup>(22)</sup> show that the power to detect  $\alpha_1$  at significance level  $\gamma$  is approximately:

$$\begin{aligned} \text{Power} &\approx \Phi\left(\frac{|\alpha_1|}{\sqrt{\text{var}(\hat{\alpha}_1)} - z_{\gamma/2}}\right) \\ &\approx \Phi\left(\frac{|\alpha_1| \rho_{TQ} \sigma_{T^*} \sqrt{D} - z_{\gamma/2}}{\sigma_{T^*}}\right), \tag{13} \end{aligned}$$

where  $\sigma_{T^*}$  is the standard deviation of  $T_i^*$ ,  $\Phi(z)$  is the standard normal distribution and  $z_{\gamma/2} = \Phi^{-1}(1 - \gamma/2)$ . The power to test the hypothesis that the odds ratio comparing the 90th to 10th percentile of true intake is equal to 1.5, i.e. that  $2.56\sigma_{T^*} \times \alpha_1 = \log(1.5)$ , is then:

$$\text{Power} \approx \Phi\left(\frac{\{\log(1.5)/2.56\} \rho_{TQ} \sqrt{D} - z_{\gamma/2}}{\sigma_{T^*}}\right). \tag{14}$$