

**Design a preliminary research proposal to study the  
relationship between breast cancer and fruit & vegetable  
consumption.**

BACKGROUND

Breast cancer is the most widespread malignant tumour in women (30 per cent of all new cases of cancers). There are more than ten percent of women who are affected with breast cancer their lifetime. By the expert estimation there were almost 36,000 new cases of breast cancer diagnosed in 2000. Breast cancer also is the most common cause of cancer death in females, e.g. the age-standardised rate of mortality from breast cancer was 30.1 per 100,000 of population in 2002 (NS, 2004) Consequently all studies dedicated to the problem of the prevention and control of breast cancer have the highest importance and can be considered as one of the priorities of the national health care system. In the last decade there were determined following risk factors for breast cancer occurrence: family history of breast cancer, early menarche, personal history of breast cancer and previous abnormal breast biopsy, late menopause, race, age, the continuous use of the hormonal preparations, null parity, refuse of breast feeding, obesity etc. Nevertheless the role of dietary factors in breast cancer epidemiology is still unclear.

The recent national survey conducted in the United Kingdom was designed as a cohort study assessing the diet of British

citizens. This survey (Cancer Research UK EPIC-Oxford study) was aimed to determine the impact of dietary factors on the hormone production and relative risk of breast cancer (van Gils et al., 2005). The primary statistical analysis showed the presence of correlation between the blood levels of estrogens and intake of phytoestrogens, food fibres, saturated and trans fatty acids, refined sugars etc. Contrarily, there were determined characteristics of the diet having preventive impact on the breast health. Thus low-protein diet and diet restricted in fat and sugars was correlated with lower breast cancer risk. Some investigators discussed the role of fruits and vegetables in the prevention of breast cancer (Freudenheim et al., 1996, Gandini et al., 2000; Smith-Warner S., 2001; Adzersen et al., 2003). The WCRF/AICR guideline (1997) includes recommendations to decrease the quantities of animal food (red meat, fat) and alcohol and to enrich diet with fruits and vegetables (see table 1 in the Appendix).

Some researchers found that both vegetables and fruits intake can decrease breast cancer risk. For example Freudenheim et al. (1996) suggested that the intake of vegetables decreases a risk of premenopausal breast cancer. The authors explain protective impact of the vegetables by the high content of the antioxidant agents (e.g. beta-carotene, lutein and zeaxanthin) in green vegetables. Similar data were obtained in the meta-analysis performed by the group of Italian researchers

(Gandini et al., 2000). They confirmed the existence of the association between lower breast cancer risk and the intake of vegetables and fruits.

On the other hand van Gils et al. (2005) stated in their prospective study that there were "observed no association of risk with either total consumption of vegetables and fruits or with vegetable subgroups" (p. 191). Accordingly to this statement there was developed a deduction about the absence of any significant protective effects for vegetable or fruit intake in relation to breast cancer risk. But the hypothesis about the possibility of such effects is still applicable for the observation of specific subgroups of females.

Yorkshire Breast Cancer Research Group reported about the significant declining mortality from breast cancer in this England's largest county (Pisani & Forman, 2004). Authors did not discuss the role of diet in this declining, but they considered that it was rather a result of the long-term effects of both mammography screening and increased use of systemic therapy (p. 152).

Yorkshire community took part in the EPIC-Oxford study; nevertheless there were not assessed issues of traditional Yorkshire meal characterising with prevalence of starchy and fatty food (e.g. Yorkshire pudding, Yorkshire parkin, fat rascals etc). Thus regional peculiarities of diet can influence on the incidence of breast cancer and other diseases.

The work hypothesis is based on the supposition that fruit and vegetables intake is inversely related to the risk of breast cancer development.

The main distinct of planned study is focus on the regional pattern of the diet and risk assessment only for breast cancer. Contrarily to the national stage of European Prospective Investigation into Cancer (EPIC) where there were discussed the role of Mediterranean diet, rich in fresh vegetables and fruits, as the preventive factors for the risk of different nosological forms of cancer, the goal of this study is the evaluation of the role of vegetables and fruits intake for the risk of breast cancer (by the example of Yorkshire community).

