Acta Oncologica, 2007; 46: 286–307

informa
healthcare

ORIGINAL ARTICLE

Nordic biological specimen banks as basis for studies of cancer causes and control – more than 2 million sample donors, 25 million person years and 100 000 prospective cancers

EERO PUKKALA[1,2], AAGE ANDERSEN[3], GÖRAN BERGLUND[4], RANDI GISLEFOSS[3], VILMUNDUR GUDNASON[5], GÖRAN HALLMANS[6], EGIL JELLUM[7], PEKKA JOUSILAHTI[8], PAUL KNEKT[8,9], PENTTI KOSKELA[8], P. PENTTI KYYRÖNEN[1], PER LENNER[10], TAPIO LUOSTARINEN[1], ARTHUR LÖVE[11], HELGA ÖGMUNDSDÓTTIR[12], PÄR STATTIN[13], LEENA TENKANEN[14], LAUFEY TRYGGVADÓTTIR[15], JARMO VIRTAMO[8], GÖRAN WADELL[16], ANDERS WIDELL[17], MATTI LEHTINEN[2] & JOAKIM DILLNER[17]

[1]*Finnish Cancer Registry, Institute for Statistical and Epidemiological Cancer Research, Helsinki, Finland,* [2]*School of Public Health, University of Tampere, Finland,* [3]*The Cancer Registry of Norway, Institute of Population-Based Cancer Research, Oslo, Norway,* [4]*Malmö Diet and Cancer Study, Lund University, Malmö, Sweden,* [5]*Icelandic Heart Association, Kópavogur, Iceland,* [6]*Department of Public Health and Clinical Medicine, Nutritional Research, Umeå University, Umeå, Sweden,* [7]*Institute of Clinical Biochemistry, Rikshospitalet University Hospital, Oslo, Norway,* [8]*National Public Health Institute, Helsinki and Oulu, Finland,* [9]*Social Insurance Institution, Helsinki and Turku, Finland,* [10]*Cancer Registry of Northern Sweden and Department of Radiation Sciences, Umeå University Hospital, Umeå, Sweden,* [11]*Department of Medical Virology, Landspitali-University Hospital, University of Iceland, Reykjavik, Iceland,* [12]*Molecular and Cell Biology Laboratory, Icelandic Cancer Society, Reykjavik, Iceland,* [13]*Department of Urology, Umeå University Hospital, Umeå, Sweden,* [14]*Helsinki Heart Study, Helsinki, Finland,* [15]*Icelandic Cancer Registry, Reykjavik, Iceland,* [16]*Department of Virology, University of Umeå, Sweden and* [17]*Department of Medical Microbiology, Lund University, University Hospital at Malmö, Sweden*

Abstract

The Nordic countries have a long tradition of large-scale biobanking and comprehensive, population-based health data registries linkable on unique personal identifiers, enabling follow-up studies spanning many decades. Joint Nordic biobank-based studies provide unique opportunities for longitudinal molecular epidemiological research. The purpose of the present paper is to describe the possibilities for such joint studies, by describing some of the major Nordic biobank cohorts with a standardised calculation of the cancer incidence in these cohorts. Altogether two million donors have since 1966 donated more than four million biological samples, stored at -20°C to -135°C, to 17 biobank cohorts in Finland, Iceland, Norway and Sweden. As a result of joint database handling principles, the accuracy of personal identifiers and completeness of follow-up for vital status in all participating biobanks was improved. Thereafter, the cancer incidence was determined using follow-up through the national cancer registries. Biobanks based on random samples of population typically showed slightly lower cancer incidence rates than the general population, presumably due to better participation rates among health-conscious subjects. On the other hand, biobanks including samples for viral screening or clinical testing showed 1.5 to 2.1 fold increased incidence of cancer. This excess was very high immediately after sampling, but for some cancer sites remained elevated for years after clinical sampling. So far, more than 100 000 malignant neoplasms have occurred after sample donation, and the annual increase of the cancer cases in these cohorts is about 10 000. The estimates on the population-representativity of the biobanks will assist in interpretation of generalizability of results of future studies based on these samples, and the systematic tabulations of numbers of cancer cases will serve in study power estimations. The present paper summarizes optimal study designs of biobank-based studies of cancer.

Correspondence: Eero Pukkala, Finnish Cancer Registry, Institute for Statistical and Epidemiological Cancer Research, Liisankatu 21 B, FI-00170 Helsinki, Finland. Tel: +358 9 135 33 224. Fax: +358 9 135 5378. E-mail: eero.pukkala@cancer.fi

(Received 20 October 2006; accepted 5 January 2007)

ISSN 0284-186X print/ISSN 1651-226X online
DOI: 10.1080/02841860701203545

In their classical assessment of the quantitative importance of avoidable causes of cancer, Doll and Peto estimated that a majority of human cancer was attributable to avoidable causes [1]. They concluded that most of these avoidable causes remained unidentified. For risk factor identification and causality inference as well as for studies searching for mechanisms behind increases or decreases in cancer incidence they recommended the use of prospective studies nested in cohorts of stored biological specimens. In the Nordic countries, there exists a series of established biological specimen banks with many decades of follow-up that enable performing prospective epidemiological studies with adequate statistical power even for diseases and exposures that are not common.

This paper introduces the Nordic biobank network NBSBCCC (Nordic Biological Specimen Banks working group on Cancer Causes & Control) to new potentially interested partners and serves as a general reference for specific studies based on these biobanks. NBSBCCC is a network of excellence that contains 17 independent biobank cohorts, five cancer registries and numerous expert user groups. The purpose of the network is to provide a concerted resource for etiologic studies of cancer, with a focus on longitudinal studies addressing unexplained causes and trends over time.

Since 1995, more than 30 joint network articles have been published [2–33]. The majority of studies so far have been aimed at elucidating infections as causes of cancer using prospective study designs, with some of the most notable findings being prospective studies on the role of Human Papillomavirus (HPV) infection and cofactors to HPV in relation to a number of cancer forms. In addition to the joint Nordic studies, the biobanks operate independently with several hundred publications based on one or several of the biobanks described in this paper. Major subject areas for study have been hormones, nutrition, smoking, organochlorine compounds and genetic polymorphisms as causes of cancer in addition to a number of studies evaluating tumour markers. There has so far not been any systematic evaluation on characteristics and quality of the biobank cohorts or features of cancer risk pattern among the donors.

This paper includes systematic descriptions of the participating biobanks: background, organisation, size, years of sample collection, and administrative aspects. Numbers of cancer cases found among persons in the serum banks after serum drawing are given, advertising the unique possibilities of the national cancer registration systems in the Nordic countries. Population representativeness of the serum bank cohorts is estimated by comparing cancer incidence in the biobank cohorts with the respective national rates. Finally, issues to be taken into account in designing case-control studies nested in the Nordic biobanks are discussed.

Participating biobanks

The network of Nordic biobanks used in nested case control studies under NBSBCCC programme so far consists of 12 biobanks in Finland, Iceland, Norway and Sweden, three of which are split into 2–3 independent subcohorts (Figure 1, Table I). Participating biobanks are independent entities that make their own decisions, but are committed to facilitate joint studies by working towards similar policies for quality assurance, logistics and study designs as well as for permission and terms of collaboration. NBSBCCC is funded by the Nordic Council of Ministries and as a European Union sixth framework programme Network of Excellence.

Research projects using the biobanks need appropriate permissions from both the national Data Protection Authorities, National or Local Ethical Committees and from the boards of the biobanks. Today, an informed consent is collected from all persons donating samples, making it clear to the donors that the material will be used for future research purposes. Details of the permission procedure can be achieved via contact email addresses given in Table I.

All samples have been stored at -20°C to -25°C except those of the Alpha-Tocopherol, Beta-Carotene Cancer (ATBC) Prevention Study, the Northern Sweden Health and Disease Study, the Malmö Diet and Cancer study and FINRISK Study (since 1997) which are stored at -70°C. Malmö Diet and Cancer biobank also has aliquots stored at -135°C.

Every resident of the Nordic countries has a unique personal identification number or code (PID) that is used in all main registers in these countries. The PID allows automatic and precise linkage of registers, without need to use names. For meaningful research use, the PIDs have to be available for each person in the biobanks. Biobank cohorts are typically linked with the population-based cancer registries shortly before a new case-control set will be extracted for a specific study.

Finnish Maternity Cohort

Sera collected during the first trimester of pregnancy (two thirds at 8 to12 weeks) for screening of congenital infections and rubella immunity have been stored since late 1983 by the National Public Health Institute of Finland. The biobank covers more than 98% of all pregnant women in Finland.

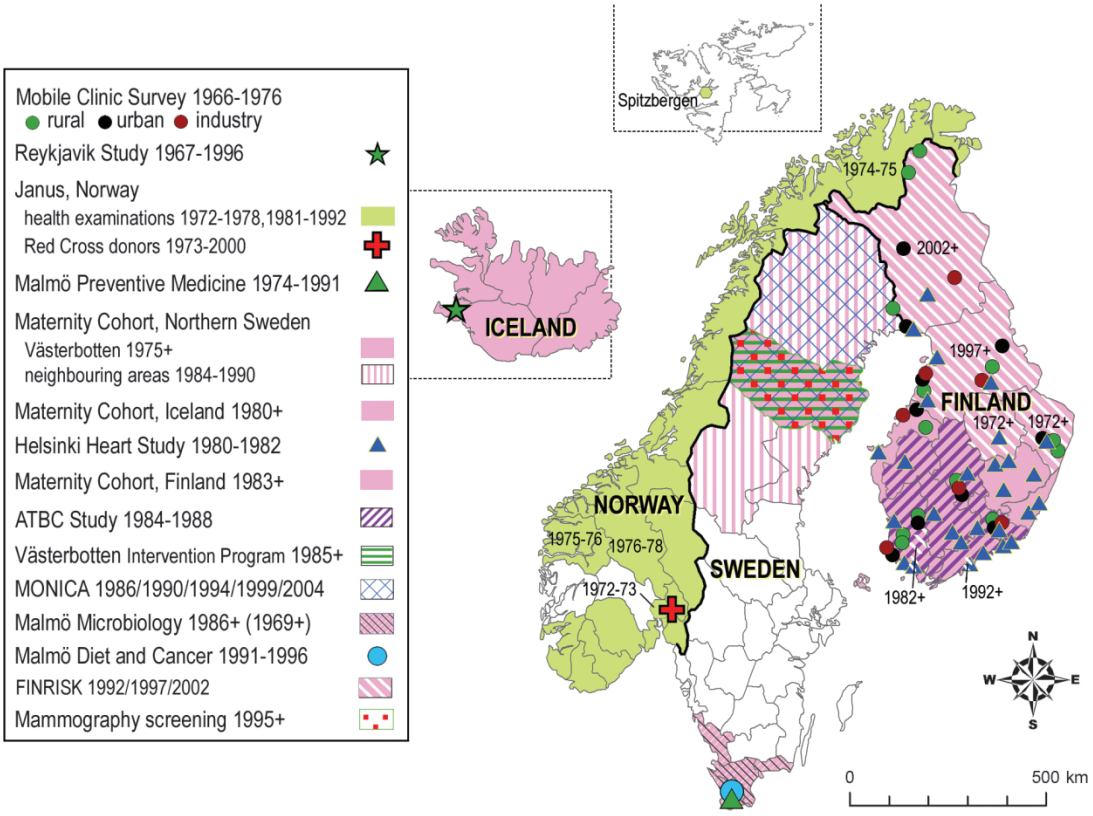

Figure 1. Map of Nordic countries indicating the coverage areas of the serum banks.

