Public Health Nutrition: 6(4), 385–391 DOI: 10.1079/PHN2002438

Seasonal misclassification error and magnitude of true between-person variation in dietary nutrient intake: a random coefficients analysis and implications for the Japan Public Health Center (JPHC) Cohort Study

Michael T Fahey[1,2,*], Satoshi Sasaki[1], Minatsu Kobayashi[1], Masayuki Akabane[3] and Shoichiro Tsugane[1]

[1]Epidemiology and Biostatistics Division, National Cancer Center Research Institute East, 6-5-1 Kashiwanoha, Kashiwa-shi, Chiba-ken, 277-8577 Japan: [2]Current address: Unit of Nutrition and Cancer, International Agency for Research on Cancer, 150 cours Albert-Thomas, 69372 Lyon Cedex 08, France: [3]Department of Nutrition, Tokyo University of Agriculture, Japan

Submitted 15 May 2002: Accepted 21 October 2002

Abstract

Objective: We examine (1) the extent to which seasonal diet assessments correctly classify individuals with respect to their usual nutrient intake, and (2) whether the magnitude of true variation in intake between individuals is seasonal. These effects could lead, respectively, to bias in estimates of relative risk for associations between usual nutrient exposure and disease, and to an increase in required sample size.
Subjects and setting: One hundred and twenty-seven families in four regions of the Japan Public Health Center (JPHC) Cohort Study.
Design: On average, 48 weighed daily food records were collected per family over six seasons of 1994 and 1995.
Results: A random slopes regression model was used to predict the correlation between seasonal and annual average intakes, and to estimate true between-person variation in intakes by season. Mean vitamin C intake was greatest in summer and autumn, and seasonal variation was attributable to the consumption of fruit and vegetables. Predicted correlations between seasonal and annual average vitamin C intake ranged from 0.62 to 0.87, with greatest correlations in summer and autumn. True between-person variation in vitamin C intake was also strongly seasonal, ranging from 45 to 78% of total variance, and was again greatest in summer and autumn. These effects were less seasonal among energy and 13 other nutrients.
Conclusions: It may be possible substantially to reduce both seasonal misclassification of individuals with respect to their usual vitamin C intake, and required sample size, by asking subjects to report high-season intake of fruit and vegetables in the JPHC Study.

Keywords
Seasons
Vitamin C
Misclassification error
True variance
Epidemiological methods
Multilevel analysis
Food records
Measurement error
Japan

US studies have reported little variation by season in mean total energy or macronutrient intake[1–4]. However, seasonal variation in fruit and vegetable consumption has been found in the USA[2,4] and Europe[5]. Several cross-sectional studies have reported that seasonal differences in average intake and seasonal misclassification of individual intake were greatest for vitamin C and fruit[1,4,5].

If seasonal diet assessments misclassify individuals with respect to their usual intake, then estimates of relative risk for associations between usual dietary exposure and disease could be biased. Moreover, the precision of relative risk estimates could depend upon the season in which diet is assessed, if the magnitude of true variation in intake between individuals is seasonal.

This paper presents data collected in part to validate the use of a diet questionnaire in the Japan Public Health Center (JPHC) Study Cohort I in four regions of Japan[6]. The primary objectives were to examine the effect of seasonal diet assessment on (1) misclassification of an individual's long-term (or usual) intake, and (2) the magnitude of true between-person variation in intake.

Materials and methods

Study design, measures and subjects

Subjects were 127 married couples who attended a public health centre (PHC) in one of four regions of Japan (Yokote PHC in Akita, Ninohe PHC in Iwate, Saku PHC in Nagano, Ishikawa PHC in Okinawa) and who were invited to participate in a study to validate the use of a diet questionnaire in the JPHC Study Cohort I[6],

*Corresponding author: Email fahey@iarc.fr

in which 124 of these couples were enrolled. These regions were chosen to reflect the three-fold nation-wide variation in gastric cancer mortality rates, and also differ in environmental, anthropometric and dietary characteristics[7–9]. Weighed food records were completed by women, for both family members, on seven consecutive days in four seasons of 1994 in three regions of Honshu (Akita, Iwate, Nagano), and in two seasons of 1995 in Okinawa (winter, summer). Recording began in winter in all regions.

Subjects were given a scale and instructed by research dietitians how to record, in prepared booklets, all food and beverages consumed. Records were checked during the survey and reviewed after completion. Subjects were permitted to record portion sizes in household units for frequently consumed foods or when it was not convenient to use the scale. Daily intakes of energy and 14 nutrients were calculated from the food records using the standard food composition table published by the Science and Technology Agency of Japan[10].

