Estimating the Genetic Epidemiology Parameters of Selected Cancers in Korea Population - The Korean Twin Study -

Joohon Sung*

Department of Preventive Medicine, Kangwon National University College of Medicine, Kangwon 200-701, Korea

Abstract

The Korean Twin Register (n=154,783 pairs) was reported in 2002 as the first nationwide twin study in Korea and the largest study in Asia. The Twin Register has the information of disease outcomes since 1990, and basic clinical and questionnaire data from biennial health examination provided by Korea National Health Service. The author attempted to calculate some of the genetic parameters of cancers in this population. Common cancers in Korea known to have familial aggregation (colon and breast) and cancers of which familial aggregation is unclear (stomach cancer) were examined for their familial recurrence risks. There were 699 stomach cancers, 438 breast and 491 colorectal cancers cases in the twin register between 1991 and 2003. Like-sex twins showed recurrence risks (λ_{LS}) of 5.1 (95% CI 3.7 - 6.9) for stomach cancers, 15.5 (95% CI 10.9-20.2) for female breast cancers, and 28.1 (95% CI 23.5-34.4) for colon cancers. Colorectal cancers of female like-sex twins show significantly higher familial recurrence risk 40.7 (95% CI 34.6-47.4), suggesting higher genetic contribution in women than in men. The results show increased familial risks compared with previous studies from the same register and are largely compatible with other studies. The data of the Twin Register could be used for estimating population level genetic parameters, as well as base of the various studies.

Keywords: genetic epidemiology, colorectal cancer, breast cancer, stomach cancer, twin study, heritability, familial recurrence risk

Introduction

There is growing body of evidence that a range of environmental and genetic risk factors contribute to the etiology of cancers. In addition, epigenetic mechanisms, such as imprinting and DNA methylation are also known to be involved (Shields and Harris, 1991; Feinberg, 2004). Efforts to elucidate the genetic cause of cancers have continued, to successfully find genetic variations responsible for certain familial cancers: BRCA1, BRCA2, genes for familial adenomatous polyposis (FAP), attenuated familial adenomatous polyposis (AFAP), and hereditary non-polyposis colorectal cancer (HNPCC), (Varesco, 2004; Jo and Chung, 2005). And molecular genetic studies to characterize discovered genetic variations are underway. These rare genetic variations are known to have high penetrance, so that almost 100% of those who carry the genetic variation would develop the cancers. However, in general population, carriers of those genetic variations are very rare, and even among the breast or colorectal cancer patients, those genes comprise less than 3-10% (Varesco, 2004; Jo and Chung, 2005). Genetic variations with high penetrance, but low frequency can only explain small proportion of total cancers. This, however, does not necessarily imply that environments are overwhelming cause of cancers, as was previously advocated (Fearon, 1997). So far, the roles of genetic variations with high frequency and low penetrance, recessively transmitted genes, and oncogenes expressed as a consequence of sequential genetic actions are largely unknown (Feinberg, 2004; Lichtenstein et al., 2000).

It is one of the first tasks of genomic epidemiology, to assess a parameter which can quantitate the overall genetic contribution to carcinogenesis. Knowledge of quantitative genetic contribution permits strategy to screen diseases/traits which have better opportunity to find underlying genetic variations. For example, if the overall genetic parameter suggests major gene or oligogenic involvement, there should be higher chance of detecting it, than the case where polygenic involvement with tiny influences are interacting with environments.

Overall genetic contribution is estimated by heritability or familial recurrence risk. Unlike cardiovascular diseases or traits related to metabolic syndrome, heritability or familial recurrence risk was difficult to estimate for cancers, and often could not be measured in classical family based studies (Aarino, et al., 1999). Studies using twins, however, have been have been successful in estimating overall genetic contribution to cancers (Sung et al., 2002).

