BMJ Open

Cohort profile: The Health and Prevention Enhancement (H-PEACE), a retrospective, population-based cohort study conducted at the Seoul National University Hospital Gangnam Center, Korea

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019327
Article Type:	Cohort profile
Date Submitted by the Author:	25-Aug-2017
Complete List of Authors:	Lee, Changhyun; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Choe, Eun Kyung; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Choi, Ji Min; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Hwang, Yunji; Seoul National University College of Medicine, Department of Preventive Medicine; Seoul National University Graduate School, Department of Biomedical Science Lee, Young; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Park, Boram; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Chung, Su Jin ; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Kwak, Min-Sun ; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Lee, Jong-Eun ; DNA link Kim, Joo Sung ; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Park, Sue K.; Seoul National University College of Medicine, Department of Preventive Medicine; Seoul National University Graduate School, Department of Biomedical Science Cho, Sang-Heon; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Patient-centred medicine
Keywords:	cohort, non-communicable disease, PREVENTIVE MEDICINE, EPIDEMIOLOGY

Cohort profile: The Health and Prevention Enhancement (H-PEACE), a retrospective, population-based cohort study conducted at the Seoul National University Hospital Gangnam Center, Korea

Short title: Cohort profile of the H-PEACE study

Changhyun Lee[1¶], Eun Kyung Choe[1¶], Ji Min Choi[1], Yunji Hwang[2,3,4], Young Lee[1], Boram Park[1], Su Jin Chung[1], Min-Sun Kwak[1], Jong-Eun Lee[5], Joo Sung Kim[1], Sue K. Park[2,3,4&*], Sang-Heon Cho[1&*], and the H-PEACE study investigators

[1]Healthcare Research Institute, Healthcare System Gangnam Center, Seoul National University Hospital, Seoul, Korea

[2]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea

[3]Department of Biomedical Science, Seoul National University Graduate School, Seoul, Korea

[4]Seoul National University Cancer Research Institute, Seoul, Korea

[5]DNA link, Inc., Seoul, Korea

[¶]Changhyun Lee and Eun Kyung Choe contributed equally to this paper.

[&]Sue Kyung Park and Sang-Heon Cho contributed equally to this paper.

***Corresponding authors**

1. Sang-Heon Cho

Gangnam Finance Center, Teheran-ro 152, Gangnam-gu, Seoul, 06236, Korea

1 Tel: +82-2-2072-2971; E-mail: shcho@snu.ac.kr

2 2. Sue K. Park

3 Dahangno 103, Chongno-gu, Seoul 03080, Korea

4 Tel: +82-2-740-8333

5 E-mail: suepark@snu.ac.kr

6

7 **Word count:** 2,807

8

ABSTRACT

Purpose: The Health and Prevention Enhancement (H-PEACE) study was designed to investigate the association of diagnostic imaging results, biomarkers and the pre-disease stage of non-communicable diseases (NCDs), such as malignancies and metabolic diseases, in an average-risk population in Korea.

Participants: This study enrolled a large-scale retrospective cohort at the Healthcare System Gangnam Center, Seoul National University Hospital, from October 2003 to December 2014.

Findings to date: The baseline and follow-up information collected in the pre-disease stage of NCDs allows for evaluation of an individual's potential NCD risk, which is necessary for establishing personalized prevention strategies. A total of 91,336 health examinees were included in the cohort, and we repeatedly measured and collected information for 50.9% (N=46,484) of the cohort members. All participants completed structured questionnaires (lifestyle, medical history, mini-dietary assessment index, sex-specific variables, and psychiatric assessment), doctors' physical examinations, laboratory blood and urine tests and digital chest X-ray imaging. For participants with available data, we also obtained information on specific diagnostic variables using advanced diagnostic tests, including coronary computed tomography (CT) for coronary calcium scores, colonoscopy, and brain magnetic resonance imaging (MRI). Furthermore, 17,455 of the participants who provided informed consent and donated blood samples were enrolled into the Gene-environmental interaction and phenotype (GENIE) study, a subcohort of the H-PEACE, from October 2013 and we analyzed genome-wide single nucleotide polymorphism (SNP) array data for 6,579 of these blood samples.

Future plans: The data obtained from this cohort will be used to facilitate advanced and accurate diagnostic techniques related to NCDs while considering various phenotypes. Potential collaborators can access the dataset after receiving approval from our institutional review board. Applications can be submitted on the study homepage

(http://healthcare.snuh.org/HPEACEstudy).

Keywords: epidemiology; cohort; non-communicable disease; preventive medicine

STRENGTHS AND LIMITATIONS OF THIS STUDY

- The strengths of the H-PEACE study include a large number of healthy subjects (N=91,336 healthy examinees, from the Healthcare System Gangnam Center, Seoul National University Hospital, between 2003 and 2014) and a structured and organized database.
- This study not only includes data widely used in medical check-ups but also data from sophisticated high-quality advanced examinations to investigate clinical effectiveness in predicting the pre-disease stage of non-communicable diseases, including malignancies and metabolic diseases, in an average-risk population in Korea.
- Another strength includes the active and passive follow-ups and the ability to obtain complete data, including deaths and incidental cancer cases, even among those who discontinued visiting our center.
- The data from this study will allow us to contribute to active, effective prevention of the development of cancers and non-communicable diseases.
- The major weakness of this cohort is that it may show selection bias because only subjects who voluntarily visited our center were included in the study.

INTRODUCTION

In recent decades, the prevalence of non-communicable diseases (NCDs), such as malignancy, metabolic disease, and cardiovascular disease, has rapidly increased in Korea.[1,2] To address this problem, a comprehensive approach that accounts for lifestyle, environmental factors and genetic variability is needed, as NCDs are known to be caused by both genetic and environmental factors.[3,4] Precision medicine is emerging as a potential solution.[5,6] This type of medicine categorizes individuals into different subgroups based on their susceptibility to disease and then focuses on individuals in whom interventions will be helpful.[7] One of the major components of efficient prevention and early detection of NCDs is thus identifying high-risk populations among those in a pre-disease or asymptomatic stage.

The Seoul National University Hospital (SNUH) Healthcare System Gangnam Center provides comprehensive medical check-ups and screening, and nearly 20,000 people visit this center each year. A data warehouse, HEALTH-WATCH®, was built as a prototype database for this cohort study within the healthcare research institute from the start of the center in 2003.[8] Using HEALTH-WATCH®, we have published many articles that have primarily been focused on cancer screening and metabolic diseases.[8-19] In 2013, we re-organized the cohorts to support our research and started collecting blood samples to analyze the genetic factors involved in NCDs. We then summarized and integrated our clinical and genetic data in the Health and Prevention Enhancement (H-PEACE) study.

The H-PEACE study was designed to investigate the pre-disease stage of NCDs, which helps us assess an individual's risk of NCDs and establish personalized prevention strategies. Through analyses of longitudinal data from comprehensive questionnaires and clinical and laboratory tests, the H-PEACE study can expand our knowledge of NCD prevention, as well as our knowledge of high-risk populations, to improve the early detection of NCDs. The findings can inform treatment priorities and therapeutic guidelines related to

NCDs. Finally, the H-PEACE study can contribute to enhancing public health and improving quality of life.

COHORT DESCRIPTION

Participants

The H-PEACE study collected data from 91,336 individuals aged 45.5 ± 11.7 years (50,507 men and 40,829 women) who received a health check-up between October 2003 and December 2014 at the SNUH Gangnam Center (IRB No H-1311-031-531). We prospectively collected blood samples from 17 455 of those individuals (9396 men and 8059 women) from October 2013 to form the Gene-environmental interaction and phenotype (GENIE) study, a subcohort of the H-PEACE study. The informed consent and study protocols were approved (IRB No H-1103-127-357 & H-1505-047-671). Furthermore, we obtained data for genome-wide single nucleotide polymorphism (SNP) arrays using 6579 donated blood samples. The flowchart of the H-PEACE study subjects is illustrated in Fig 1, and the distribution of the participants in Korea in Fig 2.

Follow-up

Annual health exams are mandatory for all workers under the Industrial Safety and Health Law in Korea. Annual or biannual health screening exams are recommended in our center. Every year, we provide reminder calls to encourage individuals to attend their health check-ups. In the H-PEACE study, 46 484 of 91 336 (50.9%) recipients completed the 1st follow-up assessment, and 29 820 recipients (32.6%) completed the 2nd (Fig 1). The median follow-up was 4.04 years (interquartile range [IQR] 2.1 – 6.5). In the GENIE study, 11 286 of 17 455 (64.7%) recipients completed the 1st follow-up test and 10 184 recipients (58.4%) completed the 2nd. The median follow-up of the GENIE study was 5.23 years (IQR 2.7 – 8.0). Furthermore, we have annually updated the survival data by linking to national death certificates through requests to Statistics Korea.