Statistical analysis

The mean of a maximum of seven daily food records in each season was used as the unit of observation for each person. This number of records is sufficient to estimate group means reliably and to correlate with long-term average intake[5,11] based upon within-person divided by between-person variance ratios for most nutrients[12].

Nutrients whose winter and summer median intakes varied by more than 15% among men and women of at least one region were identified for detailed study. Food groups were categorised according to the standard Japanese food composition table[10].

A random linear coefficients regression model[13,14] was used to estimate individual mean nutrient intake by season, and to predict the correlation between seasonal and annual average individual intakes. For $k = 1, 2$ individuals belonging to $i = 1, \ldots, 127$ families observed in $j = 1, \ldots, 4$ seasons, the relationship between nutrient intake, y_{ijk}, and season was modelled as:

$$y_{ijk} = \beta_{0i} + \beta_{1i}SPR_j + \beta_{2i}SUM_j + \beta_{3i}AUT_j + \beta_{4i}LT_j + e_{jk(i)}, \quad (1)$$

where SPR_j, SUM_j and AUT_j are indicator variables for the season of diet assessment, and LT_j takes values 0, 1, 2 and 3 for the respective seasons of winter through to autumn. The term $\beta_{4i}LT_j$ may appear to be confounded with the effects of the fixed seasonal indicator variables. However, this is not the case, because (as explained below when specifying the family-level equations) this term represents the random part only of the linear time effect, and does not require the estimation of a fixed effect.

The individual-level regression coefficients in equation (1) depend upon i and can be defined at the family level. The family-level equation for β_{0i}, the mean intake for the ith family in winter, is:

$$\beta_{0i} = Z_{00} + Z_{01}IWATE_i + Z_{02}NAGANO_i + Z_{03}OKINAWA_i + u_{0i}, \quad (2)$$

where $IWATE_i$, $NAGANO_i$ and $OKINAWA_i$ are indicator variables denoting region of residence. The $e_{jk(i)} \sim N(0, \sigma_E^2)$ and the $u_{0i} \sim N(0, \sigma_I^2)$, where σ_E^2 is the error variance, and in family winter mean intakes, i.e. the intercepts, around the grand winter mean. The equations for β_{1i}, β_{2i} and β_{3i} were written analogously to equation (2), but without random components, i.e. no terms analogous to u_{0i} in equation (2) were included.

The family-level equation for the linear time effect is $\beta_{4i} = u_{4i}$ and defines this effect to be purely random with $u_{4i} \sim N(0, \sigma_S^2)$, where σ_S^2 is the variance in the linear component of change in family intake over time, i.e. the slopes. The random effects u_{0i}, u_{4i} are assumed to be distributed bivariate normal, and to be independent of $e_{jk(i)}$[14]. Individual age in winter and sex were also included in the model. A single equation defining both the individual and family levels of the hierarchical model and their interactions can be obtained by substituting equation (2) and the other family-level equations into equation (1).

A random quadratic coefficients model was fit by adding an LT_j^2 term to equation (1) and by setting its family-level coefficient to be purely random in a fashion analogous to the random linear coefficients model (see Appendix A). Model parameters were estimated using restricted maximum likelihood, and model checking was done by residual analysis at the individual and family levels[14,15].

The total variance in an observation under equation (1) is given by:

$$\mathrm{var}(y_{ijk}) = \sigma_I^2 + LT_j^2\sigma_S^2 + 2LT_j\sigma_{IS} + \sigma_E^2, \quad (3)$$

where σ_{IS} is the covariance between u_{0i}, u_{4i}[14]. The proportion of the total due to true variation in individual nutrient intake was predicted by season from equation (3) as $(\mathrm{var}(y_{ijk}) - \sigma_E^2)/\mathrm{var}(y_{ijk})$ and expressed as a percentage. Correlations between seasonal and annual average individual intake, $\bar{y}_{i\bullet k}$, were predicted from equation (4) (see Appendix A) as:

$$\mathrm{corr}(y_{ijk}, \bar{y}_{i\bullet k}) = \mathrm{cov}(y_{ijk}, \bar{y}_{i\bullet k}) / \left(\sqrt{\mathrm{var}(y_{ijk})}\sqrt{\mathrm{var}(\bar{y}_{i\bullet k})}\right), \quad (4)$$

where $\mathrm{cov}(y_{ijk}, \bar{y}_{i\bullet k})$ is the covariance between an observation of nutrient intake in the jth season and the mean of seasonal observations for the kth individual in the ith family, and $\mathrm{var}(\bar{y}_{i\bullet k})$ is the variance in the mean for the kth individual in the ith family.