Several cancers are reported to have familial

*Corresponding author: Email sungjohn@kangwon.ac.kr,
Tel +82-33-250-8870, Fax +82-33-250-8875

aggregations; breast colorectal ovarian and prostate cancers. In Korea, the age-standardized incidence rate of breast colorectal ovarian and prostate cancers were 21.7 for women, 27.3 for men and 16.7 for women, 4.94 for women, and 7.88 for men, out of 100,000 person year, respectively. Compared with the incidences of western countries, those of ovarian cancer and prostate cancer are much lower in Korea, as most Asian countries. Breast and colorectal cancers also show lower incidences, but those are one of the most rapidly increasing cancers in Korea. On the other hand, stomach cancers, the most common malignancy in Korea (65.7for men, 25.7 for women per 100,000 person years) does not show strong familial aggregation. In the study, the author attempted to compare the parameters of genetic contribution among the common cancers in Korean population; breast, colorectal, and stomach cancers. So far, there have been only a few reports about the overall genetic contribution to breast and colorectal cancers and most of them are reports in western populations (Fearon, 1997; Aarino, et al., 1999; Lichtenstein et al., 2000).

The author and other collaborators have been organizing the Korean Twin Registry which has 154,783 pairs of twins, and 63,666 pairs of adult twins among them. And the registry is linked to the medical utilization, death certificate, and cancer registry data since 1992, resulting in more than 12 years of cancer occurrence follow-up (Sung et al., 2002). The registry, with information on twins, and a range of disease outcomes allows to estimate population level genetic parameters. For relatively common diseases, it was possible to demonstrate preliminary results (Sung et al., 2002). In the study, with longer follow-up period, the author attempted to estimate more reliable genetic parameter from large twin registry which can represent general population.

Methods

Subjects

Total of 63,666 adult twin pairs more than 30 years of age, were analyzed. The details of the Korean Twin Registry were described in previous report (Sung et al., 2002). In the register, disease outcomes are followed since 1990s. Cancer occurrence was identified from three data sources: national cancer registry, medical utilization of National Health Insurance, and cause of death report data.

To estimate population level prevalence, which is an independent parameter for calculating familial recurrence risk, a normative cohort representing Korean population was used. The normative cohort was reconstructed from

Korean National Health Service database, by stratified and random sampling one fortieth of all Koreans. Strata for sampling included 1) age group by 5 years 2) sex 3) province level geographical area 4) type of health insurance. The cohort was over-sampled for elderly persons to stabilize estimated epidemiologic parameters such as prevalence and incidence rate. The normative cohort was reconstructed to estimate a global burden of disease and death, and various descriptive epidemiologic parameters had been calculated from the cohort. The cohort consists of 1,205,470 persons. All medical utilization of the cohort members between the year of 1995 and 2002 were linked. After the data linkage, any information that can identify individuals was deleted.

Case definition

All cancers occurring after 1991 were included in this analysis. The following codes in the International Statistical Classification of Diseases, 9th and 10th Revision (ICD-9 and ICD-10) were used to identify cancers: stomach cancer (151 as ICD-9 and C16 as ICD-10), breast cancer (174 as ICD-9, C50 as ICD-10), and colorectal cancer (C18-C21 as ICD-10, 153-154 as ICD-9). Operational definitions for cancer cases were made to reduce spurious diagnosis. The disease code in the medical utilization data are often made just to rule out the diagnosis rather than to treat the confirmed diseases. While cases in cancer registry data were all confirmed cases often with pathologic diagnosis, the cases in medical utilization data, as they are, can inflate the epidemiologic parameters. The operational definition of cancer cases used in the study was 1) any case in the cancer registry data, 2) any case only found in both medical utilization and death certificate data 3) cases only found in the medical utilization data, which were reported more than twice as the same primary diagnosis, and admitted more than once with the same primary diagnosis. 4) any cases only found in medical utilization data, which were found more than three times with the same primary diagnosis, and at least one of the diagnosis was reported from general hospitals. The accuracy of diagnosis for all three cancers was considered to the same, and the operational definitions were applied to all three cancers. The effective follow-up period used in the study differed by the data sources; cancer registry data was available between the year of 1992 and 2001, medical utilization data 1991-2003, and death certificate data 1991-2002.

