

# Reproducibility of Dietary and Other Data From a Self-administered Questionnaire

Shoji Tokunaga,<sup>1</sup> Tomio Hirohata,<sup>1</sup> and Itsuyo Hirohata<sup>2</sup>

<sup>1</sup>Department of Public Health, School of Medicine, Kyushu University, Fukuoka, Japan; <sup>2</sup>Kurume Shin-Ai Women's College

This study examined the reproducibility of information obtained from a questionnaire covering dietary and other information by comparing the answers to the questionnaires with answers to the same questions one year later. Answers from 191 men and 220 women, aged 40 to 65, in Fukuoka prefecture, Japan were compared. The surveys were conducted in November 1989 and November 1990. In the second survey, 97.2% of the first respondents answered. The concordance between the two responses was high, and the differences between the mean intakes from the two surveys were within 5% for 7 food items and within 15% for 14 food items among the 20 items about which the respondents were asked. The intraclass correlation coefficients varied from 0.30 for eggs to 0.62 for milk, with 16 items greater than 0.4. The differences in reproducibility between the sexes and two age categories were not significant. Close values also were obtained for the estimated consumption of salt. A substantially high reproducibility was observed on items regarding drinking and smoking; most of the kappa statistics and the intraclass correlation coefficients were between 0.5 and 0.9. The above reproducibility on individual food items was comparable to or better than those reported from other studies. The results of the present study thus indicate that the self-administered semiquantitative food intake questionnaire used for our cross-sectional study is useful for epidemiologic studies to assess the association between diet and various diseases. In particular, the present questionnaire is highly dependable regarding the overall group intake of foods. The information on smoking, drinking, and other items obtained by this questionnaire also seemed to be sufficiently reliable for epidemiologic surveys. — *Environ Health Perspect* 102(Suppl 8):5–10 (1994)

Key words: reproducibility, self-administered questionnaire, dietary method, gastric cancer, cross-sectional study, salt, intraclass correlation coefficient, kappa

## Introduction

It is well known that dietary habits, smoking, and drinking are highly related to the occurrence of heart disease, cerebrovascular accidents, cancer, diabetes, and other serious diseases. To investigate the association between diet and various diseases, the precise assessment of dietary habits in humans is critical. We developed a self-administered questionnaire for a cross-sectional study in six prefectures in Japan to clarify some of the risk factors for cancer, especially gastric cancer. The self-administered questionnaire was designed primarily to assess dietary habits as well as drinking, smoking, and other practices. The present reproducibility study evaluated the reliability of the information obtained using that

questionnaire. High reproducibility would show that the data of our cross-sectional study are reliable, and would give general indications of whether a dietary assessment using the questionnaire is useful for epidemiologic studies on the relationship between diet and cancer, and possibly diet and some other diseases. This is particularly important for Japan because only a few studies on the reliability of information from a questionnaire have been reported previously in this country.

## Subjects and Methods

The subjects consisted of 411 residents (191 men and 220 women) in Fukuoka prefecture, Japan. The age and sex distribution of the subjects is shown in Table 1. The survey was designed primarily to assess the lifestyle of adults between 40 and 65 years old. They were randomly selected residents in the districts of three health centers in Fukuoka prefecture. We questioned them regarding their lifestyle in November 1989 and again in November 1990 using

the same questionnaire. In the second survey, 97.2% of the first respondents answered the questionnaire.

The 20 food items surveyed are shown in Table 2. We tried to make a comprehensive investigation of dietary habits, and included all food items reported to have an association with gastric cancer. For each food item, we determined the frequency of consumption by asking the number of times each food item was consumed within a given interval. The responses were categorized into no consumption, 1 to 3 times a month, 1 to 3 times a week, 4 to 6 times a week, daily, twice a day, and 3 times a day. The respondents were asked to circle the appropriate box. Next, we asked the respondents about portion size. To accurately estimate portion size, we showed participants a color photograph of the actual portion for each food. The respondents could indicate that the amount of the food consumed was about half the size of the photograph, 70 to 80%, the same, or 1.5 times as much. If consumption did not match the answer choices, then the respondents were asked to specify exactly how large the portion was in comparison to the portion size shown in the photo. The consumption of all food items was then estimated by multiplying the frequency by the portion size.

This paper was presented at the 4th Japan-US Biostatistics Conference on the Study of Human Cancer held 9–11 November 1993 in Tokyo, Japan.

The authors wish to thank the Department of Health, Fukuoka Prefecture, and the three cooperating health centers for helping us with conducting this survey. This study was supported in part by a Grant-in-Aid for Cancer Research from the Ministry of Health and Welfare of Japan (No. 1–6), and a Grant-in-Aid from the Ministry of Education, Science and Culture, Japan (No. 01480207).