Results

Study subjects and patterns in crude data

Daily food records were complete for 221 persons (87%), with an average of 48 records collected per family.

One male's three days of food records in summer, for which daily average total energy was only 2427 kJ, were excluded. Seven days of food records were completed on 854/871 person-season occasions, and on all occasions three or more days were completed.

Approximately 80% of subjects were aged between 50 and 64 years and more than half was involved in agricultural or family-related activity (Table 1). Twenty-eight per cent had only high school education or lower. Current smoking was reported by 32% of men and 1% of women.

Of total energy and 14 nutrients, variation between winter and summer median intakes exceeded 15% among men and women of at least one region only for carotene, vitamin C and retinol. Median retinol intake was greatest in winter, except for men in Iwate and women in Akita, for whom it was greatest in summer or spring, respectively. Mean retinol intake was not studied in detail, because of the poor precision in its estimate[12]. Plots of individual carotene and vitamin C intakes by region and sex indicated that seasonal changes were similar for men and women, and that total variation in intake was also seasonal (not shown). The intra-class correlation between any two intakes from the same family, estimated using a random intercepts model by setting $\beta_{4i} = 0$ in equation (1), was 0.50 for carotene and 0.48 for vitamin C.

Table 1 Characteristics of study subjects by sex, Japan, 1994–1995

	Men		Women	
Characteristic	*n**	%	*n**	%
Maximum sample size	127	100	127	100
Region of residence				
Akita	36	28	36	28
Iwate	31	24	31	24
Nagano	30	24	30	24
Okinawa	30	24	30	24
Age distribution (years)				
<50	11	9	28	22
50–64	105	83	95	75
≥65	10	8	4	3
Education (years)				
9–11	33	30	31	27
12–15	67	60	83	72
≥16	11	10	1	1
Occupation				
Agriculture	39	32	30	25
Administrative or professional	56	46	32	26
Family business or activity	26	21	59	49
Smoking				
Never smoker	59	49	121	99
Ex-smoker	23	19	0	0
Current smoker	39	32	1	1
Body mass index ($kg\,m^{-2}$)				
<22	32	26	31	26
22–26	59	48	62	51
>26	31	25	29	24

* Sample size varies due to missing data.

Crude individual mean carotene intake was greatest in winter among all regions. Among the three regions of Honshu, crude individual mean vitamin C intake was greatest in autumn and strongly non-linear, and in Okinawa was greater in summer than in winter (Figs 1 and 2). For both carotene and vitamin C intakes, seasonal variation among the regions of Honshu was similar; therefore adjusted means were reported for the three regions of Honshu combined and Okinawa (Table 2).

Adjusted individual mean carotene and vitamin C intakes by season, region and food group

The pattern of seasonal variation in adjusted mean carotene and vitamin C intakes closely reflected the crude data. Mean carotene intake was 29–41% greater in winter than in the other seasons, after adjusting for age, sex, random differences among family winter means, and random variation in the linear component of random over time (Table 2). Vegetables were the dominant food source of carotene, and intake due to vegetable sources showed similar seasonality to carotene intake from all food sources.

In Honshu, adjusted mean vitamin C intake was 12–35% greater in autumn than in the other seasons, and by

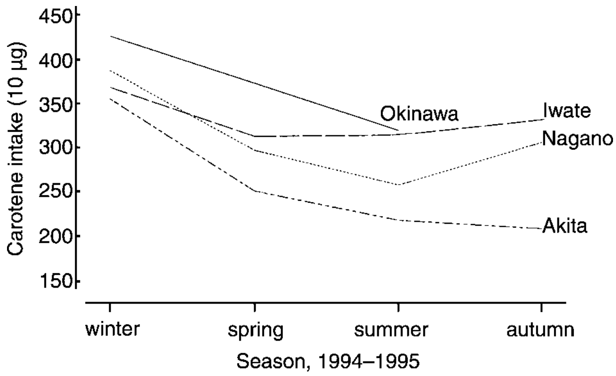

Fig. 1 Crude individual mean carotene intake by season and region, Japan, 1994–1995

