Gender difference in colorectal cancer indicators for exercise interventions: the National Health Insurance Sharing Service-Derived Big Data Analysis

Hyunseok Jee[1,2,*], Jong-Hee Kim[3]

[1]Department of Sports Science, Hannam University, Daejeon, Korea
[2]School of Kinesiology, Yeungnam University, Gyeongsan, Korea
[3]Department of Physical Education, Hanyang University, Seoul, Korea

We aimed to examine various characterized features and effects of gender-associated different parameters including exercise on the prevalence of colorectal cancer by using data from the National Health In-surance Sharing Service Database (NHISS DB). Data from NHISS were collected on Koreans aged from 40 to 85 years and were subjected to thematic analysis. The colorectal cancer codes (C19, C20, D011, and D012) from Korean Standard Classification of Disease and Causes of Death selected a target study group, and t-test and logistic regression were used. As results, the age was higher for men who had colorectal cancer than the noncancer group; however, high and low blood pres-sure, hemoglobin, and age had lower values for the cancer group com-pared to their counterparts in women. Only total cholesterol in men and waist size in women between cancer and noncancer groups were shown to have significant differences. Serum glutamic pyruvic trans-aminase and alanine aminotransaminase (SGPT_ALT) showed signifi-cant differences for both sexes. In exercise-related parameter, the re-sponse number 2 (1–2 times/wk, 0.535 for odd ratio) in women and re-sponse number 3 (3–4 times/wk, 0.466 for odd ratio) in men were asso-ciated with a reduced incidence of colon cancer. There was a dif-fer-ence in parameters in colorectal cancer patients over 40 years old for both sexes, but not in SGPT_ALT. Regular physical activity might be one of strong factors affecting or predicting colorectal cancer incidence.

Keywords: Colorectal cancer patient, Big data analysis, National Health Insurance Sharing Service Database, Exercise program development

INTRODUCTION

Westernized unhealthy and unbalanced dietary habits and life-styles contribute to the increased risk for cancer (Haydon et al., 2006). Cancer is the most common cause of death, and colorectal cancer (16.5/100,000 persons) is the third-leading cause of death according to the 2016 report from Statistics Korea (http://kostat. go.kr).

The National Health Insurance Sharing Service (NHISS; https:// www.nhis.or.kr) was implemented in 1963, and the whole Korean population is currently compulsorily enrolled in the NHISS (Choi et al., 2015). Korean population's medical records, thus, can pos-sibly be electronically tracked from the NHISS database (DB).

This is one of the coordinated strategies to solve global health is-sues (Chalkidou and Vega, 2013). NHISS data-derived results have been increasingly reported (Jee et al., 2018).

Various unspecified etiologies of cancer, such as physical stimu-lations (e.g., viruses, radiation, and ultraviolet rays), smoking, and heredity, may cause genetic mutations and ultimately, inhibitor genes (Mafune et al., 2015; Upadhyay et al., 2016). Owing to un-identified patients with cancer in this study, the cure for colorectal cancer was not provided despite conducting numerous trials. In those trials, retrospective patient-oriented studies were also tracked to identify a reasonable solution to prevent cancer symptoms and to understand the various characteristics of colorectal cancer from diverse perspectives. This is important for deciding the factors in-

fluencing the incidence of colorectal cancer, and such efforts play a critical role in developing beneficial intervention programs such as tailor-made exercise prescription programs.

Thus, in this study, we aimed to examine the various characteristic features and the effect of parameters on the prevalence of colorectal cancer using big data sourced from the NHISS DB. To accomplish the goals of this study we conducted the followings:

(a) Patients with colorectal cancer included in the NHISS DB and their demographic characteristics were determined.

(b) Sex-specific parameters based on the characteristics of patients with colorectal cancer were detected using *t*-test and logistic regression analyses, which may provide beneficial information for designing an evidence-based program for alleviating colorectal cancer symptoms.