So far, basic data for the sera until 21 August 2005 have been transferred to Finnish Cancer Registry to be used in case-control studies.

The Finnish Cancer Registry takes care of quality control of the data and also has a program for precise random case-control selection. Record linkages for both incident cancers with cancer registry data and causes of death through Statistics Finland are administered from the Finnish Cancer Registry. This biobank has been used in more NBSBCCC studies than any other [6,7,10–17,19,21–25,27,28,30,31,33].

Helsinki Heart Study

The sera were collected during 1980–1982 for a trial to test the hypothesis that lowering serum LDL-cholesterol and triglyceride levels and elevating serum HDL-cholesterol levels with gemfibrozil (a fibric acid derivative) reduces the incidence of coronary heart disease (CHD) in middle aged dyslipidemic men [34]. The volunteers for the trial were selected from men aged 40 through 55 years, employed by two government agencies and five industrial companies and living in different parts of Finland. About 19 000 men participated in the first screening and to be selected for the trial the participants (n = 4081) had to have their non-HDL cholesterol $\geq$5.2 mmol/l and no evidence of CHD or other major disease.

Serum samples were collected from the participants at the first screening and from the participants in the trial at each follow-up visit during the trial. As the participants were followed-up four times per year during the 5-year trial and twice a year during a subsequent extension of the trial, there are 28 serial samples from about 3 500 of those 4 081 that initially attended the trial. From the last trial follow-up visit also blood samples were stored. This biobank has participated in NBSBCCC studies since 1999 [16,18,20,26,29,33].

Alpha-Tocopherol-Beta-Carotene (ATBC) cancer prevention study

The ATBC study was a randomised, double-blind, placebo-controlled, primary prevention trial conducted in Finland by the National Public Health Institute in collaboration with the U.S. National Cancer Institute. The main aim of the study was to evaluate whether daily supplementation with alpha-tocopherol or beta-carotene would reduce the incidence of lung cancer and other cancers [35].

In 1985–1988, a questionnaire on current smoking and willingness to participate in the trial was sent to the total male population of 50 to 69 years living

Table I. Characteristics of the serum banks included in the Nordic Biological Specimen Banks for Cancer Causes and Control (NBSBCCC) network. Status as of June 2005. Type: R = random sample of population or other systematic invitation based on population register; S = specific group with clearly defined enrolment criteria; C = Clinical samples.

Name, country [contact address]	Type	Years of first serum donation and subsequent samples of same individuals	Number of persons [+annual increase]	Number of sampling occasions [+annual increase]	Closing year in this study (complete cancer incidence & vital status)1	Number of person-years
Finnish Maternity Cohort [pentti.koskela@ktl.fi]	S	1983 (continues) [samples from 1997+ not included in the present analyses]	723 000 women (Aug 2005) [+30 000 /year]	1.47 million [+ 60 000/year]	2003	6.69 million
Helsinki Heart Study, Finland [leena.tenkanen@uta.fi]	S	1980–1982	19 000 men	117 000	2003	390 000
Alpha-Tocopherol-Beta-Carotene (ATBC) Study, Finland [jarmo.virtamo@ktl.fi]	S	baseline sera 1984–1988, follow-up sera from all 1986–1993, annual sera from 800 men	29 133 male smokers	55 000	2003	393 000
Finnish Mobile Clinic Health Examination Survey [paul.knekt@ktl.fi]	R	1966–1976	50 448	60 000	2003	1.36 million
FINRISK, Finland [pekka.jousilahti@ktl.fi]	R	1992/1997/2002 (samples from years 1972/1977/1982/1987 incomplete)	22 910	22 910	2003	143 000
Icelandic Maternity Cohort [arthur@landspitali.is; Arthur Löve]	S	1980+	47 820 women [+1,700/year]	91 000 [+2,500/year]	2002	618 000
Reykjavik Study, Icelandic Heart Association [v.gudnason@hjartavernd.is]	R	1967–1996	19 300	60 000	2002	423 000
Janus, Norway [randi.elin.gislefoss@kreftregisteret.no]		1972–2005	331 801	493 400	2001	
* Health examinations	R	1972–1978, 1980–1992 (last samples from Finmark and Troms counties in 2002)	304 342	376 600	2001	5.23 million
* Blood donors	S	1973–1991, 1998–2000 (last samples 2005)	31 922	116 800	2001	693 000
Northern Sweden Health and Disease Study (NSHDS) [goran.hallmans@nutrires.umu.se]	R	1985+	86 000	114 000	2003	
* VIP	R	1985+	70 000 [+2000/year]	83 000	2003	560 000
* MONICA	R	(1986)/ 1990/ 1994 / 1999/ 2004	9 000	14 000	2003	51 000
* Mammography	R	1995+	27 500 women [+1500/year]	48 000	2003	158 000

Table I (*Continued*)

Name, country [contact address]	Type	Years of first serum donation and subsequent samples of same individuals	Number of persons [+annual increase]	Number of sampling occasions [+annual increase]	Closing year in this study (complete cancer incidence & vital status)1	Number of person-years
Northern Sweden Maternity Cohort (NSMC) [goran.wadell@climi.umu.se]	S	1975+	86 000 women [+2000/year]	118 000 samples	2003	1.24 million
Preventive Medicine in Malmö (PMM), Sweden [goran.berglund@medforsk.mas.lu.se]	R	1974–1991	33 000	8 000	1999	560 000
Malmö Diet and Cancer, Sweden [goran.berglund@medforsk.mas.lu.se]	R	1991–1996	29 098	–	1999	159 000
Malmö Microbiology, Sweden [joakim.dillner@med.lu.se]		1986+ (incomplete 1969+)	454 000 [+40 000/year]	1.24 million [+120 000/year]	1999	1.84 million
* Malmö Maternity Cohort	S	1985, 1989+ (incomplete 1969+)	70 000 women	115 000	1999	
* Bloodborne virus screening	C	1986+			1999	
* Other virus testing	C	1990+			1999	
Swedish Institute for Infectious Disease Control (SIIDC) [joakim.dillner@med.lu.se]		1957+ (complete 1977+)	358 000 in computerised files	>900 000 (629 000 computerised)	–	–
* Population sample	R	1968, 1977–1978, 1990–1991, 1997	12 045	12 000	–	–
* Diagnostic microbiological testing	C	1990+		617 000 computerised	–	–

in south-western Finland (n = 290 000). Of them, 43 000 men smoked at least five cigarettes per day and were willing to participate. Men with prior cancer (except non-melanoma skin cancer and carcinoma-in-situ), severe angina pectoris, chronic renal insufficiency, alcoholism, or liver cirrhosis were excluded as well as those taking anticoagulants, beta-carotene, or vitamin A/E supplements in excess of defined doses. After exclusions and written informed consent, 29 133 eligible men were randomly assigned to receive either alpha-tocopherol 50 mg per day, or beta-carotene 20 mg per day, or both alpha-tocopherol and beta-carotene, or placebo.

At baseline, serum samples were collected. New serum samples were collected from all participants at the 3-year follow-up visit, and from about 800 randomly selected men a serum sample was collected annually throughout the trial. A whole blood sample was collected from the participants at the end of the trial between August 1992 and April 1993. This biobank was used for the first time in a NBSBCCC study just lately [33].

Finnish Mobile Clinic Health Examination Survey

The Mobile Clinic Health Examination Survey was carried out by the Social Insurance Institution during 1966–1972 in 34 rural, industrial or semi-urban subpopulations (Figure 1). Total populations aged 15 years or older or random samples of them were invited to participate in the study. On average 83% (57 400 men and women) participated in the health examination. Blood samples have been stored from 40 200 individuals in the baseline examination and from all 19 500 individuals in the re-examination survey of 12 subpopulations four to seven years later (1973–1976). This biobank participated particularly in early NBSBCCC studies [2–4,9].

FINRISK

The National FINRISK Study has been conducted in Finland every five years since 1972. At the beginning the Study was done only in eastern Finland as part of the North Karelia Project. The study area was expanded gradually. The serum samples are sysmatically available since 1992. In 1992 the Study was carried out in four areas: North Karelia and Kuopio Provinces in Eastern Finland, Turku-Loimaa region in Southwest Finland, and cities of Helsinki and Vantaa in Southern Finland. Oulu province in Northern Finland was included in 1997 and Lapland province in 2002. In each study year, a random sample of 2 000 individuals aged 25 to 64 years (stratified by sex and 10-year age group) has been taken in each study area according to the WHO MONICA protocol. Since 1997 a sub-sample of 1 500 men and women aged 65 to 74 years was included. Total cumulative sample size since 1992 is 33 000 and out of them 22 908 individuals (69%) have participated in the Study. DNA samples are available for most participants.

Study cohorts have been followed-up through computerized register linkage of the National Causes of Death Register, the Hospital Discharge Register and the Finnish Cancer Register. The samples of the FINRISK Study have not been used in any NBSBCCC studies so far, but the general principle of the study is that the collected samples can be utilised in large-scale collaborative studies that according to the FINRISK Steering Group are scientifically important.

Icelandic Maternity Cohort

Sera generally collected at 12 to 14 weeks of pregnancy for rubella screening from all of Iceland have been stored since 1980 in the centralized Department of Medical Virology, Landspitali University Hospital. About 6% of the cohort members cannot be used in studies because they have moved out of the country, but the date of emigration is not registered. This biobank has participated in two NBSBCCC studies [25,28].

Icelandic Heart Association, The Reykjavik Study

The Reykjavik Study by the Heart Preventive Clinic and Research Institute of the Icelandic Heart Association is a prospective cardiovascular cohort study carried out in the Reykjavik capital area in 1967 to 1996. Selected birth cohorts of 14 923 men and 15 872 women in the Reykjavik area born in 1907 to 1935 were divided into six equally sized subgroups according to the date of birth and recruited systematically for collection of sera. The first subgroup was recruited in 1967–1969 and has attended altogether six times. The second one (first invited in 1970–1972) has attended twice. The later birth cohorts have been invited once (1974–1996) or never. Altogether 19 300 persons actually provided samples (annual participation rates between 71% and 76%), but about 200 of them cannot be used in analyses because of lacking dates of emigration. This biobank has not yet been used in any of the published NBSBCCC studies.

Janus Project (Norway)

A project to collect and store blood samples from healthy persons for later scientific use was initiated in the 1960s and named Janus after the Roman god with two faces, one looking backward, and other one

looking forward (symbolizing the retrospective and prospective directions of epidemiological research). The first collection, related to a survey of risk factors for cardiovascular disease in ages of 35–49 years, covered four counties (Oslo 1972–1973, Finmark 1974–1975, Sogn og Fjordane 1975–1976 and Oppland 1976–1978; see Figure 1). More subjects were added during 1985–1992 in the context of cardiovascular health examination of 40–42 years-old Norwegians from all of the country except two counties (Hordaland and Buskerud; Figure 1).

Red Cross blood donors in capital Oslo and surrounding areas were enrolled in 1973–1991 and 1999–2000. Every second year, these Janus donors donated 20 ml of extra blood to the biobank. Collection of later samples from these individuals ended in spring 2005.