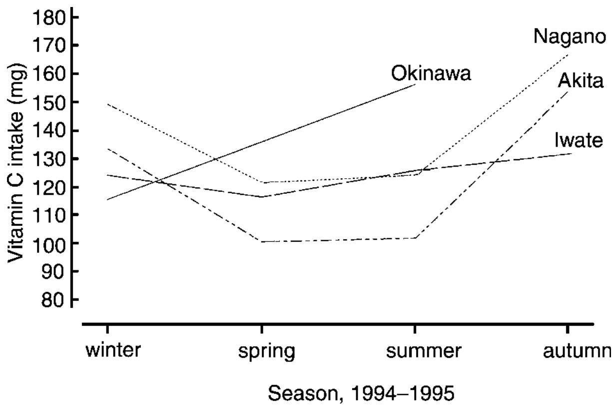

Fig. 2 Crude individual mean vitamin C intake by season and region, Japan, 1994–1995

Table 2 Adjusted* individual mean intakes of dietary carotene and vitamin C by season, region and food group, Japan, 1994–1995

Nutrient, season and food group	Honshu Mean	Honshu 95% CI	Okinawa Mean	Okinawa 95% CI
Carotene, all food sources (10 μg)				
Winter	369	341–396	433	383–482
Spring	287	261–313		
Summer	262	236–288	333	285–381
Autumn	281	254–308		
Carotene, vegetable sources only (10 μg)				
Winter	337	310–364	415	367–463
Spring	261	237–286		
Summer	222	198–246	279	235–323
Autumn	247	222–271		
Vitamin C, all food sources (mg)				
Winter	135	127–142	118	104–132
Spring	112	104–121		
Summer	117	106–128	161	141–182
Autumn	151	137–164		
Vitamin C, vegetable sources only (mg)				
Winter	83	77–89	74	64–84
Spring	70	64–75		
Summer	81	75–87	94	83–105
Autumn	75	68–81		
Vitamin C, fruit sources only (mg)				
Winter	35	31–40	27	19–35
Spring	31	26–36		
Summer	21	14–29	62	48–76
Autumn	61	51–71		

CI – confidence interval.
*Adjusted for the fixed effects of age and sex, and the random intercept and slope.

contrast in Okinawa, was 36% greater in summer than in winter. Vitamin C intake from vegetable sources accounted for half or more of total vitamin C intake and reflected seasonal variation in total intake in Okinawa, but not in Honshu. The seasonal pattern of variation in fruit sources of vitamin C intake was similar to the variation in total intake, in both Honshu and Okinawa, for which autumn and summer intake was greatest, respectively.

Correlation between seasonal and annual average individual intakes, and the magnitude of true variation in intake between individuals

Predicted correlations were of magnitude 0.70 or greater (Table 3), except for vitamin C (in winter), fat (in winter and autumn), thiamin (in autumn) and retinol (whose values may have been underestimated due to random measurement error[5,12]). Correlation with annual average intake of vitamin C was smaller in winter (0.62) than in the other seasons (range: 0.80–0.87), and there was little variation by season among other nutrients, with the exception of retinol.

Predicted true variation in vitamin C intake as a proportion of total variance was highly seasonal and ranged from 45% of total variance in winter to 78% in autumn. As error variance was constant by season, these percentages reflected seasonal changes in true variance. Predicted absolute true variation in vitamin C intake was more than four times greater in autumn (4016 mg^2) than winter (928 mg^2). True variation in thiamin and retinol intake was also seasonal, but among other nutrients it was less seasonal and was greatest in winter or autumn.

Predicted correlations and true variation from random quadratic coefficients models for carotene and vitamin C intakes were similar to the predictions in Table 3. When correlations were predicted for vitamin C from a random linear coefficients model excluding residents of Okinawa, they were 0.70, 0.79, 0.83 and 0.82 for winter through to autumn, respectively.

Contribution of specific foods to total individual nutrient intake

Carrots and spinach together accounted on average for more than 50% of total individual carotene intake by sex, region and season. Among men and women of Honshu, the food contributing most to vitamin C intake in winter to summer (cabbage or tomatoes) accounted on average for 10–15% of total individual vitamin C intake. However, in autumn, *kaki* (persimmon), which was not consumed *at all* in spring or summer, contributed most to intake and accounted on average for 21% of total individual vitamin C intake among men and for 27% among women.

In Okinawa, the food contributing most to vitamin C intake in winter was cabbage, which accounted on average for 21% of total individual vitamin C intake among men and women. However, in summer, a local vegetable, *goya* (bitter gourd), contributed most to intake and accounted on average for 36% of total individual vitamin C intake among men and 26% among women. In winter, *goya* accounted on average for only 0.3% of total individual vitamin C intake among men and for 1.6% among women.

