
MINI-REVIEW

Genomic Epidemiology Cohorts in Korea: Present and the Future

Keun-Young Yoo^{1*}, Hai-Rim Shin², Soung-Hoon Chang³, Bo Youl Choi⁴, Yun-Chul Hong¹, Dong-Hyun Kim⁵, Daehee Kang¹, Nam Han Cho⁶, Chol Shin⁷, Young-Woo Jin⁸ for the Korean Genome Epidemiology Society

Abstract

Human genome epidemiology involves the application of genetic technology to assess the impact of variations at the DNA level on health and disease. Recent developments in molecular biology allow epidemiologists to use biomarkers to determine an individual's predisposition to disease and to detect disease at an early stage. Moreover, advances in genomics and proteomics could play a central role in research into disease prediction and prevention. Large scale population-based cohort prospective studies offer the most comprehensive approach to the delineation of gene function, the effects of the environment, and their interactions. The Korean Multi-center Cancer Cohort (KMCC), under construction since 1993, is the first multi-center prospective cohort to identify risk factors for cancer in Korea. Data on general lifestyle, physical activity, diet, reproductive factors, and agricultural exposure are obtained through direct interview using a structured questionnaire. Anthropometric measurements and clinical laboratory findings are also collected using a web-based data entry system. Moreover, biological materials have been banked [blood (serum, plasma, buffy coat, packed erythrocytes) is stored at -70°C and urine at -20°C] for future analysis. Several other cohorts including the Korean National Cancer Center (KNCC) Cohort, the Korean Health Examinees (KOEX) Cohort, the Korean Health and Genome Epidemiologic Study (KHGES), and the Yang Pyeong Cohort have also been launched since the KMCC cohort was initiated. Even though these cohorts have collected similar data and biospecimen, questionnaires and protocols used have not been standardized. However, these cohort studies are of increased scope and have been designed to detect risk factors for cardiovascular disease, metabolic syndrome, and cancer. Subjects have been followed up actively by health personnel in different regions and by using record linkages with the central cancer registry, and the national death certificate and national health insurance claim databases. As of August 2004, the total number of subjects enrolled in all cohorts with archived biologic specimens was around 80,000. A new genomic cohort has been launched since 2001 in Korea, for which the target number of subjects is 250,000 men and women over the next 5 years. This article describes the goals and the designs of each of the above-mentioned cohorts.

Asian Pacific J Cancer Prev, 6, 238-243

Introduction

Cancer has been the leading cause of death in Korea since 1983, and accounts for about 26% of all deaths. The six leading cancer sites are the stomach, liver, lung, colo-rectum, female breast and the uterine cervix (Korean Central Cancer Registry, 2003; National Statistical Office, 2004; Shin et al., 2004) (fig. 1). With the objective of establishing a new large-scale genomic cohort, the Korean Genomic Epidemiology Society recently launched the Korean Health and Genome Study, which will provide evidence on cancer etiologies and on cancer prevention strategies in Korea.

Mortality statistics on the above six leading cancers over the past 20 years (1984-2003) were obtained from the Korean National Statistical Office (National Statistical Office, 2004) (fig. 2). Age-standardized mortality rates of lung, colo-rectal, female breast, and uterine cervical cancers have been increasing since 1984; 2.7-fold in men and 2.5-fold in women for lung cancer, 4.0-fold in men and 3.6 in women for colo-rectal cancer, 2.0-fold in female breast cancer, and 4.1-fold for uterine cervical cancer. On the other hand, stomach cancer decreased by 45% in men and by 53% in women over the same period. Liver cancer mortality showed no significant change over this period.