The Janus bank consists of serum samples from 331 801 persons, 10% of them Red Cross donors. The average is 2–3 samples per donor, but some donors have given samples more than 10 times. The Janus biobank is also collecting follow-up samples from cohort members who develop cancer. Before any treatment, a sample is collected when the donor is hospitalized at the Radium Hospital in Oslo (a nationally centralized cancer treatment hospital).

The Janus Project is today funded by the Norwegian Cancer Registry which is also responsible for the data handling. This also allows frequent updates for incident cancer cases; several thousands of new prospective cancer cases have been registered after the closing date used in the present study, and the addition in 2004 exceeded 3 000. Samples of the Janus health examination cohort have been used in 20 NBSBCCC publications [6–8,11,14–24,26, 27,29,31,33], and the blood donors' sera in eight studies [6,8,18–20,26,29,33].

The Northern Sweden Health and Disease Study (NSHDS) Cohort

The Northern Sweden Health and Disease Study Cohort contains three subcohorts: the Västerbotten Intervention Program (VIP), the MONICA (Monitoring Trends and Determinants in Cardiovascular Disease) and the Mammography Screening in Västerbotten. The cohorts represent a population-based sample of the county of Västerbotten in Northern Sweden (254 000 inhabitants). The Monica study also contains a population-based sample from the adjacent county of Norrbotten.

The VIP is a long-term project intended for health promotion. Since 1985, all individuals of 40, 50 and 60 years of age are invited for screening. They are also asked to donate a blood sample for later research purposes. In June 2004, the cohort included 74 000 individuals, of whom 70 000 had donated blood. A second sample is taken after 10 years; this has produced 13 000 re-sampling occasions.

Samples taken in the context of the population-based mammography screening have been stored since 1995. Screening is done every second year among all women in the age group 50–69 years in the county. There have been 48 000 sampling occasions from 27 500 women. About 50% of the women in the mammography cohort have also attended VIP.

The Northern Sweden MONICA project contains material from population-based screenings for risk factors of cardiovascular diseases that were carried out in 1986, 1990, 1994, 1999 and 2004. There are 14 000 sampling occasions of 9 000 individuals, 50% of whom are also included in VIP. Samples from 1986 have not been used in NBSBCCC studies and they are not included in this SIR analysis, either.

The VIP cohort has been used most frequently out of the numerous Swedish biobanks in NBSBCCC studies [7,11,12,14–18,20–24,26,27,29,31–33]. MONICA cohort has been utilised in 15 studies [7,11,12,14,17,18,20–24,26,27,29,31].

Northern Sweden Maternity Cohort (NSMC)

Northern Sweden Maternity cohort consists of sera collected since 1975 from pregnant women screened for rubella immunity during week 14 of pregnancy in the Västerbotten county and especially in the 1980s also for some of the adjacent counties in Northern Sweden. So far, almost 120 000 samples from 86 000 women have been stored at the virus laboratory of Umeå University. This biobank has not yet been used in any of the published NBSBCCC studies.

Preventive medicine in Malmö, Sweden

The prospective, population-based Preventive Medicine study, with main focus on cardiovascular disease, diabetes and cancer, includes sera from a population-based sample of 33 444 persons 40–60 years of age, resident in the city of Malmö. The samples were donated at baseline examination in 1974–1991. The biobank is owned by Lund University.

Malmö Diet and Cancer study, Sweden

The prospective population-based Malmö Diet and Cancer study started with a baseline examination in 1991–1996. Main focus is on cancer and cardiovascular diseases. All men born between 1923 and 1945 and all women born between 1923 and 1950 living at the time in the city of Malmö were invited to participate. The participation rate was 40% (28 098

participants). Mean age at enrolment was 58.2 years. The biobank is owned by Lund University.

Malmö Microbiology Biobank (MMB), Sweden

The Malmö Microbiology Biobank is owned by the County Council of Skåne and contains samples submitted for clinical microbiological analyses to the University Hospital in Malmö that today serves the entire county of Skåne in southernmost Sweden. Samples have been saved for clinical diagnostic and documentation purposes, the majority of them taken for diagnosis of blood-borne viral infections such as hepatitis viruses. The oldest samples are from 1969 and were submitted from the city of Malmö. The annual number of samples increased in 1986 when HIV testing started and the catchment area extended to cover most of Skåne county. Since 1990, also the samples submitted for virus serology (typically because of clinical suspicion of virus infection or desire to investigate viral immunity) have been stored. In recent years a large number of samples has been submitted from the microbiology laboratories of adjacent counties in southern Sweden (Blekinge and Halland), raising the annual number of samples added to the biobank to about 60 000.

The Malmö Microbiology Biobank also includes samples of the population-based serological screening for virus infections and rubella immunity during pregnancy scheduled to be taken during week 14 of pregnancy (Malmö Maternity Cohort). The maternity cohort contains all samples from 1985 and from 1989 onwards, altogether more than 100 000 samples from 74 000 mothers.

Malmö Microbiology Biobank was computerised in 1997. NBSBCCC studies with MMB participation have as yet not been published.

Swedish Institute for Infectious Disease Control (SIIDC) Biobank

The Swedish Institute for Infectious Disease Control has performed a series of population-based, nationwide investigations on the immunity against infections in the Swedish population.

A small fraction of the biobank consists of randomly selected persons sampled in 1968 (3 000 subjects), 1977–1978 (1 845), 1990–1991 (4 800) and 1997 (2 400) and analysed to estimate age-specific population immunity rates of, e.g., polio, parotitis, measles, rubella, diphtheria and tetanus.

Most of the about 900 000 biological samples in the SIIDC biobank are diagnostic ones, submitted for microbiological analyses from all over Sweden. The oldest stored samples are from 1957, and complete series exist since 1977.

The information on the samples has been transferred from paper documentation to computerised files for about 629 000 samples. The biobank has recently been linked with the Swedish Cancer Registry, and the quality control of the result of the linkage is on-going. Samples of the Swedish Institute for Infectious Disease Control have been utilised in one NBSBCCC study [5].

Cancer incidence among sample donors

The cancer cases among the serum donors included in the above Nordic serum banks have been traced through automatic record linkages with the national Cancer Registries. Every resident of the Nordic countries has a unique personal identification (PID) code that is used in all main registers and makes computerised linkages accurate and effective [36]. The biobank cohorts should have been compared with the national Population Register data to check that the personal identifiers are the correct ones and persons really exist in the population. Information on vital status and emigration should also have been achieved for every cohort member.

In the person-year calculation needed for calculation of expected numbers of cancer cases, the follow-up started at the date of first serum donation and ended at death, emigration or on the general closing date (depending on the lag of national cancer registration), whichever was first. Because the dates of emigration were not known in the Icelandic biobanks, about 4 000 emigrated persons of the Icelandic biobanks had to be excluded.

The numbers of observed cases and person-years at risk were counted for each calendar year, by gender and five-year age group. Further stratification was made by the time elapsed since the sample donation. The expected numbers of cases for total cancer and for selected specific cancer types were calculated by multiplying the number of person-years in each stratum by the corresponding cancer incidence rate in the national population. The specific cancer types selected *a priori* for the analysis (see Table II) included cancer sites with known risk factors that reveal deviating risk behaviour among the cohort members, and other common cancer types selected to give a representative picture of the cancer situation among the cohorts.

The standardized incidence ratio (SIR) was defined as the ratio of the observed to expected number of cases. The 95% confidence intervals (CI) for the SIR were based on the assumption that the number of observed cases followed a Poisson distribution.

Table II. Numbers of observed (O) and expected (E) cancer cases diagnosed by 31 December 2001 among the participants of the population health examinations in Norway who donated sample to the Janus biobank. Expected numbers based on national population; standardised incidence ratios (SIR = O/E) given with 95% confidence intervals (CI).

ICD-7	Cancer site	O	E	SIR	95% CI
140–207	All malignant neoplasms	21 890	24 087	0.91	0.90–0.92
140	Lip	98	109	0.90	0.73–1.09
143–144	Oral cavity	121	144	0.84	0.70–1.00
145–148	Pharynx	131	169	0.77	0.65–0.92
150	Oesophagus	156	189	0.83	0.70–0.97
151	Stomach	709	702	1.01	0.94–1.09
153	Colon	1868	1914	0.98	0.93–1.02
154	Rectum, rectosigma	1041	1092	0.95	0.90–1.01
155	Primary liver	81	106	0.77	0.61–0.95
155.1	Gall-bladder, biliary tract	96	105	0.91	0.74–1.11
157	Pancreas	524	530	0.99	0.91–1.08
161	Larynx	187	203	0.92	0.79–1.06
162–163	Lung	2178	2516	0.87	0.83–0.90
170	Breast	3364	3666	0.92	0.89–0.95
171	Cervix uteri	440	536	0.82	0.75–0.90
172	Corpus uteri	673	679	0.99	0.92–1.07
177	Prostate	2197	2266	0.97	0.93–1.01
178	Testis	181	192	0.94	0.81–1.09
180	Kidney	671	723	0.93	0.86–1.00
190	Melanoma of the skin	1454	1536	0.95	0.90–1.00
193	Brain and nervous system	804	898	0.90	0.83–0.96
194	Thyroid	291	266	1.09	0.97–1.23

Results

Observed vs. expected numbers of cancers

There were altogether 1.5 million subjects under follow-up in the 16 biobank cohorts for which we were able to calculate person-years at risk. The accumulated number of person-years from the date of first donation until the closing date (1999–2003, depending on the biobank) was 19.7 million (Table I). The mean length of follow-up of a person was 13.2 years. The number of malignant cancer cases diagnosed between sampling and closing date exceeds 75 000.

The above numbers exclude the subjects from the Finnish Mobile Clinic and the Swedish Institute for Infectious Disease Control biobanks and those donors from other biobanks who donated their first sample after the closing date, altogether more than one million donors.

Biobanks based on invitation of the general population. Persons participating in health examinations in Norway (and allowing use of their sera for anonymous cancer research) in the Janus biobank cohort had less cancer than the general Norwegian population (21 889 cases observed vs. 24 086 expected; Table II). Incidence of cancers of the oral cavity, pharynx, oesophagus, primary liver, lung and cervix uteri was significantly decreased (SIRs from 0.77 to 0.87). None of the SIRs was significantly elevated.

The observed numbers of malignant neoplasms among both men (2 955) and women (2 466) in prospective cardiovascular Reykjavik Study exceeded slightly the expected rates based on Icelandic national rates, yielding SIRs 1.07 (95% CI 1.03–1.11) and 1.05 (1.01–1.09), respectively. Men had significantly elevated incidence of cancers of the prostate (SIR 1.15; 1.08–1.23) and kidney (1.22; 1.02–1.45) and significantly low risk of lip cancer (SIR 0.46; 0.23–0.83). Women had significant excess risk of breast cancer (SIR 1.10; 1.02–1.19) and lung cancer (1.13; 1.01–1.25). From the malignancies not included above, basal cell carcinoma of the skin showed elevated incidence (SIR in men 1.35; 1.21–1.51) and in women 1.17 (1.05–1.31).