Discussion

Seasonal variation in dietary intake may be attributable in part to eating customs and to the availability of foods. Carrots and spinach accounted for more than half of the total individual carotene intake in all seasons, and the greater intake of carotene in winter appeared to reflect an increase in consumption of the same carotene-rich foods eaten in the other seasons.

However, seasonal variation in vitamin C intake was attributable to both fruit and vegetables, and the foods contributing most to high-season intake (*kaki* in Honshu and *goya* in Okinawa) were eaten rarely or not at all in the other seasons. This finding suggested seasonal availability of important food sources of vitamin C, and individual preferences for these foods may in part explain the poorer correlation of winter intake with annual average intake than either summer or autumn intake, and the greater true variation in intake among the latter seasons.

Table 3 Predicted* correlation between seasonal and annual average individual intakes of energy and nutrients, and true variation (%)† in intake, Japan, 1994–1995

	Season			
Nutrient	Winter	Spring	Summer	Autumn
Carotene				
Predicted correlation	0.78	0.81	0.81	0.77
True variation (%)	59	55	55	58
Vitamin C				
Predicted correlation	0.62	0.80	0.86	0.87
True variation (%)	45	54	68	78
Energy				
Predicted correlation	0.76	0.78	0.79	0.81
True variation (%)	44	48	52	55
Protein				
Predicted correlation	0.77	0.79	0.79	0.77
True variation (%)	52	50	50	52
Total fat				
Predicted correlation	0.68	0.72	0.72	0.69
True variation (%)	44	38	38	45
Total carbohydrate‡				
Predicted correlation	0.80	0.80	0.80	0.80
True variation (%)	51	51	51	51
Calcium				
Predicted correlation	0.74	0.79	0.81	0.78
True variation (%)	54	52	55	61
Sodium				
Predicted correlation	0.78	0.82	0.82	0.77
True variation (%)	62	57	57	61
Retinol				
Predicted correlation	0.66	0.69	0.64	0.50
True variation (%)	52	34	26	36
Iron				
Predicted correlation	0.81	0.82	0.81	0.78
True variation (%)	62	58	56	56
Potassium				
Predicted correlation	0.81	0.85	0.84	0.81
True variation (%)	67	63	63	66
Phosphorus				
Predicted correlation	0.76	0.79	0.80	0.79
True variation (%)	51	50	52	56
Niacin				
Predicted correlation	0.75	0.76	0.76	0.74
True variation (%)	48	45	44	46
Thiamin				
Predicted correlation	0.77	0.76	0.73	0.69
True variation (%)	50	43	38	35
Riboflavin				
Predicted correlation	0.75	0.79	0.80	0.78
True variation (%)	51	50	53	57

* Predicted from the random linear coefficients regression model using equation (4) in the text.
† Predicted as total minus error variation expressed as a percentage of total variation using equation (3) in the text.
‡ For carbohydrate, the random slope was estimated as zero and results were based upon a random intercept (only) model.

Other foods contributing to total vitamin C intake[16] may also provide opportunities for seasonal preferences.

In large cohort studies diet questionnaires are used. There is evidence that subjects do not easily distinguish current from usual intake assessed by questionnaire, and that there is a reporting bias towards current seasonal intake of fruit and vegetables[4]. Similar results were found among participants in the present study who responded to a pilot food-frequency questionnaire that asked about usual intake of vegetables (in the previous year) and high-season intake of fruit. Reported *usual* intake of many frequently consumed vegetables varied between a winter and summer administration of the questionnaire, while there were few differences in the reported high-season intake of fruit by season (unpublished data).

If a questionnaire used among the JPHC Study regions reflects the ranking of individual nutrient intakes by food records, then a seasonal reporting bias towards winter intake of fruit and vegetables could result in relatively poor categorisation of individuals with respect to usual vitamin C intake. One way to avoid this consequence and to reduce seasonal misclassification is to administer the questionnaire in the high seasons of vitamin C intake, for which predicted correlations with annual average intake were greatest. This approach could be difficult in a multi-region study if the high season varies by food and region. An alternative strategy may be to question subjects about their high-season intake of fruit and vegetables, and possibly foods rich in retinol. For the latter approach, it would be important for future studies to demonstrate that subjects can distinguish high-season from current intake.