¹Seoul National University College of Medicine, ²National Cancer Center Research Institute for National Cancer Control & Evaluation, ³Konkuk University College of Medicine, ⁴Hanyang University College of Medicine, ⁵Hallym University College of Medicine, ⁶Ajou University College of Medicine, ⁷Korea University College of Medicine, ⁸Radiation Health Research Institute
*Address for Correspondence: Department of Preventive Medicine, Seoul National University College of Medicine, 28 Yongon-dong, Chongno-gu, Seoul 110-799, Korea. Tel: 82-2-740-8324 Fax: 82-2-3673-3540 Email: kyyoo@plaza.snu.ac.kr

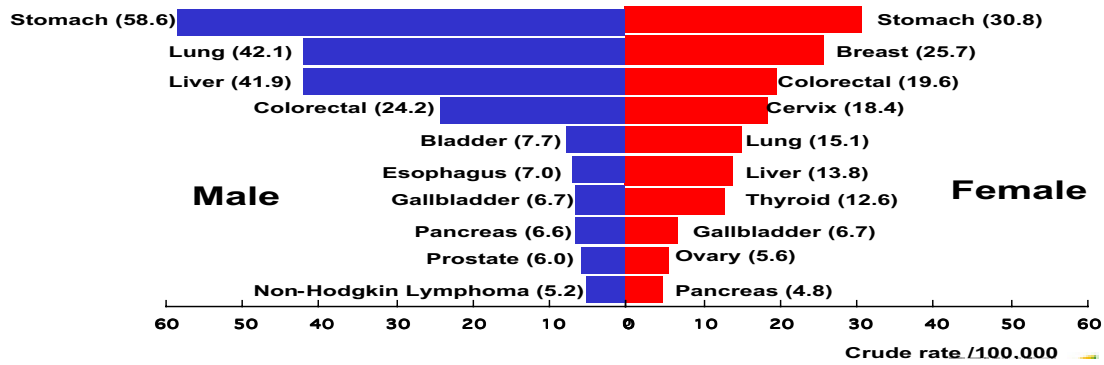


Figure 1. The Top Ten Leading Cancer Sites in Korea (adopted from the Korean Central Cancer Registry, 2003).

The two major determinants of the risk of disease, including cancer, are the environment and genetic variation (Potter, 2004). A number of previous studies have indicated the importance of the environment in terms of disease risk by providing evidence of geographical variations in disease frequencies (e.g., international, urban, rural, etc.), changes in disease incidences among genetically homogenous stable populations, or by observations of disease incidences among migrants. In addition, genetic susceptibilities to environmental exposures vary widely. Recently, there has been growing concern about low penetrant genes associated with common complex diseases, because interactions with other genes and environmental factors could affect disease risk.

Human genome epidemiology involves the epidemiological application of genome technology to the assessment of the impact of genetic variations on health and disease. Recent developments in molecular biology allow epidemiologists to use biomarkers to determine a predisposition to or to detect disease at an early stage, and advances in genomics and proteomics are likely to play central roles in research on disease prediction and prevention in the future. At the moment, the best approach to determining the roles of genes, the environment, and the effects of interactions between the two are large scale population-based prospective cohort studies (Potter, 2004; Barbour, 2003). This paper will explain the on-going

genomic cohorts in Korea (Table 1). After discussing current statuses and limitations, we address issues concerning future genomic cohort studies.

Korean Multi-Center Cancer Cohort (KMCC)

The Korean Multi-center Cancer Cohort (KMCC) is the oldest genomic cohort instituted in Korea (Yoo et al., 2002). This is a multi-center prospective cohort that was designed to meet the requirements of genome epidemiological studies on cancer etiology, which had been conducted since 1993. The principle investigator is the KMCC study group, and the objectives of this study are to establish a large prospective community-based cohort that allows studies on risk factors, biomarkers of exposures and effects, genetic susceptibility and gene-environmental interactions in certain cancers. All of the participants provided informed consent, and the study protocol was approved by the Institutional Review Board of Seoul National University Hospital. This cohort has been characterized as follows; to establish a cancer-free cohort by screening against cancer at the time of recruitment, to provide a complete set of baseline information by direct interview, by establishing a biologic materials bank and a surveillance system to determine cancer outcome. Data on general lifestyle, physical activity, diet, reproductive factors, and agricultural exposures are obtained by direct interview using a structured questionnaire. Anthropometric

