RESEARCH ARTICLE

Hospital Outpatients are Satisfactory for Case-control Studies on Cancer and Diet in China: A Comparison of Population Versus Hospital Controls

Lin Li, Min Zhang*, C D'Arcy J Holman

Abstract

Background: To investigate the internal validity of a food-frequency questionnaire (FFQ) developed for use in Chinese women and to compare habitual dietary intakes between population and hospital controls measured by the FFQ. Materials and Methods: A quantitative FFQ and a short food habit questionnaire (SFHQ) were developed and adapted for cancer and nutritional studies. Habitual dietary intakes were assessed in 814 Chinese women aged 18-81 years (407 outpatients and 407 population controls) by face-to-face interview using the FFQ in Shenyang, Northeast China in 2009-2010. The Goldberg formula (ratio of energy intake to basal metabolic rate, EI/BMR) was used to assess the validity of the FFQ. Correlation analyses compared the SFHQ variables with those of the quantitative FFQ. Differences in dietary intakes between hospital and population controls were investigated. Odds ratios (ORs) and 95% confidence intervals (CIs) were obtained using conditional logistic regression analyses. Results: The partial correlation coefficients were moderate to high (0.42 to 0.80; all p<0.05) for preserved food intake, fat consumption and tea drinking variables between the SFHQ and the FFQ. The average EI/BMR was 1.93 with 88.5% of subjects exceeding the Goldberg cut-off value of 1.35. Hospital controls were comparable to population controls in consumption of 17 measured food groups and mean daily intakes of energy and selected nutrients. Conclusions: The FFQ had reasonable validity to measure habitual dietary intakes of Chinese women. Hospital outpatients provide a satisfactory control group for food consumption and intakes of energy and nutrients measured by the FFQ in a Chinese hospital setting.

Keywords: Food-frequency questionnaire - validity - Chinese women - control selection

Asian Pacific J Cancer Prev, **14 (5)**, 2723-2729

Introduction

Diet is one of the most important contributing factors to cancer risk, being ranked second only to tobacco smoking (Akhter et al., 2009). Epidemiological studies of diet and disease rely on the accurate determination of dietary intakes and subsequent estimates of nutrient intakes. Food-frequency questionnaires (FFQs) have been considered an appropriate method of dietary assessment in nutritional epidemiology studies because they measure average, long-term, habitual dietary intakes (Willett 1998). A recurrent issue in dietary self-reports, however, is the extent to which participants underreport their energy intake, and this is a particular problem when assessing habitual diet (Cook et al., 2000). Underreporting of energy intake may be caused by lack of precision in the assessment instrument (i.e. not enough food items in the FFQ) or by the respondents' lack of motivation to report their intakes accurately (Johansson et al., 2001).

The validity of FFQs is a material issue in hospital-based case-control studies of dietary risk factors and cancers, because many causes of hospitalization are associated with or may cause selective dietary patterns. Despite this challenge to validity, few studies have examined whether dietary intakes measured in hospital controls tend to depart systematically from those in population controls, given that correct control selection is crucial to the internal validity of case-control studies (Miettinen 1985; Miettinen 1990; Wacholder et al., 1992; Grimes et al., 2005; Rothman et al., 2008). Concern generally exists about the potential for bias introduced by hospital controls in case-control studies, because they are susceptible to both information and selection biases (Miettinen 1985; Miettinen 1990; Wacholder et al., 1992; Grimes et al., 2005; Rothman et al., 2008). When we applied successfully to Australia's National Health and Medical Research Council to perform case-control studies in China on dietary risk factors for the incidence rates of colorectal cancer, breast cancer and adult leukaemia, it became a condition of funding that we were to recruit

*School of Population Health, Faculty of Medicine, Dentistry and Health Sciences, University of Western Australia , Perth *For correspondence: min.zhang@uwa.edu.au*

both hospital and population controls for around one fifth of the case series to determine if there was any difference of practical importance.

A handful of studies have compared dietary intakes between the two types of control group in western countries, where non-emergency access to hospital care generally depends on referral from medical practitioners in non-institutional settings (GonzÁLez et al., 1990; Amadori et al., 1995; Almendingen et al., 2001; Malagoli et al., 2008). Some studies have specifically reported differences between hospital and population controls in dietary exposure distributions (Amadori et al., 1995; Almendingen et al., 2001; Malagoli et al., 2008), but elsewhere the two control groups were found to have quite similar dietary exposure distributions (GonzÁLez et al., 1990; Inoue et al., 1997).

The two aims of this study were, firstly, to assess the internal validity of a FFQ for use in Chinese women in case–control studies on cancers and, secondly, to compare habitual dietary intakes measured by the FFQ between two control groups: one selected from the general population and one selected from outpatients attending the same hospital as the cases, to determine if there was any difference of practical importance.