Address correspondence to Shoji Tokunaga, Department of Public Health, School of Medicine, Kyushu University, 3-1-1, Maidashi, Fukuoka 812, Japan. Telephone 81 92 641 1151 ext. 3373. Fax 81 92 632 6402. Email e76059@kyu-cc.cc.kyushu-u.ac.jp

**Table 1.** Age and sex distribution of subjects.

	Age		Total (%)
	40–49 (%)	50–65 (%)	
Men	81 (19.7)	110 (26.8)	191 (46.5)
Women	103 (25.1)	117 (28.4)	220 (53.5)
Both sexes	184 (44.8)	227 (55.2)	411 (100.0)

**Table 2.** Mean dietary intakes (g/day), their 95% confidence intervals (CI), and intraclass correlation coefficients (*r*) from two questionnaires.

	Questionnaire 1		Questionnaire 2		Ratio <sup>c</sup>	<i>r</i>
	Mean-1 <sup>a</sup>	95% CI	Mean-2 <sup>b</sup>	95% CI		
Rice	562	536–589	536	510–563	0.95	0.59
Miso soup <sup>d</sup>	136	123–150	124	114–135	0.91*	0.54
Potatoes	22.8	20.7–25.3	21.4	19.3–23.6	0.94	0.43
Green vegetables	33.9	30.7–37.6	33.6	30.6–36.8	0.99	0.45
Raw vegetables	18.9	16.6–21.4	17.8	15.7–20.1	0.94	0.42
Fruits	79.9	67.8–94.2	90.9	81.4–101.6	1.14	0.36
Bean curd	44.9	40.5–49.8	46.3	42.6–50.2	1.03	0.35
Products with soybean	5.10	4.27–6.09	5.83	4.94–6.89	1.14	0.41
Strongly pickled vegetables <sup>e</sup>	9.93	8.66–11.37	7.96	6.87–9.20	0.80*	0.47
Lightly pickled vegetables <sup>d</sup>	9.37	7.90–11.11	7.54	6.33–8.99	0.80	0.46
Plum pickled in salt <sup>e</sup>	1.12	0.95–1.32	1.30	1.11–1.51	1.16	0.55
Milk	19.1	14.4–25.4	30.5	23.5–39.5	1.60*	0.62
Eggs	19.5	16.9–22.6	24.3	22.4–26.4	1.25	0.30
Meats	22.6	20.5–24.8	25.9	23.5–28.6	1.15**	0.52
Fresh fish	42.2	39.0–45.7	42.6	39.4–46.1	1.01	0.47
Salted fish <sup>e</sup>	3.51	3.07–4.01	4.07	3.56–4.66	1.16	0.34
Dried fish <sup>d</sup>	2.38	2.06–2.74	2.28	2.00–2.61	0.96	0.43
Boiled foods in sweetened soy sauce <sup>e</sup>	1.09	0.92–1.29	1.07	0.91–1.25	0.98	0.50
Salted fish eggs <sup>e</sup>	0.841	0.710–0.991	0.823	0.701–0.961	0.98	0.53
Squid pickled in salt, sea-urchin eggs pickled in salt <sup>e</sup>	0.265	0.216–0.320	0.296	0.244–0.356	1.12	0.59
Salt-1 <sup>f</sup>	4.37	4.12–4.63	4.03	3.79–4.28	0.92	0.55
Salt-2 <sup>g</sup>	2.25	2.09–2.41	2.17	2.01–2.34	0.96	0.51

<sup>a</sup>Mean intake from questionnaire 1. <sup>b</sup>Mean intake from questionnaire 2. <sup>c</sup>Ratio = questionnaire 1 mean intake/questionnaire 2 mean intake. <sup>d</sup>Salted food (salt concentration 0.8–5.0%). <sup>e</sup>Highly salted food (salt concentration > 5%). <sup>f</sup>Salt-1 is the estimated intake from all salted foods (*d* and *e*). <sup>g</sup>Salt-2 is the estimated intake of salt from highly salted foods (*e*). \**p* < 0.05, \*\**p* < 0.01 (Wilcoxon rank-sum test).

**Table 3.** Mean dietary intakes (g/day) and intraclass correlation coefficients (*r*) from two questionnaires for males and females.