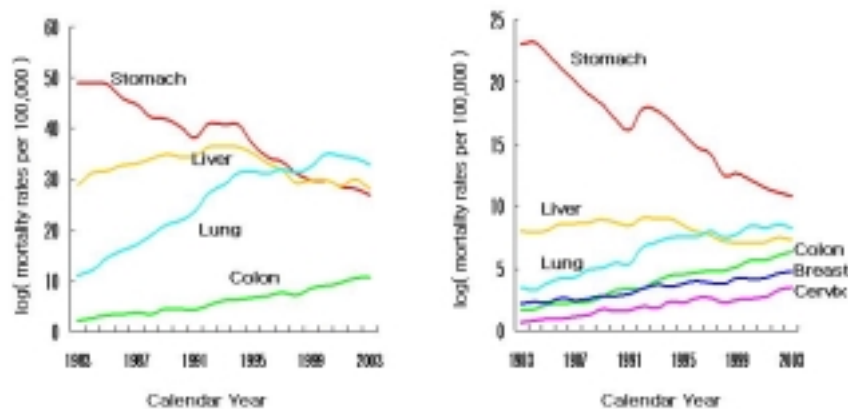


Figure 2. Trends in Age-standardized Cancer Mortality Rates in Korea, 1983-2003 (adapted from National Statistical Office, 2004). Left males and Right Females.

Table 1. Genomic cohorts constructed in Korea (as of December, 2004)

Cohorts	Principle investigator	Target diseases	Subjects	No. of subjects	Year began	Questionnaire on lifestyles and diet	Biologic materials bank
KMCC ¹⁾	Keun-Young Yoo	Cancers	Urban/rural, both genders	19,205	1993	Direct interview	serum/plasma/WBC buffy/RBC clots/urine
KNCCC ²⁾	Hai-Rim Shin	Cancers both genders	Rural/urban, both genders	13,971	2001	Direct interview	serum/plasma/WBC buffy/RBC clots/urine
KOEX ³⁾ (IHEC)	Yun-Chul Hong Dong-Hyun Kim	Cancer and metabolic syndrome	Urban, both genders	19,662	2001	Direct interview	serum/plasma/WBC buffy/RBC clots/urine
Yang Pyeong	Bo Youl Choi	CVD and cancer	Rural, both genders	6,794	1990	Direct interview	serum/plasma/DNA/urine
KHGES ⁴⁾	Nam H. Cho Chul Shin	Diabetes and Hypertension	Urban/rural, both genders	10,038	2001	Direct interview	Whole blood/buffy
KHNP ⁵⁾	Young-Woo Jin	Cancers	Nuclear workers	7,335	2001	Self-administration	serum/plasma/WBC buffy/RBC clots/urine

1) KMCC: Korean Multi-center Cancer Cohort 2) KNCCC: Korean National Cancer Center Cohort

3) KOEX: Korean Health Examinees Cohort (IHEC: formerly Incheon Health Examinees Cohort)

4) KHGES: Korean Health and Genome Epidemiology Study 5) KHNP: Korean Hydro-Nuclear Power

measurements and some clinical laboratory findings have also been collected and stored in a web-based database system. Moreover, a biological materials bank containing blood (serum, plasma, buffy coat, packed erythrocytes) stored at -70°C and urine at -20°C has been established for genomic epidemiological studies on cancer etiology. A DNA yield study using buffy coat stored long term revealed PCR products for β -globin in nearly all samples (98%). Follow-up for cancer occurrence has been implemented based on an active surveillance system by regional health researcher, and a passive surveillance system by record linkages with the central cancer registry, and the national death certificate and national health insurance claim databases. As of December 2004, the total number of observation for the cohort including both questionnaires and biologic specimen was 19,205. As of December 2002, a total of 400 incident cancer cases (256 cases in men, 144 cases in women) have been identified by the passive surveillance system and the

total number of follow-up was 67,676 person-years (PY) (Table 2). The overall incidence rate of all cancers was 591.1/100,000 PY; 973.0 for men and 348.1 for women. Five leading sites of cancer incidence were stomach (102 cases, 150.7/100,000 PY), lung (72 cases, 106.4/100,000 PY), liver (51 cases, 75.6/100,000 PY), colo-rectum (30 cases, 44.3/100,000 PY), and biliary tract (15 cases, 22.2/100,000 PY).