Data collection

The H-PEACE study consists of core and specific variables (Tables 1 and 2). The core variables were tests performed in all of the enrolled participants at the regular health check-ups. The specific variables were the tests selectively performed in the participants upon request or based on the recommendation of their medical provider according to the participant's symptoms or disease risk factors. These tests were sophisticated and expensive tests that are rarely utilized in other cohort profiles. In Table 3, we describe the representative test for each variable and its participation rate. The tests performed for the covariables and specific variables are described below. All data were collected as part of the comprehensive health check-ups. The flow diagram of health check-ups in our study is illustrated in Fig 3. All study protocols are available upon request at our homepage (http://healthcare.snuh.org/HPEACEstudy).

Table 1. Core variables collected at baseline and follow-up in the Health and Prevention Enhancement (H-PEACE) study from 2003 to 2014

Item	List
Structured questionnaires	Demographic and lifestyles questionnaire: socioeconomic variables, including education, household income, marital status, and occupation; past medical history of diabetes, hypertension, dyslipidemia, and cancer; family history of diabetes, hypertension, dyslipidemia, and cancer in first relatives; alcohol consumption and cigarette smoking Physical activity: Korea-validated version of the International Physical Activity Questionnaire (IPAQ) short form Dietary factors: Mini-Dietary Assessment Index (MDAI) Sex-specific questionnaire: International Prostate Symptom score (IPSS) and the International Index of Erectile Function (IIEF-5) for men; reproductive factors for women, including menstrual history (age at menarche, length and regularity of menstrual cycle), menopausal status and age, pregnancy history (including the number/methods of pregnancies and abortions), and breastfeeding Psychiatric assessment questionnaire: Beck Depression Inventory (BDI) and Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR 16)
Anthropometric measurements/ Physical examination	Systolic and diastolic blood pressure Height, weight, and waist circumference Body composition using a multi-frequency bio-impedance analyzer Physicians' physical examination Eye examination, including visual acuity, ocular tonometry, and slit lamp test Fundus photography for persons ≥ 35 years of age

	Hearing test for persons ≥ 50 years of age
Laboratory blood tests	Complete blood cell count (WBC, RBC, and platelet count, hemoglobin, hematocrit); Coagulation profiles (PT, PT-INR, aPTT); Diabetes profiles (fasting glucose, Hemoglobin A1c); Liver function profiles (protein/albumin, AST/ALT/Gamma-GTP, total bilirubin); Kidney function profiles (BUN, creatinine); Electrolytes (calcium, phosphate, sodium, potassium, chloride, total CO_2); Lipid profiles (total cholesterol, triglycerides, LDL/HDL cholesterol levels); Uric acid
Serum infections/Inflammation markers	HBsAg, Anti-HBs, Anti-HCV, *H. pylori* IgG Ab, VDRL (Rapid Plasma Reagin Card Test) test, HIV Ag/Ab combo test; Hs-CRP; HAV Ab IgG in persons < 50 years of age
Hormones	Thyroid function tests (TSH, T3, total-T4, Free-T4)
Tumor markers	AFP, CA19-9, CEA, CA-125 (female); PSA for males ≥ 35 years of age
Urine tests	Protein, glucose, ketone, bilirubin, blood, and pH
Stool examinations	Occult blood in stool, parasite assay
Other tests	Chest X-ray (digital imaging); Pulmonary function test; Electrocardiography; Gastroendoscopy with pathological reports; Pap smear and gynecological examination for females ≥ 35 years of age;

	Mammography for females ≥ 35 years of age

WBC, white blood cell; RBC, red blood cell; PT, prothrombin time; INR, International Normalized Ratio; aPTT, activated partial thromboplastin time; AST, aspartate transaminase; ALT, alanine transaminase; GTP, gamma-glutamyl transferase; BUN, blood urea nitrogen; LDL, low-density lipid; HDL, high-density lipid; HBsAg, hepatitis B surface antigen; Anti-HBs, anti-hepatitis B surface antibody; Anti-HCV, anti-hepatitis C antibody; *H. pylori* IgG Ab, hepatitis A immunoglobulin G antibody; VDRL, venereal disease research laboratory; HIV, human immunodeficiency virus, Hs-CRP, high sensitivity C-reactive protein; TSH, thyroid-stimulating hormone; AFP, alpha-fetoprotein; CA19-9, cancer antigen 19-9; CEA, carcinoembryonic antigen; CA-125, cancer antigen 125; PSA, prostate-specific antigen.

Table 2. Specific variables* collected at baseline and follow-up in the Health and Prevention Enhancement (H-PEACE) study

Item	List
SNP-specific	SNP variation by genome-wide SNP arrays (Affymetrix platform of Axiom™ Customized Genome-Wide Human Assay) †
Heart-specific	Coronary calcium score CT; Echocardiography, treadmill test, coronary CT angiography; Holter monitoring
Stroke-specific	Lp (a) lipoprotein (a), and homocysteine; Echocardiography, carotid Doppler; Brain MRI, brain and carotid MRA
Kidney-specific	24-h urinary uric acid, creatinine, urea nitrogen, protein, and microalbumin; 24-h urinary electrolytes including sodium, potassium, calcium, phosphorus, and magnesium; 24-h urinary citrate and oxalate; Serum cystatin C; Dynamic kidney CT
Intestine-specific	Abdominal sonography; Colonoscopy with pathologic reports; Abdomino-pelvic CT
Obesity-specific	24-h recall diet questionnaire; Serum insulin levels; Abdomen and thigh CT for measuring visceral fat; Whole body DEXA
Dementia-specific	Apo-E genotyping; Brain MRI, brain MRA, hippocampus (non-contrast);

	MR double time
Prostate-specific	Free PSA; Uroflowmetry; Ultrasonography for measuring residual urine; Transrectal sonography, prostate MRI (non-contrast)
Sex-specific hormones	Testosterone for males ≥ 60 years of age; Total and free E2, LH, FSH for postmenopausal females after 5 years from menopause
Cigarette smoking-specific	CO levels in exhalation, urine cotinine levels; Laryngoscopy, low-dose screening chest CT
Bone-specific	Bone densitometry for males ≥ 50 years of age and postmenopausal females; Site-specific X-ray digital imaging (AP and Lat.), site-specific bone CT
Dental-specific	Dental, peridental and periodontal examination
Nutrition-specific	Serum markers of selenium, zinc, copper; Plasma vitamin B6 profile (pyridoxal phosphate (PLP), pyridoxic acid (PA)); Plasma vitamin A and E (HPLC); Plasma 25(OH)D; Vitamin C, B12, and folate levels
Allergy-specific	*MAST* (Multiple *allergen* simultaneous *test*); IgE (PRIST); Skin prick test (inhalant); Methacholine bronchial c*hallenge test;* Histograms of Categorized Shapes (HCS) ear detection; *X-ray of PNS (paranasal* sinus)
Hepatitis-specific	HBV viral load, HBeAg (quantitation), HBeAb;

	HCV PCR; Liver fibrosis scan
Heavy mineral-specific	Cd, Pb, Hg, As, Al, Se, Cu, Zn, Cr, Co, Mn, Mo
Other	HPV genotyping; CLO test or UBT test for current *H. pylori* infection

*The specific variables were selected for the health screenees according to the specific test or physician's request.

†GWAS genotyping was performed among the 17 455 cohort participants who provided informed consent and donated blood samples. We categorized the 17 455 participants with blood specimens as the Gene-environmental interaction and phenotype (GENIE) study (a specific subcohort derived from the H-PEACE study). SNP testing was performed for research purposes and was not a routinely performed test.

SNP, single nucleotide polymorphism; CT, computed tomography; MRI, magnetic resonance imaging; MRA, magnetic resonance angiogram; DEXA, dual-energy X-ray absorptiometry; PSA, prostate-specific antigen; LH, luteinizing hormone; FSH, follicle-stimulating hormone; HPLC, high-performance liquid chromatography; PRIST, paper radioimmunosorbent test, PCR, polymerase chain reaction; CLO, Campylobacter-like organism; UBT, urea breath test.