In the Malmö Diet and Cancer Study cohort there were 1 852 cancer cases, while the expected number based on incidence rates of the entire Swedish population was 1 568 (SIR 1.18; 1.13–1.24). This significant excess was mainly attributable to excesses in prostate cancer (84 excess cases, SIR 1.40; 1.25–1.57), breast cancer (59 excess cases, SIR 1.22; 1.09–1.38), skin melanoma (39 excess cases, SIR 1.72; 1.39–2.11) and bladder cancer (30 excess cases, SIR 1.42; 1.16–1.72). There were no significantly decreased SIRs in the cohort.

The other invitational Southern Swedish cohort, that of the Preventive Medicine in Malmö project, produced more cancers (4 343), but the SIR was similar (1.17; 1.13–1.20). The pattern of cancer sites

with increased incidence was partly similar to that of Malmö Diet and Cancer Study – breast cancer (SIR 1.24; 1.13–1.36), bladder cancer (1.46; 1.30–1.63), and skin melanoma (1.33; 1.16–1.53) – but some other cancers also had increased SIRs: lung cancer 1.48 (1.36–1.61), laryngeal cancer 1.41 (1.02–1.89), pharyngeal cancer 1.56 (1.14–2.08) and pancreatic cancer 1.23 (1.02–1.48). Despite the large numbers of cases, none of the 22 primary sites studied separately showed a SIR significantly below unity.

The Northern Sweden Health and Disease Study consists of three cohorts randomly selected from the population of given ages in that region. The largest number of cancer cases (2 426) was found among members of the Västerbotten Intervention Program (VIP). The expected number was slightly higher (2 531). The SIR was significantly decreased for lung cancer (0.82; 95% CI 0.68–0.98), otherwise there were no major aberrations from 1.0.

There were 289 cancer cases in the smaller MONICA cohort as compared to 310 cases expected. The difference is not significant, and none of the site-specific SIRs was significantly different from unity.

The mammography screening cohort (also part of Northern Sweden Health and Disease Study) showed a SIR of 0.99 (1 159 observed cases vs. 1 174 observed). These women – age range 50–69 years – had a significantly lowered SIR for lung cancer (0.75; 0.56–0.98) while none of the other sites showed a SIR significantly different from unity. Incidence of breast cancer was significantly increased during the first year after mammography and serum sampling date (SIR 1.89; 1.58–2.24) but this excess was compensated by a significantly decreased incidence in the later years (SIR 0.88; 0.78–0.98).

Maternity cohorts. There were 11 078 cancer cases observed after sampling (1983–1996) and before 31 December 2003 in the Finnish Maternity Cohort which is the biggest of the four biobanks based on the population screening of pregnant women. The expected number based on average Finnish female population was 11 628 and the SIR was 0.95 (95% CI 0.94–0.97).

There were 5 042 cases of breast cancer, very close to the expected number (SIR 1.01; 0.98–1.03). The incidence of breast cancer was above the national average after sera drawn in the context of the first or second pregnancy but the SIR declined after the third pregnancy to 0.91 (0.83–0.98). The SIR for endometrial cancer among all pregnant women was 0.57 (0.49–0.65) and decreased after the third pregnancy to only 0.38 (0.20–0.64).

There was an excess of the rare placental choriocarcinoma during the first year after sampling (6 cases; SIR 4.27; 1.57–9.29), which is by definition related to pregnancy. Borderline tumours of the ovary were less frequent than in the population on average (SIR 0.84; 0.74–0.94). The SIR for lung cancer was 0.80 (0.68–0.93), with strongest decrease in adenocarcinoma (SIR 0.62; 0.48–0.79). The SIR for stomach cancer was 0.86 (0.74–0.98).

In the Icelandic Maternity Cohort there were 1015 malignant neoplasms observed vs. 1 019 expected (SIR 1.00; 0.94–1.06). The SIRs for single cancer sites were similar as those reported above for the Finnish Maternity cohort but none of them reached statistical significance in this ten times smaller data set.

Women in the Malmö Maternity Cohort (part of Malmö Microbiology biobank) also had overall cancer incidence similar to the national population (493 observed cases vs. 498 expected, SIR 0.99; 0.91–1.08), but there was a tendency for higher lung cancer incidence than the reference population (SIR 1.28; 0.68–2.18). None of the other cancer sites deviated significantly from the expected incidence.

In the Northern Sweden Maternity Cohort there were 1 625 cancer cases observed after sampling and before end of follow-up. The expected number was 1 717 and the SIR 0.95 (0.90–0.99). Significantly decreased SIRs were seen for lung cancer (0.59; 0.40–0.83) and endometrial cancer (0.69; 0.49–0.94).

Specific cohorts with clearly defined enrolment criteria. Men in the Helsinki Heart Study had 2 998 cancer cases, less than expected (SIR 0.91, 95% CI 0.88–0.94). The SIRs were significantly decreased for cancers of the esophagus (SIR 0.69; 0.46–0.99), stomach (0.76; 0.63–0.90), liver (0.68; 0.47–0.95), pancreas (0.82; 0.66–0.99), nose (0.16; 0.00–0.88), lung (0.68; 0.62–0.74) and unspecified sites (0.70; 0.51–0.93).

Incidence of non-melanoma skin cancer (SIR 1.33; 95% CI 1.09–1.59) and basal cell carcinoma of the skin (1.23; 1.15–1.32) was significantly above the national average. SIR for chronic lymphatic leukaemia was 1.50 (95% CI 1.08–2.01), four of the cases being diagnosed during the first year after sampling (SIR 7.91; 95% CI 2.15–20.2). Also meningeomas of the brain were in excess (SIR 1.59; 95% CI 1.05–2.29).

The smoking men in the Alpha-Tocopherol-Beta-Carotene (ATBC) study had a very different cancer pattern: there was an excess in most sites (Figure 2). The observed number of cancers was 8 262 while the expected number was only 5 412 (SIR 1.53;

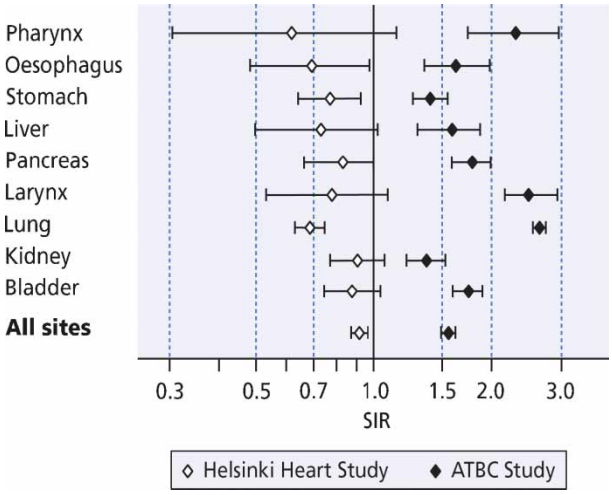

Figure 2. Standardised incidence ratios (SIR) of selected cancers among the 19 000 Finnish men in Helsinki Heart Study, and the 29 000 men in Alpha-Tocopherol-Beta-Carotene (ATBC) Study, with 95% confidence interval bars.

1.49–1.55). The SIRs were significantly increased for cancers of the tongue (SIR 1.64; 95% CI 1.11–2.34), other oral cavity (1.81; 1.24–2.53), pharynx (2.28; 1.71–2.97), oesophagus (1.63; 1.33–1.94), stomach (1.38; 1.24–1.53), liver (1.48; 1.23–1.75), pancreas (1.77; 1.58–1.96), larynx (2.48; 2.11–2.89), lung (2.63; 2.54–2.73), prostate (1.13; 1.08–1.18; for non–localized prostate cancers 1.22; 1.09–1.36), kidney (1.35; 1.20–1.50), bladder (invasive 1.73; 1.59–1.88; papilloma 2.16; 1.26–3.46), and unspecified sites (1.78; 1.54–2.04). There was no cancer with the SIR below 1.0.

The Red Cross blood donors in capital Oslo and surrounding areas (the smaller part of Janus biobank) had lower than average overall cancer incidence (2 286 cases observed vs. 2 399 expected; SIR 0.95; 95% CI 0.91–0.99). The SIRs of cancers of the stomach, primary liver and larynx are as low as 0.36–0.46, all significantly decreased. The SIR for lung cancer was 0.76 (95% CI 0.65–0.88). The SIR for breast cancer was significantly elevated (1.29; 1.17–1.42), and so was the SIR for skin melanoma (1.24, 1.07–1.42).

Viral screening and clinical testing biobanks. In the part of the Malmö Microbiology cohort including samples submitted for testing because of clinical suspicion of infection with blood-borne viruses (e.g. jaundice or impaired liver function, drug addicts, hemophiliacs and dialysis patients), there were 2 055 cancer cases more than the expected number 4 455 (SIR 1.46; 95% CI 1.43–1.50). All SIRs were above 1.0, except those for breast cancer and endometrial cancer. The highest SIRs were seen for primary liver cancer (5.58; 4.87–6.36), pancreatic cancer (3.28, 2.93–3.67) and gall-bladder cancer (2.52; 1.94–3.22).

The Malmö Microbiology subcohort consisting of sera submitted for other virus serology had even higher relative overall cancer risk (SIR 2.08; 95% CI 1.97–2.20; 1 328 cases observed vs. 638 expected). Very high SIRs were seen in primary liver cancer (4.15; 2.63–6.22), pancreatic cancer (2.71, 1.91–3.74), lung cancer (2.95; 2.46–3.53) and cancers of the brain and nervous system (3.05; 2.37–3.88).

Prospective cancers – basis for nested case control studies

The number of cancer cases diagnosed after serum donation among persons in each serum bank is given in Table III for all cancers combined and for 64 subcategories. These numbers are based on the routine linkages between the serum banks and cancer registries (status in early 2005), and include cancers until 1999 for the biobanks in Malmö and until 2001–2003 for the other cohorts (Figure 3).

In addition to the cases listed in Table III, there were 7 754 malignant neoplasms diagnosed between serum donation (1966–1972) and 31 December 2003 among the 50 448 subjects of the Finnish Mobile Clinic Health Examination Survey for whom serum sample is available. The largest site-specific numbers of cancer cases in this biobank were in lung (1 098), breast (914), prostate (893), colorectum (706), urinary organs (578), stomach (492), skin (420) and pancreas (336). The respective number of cancer cases among the FINRISK study members who have donated serum in 1992, 1997 or 2002 was 796. The leading cancer among the women was breast cancer (130 cases) and among the men prostate cancer (146 cases).

The test linkage of the 629 000 computerized records with the Swedish Institute for Infectious Disease Control data base with the cancer registry revealed more than 21 000 prospective cancer cases by the end of 2003. This addition will raise the Swedish number of subsequent cancer cases higher than the number observed in Finland (30 000 cases by the end of 2003) and Norway (27 000), giving a balanced three-country setting in the future studies; the Icelandic number (7 000) is smaller but very high as compared to the small population size in Iceland (less than 300 000).

Discussion

Strengths of biobank-based study designs

We have described an infrastructure that allows multinational and multidisciplinary networking for comprehensive prospective epidemiological studies nested in several biological specimen banks. In the

Table III. Numbers of cancer cases registered among serum donors after donation to the Nordic Biological Specimen Banks for Cancer Causes and Control (NBSBCCC) according to the latest cancer registry linkages (year covered given below the name of each biobank), by cancer site and serum bank.

