# RESEARCH ARTICLE

# Assessment of a Questionnaire for Breast Cancer Case-Control Studies

**Loreta Strumylaite[1]\*, Rima Kregzdyte[1], Danguole Ceslava Rugyte[2], Algirdas Bogusevicius[3], Kristina Mechonosina[3]**

## Abstract

    **The aim of the present study was to assess criterion validity and external reliability of a questionnaire on risk factors for breast cancer.** <u>**Materials and Methods**</u>**. Women with breast cancer diagnosis (the cases) (N=40) and matched individuals without cancer (the controls) (N=40) were asked to fill in a questionnaire twice: on a day of admission to hospital (Q1) and on a day before discharge (Q2), with a time interval of 4-6 days. The questionnaire included questions (N=150) on demographic and socioeconomic factors, diseases in the past, family history of cancer, woman's health, smoking, alcohol use, diet, physical activity, and work environment. Criterion validity of the questionnaire Q2 relative to reference questionnaire Q1 was assessed with the Spearman correlation coefficient (SCC); external reliability of the questionnaire was measured in terms of the intraclass correlation coefficient (ICC). Statistical analysis was performed with SPSS 16.** <u>**Results**</u>**. The responses to most of the questions on socioeconomic factors, family history on cancer, female health, lifestyle risk factors (smoking, alcohol use, physical activity) correlated substantially in both the cases and the controls with SCC and ICC>0.7 (p<0.01). Statistically non significant relationships defined only between the responses on amount of beer the cases drank at the ages up to 25 years and 26-35 years as well as time of use of estrogen and estrogens-progestin during menopause by the cases. Moderate and substantial SCC and ICC were determined for different food items. Only the response of the cases on veal consumption did not correlate significantly.** <u>**Conclusions**</u>**. The questionnaire on breast cancer risk factors is valid and reliable for most of the questions included.**

**Keywords:** A questionnaire - criterion validity - reliability - test–retest - a case-control study - breast cancer

## Introduction

    Breast cancer is the second most common cancer in women with nearly 1.4 million new cases in 2008 (WHRF). The main risk factors of breast cancer such as inheritance (mutations of BRCA1 and BRCA2 genes) (Ford et al., 1998; Begg et al., 2008), menstrual (early age at menarche, late menopause) and reproductive (null parity, late age at first full-term pregnancy) histories, late age at lactation and short duration are well established but are generally difficult to modify (Ma et al., 2006; Bao et al., 2011).

    A substantial amount of research has explored the influence of lifestyle factors such as diet (Pala et al., 2009; Hu et al., 2012), smoking (Luo et al., 2011a; 2011b), alcohol use (Zhang et al., 2007; Beasley et al., 2010) and physical activity (Peters et al., 2009; Eliassen et al., 2010), i.e. modifiable risk factors on breast cancer risk. However, convincing evidence has only been shown for high alcohol consumption (WHRF, 2007).

    It has been recognized that one of the main causes of uncertainty regarding the role of lifestyle factors, especially diet, in cancer causation is the lack of accuracy of the method used to obtain information on possible risk factors of cancer (Pasanisi et al., 2002). However, the relationships between diet, smoking, alcohol consumption, physical activity and cancer as an outcome are usually assessed by the questionnaires (Zhang et al., 2007; Pala et al., 2009; Peters et al., 2009; Beasley et al., 2010; Eliassen et al., 2010; Luo et al., 2011a; 2011b).

    A questionnaire to be used assessing the relationships between lifestyle risk factors and breast cancer was validated in the feasibility part of breast cancer case-control study. The article presents the results of testing the criterion validity and external reliability of the questionnaire used in the study.

## Materials and Methods

*Study population*

    The hospital based case-control study included 40 patients (cases) with new histologically confirmed breast cancer diagnose according to ICD 10, who required surgical intervention at the Department of Surgery,

[1]*Institute of Neurosciences,* [2]*Department of Anesthesiology,* [3]*Department of Surgery, Medical Academy, Lithuanian University of Health Sciences, Kaunas, Lithuania* \**For correspondence: loreta.strumylaite@lsmuni.lt*

Lithuanian University of Health Sciences and 40 controls without cancer diagnose from the other departments of the University's hospital that agreed to participate in the survey. Response rate (RR) of the cases was 71.4%, i.e. 42 out of 56 women agreed to fulfill a questionnaire, but 2 of them were omitted because of the diagnose changed; RR of the controls was 64.5% (40 out of 62 women agreed).

Each subject was asked to complete the questionnaire twice, on a day of admission (Q1) and on a day before discharge (Q2) from the hospital, 4-6 days apart. None of agreed patients knew about the second interview in advance. Both questionnaires were completed by the patients. The study was approved by the Kaunas Regional Biomedical Research Ethics Committee (01-10-2007 No. BE-2-1, Report No.5/2007). Written consent to participate in the study was received from each patient.