Materials and Methods

Study design and participants

The study, as a validation component of three large case-control studies of malignancies, was conducted in Shenyang, the capital city of Liaoning Province, Northeast China between August 2009 and July 2010. For each incidence cancer case of permanent residents of urban Shenyang, two controls were selected: one from the same hospital where the case was identified and one from population in the hospital catchment area to separately match cases in a 1:1 ratio (Li et al., 2011).

The methods of recruiting population controls were similar to those used in case–control studies in Shanghai, China (Chang et al., 2008; Hsing et al., 2008). Population household registries, which kept records of all permanent residents in urban Shenyang, were used to select controls from the five major metropolitan districts in Shenyang, namely Heping, Shenhe, Dadong, Huanggu, and Tiexi. With the assistance of the local community councils, residents who lived at their registered address during the study period were randomly selected from household registry rolls. Residents were eligible as population controls if they matched to individual cases by gender and year-of-birth quinquennium on a given selection day.

Hospital controls were drawn from the population of outpatients in the hospital (Porta et al., 2008). A systematic selection process used in our previous studies was adopted for hospital control recruitment (Zhang et al., 2007). They were selected from outpatients who attended the Medical Examination Centre at the First Hospital of China Medical University, a public teaching hospital with 2,249 beds, around 32,000 inpatients annually and 3,000 outpatients daily, and were permanent residents of urban Shenyang. The eligible hospital controls were those without any malignancy after they had consulted their doctors. Each hospital outpatient control was selected as the first attendee on a given selection day to match the next case on a daily updated list of cases by sex and 5-year age group. Hospital outpatient controls were excluded if they had a diagnosis of any malignancy after the recruitment. Interviews were completed for 407 (90.0%) of the 452 eligible female population controls who were approached to participate. Of 423 eligible female hospital outpatient controls who were approached to participate, interviews were completed also for 407 (96.2%). The project protocol had received ethics approval from both the Human Research Ethics Committee of The University of Western Australia and the First Hospital of China Medical University authority.

Questionnaire and interview

Subjects were briefed regarding confidentiality and anonymity issues and the general aims of the study to investigate lifestyle factors and habitual dietary intakes. An appointment for an interview was made after obtaining the respondent's consent via an initial contact. A face-to-face interview was then conducted by the first author, using a structured questionnaire and usually took 30-40 minutes. The structured questionnaire, available from the authors upon request, was used to collect the information on: (i) demographic and lifestyle characteristics, e.g., area of residence, education, smoking, alcohol and tea consumption and physical activity; (ii) habitual dietary intakes assessed by a 119-food frequency questionnaire (FFQ); and (iii) factors relevant to hormonal status and family history of cancer. The questionnaire was adapted from that used in our previous studies on cancers (Zhang et al., 2009). This instrument was originally modified from one used for studying cancers in Shanghai in order to ensure cultural relevance (Ji et al., 1998). The questionnaire was translated into Chinese and checked using back-translation by professional Chinese translators. The internal consistency and reliability of the questionnaire was assessed in a preliminary study and then evaluated by a test-retest. For the reproducibility of the FFQ, the intraclass correlation coefficients for mean daily intakes of food were 0.81 (Zhang et al., 2005).

To assess predictive validity of mean dietary intakes derived from the FFQ, a short food habit questionnaire (SFHQ) soliciting categorical information on food habits was also administered. The SFHQ contained items on total preserved food (cured food) intake, which was classified into four levels: never or seldom, once a month, once a week and every day. Fat consumption was described as never or seldom, sometimes and often. Information on tea drinking frequency and new batches of tea were classified into three levels: never or seldom, ≤6 times a week, and ≥1 times per day.

Statistical analysis

All data were checked at the end of each interview for completeness and were coded and analysed using SPSS version 18.0. Participants' self-reported current height in meters and weight in kilograms were used to calculate body mass index (BMI) (weight/height^2). Daily energy intake and alcohol consumption were assessed using the FFQ. The frequencies of 119 food items, including beer, wine, and liquor intake, were assigned into nine categories: never or hardly ever; once a month; 2-3 times a month; once a week; 2-3 times a week; 4-6 times a week; once a day; twice a day; and ≥3 times a day. Food and alcohol consumption was based on habitual diet and a 'reference' recall period was set as one year prior to interview for controls. If there was any recent change in habits, only information on the habits before the change was used in data analysis. Information was sought on the usual amount of each food consumed per meal as well as cooking methods used and vitamins or mineral supplements taken. Amounts of consumed items were quantified using the Chinese common measure liang (equivalent to 50 grams). The frequency and quantity variables derived from the FFQ were converted into daily food consumption, adjusted for the edible portions of foods, cooking methods, seasonal factors, and market availability (Whitemore et al., 1990). The frequency and quantity variables for beer, wine, and liquor were converted into daily intake in ml. Amounts of ethanol ingested were calculated by assuming 10g of ethanol per 285ml of beer, per 100ml of wine, and per 30ml of liquor based on a method used in a previous study (Kropp et al., 2001). Physical activity was expressed in terms of weekly metabolic equivalent task hours (MET hrs/week) (Zhang et al., 2003). MET scores of 6, 4.5, and 2.5 were assigned respectively for vigorous, moderate, and walking activity based on a compendium of physical activities (Ainsworth et al., 2000).