		Number of cancer cases																
		Finland			Iceland		Norway	Sweden										
							Janus		Northern Sweden Health and Disease Study					Malmö Microbiology Biobank				
ICD-7 (or internal code)	Cancer site/type [* also expected numbers of cases have been calculated]	Maternity cohort (females)	Helsinki Heart Study (males)	ATBC (males)	Maternity cohort (females)	Heart Preventive Clinic	Health examinations	Blood donors	Northern Sweden Maternity Cohort (females)	VIP	MONICA	Mammography (females)	Diet and cancer	Preventive medicine	Maternity cohort (females)	Diagnostic (blood–borneviral infections)	Other	TOTAL NORDIC
		1983–2003	1980–2003	1985–2003	1980–2002	1967–2002	1973–2001	1973–2001	1975–2003	1885–2003	1986–2003	1995–2003	1991–1999	1974–1999	1985–1999	1969–1999	1990–1999	
140–207	All malignant neoplasms	11078	2998	8262	1015	5421	23577	2406	1797	2792	327	1364	1852	4343	493	6511	1328	75564
140	*Lip	4	25	62	–	13	100	7	2	9	–	1	7	18	–	32	10	430
141	Tongue	35	8	30	1	15	116	14	3	5	2	3	3	18	2	25	2	423
142	Salivary glands	43	9	8	5	10	53	4	3	7	–	4	3	7	–	18	6	322
143–144	*Oral cavity	23	14	33	–	13	130	8	2	7	1	4	10	22	1	26	6	300
145–148	*Pharynx	19	11	54	4	26	139	17	4	13	2	4	10	46	1	40	7	397
150	*Oesophagus	9	29	107	1	76	178	20	1	16	–	7	18	53	1	51	10	727
151	*Stomach	187	116	342	5	324	762	38	23	78	9	25	36	125	6	157	26	2410
152	Small intestine	26	11	17	2	23	97	13	7	11	2	4	10	23	–	37	5	440
153	*Colon	369	185	310	27	444	2076	195	56	172	22	110	105	232	16	311	75	4858
154	*Rectum, rectosigma	171	140	241	19	154	1142	118	28	111	15	54	85	172	6	205	28	2843
155	*Primary liver	29	34	124	1	43	89	4	5	16	4	8	15	39	–	223	23	812
155.1	*Gall-bladder, biliary tract	41	21	52	–	41	102	8	4	9	–	5	9	16	–	65	6	379
157	*Pancreas	70	96	326	7	167	573	49	18	75	8	32	42	119	3	316	37	2095
160	Nose	4	1	11	1	7	43	5	4	5	–	2	1	6	–	5	4	259
161	*Larynx	1	34	152	1	33	198	9	1	7	1	1	4	44	–	43	1	691
162–163	*Lung	164	429	2772	44	702	2432	190	37	135	17	65	137	535	13	523	127	8322
(162A)	lung, adenoca	63	78	377	23	235	673	60	17	45	6	27	45	149	2	160	40	2000
(162S)	lung, small cell	29	69	474	7	136	479	28	4	20	7	13	27	93	2	80	20	1488
(162E)	lung, squamous cell	23	144	863	6	154	841	74	4	11	1	7	25	125	2	115	21	2416
170	*Breast	5042	4	10	350	632	3369	413	731	540	46	471	324	466	165	635	112	13480
(170D)	breast, ductal	3968	3	9	270	420	1978	192	434	358	32	324	138	117	67	243	57	8610
(170L)	breast, lobular	736	1	–	34	38	335	52	73	76	2	76	66	50	13	75	12	1639
171	*Cervix uteri, invasive	434	..	..	107	40	469	54	161	31	5	13	13	27	56	101	10	1692
172	*Corpus uteri	194	..	..	17	135	739	59	41	124	11	124	48	75	2	76	18	1835

Table III (*Continued*)

ICD-7 (or internal code)	Cancer site/type [* also expected numbers of cases have been calculated]	Number of cancer cases																
		Finland			Iceland		Norway		Sweden									
							Janus			Northern Sweden Health and Disease Study					Malmö Microbiology Biobank			
		Maternity cohort (females)	Helsinki Heart Study (males)	ATBC (males)	Maternity cohort (females)	Heart Preventive Clinic	Health examinations	Blood donors	Northern Sweden Maternity Cohort (females)	VIP	MONICA	Mammography (females)	Diet and cancer	Preventive medicine	Maternity cohort (females)	Diagnostic (blood–borneviral infections)	Other	TOTAL NORDIC
		1983–2003	1980–2003	1985–2003	1980–2002	1967–2002	1973–2001	1973–2001	1975–2003	1885–2003	1986–2003	1995–2003	1991–1999	1974–1999	1985–1999	1969–1999	1990–1999	
173	Choriocarcinoma	20	..	..	–	–	2	2	5	–	1	–	–	–	–	2	–	205
175.0	Ovary	388	..	..	27	93	784	59	81	89	5	59	35	59	13	103	16	1811
175.1	Tuba	13	..	..	1	2	–	–	2	2	–	3	4	4	–	7	–	38
176.0	Vulva	38	..	..	3	9	56	6	6	2	–	3	5	8	2	6	3	147
176.1	Vagina	7	..	..	3	3	15	1	4	–	–	–	1	–	1	2	–	37
177	*Prostate	..	831	1668	..	878	2480	266	..	483	64	..	292	606	..	694	128	8390
178	*Testis	..	8	7	..	2	187	30	..	15	1	..	2	17	..	70	2	341
(178S)	testis, seminoma	..	7	6	..	1	128	20	..	10	0	..	1	12	..	36	1	222
(178N)	testis, non-seminoma	..	1	1	..	1	34	3	..	5	1	..	1	5	..	34	1	87
179.0	Penis	..	10	8	..	9	23	–	..	4	1	..	1	4	..	8	2	70
180	*Kidney	142	137	301	12	197	742	71	19	54	9	29	43	141	5	174	41	2297
(180.1)	renal pelvis	6	9	31	–	19	12	1	1	4	2	5	7	25	–	21	2	145
181	*Bladder, ureter, urethra	63	153	540	14	302	490	60	19	113	20	37	102	310	5	292	28	2729
190	*Melanoma of the skin	591	92	90	101	69	1512	200	118	96	8	32	93	208	65	256	33	3754
191	Non-melanoma skin cancer	98	107	155	15	162	577	83	12	23	7	16	71	171	8	489	144	2329
192	Eye	30	7	9	3	13	67	11	7	11	–	4	1	5	1	8	4	373
193	*Brain and nervous system	876	102	120	43	144	887	95	117	119	16	55	58	162	40	242	67	3336
(193G)	glioma	338	44	53	14	64	407	49	22	38	8	21	1	8	4	30	10	1111
(193M)	meningeoma	271	28	26	23	55	272	23	38	39	4	22	19	48	8	70	12	958
194	*Thyroid	870	15	22	97	115	310	27	40	14	1	8	10	28	17	60	11	1839
(194F)	follicular	41	5	3	9	17	39	4	2	1	–	1	2	2	–	5	2	133
(194P)	papillary	799	9	12	64	54	131	11	16	8	1	5	4	6	9	23	4	1156
195.0	Glandula suprarenalis	17	3	5	1	3	26	5	11	7	–	6	3	11	2	24	4	128
195.1	Glandula parathyreiodea	–	–	–	–	3	–	–	27	11	2	4	16	53	8	75	13	212
195.2	Thymus	10	1	1	–	1	–	–	2	2	–	–	1	4	2	10	5	39

Table III (*Continued*)

ICD-7 (or internal code)	Cancer site/type [* also expected numbers of cases have been calculated]	Number of cancer cases																
		Finland			Iceland		Norway		Sweden									
							Janus			Northern Sweden Health and Disease Study					Malmö Microbiology Biobank			
		Maternity cohort (females)	Helsinki Heart Study (males)	ATBC (males)	Maternity cohort (females)	Heart Preventive Clinic	Health examinations	Blood donors	Northern Sweden Maternity Cohort (females)	VIP	MONICA	Mammography (females)	Diet and cancer	Preventive medicine	Maternity cohort (females)	Diagnostic (blood–borneviral infections)	Other	TOTAL NORDIC
		1983–2003	1980–2003	1985–2003	1980–2002	1967–2002	1973–2001	1973–2001	1975–2003	1885–2003	1986–2003	1995–2003	1991–1999	1974–1999	1985–1999	1969–1999	1990–1999	
195.3	Hypophysis	–	–	–	8	22	1	–	13	12	–	5	9	17	6	40	8	141
195.4	Corpus pineale	4	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	5
196	Bone	46	4	4	6	13	34	4	3	3	–	1	3	5	1	15	3	341
197	Soft tissue	97	19	26	7	14	96	18	15	18	1	5	14	26	4	46	12	615
199	Other/unknown site	86	45	196	9	101	547	53	25	71	13	40	37	103	3	211	40	1779
200,202	Non-Hodgkin's lymphoma	341	143	204	26	125	776	72	51	107	13	39	69	138	13	271	111	2699
201	Hodgkin's disease	175	10	14	16	12	90	12	21	8	1	5	3	15	9	52	40	684
203	Multiple myeloma	52	40	50	6	75	100	12	7	61	6	20	28	57	1	83	17	818
204–207	Leukaemia	183	77	113	20	113	393	40	34	47	8	22	42	91	11	158	67	1419
(204CLL)	chronic lymphocytic	23	43	46	1	44	48	4	6	20	5	11	14	32	1	41	17	356
(204AML)	acute myeloid	89	14	41	7	38	73	10	12	12	1	5	10	28	5	54	23	422
Not included above:																		
171C	Cervix, CIN3/in situ/dysplasia gravis	4063	..	..	–	–	5		2254	197	21	43	–	–	–	–	–	6583
175B	Ovary, borderline tumour	269	..	..	41	18	111	7	48	22	5	12	–	–	–	–	–	533
181P	Bladder, papilloma	8	9	17	2	19	655	60	15	107	19	35	90	271	4	252	22	1585
191B	Skin, basal cell carcinoma	1606	758	907	171	657	–	–	–	–	–	–	–	–	–	–	–	4099

.. Not applicable.
– Not registered by the national cancer registry.

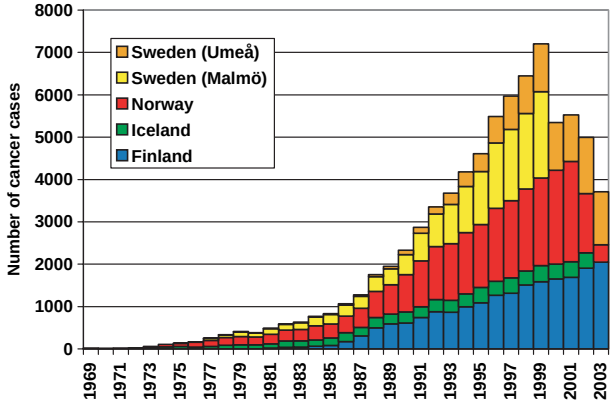

Figure 3. Annual numbers of registered cancer cases among subjects in the Nordic biobanks diagnosed after serum donation, by region. The numbers refer to cancer update status in Spring 2005, when coverage was complete only until 31 December 1999.

following, we compare strengths and weaknesses of studies based on samples readily collected in biobanks to the alternative situation that there is no biobank, i.e., samples from cases and controls have to be collected after the disease of the case has been diagnosed:

(a) Use of biobank data offers proper time order of data collection of exposure and outcome and decreases the possibility of "reverse causality bias", i.e. the mixing up of cause and effect. For instance, herpesviruses are frequently reactivated by severe diseases such as cancer and may indeed induce cellular genes related to cellular proliferation. [4,21]. If the virus is measured from a sample taken at the time of cancer diagnosis, it is difficult to assess whether associations between reactivatable viruses and cancer are causal or mere secondary associations with opportunistic infectious agents. In the biobank-based design we have been able to show that cancer reactivates herpes simplex virus type 2 and not vice versa [21,37,38].