Familial recurrence risk

Familial recurrence risk. or relative risk is defined disease

concordance rate between specific family pair, over the prevalence in general population. i.e., the excess risk of disease occurrence, conditional to the presence of diseased (same disease) family pair.

$$familial\ recurrence\ risk = \frac{probability[\,family1\ |\ family2\,]}{prevalence\,of\,the\ disease}$$

(formula 1) (Risch, 1990a; 1990b)

There are two methods to calculate the concordance rate in the family pair; pairwise concordance rate and casewise concordance rate. When the cases were ascertained from a particular phenotype, pairwise concordance rate is the choice, while casewise concordance rate is preferred when there is no problem of ascertainment. For instance, if two by two table is made according to the disease status (D+ - diseased, D- not diseased) of each twin (twin 1 and twin 2),

twin 1

D+   D-

twin 2   D+ | a | b |

D- | c | d |

let a,b,c, and d to be the numbers of each cell,

the concordance rates are defined as formula 2 and 3.

$$pairwise\ concordance\ rate = \frac{a}{a+b+c}$$

(formula 2)

$$casewise\ concordance\ rate = \frac{a}{a+b} = \frac{a}{a+c}$$

(formula 3) (note that b=c here)

In the study, it was reasonable to assume there is no ascertainment bias, and familial recurrence risks were calculated by the casewise concordance rate (formula 3) over prevalence of general population (formula 1). Because the familial recurrence risk is a function of specific family relationship, rather than generalized measurement, it is denoted λ with subscript indicating the particular relationship; λs for a recurrence risk of the sibling, $\lambda 1$ for a recurrence risk of pooled primary relationship, etc. λs of a disease, the recurrence risk of the sibling for the disease, is most widely used. Comparing the λ_{twin} and λs can provide clue about the overall genetic profile. In this study, a familial recurrence risk for like-sex twin pairs (λ_{LS}) and also familial recurrence

risk for opposite-sex twin pairs (λ_{OS}), except breast cancer, were calculated, to give a broad picture about the underlying genetics of selected common cancers in Korean population.

Population Prevalence

In conventional epidemiology, prevalence is measured as a unit of person and year. In contrast, the genetic concept of prevalence implies lifetime risk of the diseases. In the study, prevalence was measured for all the cases during the overall follow-up period (1995-2003), without taking the time span into account. However, since the follow-up period for estimating population level prevalence (=denominator, 1995-203) and concordance rate within twin pairs (=nominator, maximum 1991-2003) does not agree, the prevalence was adjusted by a factor which was calculated from the ratio of two follow-up periods. Follow-up period was re-examined during the study to define effective follow-up period, where the completeness of data sources are satisfactory, by graphical comparisons.

Results

The prevalences of selected cancers were 162.9 for men and 98.6 for women (stomach), 121.0 (breast, women), and 76.8 for men and 60.5 for women (colorectal) out of 100,000 person-year, after adjusting for the correction factor of 1.32 which was calculated form two effective follow-up periods of 8.2 and 10.8 years. Actually estimated prevalence and adjusted prevalence was demonstrated in Table 1. By applying the operational definition of the study (see the method), the total number of estimated cancer cases fell by 1/3 compared with the total cases in the data sources. Fig.1 shows the difference between total cases and cases that fell in the operational definition when the first occurrence of each individual was plotted along the follow-up period.

Defined cancer cases were calculated pairwisely, with discrimination of like-sex and opposite-sex twin

Table 1. Estimated population prevalence and mortality of stomach cancer, breast cancer and colorectal cancer.

	prevalence (100,000)		adjusted prevalence (100,000)*		mortality (100,000/year)	
	men	women	men	women	men	women
Stomach cancer	123.4	74.7	162.9	98.6	23.5	18.5
Colorectal cancer	58.2	45.8	76.8	60.5	9.1	8.2
Breast cancer	0.67	91.7	0.87	121.0	-	3.3