An additional advantage of assessing high-season intake of foods rich in vitamin C is much greater true variation, since the precision in an estimated association between an exposure at baseline and a subsequently observed outcome is directly proportional to the exposure variance[13,17–21]. For example, consider a case–control study nested within a cohort. The relative efficiency of estimating an association between a continuous exposure and disease in two populations can be expressed as the sample size required in the population with greater, relative to the population with lesser, exposure variance. Under the assumptions that the size of the association and the error in exposure measurement are equal in both populations, the relative sample size required can be expressed as the product of two terms: (1) the ratio of the absolute true variances in the populations with lesser and greater variance, respectively, and (2) the true variance proportion (of total variance) in the population with lesser variance divided by the same proportion in the population with greater variance[20]. The above assumptions are rather strong when different populations are compared, but may be appropriate when one population's exposure is assessed in different seasons. Our results showed that estimation of an association between disease and high-season vitamin C intake in autumn would require only 13% of the sample size required when vitamin C intake is assessed in winter, e.g. 928/4016, as given in the text of the Results section, multiplied by 45/78 from Table 3. The maximum savings in sample size associated with dietary assessment in winter, relative to another season, for fat, thiamin and retinol are 22, 46 and 68%, respectively.

Increased efficiency also can be expected when vitamin C intakes are predicted from diet questionnaire measurements calibrated against the mean of repeated food records. Under the calibration model assumptions[21], in the previous example there will be sample size savings in autumn, compared with winter, as long as the correlation between questionnaire and food record intakes in autumn is greater than approximately half of this correlation in winter (see Appendix B). Savings will exceed 75% if the questionnaire intakes correlate better with food records in autumn than winter (see equation (9) in White *et al.*[20]).

Although data were not collected in the spring or autumn in Okinawa, there was strong evidence that seasonal patterns in average vitamin C intake differed between Honshu and Okinawa. Moreover, there was less seasonal variation among correlations with annual average vitamin C intake, predicted from individuals residing in Honshu only, than in the full sample. This result may indicate that the magnitude of seasonal misclassification in intake varies by region. Greatest misclassification of usual vitamin C intake by seasonal assessment in winter was, however, observed in all regions.

The lower rates of current smoking in the present study, compared with a random sample of younger men (range: 49–59%) and women (range: 4–9%) aged 40–49 years from the same regions[9], were consistent with preferential participation by health-conscious individuals. Thus, the seasonal carotene and vitamin C mean intakes we report may overestimate the true population values, although one would not expect selection bias to have greatly affected the patterns of seasonal variation.

Random coefficients regression models were used to take into account the clustering of dietary data among families and to allow random variation in the linear component of change in intake over time. The inclusion of a random linear coefficient permitted the prediction of correlations and true variation by season, and introducing a non-linear random component did not alter conclusions.

Categorisation of individuals by their usual nutrient intakes is important in cohort studies for the estimation of relative risks, and a previous report has suggested assessing diet in one season or probing for seasonal variation[4]. Our results indicated that, for most nutrients, seasonal misclassification of usual intake was small, but that assessment of vitamin C intake in winter would be a poor strategy. Vitamin C is of interest partly because of its antioxidant effect, and low dietary intake has been associated with the risk of seven types of cancer[22]. Therefore, if subjects are able to distinguish high season from current intake, there appears to be merit in asking them to report high-season intake of fruit and vegetables in the JPHC Study. Moreover, the magnitude of true between-person variation in nutrient intake could bear upon the choice of the season in which to assess food consumption.

Acknowledgements

M. Fahey was supported by a fellowship granted by the Foundation for the Promotion of Cancer Research, Ministry of Health and Welfare, Japan.

The data collection phase of this study was supported by the Japanese Ministry of Health and Welfare's Grants-In-Aid of Cancer Research for the Second Term Comprehensive Ten-Year Strategy for Cancer Control.

The authors thank local staff in the JPHC Study regions and especially the following dietitians for collecting food record data: Yuko Hatakeyama, Mayumi Morimoto, Masako Makino, Manami Nishizawa, Misako Tawada, Noriko Kamiunten and Satako Takamori. We also thank Drs Youji Iitoi, Yasuhiko Iwase and Tosei Takahashi from the Department of Nutrition, Tokyo University of Agriculture, for data management.

References

1 Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, *et al.* Reproducibility and validity of a semi-quantitative food frequency questionnaire. *Am. J. Epidemiol.* 1985; **122**: 51–65.

2 Ziegler RG, Wilcox HB, Mason TJ, Bill JS, Virgo PW. Seasonal variation in intake of carotenoids and vegetables and fruits among white men in New Jersey. *Am. J. Clin. Nutr.* 1987; **45**: 107–14.