Demographic characteristics and lifestyle factors between the two control groups were compared using t-test for continuous variables and Chi square test for categorical variables. Partial correlation coefficients between the continuous variables in the quantitative FFQ and the categorical variables in the SFHQ were compared separately for two control groups together and separated, controlling for age and family history of cancer. Average daily energy and fat intakes from the 119 food items were calculated using data from the Chinese nutrient database established by the Institute of Nutrition and Food Health, Chinese Academy of Preventive Medicine (Institute of Nutrition and Food Hygiene 1999). EI was expressed in kcal/day. BMR was calculated based on the following equations (Food and Agriculture Organization 1985), accounting for age and weight of the subjects (James et al., 1990): BMR=2.08+0.0615 weight, for women 18–29 years; BMR=3.47+0.0364 weight, for women 30–59 years; BMR=2.49+0.0434 weight, for women 60 years and over.

The Goldberg equation was used to evaluate the overall bias for underreporting at the group level (Goldberg et al., 1991), with a cut-off value of 1.35 for EI/BMR to classify underreporting and normal groups (Cook et al., 2001). When univariate statistics showed no significant differences in EI and MET between them, the two control groups were combined together for further analysis for underreporting.

Food items were grouped into 17 major groups and subgroups (Shannon et al., 2005). Differences in mean intake of food groups between the two control groups were compared by using a t-test. Energy (kcal) and selected nutrients intakes were divided into quartiles based on the corresponding distribution of the population controls, with the lowest quartile being the reference category. Univariate analysis was undertaken to screen potential explanatory variables for subsequent multivariate analysis. These potential confounders were included in the models, because either they emerged as risk factors in previous studies (Dai et al., 2001; Suzuki et al., 2008) or because we observed evidence of potential confounding in our data set by comparisons between univariate and multivariate analyses (Zhang et al., 2007). Associations of hospital/population control status with energy and selected nutrients intakes were assessed using adjusted odds ratios (ORs), 95% confidence intervals (95% CIs) and p-values for trend estimated from conditional logistic regression, adjusted for education, income, BMI, smoking, passive smoking, alcohol and tea consumption, energy intake (kcal), physical activity and cancer in first degree relative.

Results

Table 1 shows demographic and lifestyle characteristics of the 814 participants aged 18 to 81years by hospital and population control status. The two control groups were remarkably similar in their distributions of age (forced by matching), marital status, education, income, BMI, smoking, passive smoking, alcohol consumption, tea drinking, physical activity, and family history of malignancy.

Table 2 presents the mean daily intakes of preserved dietary intakes, fat consumption and tea drinking measured by the quantitative FFQ within each of the relevant SFHQ categories. The partial correlation coefficients were: 0.47, 0.57 and 0.42 for total preserved foods versus cured foods; 0.64, 0.58 and 0.69 for fat intake versus fat consumption; 0.76, 0.74 and 0.80 for dried tea leaves consumption

Table 1. Demographic and Lifestyle Characteristics of Participants by Hospital and Population Control Status

Factor	Hospital controls (n=407)	Population controls (n=407)	P*
Age (years)	50.5±10.4	50.7±10.4	
Marital status			0.80
Married	372 (91.4)	374 (91.9)	
Others	35 (8.6)	33 (8.1)	
Education			0.71
No or primary school	32 (7.9)	28 (6.9)	
Junior high school	159 (39.1)	168 (41.3)	
Senior high school	88 (21.6)	77 (18.9)	
Tertiary education	128 (31.4)	134 (32.9)	
Income (per capita, Yuan/month)			0.26
≤1000	53 (13.0)	63 (15.7)	
1001-2000	269 (66.1)	252 (61.9)	
>2001	85 (20.9)	92 (22.6)	
BMI (kg/m^2)	23.5±3.1	23.8±3.0	0.16
Smoking (20 packs in lifetime)	20 (4.9)	15 (3.7)	0.39
Passive smoking	129 (31.7)	141 (34.6)	0.37
Alcohol consumption	110 (27.0)	104 (25.6)	0.63
Green tea drinking	128.8±388.8	102.6±297.8	0.28
Physical activity (MET hrs/week)	53.8±33.9	48.2±39.4	0.03
Malignancies in first degree relatives	60 (14.7)	65 (16.0)	0.63

*Two-sided t-test for continuous variables and Chi-square test for categorical variables.Values expressed as mean±SD or number (percent). BMI, body mass index; MET, metabolic equivalent tasks