*a factor of 1.32 were multiplied to adjust the difference in follow-up (average 10.8 years follow-up for twins and 8.2 year follow-up for normotive cohort)
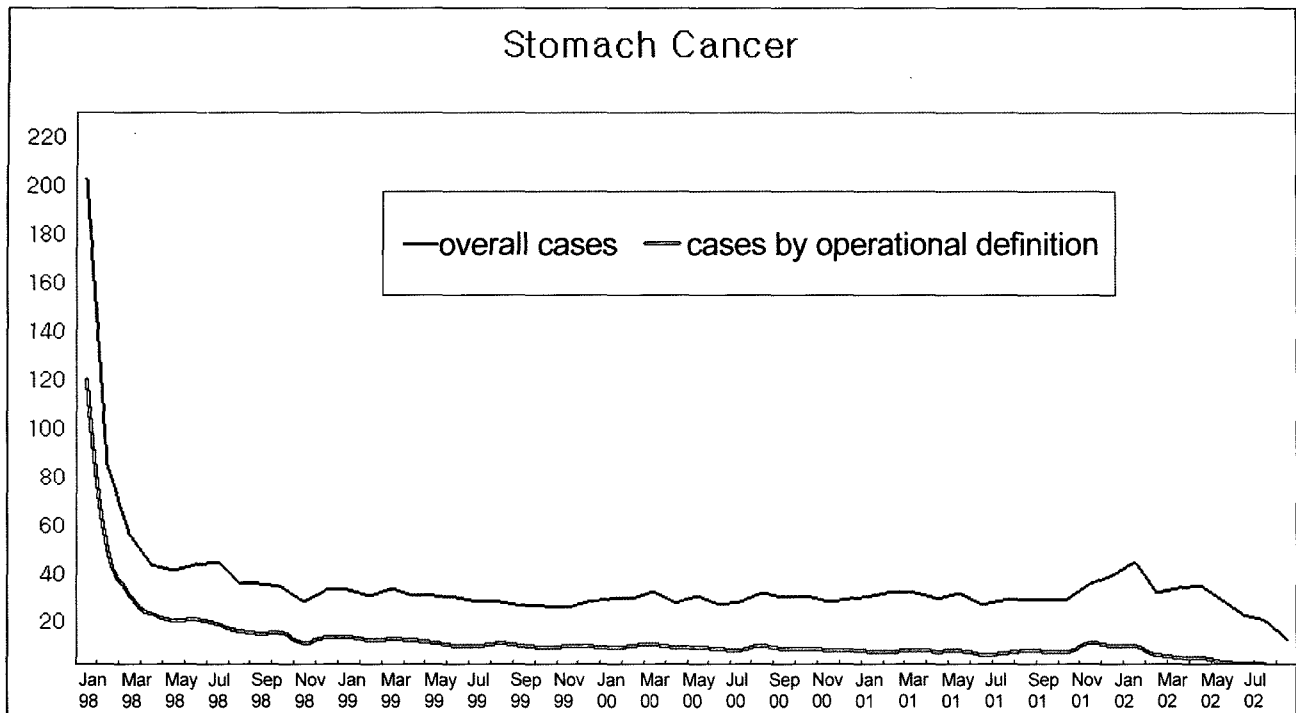
Stomach Cancer

—overall cases   —cases by operational definition

Fig.1. Comparisons between the overall case occurrence and number of cases under operational case definition of stomach cancer.

Table 2. Stomach, breast and colorectal cancer cases among the Korean Twin Register.

Results in Registry	Stomach cancer	Breast cancer	Colorectal cancer
like-sex twins (All)			
Both affected	4	7	8
One affected	598	367	407
None affected	54,119	54,347	54,306
Male like-sex pairs			
Both affected	3	0	3
One affected	279	1	209
None affected	31,211	31,493	31,281
Female like-sex pairs			
Both affected	1	7	5
One affected	319	366	198
None affected	22,907	22,855	23,023
opposite-sex twins			
Both affected	0	0	1
One affected	93	57	68
(Male affected)	56	0	25
(Female affected)	37	57	43
None affected	8,844	8,880	8,868

Number of the cases were organized according to like-sex and opposite-sex pair structures.

pairs. Each like-sex and opposite-sex pairs were further divided into male-male pairs and female-female pairs for like-sex twins, and then once again subdivided according to their disease occurrence; both affected / one affected / none affected (like-sex), both affected / male affected / female affected / none affected (opposite-sex). There were 4, 7, and 8 concordantly affected like-sex twins for stomach, breast and colorectal cancers, while none concordantly affected pairs of opposite-sex twins (Table 2). There were total of 699 stomach cancer, 438 breast cancer, and 491 colorectal cancer cases, which fell in the operational case definition of the study.

Familial recurrence risk for opposite-sex twins could not be calculated owing to lack of concordantly affected pairs. The familial recurrence risk for like-sex twins (λ_{LS}) were estimated to be 6.5 (men, 95% CI 4.0-9.9) and 5.1 (women, 95% CI 3.7 - 6.9) for stomach cancer, 15.5 (95% CI 10.9-20.2) for female breast cancer, 18.4 (men, 95% CI 12.6-25.1) and 40.7 (women, 95% CI 34.6-47.4) for colorectal cancer (Table 3). This means, if one twin was affected above cancers the co-twin has increased risks by the factor of the familial recurrence risk compared with that of the general population. Stomach cancer did not demonstrate significant sex difference in the familial recurrence risk. However, for colorectal cancer, the familial recurrence risk in women was significantly greater than that of men.