	Males				Females			
	Mean-1 <sup>a</sup>	Mean-2 <sup>b</sup>	Ratio <sup>c</sup>	<i>r</i>	Mean-1 <sup>a</sup>	Mean-2 <sup>b</sup>	Ratio	<i>r</i>
Rice	687	669	0.97	0.45	473	442	0.93	0.57
Miso soup <sup>d</sup>	140	129	0.92	0.50	132	119	0.90	0.60
Potatoes	19.6	19.2	0.98	0.41	26.1	23.4	0.90	0.44
Green vegetables	32.2	31.3	0.97	0.50	35.6	35.7	1.00	0.40
Raw vegetables	16.8	16.3	0.97	0.41	20.8	19.1	0.92	0.43
Fruits	60.2	69.8	1.16	0.36	102	114	1.12	0.33
Bean curd	48.7	49.1	1.01	0.36	41.8	44.0	1.05	0.33
Products with soybean	4.65	5.25	1.13	0.41	5.54	6.39	1.15	0.42
Strongly pickled vegetables <sup>e</sup>	10.9	8.25	0.76	0.52	9.16	7.71	0.84	0.42
Lightly pickled vegetables <sup>d</sup>	8.88	6.88	0.77	0.44	9.82	8.17	0.83	0.49
Plum pickled in salt <sup>e</sup>	1.29	1.38	1.07	0.54	0.992	1.23	1.24	0.55
Milk	18.5	29.4	1.59	0.61	19.7	31.5	1.60	0.62
Eggs	20.4	25.8	1.26	0.33	18.8	23.1	1.23	0.27
Meats	24.4	26.9	1.10	0.47	21.1	25.1	1.19*	0.56
Fresh fish	48.5	48.2	0.99	0.49	37.4	38.2	1.02	0.44
Salted fish <sup>e</sup>	3.58	4.56	1.27	0.39	3.46	3.69	1.07	0.28
Dried fish <sup>d</sup>	2.69	2.38	0.88	0.40	2.14	2.20	1.03	0.45
Boiled foods in sweetened soy sauce <sup>e</sup>	1.28	1.08	0.84	0.47	0.944	1.05	1.11	0.52
Salted fish eggs <sup>e</sup>	1.04	0.875	0.84	0.52	0.695	0.780	1.12	0.53
Squid pickled in salt, sea-urchin eggs pickled in salt <sup>e</sup>	0.388	0.446	1.15	0.61	0.181	0.197	1.09	0.52
Salt-1 <sup>f</sup>	4.70	4.32	0.92	0.58	4.10	3.79	0.92	0.51
Salt-2 <sup>g</sup>	2.46	2.35	0.96	0.52	2.07	2.03	0.98	0.50

<sup>a</sup>Mean intake from questionnaire 1. <sup>b</sup>Mean intake from questionnaire 2. <sup>c</sup>Ratio = questionnaire 1 mean intake/questionnaire 2 mean intake. <sup>d</sup>Salted food (salt concentration 0.8–5.0%). <sup>e</sup>Highly salted food (salt concentration > 5%). <sup>f</sup>Salt-1 is the estimated intake from all salted foods (*d* and *e*). <sup>g</sup>Salt-2 is the estimated intake of salt from highly salted foods (*e*). \**p* < 0.05 (Wilcoxon rank-sum test).

Since the distribution of the intake values of most food items was skewed toward higher values, all food intake values were log<sub>10</sub> transformed before analysis. Some people reported no consumption for uncommon food items, so 1.0 was added to all values before transformation. Although the resulting distributions were not strictly normal, they did not include any serious outliers. The mean food intakes calculated by the two questionnaires were then compared after the log transformation. The homogeneity of the distributions of the repeated questionnaire was then evaluated by the Wilcoxon rank-sum test for each food item. We computed the intraclass correlation coefficient as a measure of reproducibility for dietary consumption calculated from the two questionnaires. The coefficient quantifies the extent of the within-subject agreements relative to the between-subject agreements (*I*). For all categories except dietary intake, we calculated the proportion of agreement. The intraclass correlation coefficients also were computed for ordinal data such as the amount of alcohol consumption and the number of cigarettes smoked. The reproducibility of the categorical data was evaluated by the kappa statistic. Kappa is the proportion of agreement after chance agreement is removed (*2*).

The following guidelines were used to evaluate kappa: below 0.20 indicated slight agreement, 0.21 to 0.40 fair, 0.41 to 0.60 moderate, 0.61 to 0.80 substantial and 0.81 to 1.00 almost perfect (*3*). Although there is still some controversy as to whether or not this guideline, along with a similar one (*4*) for the range of acceptability, is adequately informative (*5*), we used it because there seems to be no appropriate alternative. Since the intraclass correlation coefficient is analogous to kappa (*6*), this guideline was also used for the intraclass correlation coefficient. Because the purpose of this study was to quantify reproducibility rather than to test the null hypothesis of agreement by chance, the *p* values were not presented.