MATERIALS AND METHODS

Study design

The NHISS-derived data contributed to a longitudinal study in the population registered from 2002 to 2008. Patients with colorectal cancer were tracked using codes from the Korean Standard Classification of Disease and Causes of Death (KSCDCD; http://kssc.kostat.go.kr/ksscNew_web/index.jsp); the code numbers were C19, C20, C218, D011, and D012. From the data of extracted patients with colorectal cancer, sex-specific values of 12 selected parameters were analyzed using *t*-test. A total of 1,740 patients who registered as those harboring benign colon lesions were subsequently tracked to compare with patients in the extracted colorectal cancer group. The codes for colon benign lesions were M814
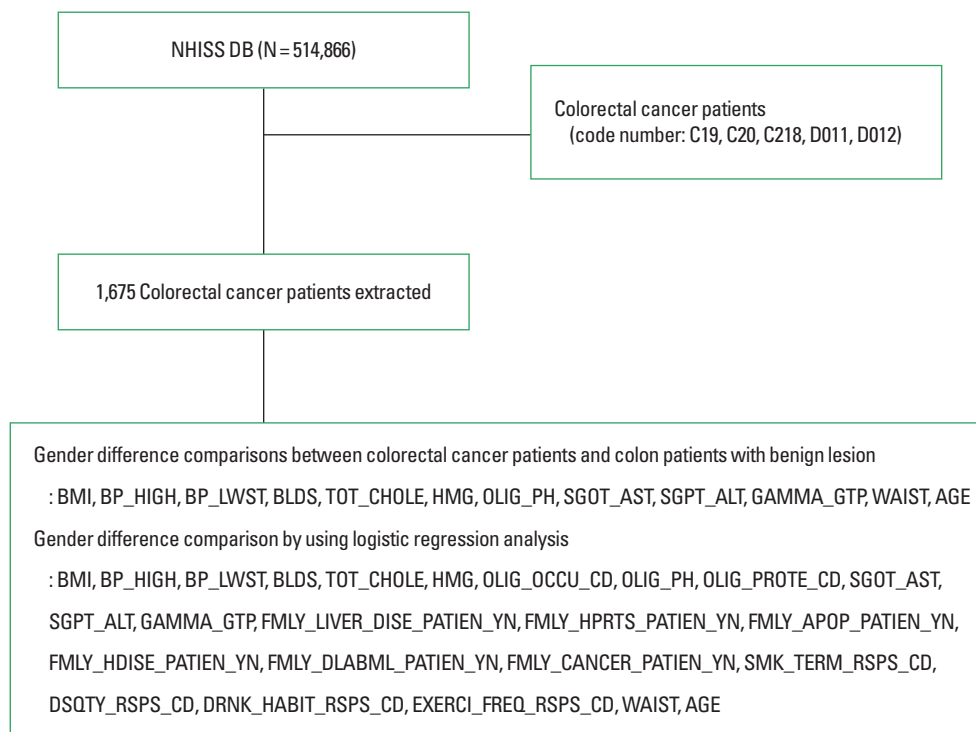
Fig. 1. Flow diagram of the study design. A total of 16,75 patients with a one-time colorectal cancer diagnosed were extracted from 514,866 patients enlisted in the NHISS by using KSCDCD-provided codes (C19, C20, C218, D011, D012). In BMI, BP_HIGH, BP_LWST, TOT_CHOLE, HMG, OLIG_PH, SGOT_AST, SGPT_ALT, GAMMA_GTP, WAIST, AGE were used as parameters to compare between colorectal cancer and noncancer groups (patients with benign lesion in colon) in men and women. Logistic regression was also used to analyze. NHISS, National Health Insurance Sharing Service; KSCDCD, Korean Standard Classification of Disease and Causes of Death; BMI, body mass index (kg/m2); BP_HIGH, high blood pressure (mmHg); BP_LWST, low blood pressure (mmHg); BLDS, fasting blood sugar level (mg/dL); TOT_CHOLE, total cholesterol (mg/dL); HMG, hemoglobin (g/dL); OLIG_OCCU_CD, occult hematuria; OLIG_PH, urine pH level (pH); OLIG_PROTE_CD, protein in urine; SGOT_AST, serum glutamic oxaloacetic transaminase and aspartate aminotransferase (U/L); SGPT_ALT, serum glutamic pyruvic transaminase and alanine aminotransferase (U/L); GAMMA_GTP, gamma glutamyl transpeptidase (U/L); FMLY_LIVER_DISE_PATIEN_YN, hepatism or not; FMLY_HPRTS_PATIEN_YN, family history of high blood pressure; FMLY_APOP_PATIEN_YN, family history of cerebral apoplexy; FMLY_HDISE_PATIEN_YN, family history of cardiopathy; FMLY_DIABML_PATIEN_YN, family history of diabetes; FMLY_CANCER_PATIEN_YN, family history of cancers; SMK_TERM_RSPS_CD, smoking period in the past; DSQTY_RSPS_CD, smoking in a day (pack/day); DRNK_HABIT_RSPS_CD, drinking habit in a week (time/month; EXERCI_FREQ_RSPS_CD, the frequency of doing exercise in a week (time/wk); WAIST, waist circumference (cm); AGE, age (yr).

(adenomas and adenocarcinomas), M838 (adenomas and adenocarcinomas), K621 (rectal polyps), K635 (polyps of the colon), K6358 (other polyps of the colon), D12 (benign neoplasms of the colon, rectum, anus, and anal canal), D123 (benign neoplasms of the transverse colon), and D126 (benign neoplasms of the colon, unspecified). This part of the study is described in Fig. 1.