The target group will be formed from women of fertile age. There following objectives for achieving stated goal could be offered:

- Perform the postal survey for the assessment of daily intake of fruits and vegetables in the group of Yorkshire women of fertile age
- Determine the pattern of daily consumption of the essential nutrients containing in the vegetables and fruits (antioxidant agents: carotenes, vitamin E, vitamin C, bio-flavones, food fibers, lycopene, genicysteine etc)
- Divide group of respondents into the subcohorts by the level of intake of vegetables and fruits and by the ethnical bacground.

- Determine women having pathological changes in their breasts during regularly scheduled screenings for breast cancer
- Provide statistical analysis of the dependence of breast cancer occurrence on the pattern of vegetable and fruit intake

#### METHODS

The optimum study design for the survey of the role of dietary factors in the epidemiology of breast cancer is longitudinal cohort prospective study. The reason of this choice can be explained by the necessity of the assessment of the delayed effect (i.e. case of breast cancer) of dietary factors (vegetable and fruit intake). The advantages of this design are presented by the possibility to assess dynamical processes, e.g. pathogenesis or risk releasing, because the procedure of risk assessment requires prolonged follow up supervision (Bijleveld et al., 1999). Thus to detect releasing risk (occurrence of cancer) the control of mammography screening results should be used. The alternative designs (e.g. case-control or cross-sectional study) cannot determine the changes in time and consequently they are useless for planned survey.

Only women of fertile age constantly living in the area of Yorkshire county will be included in the survey. They will get mails with the proposal to take part in the survey, the form of consent and questionnaire.

The local survey will be conducted in the Yorkshire, a county in the Northern England. By the statistical data of 2001 census there were 1,022,970 women of fertile age (or 40.1% of total female population) in Yorkshire and the Humber (see fig. 1 in the Appendix). Because the average age-standardised rate of the incidence of new cases of female breast cancer in England was 114 cases per 100,000 of population than this rate will be used for sample calculations.

The roster of the respondents will be developed from the databases of the UK census 2001. All participants of planned survey should give written informed consent to contribute in the study. The ethical standards of the International Agency for Research on Cancer and national agencies (Research Ethics Committee) will be applied during all stages of the survey. Before survey conducting there would be granted permission by an ethical review committee. The confidentiality of personal data will be warranted.

The referent groups in this survey will be presented as the basic (women of fertile age constantly living in Yorkshire) and control (women of fertile age living in Yorkshire less than 2 years). Diet pattern will be evaluated by the special questionnaire based on the new version of Food Pyramid introduced by USDA on 26<sup>th</sup> of April, 2005. Accordingly to this version there are following groups of vegetables: dark green vegetables, orange vegetables, dry beans and peas (leguminous), starchy vegetables and other vegetables (see

fig. 2 in the Appendix). The distinct of the revised food pyramid is the rejection of the use of so called 'servings' and their replacement with more understandable units: cups, oz-egs (ounces-equivalents), teaspoons etc (USDA, 2005). Because these units are not related directly to the metric or Avoirdupois and Imperial systems they could be applied in the different cultural environment. Thus teaspoon is equivalent to 5 ml, and a cup - to 250 ml.

Proposed survey could be the first experience of the use of updated food pyramid for scientific purposes. Because the new variant of the USDA pyramid was introduced only three months ago the adaptation of this dietary guideline for pertinent needs of medical community could be very helpful and promising. The necessity of small pilot study could be discussed also.

All data in the survey will be collected by the direct mails. The mailed materials will include the form of written consent with explanation of confidentiality policy, its terms and conditions, as well as a questionnaire. If the woman agrees to become a participant of the survey she should sign the consent form and fill in the questionnaire.

The main advantage of the postal surveys includes their economic affordability and possibility of self-administration of time and order of answering for the respondents. But there are some weaknesses: these surveys need pre-coding of the distracters (answers) and frequently response rate is poor and

it is not exceed 30% of total number of mailed questionnaires (Dillman D., 1999). On the other hand the response rate can be used also for evaluating motivation of health related activities.

When the pattern of daily consumption of vegetables and fruits will be identified than the diet content of essential nutrients including antioxidant agents (carotenes, vitamin E, vitamin C, bio-flavones etc) will be determined by the use of specialized software (e.g. NutriBase or NUT). Finally the groups of women of fertile age and consuming the different amounts of vegetables and fruits will be divided into the sub-cohorts for further analysis.