Table 2. Preserved Foods, Fat and Tea Intakes Reported in the Quantitative FFQ and the SFHQ

	All controls (n=814)	Hospital Controls (n=407)	Population Controls (n=407)
	Mean preserved foods (g/day)[a]		
Cured food[b]			
Never or seldom	19.5 (25.5)[c]	15.2 (16.8)	23.1 (30.5)
Once a month	26.8 (24.4)	24.7 (25.5)	28.8 (23.3)
Once a week	39.0 (32.5)	32.8 (25.2)	45.7 (37.8)
Everyday	62.9 (41.8)	60.1 (29.2)	66.4 (53.5)
Correlation coefficients	0.47	0.57	0.42
p[d]	<0.01	<0.01	<0.01
	Mean fat intake (g/day)[a]		
Fat consumption[b]			
Never or seldom	53.3 (15.0)[c]	55.2 (16.0)	51.3 (13.8)
Sometimes	73.3 (22.0)	73.3 (21.5)	73.2 (22.6)
Often	111.7 (34.4)	100.2 (23.6)	119.3 (38.1)
Correlation coefficients	0.64	0.58	0.69
p[d]	<0.01	<0.01	<0.01
	Mean dried tea leaves (g/day)[a]		
Tea drinking frequency[b]			
Never or seldom	0	0	0
≤6 times a week	0.92 (0.56)[c]	0.92 (0.51)	0.91 (0.63)
≥1 times per day	2.14 (1.62)	2.19 (1.77)	2.06 (1.40)
Correlation coefficients	0.76	0.74	0.80
New batches of tea[b]			
Never or seldom	0	0	0
≤6 times a week	0.91 (0.56)[c]	0.91 (0.51)	0.91 (0.63)
≥ 1 times per day	2.13 (1.61)	2.18 (1.76)	2.06 (1.40)
Correlation coefficients	0.76	0.74	0.80
p[d]	<0.01	<0.01	<0.01

*FFQ, food-frequency questionnaire; SFHQ, short food habit questionnaire. [a]Quantitative variable of mean daily intake from FFQ. [b]Categorical variable from SFHQ. [c]Values expressed as mean (standard deviation). [d]Two-sided

versus tea drinking frequency, and new batches of tea in all, hospital, and population controls respectively. The correlations were moderate to high and all associated p-values were less than 0.05, confirming internal validity of the questionnaire in both control groups.

Table 3 compares the characteristics between the low energy reporters (EI/BMR<1.35) and the normal group. The EI/BMR exceeded 1.35 in 88.5% of all controls (90.7% of hospital controls and 86.2% of population controls). There was no statistically significant difference in age, income, marital status, smoking, passive smoking, alcohol consumption, tea drinking, malignancies in first degree relatives, height, and physical activity. However, significant differences were found in education, weight and BMI.

Table 4 presents mean daily intakes of the 17 food groups in hospital outpatient and population controls with t-test p-values from comparisons of the control groups. There was no statistically significant difference in any food groups between the two control groups.

Table 5 presents adjusted ORs for energy and selected nutrient intakes, representing the tendency for each factor to associate independently with status as a hospital outpatient rather than population control. The ORs ranged from 0.83 to 1.45 with only two confidence intervals excluding the null value for carbohydrates compared with the lowest quartile intake; however, the trend was only on the margin of statistical significance (p=0.05). There was no statistically significant difference in trend between the two control groups with regard to the daily

Table 3. Characteristics of Participants with Low (<1.35) and Normal EI/BMR Ratio

	EI/BMR <1.35 (n=94)	EI/BMR ≥1.35 (n=720)	P*
Age (years)			0.14
18-49	46 (48.9)	323 (44.6)	
50-59	37 (39.4)	257 (35.4)	
≥60	11 (11.7)	145 (20.0)	
Education			<0.01
No or primary school	10 (10.6)	50 (6.9)	
Junior high school	51 (54.3)	276 (38.3)	
Senior high school	15 (16.0)	150 (20.8)	
Tertiary education	18 (19.1)	244 (33.9)	
Income (per capita, Yuan/month)			0.14
≤1000	14 (14.9)	99 (13.7)	
1001-2000	66 (70.2)	457 (63.5)	
≥2001	14 (14.9)	164 (22.8)	
Marital status			0.74
Married	87 (92.6)	659 (91.5)	
Others	7 (7.4)	61 (8.5)	
Smoking (20 packs in lifetime)			0.98
No	90 (95.7)	689 (95.7)	
Yes	4 (4.3)	31 (4.3)	
Passive smoking			0.67
No	61 (64.9)	483 (67.1)	
Yes	33 (35.1)	237 (32.9)	
Alcohol consumption			0.75
No	68 (72.3)	532 (73.9)	
Yes	26 (27.7)	188 (26.1)	
Tea drinking			0.27
No	76 (80.9)	545 (75.7)	
Yes	18 (19.1)	175 (24.3)	
Malignancies in first degree relatives			0.44
No	77 (81.9)	612 (85.0)	
Yes	17 (18.1)	108 (15.0)	
Height (cm)	161.6±4.65	160.9±4.53	0.18
Weight (kg)	63.5±7.67	61.04±8.85	0.01
BMI (kg/m²)	24.3±2.80	23.5±3.10	0.02
<25	55 (58.5)	523 (72.6)	0.01
≥25	39 (41.5)	197 (27.4)	
Physical activity (MET hrs/week)			0.78
≤27.00	23 (24.5)	181 (25.1)	
>27.00-42.00	20 (21.3)	185 (25.7)	
>42.00-67.50	26 (27.7)	178 (24.7)	
>67.50	25 (26.6)	176 (24.4)	