The homogeneity of the intraclass correlation coefficients across the groups of subjects was tested with the *t*-approximation approach described by Kraemer (*7*). The statistical analyses were conducted using SAS, SAS Institute (Cary, NC) at the Computer Center, Kyushu University. The significance level of all statistical tests was 0.05.

**Results**

Table 2 shows the means, their 95% confidence intervals, the ratios of means, and intraclass correlation coefficients regarding intakes of 20 food items and the consumption of salt; the latter was estimated in two ways. The two questionnaire surveys produced closely similar results. The differences in the mean intakes from the two surveys were within 5% for 7 food items and within 15% for 14 items. (Table 2 shows the ratio of the two means.) The intakes from the two surveys, tested with the Wilcoxon rank-sum test, were not significantly different for 16 items (two-sided, the significance level was 0.05.)

The intraclass correlation coefficients varied from 0.30 for eggs to 0.62 for milk, with 16 items greater than 0.4. The mean ( $\pm$ SE) of the correlation coefficient was 0.47 ( $\pm$ 0.02). The correlation coefficients showed a moderate reproducibility for most food items.

We estimated consumption of salt in two ways. Estimation 1 represented the intake of salt from all salted foods listed on the questionnaire, while estimation 2 was from highly salted foods only, such as strongly-pickled vegetables and salted fish. They are shown as Salt-1 and Salt-2 in Table 2. The differences in the mean intakes at the two surveys were 8% for Salt-1 and 4% for Salt-2. Their intraclass correlation coefficients were 0.55 and 0.51, respectively. Thus, the two surveys provided similar values for the consumption of salt, and the reproducibility was moderate.

It is also of interest to see whether or not dietary reproducibility differs by gender. The mean and the intraclass correlation coefficient for the intake of the 20 food items listed and the estimated consumption of salt are shown separately for men and women in Table 3. The reproducibility of the mean food intake was similar between men and women. Of the 20 food items, the difference in the mean intake from the two surveys was less than 15% for 12 food items in men, and for 13 food items in women. The gender difference in the reproducibility of the mean intake of salt was also small. For men and women, the mean salt intake from the first questionnaire and the second were almost the same.

The difference in the intraclass correlation coefficients between the sexes was also small for most food items. The intraclass correlation coefficients of 13 items were greater than 0.4 in men while those of 15 items were greater than 0.4 in women. For the intake of all food items and salt, no significant difference was found between men

**Table 4.** Mean dietary intakes (g/day) and intraclass correlation coefficients (*r*) from two questionnaires for two age groups.

	Age 40-49				Age 50-65			
	Mean-1 <sup>a</sup>	Mean-2 <sup>b</sup>	Ratio <sup>c</sup>	<i>r</i>	Mean-1 <sup>a</sup>	Mean-2 <sup>b</sup>	Ratio <sup>c</sup>	<i>r</i>
Rice	579	546	0.94	0.53	549	527	0.96	0.65
Miso soup <sup>d</sup>	124	116	0.94	0.59	146	131	0.90	0.50
Potatoes	21.1	21.1	1.00	0.35	24.4	21.6	0.89	0.49
Green vegetables	31.0	30.3	0.98	0.55	36.5	36.5	1.00	0.37
Raw vegetables	18.9	17.7	0.94	0.45	18.8	17.8	0.95	0.40
Fruits	71.4	87.9	1.23	0.30	87.7	93.5	1.07	0.42
Bean curd	42.6	45.8	1.08	0.36	46.9	46.7	1.00	0.34
Products with soybean	5.05	5.66	1.12	0.35	5.14	5.98	1.16	0.46
Strongly pickled vegetables <sup>e</sup>	10.37	6.85	0.66*	0.44	9.58	8.99	0.94	0.48
Lightly pickled vegetables <sup>d</sup>	8.31	6.20	0.75	0.39	10.3	8.83	0.86	0.51
Plum pickled in salt <sup>e</sup>	1.05	1.23	1.17	0.53	1.18	1.35	1.14	0.56
Milk	18.2	38.1	2.09**	0.50	19.9	25.4	1.28	0.69
Eggs	19.6	23.8	1.21	0.29	19.5	24.8	1.27	0.30
Meats	29.0	32.2	1.11	0.40	18.4	21.7	1.18*	0.55
Fresh fish	41.7	40.4	0.97	0.56	42.7	44.5	1.04	0.40
Salted fish <sup>e</sup>	4.22	4.21	1.00	0.40	3.03	3.97	1.31	0.29
Dried fish <sup>d</sup>	2.86	2.49	0.87	0.42	2.04	2.13	1.04	0.43
Boiled foods in sweetened soy sauce <sup>e</sup>	1.11	0.983	0.89	0.51	1.07	1.14	1.07	0.49
Salted fish eggs <sup>e</sup>	1.03	0.943	0.92	0.55	0.708	0.735	1.04	0.50
Squid pickled in salt, sea-urchin eggs pickled in salt <sup>e</sup>	0.345	0.418	1.21	0.53	0.209	0.217	1.04	0.63
Salt-1 <sup>f</sup>	4.24	3.69	0.87*	0.56	4.48	4.33	0.97	0.53
Salt-2 <sup>g</sup>	2.27	2.00	0.88	0.56	2.22	2.32	1.05	0.48