Data source and subject population

The randomly selected number of patients older than 40 years and representing almost 10% of the whole Korean population in the NIHSS DB pool was 514,866. Whole Koreans are mandatorily required to enregister (Lee et al., 2016), and Korean people older than 40 years should especially undergo regular health checkups (there is unnecessary to recruit the medical record since it has been automatically recorded and there is basically no missing data in the NHISS DB). The pathogenesis of colorectal cancer abruptly increased from the age of 40 years; thus, we focused on analyzing data from the NHISS DB from this age. Patients with cancer were tracked using colorectal cancer-related codes such as C19 (malignant neoplasm of the colon with the rectum), C20 (malignant neoplasm of the rectal ampulla), C21.8 (malignant neoplasm of the anorectal junction), D01.1 (carcinoma *in situ* of the rectosigmoid junction), or D01.2 (carcinoma *in situ* of the rectum) from the KSCDCD. The institutional review board of Seoul National University Bundang Hospital (X-1707-411-903) approved this study.

Categorization of variables

The following variables were used to analyze the obtained data, which were divided as follows: body mass index (BMI; kg/m^2); high blood pressure (BP_HIGH; mmHg); low blood pressure (BP_LWST; mmHg); fasting blood sugar levels (BLDS; mg/dL); total cholesterol levels (TOT_CHOLE; mg/dL); hemoglobin levels (HMG; g/dL); urine glucose (GLY_CD; 1, negative; 2, weakly positive; 3, positive [+1]; 4, positive [+2]; 5, positive [+3]; 6, positive [+4]); occult hematuria (OLIG_OCCU_CD; 1, negative; 2, weakly positive; 3, positive [+1]; 4, positive [+2]; 5, positive [+3]; 6, positive [+4]); urine pH levels (OLIG_PH; pH); protein in urine (OLIG_PROTE_CD; 1, negative; 2, weakly positive; 3, positive [+1]; 4, positive [+2]; 5, positive [+3]); serum glutamic oxaloacetic transaminase and aspartate aminotransferase levels (SGOT_AST; U/L); serum glutamic pyruvic transaminase and alanine aminotransaminase levels (SGPT_ALT; U/L); gamma glutamyl transpeptidase levels (GAMMA_GTP; U/L); individual disease history (HCHK_PMH_CD; 1, tuberculosis; 2, hepatitis; 3, hepatism; 4, high blood pressure; 5, cardiopathy; 6, cerebral apoplexy; 7, diabetes; 8, cancer; 9, others); hepatism (FMLY_LIVER_DISE_PATIEN_YN; 1, no; 2, yes); family history of high blood pressure (FMLY_HPRTS_PATIEN_YN; 1, no; 2, yes); family history of cerebral apoplexy (FMLY_APOP_PATIEN_YN; 1, no; 2, yes); family history of cardiopathy (FMLY_HDISE_PATIEN_YN; 1, no; 2, yes); family history of diabetes (FMLY_DIABML_PATIEN_YN; 1, no; 2, yes); family history of cancers (FMLY_CANCER_PATIEN_YN; 1, no; 2, yes); current smoking status (SMK_STAT_TYPE_RSPS_CD; 1, no; 2, quit; 3, smoking); smoking period in the past (SMK_TERM_RSPS_CD; 1, within 6 years; 2, 6–9 years; 3, 10–19 years; 4, 20–29 years; 5, over 30 years); smoking in a day (DSQTY_RSPS_CD; 1, half-pack/day; 2, half-pack/day–1 pack/day; 3, 1 pack/day–2 packs/day; 4, more than 2 packs/day); alcohol consumption habits per week (DRNK_HABIT_RSPS_CD; 1, almost nondrinking; 2, 2–3 times/month; 3, 1–2 times/week; 4, almost every day); alcohol drinking quantity (360 mL) of 20% alcohol by volume at once, (TM1_DRKQTY_RSPS_CD; 1, less than half a bottle; 2, one bottle; 3, one and a half bottle; 4, more than 2 bottles); frequency of moderate-intensity exercise per week (EXERCI_FREQ_RSPS_CD); waist circumference (WAIST; cm); and age (AGE; years).

Statistical analysis

All data are presented as means ± standard deviations. In the sex-specific comparison, *t*-test was used to analyze the mean difference between the cancer and noncancer groups of both sexes. Using logistic regression analysis between patients of different sexes, the effect of various parameters on colorectal carcinogenesis and odds ratios (ORs), *P*-values, and 95% confidence intervals (CIs) were calculated. SAS ver. 9.4 (SAS Institute, Cary, NC, USA) and IBM SPSS ver. 18.0 (IBM Co., Armonk, NY, USA) were used for all statistical analyses. A *P*-value of < 0.05 was considered to indicate a statistically significant difference in all analyses.