*Two-sided t-test for continuous variables and Chi-square test for categorical variables. Values expressed as mean±SD or number (percent). EI, energy intake; BMR, basal metabolic rate; BMI, body mass index; MET, metabolic equivalent tasks

Table 4. Comparison of Mean Daily Intake (g/d) of Food Groups between Hospital and Population Controls

Food groups	Hospital controls (n=407)	Population controls (n=407)	P*
Preserved foods	35.2 (30.6)	39.8 (41.2)	0.08
Staple foods	420.1 (139.8)	438.2 (182.9)	0.11
Meat	67.6 (63.0)	74.5 (68.6)	0.14
Poultry	8.9 (12.4)	10.3 (16.6)	0.18
Eggs	36.0 (24.4)	38.9 (30.2)	0.13
Fish and shellfish	26.0 (28.4)	23.9 (23.7)	0.26
Milk and its products	127.6 (127.5)	139.4 (129.8)	0.19
Vegetables	399.9 (179.7)	419.3 (256.0)	0.21
Dark green leafy vegetables	54.3 (35.2)	56.6 (48.8)	0.43
Cruciferous vegetables	37.4 (25.3)	38.3 (35.3)	0.67
Allium vegetables	73.9 (45.7)	80.2 (68.5)	0.12
Soy foods	155.8 (123.9)	151.7 (121.3)	0.64
Legumes	36.1 (28.3)	39.8 (35.8)	0.10
Soybean products	119.7 (116.3)	111.9 (104.3)	0.32
Fruits	269.8 (165.4)	259.3 (166.3)	0.37
Vegetable oil	38.9 (17.2)	40.4 (18.2)	0.22
Lard	0.24 (2.36)	0.05 (0.64)	0.12

*Two-sided t-test for continuous variables. Values expressed as mean (SD)

Table 5. Adjusted Odds Ratios and 95% Confidence Intervals for the Odds of being a Hospital Control According to Quartile Daily Intake of Energy and Nutrients

Nutrient	No. hospital/ population controls	OR[a]	95% CI	P[b]
Energy intake (kcal)				0.09
≤2126.23	79/101	1.00[c]		
2126.24-2575.12	114/102	1.23	0.92, 1.64	
2575.13-3370.60	140/102	1.32	1.00, 1.75	
>3370.60	74/102	0.98	0.70, 1.36	
Protein (g)				0.26
≤65.69	85/102	1.00		
65.70-85.21	128/102	1.29	0.97, 1.73	
85.22-114.10	117/102	1.31	0.95, 1.81	
>114.10	77/101	1.19	0.74, 1.92	
Fat (g)				0.11
≤60.02	90/101	1.00		
60.03-74.95	121/102	1.20	0.91, 1.60	
74.96-101.36	132/102	1.25	0.94, 1.67	
>101.36	64/102	0.90	0.59-1.35	
Dietary fiber (g)				0.15
≤13.30	125/140	1.00		
13.31-17.92	128/140	0.98	0.73, 1.33	
17.93-24.33	152/140	1.16	0.87, 1.54	
>24.33	155/140	1.38	0.99, 1.90	
Carbohydrate (g)				0.05
≤302.81	66/101	1.00		
302.82-382.37	126/102	1.40	1.03-1.89	
382.38-504.73	135/102	1.45	1.08-1.94	
>504.73	80/102	1.15	0.83-1.61	
Cholesterol (mg)				0.49
≤202.72	104/101	1.00		
202.73-346.71	129/102	1.08	0.83, 1.41	
346.72-462.00	104/102	0.99	0.75, 1.30	
>462.00	70/102	0.83	0.59, 1.18	

*OR, odds ratio; CI, confidence interval. [a]Estimates from conditional logistic regression models included terms for education (no or primary school, junior high school, senior high school, tertiary education), income (per capita, ≤1000, 1001-2000, ≥2001 Yuan/month), BMI now (continuous), smoking (no, yes), passive smoking (no, yes), alcohol consumption (no, yes), tea drinking (no, yes), physical activity (weekly MET-hours, continuous), energy intake (continuous, kilocalories), cancer in first degree relative. [b]Two-sided test for trend across quantitative variables. [c]Reference group

intakes of energy, protein, fat, dietary fiber, carbohydrate, and cholesterol.