The measurable features of the studies includes diet facts datasheets (i.e. amounts of vegetables and fruits consumed per week, the preferences in the consumption of the various types vegetables and fruits), the percentage of women consuming less fruits and vegetables than it's recommended by the existing guidelines (USDA, 2005), the occurrence of breast cancer among members of the different referent groups in the first five years after postal survey conducted, the statistical differences between these referent groups, the correlation between breast cancer incidence and fruit and vegetables intake.

The control on the survey conduct will be provided by the follow-up analysis of the response rate and surveillance on the female breast cancer incidence in the Yorkshire County.

To investigate whether there was an effect of very high fruit and vegetable consumption, there will be categorized fruits and vegetables intake into the percentiles. The level of statistical significance will be taken as 95%.

To calculate the sample size we need the following information: general size of the population, the occurrence of the issues (object, action) related to the goal and objectives of the study, level of the confidence etc. There are a formula widely using for such calculations (Sabo D., 1999):

$$n \geq \frac{z_{\alpha/2}^2 \pi (1-\pi)}{\Delta^2},$$

where  $z_{\alpha/2}^2$  - is probability factor (for 95% power (5% alpha) the probability factor is equal to 1.96)  
 $\pi$ - population parameter, it could be calculated as  $\pi = p \pm z_{\alpha/2} \sigma_p$ , where correspondingly  $p$  - is a point estimator and  $\sigma_p$  - the standard deviation of the sample.

$\Delta^2$  - precision level (accepted as 0.001)

As it was stated above there are 1,022,970 women of fertile age in Yorkshire County. The age-standardised rate of female breast cancer morbidity is 114.0 cases per 100,000 population (NS, 2004) or 0.09%, consequently 1167 new cases would be registered in Yorkshire County annually.

As it was stated above the power of statistical hypothesis in the sample of planned survey is taken as 95%. In other words

the probability of the absence of a type II error is 95%. The power of statistical hypothesis can be determined by the formula:

$$Power = 1 - p,$$

where  $p$  - is a type II error.

Consequently, the maximum power of statistical hypothesis theoretically could be equal to 1, and the minimum - to 0. 95% power is usual for researches in health sciences because it is very close to 1, i.e. this power is enough high for the exclusion of null hypothesis.

Because the rate of responses in the postal surveys is lower than in other types of researches we should take this circumstance in the consideration and the final sample size can be significantly higher than it will be calculated using mentioned formula.

$$n \geq \frac{1.96 \times 0.1 \times 0.9}{0.001^2} = \frac{0.1764}{0.00001} = 17640$$

There was determined that 17640 randomly selected women of fertile age constantly living in Yorkshire should participate in the survey as the basic group and the equal number of women arrived to Yorkshire in 2003-2005 - as the control group.

The alternative approach in calculating sample size could be described by the following formula:

$$n = \frac{\frac{t^2_{n,\alpha} S^2}{d^2}}{1 + \frac{1}{N} \left( \frac{t^2_{n,\alpha} S^2}{d^2} - 1 \right)}$$

where N- population,

d - admissible bias

S<sup>2</sup> - dispersion of the sample

t<sup>2</sup><sub>n,α</sub> - critical value (=1.96)

If apply this formula for the conditions described than sample size could be estimated as:

$$n = \frac{\frac{1.96 \times 10000}{10^2}}{1 + \frac{1}{1022970} \left( \frac{1.96 \times 10000}{10^2} - 1 \right)} = 196.0$$

but because the response rate was assessed as 10% only the sample size will be more or equal to 1960 persons.

The statistical analysis of obtained data could be conducted with the use of software like SPSS 13.0. This software is helpful for both parametric and non-parametric calculations, for determining the interrelations of analyzed numeral and attributive data. Because SPSS 13.0 includes enhanced module of SPSS Regression Models, which could be useful for assessment of the concordance of cancer incidence and fruit and/or vegetables intake. Other methods, which will be used for statistical processing, may include dispersion analysis and correlation analysis as well as descriptive statistics.

ETHICAL CONSIDERATIONS.

All researches related to the issues of cancer prevention or treatment may require very strict ethical policy (GMC-UK, 1998). The main ethical concerns of planned research are presented by the necessity of randomisation, preserving confidentiality and obtaining informed consent.

As it was stated above the ethical standards of the International Agency for Research on Cancer and national agencies (Research Ethics Committee) will be applied during all stages of the survey. Before survey conducting there would be granted permission by an ethical review committee. The confidentiality of personal data will be warranted during all stages of the survey.