RESULTS

Subject characteristics

Medical histories of all Korean people who were mandatorily registered were tracked through the NHISS DB, which included 514,866 randomly selected representatives of the Korean population who were older than 40 years (a 10th ratio; thus, the whole population older than 40 years was approximately 5 million) were analyzed. We found that among the 514,866 people, the men and women's representative populations were 279,125 and 235,741

people, respectively. Their average age was 58.56 ± 9.58 years, which was calculated based on the age range from 40 to 85 years. Importantly, we could extract 1,675 patients with colorectal cancer from the DB using the cancer code decided from the KSCDC (Table 1).

Sex-based differences in 12 parameters according to colorectal cancer pathogenesis

Table 2 shows that the cancer and noncancer groups were compared based on sex (Table 2). Parameters that required selection of numbers for responses (questionnaires) were excluded in Table 2. Among the 12 parameters, significant differences ($P < 0.01$) were observed for BMI, BLDS, HMG, SGPT_ALT, WAIST, and AGE in men with cancer. Only BLDS was significantly higher in the men of the cancer groups, whereas higher P_HIGH, BP_LWST, BLDS, and AGE were observed in the women of the cancer group than in those of the noncancer group ($P < 0.05$ and $P < 0.01$, respectively). BLDS in men with colorectal cancer increased by almost 5% ($P < 0.01$). Approximately 12% of women in the cancer group had increased AGE ($P < 0.01$).

Factors affecting colorectal cancer analyzed using logistic regression analysis

Significant factors affecting colorectal cancer were dependent on sex-based differences, except family-related factors, such as FMLY_HPRTS_PATIEN_YN, FMLY_APOP_PATIEN_YN, and FMLY_CANCER_PATIEN_YN. DRNK_HABIT_RSPS_CD, EXERCI_FREQ_RSPS_CD, and AGE were significant factors (Table 3). Significant factors were as follows: BMI (men), BP_LWST (women), BLDS (men), HMG (men), GLY_CD (men), OLIG_OCCU_CD (men), OLIG_PROTE_CD (men), SGPT_ALT (men), HCHK_PMH_CD (women), FMLY_LIVER_DISE_PATIEN_YN (men), SMK_STAT_TYPE_RSPS_CD (men), and WAIST (men) ($P < 0.05$ and $P < 001$). Interestingly, the protein in urine (OLIG_PROTE_CD) in men who responded as number 2 (weakly positive) conferred 2.173 times greater colorectal carcinogenesis than that in control men ($P < 0.035$; 95% CI, 1.054–4.479). Regarding individual disease history (HCHK_PMH_CD), women with cardiop-

Table 1. Demographic characteristics of colorectal cancer patients derived from NHISS DB

Parameter	Value
Total population	514,866
Men:women	279,125: 235,741
Age (yr), mean± SD	58.56± 9.58
Age range (yr)	40–85
Following year of the study	2002–2008
No. of subtracted colorectal cancer patients	1675
Used colorectal cancer code	C19, C20, C218, D011, D012

The codes indicating the colorectal cancer were traced from Korean standard classification of disease and causes of death.
NHISS DB, National Health Insurance Sharing Service Data Base; SD, standard deviation.

Table 2. Sexual difference-based parameters result between cancer and noncancer groups

Variable	Men		Women	
	Cancer	Noncancer	Cancer	Noncancer
BMI	23.55±2.92**	24.40±2.63	23.92±3.21	24.06±3.05
BP_HIGH	129.97±17.67	128.64±15.48	125.96±18.04**	122.08±16.20
BP_LWST	80.06±10.85	77.92±10.21	76.75±11.36**	74.93±10.51
BLDS	105.26±42.59**	100.94±23.29	101.84±48.80*	96.25±23.65
TOT_CHOLE	192.80±38.41	193.54±34.22	203.32±51.46	204.63±37.43
HMG	14.24±1.48**	14.63±1.24	12.57±1.43**	12.92±1.12
OLIG_PH	6.08±0.61	6.06±0.58	6.08±0.63	6.10±0.60
SGOT_AST	26.76±13.35	27.49±17.88	27.90±46.34	24.40±12.06
SGPT_ALT	24.60±15.52**	26.72±15.40	23.76±47.32	22.20±18.58
GAMMA_GTP	47.44±73.90	43.45±43.35	25.13±47.56	24.65±26.52
WAIST	83.97±7.78**	86.66±7.26	79.20±11.72	79.45±8.41
AGE	63.67±9.31*	62.74±7.90	64.59±9.22**	57.52±8.00