<sup>a</sup>Mean intake from questionnaire 1. <sup>b</sup>Mean intake from questionnaire 2. <sup>c</sup>Ratio=questionnaire 1 mean intake/questionnaire 2 mean intake. <sup>d</sup>Salted food (salt concentration 0.8-5.0%). <sup>e</sup>Highly salted food (salt concentration > 5%). <sup>f</sup>Salt-1 is the estimated intake from all salted foods (*d* and *e*). <sup>g</sup>Salt-2 is the estimated intake of salt from highly salted foods (*e*). \**p*<0.05, \*\**p*<0.01 (Wilcoxon rank-sum test).

and women in the intraclass correlation coefficients. In addition, the mean of the intraclass correlation coefficients for 20 food items did not differ significantly between the sexes (matched paired *t*-test).

In Table 4, the means of the food and salt intakes and the intraclass correlation coefficients from the repeated questionnaires are shown separately for age groups 40 to 49 and 50 to 65. The reproducibility of the mean food intakes was similar between the two age groups. Regarding the 20 food items, the difference in the mean intakes from the two surveys was less than 15% for 13 food items in the 40 to 49 group, and for 15 food items in the 50 to 65 group. For both estimations of salt, the difference in the mean consumption was 12 to 13% for the 40 to 49 group, whereas the difference was less than 5% for the 50 to 65 group.

The difference in the intraclass correlation coefficients was similar for the two age groups for most food items. The intraclass correlation coefficients of 14 items were greater than 0.4 in the 40 to 49 group, while those of 16 items were greater than 0.4 in the 50 to 65 group. For some food items, however, a significant difference in the intraclass correlation coefficient was

found between the two age groups. The intraclass correlation coefficients of green vegetables and fresh fish for the 40 to 49 group were significantly higher than those for the 50 to 65 group (*p*<0.05), whereas the correlation coefficient of milk for the 50 to 65 group was significantly higher than that for the 40 to 49 group (*p*<0.01). The difference in the mean intraclass correlation coefficients between the age groups was not significant (*p*>0.05, matched paired *t*-test). On the whole, the reproducibility was similar between the two age groups.

Table 5 shows the reproducibility of the items concerning drinking alcohol and smoking. Reproducibility was presented separately for men and women. The agreement for drinking was approximately 0.9 and its kappa was greater than 0.7 for both men and women. For most types of alcoholic beverages, the values of the kappa were between 0.4 and 0.7. The intraclass correlation coefficients for the amount of drinking and the age when individuals first started to drink were between 0.57 to 0.72. For smoking, the kappa values were 0.8 and 0.89 for men and women, respectively. These data demonstrated a substantially high to almost perfect reproducibility

**Table 5.** Proportion of agreement, kappa, and intraclass correlation coefficients (*r*) for two questionnaires on drinking and smoking.

Item	Males				Females			
	<i>n</i>	Proportion of agreement	Kappa	<i>r</i>	<i>n</i>	Proportion of agreement	Kappa	<i>r</i>
Liquor drinking habit	190	0.91	0.77		218	0.90	0.73	
Sake	142	0.83	0.64		41	0.78	0.56	
Distilled liquor (Shochu)	142	0.82	0.61		41	0.88	0.38	
Distilled liquor (Awamori)	142	0.99	— <sup>a</sup>		41	1.00	1.00	
Beer	142	0.75	0.49		41	0.78	0.38	
Whiskey	142	0.85	0.13		41	0.90	0.66	
Miscellaneous	142	0.98	— <sup>a</sup>		41	0.93	— <sup>a</sup>	
Amount of drinking	135	0.57		0.57	41	0.78		0.60
Age at first drink	135	0.39		0.59	38	0.37		0.72
Smoking habit	189	0.88	0.80		219	0.99	0.89	
Year when last smoked	41	0.27		0.82	— <sup>b</sup>	— <sup>b</sup>	— <sup>b</sup>	
Number of cigarettes smoked	143	0.51		0.61	13	0.54		0.81
Age when first began smoking	136	0.40		0.80	13	0.39		0.76

<sup>a</sup>Kappa not shown due to large variance. <sup>b</sup>Sample size too small to calculate.