Table 3. Participation rate for the core and specific variables collected in the Health and Prevention Enhancement (H-PEACE) study at baseline

Variables	Participants, N	Participation rate, %*
Core variables		
Most core variables	≥ 90 377	≥ 99
e.g., Electrocardiography	91 197	99.8
Estimated GFR	91 197	99.8
Stool parasite examination	82 493	90.3
H. pylori IgG Ab	81 142	88.8
Thyroid function test	77 313	84.6
CLO test or UBT test for current *H. pylori* infection	67 986	74.4
Gastroendoscopy with pathologic report	66 451	72.8
Specific variables		
Dental examination	43 232	47.3
Colonoscopy with pathologic reports	46 050	50.4
Abdomen CT with visceral fat CT	34 771	38.1
Abdomen sonography	33 446	36.6
Echocardiography	20 688	22.7
Coronary CT	12 846	14.1
Echocardiography, Treadmill test	9910	10.9
SNP from GWAS	6579	7.2
Bone densitometry (postmenopausal females)	26 376	99.4
E2 levels (postmenopausal females after 5 years from menopause)	4742	49.9
Bone densitometry (males ≥ 50 years)	9488	48.9
T levels (males ≥ 60 years)	6697	52.9
HPV genotyping (females)	12 999	31.8

Pap smear (females ≥ 35 years)	12 951	49.1
Blood collection for the Gene-environmental interaction and phenotype (GENIE) study	17 455	19.1
SNP variations in the GWAS	6579	37.7[†]

[*]Participation rate of all 91 336 H-PEACE cohort members.

[†]Participation rate of all 17 455 GENIE study members (a subcohort of the H-PEACE study).

GFR, glomerular filtration rate; CLO, Campylobacter-like organism; UBT, urea breath test; CT, computed tomography; GWAS, genome-wide association study; SNP, single nucleotide polymorphism.

Structured questionnaires

Self-reported questionnaires were used to obtain socio-demographic data, personal and familial medical history, health-related behaviors (such as smoking status, alcohol consumption, physical activity, dietary behavior), international prostate symptom score (IPSS), female reproductive factors, and psychological status. Physical activity levels were assessed using the validated Korean version of the International Physical Activity Questionnaire (IPAQ) short form and were classified into three categories: inactive, minimally active and health-enhancing physically active (HEPA).[20,21] Dietary habits were evaluated using the mini-dietary assessment index (MDAI), which was validated in Korean.[22] Psychological status was assessed by the Beck Depression Inventory (BDI) or Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR 16).[23]

Physical examinations

Blood pressure was measured using sphygmomanometers with patients in a seated position after a resting period. If systolic blood pressure was ≥140 mm Hg or diastolic blood

pressure was ≥90 mm Hg after a rest period and two measurements, we recorded the values and calculated their averages. Height (cm), weight (kg), waist circumference (cm) and body fat composition (%) were measured by trained nurses with participants wearing a lightweight hospital gown and in bare feet. Height and weight were measured using digital scales in a standing position. Waist circumference was obtained by measuring the smallest area below the rib cage above the umbilicus using a non-stretch tape measure, without any pressure to the body surface during measurements. Percentage of body fat and visceral fat area were estimated with a multi-frequency bio-impedance analyzer with 8-point tactile electrodes (Inbody 720, Biospace co, Seoul, Korea). Comprehensive eye examinations (visual acuity, ocular tonometry, slit lamp test, fundus photography) and hearing tests were also performed.

Laboratory tests

Blood samples from the antecubital vein were collected after at least ten hours of fasting. The blood parameters assessed included complete blood cell count, fasting blood glucose, glycated hemoglobin (HbA1c), uric acid, blood urea nitrogen, creatinine, total calcium, inorganic phosphorus, glucose, sodium, potassium, chloride, total CO2, total protein, albumin, total bilirubin, AST, ALT, gamma-GTP, total cholesterol, low-density lipid cholesterol (LDL-C), high-density lipid cholesterol (HDL-C), triglycerides, high sensitivity C-reactive protein (hs-CRP) concentration, prothrombin time, aPTT, HBsAg, anti-HBs, HAV Ab IgG, AFP, CA19-9, CEA and PSA in men and CA 125 in women. The laboratory medicine department at the SNUH has been certified by the Korean Society of Laboratory Medicine and participated in the College of American Pathologist's Survey/Proficiency Testing program. Urine tests were performed by stick test using spot urine. Semi-quantitative variables included pH, protein, glucose, ketone, bilirubin and blood. Stool samples were collected to conduct fecal occult blood tests and parasite assays.

Digital imaging and specific diagnostic variables

Chest X-ray was included as a core variable of the H-PEACE study, which included specific information on diagnostic variables (Table 1). First, the GENIE study collected donated blood samples from 17 455 recipients, and genome-wide SNP arrays from 6579 donated blood samples have already been analyzed using the Affymetrix platform (Axiom™ Customized Genome-Wide Human Assay) (Table 3). We plan to increase the available genetic information by 2020 using an SNP array. The PLINK program (Ver. 1.9) and R statistics (Ver. 3.3.0) were used for quality control procedures and data analysis. SNP genotype data combined with clinical data from the H-PEACE study were used to evaluate gene-environment interactions and to define the related phenotypes.

In total, 12,846 participants underwent coronary CT with a coronary calcium score to assess coronary calcification. The calcium score of the coronary artery is a strong predictor of myocardial infarction and sudden cardiac death.[24] In addition, 34,771 participants underwent visceral fat CT to measure visceral adipose tissue (VAT) and subcutaneous adipose tissue (SAT). The detailed methods used to measure VAT area and SAT area on abdominal fat CT images have been described elsewhere.[17] This quantitative assessment of intra-abdominal adipose tissue is considered the gold standard for measuring the amount of visceral fat.[25] With advanced imaging techniques including echocardiography, brain MRI/MRA, carotid Doppler ultrasound, abdominal ultrasonography and abdominal CT, we elucidated the correlation of visceral obesity with vascular disease using both data in a complementary manner. The core variables of the H-PEACE study including the blood test results, questionnaire findings, and depression scores can contribute to determining the correlation between visceral obesity and metabolic phenotype.

We collected electrocardiogram (ECG) reports from 91,197 consecutive recipients to

evaluate the incidence and risk factors of atrial fibrillation, a significant risk factor for stroke. The effects of a bundle branch block or atrioventricular block on stroke and cardiac disease were also analyzed. Furthermore, we collected Holter monitor results to confirm the ECG reports.

Human papillomavirus (HPV) is known to lead to cervical cancer in women.[26] Of the participants in the H-PEACE study, 12,951 women underwent both a liquid-based cervical cytology (SurePath LBC, Becton Dickinson, Franklin Lakes, NJ, USA) and an HPV genotyping test using an HPV DNA chip (MyHPV Chip, Biomedlab Co., Seoul, Korea) for cervical cancer screening.[27] We also included the results of gynecologic sonography, as well as VDRL and HIV tests.

The World Health Organization (WHO) considers *Helicobacter pylori* infection a class I carcinogen for gastric cancer.[13] The diagnosis of *H. pylori* infection was based on the detection of serum *H. pylori* immunoglobulin G antibody using a kit (*H. pylori*-EIA-Well, Radim, Rome, Italy) that was previously validated in a nationwide Korean sero-epidemiologic study.[28] Furthermore, 66,451 recipients had available upper endoscopy data with pathology results. Serum pepsinogen data and eradication history were also collected.

A total of 46,050 participants in the H-PEACE study received a colonoscopy in our center. All colonoscopies were conducted by board certified endoscopists, and the average adenoma detection rate in recipients 50-70 years old was over 30%.[12] Endoscopes including the CF-H260 and CF-HQ290 series (Olympus, AIZU, Japan) and the EC-450HL5, EC-450WM5, and EC-590ZW series (Fujinon, Saitama, Japan) were used. Histological diagnoses at our center were determined according to the WHO classification of tumors of the digestive system.[29,30] All colonoscopy results and corresponding pathology reports were collected.

We calculated estimated glomerular filtration rate (GFR) using the chronic kidney

disease epidemiology collaboration equation and the modification of diet in renal disease formula. We collected GFR data successively to identify the risk factors for chronic kidney disease. We also included spot urine sodium and creatinine ratio and 24-hour urine data including GFR.

Dental exams were conducted in 43,232 participants. Periodontitis is highly prevalent among adults and is one of the most common causes of teeth loss after 40 years.[31-33] Furthermore, periodontitis can induce systemic inflammation, as well as masticatory dysfunction and poor nutritional status.[32] This is the first cohort study to include a large number of dental exams to evaluate the progression of periodontitis. In this study, 43,232 participants received a dental exam; 33.1% had mild to moderate periodontitis, and 2.1% had severe periodontitis. We also included participants' colonoscopy and coronary CT results, including coronary calcification and colon polyps, to assess the correlation between periodontitis and systemic inflammation.

The presence of osteoporosis was determined using bone densitometry in 26,207 women and 9488 men. We measured bone mineral density (BMD) of the lumbar spine (L1-L4) and femur using dual-energy X-ray absorptiometry (GE Medical, United Kingdom). Based on participants' lowest T scores, normal BMD, osteopenia, and osteoporosis were defined as T scores of −1.0 SD and above, −1.0 to -2.5 SDs, and -2.5 SDs and below, respectively. In this cohort, 26.6% of the participants who underwent bone densitometry were male. Of those, osteopenia was found in 24.75% and osteoporosis was found in 3.21%. The information collected in this cohort also included hormone levels, estrogen levels in women and testosterone levels in men.