Values are presented as mean± standard deviation.
BMI, body mass index (kg/m²); BP_HIGH, high blood pressure (mmHg); BP_LWST, low blood pressure (mmHg); BLDS, fasting blood sugar level (mg/dL); TOT_CHOLE, total cholesterol (mg/dL); HMG, hemoglobin (g/dL); OLIG_PH, urine pH level (pH); SGOT_AST, serum glutamic oxaloacetic transaminase and aspartate aminotransferase (U/L); SGPT_ALT, serum glutamic pyruvic transaminase and alanine aminotransaminase (U/L); GAMMA_GTP, gamma glutamyl transpeptidase (U/L); WAIST, waist circumference (cm).
*$P < 0.05$, **$P < 0.01$, statistical significances between cancer and noncancer groups.

athy (response of number 5) harbored 22 times greater colorectal carcinogenesis than control women ($P < 0.01$; 95% CI, 6.440–75.161). Unexpected interesting data of BMI (men), SMK_STAT_TYPE_RSPS_CD (men), and WAIST (men) revealed ORs of

Table 3. Logistic regression analysis of the parameters on colorectal carcinogenesis

Parameter	Men			Women		
	OR	*P*-value	95% CI	OR	*P*-value	95% CI
BMI	0.895	0.000	0.868–0.922	1.028	0.683	0.902–1.172
BP_HIGH	1.005	0.054	1.000–1.010	0.980	0.146	0.954–1.007
BP_LWST	1.004	0.260	0.997–1.012	1.050	0.015	1.009–1.092
BLDS	1.004	0.003	1.001–1.007	1.010	0.083	0.999–1.020
TOT_CHOLE	0.999	0.623	0.997–1.002	1.000	0.974	0.993–1.007
HMG	0.807	0.000	0.758–0.860	0.828	0.102	0.660–1.038
GLY_CD (3)<1>	3.327	0.003	1.494–7.049	3.320	0.299	0.344–32.020
OLIG_OCCU_CD (2) <3>	0.604	0.033	0.381–0.959	0.594	0.126	0.305–1.158
OLIG_PH	1.068	0.347	0.932–1.224	1.109	0.625	0.733–1.676
OLIG_PROTE_CD (2)<2>	2.173	0.035	1.054–4.479	0.633	0.341	0.247–1.622
SGOT_AST	0.997	0.277	0.991–1.003	1.023	0.207	0.988–1.059
SGPT_ALT	0.991	0.001	0.985–0.995	0.977	0.265	0.937–1.018
GAMMA_GTP	1.001	0.116	1.000–1.003	0.991	0.371	0.973–1.010
HCHK_PMH_CD (5)<7>	0.130	0.059	0.016–1.079	22.000	0.000	6.440–75.161
FMLY_LIVER_DISE_PATIEN_YN	0.505	0.006	0.310–0.822	0.367	0.054	0.132–1.016
FMLY_HPRTS_PATIEN_YN	0.546	0.000	0.403–0.739	0.537	0.005	0.346–0.833
FMLY_APOP_PATIEN_YN	0.684	0.032	0.483–0.968	0.481	0.047	0.234–0.990
FMLY_DIABML_PATIEN_YN	0.800	0.246	0.549–1.166	0.788	0.352	0.477–1.302
FMLY_CANCER_PATIEN_YN	0.772	0.034	0.607–0.980	0.533	0.000	0.382–0.743
SMK_STAT_TYPE_RSPS_CD (1)<1>	1.413	0.004	1.117–1.787	0.729	0.730	0.121–4.382
SMK_TERM_RSPS_CD (2)<4>	1.858	0.107	0.876–3.942	6.000	0.224	0.335–107.420
DSQTY_RSPS_CD (3)<1>	0.380	0.172	0.095–1.521	0.292	0.112	0.064–1.333
DRNK_HABIT_RSPS_CD (2)<1>	0.575	0.000	0.462–0.715	0.478	0.003	0.294–0.776
TM1_DRKQTY_RSPS_CD (1)<2>	0.779	0.051	0.607–1.001	1.737	0.699	0.106–28.573
EXERCI_FREQ_RSPS_CD (3)<2>	0.466	0.000	0.310–0.699	0.535	0.002	0.362–0.792
WAIST	0.952	0.000	0.934–0.971	0.980	0.397	0.935–1.027
AGE	1.013	0.010	1.003–1.022	1.095	0.000	1.058–1.134