*Questionnaire*

The questionnaire on breast cancer risk factors was based on modified and adapted English version of Aichi Cancer Center Research Institute Lifestyle and Health Questionnaires (Hirose et al., 1995) with exception the parts of physical activity measured by Baecke (1982), consumption of alcoholic beverages (Horn-Ross et al., 2004) and a part of work environment with original questions. Translation of the questionnaire from English to Lithuanian and retranslation back to English and again translation of English version to Lithuanian by different person was carried out to ensure the uniformity across the questionnaires used. The questionnaire was pretested before validation study.

The self-administered questionnaire consisted of 150 items concerning the following nine components: 1) general and socio-economic factors (date of birth, residence (city/town/village), the level of education (basic/incomplete secondary/secondary/special secondary/university/incomplete university), education years (≤4/5-8/9-12/≥13 years), income per family member per month (≤400/401-800/801-1000/>1000 Lt), marital status (single/married/live as married/divorced/widow)); 2) history of diseases (stomach/thyroid/diabetes mellitus/asthma/tuberculosis/kidneys/arterial hypertension/angina pectoris/myocardial infarction/stroke/breast/oophorectomy/hysterectomy/cancer/other) and anthropometrical indices (height and weight at present/weight at 20 and 50 years of age); 3) family history on cancer (no/yes/do not know; if yes, mother/father/grandmother/grandfather; localization of tumor; age at the diagnose); 4) woman's history (menstrual and reproductive history, lactation, use of contraceptives, medicines against infertility, and hormone replacement therapy in menopause); 5) smoking (active and passive); 6) alcohol consumption (frequency and quantity); 7) physical activity (occupation, movements, sport, leisure time activities excluding sport, and sleeping); 8) diet (food frequency questionnaire (FFQ)); 9) work environment (dust, chemicals, radiation, stress, other factors).

The responses to the questions on woman's history were "yes", "no" or a figure showing age of the beginning of menstruation, first pregnancy, delivery, beginning of menopause, number of pregnancies and deliveries, miscarriages and abortions, breast-fed children, duration of the use of estrogens/estrogens and progestin during menopause. The question "Was your menses regular at 18 years of age?" had following responses: "Almost regular (±3 days)/sometimes irregular (one cycle in six month is untypical, i.e. variation is more than three days or cycle appears not every month/irregular (variation of cycle more than three days or cycle appears not every month)/never have had". The question "Do you still have menses?" had three possible responses: "Yes/yes, but the menses are irregular during the last 12 months/no, did not have during the last 12 months and more". The third response "No, did not have during ..." raised question "How did menopause occur?" with following responses: "Naturally at age ... years/after surgical treatment of ovaries or uterus at age ... years/other reasons at age ... years". The questions "Have you ever used following contraceptives? and "Have you ever used hormones for treatment of infertility?" had responses "No, never/yes, almost a month/2-6 months/>6 months".

The question "Do you smoke?" had following responses: "Yes, I do .... years/yes, I do sometimes/gave up ... years or ... months ago/no, I have never smoked". The other responses were the figures related to the number of cigarettes smoked a day or age, when they started to smoke or length of time spent in smoke-filled apartment. The question "Does/did someone smoke at your workplace?" was assessed on a rating scale "Never/sometimes/half of a day/almost all working day/never worked". The response on length of time spent at smoke-filled workplace was estimated on five-point scale "<1/1-4/5-9/10-19/≥20 years".

Alcohol consumption was estimated by frequency and quantity of the use of strong alcoholic beverages (liqueur, brandy, vodka and etc.), wine/champagne and beer at the age under 25 years, 26-35 years of age and a year before the survey (Horn-Ross et al., 2004). Frequency of drinks was evaluated on a rating scale "Every day/4-6 times a week/2-3 times a week, 1 time a week/1-3 times a month/once in 2 months/do not use". Quantity of different alcoholic beverages was assessed by ml per drink.

All responses on physical activity were estimated on five-point scale "Never/seldom/sometimes/often/always" with the exception of the questions on the name of the main occupation, the types of sport played and duration of sleep in average (Baecke et al., 1982).

FFQ with a rating scale "Almost do not eat or less than once a month/1-3 times a month/1-2 times a week/3-4 times a week/5-6 times a week/every day" was used to assess the use of 35 food items from the main Lithuanian diet (Kriaucioniene et al., 2008).

The possible responses to the questions on dust, chemicals, radiation, stress, and other factors at workplace were "yes" and "no" with the exception of the questions on duration of exposure to the factors.