Korean National Cancer Center Cohort (KNCCC)

The Korean National Cancer Center Cohort (KNCCC) is a prospective cohort for genome epidemiological studies on cancer etiology, and has been operational since 2001. The KNCCC is composed of two different cohorts, i.e., community-based and hospital-based cohorts. The community-based cohort targets residents on rural areas, and the hospital-based cohort recruits examinees mainly from

Table 2. Number of Cancer Patients Among the KMCC Cohort Members during the Follow-up Period of 1983~2002

Rank	Male		Female		Total	
	Site	No.	Site	No.	Site	No.
1	Stomach	68	Stomach	34	Stomach	102
2	Lung	56	Colorectum	17	Lung	72
3	Liver	39	Lung	16	Liver	51
4	Colorectum	13	Liver	12	Colorectum	30
5	Esophagus	12	Breast	11	Biliary tract	15
6	Biliary tract	11	Thyroid	8	Esophagus	12
7	Bladder	11	Uterine cervix	8	Pancreas	12
8	Pancreas	7	Pancreas	5	Bladder	12
9	Oral cavity, Pharynx	7	Biliary tract	4	Breast	11
10	Leukemia	6	Lymphoma	3	Thyroid	9
	Others	39	Others	32	Others	87
Total		256		144		400

(as of 2002. 12. 31)

urban areas who participate in cancer screening at the National Cancer Center. The cohort study was approved by the Institutional Review Board of the National Cancer Center, and only those that provided informed consent were included. In the community-based cohort, data on lifestyle, physical activity, diet, reproductive factors, and agricultural exposures are obtained by direct interview using a structured questionnaire. Anthropometric measurements and some clinical laboratory findings have also been collected and stored in a web-based database. In the hospital-based cohort, data on lifestyle, physical activity, diet, reproductive factors, and a past history of medical illness are obtained through a self-administered questionnaire. Medical records of cancer screening have also been collected and stored in a computer-based database. A biological materials bank for blood (serum, plasma, buffy coat, packed erythrocytes) stored at -70°C and urine at -20°C has been established for future genome epidemiological studies on cancer etiology. A DNA yield study using buffy coat specimens stored long term revealed PCR products for p53 in most samples (80-90%). Follow-up for cancer occurrence has been implemented based on an active surveillance system by regional health personnel and a passive surveillance system through record linkages with the National Cancer Incidence Database (constructed in 2005 using the DB from the Korea Central Cancer Registry and all the Regional Cancer Registries). As of August 2004, the total number of observations for the community-based cohort including biologic specimen were 5,971 and that for the hospital-based cohort was 8,000. The cohort study is funded by the National Cancer Center of Korea and has been collaborating with other research teams. Currently, the total number of study subjects is insufficient to comment on cancer incidence, but there are plans to expand the cohort size.

Korea Health Examinees Cohort (KOEX)

The Korean Health Examinee Cohort (KOEX) was launched in 2001. It is a large multi-centre study designed to examine the relationships between cardiovascular disease, metabolic syndrome, or cancer and lifestyle factors, and the modification of these by genetic profiles. The primary aim of the KOEX study is to examine the relationships between lifestyle factors such as smoking, alcohol consumption, exercise, or dietary intake and cancer incidence. Its secondary aim is to study the relationships between lifestyle factors and other diseases, such as, cardiovascular diseases, metabolic syndrome, and others. The effect of genetic variations on relationships between lifestyle factors and disease is also being studied. The recruitment goal is to establish a cohort of more than 100,000 men and women from the general population. KOEX subjects are men and women aged between 20 and 79 at enrollment, who live in 8 major Korean cities. The principal investigator is the KOEX study group, which is composed of 12 investigators from health centers who provide health examinations to residents and to employees in their respective catchment

areas. Demographic information, life style, family history, and previous medical history information is obtained using a self-administered questionnaire. Anthropometry and clinical laboratory data (i.e., blood pressure, CBC, fasting blood sugar, lipid profile, liver function test, renal function test, C-reactive protein, urinalysis) have also been obtained. With regard to biologic specimen storage for future analysis for genomics and proteomics, whole blood and urine samples were stored during 2001-2003 and serum, plasma, buffy coat, and packed erythrocytes were added in 2004. Urine has been stored at -20°C and the other biological materials at -70°C . Because health examinees are able to receive free check-ups annually if workers, and biannually if community residents, this cohort offers the possibility of regular follow-up. Passive surveillance by cancer registry, health insurance claim, and mortality data has also been detailed. At the end of 2004, the KOEX cohort totaled 19,662 participants and this is expected to increase to more than 100,000 by 2008.