The data management policy includes full anonymity of the questionnaire sheets. The sample list with the addresses selected for the survey will be kept in the safe place, which is unavailable for unauthorized persons. During the follow-up surveillance all facts and findings about health and other personal information must be never disclose.

References:

Adzersen KH, Jess P, Freivogel KW, Gerhard I, Bastert G. (2003) Raw and cooked vegetables, fruits, selected micronutrients, and breast cancer risk: a case-control study in Germany. *Nutr Cancer*. Vol. 46, No. 2 pp. 131-137.

Bijleveld C., van der Kamp L., Ab Mooijaart, van der Kloot W., van der Leeden R., van der Burg E. (1999) *Longitudinal Data Analysis: Designs, Models and Methods* SAGE Publications 448 p.

Dillman D. (1999) *Mail and Internet Surveys : The Tailored Design Method*. Wiley. 2<sup>nd</sup> edition. 480 p.

Freudenheim JL, Marshall JR, Vena JE, Laughlin R, Brasure JR, Swanson MK, Nemoto T, Graham S. (1996) Premenopausal breast cancer risk and intake of vegetables, fruits, and related nutrients. *J Natl Cancer Inst*. Vol. 88, No. 6 pp. 340-348.

Gandini S., Merzenich H, Robertson C, Boyle P. (2000) Meta-analysis of studies on breast cancer risk and diet: the role of fruit and vegetable consumption and the intake of associated micronutrients. *Eur J Cancer*. Vol. 36, No. 5, pp. 636-646.

Gils van CH, Peeters PH, Bueno-de-Mesquita HB, Boshuizen HC, Lahmann PH, Clavel-Chapelon F, Thiebaut A, Kesse E, Sieri S, Palli D, Tumino R, Panico S, Vineis P, Gonzalez CA, Ardanaz E, Sanchez MJ, Amiano P, Navarro C, Quiros JR, Key TJ, Allen N, Khaw KT, Bingham SA, Psaltopoulou T, Koliva M, Trichopoulou A, Nagel G, Linseisen J, Boeing H, Berglund G, Wirfalt E, Hallmans G, Lenner P, Overvad K, Tjonneland A, Olsen A, Lund

E, Engeset D, Alsaker E, Norat T, Kaaks R, Slimani N, Riboli E (2005) Consumption of vegetables and fruits and risk of breast cancer. JAMA. Vol. 293, No. 2 pp. 183-193.

GMC-UK (1998) Seeking patients' consent: the ethical considerations. Available at the web-site <http://www.gmc-uk.org/standards/consent.htm> (Accessed on 27.05.2005)

Jansen MC, Bueno-de-Mesquita HB, Feskens EJ, Streppel MT, Kok FJ, Kromhout D. (2004) Quantity and variety of fruit and vegetable consumption and cancer risk. Nutr Cancer. 2004;48(2):142-8.

NS (2002) 2001 Census data: Yorkshire and the Humber. <http://www.statistics.gov.uk/census2001/pyramids/pages/d.asp>

NS (2004) Breast Cancer Factsheet - February 2004. Available at the web-site <http://www.statistics.gov.uk> (Accessed on 26.05.2005)

Pisani P, Forman D. (2004) Reduced mortality from breast cancer in Yorkshire, 1983-1998: How large an effect and what are the causes? Br J Cancer Vol. 90, pp. 652-656.

Sabo D. (1999) Confidence Interval Estimates of the Population Proportion. Available at the web-site [http://www.math.bcit.ca/faculty/david\\_sabo/apples/math2441/section8/lrgsampprops/largesampprops.htm](http://www.math.bcit.ca/faculty/david_sabo/apples/math2441/section8/lrgsampprops/largesampprops.htm) (Accessed on 27.05.2005)

Smith-Warner S, Spiegelman D, Yaun SS, Albanes D, Beeson WL, van den Brandt PA, Feskanich D, Folsom AR, Fraser GE, Freudenheim JL, Giovannucci E, Goldbohm RA, Graham S, Kushi LH, Miller AB, Pietinen P, Rohan TE, Speizer FE, Willett WC,

Hunter DJ. (2001) Intake of fruits and vegetables and risk of breast cancer: a pooled analysis of cohort studies. JAMA. Vol. 285, No. 6 pp. 769-776.