Variables with significant *P*-values were selected via logistic regression analysis.
(Number) following the parameters (e.g., EXERCI_FREQ_RSPS_CD (1)) indicates (number) selected in the questionnaire of the parameter, which is the nearest to the *P*-value (it means *P*-value of (1) is the nearest to the significant difference among the whole (number) used in EXERCI_FREQ_RSPS_CD parameter). In EXERCI_FREQ_RSPS_CD (1), EXERCI_FREQ_RSPS_CD indicates the frequency of doing exercise in a week and number (1) indicates doing no exercise (described below). () and < > indicate the questionnaire of men and women, respectively.
OR, odds ratio; CI, confidence interval; BMI, body mass index (kg/m2); BP_HIGH, high blood pressure (mmHg); BP_LWST, low blood pressure (mmHg); BLDS, fasting blood sugar level (mg/dL); TOT_CHOLE, total cholesterol (mg/dL); HMG, hemoglobin (g/dL); GLY_CD, urine glucose (1, negative; 2, weak positive; 3, positive (+1); 4, positive (+2); 5, positive (+3); 6, positive (+4)); OLIG_OCCU_CD, occult hematuria (1, negative; 2, weak positive; 3, positive (+1); 4, positive (+2); 5, positive (+3); 6, positive (+4)); OLIG_PH, urine pH level (pH); OLIG_PROTE_CD, protein in urine (1, negative; 2, weak positive; 3, positive (+1); 4, positive (+2); 5, positive (+3)); 6, positive (+4)); SGOT_AST, serum glutamic oxaloacetic transaminase and aspartate aminotransferase (U/L); SGPT_ALT, serum glutamic pyruvic transaminase and alanine aminotransaminase (U/L); GAMMA_GTP, gamma glutamyl transpeptidase (U/L); HCHK_PMH_CD, individual disease history (1, tuberculosis; 2, hepatitis; 3, hepatism; 4, high blood pressure; 5, cardiopathy; 6, cerebral apoplexy; 7, diabetes; 8, cancer; 9, others); FMLY_LIVER_DISE_PATIEN_YN, hepatism or not (1, no; 2, yes); FMLY_HPRTS_PATIEN_YN, family history of high blood pressure (1, no; 2, yes); FMLY_APOP_PATIEN_YN, family history of cerebral apoplexy (1, no; 2, yes); FMLY_DIABML_PATIEN_YN, family history of diabetes (1, no; 2, yes); FMLY_CANCER_PATIEN_YN, family history of cancers (1, no; 2, yes); SMK_STAT_TYPE_RSPS_CD, current smoking state (1, no; 2, quit; 3, smoking); SMK_TERM_RSPS_CD, smoking period in the past (1, within 6 years; 2, 6–9 years; 3, 10–19 years; 4, 20–29 years; 5, over 30 years); DSQTY_RSPS_CD, smoking in a day (1, half-pack/day; 2, half-pack/a day–a pack/day; 3, a pack/day–two packs/day; 4, more than 2 packs/day); DRNK_HABIT_RSPS_CD , drinking habit in a week (1, almost nondrinking; 2, 2–3 times/month; 3, 1–2 times/wk; 4, almost everyday); TM1_DRKQTY_RSPS_CD, alcohol drinking quantity (360 mL) of 20 percent alcohol by volume at once, (1, below the half bottle; 2, one bottle; 3, one and half; 4, over 2 bottles); EXERCI_FREQ_RSPS_CD, the frequency of doing exercise in a week (1, no exercise; 2, 1–2 times/wk; 3, 3–4 times/wk; 4, 5–6 times/wk; 5, everyday); WAIST, waist circumference (cm); AGE, age (yr)

0.895 (95% CI, 0.868–0.922), 1.413 (95% CI, 1.117–1.787), and 0.952 (95% CI, 0.934–0.971), respectively ($P < 0.01$).

DISCUSSION

The results of this retrospective cohort of patients with colorectal cancer who were aged older than 40 years and registered in the NHISS DB were as follows.

(a) A total of 1,675 patients with colorectal cancer who were traced through C19, C20, C218, D011, and D012 codes were found from the NHISS DB during the study years from 2002 to 2008.

(b) BMI, BLDS, HMG, SGPT_ALT, WAIST, and AGE were higher in the men of the cancer group than in those of the noncancer group ($P < 0.01$). All values, except those of BLDS, were lower in the men of the cancer group than in those of the noncancer group. However, BP_HIGH, BP_LWST, BLDS, and AGE were higher in the women of the cancer group than in those of the noncancer group ($P < 0.05$ and $P < 0.01$, respectively).

(c) Familial disease-related history (high blood pressure, cerebral apoplexy, and cancer), alcohol consumption habits, exercise frequency, and age differences significantly affected colon carcinogenesis without sex differences ($P < 0.05$). Twelve sex difference-dependent factors, such as waist circumference in men ($P < 0.01$; OR, 0.952; 95% CI, 0.934–0.971) and diastolic pressure in women ($P < 0.05$; OR, 1.050; 95% CI, 1.009–1.092), that affected colorectal cancer were also reported.