*Statistical analysis*

Since variables of possible risk factors were not distributed normally, criterion validity of the questionnaire Q2 relative to the reference questionnaire Q1 was estimated by Spearman's correlation coefficient (SCC) (Pasanisi et al., 2002). Intraclass correlation coefficient

(ICC) was calculated to measure external reliability (test-retest) of the questionnaire. We used the cut offs to interpret the level of relationship associated with a given statistic: r≤0.3=weak, 0.3<r≤0.7=moderate, r>0.7= substantial (Agresti, 1996). The data analysis was performed with the SPSS 16 software program.

## Results

Table 1 lists some characteristics of a study group. Breast cancer patients (cases) were older than controls, i.e., patients without cancer diagnose. Significantly more subjects with secondary/special secondary education were defined in a group of the controls. A difference according to residence between cases and controls was not defined.

SCC of the responses on general (date of birth, living place, residence) and socio-economic factors (education, income, marital status) were high and varied from 0.95-1.00 (p<0.01) in the cases and from 0.91-1.00 (p<0.01) in the controls. The lowest ICC was for income per family member per month in the cases (0.97, p<0.01) and education years in the controls (0.94, p<0.01).

SCC for different diseases in the past was 0.64-1.00 (p<0.01) in case group and 0.61-1.00 (p<0.01) in control group; ICC for the diseases of the cases and controls were 0.76-1.00 (p<0.01) and 0.75-1.00 (p<0.01), respectively.

Both SCC and ICC for height and weight of cases and controls at present, at the 20 years and 50 years of age were 0.9 or higher (p<0.01).

SCC and ICC of the responses to the questions on family history on cancer are given in Table 2. The responses to family history on cancer questions substantially correlated in both the cases and the controls.

Correlation of the responses regarding woman's health was substantial with SCC and ICC about 0.7 or higher for both the cases and the controls (p<0.01), except the

**Table 1. The Main Characteristics of the Study Group**

| Variables | Cases (N=40) | Controls (N=40) |
|---|---|---|
| Age (mean±SD) | 57.7±11.7* | 48.9±12.4 |
| Education (%, N) | | |
|   University/incomplete university | 47.5 (19) | 30.0 (12) |
|   Secondary/special secondary | 40.0 (16) | 65.0 (26)* |
|   Basic/incomplete secondary | 12.5 (5) | 5.0 (2) |
| Residence (%, N) | | |
|   City | 67.5 (27) | 55.0 (22) |
|   Town | 20.0 (8) | 25.0 (10) |
|   Village | 12.5 (5) | 20.0 (8) |

*p<0.05

**Table 2. Spearman (SCC) and Intraclass Correlation Coefficients (ICC) for the Responses to the Questions on Family History on Cancer**

| Variables | Cases | | Controls | |
|---|---|---|---|---|
| | SCC* | ICC* | SCC* | ICC* |
| Someone in a family was diagnosed with cancer | | | | |
| | 0.78 | 0.86 | 0.71 | 0.77 |
| Mother | 0.88 | 0.93 | 1 | 1 |
| Father | 0.93 | 0.96 | 1 | 1 |
| Grandfather/grandmother | 1 | 1 | 1 | 1 |
| Other relatives | 0.87 | 0.93 | 1 | 1 |

*p<0.01

**Table 3. Spearman (SCC) and Intraclass Correlation Coefficients (ICC) for Responses to Questions on Woman's Health**

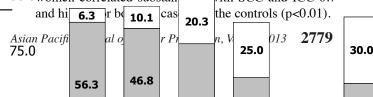| Variables | Cases | | Controls | |
|---|---|---|---|---|
| | SCC* | ICC* | SCC* | ICC* |
| How old were you when you began to menstruate? | | | | |
| | 0.96 | 0.97 | 0.95 | 0.96 |
| Was your menses regular at 18 years of age? | | | | |
| | 0.74 | 0.92 | 0.95 | 0.99 |
| Do you still have menses? | 1 | 0.99 | 0.99 | 0.99 |
| If NO, how did menopause occur? | 1 | 1 | 1 | 1 |
| How old were you when menopause occurred? | | | | |
| | 0.92 | 0.96 | 0.94 | 0.97 |
| Have you ever been pregnant? | 0.88 | 0.94 | 1 | 1 |
| If YES, how many times? (include live birth, still birth, miscarriages and abortions) | 1 | 1 | 1 | 1 |
| How old were you, when you were pregnant for the first time? | | | | |
| | 0.99 | 1 | 0.99 | 1 |
| Did you give birth? | 1 | 1 | 1 | 1 |
| How many times have you given birth? (include still births after 8th month) | 1 | 1 | 1 | 1 |
| How old were you when you delivered first baby? (include still birth after 8th month) | 0.99 | 1 | 0.99 | 1 |
| Did you breast-feed? | 0.87 | 0.81 | 0.91 | 0.95 |
| If YES, how many children? | 0.85 | 0.9 | 1 | 1 |
| How long did you breast-feed, including mixed feeding? | | | | |
|   1st baby | 0.92 | 0.97 | 0.99 | 0.99 |
|   2nd baby | 0.94 | 0.98 | 0.97 | 1 |
|   3rd baby | 0.96 | 1 | 1 | 1 |
| Did you have miscarriages or abortions? | | | | |
| | 1 | 1 | 1 | 1 |
| If YES, how many? | 1 | 1 | 1 | 1 |
| Have you ever used following medicines? | | | | |
| Contraceptives | 0.71 | 0.8 | 1 | 0.99 |
| Hormones for treatment of infertility | 1 | 1 | 1 | 1 |
| Have you used following medicines during menopause? | | | | |
|   Estrogens | 0.81 | 0.9 | 1 | 1 |
|   Estrogens-progestin | 0.9 | 0.94 | 1 | 1 |
|   Other hormones | 0.7 | 0.8 | 1 | 1 |
| How long did you used following medicines during menopause? | | | | |
|   Estrogens | 0.82[a] | 0.76[a] | 1 | 1 |
|   Estrogens-progestin | 0.65[a] | 0.16[a] | 1 | 1 |
|   Other hormones | 0.8 | 0.8 | 1 | 1 |