**Table 6.** Proportion of agreement, kappa, and intraclass correlation (*r*) for two questionnaires on miscellaneous items.

Item	<i>n</i>	Proportion of agreement	Kappa	<i>r</i>
Blood recipient or not	409	0.97	0.74	
Age when blood transfusion received	21	0.48		0.63
Prefecture of residence during childhood	407	0.99		
Ever employed	403	0.93	0.48	
Type of job	253	0.81		

regarding drinking and smoking.

On other miscellaneous items, reproducibility was also analyzed and is shown in Table 6. The items on blood transfusion showed moderate to substantial reproducibility. The kappa for previous blood transfusions received was 0.74, while the intraclass correlation coefficient for the age at the time of receiving a blood transfusion was 0.63.

## Discussion

As stated, the precise measurement of dietary habits is by no means an easy task. Therefore, many investigators have examined reproducibility to check the reliability of dietary information using various methods of assessments (8–21). Unfortunately, only a few studies have yet been reported on this subject in Japan (10, 22); and these studies have been greatly limited in that the dietary questions were very simple and only the frequency of dietary intake usually was assessed. The present study is probably the first reproducibility study conducted in Japan using a comprehensive, semiquantitative questionnaire. Hankin et al. (11,12) examined reproducibility among Japanese-Americans in Hawaii; however the dietary habits of Japanese-Americans

may be substantially different from those of indigenous Japanese living in Japan.

In this study, we assumed that the dietary habits in Fukuoka prefecture are stable between the two questionnaires. Although the dietary habits were not surveyed specifically in Fukuoka prefecture from 1989 to 1990, the Ministry of Welfare in Japan surveys the dietary and nutritional intake of Japanese on an annual nationwide scale known as the National Nutritional Survey (NNS). According to the NNS (23,24), the mean ( $\pm$  SE) ratio of consumption of some food items common to the food items used in this study in 1989 and 1990 was 0.99 ( $\pm$  0.08) in the northern Kyushu district where the study took place. Thus, the respondents' dietary habits during the 2 years seemed to be the same; if they changed, the magnitude is probably very slight.

The present study was concerned primarily with the reproducibility of studies of individual food intake; similar studies have been reported elsewhere (8–10,13,19). All previous authors adopted different methods, such as interviewing the subjects or using self-administered questionnaires. Some questionnaires used color pictures to show portion size and some merely stated a

unit size. For the sake of comparison, we computed Pearson's correlation coefficient (*r*), Spearman's *r*, and Kendall's *r* whenever necessary.

Nomura et al. (14) reinterviewed 109 men using the same questionnaire after a 6 month interval and 111 men after a 2 year interval. For 19 food items common to those used in our study, the means ( $\pm$  SE) of Kendall's *r* were 0.32 ( $\pm$  0.04) for the 6-month group and 0.29 ( $\pm$  0.03) for the 2-year group, while the present study showed that the mean Kendall's *r* was 0.39 ( $\pm$  0.02).

Fukao et al. (10) sent out the same self-administered questionnaire on food consumption to 3157 men and women after an interval of 5 month. The means ( $\pm$  SE) of the proportion of exact agreement and kappa were 0.43 ( $\pm$  0.02) and 0.20 ( $\pm$  0.02), respectively, for the 19 food items that overlapped with ours. Since Fukao et al. surveyed only the frequency of food intake, we also calculated the proportion of exact agreement and kappa for frequency of consumption. In the present study, the means ( $\pm$  SE) of the proportion of exact agreement and kappa were 0.46 ( $\pm$  0.02) and 0.25 ( $\pm$  0.02), respectively, and showed better agreement than their studies.

Salvini et al. (19) evaluated reproducibility on a self-administered food frequency questionnaire administered twice at a 12 month interval for 173 women from the Nurses' Health Study. In their questionnaire, a standard serving size was specified for each food. Of the 24 food items that overlapped with ours, the mean ( $\pm$  SE) of Pearson's *r* was 0.51 ( $\pm$  0.02), which showed a slightly better agreement than ours, 0.47 ( $\pm$  0.02). The subjects in the study by Salvini et al., however—all nurses—were well educated and this factor may be associated with the higher reproducibility. Colditz et al. (9) reported the reproducibility of a semiquantitative food frequency questionnaire completed by the participants in the Nurses' Health Study and repeated 9 months later. For 24 food items common to those used in the present study, the mean Pearson's *r* ( $\pm$  SE) was 0.52 ( $\pm$  0.02), a slightly higher value than ours, 0.49 ( $\pm$  0.02).