Table 3. Estimated familial recurrence risk of like-sex (λLS) and opposite-sex (λOS) twins stomach, breast and colorectal cancers

	stomach cancer			breast cancer			colorectal cancer		
	casewise concordance	prevalence (per 10^5)	familial recurrence risk (95% CI)†	casewise concordance	prevalence (per 10^5)	familial recurrence risk (95% CI)†	casewise concordance	prevalence (per 10^5)	familial recurrence risk (95% CI)†
Like-sex all	0.007	130.8*	5.1 (3.7 - 6.9)	0.019		-	0.019	68.7*	28.1 (23.5-34.4)
like-sex male	0.011	162.9	6.5 (4.0-9.9)	0.000	0.9	-	0.014	76.8	18.4 (12.6-25.1)
like-sex female	0.003	98.6	3.2 (1.3-5.7)	0.019	121.0	15.5 (10.9-20.2)	0.025	60.5	40.7 (34.6-47.4)
opposite sex	0.000	-	-	0.000	-	-	0.000	-	-

*estimated prevalence of general population
†95% confidence intervals (for ratio scale)

Because λ is a parameter which is determined by both the concordance rate and the population prevalence, if the disease has high prevalence, the value is underestimated. A logarithmic scale of λ values were plotted against the logarithm of prevalence, which can directly show the position in terms of heritability (Fig. 2). Breast cancer of women showed less genetic effects than colorectal cancer in terms of λ, but breast cancer and colorectal cancers demonstrated same level of genetic contribution when they were plotted logarithmic scale, implying similar level of heritability. However, the absolute level of heritability in Fig. 2 can not be applied, since the familial recurrence risks were measured from like-sex twins not from siblings.

Discussion

Familial recurrence risk

Familial recurrence risk it self was not new, but first introduced to the linkage analysis to discover genetic variation in 1990 by Risch (Cavalli-Sforza, 1978; Risch, 1990a). The parameter is particularly useful for common complex diseases when the mode of transmission is not understood, or when the penetrance is very low. Familial recurrence risk (λ) differs from classical heritability in several points; Familial recurrence risk (λ) is dependent on the specific familial relationship while heritability is not, λ can be calculated given specific family relationship and population data while heritability need at least more than two different family relationships to be measured, λ represents two genetic parameters - frequency of genetic variation (as prevalence in general population) and the effect size of the genetic variation while heritability is determined by the relative influence of genetic and environmental contributions. λ can be exaggerated when the prevalence of the disease is low, underestimated when it is high. The logarithmic plotting in Fig. 2 can linearly indicate relative genetic position between the λ and heritability. However, the reference heritability levels

are for λs (familial recurrence risk of siblings) and not for λTWIN, each heritability value of cancers only has relative meanings to each others (Cavalli-Sforza, 1978). Against this background, we cannot exactly estimate heritabilities of selected cancers but relative order of them, which show similar level of heritability between colorectal cancer and breast cancer, higher heritability for colorectal cancer of women, and lower heritability for stomach cancer.

Familial recurrence risk is a parameter which is needed for non-parametric linkage analysis. In non-parametric linkage analysis or penetrance-model free linkage analysis, identity by descent (IBD) value is calculated to examine the specific effect of markers, and λ value is used for estimating overall genetic effects. Although classical parametric linkage analysis has been very powerful to detect genetic variations of Mendelian diseases/ traits, IBD-based linkage methods better suit for many common complex diseases. Given that linkage studies are still powerful and useful study for common complex human disease, and IBD-based linkage methods have higher flexibilities and more power in many cases (Klein et al., 2005; Edwards et al., 2005), genomic epidemiologic parameters like λ in Korean population can be used as a future reference of various genetic studies.

In the study λ of like-sex twins are determined by λ of both monozygotic twins and dizygotic twins, however considering that monozygotes were majority the results should be close to λ of monozygotes. λ of dizygotic twins can be a maximal expected value of λs. If λ of monozygotic twins are compared with λs, it can give clues about the underlying genetic mechanism and number of genes involved.