USDA (2005) Food Intake Pattern Calorie Levels. Available at the [web-site](http://www.mypyramid.gov/downloads/MyPyramid_Food_Intake_Patterns.pdf) [http://www.mypyramid.gov/downloads/MyPyramid\\_Food\\_Intake\\_Patterns.pdf](http://www.mypyramid.gov/downloads/MyPyramid_Food_Intake_Patterns.pdf) (Accessed on 27.05.2005)

WCRF/AICR (1997) Food, Nutrition and the prevention of Cancer: a global perspective. World Cancer Research Fund and the American Institute of Cancer Research. Chapter 4. Available at the [web-site](http://www.wcrf-uk.org/report/chapter4/chapter4page186.lasso?WCRFS=D992F6010cd7d33476GQP3314A66) <http://www.wcrf-uk.org/report/chapter4/chapter4page186.lasso?WCRFS=D992F6010cd7d33476GQP3314A66> (Accessed on 27.05.2005)

Appendix.

Table 1. The evidence of dietary factors impact on the breast health (source: WCRF/AICR (1997) Food, Nutrition and the prevention of Cancer: a global perspective. World Cancer Research Fund and the American Institute of Cancer Research)

Evidence	Decreases risk	No relationship	Increases risk
Convincing	Vegetables	Coffee	Rapid Growth and Greater Adult Height
Probable	Vegetables and Fruits	Cholesterol	High Body Mass Adult Weight Gain
Possible	NSP/Fibre Carotenoids	Monounsaturated Fat Polyunsaturated Fat Retinol Vitamin E Poultry Black Tea	Alcohol Total Fat Saturated/ Animal Fat meat
Insufficient	Vitamin C Isoflavones and Licnans Fish	Calcium Selenium Fish	Animal Protein DDT Residues

Figure 1. The demographic pattern of Yorkshire and the Humber (NS, 2004)

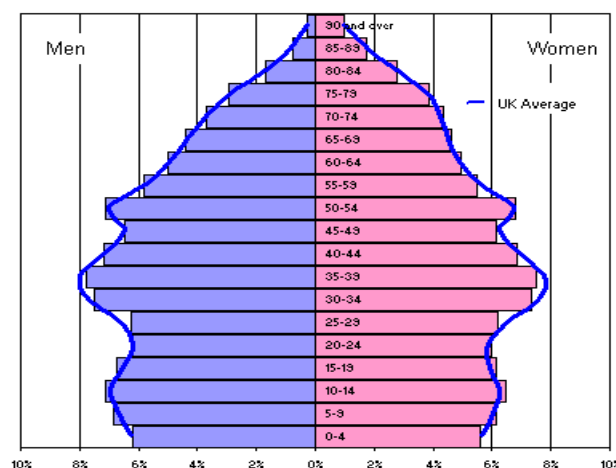


Fig. 2 Recommended vegetable intake per week (USDA, 2005)

Vegetable Subgroup Amounts are Per Week												
Calorie Level	1,000	1,200	1,400	1,600	1,800	2,000	2,200	2,400	2,600	2,800	3,000	3,200
Dark green veg.	1 c/wk	1.5 c/wk	1.5 c/wk	2 c/wk	3 c/wk	3 c/wk	3 c/wk	3 c/wk	3 c/wk	3 c/wk	3 c/wk	3 c/wk
Orange veg.	.5 c/wk	1 c/wk	1 c/wk	1.5 c/wk	2 c/wk	2 c/wk	2 c/wk	2 c/wk	2.5 c/wk	2.5 c/wk	2.5 c/wk	2.5 c/wk
Legumes	.5 c/wk	1 c/wk	1 c/wk	2.5 c/wk	3 c/wk	3 c/wk	3 c/wk	3 c/wk	3.5 c/wk	3.5 c/wk	3.5 c/wk	3.5 c/wk
Starchy veg.	1.5 c/wk	2.5 c/wk	2.5 c/wk	2.5 c/wk	3 c/wk	3 c/wk	6 c/wk	6 c/wk	7 c/wk	7 c/wk	9 c/wk	9 c/wk
Other veg.	3.5 c/wk	4.5 c/wk	4.5 c/wk	5.5 c/wk	6.5 c/wk	6.5 c/wk	7 c/wk	7 c/wk	8.5 c/wk	8.5 c/wk	10 c/wk	10 c/wk