(b) A related type of bias is the differential measurement bias, i.e., situations where the fact that the patient has disease influences measurements. Even existent (pre)cancer may influence both antibody levels and cellular immunity because of the immune dysfunctions seen in cancer [11,25,27]. Also, it may be easier to obtain cancer tissue than control tissue. When measurement biases are related to case status, their effect is particularly unpredictable. Studies using samples taken from individuals long time before the cancer diagnosis suffer only from misclassification bias that is non-differential with regard to case status, which may result in a conservative and readily quantifiable bias.

(c) Many exposures are associated with non-attendance in retrospective case-control studies, biasing results. In biobank-based studies there may be baseline selection in the formulation of the study base (that makes the study base different as compared to the population from which it was originally drawn), but after that all samples from the study base are available for testing, and there is no selection related to later case-control status. Participants in cohort studies do not have to be randomly selected from the general population in order to make valid inference about exposure-disease associations.

(d) Cost-effectiveness and time required to complete the study: The classical prospective cohort study – where samples are not stored but analysed immediately after sampling – requires very long follow-up, often decades, at very high expenses. Study hypotheses and measurement assays may be outdated when the outcomes are finally obtained. The establishment and maintenance of population-based biological specimen banks is costly, but when such banks are established they can be used for a variety of prospective studies on the aetiology of several reasonably common diseases, e.g., association of HPV infections to various human cancers [2,3,6–9,16,19]. The marginal cost for a prospective study can be reduced to the level where also rather unlikely, innovative hypotheses (that may result in breakthroughs) can be reliably evaluated., e.g., the role of *C. trachomatis* in cervical cancer causation [14,17,39]. Since biological specimen banks are already established, the time required for completion of a reliable prospective study with decade-long follow-up of a recently emerged epidemiological problem is short.

(e) Objective measurements: The extent of misclassification of self-reported prior exposures can be considerable, especially for sensitive questions such as addictions. Even a very modest amount of misclassification may lead to very misleading conclusions. There now exists an increasing arsenal of biochemical measurements that can be used for *objective* measurement of exposures in stored biological specimens, e.g. serum cotinine measurements for assessment of smoking habits [40]. The prospectively collected, extensive questionnaires on environmental exposures, diet and life style that most of the NBSBCCC banks

contain may also be compared with the biochemical measurements for mutual validation.

(f) Advantages of serial measurements: For studies of chronic diseases, such as cancer, that develop over a very long timespan, a considerably more reliable and complete assessment of the importance of various exposures can be obtained by studying multiple serial measurements of the same person [10,13]. Furthermore, the unavoidable variability of measurements in a single sample will cause a systematic underestimation of the importance of a risk factor (regression dilution towards the mean) [41], which can be corrected for using serial measurements. Serial samples can also be used to pin-point the time-point of exposure [13]. There are many examples of disease causation by an exposure that occurs only if the exposure occurs at a certain time-point. Poliomyelitis as a result of delayed exposure to poliovirus is well known example. Only if samples taken at many different time-points preceding development of disease have been stored can one attempt to study time-point of exposure by biochemical and molecular assays. The Nordic biological sample banks contain a very high proportion of serial samples; the mean number of samples per person is 2–3 (Table I). For instance, the maternity cohorts the sets of serial samples related to pregnancies become almost complete [13], and some specific research cohorts may include very tight set of samples, e.g. there are up to 28 samples from part of the Helsinki Heart Study subjects.

Stability and validity of old samples

A potential weakness of studies based on historical biobank samples is the stability and validity of the old samples: The oldest samples in the Nordic biobanks are more than 30 years old and many are stored at -25°C. Validations of the Janus biobank have shown that most of the substances commonly analysed in epidemiological studies, for instance proteins (in particular antibodies), organic acids, carbohydrates, trace metals, inorganic salts and polyunsaturated fatty acids are stabile when they are stored at -25°C. However, not all enzymes and vitamins are stabile under these conditions [42].

Genotyping from archival serum and plasma samples is, following the development of efficient whole genome amplification methods, a fairly routine method also from very old samples stored at -25°C [43]. However, investigators contemplating amplification-based methods such as PCR should be aware that in the 1960s and 1970s disposable pipettes and tips may not always have been used in all biobanks. Possible deterioration of the oldest sera is commonly outweighed by consideration of increased statistical power, reduced reverse causality biases with longer follow-up and possibility to detect causative exposures that occur many years before diagnosis of disease and may not be detectable in samples taken at or close to diagnosis.

Follow-up procedures

Initial calculations of SIRs in some of the biobanks did not include *follow-up for vital status*, which produced erroneous, markedly lowered SIRs in older ages. The problem with missing data on vital status would have slowly become a serious problem also in case-control settings: a control subject that was registered as being alive may actually have had died before the respective case is diagnosed with cancer. In the data presented in this paper, all biobanks were linked with national population registers to keep dates of death up-to-date, and the procedure will from now on become a regular routine procedure. Follow-up for emigration has not been considered very important because its magnitude has been rather small. However, in younger cohorts of modern Europeans emigration really has an effect: for instance, almost 4 000 women (6%) of the Icelandic maternity cohort had emigrated after serum sampling and have to be excluded from all studies (because dates of emigrations are not known).

Incorrect PIDs is another source of errors on cancer risk estimates and control selection. The practice to check all PIDs against the population registries was not in routine use by all biobanks before this study, but the procedure will from now on become a regular routine procedure.

The data quality requirements for the standardised incidence ratio calculation were a good way to improve accuracy of identifiers and completeness of follow-up for vital status, which is crucial in case-control studies for picking up controls that really are at risk of getting the cancer. Lack of follow-up for vital status and presence of some incorrect identifiers are likely to have caused minor errors in control selection in previous studies: controls might have died or got cancer which was not known to the researchers. This type of error would have reduced the risk estimates towards unity, i.e., any excess risks published so far are rather under – than over – estimates of the true risk. Computerised record linkage procedures based on the unique PIDs are unambiguous [36]. Therefore, linking failures do not bias cancer risk estimates.

Registers on cancer incidence and other outcomes

The tabulation of observed numbers of cancer cases given in Table III demonstrate that the Nordic cancer registries are able to produce data also by cancer classifications based on variables other than the topography alone (such as subtypes of leukaemia, histology and stage-specific categories) and tabulations of certain precancerous lesions. These specific categories are often useful for focused hypotheses testing. The data collection procedures prepared for this paper will serve as the tool for preparing annual tabulations of actual numbers of cancer cases diagnosed after sample donation. Such tabulations must be made in order to be able to design nested case-control studies, as knowledge of the number of cases is required to estimate statistical power.

The numbers of cancer cases accumulated to Table III were based on the linkages between the biobanks and cancer registries done in early 2005: and cancer registries: 10 000 to 20 000 of newly diagnosed cases are missing due to the normal delay of cancer registration and about 10 000 are missing because some biobanks are not linked with cancer registry very often. In some countries, each linkage for a specific research purpose requires a new ethical permission.

The Nordic health data infrastructure and the unique personal identifiers are utilised in all important registers to allow electronic linkages of numerous register-based health indicators. The nationwide Nordic cancer registries have been in operation since the 1950s and have virtually complete coverage for cancer incidence [44]. Data from many other registries may be used as outcome variables or as co-factors, e.g. proxy variables for confounding or to determine the indication of sampling in clinical biobanks. Hospital discharge registries, perinatal outcome registries, cause of death registries, registries of infectious diseases and various disease-specific registries (diabetes, AIDS, etc) are commonly used.

Cancer incidence rates in cohorts in relation to national cancer incidence rates

None of the biobank cohorts had exactly the incidence pattern of the national general population. Some of them were known to deviate from the general population by enrolment design. For instance, the maternity cohorts included only pregnant women who are known to have lower risk of cancers of breast, corpus uteri and ovary than nulliparous women. Information on parity and age at first pregnancy are available from the databases and can be taken into account when designing studies on diseases related to reproductive parameters.

Studies on samples taken during pregnancy are not necessarily generalizable to non-pregnant women. On the other hand, these samples offer a unique possibility to study the effect of *in utero* exposures to the health of the children [25,28]. The large Nordic Maternity cohorts are the main source of prospective cancer cases diagnosed in ages before the age of 50 (Figure 4).

The most extreme example of an a priori known selection was the ATBC cohort which included only smoking men, who have a more than two-fold excess incidence of numerous cancer types than the average male population. Clinical biobanks also deviated from population averages due to the clinical diagnostics selection process, the impact of which could not have been estimated in advance.

The overall cancer incidence among men increases and among women decreases towards the lower socio-economic position [45,46]. Typical cancers associated with low socio-economic status or educational level are cancers of the lip, oesophagus, stomach, larynx and nose, and multiple myeloma in both sexes, cancers of cervix uteri and vagina in women and lung cancer in men. Cancers of the colon, breast, testis and soft tissue, and skin mela-

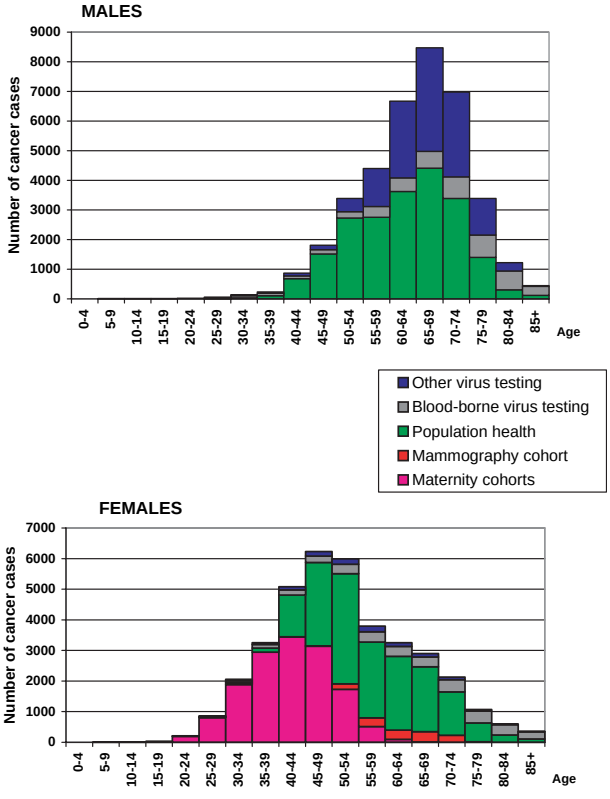

Figure 4. Numbers of registered cancer cases among subjects in the Nordic biobanks diagnosed after serum donation, by sex, age and type of biobank.

noma (especially in the trunk and limbs) are most common in high social strata. A person who knows the variation of cancer incidence over socio-economic or health habit strata can estimate from the cancer pattern whether a cohort is representative of the general population in terms of these factors.

