

## RESEARCH COMMUNICATION

# A Prospective Study of Diet, Lifestyle, and Genetic Factors and the Risk of Cancer in Khon Kaen Province, Northeast Thailand: Description of the Cohort

Supanee Sriamporn<sup>1</sup>, DM Parkin<sup>2</sup>, P Pisani<sup>3</sup>, V Vatanasapt<sup>4</sup>, K Suwanrungruang<sup>4</sup>, S Kamsa-ard<sup>4</sup>, P Pengsaa<sup>5</sup>, O Kritpetcharat<sup>6</sup>, V Pipitgool<sup>7</sup>, P Vatanasapt<sup>4</sup>

## Abstract

Cohort studies are the preferred design in observational epidemiology, but few involving the general population have been performed in Asia, and most concern affluent urban populations. The Khon Kaen study has recruited about 25,000 subjects, aged mainly 35-64, from villages in the relatively underdeveloped north-east of Thailand. All subjects underwent simple physical examination, completed an interviewer-administered questionnaire (including sections on lifestyle, habits, and diet) and donated specimens of blood, which were processed and stored in a biological bank at -20°C. Female subjects (about 16,500) were offered screening by Pap smear, and specimens of cells from the cervix were stored at -20°C. This paper describes the methodology of the study, and the characteristics of the participants. Almost all subjects are peasant farmers, with low annual income and body mass, although 14.6% of women had a BMI in the obese range ( $>30 \text{ kg/m}^2$ ). Smoking was common among men (78% regular smokers, most of whom used home-produced cigarettes), but rare among women. Fertility levels were relatively high, with a more than half the women having four or more live births. 23.4% of subjects were infected with the liver fluke *Opisthorchis viverrini*, known to be highly endemic in this region. Follow-up of the cohort is by record-linkage to the provincial cancer registry. By 2003, 762 cancer cases had occurred, the most common being cancers of the liver (363 cases) and cervix uteri (44 cases). The antecedents of these cancers are being investigated using a nested case-control approach. The cohort will yield increasing numbers of cancers for study in the next decade, giving important information on the relative importance of dietary and lifestyle factors in a rural population, undergoing gradual transition to a more westernised lifestyle.

*Asian Pacific J Cancer Prev*, 6, 295-303

## Introduction

Cohort studies are generally accepted to be the most useful and valid type of investigation in observational epidemiology. Their strength lies not only in permitting the investigation of multiple disease endpoints, but also in providing an unbiased measurement of exposure before the onset of clinical disease. Large scale, multi-purpose prospective studies are currently being conducted in several western countries. Questionnaire data are integrated with biological samples that are stored at low temperature for future assessment. This strategy has the advantage that the study is not limited by the extent of current knowledge or

techniques, rather, hypotheses can be refined new and more advanced technology exploited at the time of analysis. This may well be after many years of follow-up.

In Asia, rather few cohort studies of the general population have been carried out; they include the pioneering study of Hirayama in Japan (Hirayama, 1990), which was limited to questionnaire data, but which documented the association between some dietary habits and all causes mortality as well as between some reproductive factors and breast and gynaecological cancers. More recent studies include the Japan Public Health Center-based prospective study on cancer and cardiovascular diseases (JPHC study) (Iwasaki et al., 2003), and ongoing studies in Shanghai,

<sup>1</sup> Department of Epidemiology, Faculty of Public Health, Khon Kaen University, Thailand <sup>2</sup> Clinical Trials Service Unit and Epidemiological Studies Unit, University of Oxford, England <sup>3</sup> Descriptive Epidemiology Group, International Agency for Research on Cancer, Lyon, France <sup>4</sup> Cancer Unit, Srinagarind Hospital, Khon Kaen University <sup>5</sup> Department of Obstetrics & Gynaecology, Faculty of Medicine, Khon Kaen University <sup>6</sup> Department of Pathology, Faculty of Medicine, Khon Kaen University <sup>7</sup> Department of Parasitology, Faculty of Medicine, Khon Kaen University. Corresponding Author: Dr Supanee Sriamporn, Department of Epidemiology, Faculty of Public Health, Khon Kaen University, Khon Kaen 40002, Thailand - Tel: + 66 (0)43-347057; Fax: + 66 (0)43-347058; e-mail: supanee@kku.ac.th

China (Yuan et al., 1996; Zhang et al., 2005), and Korea (Yoo et al., 2002). These have all been concerned with relatively urbanised populations.

The Khon Kaen cohort study began in 1990. It was initially conceived as a community-based early detection and health education project, designed to improve the outcome of cancer and some other non-communicable diseases, in the rural population of Khon Kaen Province in Northeast Thailand. The north-eastern region is the least developed part of Thailand. In 1992 the protocol was revised, with improvement and extension of the questionnaire in particular by the addition of a dietary section, and the storage of biological specimens on all participants.

An initial objective of the study was to investigate the epidemiology of liver cancer, and specifically, cholangiocarcinoma, which is particularly common in this population (Vatanasapt et al., 1990). Later, the scope of the study was expanded to permit study of disease endpoints and causes of mortality other than liver cancer.

In this paper, we present the methods of recruitment of the study subjects, and describe the characteristics of the cohort, in terms of frequencies of the variables collected, and their interrelationships. The methods of follow up, and likely study power are also presented.

## Materials and Methods

### Recruitment

In the first two years the target group was the population age 30 and above; in 1992 this was changed, so that the study concentrated upon the age groups providing the majority of cases of liver cancer and cervical cancer (age 35-64), as the cancers most likely to be the focus of the initial analyses.