Comparison with previous studies

Most of the genetic parameters of cancers have been drawn in Caucasian populations (Fearon, 1997; Jo and Chung, 2005). Lichtenstein et al. (2000) pooled existing twin registries in European countries and measured heritability for a range of cancers. Compared with the
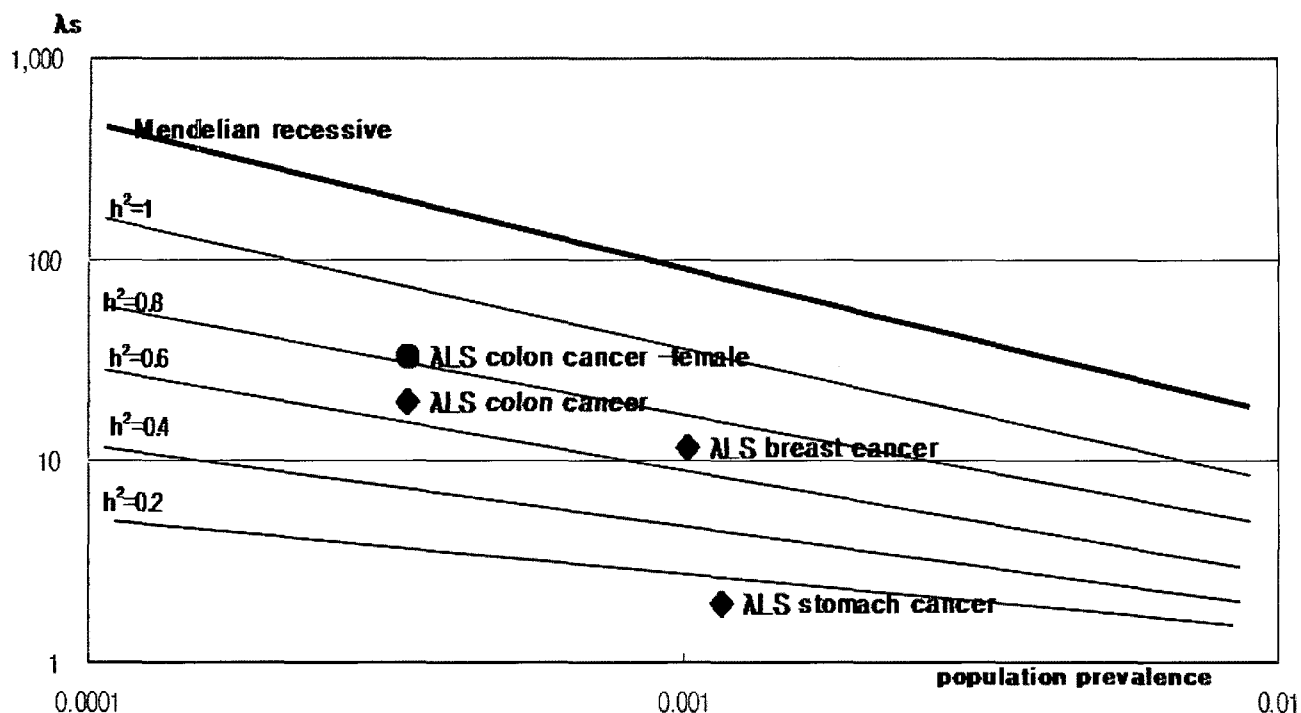
Fig. 2. Relative genetic effects for the familial recurrence risk (like-sex twin pair) of three cancers (stomach, breast, colon) after adjusting the prevalence effects (logarithmic plot). Note that, the results are from like-sex twins not from the sib-pairs.

study of Lichtenstein *et al.,* the prevalences of cancers are as much as ten times lower in this study. In addition, the concordance rate of each cancers are also lower so that stomach cancer showed a sixth to eighth, breast cancer a fifth to seventh, and colorectal cancer a forth to fifth (monozygote and dizygote). The difference of the parameters can be explained by the fact that 1) difference in the incidence of each cancers, 2) the average age of the subjects in Korean Twin Registry is younger than that of the European study so that the concordance rate could have been underestimated, 3) Relatively conservative operational definition of cancers used in the study, 4) the use of like-sex twins rather than exact monozygotic or dizygotic twins in the study. λ values were estimated to be larger by 2-2.5 for breast and colorectal cancers but less for stomach cancer, mainly because the smaller population prevalence.