FINDINGS TO DATE

The baseline characteristics are summarized in Table 4, and the morbidity rate of the major outcomes is shown in Table 5. The morbidity rates were calculated as the number of findings divided by the number of participants tested. The prevalence of major outcomes (mortality and cancer incidence rate) is provided in Table 6. The H-PEACE study has contributed to the research community by publishing >200 articles since the initiation of the data warehouse, HEALTH-WATCH®, as a prototype in 2003.[8] Our studies have primarily focused on cancer screening and identifying the risk factors and prognosis of metabolic disease.

Table 4. Baseline characteristics of the participants in the Health and Prevention Enhancement (H-PEACE) study

	Total	Men	Women	p-value
N of participants	91 336	50 507	40 829	
N of genome-wide association study (GWAS) data	6579	4011	2568	
	N (%)	**N (%)**	**N (%)**	
Age (years)				<0.001
< 35	16228 (17.8)	7666 (15.2)	8562 (21.0)	
35 – 44	26963 (29.5)	15139 (30.0)	11824 (29.0)	
45 – 54	27915 (30.6)	15776 (31.2)	12139 (29.7)	
55– 64	14937 (16.4)	8717 (17.3)	6220 (15.2)	
65 +	5293 (5.8)	3209 (6.4)	2084 (5.1)	
Sex				
Men	55.3			
Women	44.7			

Questionnaires				
Smoking status				<0.001
Non-smokers	48 521 (53.1)	11 779 (23.3)	36 742 (90.0)	
Ex-smokers	22 564 (24.7)	20 469 (40.5)	2095 (5.1)	
Current smokers	20 251 (22.2)	18 259 (36.2)	1992 (4.9)	
Drinking status				<0.001
Excess alcohol (>140 g/week)	20 421 (22.3)	18 934 (37.5)	1469(3.6)	
Regular exercise				<0.001
Inactive	36 343 (39.8)	17 843 (35.3)	18 500 (45.3)	
Minimally active	26 189 (28.7)	16 428 (32.5)	9761 (23.9)	
HEPA	28 804 (31.5)	16 236 (32.1)	12 568 (30.8)	
Education				<0.001
High school graduate	78 821 (86.3)	45 538 (90.2)	33 283 (81.5)	
Comorbidity				
Past hypertension	13 279 (14.5)	9003 (17.8)	4276 (10.5)	<0.001
Past diabetes	4377 (4.8)	3181 (6.3)	1196 (2.9)	<0.001
Current HBsAg	3810 (4.2)	2377 (4.7)	1433 (3.5)	<0.001
Current Anti-HCV	887 (0.97)	492 (0.97)	395 (0.97)	0.946
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	
Anthropometric measurements				
Height (cm),	165.6 (11)	171.7 (5.6)	159.9 (5.2)	<0.001
Weight (kg),	64.3 (12.1)	71.5 (9.1)	55.0 (7.2)	<0.001
Body mass index (kg/m^2)	23.4 (12.1)	24.2 (2.6)	21.5 (2.8)	<0.001
Waist circumference (cm)	83.7 (9.3)	86.5 (7.1)	76.9 (7.5)	<0.001
Skeletal muscle mass (kg)	26.6 (6.0)	30.8 (3.7)	20.6 (2.4)	<0.001
Body fat mass (kg)	26.9 (5.9)	30.8 (3.6)	20.7 (2.3)	

Body fat percent (%)	25.6 (6.1)	23.0 (4.8)	29.5 (5.7)	
Visceral fat area (cm^2)*	79.9 (24.2)	82.3 (23.6)	76.1 (24.7)	<0.001
Systolic blood pressure (mm Hg)	115.4 (15.9)	118.0 (12.2)	111.2 (13.3)	<0.001
Diastolic blood pressure (mm Hg)	76.01 (11.9)	79.0 (9.45)	71.36 (9.56)	<0.001
Clinical examination				
Fasting blood glucose (mg/dL)	97 (18.6)	100.2 (17.6)	90.7 (16.2)	<0.001
HbA1c (%)	5.56 (0.6)	5.6 (0.6)	5.4 (0.4)	<0.001
Creatinine (mg/dL)	0.84 (0.17)	0.93 (0.13)	0.68 (0.10)	<0.001
Albumin (g/dL)	4.62 (0.25)	4.64 (0.24)	4.58 (0.25)	0.002
Total cholesterol (mg/dL)	191.03 (33.02)	192.5 (33.1)	188.6 (32.8)	0.105
Triglycerides (mg/dL)	101.2 (64.2)	117.9 (68.8)	74.6 (44.9)	<0.001
LDL cholesterol (mg/dL)	119.8 (31.6)	126.3 (31.3)	112.6 (30.0)	<0.001
HDL cholesterol (mg/dL)	54.3 (13.2)	49.4 (9.5)	59.7 (12.6)	<0.001
Hs-CRP (mg/dL)	0.12 (0.26)	0.12 (0.26)	0.10 (0.26)	0.285

*Visceral fat area is defined here as the cross-sectional area of visceral fat found in the abdomen.

HEPA, health-enhancing physically active; HBsAg, hepatitis B surface antigen; Anti-HCV, anti-hepatitis C antibody; LDL, low-density lipid; HDL, high-density lipid; Hs-CRP, high sensitivity C-reactive protein.

Table 5. Abnormal health findings detected through active follow-up with specific measurements in the Health and Prevention Enhancement (H-PEACE) study

	Definition	Participants tested	Follow-up years (median (IQR))	N of finding*
Coronary calcification	Coronary artery calcium score > 400 at coronary CT	12 846	5.6 (3.2-8.0)	805
Atrial fibrillation	ECG and/or ambulatory ECG	91 076	4.1 (2.1-6.5)	446
Left bundle branch block	ECG and/or ambulatory ECG	91 076	4.1 (2.1-6.5)	64
Chronic kidney disease	Estimated GFR† < 60 mL/min/1.73 m^2	91 197	4.0 (2.1-6.5)	5372
Visceral obesity	Men > 10000 cm^2, Women > 8000 cm^2 in abdomen visceral fat CT	34 771	5.0 (3.0-7.4)	23 126
Fatty liver	Abdominal sonography and/or abdominal CT	34 771	5.0 (3.0-7.4)	Mild 10 244 Moderate 5963 Severe 688
Abnormal thyroid function	Thyroid hormone measurement	77 313	4.1 (2.2-6.6)	Hypothyroidism 363 Hyperthyroidism 1260

H. Pylori infection	*H. pylori* IgG Ab and/or CLO test or UBT test for current *H. pylori* infection	81 142	4.2 (2.3-6.7)	44 304
High-risk HPV infection	HPV 16, 18 positive finding in HPV genotyping	12 999	5.0 (3.1-7.2)	1462
Parasite infection	Stool examination	82 493	4.1 (2.2-6.6)	*Clonorchis sinensis* 1262 *Ascaris lumbricoides* 124 *Trichuris trichiura* 256 *Metagonimus yokogawai* 412
Premalignant conditions of stomach	Gastroendoscopy with pathological reports	66 451	4.2 (2.3-6.7)	Atrophic gastritis 23 896 Intestinal metaplasia 10 187
Malignant gastric cancer	Gastroendoscopy with pathological reports	66 451	4.2 (2.3-6.7)	299
Squamous intraepithelial lesion (SIL)	Pap smear and pathological report	12 951	5.0 (3.1-7.2)	Total SIL 375 Low grade 315 High grade 60
Premalignant colonic neoplasms	Colonoscopy with pathological reports	46 746	4.83 (2.89-7.13)	Adenoma 18 656‡ Serrated polyp 1178‡ Carcinoid 150

Malignant colon cancer	Colonoscopy with pathological reports	46 746	4.83 (2.89-7.13)	262
Low bone density	Bone densitometry	35 695	4.86 (2.83-7.18)	Osteopenia 10 601 Osteoporosis 1653
Periodontitis	Dental examination	43 232	4.9 (2.8-7.2)	Total periodontitis 15 204 Mild to moderate 14 292 Severe 912

*The number of findings among the participants who were at risk.

†The glomerular filtration rate was estimated by the chronic kidney disease epidemiology collaboration (CKD-EPI) equation and the modification of diet in renal disease (MDRD) formula.

‡Adenoma and serrated polyp detection rate were calculated in participants 50-75 years old

ECG: electrocardiogram; GFR, glomerular filtration rate; CT, computed tomography; CLO, Campylobacter-like organism; UBT, urea breath test; HPV, human papillomavirus.