Korean Health and Genome Epidemiologic Study (KHGES)

The Korean Health and Genome Epidemiologic Study (KHGES) is a prospective cohort epidemiologic study devised to investigate genetic-environment-diabetes/hypertension relationships. Initially, Ansung (representing a rural community) and Ansan (an urban community) were selected for a Korean Health and Genome Epidemiology Study (KHGES) in 2001, and data on lifestyle, physical activity, diet, reproductive factors, and habitual factors were collected by face-to-face interviews using a standardized questionnaire. However, this has become an ongoing prospective study with a biennial examination. As of 2005, 10,038 subjects aged from 40 to 69 years had been recruited; 5,020 from Ansung and 5,018 from Ansan. All subjects participated in this study voluntarily and informed consent was obtained in all cases. The study protocol was approved by the National Institute of Health Ethics and the Institutional Review Board of the Korean Health and Genomic Study. All participants provided written informed consent. After 8-14 hours overnight fasting, plasma concentrations of glucose, insulin, total cholesterol, triglyceride, and HDL and blood pressure were measured. All participants underwent a 2-hour 75gm Oral Glucose Tolerance Test (OGTT) and WHO and NDDG criteria are used to interpret OGTT results. Biological materials, i.e., serum, plasma, buffy coat, and whole blood are stored at -70°C and urine at -20°C for future genome epidemiological studies.

Yang Pyeong Cardiovascular Disease Cohort

The Yang Pyeong Cardiovascular Disease Cohort is a longitudinal community-based cohort which was designed to study the incidence and determinants of cardiovascular disease and cancer in a small rural area, i.e., in Yang Pyeong County, Kyunggi-do province. The study area, Yang Pyeong, is located 45km east of Seoul and consists of 12 districts.

Population movement in the area is low, and 50% of the population is farmers. The study has been conducted on healthy people since 1990, and includes information on environmental (demographic, social, and life style factors, and physical activity and diet), psychological (stress, depression, social support), physiological (anthropometrics, blood, urine) and genetic factors related to the risk of cardiovascular disease (especially, ischemic heart disease and stroke) and cancer. Intermediate target diseases related to ischemic heart disease and stroke, such as, atherosclerosis, hypertension, and metabolic syndrome were also evaluated. Data on environmental and psychological factors were obtained by direct interview using a structured questionnaire. About 6,794 subjects had been included in this cohort by January 2005.

Korea Hydro and Nuclear Power Cohort (KHNP)

The KHNP (Korea Hydro and Nuclear Power Co.) Cohort, which also has a biological materials bank, was constructed in 2001 to assess the effects of low-dose radiation on the human body, particularly in association with cancer. Subjects in the main are the employees of KHNP and its contractors. These people were recruited during periodic annual health examinations (part of a comprehensive health management program) established by KHNP. The cohort consists of 7,011 men and 324 women of mean age 40.8(\pm 7.8) and 34.4(\pm 7.4), respectively. Subjects complete a self-report questionnaire on history of disease, smoking, drinking, diet, physical activity, and occupation. After drawing 10ml of blood, samples are divided into 3 samples of buffy coat, 5 of serum, and 4 of erythrocytes, and then allocated to long-term or mid-term storage. Samples are stored in different deep freezers with 1 sample of urine in each. During 2001-2003, 6,893 subjects had submitted a urine sample, and 4,550 had twice submitted blood samples. All results are managed using a web-based database system. Informed consent was obtained from all subjects. DNA yield based on β -globin showed that DNA was of high quality without microorganism contamination. The KHNP Cohort is self-funded, and continuous recruitment is possible. Follow up for cancer occurrence is conducted through periodic health examinations. One of the strengths of this cohort is that clinical and questionnaire data are updated annually, and follow-up losses are low.