3 Hunter DJ, Rimm EB, Sacks FM, Stampfer MJ, Colditz GA, Litin LB, *et al.* Comparison of measures of fatty acids intake by subcutaneous fat aspirate, food frequency questionnaire, and diet records in a population of free living US living men. *Am. J. Epidemiol.* 1992; **135**: 418–27.

4 Subar AF, Frey CM, Harlan LC, Kahle L. Differences in reported food frequency by season of questionnaire administration: the 1987 National Health Interview Survey. *Epidemiology* 1994; **5**: 226–33.

5 Hartman AM, Brown CC, Palmgren J, Pietinen P, Verkasalo M, Myer D, *et al.* Variability in nutrient and food intakes among older middle-aged men. Implications for design of epidemiologic and validation studies using food recording. *Am. J. Epidemiol.* 1990; **132**: 999–1012.

6 Tsugane S, Fahey MT, Sasaki S, Baba S for the JPHC Study Group. Alcohol consumption and all-cause mortality among middle-aged Japanese men: seven-year follow-up of the JPHC Study Cohort I. *Am. J. Epidemiol.* 1999; **150**: 1201–7.

7 Tsugane S, Akabane M, Inami T, Matsushima S, Ishibashi T, Ichinowatari Y, *et al.* Urinary salt excretion and stomach cancer mortality among four Japanese populations. *Cancer Causes Control* 1991; **2**: 165–8.

8 Tsubono Y, Kobayashi M, Tsugane S. Food consumption and gastric cancer mortality in five regions of Japan. *Nutr. Cancer* 1997; **27**: 60–4.

9 Tsugane S, Gey F, Ichinowatari Y, Miyajima Y, Ishibashi T, Matsushima S, *et al.* Cross-sectional epidemiologic study for assessing cancer risks at the population level: II baseline data and correlation analysis. *J. Epidemiol.* 1992; **2**: 83–9.

10 Science and Technology Agency, Japan. *Standard Tables of Food Consumption in Japan*, 4th revised ed. Tokyo, Japan: Ministry of Finance Printing Bureau, 1999 [in Japanese].

11 Beaton GH, Milner J, Corey P, McGuire V, Cousins M, Stewart E, *et al.* Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. *Am. J. Clin. Nutr.* 1979; **32**: 2546–59.
12 Tsubono Y, Fahey MT, Takahashi T, *et al.* Interpopulation and intrapopulation variability of nutrient intake in five reigons of Japan. *Eur. J. Clin. Nutr.* 1998; **52**: 176–9.
13 Diggle PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. Oxford: Clarendon Press, 1994.
14 Longford NT. *Random Coefficient Models*. Oxford: Clarendon Press, 1993.
15 Byrk AS, Raudenbush SW. *Hierarchical Linear Models*. CA: Sage, 1992.
16 Tsubono Y, Takamori S, Kobayashi M, Takahashi T, Iwase Y, Iitoi Y, *et al.* A data-based approach for designing a semiquantitative food frequency questionnaire for a population-based prospective study in Japan. *J. Epidemiol.* 1996; **6**: 45–53.
17 Wynder EL, Hebert JR. Homogeneity in nutrient exposure: an impediment in cancer epidemiology. *J. Natl. Cancer Inst.* 1987; **79**: 605–7.
18 Prentice RL, Pepe M, Self SC. Dietary fat and breast cancer: a quantitative assessment of the epidemiologic literature and discussion of methodologic issues. *Cancer Res.* 1989; **49**: 3147–56.
19 McKeown-Eyssen GE, Tibshirani R. Implications of measurement error in exposure for the sample size of case–control studies. *Am. J. Epidemiol.* 1994; **139**: 415–21.
20 White E, Kushi LH, Pepe MS. The effect of exposure variance and exposure measurement error on study sample size: implications for the design of epidemiologic studies. *J. Clin. Epidemiol.* 1994; **47**: 873–80.
21 Kaaks R, Riboli E. Validation and calibration of dietary intake in the EPIC project: methodological considerations. *Int. J. Epidemiol.* 1997; **26**: S15–25.
22 Riboli E, Norat T. Cancer prevention and diet: opportunities in Europe. *Public Health Nutr.* 2001; **4**: 475–84.