Discussion

In this study, we investigated the internal validity of the FFQ for use in Chinese women and obtained very similar results using hospital and population controls separately and together. The partial correlation coefficients were moderate to high when the SFHQ items for preserved foods, fat intake and tea drinking were compared with the corresponding quantitative FFQ variables, thus confirming agreement between the two control groups. Reported EI is an important benchmark of validity in nutritional epidemiology (Livingstone et al., 2003). The Goldberg formula, the ratio of reported EI to BMR (EI/BMR), has frequently been used to assess the validity of dietary methods at the group level (Goldberg et al., 1991). An EI/BMR ratio of 1.35 and above has been considered as the maintenance requirement for energy (Cook et al., 2001). Our results showed that the average EI/BMR for all participants was 1.93, indicating sufficient energy intake by Chinese women and only 11.5% of all controls were apparently underreporting energy intake measured by the FFQ when evaluated by Goldberg's technique. Similarities in most demographics and lifestyle factors between low and normal EI/BMR groups provided additional evidence of homogeneity of the study population. The findings confirmed that the FFQ had reasonable validity, and can be used to measure habitual dietary intakes for Chinese women, which was consistent with a previous study conducted by our research team in elderly Chinese men (Jian et al., 2006).

In addition, our results showed that with respect to the food groups and selected nutrients measured by the FFQ, the habitual dietary intakes in hospital outpatient controls were similar to those in population controls. Therefore, our data provided evidence that regardless of whether controls were selected from hospital outpatient attendees without malignancy or drawn from population household registries covering the catchment area of the participating hospital, the controls appeared to function mostly as if they were having similar habitual dietary intakes assessed by the FFQ. The findings of this investigation were consistent with a smaller pilot study conducted in 1999-2000 by our research team (Zhang et al., 2002), and also with a case-control study that compared hospital controls dietary intakes with neighbourhood controls in Spain (GonzÁLez et al., 1990), as well as with another case-control study that compared non-cancer outpatients food consumption with population controls in Japan (Inoue et al., 1997).

To our knowledge, the present study is the first to investigate differences in food consumption and intakes of energy and nutrients between population and hospital controls in Chinese women. The strengths of the study were that a reliable instrument specifically for Chinese people was used to collect the information (Ji et al., 1998; Zhang et al., 2005; Jian et al., 2006), and all of the interviews were conducted by a single investigator to avoid inter-interviewer variation. Selection bias was minimised by the high response rates (96.2% for hospital outpatient controls and 90.0% for population controls), by systematic recruitment procedures and evidenced by the remarkably similar distributions of demographic characteristics and lifestyle factors between the two control groups.

Validity is defined as the degree to which a study meets basic logical criteria for the absence of bias (Greenland 1997). A valid FFQ should accurately reflect typical food consumption over a designated period of time, undistorted by behavioural patterns or false memory (Livingstone et al., 2003). EI is an important measure because nutrients must be provided within the quantity of food consumed to fulfil energy requirements. Therefore, reported EI may be considered a surrogate measure of the total quantity of food intake (Livingstone et al., 2003). Underreporters might have deliberately or unconsciously erred when estimating frequencies or portion sizes. The structure of the FFQ itself (food items, frequency categories and reference portion sizes) could also be a source of error. BMI seems to be one of the most consistent factors in predicting underreporting of energy intake in nutritional assessment studies (Johansson et al., 2001). Women underreporting their energy intake in our study had lower education, higher body weight and BMI than their normal reporting counterparts, which were consistent with a study conducted in Canada (Bedard et al., 2004).

In China, all public hospitals have preventive health care branches, which are responsible for reporting infectious diseases, health checks, health counseling, community prevention services, health education, disease screening, family planning and birthing guidance, and care of staff (Tao et al., 2010). Patients living in cities readily visit hospitals as non-referred outpatients for check-ups. Cultural factors and insurance arrangements lead patients to maintain a strong relationship with one particular hospital, where they receive a complete range of health care (Wang et al., 2007; Hu et al., 2008). Survey data in 2008 showed low levels (14%) of community health service utilization, suggesting that community health services are not yet the first point of contact with the health system in China (Bhattacharyya et al., 2011). Therefore, hospital outpatients in China are somewhat similar to ambulatory patients visiting GP clinics in Western counties.