Table 6. Major cancer incidence and outcomes based on record linkage to the Nationwide Death Certificate Database in the Health and Prevention Enhancement (H-PEACE) study*

	N of cancer incidence			N of all-cause death†		
	Total	**Men**	**Women**	**Total**	**Men**	**Women**
Total cancer	2814	1402	1412	207	150	57
Thyroid	956	390	566	3	3	0
Stomach	513	337	176	37	29	8
Breast	343	1	342	6	0	6
Colorectal	284	182	102	28	20	8
Lung	186	109	77	45	32	13
Prostate	127	127	0	3	3	0
Kidney	91	71	20	1	1	0
Liver	87	71	16	31	28	3
Gynecology	52	0	52	7	0	7
Pancreas	42	25	17	27	19	8
Brain	39	20	19	0	0	0
Hematologic	39	24	15	5	3	2
Biliary tract	32	26	6	11	9	2
Esophagus	23	19	4	3	3	0

*Median (interquartile range) follow-up years of the 91 336 participants=4.04 (2.13-6.5).

†Since the Nationwide Death Certificate Database does not provide information about cause of death, N of death includes death from not only cancer but also other causes. Among the 91 336 H-PEACE participants, the all-cause mortality rate was 969: 715 men, 254 women.

Regarding colon cancer screening, we demonstrated that colonoscopic surveillance

should be targeted to high-risk patients, and we recommended a 3-yr follow-up after initial polypectomy.[12] This finding has been reflected in the guidelines for colonoscopic surveillance.[34] We also examined the quality of colon cancer screening using metrics such as the adenoma detection rate.[35] Additionally, we evaluated the influence of image-enhanced endoscopy, such as narrow band imaging and Fuginon Intelligent Color Enhancement, using our cohort data.[10,36,37]

For gastric cancer screening, the effect of screening endoscopy remains controversial, but population-based screening has been undertaken in Korea and Japan. We demonstrated that annual screening endoscopy improved the detection rate of early-stage and endoscopically treatable gastric cancer and further showed that high-risk populations with intestinal metaplasia should receive intensive screening.[11] Gastric mucosa-associated lymphoid tissue lymphomas (MALToma) and suspicious MALToma lesions were also detected in screening endoscopies, and the prevalence of gastric MALToma was high in middle-aged women.[38] Active endoscopic screening for gastric cancer had the additional benefit of detecting early-stage MALToma.[38,39] Furthermore, we established that small subepithelial tumors that were incidentally found during upper endoscopy screening showed no size change in subsequent follow-up periods.[36]

Our research focus was not limited to cancer screening. We also examined metabolic diseases such as non-alcoholic fatty liver disease (NAFLD) and metabolic syndrome. We reported that metabolic syndrome was an independent risk factor of silent brain infarction using brain MRI data.[40] We also found an effect of body fat distribution on the incidence of NAFLD and reflux esophagitis using abdominal fat CT data.[9,14] We also identified a relation between NAFLD and the risk of coronary heart disease and arterial stiffness based on coronary CT and cardio-ankle vascular index, respectively.[16-19] The effect of physical activity on NAFLD was also demonstrated using data from the questionnaires.[41]

The list and number of publications can be requested on our homepage (http://healthcare.snuh.org/HPEACEstudy).

STRENGTHS AND LIMITATIONS

A major strength of the H-PEACE study is its structured and organized database, which not only includes data widely used in medical check-ups but also data from sophisticated high-quality advanced examinations, such as colonoscopy, brain MRI, abdominal CT and abdominal fat CT. Particularly in our study, comprehensive biomarkers related to NCDs, including cancer antigens and infection markers, were included, which have not been included in the medical check-ups of other cohort profiles. As we obtained a large sample size, we could confirm disease statuses as well as numerous phenotypes of NCDs. Furthermore, the study protocols were conducted and monitored with intensive quality control procedures. Additional strengths include the active and passive follow-ups and the ability to obtain complete data, covering deaths and incidental cancer cases even among those who discontinued visiting our center. We believe that these data could contribute to active, effective prevention against the development of cancers and NCDs. This study also has potential limitations. The participation was restricted to individuals who visited our center. In addition, information could be missing because data collection for the repeated measurements relied on voluntary visits to our center, and this procedure may have introduced selection bias.

COLLABORATION

The genotype (SNP) data from the GENIE study and the epidemiologic data from the other cohorts are available to researchers after a quality control process has been completed. Potential collaborators can access the dataset after receiving approval from the SNUH Gangnam Center institutional review board. Applications can be submitted via our homepage. Further information is available at the H-PEACE study website (http://healthcare.snuh.org/HPEACEstudy).

ACKNOWLEDGEMENTS

The authors thank You Mi Kwon and Si Won Choi for their technical assistance.

COMPETING INTERESTS

J.E.L is an employee of DNALink, Inc.

FUNDING

This work was supported by the SK Telecom Research Fund through Seoul National University Hospital [34-2013-0390 and 34-2014-0210 to S.J.C.] and the Seoul National University Hospital Research Fund [04-2011-0700 to S.J.C.].

CONTRIBUTORS

CL, EKC, JMC, SKP and SC were involved in the conception and design of the study. CL, EKC, JMC, YH, YL, and BP contributed to development of methods and data collection. CL, EKC, JMC, YL, and BP were involved in data analysis and interpretation. CL, EKC, JMC, SKP and SC drafted the manuscript. All the authors have critically revised the article and approved the final version and the findings.

ETHICS APPROVAL

The H-PEACE study was approved by the ethics committee of Seoul National University Hospital.

DATA SHARING STATEMENT

Researchers can apply for the dataset after receiving approval from the SNUH Gangnam Center institutional review board. The H-PEACE study website provides information on the application process (http://healthcare.snuh.org/HPEACEstudy).

REFERENCES

1 Beaglehole R, Bonita R, Horton R, et al. Priority actions for the non-communicable disease crisis. *Lancet* 2011;377:1438-47.

2 Ministry of Health and Welfare of Korea, Korea Centers for Disease Control and Prevention. *2010 Korea Health statistics.* Seoul: Ministry of Health and Welfare of Korea, 2011.

3 Hunter DJ. Gene-environment interactions in human diseases. *Nat Rev Genet* 2005;6:287-98.

4 Renz H, von Mutius E, Brandtzaeg P, et al. Gene-environment interactions in chronic inflammatory disease. *Nat Immunol* 2011;12:273-7.

5 Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015;372:793-5.

6 Jameson JL, Longo DL. Precision medicine — personalized, problematic, and promising. *N Engl J Med* 2015;372:2229-34.

7 Elliott P, Peakman TC, UK Biobank. The UK biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol* 2008;37:234-44.

8 Park MJ, Kim DH, Lim SH, et al. Features of gastric neoplasm detected during the screening examination. *Gut Liver* 2007;1:33-9.

9 Kim D, Chung GE, Kwak MS, et al. Body fat distribution and risk of incident and regressed nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2016;14:132–8.e4.

10 Chung SJ, Kim D, Song JH, et al. Comparison of detection and miss rates of narrow band imaging, flexible spectral imaging chromoendoscopy and white light at screening colonoscopy: a randomised controlled back-to-back study. *Gut*

2014;63:785-91.

11 Chung SJ, Park MJ, Kang SJ, et al. Effect of annual endoscopic screening on clinicopathologic characteristics and treatment modality of gastric cancer in a high-incidence region of Korea. *Int J Cancer* 2012;131:2376-84.

12 Chung SJ, Kim YS, Yang SY, et al. Five-year risk for advanced colorectal neoplasia after initial colonoscopy according to the baseline risk stratification: a prospective study in 2452 asymptomatic Koreans. *Gut* 2011;60:1537-43.

13 Chung SJ, Lim SH, Choi J, et al. Helicobacter pylori serology inversely correlated with the risk and severity of reflux esophagitis in helicobacter pylori endemic area: a matched case-control study of 5,616 health check-up Koreans. *J Neurogastroenterol Motil* 2011;17:267-73.

14 Chung SJ, Kim D, Park MJ, et al. Metabolic syndrome and visceral obesity as risk factors for reflux oesophagitis: a cross-sectional case-control study of 7078 Koreans undergoing health check-ups. *Gut* 2008;57:1360-5.

15 Chung GE, Kim D, Kim W, et al. Non-alcoholic fatty liver disease across the spectrum of hypothyroidism. *J Hepatol* 2012;57:150-6.

16 Chung GE, Choi SY, Kim D, et al. Nonalcoholic fatty liver disease as a risk factor of arterial stiffness measured by the cardioankle vascular index. *Medicine (Baltimore)* 2015;94:e654.

17 Kim D, Choi SY, Park EH, et al. Nonalcoholic fatty liver disease is associated with coronary artery calcification. *Hepatology* 2012;56:605-13.