At the beginning of the study (1990), Khon Kaen Province comprised 20 districts (amphur). In 1994, local government reorganisation increased the number of districts to 25, but in this presentation, the original 20 were retained (Figure 1). Each district contains about 10 sub-districts (tambon). One or two tambon were randomly selected in each district, and recruitment took place in all of the villages within it, the eligible population being interviewed and examined during 2-3 weeks of field-work, depending on the size of the resident population. First the Head of the village was contacted and the aim of the intervention explained. Lists of the resident population were obtained, and files of eligible people prepared. The village was visited by the study team, with a mobile unit equipped with an ultrasound machine. The unit was then installed in the village, and the population invited to participate. Village residents willing to participate were informed of the examinations and procedures they would undergo, as well as details of tests and exams performed and rationale of keeping part of their biological samples for future research investigations. Those who accepted to participate signed a consent form.

### Data collection

Information collected from subjects was in the form of responses to a structured interview, and findings on examination; biological samples were collected for later analysis.

### Interview

The structured questionnaire included the following sections:

- (1) Demographic and socio-economic characteristics
- (2) Usual diet including alcohol consumption, by means of a food-frequency questionnaire. The questionnaire was developed taking advantage of the results of a dietary survey conducted in 1992 by means of 24-hour dietary recall interviews of a representative population sample of 172 subjects. Food composition tables for the Thai population were obtained, and analysis of the 24-hour recall data allowed identification of the most important dietary sources of nutrients, fibre, vitamins, etc. The food frequency questionnaire was designed to include items that are common in the Thai diet. It is oriented to the estimation of usual intake of energy, animal and vegetable proteins, fat, fibre, carbohydrates, equivalent vitamin A and vitamin C, using food composition tables developed in Thailand (Viriyanich et al., 1990). In addition, the questionnaire includes food items which do not contribute significantly to the intake of these nutrients, but which are a source of aflatoxins, pre-formed nitrosamines, nitrate, nitrite or opisthorchis viverrini (OV).
- (3) The subject's current or last occupation, recorded according to 48 pre-defined categories of the International Standard Classification of Occupations (International Labour Organisation, 1988).
- (4) Tobacco smoking and betel nut chewing
- (5) Past history of active infection with OV and of treatment with the drug praziquantel
- (6) Reproductive & sexual history and practices of contraception and spouse's occupation (for female subjects)

### Medical Examination

The subjects underwent visual inspection of the oral cavity and ultrasound examination of the abdomen. Women received a gynaecological examination during which a cervical Pap smear was taken and their breasts were examined by palpation. Blood pressure, height and weight were measured and recorded.

### Biological specimens

**Blood:** Subjects enrolled donated ~10 ml of venous blood. This was separated into two sets of three 1 ml aliquots of plasma, buffy coat and clot respectively. Each set is stored in separate freezers at -20 °C. An additional plasma sample was used to assess anti-OV antibody, and fasting blood sugar. **Faeces:** Presence of OV eggs was also assessed in faeces using a semi-quantitative count (Stoll's egg count). **Cervical cells:** Cervical cells were collected from women using a spatula. A Pap smear was prepared, and examined

for abnormal cytology in the laboratory. All Pap smears were archived for future reference. A sample of cells was suspended in Phosphate Buffered Saline (PBS), centrifuged at 2,000 rpm for 15 mins, re-suspended in 50 ml of PBS and separated in two aliquots also stored at -20 °C.

### Referral

Participants who were found to be positive for OV-eggs in faeces, had an abnormal Pap test, elevated blood pressure or blood sugar were informed, and advised to seek treatment.

### Follow-up

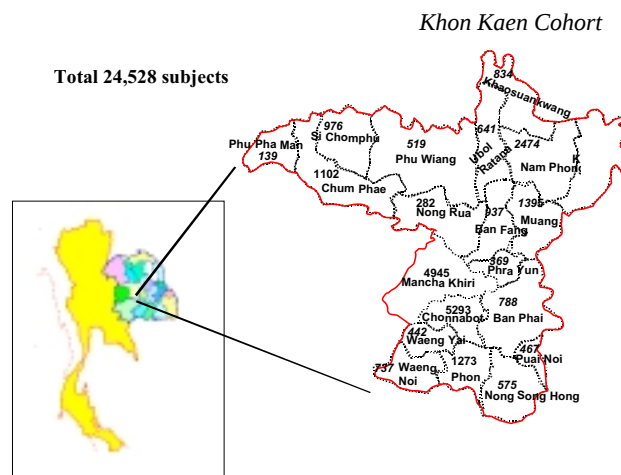
Follow-up of this cohort is being carried out by linkage of the study database with the deaths register for Khon Kaen Province, and the Khon Kaen Provincial Cancer Registry (Vatanasapt et al., 1992). The cancer registry is population-based and collects data on new cases of cancer occurring in the resident population. Record linkage between the files of the cohort and those of the cancer registry and of death certificates is carried out using the RECLINK program<sup>1</sup>.

### Data Entry

A large amount of information is available on each subject. It was anticipated that analyses of the association between different cancers and their antecedents would generally make use of the nested case-control approach (Samet & Munoz 1998), so that detailed information is required only for the relevant cases and a selected subset of non-cases (controls). For this reason, only the basic demographic data required for record linkage purposes have been entered for all subjects, although the laboratory tests of clinical relevance to participants (i.e. OV, Pap smear and blood sugar) were performed on all subjects, the results communicated to them, and recorded in the study database. To examine the characteristics of the study cohort, a random sample, stratified by age, of 500 men and 500 women aged 30-69, was selected and information from the questionnaire, physical and laboratory examinations coded and entered.