The use of a proper, representative control population is important in reducing biases in a case-control study. The function of controls is to provide valid information on the distribution of exposure within the population at risk of becoming a case (Wacholder et al., 1992). As there is almost never one ideal control group (Miettinen 1985; Miettinen 1990; Wacholder et al., 1992; Grimes et al., 2005; Rothman et al., 2008), some researchers believe that population registers provide the most valid way of sampling controls in a hospital case-control study (Amadori et al., 1995; Almendingen et al., 2001; Malagoli et al., 2008), because the main theoretical strength of population controls is the potential to provide information of exposure that is unaltered by associations with illness. Hospital controls, especially those with other diseases, may fail to provide an unbiased sample of the population at risk with respect to exposure status (Miettinen 1985; Wacholder et al., 1992; Rothman et al., 2008). Others (GonzÁLez et al., 1990; Inoue et al., 1997) suggest that generally, hospital controls should be preferred in a hospital case-control study in view of the issues of practicability, cost and travel time for face to face interviews. There may also be differences in the capacity to recall and report exposures between randomly selected population members and those who are actively engaged in the health system (Wacholder et al., 1992). Our study did find that population controls had a higher rate of underreporting of energy intake compared with hospital controls (13.8% for population controls and 9.3% for hospital outpatient controls), although there was no difference in mean energy intake between the two control groups. In addition, only hospital controls have shown some evidence that in the event of developing the cancer, they would be likely to attend the hospital and become a case in the study (Miettinen 1985; Wacholder et al., 1992; Rothman et al., 2008).

In conclusion, with respect to food groups and selected nutrients, our results suggest that the FFQ had reasonable validity to measure habitual dietary intakes in Chinese women, and hospital outpatient controls performed little different from population controls in estimating habitual dietary intakes measured by the FFQ. Therefore, even though some points of concern exist, hospital outpatients provide a satisfactory control group to assess dietary and nutrient intake using the FFQ in hospital-based case-control study in the Chinese hospital setting.

Acknowledgements

This work was supported by the National Health and Medical Research Council of Australia project grant (APP ID 572542). L Li was supported by the Australian Postgraduate Award and the University of Western Australia Establishment Award. The authors acknowledge with gratitude the participation of outpatients in Shenyang. We are grateful for the collaboration received from the participating hospital and the staff; in particular, we would like to thank Professor Liu Yun-peng and Dr Shi Jing of the First Hospital of China Medical University, for their kind assistance.

References

Ainsworth BE, Haskell WL, Whitt MC, et al (2000). Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc*, **32**, 498-504.

Akhter M, Iwasaki M, Yamaji T, et al (2009). Dietary isoflavone and the risk of colorectal adenoma: a case–control study in Japan. *Br J Cancer*, **100**, 1812-6.

Almendingen K, Hofstad B, Trygg K, et al (2001). Current diet and colorectal adenomas: a case_control study including different sets of traditionally chosen control groups. *Eur J Cancer Prev*, **10**, 395-406.

Amadori D, Nanni O, Ricci M, et al (1995). Hospital versus population controls in a retrospective study on diet and stomach cancer. *Eur J Public Health*, **5**, 209-14.

Bedard D, Shatenstein B, Nadon S (2004). Underreporting of energy intake from a self-administered food-frequency questionnaire completed by adults in Montreal. *Public Health Nutr*, **7**, 675-81.

Bhattacharyya O, Delu Y, Wong ST, et al (2011). Evolution of primary care in China 1997–2009. *Health policy*, **100**, 174-80.

Chang SC, Rashid A, Gao YT, et al (2008). Polymorphism of genes related to insulin sensitivity and the risk of biliary tract cancer and biliary stone: a population-based case-control study in Shanghai, China. *Carcinogenesis*, **29**, 944-8.

Cook A, Pryer J, Shetty P (2000). The problem of accuracy in dietary surveys. Analysis of the over 65 UK national diet and nutrition survey. *J Epidemiol Community Health*, **54**, 611-6.

Cook T, Rutishauser I, Seelig M, Australasian Food and Nutrition Monitoring Unit (2001). Comparable Data on Food and Nutrient Intake and Physical Measurements from the 1983, 1985 and 1995 National Nutrition Surveys, pp. 1-149. Canberra: Commonwealth of Australia.

Dai Q, Shu XO, Jin F, et al (2001). Population-based case-control study of soyfood intake and breast cancer risk in Shanghai. *Br J Cancer*, **85**, 372-8.

Food and Agriculture Organization/World Health Organization (WHO)/United Nations University (1985). Energy and Protein Requirements. Report of a Joint Expert Consultation. WHO Technical Report Series No. 724. Geneva: WHO.

Goldberg GR, Black AE, Jebb SA, et al (1991). Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. *Eur J Clin Nutr*, **45**, 569-81.

GonzÁLez CA, Torrent M, Agudo A, et al (1990). Hospital versus neighbourhood controls in the assessment of dietary risk factors. *Int J epidemiol*, **19**, 354-61.

Greenland S (1997). In Concepts of validity in epidemiological research, pp. 417-1016. New York: Oxford University Press.

Grimes DA, Schulz KF (2005). Compared to what? Finding controls for case-control studies. *Lancet*, **365**, 1429-33.