18 Kwak MS, Yim JY, Kim D, et al. Nonalcoholic fatty liver disease is associated with coronary artery calcium score in diabetes patients with higher HbA1c. *Diabetol Metab Syndr* 2015;7:28.

19 Choi SY, Kim D, Kim HJ, et al. The relation between non-alcoholic fatty liver disease

and the risk of coronary heart disease in Koreans. *Am J Gastroenterol* 2009;104:1953-60.

20 Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* 2003;35:1381-95.

21 Chun MY. Validity and reliability of Korean version of international physical activity questionnaire short form in the elderly. *Korean J Fam Med* 2012;33:144-51.

22 Kim WY, Cho MS, Lee HS. Development and validation of mini dietary assessment index for Koreans. Korean. *J Nutr* 2003;36:83-92.

23 Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. *Clin Psychol Rev* 1988;8:77-100.

24 Wong ND, Hsu JC, Detrano RC, et al. Coronary artery calcium evaluation by electron beam computed tomography and its relation to new cardiovascular events. *Am J Cardiol* 2000;86:495-8.

25 Seidell JC, Bakker CJ, van der Kooy K. Imaging techniques for measuring adipose-tissue distribution--a comparison between computed tomography and 1.5-T magnetic resonance. *Am J Clin Nutr* 1990;51:953-7.

26 Bosch FX, Manos MM, Muñoz N, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. *J Natl Cancer Inst* 1995;87:796-802.

27 Kim MJ, Kim JJ, Kim S. Type-specific prevalence of high-risk human papillomavirus by cervical cytology and age: data from the health check-ups of 7,014 Korean women. *Obstet Gynecol Sci* 2013;56:110-20.

28 Yim JY, Kim N, Choi SH, et al. Seroprevalence of Helicobacter pylori in South Korea. *Helicobacter* 2007;12:333-40.

29 Hamilton SR, Aaltonen LA, 2000. *WHO Classification of Tumours of the Digestive System*. 3rd ed. Geneva: IARC.

30 Bosman FT, Carneiro F, Hruban RH, et al., 2010. *WHO Classification of Tumours of the Digestive System*. 4th ed. Geneva: IARC.

31 Eke PI, Dye BA, Wei L, et al. Update on prevalence of periodontitis in adults in the United States: NHANES 2009 to 2012. *J Periodontol* 2015;86:611-22.

32 White DA, Tsakos G, Pitts NB, et al. Adult Dental Health survey 2009: Common oral health conditions and their impact on the population. *Br Dent J* 2012;213:567-72.

33 Murray H, Locker D, Kay EJ. Patterns of and reasons for tooth extractions in general dental practice in Ontario, Canada. *Commun Dent Oral Epidemiol* 1996;24:196-200.

34 Lieberman DA, Rex DK, Winawer SJ, et al. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology* 2012;143:844-57.

35 Kang HY, Kim D, Kim HJ, et al. The relationship between colonoscopy procedure order and adenoma detection rates: a prospective study. *J Clin Gastroenterol* 2015;49:683-9.

36 Song JH, Kim SG, Chung SJ, et al. Risk of progression for incidental small subepithelial tumors in the upper gastrointestinal tract. *Endoscopy* 2015;47:675-9.

37 Chung SJ, Kim D, Song JH, et al. Efficacy of computed virtual chromoendoscopy on colorectal cancer screening: a prospective, randomized, back-to-back trial of Fuji intelligent color enhancement versus conventional colonoscopy to compare adenoma miss rates. *Gastrointest Endosc* 2010;72:136-42.

38 Yang HJ, Lee C, Lim SH, et al. Clinical characteristics of primary gastric lymphoma detected during screening for gastric cancer in Korea. *J Gastroenterol Hepatol* 2016;31:1572-83.

39 Yang HJ, Lim SH, Lee C, et al. Management of suspicious mucosa-associated lymphoid tissue lymphoma in gastric biopsy specimens obtained during screening endoscopy. *J Korean Med Sci* 2016;31:1075-81.

40 Kwon HM, Kim BJ, Lee SH, et al. Metabolic syndrome as an independent risk factor of silent brain infarction in healthy people. *Stroke* 2006;37:466-70.

41 Kwak MS, Kim D, Chung GE, et al. Role of physical activity in nonalcoholic fatty liver disease in terms of visceral obesity and insulin resistance. *Liver Int* 2015;35:944-52.

FIGURE LEGENDS

Fig 1. Study flow diagram of the Health and Prevention Enhancement (H-PEACE) study

Fig 2. Regional distribution of the Health and Prevention Enhancement (H-PEACE) study in Korea

Fig 3. Flow diagram of health check-ups in the Health and Prevention Enhancement (H-PEACE) study

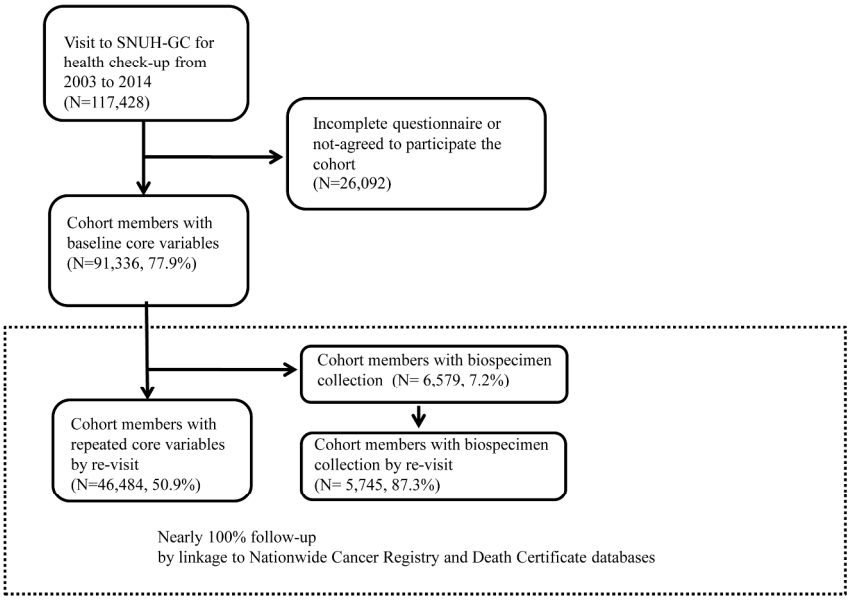
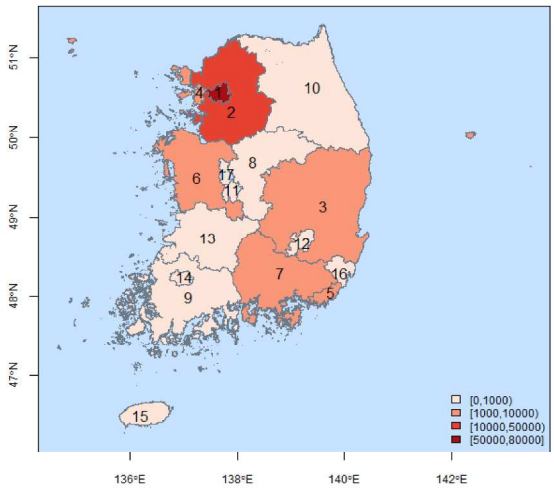


Fig 1. Study flow diagram of the Health and Prevention Enhancement (H-PEACE) study

254x190mm (300 x 300 DPI)



	Location	N
1	Seoul (Capital, Metropolitan)	60,301
2	Gyeonggi-do	18,383
3	Gyeongsangbuk-do	1,564
4	Incheon (Metropolitan)	1,478
5	Busan (Metropolitan)	1,445
6	Chungcheongnam-do	1,400
7	Gyeongsangnam-do	1,278
8	Chungcheongbuk-do	769
9	Jeollanam-do	737
10	Gangwon-do	727
11	Daejeon (Metropolitan)	692
12	Daegu (Metropolitan)	686
13	Jeollabuk-do	580
14	Gwangju (Metropolitan)	488
15	Jeju-do	338
16	Ulsan (Metropolitan)	297
17	Sejong (Metropolitan)	20
	Total	91,183

Fig 2. Regional distribution of the Health and Prevention Enhancement (H-PEACE) study in Korea

254x190mm (300 x 300 DPI)