Stomach cancer and Colorectal cancer of women

While the absolute comparisons of λ with those of other studies can be inappropriate, the relative size of λ within the study has the meaning. In this context, stomach cancer and colorectal cancer of women showed difference compared with other findings. Stomach cancer indicated less genetic contribution, while colorectal cancer of women showed larger genetic contribution. The difference

can be partially explained from the difference in the environmental exposure of stomach cancer risk factors, such as the prevalence of H pylori infection, high salt diet and smoking rate, etc. which is more prevalent in Korea. However, the relatively high prevalence of stomach cancer compared with the Western studies also could have played the role. The relative difference between male and female colorectal cancers is also new. The men can have more environmental risks of colorectal cancers, such as higher red meat consumption, alcohol intake, and smoking habits. It should be the next step to examine whether the excess genetic contribution in women has something to do with hormonal milieu or sex difference in the life styles.

Case definition and prevalence

Because this study identified cancer cases through record linkage, we could not directly confirm the case except for the cancer registry cases. We adopted operational case definitions to increase the validity of the diagnosis. By applying the operational case definition, the total cancer cases decreased to one third. While the cancer incidence is reported by the National Cancer Registry data, the prevalence of cancers is not reported. The incidence of cancers estimated by the same case definition of the study show 10-28% increase compared

with those reported by cancer registry. Cancer registry data can be underestimated, since it mainly depends of the reports from hospitals. The record linkage of the study was extensive, and it is natural that the incidence of the study exceeds that of the cancer registry. The prevalence of cancers in this study, estimated with the same case definition, is thought to reasonably represent the true level.

Conclusion

The study estimated familial recurrence risks of stomach, breast, and colorectal cancer based on the Korean Twin Registry and a cohort, both of which represent Korean populations. Although zygosity information was limited, overall findings were largely compatible with the previous studies in Caucasian population. The relative genetic contribution estimated by familial recurrence risk showed increase in female colorectal cancer and decrease in stomach cancer, which needs to be further elucidated. The study shows the potentials of the population based data sources to be used to estimate genomic epidemiology index, which can be applied to valuable parameters in gene discovery study.

References

Aarino, M., Sankila, R., and Pukkala, E. (1999). Cancer risk in mutation carriers of DNA-mismatch-repair genes. *Int. J. Cancer* 81, 214-218

Cavalli-Sforza, Bodmer (1978). The Genetics of Human Populations, 2nd Ed.(Freeman and Company)

Edwards, A.O., Ritter, RIII., Abel. K.J., Manning, A., Panhuysen, C., and Farrer, L.A. (2005). Complement factor H polymorphism and age-related macular degeneration. *Science* 308, 421-424.

Fearon, E.R. (1997). Human cancer syndromes: clues to the origin and nature of cancer. *Science* 278, 1043-1050.

Feinberg, A.P. (2004). The epigenetics of cancer etiology. *Semin.Cancer Biol.* 14, 427-432.

Jo, W.S. and Chung, D.C. (2005). Genetics of hereditary colorectal cancer. *Semin.Oncol.* 32, 11-23.

Klein, R.J., Zeiss, C., Chew, E.Y., Tsai, J.Y., Sackler, R.S., Haynes, C., Henning, A.K., SanGiovanni, J.P., Mane, S.M., Mayne, S.T., Bracken, M.B., Ferris, F.L., Ott, J., Barnstable, C., and Hoh, J. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science* 308, 385-389.

Lichtenstein, P., Holm, N.V., Verkasalo, P.K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A., and Hemminki, K. (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.

Ministry of Health and Welfare Research Report (2005). The global burden of disease in Korea.

Risch, N. (1990a). Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am. J. Hum. Genet.* 46, 229-241.

Risch, N. (1990b). Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *Am.J.Hum.Genet.* 46, 242-253.

Shields, P.G. and Harris, C.C. (1991). Molecular epidemiology and the genetics of environmental cancer. *JAMA* 266, 681-687.

Sung, J., Cho, S.H., Cho, S.I., Duffy, D.L., Kim, J.H., Kim, H., Park, K.S., and Park, S.K. (2002). The Korean Twin Registry--methods, current stage, and interim results. *Twin Res.* 5, 394-400,

Varesco, L. (2004). Familial adenomatous polyposis: genetics and epidemiology. *Tech.Coloproctol.* 8 Suppl 2, s305-s308.