## Results

At the end of the recruitment, by April 2001, some 24,800 subjects had been enrolled. Of this number, some subjects were found to have been diagnosed with cancer (from the



**Figure 1. Khon Kaen Province: Map Showing of Subjects Recruited by District (Amphur)**

registry file) before the date of recruitment. These subjects were excluded from the cohort data file. This left a total of 24,528 subjects (8,021 men and 16,507 women), with the numbers in the different districts ranging from 139 to 5,293 (Figure 1). Blood specimens were obtained and stored for 19,484 of the cohort, Pap smears from 10,553 women were read, and cervical cells from 10,002 women were stored. The age and sex distribution of the whole cohort is shown in Table 1, together with the sample of 1,000 subjects, for whom data entry was carried out.

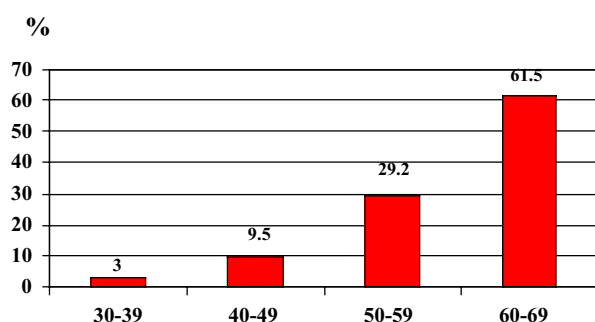
The great majority (85%) of the subjects gave their occupation (or husband's occupation for women) as farmer, and most (92%) had received only primary level education. The median income per household was Baht 21,600 (US \$ 540) per year.

Few women were smokers (1.4%), but 83.6% of men were smokers (ever or current) and 78.4% reported smoking regularly (every day). For the men who were regular smokers, 82.8% smoked yamuan (a home-made cigarette, comprising locally-cured tobacco wrapped in a banana leaf or paper (Mitacek et al., 1991). Among these yamuan smokers, 98.8% reported smoking every day, with 47.0% smoking 1-9 cigarettes per day, 34.3% smoking 10-19 cigarettes per day and 18.7% smoking 20 or more cigarettes per day.

The habit of chewing betel nut is more common in women than in men: 24.0% of women reported ever having

**Table 1. Age and Sex Composition of the Study Cohort, and of the Sample of 1,000 Subjects used for Analysis**

| Age   | Males  |      |        |      | Females |      |        |      |
|-------|--------|------|--------|------|---------|------|--------|------|
|       | Cohort |      | Sample |      | Cohort  |      | Sample |      |
|       | Number | %    | No.    | %    | Number  | %    | No.    | %    |
| <30   | 10     | 0.1  | —      | —    | 130     | 0.8  | —      | —    |
| 30-39 | 800    | 10.0 | 52     | 10.4 | 2,326   | 14.1 | 73     | 14.6 |
| 40-49 | 2,548  | 31.8 | 167    | 33.2 | 5,519   | 33.4 | 174    | 34.7 |
| 50-59 | 2,841  | 35.4 | 184    | 37.1 | 5,492   | 33.3 | 173    | 34.5 |
| 60-69 | 1,476  | 18.4 | 97     | 19.3 | 2,582   | 15.6 | 80     | 16.2 |
| 70+   | 346    | 4.3  | —      | —    | 458     | 2.8  | —      | —    |
| Total | 8,021  | 100  | 500    | 100  | 16,507  | 100  | 500    | 100  |



**Figure 2. Prevalence (%) of betel nut chewing by age (female)**

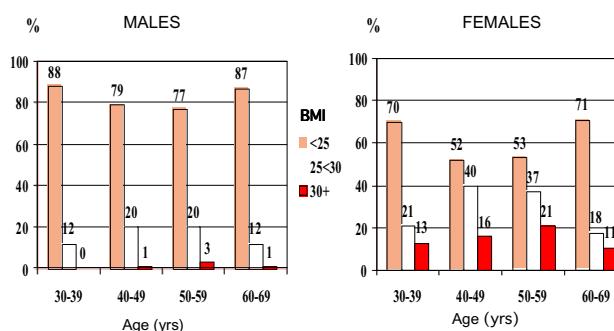
chewed betel nut, and the prevalence was strongly related to age (Figure 2). Only 1.2% of men reported betel nut chewing.

Alcohol consumption is more common in men, with 77.2% reporting alcohol drinking (ever or current), compared with 29.6% of women. Out of these men drinkers, 7.0% reported daily consumption. Less than 1% women reported daily drinking. Laokhao (rice whisky) is the most common alcoholic drink in this area, 79.5% of alcohol consumers among the men reporting Laokhao consumption. In women, beer is the most common form of alcohol consumed, with 77% of drinkers reporting beer consumption, but most were only occasional consumers (once a month).

The distribution of the cohort subjects by Body Mass Index is shown in Figure 3. Women are rather heavier than men (means for the cohort: women 24.5, men 22.5), and BMI tends to increase with age.

Among the women recruited, the average age at menarche was late ( $16.2 \pm 1.6$  years of age).

All but 11 of the women in the sample were married. Table 2 shows the distribution by parity (live and still births). 97.7% of women had one or more births, 46.2% of women had had 1–3 full term pregnancies, 29.2% had 4–5 pregnancies, and 22.2% had 6 or more. The age at first birth



**Figure 3. Distribution of BMI (% of subjects) within age groups, by sex**

is shown in Table 3; average age at first birth was  $21.7 \pm 4.1$ ). Use of contraception is shown in Table 4. About one third of the women had used oral contraceptives and 14% injectable contraceptives. Use was strongly related to age, and much less common among older women, with only 17.1% of women aged 60–69 reporting having used hormonal contraception and 2.9% having used injectables. 60.5% of women had never used any hormonal contraception.