*p<0.01, [a]p>0.05

response of the cases to a question on duration of the use of estrogens and estrogens-progestin during menopause with both statistically insignificant SCC and ICC (Table 3).

SCC and ICC of the responses to the questions on smoking (active and passive) were greater than 0.7 in both the cases and the controls (p<0.01) (Table 4). Significantly substantial and moderate correlations defined for the responses on frequency and amount of the use of different alcoholic beverages, with exception ICC for the responses of the cases on amount of beer drank at age up to 25 and 26-35 years (Table 4).

Both SCC and ICC for different food items were moderate and substantial (p<0.01) in the cases as well as in the controls, and only correlation of the response on frequency of consumption of veal of the cases was not significant (SCC=0.28, ICC=0.34, p>0.05) (Table 5).

The responses regarding physical activity of the women correlated substantially with SCC and ICC 0.7 and higher for both cases and the controls (p<0.01).

**Table 4. Spearman (SCC) and Intraclass Correlation Coefficients (ICC) for Responses to Questions on Active and Passive Smoking and use of Alcohol**

| Variables | Cases | | Controls | |
|---|---|---|---|---|
| | SCC* | ICC* | SCC* | ICC* |
| **Active smoking** | | | | |
| Do you smoke? | 0.86 | 0.84 | 1 | 0.99 |
| How many cigarettes do you smoke a day? | | | | |
| | 1 | 1 | 0.96 | 0.97 |
| What type (light, medium, strong) of cigarettes do you smoke? | | | | |
| | 1 | 1 | 0.74 | 0.82 |
| How old were you, when you start to smoke? | | | | |
| | 1 | 1 | 0.87 | 0.96 |
| **Passive smoking** | | | | |
| Does/did someone smoke inside your living apartment? | | | | |
| | 0.96 | 0.99 | 0.95 | 0.98 |
| How long does/did someone smoke inside your living apartment? | | | | |
| | 1 | 0.99 | 0.94 | 0.99 |
| How many hours per day do/did you spend in smoke-filled apartment? | 0.99 | 0.97 | 0.91 | 0.95 |
| Does/did someone smoke at your workplace? | | | | |
| | 0.88 | 0.95 | 0.9 | 0.86 |
| How long do you work at smoke-filled workplace? | | | | |
| | 1 | 1 | 0.96 | 0.97 |
| **Alcohol use** | | | | |
| Do/did you drink alcoholic beverages? | 1 | 1 | 0.91 | 0.97 |
| How old were you, when you start drinking habitually? | | | | |
| | 0.88 | 0.94 | 0.97 | 0.99 |
| How often did you drink strong alcoholic beverages (liqueur, brandy, vodka and etc.)?  >25 years | 0.85 | 0.93 | 0.93 | 0.94 |
| 26-35 years | 0.92 | 0.97 | 0.83 | 0.86 |
| 1 years before | 0.82 | 0.86 | 0.85 | 0.91 |
| Amount of strong alcoholic beverages (liqueur, brandy, vodka and etc.) (ml) used in average during one drink: | | | | |
| > 25 years | 0.85 | 0.95 | 0.92 | 0.94 |
| 26-35 years | 0.81 | 0.87 | 0.88 | 0.96 |
| 1 years before | 0.60 | 0.77 | 0.88 | 0.97 |
| How often did you drink wine/champagne? | | | | |
| > 25 years | 0.73 | 0.82 | 0.74 | 0.88 |
| 26-35 years | 0.83 | 0.94 | 0.85 | 0.9 |
| 1 years before | 0.63 | 0.8 | 0.97 | 0.98 |
| Amount of wine/champagne (ml) used in average during one drink: | | | | |
| > 25 years | 0.85 | 0.89 | 0.83 | 0.93 |
| 26-35 years | 0.83 | 0.93 | 0.67 | 0.85 |
| 1 years before | 0.58 | 0.87 | 0.93 | 0.97 |
| How often did you drink beer? | | | | |
| > 25 years | 0.46 | 0.55 | 0.78 | 0.93 |
| 26-35 years | 0.73 | 0.76 | 0.94 | 0.96 |
| 1 years before | 0.83 | 0.89 | 0.99 | 0.92 |
| Amount of beer (ml) used in average during one drink: | | | | |
| > 25 years | 0.43 | 0.26ᵃ | 0.83 | 0.91 |
| 26-35 years | 0.68 | 0.35ᵃ | 0.97 | 1 |
| 1 years before | 0.99 | 1 | 0.99 | 0.99 |