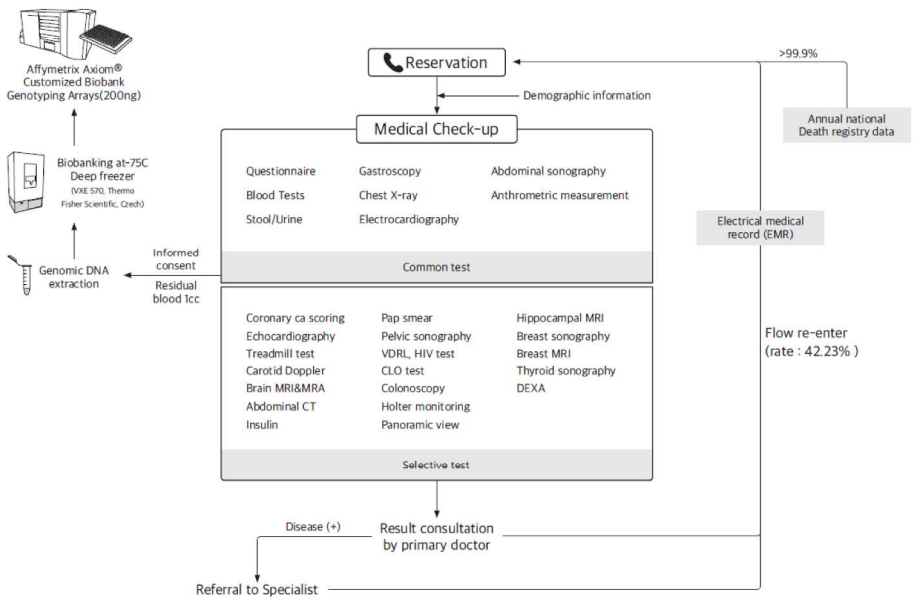


Fig 3. Flow diagram of health check-ups in the Health and Prevention Enhancement (H-PEACE) study

254x190mm (300 x 300 DPI)

STROBE (Strengthening The Reporting of OBservational Studies in Epidemiology) Checklist

A checklist of items that should be included in reports of observational studies. You must report the page number in your manuscript where you consider each of the items listed in this checklist. If you have not included this information, either revise your manuscript accordingly before submitting or note N/A.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Section and Item	Item No.	Recommendation	Reported on Page No.
Title and Abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/Rationale	2	Explain the scientific background and rationale for the investigation being reported	6
Objectives	3	State specific objectives, including any prespecified hypotheses	6-7
Methods			
Study Design	4	Present key elements of study design early in the paper	8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	8
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	8
		Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	N/A
		Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants	N/A
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed	N/A
		Case-control study—For matched studies, give matching criteria and the number of controls per case	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	9-15

Section and Item	Item No.	Recommendation	Reported on Page No.
Data Sources/ Measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	9-15
Bias	9	Describe any efforts to address potential sources of bias	31
Study Size	10	Explain how the study size was arrived at	9
Quantitative Variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	N/A
Statistical Methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	N/A
		(*b*) Describe any methods used to examine subgroups and interactions	N/A
		(*c*) Explain how missing data were addressed	N/A
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed	N/A
		Case-control study—If applicable, explain how matching of cases and controls was addressed	N/A
		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	22-24
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	Fig 1
Descriptive Data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 4
		(b) Indicate number of participants with missing data for each variable of interest	Fig 1
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	8
Outcome Data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	Table 5,6
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	N/A
		Cross-sectional study—Report numbers of outcome events or summary measures	N/A

Section and Item	Item No.	Recommendation	Reported on Page No.
Main Results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	N/A
		(*b*) Report category boundaries when continuous variables were categorized	Table 4,5
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other Analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	N/A
Discussion			
Key Results	18	Summarise key results with reference to study objectives	31
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	31
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	31
Generalisability	21	Discuss the generalisability (external validity) of the study results	31
Other Information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	32

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Once you have completed this checklist, please save a copy and upload it as part of your submission. DO NOT include this checklist as part of the main manuscript document. It must be uploaded as a separate file.

BMJ Open

Cohort profile: The Health and Prevention Enhancement (H-PEACE), a retrospective, population-based cohort study conducted at the Seoul National University Hospital Gangnam Center, Korea

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019327.R1
Article Type:	Cohort profile
Date Submitted by the Author:	22-Dec-2017
Complete List of Authors:	Lee, Changhyun; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Choe, Eun Kyung; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Choi, Ji Min; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Hwang, Yunji; Seoul National University College of Medicine, Department of Preventive Medicine; Seoul National University Graduate School, Department of Biomedical Science Lee, Young; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Park, Boram; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Chung, Su Jin ; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Kwak, Min-Sun ; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Lee, Jong-Eun ; DNA link Kim, Joo Sung ; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Park, Sue K.; Seoul National University College of Medicine, Department of Preventive Medicine; Seoul National University Graduate School, Department of Biomedical Science Cho, Sang-Heon; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Patient-centred medicine
Keywords:	cohort, non-communicable disease, PREVENTIVE MEDICINE, EPIDEMIOLOGY

Cohort profile: The Health and Prevention Enhancement (H-PEACE), a retrospective, population-based cohort study conducted at the Seoul National University Hospital Gangnam Center, Korea

Short title: Cohort profile of the H-PEACE study

Changhyun Lee[1¶], Eun Kyung Choe[1¶], Ji Min Choi[1], Yunji Hwang[2,3,4], Young Lee[1], Boram Park[1], Su Jin Chung[1], Min-Sun Kwak[1], Jong-Eun Lee[5], Joo Sung Kim[1], Sue K. Park[2,3,4&*], Sang-Heon Cho[1&*], and the H-PEACE study investigators

[1]Healthcare Research Institute, Healthcare System Gangnam Center, Seoul National University Hospital, Seoul, Korea

[2]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea

[3]Department of Biomedical Science, Seoul National University Graduate School, Seoul, Korea

[4]Seoul National University Cancer Research Institute, Seoul, Korea

[5]DNA link, Inc., Seoul, Korea

[¶]Changhyun Lee and Eun Kyung Choe contributed equally to this paper.

[&]Sue Kyung Park and Sang-Heon Cho contributed equally to this paper.

***Corresponding authors**

Sang-Heon Cho

Gangnam Finance Center, Teheran-ro 152, Gangnam-gu, Seoul, 06236, Korea

Tel: +82-2-2072-2971; E-mail: shcho@snu.ac.kr

Word count: 3,772

ABSTRACT

Purpose: The Health and Prevention Enhancement (H-PEACE) study was designed to investigate the association of diagnostic imaging results, biomarkers and the pre-disease stage of non-communicable diseases (NCDs), such as malignancies and metabolic diseases, in an average-risk population in Korea.

Participants: This study enrolled a large-scale retrospective cohort at the Healthcare System Gangnam Center, Seoul National University Hospital, from October 2003 to December 2014.

Findings to date: The baseline and follow-up information collected in the pre-disease stage of NCDs allows for evaluation of an individual's potential NCD risk, which is necessary for establishing personalized prevention strategies. A total of 91,336 health examinees were included in the cohort, and we repeatedly measured and collected information for 50.9% (N=46,484) of the cohort members. All participants completed structured questionnaires (lifestyle, medical history, mini-dietary assessment index, sex-specific variables, and psychiatric assessment), doctors' physical examinations, laboratory blood and urine tests and digital chest X-ray imaging. For participants with available data, we also obtained information on specific diagnostic variables using advanced diagnostic tests, including coronary computed tomography (CT) for coronary calcium scores, colonoscopy, and brain magnetic resonance imaging (MRI). Furthermore, 17,455 of the participants who provided informed consent and donated blood samples were enrolled into the Gene-environmental interaction and phenotype (GENIE) study, a subcohort of the H-PEACE, from October 2013 and we analyzed genome-wide single nucleotide polymorphism (SNP) array data for 6,579 of these blood samples.

Future plans: The data obtained from this cohort will be used to facilitate advanced and accurate diagnostic techniques related to NCDs while considering various phenotypes. Potential collaborators can access the dataset after receiving approval from our institutional review board. Applications can be submitted on the study homepage (http://en-

healthcare.snuh.org/HPEACEstudy).

Keywords: epidemiology; cohort; non-communicable disease; preventive medicine

STRENGTHS AND LIMITATIONS OF THIS STUDY

- The strengths of the H-PEACE study include a large number of healthy subjects (N=91,336 healthy examinees, from the Healthcare System Gangnam Center, Seoul National University Hospital, between 2003 and 2014) and a structured and organized database.
- This study not only includes data widely used in medical check-ups but also data from sophisticated high-quality advanced examinations to investigate clinical effectiveness in predicting the pre-disease stage of non-communicable diseases, including malignancies and metabolic diseases, in an average-risk population in Korea.
- Another strength includes the active and passive follow-ups and the ability to obtain complete data, including deaths and incidental cancer cases, even among those who discontinued visiting our center.
- The data from this study will allow us to contribute to active, effective prevention of the development of cancers and non-communicable diseases.
- The major weakness of this cohort is that it may show selection bias because only subjects who voluntarily visited our center were included in the study.