In total, 23.4% of study subjects were positive for the presence of ova of O.V. in the stool sample. Table 5 shows the prevalence by age and sex. Prevalence of infection is rather higher in men than in women. There is little difference according to age in men, but prevalence increases with age in women. Table 6 shows intensity of infection, as eggs per gram of faeces, in males and females. Heavy infections (counts of 1,800 per gram or more) are rather more common in men.

Of the 10,553 women who received a PAP-test at the time of recruitment into the study, cellular abnormality was observed in 236 (2.3%). The corresponding cytological diagnosis is shown in Table 7.

The average level of fasting blood sugar in this cohort subjects in men (3,202 subjects) was  $97.4 \pm 29.0$  mg per 100

**Table 2. Females: Total Births (live and still birth) by Age Group**

| Age   | Total sample | Total responding | Total births |    |    |     |    |    |    |    |    |
|-------|--------------|------------------|--------------|----|----|-----|----|----|----|----|----|
|       |              |                  | 0            | 1  | 2  | 3   | 4  | 5  | 6  | 7  | 8+ |
| 30–39 | 73           | 71               | 0            | 4  | 33 | 22  | 8  | 2  | 1  | 1  | 0  |
| 40–49 | 174          | 158              | 5            | 12 | 39 | 42  | 29 | 22 | 5  | 3  | 1  |
| 50–59 | 173          | 168              | 4            | 6  | 17 | 29  | 39 | 18 | 25 | 15 | 15 |
| 60–69 | 80           | 75               | 2            | 3  | 4  | 7   | 7  | 13 | 8  | 11 | 20 |
| Total | 500          | 472              | 11*          | 25 | 93 | 100 | 83 | 55 | 39 | 30 | 36 |

\* Unmarried women

**Table 3. Females: Age at First Birth (by age)**

| Age   | Total sample | Total responding | Age at first birth |    |    |    |    |    |    |    |    |     |
|-------|--------------|------------------|--------------------|----|----|----|----|----|----|----|----|-----|
|       |              |                  | <18                | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | >25 |
| 30–39 | 73           | 71               | 10                 | 6  | 7  | 11 | 6  | 7  | 8  | 5  | 4  | 7   |
| 40–49 | 174          | 152              | 8                  | 20 | 15 | 21 | 19 | 19 | 20 | 7  | 7  | 16  |
| 50–59 | 173          | 164              | 13                 | 15 | 10 | 36 | 26 | 13 | 13 | 5  | 14 | 19  |
| 60–69 | 80           | 72               | 2                  | 7  | 5  | 18 | 12 | 11 | 2  | 4  | 4  | 7   |
| Total | 500          | 459              | 33                 | 48 | 37 | 86 | 63 | 50 | 43 | 21 | 29 | 49  |

**Table 4. Use of Contraception among Females**

| Age   | Contraception: ever used |            |
|-------|--------------------------|------------|
|       | Oral                     | Injectable |
| 30–39 | 46.3%                    | 37.3%      |
| 40–49 | 31.8%                    | 14.0%      |
| 50–59 | 32.3%                    | 9.1%       |
| 60–69 | 17.1%                    | 2.9%       |
| Total | 31.9%                    | 14.0%      |

**Table 5. Prevalence of Opisthorchis Viverrini (OV) Infection, by Age and Sex**

| Age group | Males |               | Females |               |
|-----------|-------|---------------|---------|---------------|
|           | Total | % OV positive | Total   | % OV positive |
| 30–39     | 555   | 33.7          | 1,814   | 21.2          |
| 40–49     | 2,039 | 27.3          | 4,464   | 19.8          |
| 50–59     | 2,342 | 26.3          | 4,449   | 22.0          |
| 60–69     | 1,218 | 29.4          | 2,018   | 23.1          |
| Total     | 6,154 | 27.9          | 12,745  | 21.3          |

ml (range 51.9–393.8), in women (7,401 subjects)  $98.0 \pm 32.4$  mg per 100ml (range 51.9–391.8). The percentage of subjects who had high level of blood sugar ( $>120$  mg per 100 ml) was 11.8% and 12.3% in men and women respectively.

The pattern of cancer in Khon Kaen Province, as recorded by the cancer registry in the years 1993–1997 is shown in Figure 4, as the crude incidence (per 100,000) for the age group 30–69. Liver cancer is the most common cancer of both men and women; second and third in importance are cancers of the cervix and breast in women, and cancers of the lung and large bowel in men (Parkin et al., 2002).

Record linkage between the database of the cancer registry and the study cohort was carried out for the first time in 2002, and again in February 2004. Table 8 shows the numbers of new (incident) cancer cases identified in the cohort subjects at the end of 2001 and 2003. As expected from the incidence rates reported by the registry, liver cancer is by far the most common cancer in the study population (363 cases by the end of 2003) followed by lung cancer (46 cases), cancer of the cervix uteri (44 cases), and cancers of the breast and large bowel (42 cases of each).