*p<0.01, ᵃp>0.05

SCC of the responses to the questions on dust, chemicals, radiation and stress at workplace as well as on exposure time to these factors varied from 0.722-1.00 (p<0.01) in a case group and from 0.57-1.00 (p<0.01) in a control group; ICC of the responses were 0.84-1.00 (p<0.01) and 0.73-1.00 (p<0.01) in the cases and the controls, respectively.

## Discussion

Using a questionnaire in epidemiological surveys is one of the most frequent research methods. However, only valid and reliable questionnaire with optimal number of questions clearly formulated and understood is appropriate as a tool to get necessary information. In research studies

**Table 5. Spearman (SCC) and Intraclass Correlation Coefficients (ICC) for Responses to Questions on Consumption of Different Food Items**

| Variables | Cases | | Controls | |
|---|---|---|---|---|
| | SCC* | ICC* | SCC* | ICC* |
| White (semi white) bread | 0.84 | 0.91 | 0.98 | 0.98 |
| Dark bread | 0.87 | 0.94 | 0.95 | 0.97 |
| French bread | 0.85 | 0.94 | 0.88 | 0.94 |
| Noodles | 0.86 | 0.92 | 0.67 | 0.82 |
| Rice | 0.62 | 0.81 | 0.68 | 0.81 |
| Milk | 0.86 | 0.93 | 0.76 | 0.87 |
| Sour cream | 0.75 | 0.85 | 0.85 | 0.9 |
| Butter on bread | 0.78 | 0.89 | 0.82 | 0.88 |
| Margarine on bread | 0.92 | 0.96 | 0.86 | 0.94 |
| Curd/curd cheese | 0.67 | 0.8 | 0.8 | 0.91 |
| Cheese | 0.76 | 0.89 | 0.73 | 0.8 |
| Mayonnaise | 0.69 | 0.85 | 0.45 | 0.64 |
| Eggs (cooked, fried, omelet) | 0.76 | 0.89 | 0.53 | 0.61 |
| Beef | 0.73 | 0.83 | 0.62 | 0.82 |
| Veal | 0.28ᵃ | 0.34ᵃ | 0.73 | 0.77 |
| Pork | 0.73 | 0.81 | 0.82 | 0.88 |
| Poultry | 0.69 | 0.86 | 0.5 | 0.7 |
| Smoked meat (ham, sausages, bard) | 0.82 | 0.9 | 0.81 | 0.9 |
| Smoked fish | 0.49 | 0.62 | 0.76 | 0.83 |
| Salted fish (herring, salmon, etc.) | 0.74 | 0.85 | 0.74 | 0.85 |
| Fried/Cooked Fish/Fresh | 0.6 | 0.75 | 0.55 | 0.77 |
| Vegetables: | | | | |
| Carrots | 0.68 | 0.86 | 0.67 | 0.79 |
| Cabbages | 0.5 | 0.7 | 0.69 | 0.8 |
| Tomatoes during summer-autumn season | 0.59 | 0.76 | 0.5 | 0.63 |
| Paprika | 0.62 | 0.75 | 0.87 | 0.95 |
| Garlic | 0.87 | 0.9 | 0.78 | 0.89 |
| Onions | 0.75 | 0.83 | 0.52 | 0.67 |
| Pickled cabbages | 0.62 | 0.82 | 0.53 | 0.8 |
| Citrus fruit | 0.73 | 0.85 | 0.7 | 0.83 |
| Other fruits | 0.57 | 0.74 | 0.66 | 0.82 |
| Sweets (cakes, candies) | 0.84 | 0.91 | 0.96 | 0.98 |
| Black tea | 0.71 | 0.83 | 0.72 | 0.82 |
| Green tea | 0.8 | 0.9 | 0.78 | 0.92 |
| Other tea (herbs, etc) | 0.56 | 0.73 | 0.82 | 0.9 |
| Coffee | 0.96 | 1 | 0.94 | 0.97 |