Perspectives of Genomic Cohorts in Korea

The Korean Genomic Epidemiology Society has recently been founded to utilize the available resources, to make a uniform data collection procedure, to establish a complete validation of instruments, and to facilitate each preexisting cohort for enhancement in networking activities. A fundamental question about genomic cohort is how large it should be in order to estimate the main effect of genetic markers, and to detect gene-environmental and gene-gene

interactions. A new genomic cohort has been launched for this purpose under the supervision of the society during the end of 2004. The target number of subjects for this new cohort is 250,000 men and women for the next 5 years. The members will be recruited from urban and rural population using two distinct methods, which are the community based cohort, and health examinee cohort. Unlike the past cohort study, metropolitan citizens (ie., Seoul) will be in the new cohort. The new cohort will be funded by the Health Promotion Fund from the Ministry of Health and Welfare. Along with other cohorts in Asian countries, both the existing Korean cohorts and the new Korean Genomic Cohort could provide more convincing evidence on new etiologies of cancer and on cancer prevention strategy in the near future.

Conclusions

We have discussed the current statuses of genomic cohorts in Korea. Although networking of existing cohorts seems to be a cost-effective way of utilizing available resources, a lack of uniformity of data collection procedures and the incomplete validation of the instruments used hinder such networking activities. In addition, most genomic cohorts in Korea are limited by sample sizes. Therefore, a large cohort (>250,000) with biospecimen (e.g., blood, urine, fresh tumor tissue, etc.) constructed using sound measures of exposure characterization (e.g., valid and accurate assessment of diet, physical activity, etc.) is needed to better understand human disease pathogenesis and interplay between gene and environmental contributions to disease. Future developments by genomics and proteomics research studies will increase the value of stored biospecimen in terms of clarifying gene-environmental interactions in disease processes. Such a genomic cohort would also allow studies on early disease markers, disease outcome heterogeneities, and on treatment responses and prognoses (Potter, 2000). Ensuring public trust with continuous support from scientific societies with careful attention to ethical issues for construction of a large genomic cohort is also a critical issue in the long journey to disease prevention and health promotion (Tutton et al., 2004).

Acknowledgements

The authors thank Yeonju Kim, Jisuk Bae and Chanook Ahn for their help in preparing the manuscript. This study was supported by a grant of the National R&D Program for Cancer Control, Ministry of Health & Welfare, Republic of Korea (0220180-3).

References

- <http://ncc.re.kr/>
- Barbour V (2003). UK Biobank: A project in search of a protocol? *Lancet*, **361**, 1734-8.
- Ministry of Health and Welfare Republic of Korea (2003). Korea Central Cancer Registry, 2002 Annual Report of the Korea

- Central Cancer Registry (Based on Registered Data from 139 Hospitals). Seoul, Korea.
- Lichtenstein P, Holm NV, Verkasalo PK, et al (2000). Environmental and heritable factors in the causation of cancer-analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*, **343**, 78-85.
- National Statistical Office (2004). Ministry of Health and Welfare Republic of Korea. Yearbook of Health and Welfare Statistics. Gwacheon-city, Korea.
- Potter JD (2004). Toward the last cohort. *Cancer Epidemiol Biomarkers Prev*, **13**, 895-7.
- Shin HR, Jung KW, Won YJ, Park JG, 139 KCCR-affiliated Hospitals (2004). 2002 Annual Report of the Korea Central Cancer Registry: Based on registered data from 139 hospitals. *Cancer Res Treat*, **36**, 103-14.
- Tutton R, Kaye J, Hoeyer K (2004). Governing UK Biobank: The importance of ensuring public trust. *Trends Biotechnol*, **22**, 284-5.
- Yoo KY, Shin HR, Chang SH, et al (2002). Korean multi-center cancer cohort study including a biological materials bank (KMCC-I). *Asian Pac J Cancer Prev*, **3**, 85-92.