**Table 7. Epithelial Cell Abnormalities in Women with Positive Pap Smear at Recruitment to the Khon Kaen Cohort Study**

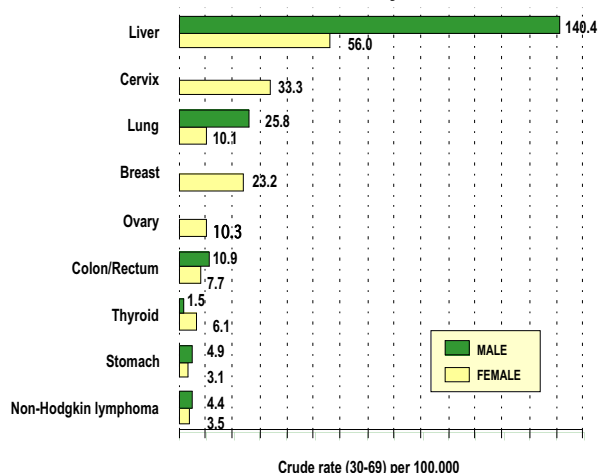
| Type of epithelial cell abnormalities                   | Number | %     |
|---------------------------------------------------------|--------|-------|
| Atypical squamous of undetermined significance (ASC-US) | 89     | 37.7  |
| Low grade squamous intraepithelial lesion (LSIL)        | 49     | 20.8  |
| High grade squamous intraepithelial lesion (HSIL)       | 44     | 18.6  |
| Squamous cell carcinoma                                 | 20     | 8.5   |
| ASC-US & Atypical endometrial cells (NOS)               | 1      | 0.4   |
| LSIL & Atypical glandular cells (NOS)                   | 1      | 0.4   |
| Atypical endometrial cells (NOS)                        | 1      | 0.4   |
| Atypical glandular cells (NOS)                          | 28     | 11.9  |
| Endocervical adenocarcinoma in situ                     | 1      | 0.4   |
| Adenocarcinoma                                          | 2      | 0.9   |
| Total                                                   | 236    | 100.0 |

**Table 6. Intensity of Opisthorchis Viverrini (OV) Egg Counts, by Sex**

| OV egg intensity per gram | Males  |      | Females |      |
|---------------------------|--------|------|---------|------|
|                           | Number | %    | Number  | %    |
| <500                      | 522    | 32.0 | 938     | 36.7 |
| 500–1,799                 | 490    | 30.1 | 837     | 32.7 |
| 1,800–9,999               | 474    | 29.1 | 616     | 24.0 |
| 10,000+                   | 144    | 8.8  | 168     | 6.6  |
| Total                     | 1,630  | 100  | 2,559   | 100  |

## Discussion

Although rather more than half of the new cancer cases in the world occur in the so-called “developing” countries (Parkin et al., 2005), the vast majority of epidemiological studies investigating genetic and environmental antecedents have been in North America and western Europe, representing only a fraction of the global population and their dietary patterns and lifestyle factors. There is an urgent need to increase our knowledge of cancer aetiology through studies elsewhere. Populations in the developing world provide opportunities to investigate a wider range of exposures (both qualitative and quantitative) than is possible in “western countries”, and many countries are in a

**Figure 4. Incidence of Major Cancers in Khon Kaen Province (ages 30–69) 1993–1997 (Parkin et al., 2002)**

**Table 8. Cancers Detected by Record Linkage between the Khon Kaen Study Cohort and the Khon Kaen Population-based Cancer Registry**

| Site                 | ICD10-Code  | By end 2001 | By end 2003 |
|----------------------|-------------|-------------|-------------|
| Lip                  | C00         | 5           | 5           |
| Tongue               | C02         | 3           | 4           |
| Mouth                | C03-C06     | 9           | 11          |
| Nasopharynx          | C11         | 7           | 9           |
| Oesophagus           | C15         | 4           | 4           |
| Stomach              | C16         | 16          | 19          |
| Colorectum           | C18-C21     | 36          | 42          |
| Liver                | C22         | 315         | 363         |
| Gallbladder etc.     | C23-C24     | 12          | 13          |
| Bronchus, lung       | C33-C34     | 36          | 46          |
| Other skin           | C44         | 19          | 20          |
| Breast               | C50         | 37          | 42          |
| Cervix uteri         | C53         | 38          | 44          |
| Ovary etc.           | C56         | 13          | 14          |
| Bladder              | C67         | 5           | 6           |
| Thyroid              | C73         | 13          | 15          |
| Non-Hodgkin lymphoma | C82-C85;C96 | 8           | 9           |
| Cervix in situ       | D06         | 15          | 16          |
| Others               |             | 67          | 80          |
| All sites            |             | 658         | 762         |

transitional period between traditional lifestyles and those characterised as “western” (Rastogi et al., 2004).

The cohort study is the best available technique in observational epidemiology for examining the association between exposures and a range of possible outcomes. It is particularly useful for exposures that are likely to be modified with the onset of disease, or for which substantial recall bias is likely to be present. The simpler case-control study design, with concurrent collection of biological specimens and questionnaire on past exposure history can be quite misleading in these circumstances. Multi-purpose large prospective studies are being conducted in several western countries. They collect questionnaire data together with biological samples stored at low temperature for future assessment; this strategy has the great advantage that the investigations that can be done are not constrained by current knowledge and technology, rather refined hypotheses and more advanced technology will be exploited at time of analysis, after years of follow-up (Rothman et al., 1995). Well-known examples are the Nurses’ Health Study in the USA (Colditz et al., 1987), and the EPIC study involving 23 centres in 10 European countries (Riboli et al., 2002).