Big data analysis to identify parameters related to colorectal cancer incidence

To possibly conquer cancer or at least ameliorate the symptoms of cancer, convergent insights from multidisciplinary aspects of cancer-related studies are indispensable. *In vivo* and *in vitro* (Jee et al., 2016) results from direct and practical experiments have been important; furthermore, big data analysis can be added as an allied aid. A study introduced a synergistic double-check of the interrelationship of cancer-regulating genes in colorectal cancer from precedent published datasets to build target gene DBs based on data mining from the IluminaGA_miRNASeq platform served by The Cancer Genome Atlas (https://cancergenome.nih.gov/) by validating *in vitro* experiments. Especially, Koreans mandatorily enregistered in the large-scale NHISS DB can be adapted for ideal research (i.e., cross-sectional, retrospective, and prospective studies

of each individualized patient) (Kim et al., 2016). We believe that these big data-related research activities can contribute to a new frontier field committed to national health and welfare.

Individualized exercise interventions to reduce the incidence of colorectal cancer

Colorectal cancer is a terrifying disease called "quiet cancer" because it metastasizes with no specific symptoms. It occurs by defying the golden rule for preventing colorectal cancer, even though patients are usually aware of their defiance. The golden rule for preventing colorectal cancer involves a healthy diet and regular lifestyle. The easiest means to prevent colorectal cancer is to maintain a healthy lifestyle by engaging in 30-min walking exercises and avoiding excessive intake of red meat, for example. In our findings, exercise-related findings (EXERCI_FREQ_RSPS_CD) were the most clear results obtained from the aspect of the statistical analysis, in that the response numbers 2 (1–2 times/wk) in women and response number 3 (3–4 times/wk) in men was associated with a reduced incidence of colon cancer ($P < 0.01$). In women, the ORs of response numbers 1 and 2 were 0.462 and 0.535, respectively; in men, the ORs of response numbers 1, 2, and 3 were 0.544, 0.496, and 0.466, respectively. These results suggest that designing tailor-made ideal exercise programs using the analyzed parameters in this study is one way to prevent or alleviate the symptoms of colorectal cancer, although the exercise-related parameter in this study did not indicate details such as the intensity or duration of exercise because these factors appear to be more important to obtain the maximal effect of exercise intervention. Jee et al. (2016) reported that high rather than moderate intensity of exercise confers a greater effect on various parameters of cancer models. To design an individual exercise program, adjusting for the markers found to be significant, such as SEX, BMI, BP_LWST, BLDS, HMG, SGPT_ALT, and WAIST, to specify an idealistic exercise program should be developed to prevent colorectal cancer.

As shown by the unexpected interesting results obtained in this study (e.g., OLIG_OCCU_CD only in men), we also consider that the appropriate response of patients with colorectal cancer is dependent on sex-based parameters, which suggests that there seem to exist proper ideal values of the relevant parameters to at least suppress the prevalence of colon cancer (Vainio et al., 2002; Thune and Lund, 1996).

Prescriptive exercise programs can be thus designed according to individualized, age-, and sex-based differences according to frequency, intensity, time (duration), and type of exercise because it has been suggested that they may reflect different susceptibilities

that induce non-identical effects of exercise interventions (Bufill, 1990; Dubrow et al., 1993).

Sex differences and other etiological patterns in logistic regression analysis for colorectal carcinogenesis

Regarding logistic regression analysis, sex-based differences are also minimally shown in this study. In Table 3, different patterns with significance are shown only in BMI, BLDS, HMG, GLY_CD, OLIG_OCCU_CD, OLIG_PROTE_CD, and SGPT_ALT in men with colorectal cancer; however, only BP_LWST and HCHK_PMH_CD are significantly different in women ($P < 0.05$). Interestingly, 1.955 times higher risk for colorectal cancer was found in women with a family history of cerebral apoplexy ($P < 0.05$; 95% CI, 1.007–3.796) and 0.435 times higher risk for the incidence of colorectal cancer was found in women with a family history of diabetes. Regarding the relation between cerebral apoplexy and colorectal cancer, some studies have found that bevacizumab (antitumor angiogenesis drug) increases the risk for cerebral apoplexy as a side effect in patients with colorectal cancer (Lv et al., 2012). Aspirin intake for colorectal cancer increases the risk for developing cerebral hemorrhages (Dehmer et al., 2016). It suggests that patients with colorectal cancer would take aspirin because of secondary diseases, which may induce blood disorders in the brain. Regarding the diabetes and colorectal cancer incidences, we obtained consistent results with those of the study by de Kort et al. (2016). However, there are discrepant results from various studies regarding the relation between colorectal cancer and family disease histories. For the pursuing the etiological relationship between the incidence of colorectal cancer and various factors described in this study such as family disease histories, longitudinal etiology-related study can be one of the way to offset the arising issue (Lin et al., 2018). This is also a way of raising the quality to be pursued for the next study.