Most biobank cohorts showed slightly lower than average cancer risk. The biobanks that were based on population registry-based invitations presumably contain a representativity bias related to better participation rate among health-conscious subjects. Participation rate was not a strong indicator of this selection; e.g. the cancer pattern for the Malmö Diet and Cancer Study, with participation rate of only 40%, was rather typical for the entire population in Southern Sweden, and similar to the population samples with higher participation rates, suggesting that selection is commonly related to a never-attending non-health-conscious population.

Some serum banks contain clearly discernible subcohorts with obviously different cancer incidences. In nested case-control studies, it is therefore recommended to consistently match for such subcohorts. Malmö Microbiology Biobank is the best example of a biobank technically collected in same place by the same organisation, but that contains clearly discernible subcohorts enrolled for different reasons. As described in the present paper, these subcohorts have clearly different background cancer risks. Matching for subcohort in case-control selection is important to maintain validity in the rate ratio estimation.

The fact that symptoms related to the outcome disease of the study may increase the likelihood for sampling will increase the likelihood to encounter *reverse causality* biases (mix-up of cause and effect). In Malmö Microbiology Biobank, the SIRs for liver, gall bladder and pancreatic cancer were extremely high during the first year after serum sampling. Symptoms from these cancers (such as jaundice) are likely to cause testing for hepatitis viruses. While the risk for gallbladder and pancreatic cancers were not elevated after the first year after sampling, the risk for liver cancer remained elevated, presumably reflecting a true etiologic link (such as infections with hepatitis B and C viruses being causes of liver cancer). When using clinical biobanks for prospective studies, we therefore suggest to not include samples who do not have at least one year of follow-up between sampling and diagnosis of the endpoint disease.

In the cohort collected in association of mammography screenings in Northern Sweden, there was an almost two-fold incidence of breast cancer during the first year after sampling. Mammography screening is indeed expected to find non-symptomatic breast cancer cases that will have a diagnosis date shortly after the screening visit. The cohort formation principle therefore produces an atypical collection of breast cancers in terms of timing of diagnosis and stage distribution that must be considered if these cases are used, e.g., in studies on natural latency times.

While calculation of observed and expected rates is very helpful for characterising cohorts and estimating generalizability, it should be pointed out that the main focus of biobank-based studies is more on studies of new etiologies than on generalizing to total cancer occurrence in national populations. When cases and controls are selected from the same prospectively followed closed cohort (strictly defined using personal identifiers and enrolment date) there is internal validity and possibility to make valid etiologic inferences regardless of the degree of population representativeness of the cohort.

Recommended study design

The nested case-control design and the case-cohort design are commonly used in molecular epidemiological studies within cohorts. They are methods of sampling from an assembled cohort study [47].

In the nested case-control design, for each case controls are randomly sampled from those eligible to be controls. In the classical case-cohort design, a simple random sample of the cohort, a subcohort, is used as a comparison group for all cases in the cohort. In the stratified case-cohort design the subcohort is selected applying stratified random sampling.

The controls for nested case-control studies are appropriately selected applying incidence-density sampling. For each case controls are randomly sampled from all control candidates alive and free of cancer at the time of case's diagnosis. A subject is eligible to be selected as a control for more than one case and a case can serve as a control for cases with earlier date of diagnosis [48]. This is called sampling with replacement. The odds ratio will then be an estimate of the incidence rate ratio in the source population between those exposed and not exposed. This holds true regardless of a disease rarity assumption, provided that the control sampling is independent of the exposure given the factors used in matching [49]. In sampling with replacement there is a small probability of multiple use of the same sample, and therefore more complicated statistical pseudo-likelihood approach is usually not necessary.

In case-cohort studies, the subcohort is selected without regard to disease status. The subcohort provides information about the person-time experi-

ence in the random sample. The case-cohort design allows direct estimation of risk ratio.

Among the advantages of the nested case-control design is that there is no need to follow-up the controls beyond case's diagnosis. Effects of analytic batch, storage time and freeze-thaw cycles can be removed by matching [50]. The major advantages of the case-cohort design are that the subcohort can be used for several diseases and for extended follow-up.

Among the drawbacks of the nested case-control design is that the controls are not a representative sample of the cohort and thus cannot necessarily be used as controls for future cases. Control for batch and storage effects and freeze-thaw cycles are cumbersome in case-cohort design compared to nested case-control design. Batch effect will cause bias, when subsequent case series are studied in case-cohort design [50].

The case-cohort design might be preferable if the biomarkers would not suffer from storage length, batch effects, and freeze-thaw cycles. The nested case-control design provides tools for dealing with such issues in principle, and is therefore more appropriate design for the NBSBCCC studies.

In conclusion, the optimal design is the nested case-control design applying incidence-density sampling with replacement.

Matching in nested case-control design

Matching is restriction on selection of control series. The goal of matching is to balance the ratio of cases to controls within matched sets, and to make controls' distributions of the potentially confounding matching variables more like those of cases'.

The network of Nordic biobanks has attempted to use uniform control selection algorithms in all biobanks participating in a given joint study. For each cancer case of interest, typically 1–4 control donors of same sex are randomly selected among persons who were alive at the time of case's diagnosis, have donated a sample around the same time as the case and were born within two years of the case's date of birth. As pointed out above, in the case of heterogeneous biobanks, matching for subcohort (e.g. Malmö Maternity Cohort and Bloodborne virus screening within Malmö Microbiology Bank) is essential. Rather exact matching for sampling date has been considered important, because different length of storage time in the bank can have profound influences on some biological markers. For some markers, seasonal variation is so large that it is also therefore important to select the control samples from same time of the year as the sample of the case. In NBSBCCC studies typically only a difference of 1–2 months in sampling date is accepted. As freezing and thawing can affect a number of biomarkers it is also highly recommended to match on the number of freeze-thawing cycles a sample has been subjected to. The biobanks have not necessarily recorded the numbers of freeze-thaw cycles. The effect freeze-thaw cycles should be in any case prevented by sufficient aliquoting or other suitable methods, for example the straws in the EPIC study [51]. Samples of the matched set are typically pipetted in random order on same panel to minimize the effects of analytic batch.

In general, matching is a means of reducing bias due to confounders. However, matching on variables intermediate in the causal pathway between exposure and disease will bias estimates [52]. This is also true for matching on variables affected by exposure and disease. Therefore, matching on other variables than those mentioned above is generally not allowed in NBSBCCC studies. Matching may increase the random error, e.g., matching on a nonconfounder associated with exposure but not disease reduces efficiency.

There are certain practices in control selection that are bound to specific features of the unique sample materials. First, because most biobank data bases do not include variables indicating how many times a sample has been used as a control and how much serum is left, it is often necessary to pick up one or two extra control candidates that will be used if the actual controls are missing or do not contain enough materials. Persons who have been diagnosed with other cancers have in some studies in some biobanks not been accepted as controls (to save these valuable samples), although formally they would be eligible at least until the date of cancer diagnosis of the respective case. This causes only a negligible theoretical error, because the pool of eligible controls for each case normally includes hundreds of subjects. In conclusion, matching for a limited number of variables, typically sex, age, storage time and subcohort, is preferable.

Quality assurance

To assure protection of integrity and ensuring equal analyses of cases and controls, the samples must be blinded before they are sent to the analysing laboratory. After the laboratory analyses are ready, the researchers receive the code key that tells them which samples are cancer cases and which are controls. The code key does *not* include personal information that could connect the samples to the donors. When research data are never linked to personal identities, individual donor will not receive any information about his or her sample and the

risks that biobank-based research would violate the integrity of individuals is minimal.

Since 2001 a system of Quality Assurance (QA) for Good Biobanking Practice is used on a routine basis at the Medical Biobank in Umeå. Quality control and auditing by an external expert or organisation is performed at regular intervals. QA is a process that aims at measuring, evaluating and continuously re-evaluating the quality and when required, improving the quality. The QA work should have a plan of activity and schedule for work, and all employees of a biobank should be involved in the QA system. The QA system is supposed to guarantee that the biological samples, questionnaires, and data have the quality that corresponds to the intended use. Data base systems that document historical storage conditions, aliquotting history, number of thawings/freezings and amounts available are highly recommended.

The QA system should include procedures for how the completeness and accuracy of the attached database (non-material part of the biobank) should be maintained, kept up-to-date and how the pitfalls of selection and follow-up biases should be traced. Many biobanks have no instruments to make basic person-year at risk calculation from their cohorts or other means to control the coverage and population representativeness of their data. We suggest that calculation of cancer incidences and SIRs should be included as a basic QA practice of essential importance in biobanking QA, that should be asked for in reviews of biobank-based studies.

Many clinical biobanks do not give high priority to such check-ups of registered data that are absolutely necessary for epidemiological follow-up studies. The system used in this paper, where the data management of clinical biobanks was entrusted to cancer registries or experienced epidemiological biobanks, is likely to be essential for valid use of clinical biobanks for epidemiological studies.

Conclusions

The high internal validity of internal comparisons within a defined biobank cohort make prospective biobank-based study designs preferable for etiological studies. Limited population-representativity implies only that generalization of results to entire national populations should be made with caution. As the described biobanks are committed to work towards joint Quality Assurance standards, including defined accessibility to external requests for samples and as the biobanks together contain a huge numbers of prospectively occurring cases of cancer, the Nordic biobank cohorts provide a solid basis for prospective studies on cancer causes and control.

Acknowledgements

Jan Ivar Martinsen at the Cancer Registry of Norway, Björn Tavelin in Umeå, and Håkan Krzeszowski and Henrik Månsson in Malmö made a great effort in creating the O/E calculation procedures for biobank cohorts in Norway and Sweden. Gudrídur Olafsdóttír took care of quality control of the data related to the Icelandic serum cohorts. Anna Törner kindly offered the numbers of cancer cases from the test linkage from the Swedish Institute of Infectious Disease Control. Kari Pasanen from the University of Kuopio prepared the informative map (Figure 1). Several network researchers – in addition to those listed as authors – gave valuable comments in topics related to this study discussed in numerous joint network meetings. This study was supported by the Nordic Council of Ministers longitudinal epidemiology programme, by the European Union fifth framework Concerted Action on Evaluation of the Role of Infections in Cancer and by the sixth framework Network of Excellence on Cancer Control using Population-based Registries and Biobanks.

References

[1] Doll R, Peto R. The causes of cancer: Quantitative estimates of avoidable risks of cancer in the United States today. J Natl Cancer Inst 1981;66:1191–308.

[2] Dillner J, Knekt P, Schiller JT, Hakulinen T. Prospective seroepidemiological evidence that Human Papillomavirus type 16 infection is a risk factor for esophageal squamous cell carcinoma. Br Med J 1995;311:1346.

[3] Lehtinen M, Dillner J, Knekt P, Luostarinen T, Aromaa A, Kirnbauer R, et al. Serologically diagnosed infection with human papillomavirus type 16 and risk for subsequent development of cervical carcinoma: Nested case-control study. Br Med J 1996;312:537–9.

[4] Lehtinen T, Luostarinen T, Dillner J, Aromaa A, Hakama M, Hakulinen T, et al. Serum p53 accumulation and altered antibody responses to Epstein-Barr virus proteins precede diagnosis of haemopoietic malignancies of lymphoid origin. Br J Haematol 1996;93:104–10.

[5] Dillner J, Kallings I, Brihmer C, Sikstrom B, Koskela P, Lehtinen M, et al. Seropositivities to human papillomavirus types 16, 18, or 33 capsids and to Chlamydia trachomatis are markers of sexual behavior. J Infect Dis 1996;173:1394–8.