Hsing AW, Sakoda LC, Rashid A, et al (2008). Variants in inflammation genes and the risk of biliary tract cancers and stones: a population-based study in China. *Cancer Res*, **68**, 6442-52.

Hu S, Tang S, Liu Y, et al (2008). Reform of how health care is paid for in China: challenges and opportunities. *Lancet*, **372**, 1846-53.

Inoue M, Tajima K, Hirose K, et al (1997). Epidemiological features of first-visit outpatients in Japan: comparison with general population and variation by sex, age, and season. *J Clin Epidemiol*, **50**, 69-77.

Institute of Nutrition and Food Hygiene, Chinese academic of preventive medicine (1999). Food composition table (National representative values), 1st ed. Beijing: People's Health Press.

James W, Schofield E (1990). Human Energy Requirements – A Manual for Planners and Nutritionists. Oxford: Oxford University Press.

Ji BT, Chow WH, Yang G, et al (1998). Dietary habits and stomach cancer in Shanghai, China. *Int J Cancer*, **76**, 659-64.

Jian L, Binns CW, Lee AH (2006). Validity of a food-frequency questionnaire for elderly men in southeast China. *Public Health Nutr*, **9**, 928-33.

Johansson G, Wikman A, Ahre´n A-M, et al (2001). Underreporting of energy intake in repeated 24-hour recalls related to gender, age weight status, day of interview, educational level, reported food intake, smoking habits and area of living. *Public Health Nutr*, **4**, 919-27.

Kropp S, Becher H, Nieters A, et al (2001). Low-to-moderate alcohol consumption and breast cancer risk by age 50 years among women in Germany. *Am J Epidemiol*, **154**, 624-34.

Li L, Zhang M, Holman CDJ (2011). Population versus hospital controls for case-control studies on cancers in Chinese hospitals. *BMC Med Res Methodol*, **11**, 167.

Livingstone MB, Black AE (2003). Markers of the validity of reported energy intake. *J Nutr*, **133**, 895-920.

Malagoli C, Vinceti M, Pellacani G, et al (2008). Diet and melanoma risk: effects of choice of hospital versus population controls. *Tumori*, **94**, 669-73.

Miettinen OS (1985). In Theoretical Epidemiology: Principles of Occurrence Research in Medicine, pp. 46-83. New York: John Wiley & Sons.

Miettinen OS (1990). The concept of secondary base. *J Clin Epidemiol*, **43**, 1017-20.

Porta M, Greenland S, Last JM (2008). In A dictionary of epidemiology, 5th ed., pp. 53. New York: Oxford University Press.

Rothman KJ, Greenland S, Lash TL (2008). Case-control studies. In Modern epidemiology, 3rd ed., pp. 111-27. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Shannon J, Ray R, Wu C, et al (2005). Food and botanical groupings and risk of breast cancer: a case-control study in Shanghai, China. *Cancer Epidemiol Biomarkers Prev*, **14**, 81-90.

Suzuki T, Matsuo K, Tsunoda N, et al (2008). Effect of soybean on breast cancer according to receptor status: a case-control study in Japan. *Int J Cancer*, **123**, 1674-80.

Tao D, Hawkins L, Wang H, et al (2010). In Fixing the public hospital system in China, Vol. 2: Main report, pp. 8-20. Washington: World Bank.

Wacholder S, Silverman DT, McLaughlin JK, et al (1992). Selection of controls in case-control studies. I principles; II types of controls; III design options. *Am J Epidemiol*, **135**, 1019-50.

Wang J, Kushner K, Frey JJ, et al (2007). Primary care reform in the Peoples' Republic of China: implications for training family physicians for the world's largest country. *Fam Med*, **39**, 639-43.

Whitemore AS, Wu-Willians AH, Lee M, et al (1990). Diet, physical activity and colorectal cancer among Chinese in North American and China. *J Natl Cancer Inst*, **82**, 915-26.

Willett W (1998). Nutritional Epidemiology. New York: Oxford University Press.

Zhang M, Binns CW, Lee AH (2005). A quantitative food frequency questionnaire for women in southeast China: Development and reproducibility. *Asia Pac J Public Health*, **17**, 29-35.

Zhang M, Holman CDJ, Binns CW (2007). Intake of specific carotenoids and the risk of epithelial ovarian cancer. *Br J Nutr*, **98**, 187-93.

Zhang M, Lee AH, Binns CW (2003). Physical activity and epithelial ovarian cancer risk: a case-control study in China. *Int J Cancer*, **105**, 838-43.

Zhang M, Yang HJ, Holman CDJ (2009). Dietary intake of isoflavones and breast cancer risk by estrogen and progesterone receptor status. *Breast Cancer Res Treat*, **118**, 553-63.

Zhang M, Yang ZY, Binns CW, et al (2002). Diet and ovarian cancer risk: a case-control study in China. *Br J Cancer*, **86**, 712-7.