Bueno de Mesquita et al. (8) tested the reproducibility of a food frequency questionnaire used in a case-control study. They conducted two studies 12 months apart, each using 63 men and 54 women. The mean ( $\pm$  SE) of the Pearson's *r* for the 18 food items overlapping with our study was 0.64 ( $\pm$  0.04) and 0.64 ( $\pm$  0.03). These coefficients are much higher than ours, probably because the subjects were inter-

viewed by the same trained dietitian and were assisted by other household members.

Jacobsen and Bønaa (13) assessed the reproducibility of a self-administered questionnaire by comparing answers from 201 men and women to questions in two surveys approximately 1 year apart. The reproducibility expressed by Spearman's  $r$  was 0.61 and the proportion of agreement in the consumption category was 0.63 for six food items, which were comparable to ours. In our study, the means ( $\pm$  SE) of Spearman's  $r$  and the proportion of exact agreement were 0.49 ( $\pm$  0.02) and 0.46, respectively. The men and women included in Jacobsen and Bønaa's study were selected from the general population, were willing to attend a health screening, and later were selected because of previously unrecognized mild hypertension and a higher-than-desirable total serum cholesterol value. These selection factors also may be associated with higher reproducibility.

Overall, the reproducibility obtained by our study was better than or comparable to other studies, even though the subjects were randomly selected from the general population. Color photographs of actual portion sizes used in our questionnaire probably contributed to the high reproducibility.

One disadvantage of this study is that the estimation of nutrient intake was not done due to practical reasons. Only salt intake was estimated, and the intraclass correlation coefficients for the two estimations were 0.55 and 0.51, respectively. Rimm et al. (18) and Pietinen et al. (15) reported that the reproducibility on sodium intake was 0.72 and 0.70, respec-

tively. However, their studies are not comparable to ours because the former was conducted on selected subjects—male health professionals—while the latter examined the reproducibility of the questionnaire distributed three different times.

Among the studies dealing with the reproducibility of nutrients (8,11,12,18,25), the intraclass correlation coefficients, Pearson's  $r$ , and Spearman's  $r$ , ranged from 0.12 to 0.87 and most fell into a range between 0.5 and 0.6. As stated, our study was not intended to estimate nutrient intake. With regard to salt intake, the intraclass correlation coefficients, Pearson's  $r$  and Spearman's  $r$  ranged from 0.49 to 0.56 for the two estimates of salt intake in our study.

Several other studies have investigated the reproducibility of responses for drinking alcoholic beverages and smoking. Kelly et al. (26) reported the reliability, expressed by kappa, of the answers to questions on the current status of cigarette smoking and alcohol drinking. In their study, the kappa values for cigarette smoking and alcohol drinking were 0.81 and 0.47, respectively. The intraclass correlation coefficient for number of cigarettes smoked per day was 0.69, and for years since stopping smoking was 0.86. Bueno de Mesquita et al. (8) reported that the reproducibility (Pearson's  $r$ ) of the daily intake of alcoholic beverages was 0.96 and 0.78 in their two studies. In the present study, the reproducibility (kappa) of drinking was 0.77 for men and 0.73 for women, while that of smoking was 0.80 for men and 0.89 for women. The intraclass correlation coefficients on

amount of drinking, number of cigarettes smoked, and year the subject last smoked ranged from 0.57 to 0.82. These high values of reproducibility were also comparable to those of the above studies.

In the current study, the reproducibility of the question on previous blood transfusion received also was evaluated. Blood transfusions have been done only rarely in recent years in Japan. Therefore, even if some subjects may have received a blood transfusion after the first questionnaire survey, the number would be very small. Other items, such as working status, type of job, and smoking and drinking history, may change after the first survey. Thus, the reproducibility on these items reported in the current study would be the minimum estimates.

In summary, the mean values of the food intake obtained by the two questionnaire surveys were close, in spite of the 1 year interval and the nature of the self-administered questionnaire. The intraclass correlation coefficients showed a moderate agreement for individual food intakes and estimated salt intakes. These results thus indicate that the semiquantitative food intake questionnaire used for our cross-sectional study is useful for assessing dietary intake and, in particular, for evaluating the dietary habits of the group. For most items regarding drinking, smoking, and blood transfusion, the reproducibility was substantially high, and thus was considered reliable when used in conjunction with the epidemiologic survey.