Appendix A

From equation (1), the covariance between any two observations on the kth individual of the ith family for the random linear coefficients model is given by:

$$\mathrm{cov}(y_{ijk}, y_{ilk}) = \sigma_I^2 + LT_j LT_l \sigma_S^2 + (LT_j + LT_l)\sigma_{IS}, \quad \text{(A1)}$$

where $j \neq l$[13]. It follows from equation (A1) that the predicted covariance between a seasonal observation and the mean of $j = 1, \ldots, 4$ observations on an individual is:

$$\begin{aligned}\mathrm{cov}(y_{ijk}, \bar{y}_{i\bullet k}) &= \sigma_I^2 + (3LT_j\sigma_S^2)/2 \\ &\quad + [(3 + 2LT_j)\sigma_{IS}]/2 + \sigma_E^2/4.\end{aligned} \quad \text{(A2)}$$

Similarly, from equation (3) it can be shown that

$$\mathrm{var}(\bar{y}_{i\bullet k}) = \sigma_I^2 + 9\sigma_S^2/4 + 3\sigma_{IS} + \sigma_E^2/4, \quad \text{(A3)}$$

and predicted correlations can be computed from equation (4) using equations (3), (A2) and (A3).

The random quadratic coefficients model was defined by adding an LT_j^2 term to equation (1) with coefficient $\beta_{5i} = u_{5i}$ at the family level. This was analogous to the linear coefficients model and $u_{5i} \sim N(0, \sigma_Q^2)$, where σ_Q^2 is the variance among families due to the quadratic component of change in nutrient intake over time. The random effects u_{0i}, u_{4i}, u_{5i} are assumed to be distributed multivariate normal, and to be independent of $e_{jk(i)}$. We let $\mathrm{cov}(u_{0i}, u_{5i}) = \sigma_{IQ}$ and $\mathrm{cov}(u_{4i}, u_{5i}) = \sigma_{SQ}$. It can be shown that the variance in an observation under this model is

$$\begin{aligned}\mathrm{var}(y_{ijk}) &= \sigma_I^2 + LT_j^2\sigma_S^2 + 2LT_j\sigma_{IS} + LT_j^4\sigma_Q^2 \\ &\quad + 2LT_j^2\sigma_{SQ} + 2LT_j^3\sigma_{SQ} + \sigma_E^2 \end{aligned} \quad \text{(A4)}$$

and that the covariance between any two observations on the same family is given by

$$\begin{aligned}\mathrm{cov}(y_{ijk}, y_{ilk}) &= \sigma_I^2 + LT_jLT_l\sigma_S^2 + (LT_j + LT_l)\sigma_{IS} \\ &\quad + LT_j^2LT_l^2\sigma_Q^2 + (LT_j^2 + LT_l^2)\sigma_{IQ} \\ &\quad + (LT_j^2LT_l + LT_jLT_l^2)\sigma_{SQ} \\ &\quad + (LT_jLT_l^2)\sigma_{SQ}, \end{aligned} \quad \text{(A5)}$$

where $j \neq l$. Equation (A4) can be used in a manner analogous to equation (3) to estimate true variation. Equations (A4) and (A5) can be used to derive $\mathrm{var}(\bar{y}_{i\bullet k})$ and $\mathrm{cov}(y_{ijk}, \bar{y}_{i\bullet k})$ for the random quadratic coefficients model, and equation (4) in the text to compute predicted correlations.

Appendix B

If a nutrient intake estimated from a diet questionnaire administered in two seasons, $Q1$ and $Q2$, is calibrated, respectively, against the mean of repeated seasonal food records, $R1$ and $R2$, then the variance of the predicted intakes can be estimated by $\rho_{Q1R1}^2\sigma_{T1}^2$ and $\rho_{Q2R2}^2\sigma_{T2}^2$, respectively[21], where ρ_{Q1R1}, ρ_{Q2R2} are the deattenuated correlations between the seasonal questionnaires and food record intakes, and σ_{T1}^2, σ_{T2}^2 are the true variations in intake estimated using seasonal food records. In a nested case–control study, there will be savings in sample size by estimating nutrient intake $Q2$, relative to $Q1$, when $\rho_{Q1R1}^2\sigma_{T1}^2/\rho_{Q2R2}^2\sigma_{T2}^2 < 1$ or when $\rho_{Q2R2} > \rho_{Q1R1}\sqrt{\sigma_{T1}^2/\sigma_{T2}^2}$[19,20]. The square root of the variance ratio in the right-hand side of this expression can be estimated from the data reported in the text of the Results section for vitamin C intake in autumn, relative to winter, as $(928/4016)^{1/2}$ or 0.48. When winter was compared with the other seasons, the minimum values of the square root of this variance ratio for fat, thiamin and retinol were 0.88, 0.74 and 0.56, respectively.