*p<0.01, ᵃp>0.05

**Table 6. Spearman (SCC) and Intraclass Correlation Coefficients (ICC) for Responses to Questions on Physical Activity**

| Variables | | Cases | | Controls | |
|---|---|---|---|---|---|
| | | SCC* | ICC* | SCC* | ICC* |
| At work I: | Sit | 0.93 | 0.96 | 0.94 | 0.97 |
| | Stand | 0.92 | 0.96 | 0.88 | 0.92 |
| | Walk | 0.86 | 0.88 | 0.92 | 0.93 |
| | Lift heavy loads | 0.86 | 0.94 | 0.95 | 0.97 |
| | After working I am tired | 0.82 | 0.91 | 0.93 | 0.95 |
| | At work I sweat | 0.77 | 0.89 | 0.9 | 0.95 |
| In comparison with others of my own age, I think my work physically is: much heavier/heavier/as heavy/lighter/much lighter | | 0.8 | 0.91 | 0.87 | 0.94 |
| Do you play sport? | | 0.89 | 0.94 | 0.92 | 0.96 |
| If yes, which sport do you play most frequently? | | 0.86 | 0.91 | 0.92 | 0.98 |
| If yes, how many hours per week? | | 0.68• | 0.8 | 0.94 | 0.97 |
| How many months a year? | | 0.75 | 0.87 | 0.71 | 0.93 |
| In comparison with others of my own age, I think my physical activity during leisure time is active: much more/more/the same less/much less | | 0.82 | 0.9 | 0.86 | 0.93 |
| During leisure time I: | Sweat | 0.78 | 0.86 | 0.91 | 0.93 |
| | Play Sport | 0.87 | 0.93 | 0.84 | 0.94 |
| | Watch Television | 0.8 | 0.9 | 0.78 | 0.88 |
| | Walk | 0.75 | 0.86 | 0.78 | 0.89 |
| | Cycle | 0.86 | 0.91 | 0.91 | 0.94 |
| How many minutes do you walk and/or cycle per day to and from work, school and shopping? | | 0.74 | 0.86 | 0.83 | 0.9 |
| How many hours do you sleep in average? | | 0.84 | 0.96 | 0.92 | 0.98 |

*p<0.01, •p<0.05

the lack of validity and reliability of the questionnaire shows the systematic and random measurement errors, respectively. Therefore, testing the validity and reliability of a questionnaire as a tool is a way to avoid or reduce the number of errors in scientific studies (Feunekes et al., 1999).

The paper presented the results of a study, which was carried out in order to test the criterion validity and external reliability of the questionnaire used in breast cancer case-control study. Inasmuch as, the "gold standard", that is the SF-36 questionnaire in quality of life studies, does not exist, testing criterion validity of the questionnaire, first completed questionnaire Q1 was used as the reference (Pasanisi et al., 2002). The data showed that the responses to most of the questions on demographic and socio-economic factors, anthropometric indices, family history on cancer, smoking, and physical activity in both the cases and the controls correlated substantially. Moderate and substantial correlation was defined for the responses on diseases in the past. This can be explained by relevant and memorable information asked of the subjects.

Substantial correlation was defined between the responses to most of the questions about woman's health in both the cases and the controls. However, both correlations (SCC, ICC) of the response of the cases to a question on duration of the use of estrogens and estrogens-progestin during menopause were not statistically significant, although the correlations were significant in the controls. A review of the questionnaires revealed that two cases-respondents that used both estrogens and estrogen-progestin medicines mixed their use time. This could be explained by the older age of the cases and personal ability to memorize information dealing with use of medicines. As the evaluation of the use of these hormones is essential, in order to avoid systematic and random errors of the measurement in a further analysis, both the use of estrogens and estrogens-progestin will be assessed as one category of female sex hormones taking into account the overall consumption of these medicines during menopause.

In epidemiological surveys alcohol consumption is determined by five main methods: quantity/frequency and extended quantity/frequency questionnaires, retrospective and prospective diaries, and 24-hour recalls. The mean level of alcohol intake differed by 20% between these methods, although specific questions on intake of beer, wine, and liquor resulted in 20% higher estimates of intake (Feunekes et al., 1999). It has been found that high drinkers tend to underestimate alcohol consumption and drinking behavior, whereas lower drinkers tended to overestimate it (Townshend and Duka, 2002). When there is sufficient evidence that alcohol intake is underestimated in a population, methods that enquire about both the frequency and amount consumed, for beer, wine, and liquor, separately, will yield the most realistic levels of intake (Feunekes et al., 1999).