[6] Björge T, Dillner J, Anttila T, Abeler V, Engeland A, Hakulinen T, et al. Prospective seroepidemiological study of the role of human papillomavirus in non-cervical anogenital cancers. Br Med J 1997;315:646–9.

[7] Dillner J, Lehtinen M, Björge T, Luostarinen T, Youngman L, Jellum E, et al. Prospective seroepidemiologic study of human papillomavirus infection as a risk factor for invasive cervical cancer. J Natl Cancer Inst 1997;89:1293–9.

[8] Björge T, Hakulinen T, Engeland A, Jellum E, Koskela P, Lehtinen M, et al. A prospective, seroepidemiological study of the role of Human Papillomavirus in esophageal cancer in Norway. Cancer Res 1997;57:3989–92.

[9] Dillner J, Knekt P, Boman J, Lehtinen M, af Geijersstam V, Sapp M, et al. Seroepidemiological association between

Human Papillomavirus infection and risk of prostate cancer. Int J Cancer 1998;75:564–7.

[10] af Geijersstam V, Kibur M, Wang Z, Koskela P, Pukkala E, Schiller J, et al. Stability over time of serum antibody levels to Human Papillomavirus type 16. J Infect Dis 1998;177:1710–4.

[11] Luostarinen T, af Geijersstam V, Bjorge T, Eklund C, Hakama M, Hakulinen T, et al. No excess risk of cervical carcinoma among women seropositive for both HPV16 and HPV6/11. Int J Cancer 1999;80:818–22.

[12] Lehtinen M, Luostarinen T, Youngman LD, Anttila T, Dillner J, Hakulinen T, et al. Low levels of serum vitamins A and E in blood and subsequent risk for cervical cancer: interaction with HPV seropositivity. Nutr Cancer 1999;34:229–34.

[13] Kibur M, af Geijerstam V, Pukkala E, Koskela P, Luostarinen T, Paavonen J, et al. Attack rates of Human Papillomavirus type 16 and cervical neoplasia in primiparous women and field trial designs for HPV16 vaccination. Sex Trans Infect 2000;76:13–7.

[14] Koskela P, Anttila T, Bjorge T, Brunsvig A, Dillner J, Hakama M, et al. Chlamydia trachomatis infection as a risk factor for invasive cervical cancer. Int J Cancer 2000;85:35–9.

[15] Sigstad E, Lie AK, Luostarinen T, Dillner J, Jellum E, Lehtinen M, et al. A prospective study of the relationship between prediagnostic Human Papillomavirus seropositivity and HPV DNA in subsequent cervical carcinomas. Br J Cancer 2002;87:175–80.

[16] Mork J, Lie AK, Glattre E, Hallmans G, Jellum E, Koskela P, et al. Human Papillomavirus infection as a risk factor for squamous-cell carcinoma of the head and neck. New Eng J Med 2001;344:1125–31.

[17] Anttila T, Saikku P, Koskela P, Bloigu A, Dillner J, Ikaheimo I, et al. Serotypes of Chlamydia trachomatis and risk for development of cervical squamous cell carcinoma. JAMA 2001;285:47–51.

[18] Stattin P, Adlercreutz H, Tenkanen L, Jellum E, Lumme S, Hallmans G, et al. Circulating enterolactone and prostate cancer risk: A Nordic nested case-control study. Int J Cancer 2002;99:124–9.

[19] Björge T, Engeland A, Luostarinen T, Mork J, Gislefoss RE, Jellum E, et al. Human Papillomavirus infection as a risk factor for anal and perianal skin cancer in a prospective study. Br J Cancer 2002;87:61–4.

[20] Stattin P, Lumme S, Tenkanen L, Alfthan H, Jellum E, Hallmans G, et al. High levels of circulating testosterone are not associated with increased prostate cancer risk: A pooled prospective study. Int J Cancer 2004;108:418–24.

[21] Lehtinen M, Koskela P, Jellum E, Bloigu A, Anttila T, Hallmans G, et al. Herpes simplex virus and risk of cervical cancer: A longitudinal nested case-control study in the Nordic countries. Am J Epidemiol 2002;156:687–92.

[22] Lehtinen M, Pawlita M, Zumbach K, Lie K, Hakama M, Jellum E, et al. Evaluation of antibody response to Human Papillomavirus early proteins in women whom cervical cancer developed 1 to 20 years later. Am J Obstet Gynecol 2003;188:49–55.

[23] Youngman LD, Jellum E, Lehtinen M, Dillner J, Björge T, Luostarinen T, et al. A prospective seroepidemiological study of smoking as a risk factor for invasive cervical cancer. Manuscript.

[24] Paavonen J, Karunakaran KP, Noguchi Y, Anttila T, Bloigu A, Dillner J, et al. Serum antibody response to the heat shock protein 60 of Chlamydia trachomatis in women with developing cervical cancer. Am J Obstet Gynecol 2003;189:1287–92.

[25] Lehtinen M, Koskela P, Ögmundsdottir H, Bloigu A, Dillner J, Gudnadottir M, et al. Maternal herpesvirus infections and risk of acute lymphoblastic leukaemia in the offspring. Am J Epidemiol 2003;158:207–13.

[26] Tuohimaa P, Tenkanen L, Ahonen M, Lumme S, Jellum E, Hallmans G, et al. Both high and low levels of blood vitamin D are associated with a higher prostate cancer risk: a longitudinal, nested case-control study in the Nordic countries. Int J Cancer 2004;108:104–8.

[27] Luostarinen T, Lehtinen M, Björge T, Abeler V, Hakama M, Hallmans G, et al. Joint effects of different human papillomaviruses and Chlamydia trachomatis infections on risk of squamous cell carcinoma of the cervix uteri. Eur J Cancer 2004;40:1058–65.

[28] Lehtinen M, Ögmundsdottir HM, Bloigu A, Hakulinen T, Hemminki E, Gudnadottir M, et al. Associations between three types of maternal bacterial infection and risk of leukemia in the offspring. Am J Epidemiol 2005;162:662–7.

[29] Anttila T, Tenkanen L, Lumme S, Leinonen M, Gislefoss R, Hallmans G, et al. Chlamydial antibodies and risk of prostate cancer. Cancer Epidemiol Biomarkers Prev 2005;14:385–9.

[30] Stolt A, Kjellin M, Sasnauskas K, Luostarinen T, Koskela P, Lehtinen M, et al. Maternal Human Polyomavirus infection and risk of neuroblastoma in the child. Int J Cancer 2005;113:393–6.

[31] Hakama M, Luostarinen T, Hallmans G, Jellum E, Koskela P, Lehtinen M, et al. Joint effect on HPV16 with Chlamydia trachomatis and smoking on risk of cervical cancer: antagonism or misclassification (Nordic countries). Cancer Causes Control 2000;11:783–90.

[32] Tedeschi R, Bidoli E, Ågren Å, Wadell G, Paoli PD, Dillner J. Epidemiology of Kaposi's sarcoma herpesvirus (HHV8) in Västerbotten country, Sweden. In press.

[33] Tedeschi R, Luostarinen T, Paoli PD, Gislefoss RE, Tenkanen L, Virtamo J, et al. Joint Nordic prospective study on Human Herpesvirus 8 and multiple myeloma risk. Br J Cancer 2005;93:834–7.

[34] Frick MH, Elo O, Haapa K, Heinonen OP, Heinsalmi P, Helo P, et al. Helsinki Heart Study: Primary-prevention trial with gemfibrozil in middle-aged men with dyslipidemia. Safety of treatment, changes in risk factors, and incidence of coronary heart disease. New Eng J Med 1987;317:1237–45.

[35] ATBC The Alpha-Tocopherol B-CCPSG. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. New Eng J Med 1994;330:1029–35.

[36] Pukkala E. Use of record linkage in small-area studies. Elliot P, Guzick J, English D, Stern R, editors. Geographical and environmental epidemiology. Oxford: Oxford University Press; 1992. p 125–31.

[37] Lehtinen M, Leminen A, Kuoppala T, Tiikkainen M, Lehtinen T, Lehtovirta P, et al. Pre- and post-treatment serum antibody responses to HPV16 E2 and HSV-2 ICP8 proteins in women with cervical carcinoma. J Med Virol 1992;37:180–6.

[38] Lehtinen M, Hakama M, Knekt P, Heinonen PK, Lehtinen T, Paavonen J, et al. Serum antibodies to the HSV-2 specified major DNA-binding protein are elevated before the diagnosis of cervical cancer. J Med Virol 1989;27:131–6.

[39] Wallin KL, Wiklund F, Luostarinen T, Hallmans G, Anttila T, Koskela P, et al. Chlamydia trachomatis infection: A risk factor in cervical cancer development-a population based prospective study. Int J Cancer 2002;101:371–4.

[40] Parish S, Collins R, Peto R, Youngman L, Barton J, Jayne K, et al. Cigarette smoking, tar yields, and non-fatal myocardial

infarction: 14 000 cases and 32 000 controls in the United Kingdom. International Studies of Infarct Survival (ISIS) Collaborators. Br Med J 1995;311:471–7.

[41] Clarce R, Shipley M, Lewington S, Youngman L, Collins R, Marmot M, et al. Underestimation of risk associations due to regressions dilution in long-term follow-up of prospective studies. Am J Epidemiol 1999;150:341–53.

[42] Jellum E, Andersen A, Lund-Larsen P, Theodorsen L, Orjasaeter H. Experiences of the Janus Serum Bank in Norway. Environ Health Perspect 1995;103(Suppl 3):85–8.

[43] Sjöholm M, Hoffmann G, Lindgren S, Dillner J, Carlson J. Comparison of archival plasma and formalin-fixed paraffin-embedded tissue for genotyping in hepatocellular carcinoma. Cancer Epidemiol Biomarkers Prev 2005;14:393–6.

[44] Teppo L, Pukkala E, Lehtonen M. Data quality and quality control of a population-based cancer registry. Acta Oncol 1994;33:365–9.

[45] Andersen A, Barlow L, Engeland A, Kjaerheim K, Lynge E, Pukkala E. Work-related cancer in the Nordic countries. Scand J Work Environ Health 1999;25(Suppl 2).

[46] Pukkala E. Cancer risk by social class and occupation. A survey of 109 000 cancer cases among Finns of working age. Contributions to Epidemiology and Biostatistics. Basel: Karger; 1995. p 7.

[47] Langholz B. Entries: Case-Cohort Study and Case-Control Sudy, Nested. In: Armitage P, Colton T editors. Encyclopedia of Biostatistics. Chichester: John Wiley & Sons; 1999. p 497–503 & 514–9.

[48] Greenland R, Thomas DC. On the need for the rare disease assumption in case-control studies. Am J Epidemiol 1982;116:547–53.

[49] Rothman KJ, Greenland S. Modern Epidemiology 2nd ed. Philadelphia: Lippincott-Raven; 1998. p 95–6.

[50] Rundle AG, Vineis P, Ahsan H. Design options for molecular epidemiology research within cohort studies. Cancer Epidemiol Biomarkers Prev 2005;14:1899–907.

[51] Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): Study populations and data collection. Public Health Nutr 2002;5:1113–24.

[52] Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. III. Design options. Am J Epidemiol 1992;135:1042–50.