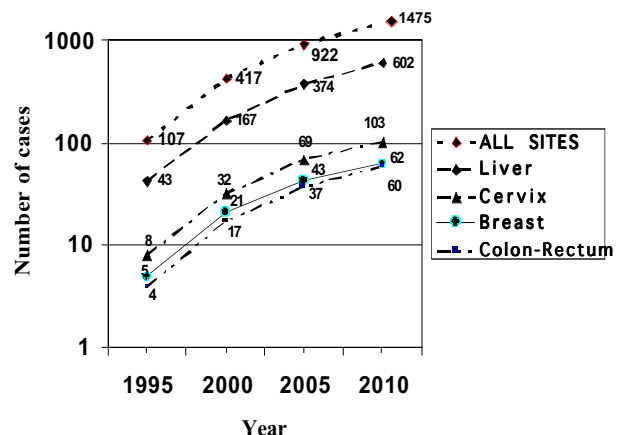
The use of cohort studies of lifestyle and cancer in developing countries has been much more limited, generally involving selected population groups, e.g. HBV carriers or EBV positive subjects, and addressing limited hypotheses (e.g. the incidence of hepatocellular carcinoma or nasopharyngeal cancer). The list of 17 large, general cohort studies collecting comprehensive food frequency questionnaires and blood samples on at least 10,000 adults, compiled by Langholz et al. (1999) includes only one in a developing country, the Shanghai Chinese Women’s Health

Study (Zhang et al., 2005). Other than this, an earlier cohort study had been set up in the population of Shanghai, China, although it originally included only men (Yuan et al., 1996; London et al., 2000), and a multi-centre study is in progress in Korea (Yoo et al., 2002).

The Khon Kaen cohort study evolved from a community project that aimed to improve health knowledge and practices in the population, and to detect asymptomatic disease. The province has an unusual cancer profile, with a very high incidence of liver cancer. These liver cancers are predominantly cholangiocarcinomas, and results from chronic infection with the liver fluke *Opisthorchis viverrini*, acquired by the local habit of consuming small raw cyprinoid fishes along with the dietary staple of sticky rice. The association between cholangiocarcinoma and liver fluke infestation has been the subject of much research locally (for review, see Vatanasapt and Sripa, 2000). Cancer of the cervix uteri is also relatively common (Sriamporn et al., 2003), but incidence rates of most of the cancers associated with western lifestyles (large bowel, lung, breast, prostate) are relatively low, as is the incidence of cancer of the stomach.

With some 25,000 subjects, the Khon Kaen cohort is modest in size, but as this relatively young population ages, it can be expected to yield increasing numbers of cancer cases for study. Figure 5 shows the a priori estimates of the number of cancer cases likely to occur in the cohort, assuming that the general life-table mortality of Thailand in the year 2000 (WHO, 2005) and the incidence rates observed in Khon Kaen Province in 1993-1997 (Parkin et al., 2002). Comparison with Table 8 suggests that these estimates may have been a little conservative, notably for cancers for which incidence rates have been increasing in recent years – breast and large bowel.

To date, the analytic strategy with respect to the relationship between exposure variables and cancer outcomes has used a nested case-control approach, whereby cases of the cancer of interest are identified by record linkage between the provincial cancer registry and the cohort database, and selection of an appropriate control group of non-cases (matching on such variables as sex, age, and date



**Figure 5. Estimated Accrual of Cancer Cases in the Cohort, 1995-2010**

of recruitment). Two studies, examining antecedents of cervix cancer in this population have been published (Sriamporn et al., 2004a, *in press*), and a study of liver cancer is in progress. The reasons for this approach are twofold. Firstly, only a selected subset of the information collected by questionnaire or physical examination has been entered into the computer database, and only those laboratory examinations of direct clinical relevance to the subjects have been carried out (Pap smear, blood sugar, OV eggs in stools). This limited dataset has been used in an ecological study of OV prevalence and liver cancer incidence (Sriamporn et al., 2004b), but otherwise, laboratory analyses on materials in the bio-bank, and data entry, have been confined to selected cases of disease and their matched controls. Secondly, no systematic follow-up of all the cohort members has been carried out, so we do not know which members are still under observation (and could be detected if they developed cancer), and which have died or emigrated. It is not possible therefore to estimate person-years of observation for different exposure categories.

Loss of cohort members due to death or emigration would result in over estimation of the power of the study with respect to different exposures and endpoints, by reducing the person-years of observation available. In the last five years, records of deaths occurring in the provincial population have become available on computer files (and are routinely linked to the cancer registry database), so that it is relatively simple to estimate losses due to deaths within the province. Losses due to emigration are likely to be rather small. The rural population, at least, those over the age of 30, is quite stable; in an ongoing study following up a sample of 1,290 women recruited in 2000-2001 we found that, in 2004, 9 had died, and only 13 (1%) had moved away or were otherwise untraceable.

The Khon Kaen cohort must be unique amongst major prospective studies in including subjects who are entirely from a rural, low-income agricultural population in a developing country. Literacy levels are low, and female fertility high, although in this community, marriage tends to take place relatively late, so that age at first birth among women in our cohort is relatively high.

As in many Asian populations, smoking prevalence among men is high (51%), although smoking is predominantly of hand-made cigarettes, using locally grown and cured tobacco. Although average daily consumption of smokers is not high, these cigarettes have been shown to have very high concentrations of tar and nicotine (Mitacek et al., 1991). In contrast, cigarette smoking is extremely rare amongst women (only 1.4%), in the National Health Survey of Thailand during 1991-1992, the average prevalence of "ever smoking" among Thai men aged 10 years and above was 68.6% (Institute of Health System Research, 1996), while in our cohort sample 83.6% men reported ever having smoked. The difference is possibly related to the older age group included in the Khon Kaen cohort. This is a finding typical of Asian populations in general (although there have been recent increases in smoking prevalence among younger

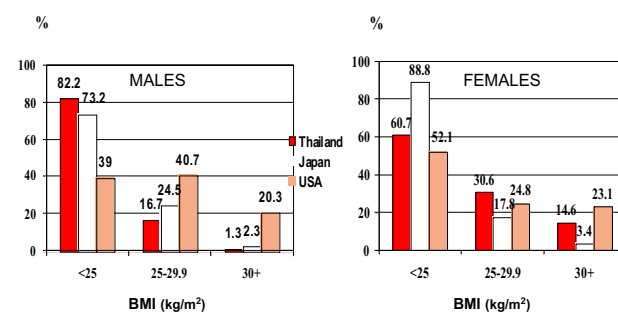
generations in Japan (WHO, 1997). In contrast, the habit of betel chewing, although rare in Eastern Asia, remains quite common among older women in North East Thailand (Simarak et al., 1977; Vatanasapt et al., 1991), although as our results show, this habit has almost entirely disappeared amongst the young.