In our case, alcohol use was measured by quantity/frequency questions for different sorts of alcohol (strong alcoholic beverages: liquor, brandy, vodka; wine/champagne; beer). Moderate and substantial correlation coefficients (SCC, ICC) were found for the responses on frequency/quantity of strong alcoholic beverages and wine/champagne. However, ICC of the responses of the cases on amount of beer drunk up to 25 and 26-35 years of age was not statistically significant, although the responses of the controls showed significant correlations. On the one hand, this can be explained by the fact, that in the past, young women at the ages up to 25 or 26-35 years consumed beer rarely. On the other hand, the cases were older than the controls. Because of that, the cases had some difficulties to remember and state correctly consumption of beer in the past.

The same issue was found using FFQ in diet assessment, i.e. the less you eat a certain food item, the more difficult is to remember the frequency of consumption. In epidemiological studies information on diet is collected by FFQ with responses on a scale indicating the frequency of use of the product, sometimes the portion size, and a 1-7 day diet diary. Biomarkers of nutrient intake (mean daily intakes of dietary energy, total fat, saturated, monounsaturated and polyunsaturated fatty acids, linoleic acid, total carbohydrate, sugars, starch, dietary fibre, protein, vitamin C, vitamin E, folate, carotenes, Fe, Ca, Mg, K and alcohol) are also measured. There is evidence of satisfactory links between a 1-7 day diet diary and food-frequency questionnaire as well as different biomarkers of nutrient status (Brunner et al., 2001; Millen et al., 2006; Barclay et al., 2008). However, the data are not consistent (Michels et al., 2005). Some data show that evaluation of diet using the questionnaires prone to measurement error, therefore, for both methods, adjustment of nutrient intake to mean dietary energy intake appears to be the optimal approach to present and analyze the data (Brunner et al., 2001; Michels et al., 2004). In our study, FFQ used represent the consumption of different food items a year before the illness. Therefore, to use a diet diary for present use of different food items or to measure some biomarkers is not correct methodologically, because the data show the information obtained at different times. Our data defined substantial and moderate correlation between the responses to most of the questions about many food items in both the cases and the controls, with exception veal consumption with both statistically not significant SCC and ICC for the responses of the cases and the controls for the use of this type of the meat. As the subjects of the study ate veal rarely, we think that this was the main reason of the different responses about the frequency of veal consumption.

One of the possible shortcomings in the study is that the information collected with the questionnaire was not verified by objective measurements. Biomarkers of smoking (cotinine in plasma, saliva, or urine) (Jarvis et al., 1984), that could supplement and clarify the questionnaire's information were not determined. We did not measure doubly labeled water energy and/or maximal oxygen uptake to validate physical activity questionnaire, however it has been done by Bonnefoy et al. (2001). Inasmuch as, information on diet was collected asking about the food products used a year before the illness, to check different metabolic substances of the diet directly measuring them in biological media of the subjects was not possible. Indirect measuring of the biomarkers

of nutritional intake using nutrient database was not possible because of a portion size was not determined by the questionnaire. However, substantial correlations (SCC, ICC) between the most of the responses to most of the questions indicate that information collected by the questionnaire is valid and reliable.

In conclusion, the questionnaire used in a hospital based case-control study on breast cancer is valid to assess risk factors of the disease. The responses to most of the questions correlated substantially or moderately with coefficients of the correlations being statistically significant. Information on the response about duration of the use of estrogens or estrogen-progestin products during menopause will be combined in further analysis and assessed as the female sex hormone use during menopause. The responses to the questions on amount of beer or consumption of veal will not be taken into analysis or treated with caution.

## Acknowledgements

## References

Agresti A (1996). Introduction to categorical data analysis. John Wiley and Sons, NY.

Baecke JAH, Burema J, Frijters JER (1982). A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *Am J Clin Nutr*, **36**, 936-42.

Bao PP, Shu XO, Gao YT, et al (2011). Association of hormone-related characteristics and breast cancer risk by estrogen receptor/progesterone receptor status in the Shanghai Breast Cancer Study. *Am J Epidemiol*, **174**, 661-71.

Barclay AW, Flood VM, Brand-Miller JC, Mitchell P (2008). Validity of carbohydrate, glycaemic index and glycaemic load data obtained using a semi-quantitative food-frequency questionnaire. *Public Health Nutr*, **11**, 573-80.

Beasley JM, Coronado GD, Livaudais J, et al (2010). Alcohol and risk of breast cancer in Mexican women. *Cancer Causes Control*, **21**, 863-70.

Begg CB, Haile RW, Borg A, et al (2008). Variation of breast cancer risk among BRCA1/2 carriers. *JAMA*, **299**, 194-201.