## REFERENCES

- Lee J. Alternate approaches for quantifying aggregate and individual agreements between two methods for assessing dietary intakes. *Am J Clin Nutr* 33:956–964 (1980).
- Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20:37–46 (1960).
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 33:159–174 (1977).
- Fleiss JL. *Statistical Methods for Rates and Proportions*. 2nd ed. New York: John Wiley & Sons, 1981.
- Seigel DG, Podgor MJ, Remaley NA. Acceptable values of kappa for comparison of two groups. *Am J Epidemiol* 135:571–578 (1992).
- Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 33:613–619 (1973).
- Kraemer CK. On estimation and hypothesis testing problems for correlation coefficients. *Psychometrika* 40:473–485 (1975).
- Bueno de Mesquita HB, Smeets FWM, Runia S, Hulshof KFAM. The reproducibility of a food frequency questionnaire among controls participating in a case-control study on cancer. *Nutr Cancer* 18:143–156 (1992).
- Colditz GA, Willett WC, Stampfer MJ, Sampson L, Rosner B, Hennekens CH, Speizer FE. The influence of age, relative weight, smoking, and alcohol intake on the reproducibility of a dietary questionnaire. *Int J Epidemiol* 16:392–398 (1987).
- Fukao A, Shimizu H, Maewawa M, Hisamichi, S. The reproducibility of answer to the questionnaire for food intake frequency in a survey (in Japanese). *Jpn J Publ Health* 37:347–352 (1990).
- Hankin JH, Nomura AMY, Lee J, Hirohata T, Kolonel LN. Reproducibility of a diet history questionnaire in a case-control study of breast cancer. *Am J Clin Nutr* 37:981–985 (1983).
- Hankin JH, Yoshizawa CN, Kolonel LN. Reproducibility of a diet history in older men in Hawaii. *Nutr Cancer* 13:129–140 (1990).
- Jacobsen BK, Bønaa KH. The reproducibility of dietary data from a self-administered questionnaire. The Tromsø study. *Int J Epidemiol* 19:349–353 (1990).
- Nomura A, Hankin JH, Rhoads GG. The reproducibility of dietary intake data in a prospective study of gastrointestinal cancer. *Am J Clin Nutr* 29:1432–1436 (1976).
- Pietinen P, Hartman AM, Haapa E, Räsänen L, Haapakoski J,

- Palmgren J, Albanes D, Virtamo J, Huttunen JK. Reproducibility and validity of dietary assessment instruments. 1. A self-administered food use questionnaire with a portion size picture booklet. *Am J Epidemiol* 128:667-676 (1988).
16. Pietinen P, Hartman AM, Haapa E, Räsänen L, Haapakoski J, Palmgren J, Albanes D, Virtamo J, Huttunen JK. Reproducibility and validity of dietary assessment instruments. 2. A qualitative food frequency questionnaire. *Am J Epidemiol* 128:667-676 (1988).
  17. Reshef A, Epstein LM. Reliability of a dietary questionnaire. *Am J Clin Nutr* 25:91-95 (1972).
  18. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *Am J Epidemiol* 135:1114-1126 (1992).
  19. Salvini S, Hunter DJ, Sampson L, Stampfer MJ, Colditz GA, Rosner B, Willett WC. Food-based validation of a dietary questionnaire: the effects of week-to-week variation in food consumption. *Int J Epidemiol* 18:858-867 (1989).
  20. Wu ML, Whittemore AS, Jung DL. Errors in reported dietary intakes. I. Short-term recall. *Am J Epidemiol* 124:826-835 (1986).
  21. Wu ML, Whittemore AS, Jung DL. Errors in reported dietary intakes. 2. Long-term recall. *Am J Epidemiol* 128:1137-1145 (1988).
  22. Kubo N, Ohno Y. Reproducibility of life-style information obtained from a self-administered questionnaire. *Jpn J Cancer Clin* 37:217-223 (1991).
  23. The Ministry of Welfare, Japan. Report of the National Nutritional Survey in 1989. Tokyo:Daiichi-Shuppan, 1991.
  24. The Ministry of Welfare, Japan. Report of the National Nutritional Survey in 1990. Tokyo:Daiichi-Shuppan, 1992.
  25. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 122:51-65 (1985).
  26. Kelly JP, Rosenberg L, Kaufman DW, Shapiro S. Reliability of personal interview data in a hospital-based case-control study. *Am J Epidemiol* 131:79-90 (1990).