The results from the large-scale NIHSS DB suggest that sex-based differences in parameters are observed in patients with colorectal cancer aged older than 40 years, except for the factors described above; thus, sex difference-adjusted parameters significantly reflect evidence-based programs (such as reasonable exercise interventions) that can be developed for alleviating prognostic colorectal cancer symptoms.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

Bufill JA. Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. Ann Intern Med 1990;113:779-788.

Chalkidou K, Vega J. Sharing the British National Health Service around the world: a self-interested perspective. Global Health 2013;9:51.

Choi Y, Kim JH, Yoo KB, Cho KH, Choi JW, Lee TH, Kim W, Park EC. The effect of cost-sharing in private health insurance on the utilization of health care services between private insurance purchasers and non-purchasers: a study of the Korean health panel survey (2008-2012). BMC Health Serv Res 2015;15:489.

de Kort S, Simons CC, van den Brandt PA, Goldbohm RA, Arts IC, de Bruine AP, Janssen-Heijnen ML, Sanduleanu S, Masclee AA, Weijenberg MP. Diabetes mellitus type 2 and subsite-specific colorectal cancer risk in men and women: results from the Netherlands Cohort Study on diet and cancer. Eur J Gastroenterol Hepatol 2016;28:896-903.

Dehmer SP, Maciosek MV, Flottemesch TJ, LaFrance AB, Whitlock EP. Aspirin for the primary prevention of cardiovascular disease and colorectal cancer: a decision analysis for the US preventive services task force aspirin to prevent CVD and CRC: a decision analysis for the USPSTF. Ann Intern Med 2016;164:836-845.

Dubrow R, Bernstein J, Holford TR. Age-period-cohort modelling of large-bowel-cancer incidence by anatomic sub-site and sex in Connecticut. Int J Cancer 1993;53:907-913.

Haydon AM, Macinnis RJ, English DR, Giles GG. Effect of physical activity and body size on survival after diagnosis with colorectal cancer. Gut 2006;55:62-67.

Jee H, Chang JE, Yang EJ. Positive prehabilitative effect of intense treadmill exercise for ameliorating cancer cachexia symptoms in a mouse model. J Cancer 2016;7:2378-2387.

Jee H, Lee HD, Lee SY. Evidence-based cutoff threshold values from receiver operating characteristic curve analysis for knee osteoarthritis in the 50-year-old Korean population: analysis of big data from the National Health Insurance Sharing Service. Biomed Res Int 2018;2018:2013671.

Kim HY, Jang EJ, Park B, Kim TY, Shin SA, Ha YC, Jang S. Development of a Korean Fracture Risk Score (KFRS) for predicting osteoporotic fracture risk: analysis of data from the Korean National Health Insur-

ance Service. PLoS One 2016;11:e0158918.

Lee K, Kim H, Lee JH, Jeong H, Shin SA, Han T, Seo YL, Yoo Y, Nam SE, Park JH, Park YM. Retrospective observation on contribution and limitations of screening for breast cancer with mammography in Korea: detection rate of breast cancer and incidence rate of interval cancer of the breast. BMC Womens Health 2016;16:72.

Lin H, Tang X, Shen P, Zhang D, Wu J, Zhang J, Lu P, Si Y, Gao P. Using big data to improve cardiovascular care and outcomes in China: a protocol for the CHinese Electronic health Records Research in Yinzhou (CHERRY) Study. BMJ Open 2018;8:e019698.

Lv J, Liu N, Liu KW, Ding AP, Wang H, Qiu WS. A randomised controlled phase II trial of the combination of XELOX with thalidomide for the first-line treatment of metastatic colorectal cancer. Cancer Biol Med 2012;9:111-114.

Mafune A, Hama T, Suda T, Suzuki Y, Ikegami M, Sakanashi C, Imai S, Nakashima A, Yokoo T, Wada K, Kojima H, Urashima M. Homozygous deletions of UGT2B17 modifies effects of smoking on TP53-mutations and relapse of head and neck carcinoma. BMC Cancer 2015; 15:205.

Thune I, Lund E. Physical activity and risk of colorectal cancer in men and women. Br J Cancer 1996;73:1134-1140.

Upadhyay P, Nair S, Kaur E, Aich J, Dani P, Sethunath V, Gardi N, Chandrani P, Godbole M, Sonawane K, Prasad R, Kannan S, Agarwal B, Kane S, Gupta S, Dutt S, Dutt A. Notch pathway activation is essential for maintenance of stem-like cells in early tongue cancer. Oncotarget 2016;7:50437-50449.

Vainio H, Kaaks R, Bianchini F. Weight control and physical activity in cancer prevention: international evaluation of the evidence. Eur J Cancer Prev 2002;11 Suppl 2:S94-100.