Bonnefoy M, Normand S, Pachiaudi C, et al (2001). Simultaneous validation of ten physical activity questionnaires in older men: a doubly labeled water study. *J Am Geriatr Soc*, **49**, 28-35.

Brunner E, Stallone D, Juneja M, Bingham S, Marmot M (2001). Dietary assessment in Whitehall II: comparison of 7 d diet diary and food-frequency questionnaire and validity against biomarkers. *Br J Nutr*, **86**, 405-14.

Eliassen AH, Hankinson SE, Rosner B, Holmes MD, Willett WC (2010). Physical activity and risk of breast cancer among postmenopausal women. *Arch Intern Med*, **170**, 1758-64.

Feunekes GI, Staveren WA, Kok FJ (1999). Alcohol intake assessment: the sober facts. *Am J Epidemiol*, **150**, 105-12.

Ford D, Easton DF, Stratton M, et al (1998). Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am J Hum Genet*, **62**, 676-89.

Huang XE, Hirose K, Wakai K, et al (2004). Comparison of lifestyle risk factors by family history for gastric, breast, lung and colorectal cancer. *Asian Pac J Cancer Prev*, **5**, 419-27.

Horn-Ross PL, Canchola AJ, West DW, et al (2004). Patterns of alcohol consumption and breast cancer risk in the California Teachers Study Cohort Cancer. *Epidemiol Biomarkers Prev*, **13**, 405-11.

Hu F, Yi BW, Zhang W, et al (2012). Carotenoids and breast cancer risk: a meta-analysis and meta-regression. *Breast Cancer Res Treat*, **131**, 239-53.

Jarvis M, Tunstall-Pedoe H, Feyerabend C, Vesey C, Salloojee Y (1984). Biochemical markers of smoke absorption and self-reported exposure to passive smoking. *J Epidemiol Community Health*, **38**, 335-9.

Kriaucioniene V, Petkeviciene J, Klumbiene J (2008). Dietary patterns and their association with sociodemographic factors in Lithuanian adult population. *Medicina (Kaunas)*, **44**, 799-804.

Luo J, Margolis KL, Wactawski-Wende J, et al (2011). Association of active and passive smoking with risk of breast cancer among postmenopausal women: a prospective cohort study. *BMJ*, **342**, 1016.

Ma H, Bernstein L, Pike MC (2006). Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies. *Breast Cancer Res*, **8**, 43.

Michels KB, Bingham SA, Luben R, Welch AA, Day NE (2004). The effect of correlated measurement error in multivariate models of diet. *Am J Epidemiol*, **160**, 59-67.

Michels KB, Welch AA, Luben R, Bingham SA, Day NE (2005). Measurement of fruit and vegetable consumption with diet questionnaires and implications for analyses and interpretation. *Am J Epidemiol*, **161**, 987-94.

Millen AE, Midthune D, Thompson FE, Kipnis V, Subar AF (2006). The National Cancer Institute Diet History Questionnaire: Validation of Pyramid Food Servings. *Am J Epidemiol*, **163**, 279-88.

Pala V, Krogh V, Berrino F, et al (2009). Meat, eggs, dairy products, and risk of breast cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. *Am J Clin Nutr*, **90**, 602-12.

Pasanisi P, Berrino F, Bellati C, Sieri S, Krogh V (2002). Validity of the Italian EPIC questionnaire to assess past diet. In: Nutrition and lifestyle: Opportunities for cancer prevention, Eds Riboli E and Lambert R. IARC Press, Lyon pp 41-4.

Peters TM, Schatzkin A, Gierach GL, et al (2009). Physical activity and postmenopausal breast cancer risk in the NIH-AARP Diet and Health Study. *Cancer Epidemiol Biomarkers Prev*, **18**, 289-96.

Steiner DL, Norman GR (2001). Health measurement scales. A practical guide to their development and use. 2nd ed. Oxford University Press, New York.

Townshend JM, Duka T (2002). Patterns of alcohol drinking in a population of young social drinkers: a comparison of questionnaire and diary measures. *Alcohol Alcohol*, **37**, 187-92.

World Cancer Research Fund (WHRF) http://www.wcrf.org/cancer_statistics/world_cancer_statistics.php

World Cancer Research Fund/American Institute of Cancer Research (2007). Food, nutrition, physical activity, and the prevention of cancer: a global perspective. AICR, Washington, DC.

Xue F, Willett WC, Rosner BA, Hankinson SE, Michels KB (2011). Cigarette smoking and the incidence of breast cancer. *Arch Intern Med*, **171**, 125-33.

Zhang SM, Lee IM, Manson JE, et al (2007). Alcohol consumption and breast cancer risk in the women's health study. *Am J Epidemiol*, **165**, 667-76.