The same survey (Institute of Health System Research, 1996), showed that in the Thai population (aged 10 years and above) 66.2% of men and 30.8% of women reported ever having consumed alcohol. On average, 5.2% of the sample reported frequent alcohol consumption, and of these frequent consumers, 81.7% were daily consumers. Men consumed more than 10 times the quantity of alcohol than women, a finding similar to our observation in the cohort subjects.

Figure 6 shows a comparison of the distribution of cohort subjects by BMI, with results from population surveys in the USA and Japan. Men in Khon Kaen are rather lean, with only 18% having a BMI of 25 or more (compared with 27% of Japanese men, and 59% of US whites). However, women are considerably heavier, with 14.6% in the "obese" category (BMI 30+); this is less than in the USA (23.1%), but a much greater proportion than in Japan (only 3.4%). Moderate, and increasing, levels of obesity have been noted in other populations of rural females in Asia (Adair, 2003; Ismael et al., 2004).

The biological specimens (blood, cervical cells) have been aliquoted and are stored in two different locations at  $-20^{\circ}\text{C}$ . The reason for not using storage at lower temperatures ( $-70^{\circ}\text{C}$ ) was simply the financial burden of the necessary equipment. It is recognised that analyses involving labile substrates (e.g. serum micronutrients) will therefore be impossible, but many highly relevant substrates, such as DNA, proteins, minerals, are adequately preserved at such temperatures, and can be studied in future.

The cohort has not yet been used to study associations between diet and cancer, despite the fact that the largest component of the questionnaire included an assessment of food intake. Nor has it been used, so far, to investigate endpoints other than cancer, although much of the exposure information collected is relevant to other diseases, such as cardiovascular disease and diabetes. The constraint here is



**Figure 6. BMI (% of subjects) in Khon Kaen cohort\* compared with Japan¹ and USA²**

\* Age standardised to European Standard Population

1. Ages 20+: from National Nutrition Survey of Japan (2000)

2. Ages 20-74: from NHANES IV (1999-2000)

the absence, to date, of an active follow-up to detect the occurrence of disease events – the cancer studies are relying upon passive surveillance by the cancer registry. However, with the efforts that have been made in the last decade to improve the quality of the cause of death statements on death certificates, the potential for linkage with the mortality register would be worth exploring.

## Acknowledgements

This research is part of the project “a prospective study of cancer and other outcomes in a rural population in Thailand”, which was approved by the research ethic committee, Faculty of Medicine, Khon Kaen University on February 11, 1997. The study was supported in part by grants from Buddhist Aid Center of Japan, and the Association for International Cancer Research Scotland, UK, (grant number 98-55), whose contributions are gratefully acknowledged. We are grateful to all the health personnel who volunteered to assist with the mobile cancer screening programme, which was the basis of this study, Dr.Winyou Wongpratoom for the blood sugar testing, the Khon Kaen Cancer Registry has been directed by Dr.Surapon Wiangnon since 2000, and we also acknowledge the contribution of the registry personnel Ms.Sujinan Horasith, Ms.Pakanant Usantia, Ms.Prasitporn Meunhast, Ms.Chaveeluck Sirikampeng.

<sup>1</sup>RECLINK is a record linkage software developed at unit of Descriptive Epidemiology, International Agency for Research on Cancer, Lyon. The software performs probabilistic linkage between records from different sources using selected personal identifiers. National ID number, name and surname, date of birth, sex and residence were used for this purpose. The program allows for differences in the spelling and in the date of birth. It computes a score of the likelihood that two records refer to the same subject. This is based on weights assigned to the variables involved expressing their discriminating value (e.g. low in the case of sex), and the likelihood that it is correctly recorded (e.g. low for date of birth).

## References

- Adair LS (2004). Dramatic rise in overweight and obesity in adult filipino women and risk of hypertension. *Obes Res*, **12**, 1335-41.
- Colditz GA, Stampfer MJ, Willett WC, et al (1987). Reproducibility and validity of self-reported menopausal status in a prospective cohort study. *Am J Epidemiol*, **126**, 319-25.
- Hirayama T (1990). A large scale cohort study on the effect of life styles on the risk of cancer by each site. *Gan No Rinsho. Spec No*: 233-42.
- Institute of Health System Research (1996). The first National Health Survey during 1991-1992. Bangkok, Thailand.
- International Labour Organisation (1988). International Standard Classification of Occupations ISCO-88. [online] 2004 [cited 2005 May 20] <http://www.ilo.org/public/english/bureau/stat/isco/isco88/index.htm>.
- Ismail MN, Chee SS, Nawawi H, et al (2002). Obesity in Malaysia. *Obes Rev*, **3**, 203-8.
- Iwasaki M, Otani T, Yamamoto S, et al (2003). Background characteristics of basic health examination participants: the JPHC Study Baseline Survey. *J Epidemiol*, **13**, 216-25.
- Langholz B, Rothman N, Wacholder S, Thomas DC (1999). Cohort studies for characterising measured genes. *J Natl Cancer Inst Monogr*, **26**, 39-42.
- London SJ, Yuan JM, Chung FL, et al (2000). Isothiocyanates, glutathione S-transferase M1 and T1 polymorphisms, and lung-cancer risk: a prospective study of men in Shanghai, China. *Lancet*, **356**, 724-9.
- Mitacek EJ, Brunnemann HD, Polednak AP, Hoffman D, Suttajit M (1991). Composition of popular tobacco products in Thailand, and its relevance to disease prevention. *Prev Med*, **20**, 764-73.
- Parkin DM, Bray F, Ferlay J, Pisani P (2005). Global cancer statistics, 2002. *CA Cancer J Clin*, **55**, 74-108.
- Parkin DM, Whelan SL, Ferlay J, Teppo L, Thomas DB eds (2002). Cancer Incidence in Five Continents, Vol. VIII. IARC Scientific Publications No. 155. Lyon:International Agency for Research on Cancer.
- Qian GS, Ross RK, Yu MC, et al (1994). A follow up study of urinary markers of aflatoxin exposure and liver cancer risk in Shanghai, Peoples Republic of China. *Cancer Epidemiol Biomarker Prev*, **3**, 3-10.
- Rastogi T, Hildeheim A, Sinha R (2004). Opportunities for cancer epidemiology in developing countries. *Nat Rev Cancer*, **4**, 909-17.
- Riboli E, Hunt KJ, Slimani N, et al (2002). European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutri*, **5**, 1113-24.
- Rothman N, Stewart WF, Sculte PA (1995). Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories. *Cancer Epidemiol Biomarkers Prev*, **4**, 310-11.
- Samet JM, Munoz A (1998). Evolution of cohort study. *Epidemiol Rev*, **20**, 1-14.
- Simarak S, de Jong UW, Breslow N, et al (1977). Cancer of the oral cavity pharynx/larynx and lung in North Thailand: case-control study and analysis of cigar smoke. *Br J Cancer*, **36**, 130-40.
- Sriamporn S, Parkin DM, Pisani P, Suwanrungruang K, Pengsaa P (2004a). Behavioural risk factors for cervical cancer from a prospective study in Khon Kaen, Northeast Thailand. *Cancer Detect and Prev*, **28**, 334-9.
- Sriamporn S, Pengsaa P, Hakama M, Suwanrungruang K, Parkin DM (2003). Cervix cancer in Khon Kaen, northeast Thailand, 1985-1999. *Asian Pacific J Cancer Prev*, **4**, 312-8.
- Sriamporn S, Pisani P, Pipitgool V, et al (2004b). Prevalence of Opisthorchis viverrini infection and incidence of cholangiocarcinoma in Khon Kaen, Northeast Thailand. *Trop Med Int Health*, **9**, 588-94.
- Sriamporn S, Snijders PJF, Pientong C, et al (in press). Human papillomavirus and cervical cancer from a prospective study in Khon Kaen, Northeast Thailand. *Int J Gynecol Cancer*.
- Vatanasapt V, Pengsaa P, Tangvoraphonkchai V, et al (1992). Thailand: Khon Kaen Cancer Registry. In Parkin DM, Muir CS, Whelan SL, Gao TT, Ferlay J and Powell J (eds). Cancer Incidence in Five Continents, Vol. VI. IARC Scientific Publications No. 120, pp 546-549. Lyon: International Agency for Research on Cancer.

- Vatanasapt V, Sriamporn S, MacLennan R (1991). Contrasts in risk factors for cancers of the oral cavity and hypopharynx and larynx in Khon Kaen, Thailand. In: Varma (Ed). Oral Oncology Volume II, Proceedings of International Congress on Oral Cancer, pp 39-42. Delhi: MacMillan Ltd.
- Vatanasapt V, Sripa B (eds) (2000). Liver Cancer in Thailand, Epidemiology, Diagnosis and Control. Khon Kaen, Siriphan Press.
- Vatanasapt V, Tangvoraphonkchai V, Titapant V, et al (1990). A high incidence of liver cancer in Khon Kaen Province, Thailand. *Southeast Asian J Trop Med Public Health*, **21**, 489-94.
- Viriyapanich T, Suansilpong P, Chittchang U, Banjong O (1990). Conversion factors: modified 24 hour dietary recall method. The Institute of Nutrition, Mahidol University, Bangkok, Thailand.
- Webb PM, Yu MC, Forman D, et al (1996). An apparent lack of association between *Helicobacter pylori* infection and risk of cancer in China. *Int J Cancer*, **67**, 603-7.
- World Health Organization (1997). Tobacco or Health. A Global Status report. Geneva: WHO.
- World Health Organization (2005). WHO Statistical Information System (WHOSIS). Life Tables for 191 countries. [online] [cited 2005 May 20] tables/life\_tablesé [http://www3.who.int/whosis/life/life\\_tables/life\\_tables](http://www3.who.int/whosis/life/life_tables/life_tables).
- Yoo KY, Shin HR, Chang SH, et al (2002). Korean Multi-centre Cancer Cohort Study including a Biological Materials Bank (KMCC-I). *Asian Pac J Cancer Prev*, **3**, 85-92.
- Yuan JM, Ross RK, Wang XL, et al (1996). Morbidity and mortality in relation to cigarette smoking in Shanghai, China. A Prospective male cohort study. *JAMA*, **275**, 1646-50.
- Zhang X, Shu XO, Yang G, et al (2005). Association of passive smoking by husbands with prevalence of stroke among Chinese women nonsmokers. *Am J Epidemiol*, **161**, 213-8.