INTRODUCTION

In recent decades, the prevalence of non-communicable diseases (NCDs), such as malignancy, metabolic disease, and cardiovascular disease, has rapidly increased in Korea.[1,2] To address this problem, a comprehensive approach that accounts for lifestyle, environmental factors and genetic variability is needed, as NCDs are known to be caused by both genetic and environmental factors.[3,4] Precision medicine is emerging as a potential solution.[5,6] This type of medicine categorizes individuals into different subgroups based on their susceptibility to disease and then focuses on individuals in whom interventions will be helpful.[7] One of the major components of efficient prevention and early detection of NCDs is thus identifying high-risk populations among those in a pre-disease or asymptomatic stage.

The Seoul National University Hospital (SNUH) Healthcare System Gangnam Center provides comprehensive medical check-ups and screening, and nearly 20,000 people visit this center each year. A data warehouse, HEALTH-WATCH®, was built as a prototype database for this cohort study within the healthcare research institute from the start of the center in 2003.[8] Using HEALTH-WATCH®, we have published many articles that have primarily been focused on cancer screening and metabolic diseases.[8-19] In 2013, we re-organized the cohorts to support our research and started collecting blood samples to analyze the genetic factors involved in NCDs. We then summarized and integrated our clinical and genetic data in the Health and Prevention Enhancement (H-PEACE) study.

The H-PEACE study was designed to investigate the pre-disease stage of NCDs, which helps us assess an individual's risk of NCDs and establish personalized prevention strategies. Through analyses of longitudinal data from comprehensive questionnaires and clinical and laboratory tests, the H-PEACE study can expand our knowledge of NCD prevention, as well as our knowledge of high-risk populations, to improve the early detection of NCDs. The findings can inform treatment priorities and therapeutic guidelines related to NCDs. Finally,

the H-PEACE study can contribute to enhancing public health and improving quality of life.

COHORT DESCRIPTION

Participants

The H-PEACE study collected data from 91,336 Koreans aged 45.5 ± 11.7 years (50,507 men and 40,829 women) who received a health check-up between October 2003 and December 2014 at the SNUH Gangnam Center (IRB No H-1311-031-531). We prospectively collected blood samples from 17 455 of those individuals (9396 men and 8059 women) from October 2013 to form the Gene-environmental interaction and phenotype (GENIE) study, a subcohort of the H-PEACE study. The informed consent and study protocols were approved (IRB No H-1103-127-357 & H-1505-047-671). Furthermore, we obtained data for genome-wide single nucleotide polymorphism (SNP) arrays using 6579 donated blood samples. The flowchart of the H-PEACE study subjects is illustrated in Fig 1, and the distribution of the participants in Korea in Fig 2.

Repeated measurement and active and passive follow-up

In the H-peace study, there are two follow-up systems (passive and active follow-up system) for tracking the development of the new diseases and death of the participants. We have annually updated the survival data by linking to National Death Certificates through requests to Statistics Korea. Under the passive follow-up systems, over 99% of participants are followed up their new cancer development and death annually.

Moreover, the study participants are followed up their new disease status by re-visit to center for repeated measurement and assessment system. This active follow-up system allows the subject to voluntarily visit to our center. The health checkup system in Korea has two systems in which the NHIC is partially paid for participants' basic health examination fee once every two years or in which the health examination participants pays all expenses for their health examination fees. The latter program includes more precise health examination

testing that individuals want, and all costs must be paid by individuals. We are under the latter system and the health examination participants have to pay for all the health checkups under their voluntary visit. The participants taken health examination at our center may do repeated measurement at different centers without revisit to our center under the partial coverage of the NHIC. Although the health check-up services conducted in our center includes a variety of tests that the subject wants, including the basic health check-up program conducted by the NHIC, we have to make a lot of effort in order for the subject to return to our center because the former health check-up system covered by the NHIC is free-paid for the basic health check-up program. We are using a number of ways to encourage our return to our center. To encourage participants, we provide reminder calls every year and send health care information about people's next health check-up date via phone and letter. So the eligibility for follow-up assessment depends on the participants' voluntary re-visit need and self-payment availability.

Under this active follow-up system, we did a repeated measurement of risk factors and an assessment of outcome variables for 46,484 individuals (50.9%) of total 91,336 cohort members participated in the baseline health check-up once within 4 years from the first visit and 74,304 individuals (83.5%) were actively followed up once within 10 years from the first visit. Moreover, 50,049 participants (54.8%) completed the two or three repeated measurements (Fig 1). The median follow-up was 4.04 years (interquartile range [IQR] 2.1 – 6.5).

Of 91,336 cohort members, a total of 17,455 members (19.1%) agreed to blood collection and had several aliquots of blood specimens. Of 17,455 members, 6,579 participants were included in the GENIE study with GWAS information (37.7%) and of them, 5,915 participants (89.9%) completed at least 1st follow-up test within 4 years of the first visit. The median follow-up of the GENIE study was 5.23 years (IQR 2.7 – 8.0).

Data collection

The H-PEACE study consists of core and specific variables (Tables 1 and 2). The core variables were tests performed in all of the enrolled participants at the regular health check-ups. The specific variables were the tests selectively performed in the participants upon request or based on the recommendation of their medical provider according to the participant's symptoms or disease risk factors. These tests were sophisticated and expensive tests that are rarely utilized in other cohort profiles. In Table 3, we describe the representative test for each variable and its participation rate. The tests performed for the covariables and specific variables are described below. All data were collected as part of the comprehensive health check-ups. The flow diagram of health check-ups in our study is illustrated in Fig 3. All study protocols are available upon request at our homepage (http://en-healthcare.snuh.org/HPEACEstudy).

Table 1. Core variables collected at baseline and follow-up in the Health and Prevention Enhancement (H-PEACE) study from 2003 to 2014

Item	List	
Self-record questionnaires	· Demographic and lifestyles questionnaire: socioeconomic variables, including education, household income, marital status, and occupation; past medical history of diabetes, hypertension, dyslipidemia, and cancer; family history of diabetes, hypertension, dyslipidemia, and cancer in first relatives; alcohol consumption and cigarette smoking · Physical activity: Korea-validated version of the International Physical Activity Questionnaire (IPAQ) short form · Dietary factors: Mini-Dietary Assessment Index (MDAI) · Sex-specific questionnaire: International Prostate Symptom score (IPSS) and the International Index of Erectile Function (IIEF-5) for men; reproductive factors for women, including menstrual history (age at menarche, length and regularity of menstrual cycle), menopausal status and age, pregnancy history (including the number/methods of pregnancies and abortions), and breastfeeding · Psychiatric assessment questionnaire: Beck Depression Inventory (BDI) and Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR 16)	
Anthropometric measurements and physical examination	· Systolic and diastolic blood pressure · Height, weight, and waist circumference · Body composition using a multi-frequency bio-impedance analyzer · Physical examination by physician · Eye examination, including visual acuity, ocular tonometry, and slit lamp test by Ophthalmologist · Fundus photography (persons ≥ 35 years old) by Ophthalmologist · Hearing test (persons ≥ 50 years old) by otolaryngologist	
Laboratory blood tests	· Complete blood cell count (WBC, RBC, and platelet count, hemoglobin, hematocrit); · Coagulation profiles (PT, PT-INR, aPTT); · Diabetes profiles (fasting glucose, Hemoglobin A1c); · Liver function profiles (protein/albumin, AST/ALT/Gamma-GTP, total bilirubin); · Kidney function profiles (BUN, creatinine); · Electrolytes (calcium, phosphate, sodium, potassium, chloride, total CO_2); · Lipid profiles (total cholesterol, triglycerides, LDL/HDL cholesterol levels); · Uric acid	
Urine tests	· Protein, glucose, ketone, bilirubin, blood, and pH	
Stool exams	· Occult blood in stool, parasite assay	
Infections and inflammation markers	· HBsAg, Anti-HBs, Anti-HCV, · HAV Ab IgG (peoples < 50 years old) · *H. pylori* IgG Ab	· VDRL (Rapid Plasma Reagin Card Test) · HIV Ag/Ab combo test; · Hs-CRP
Hormones	· Thyroid function tests (TSH, T3, total-T4, Free-T4)	
Tumor markers	· AFP, CEA, · CA19-9	· CA-125 (all females); · PSA (males ≥ 35 years old)
Other clinical tests	· Chest X-ray (digital imaging); · Pulmonary function test; · Electrocardiography; · Gastroendoscopy with pathological reports	· Pap smear and gynecological examination (females ≥ 35 years old) · Mammography (females ≥ 35 years old)

WBC, white blood cell; RBC, red blood cell; PT, prothrombin time; INR, International Normalized Ratio; aPTT, activated partial thromboplastin time; AST, aspartate transaminase; ALT, alanine transaminase; GTP, gamma-glutamyl transferase; BUN, blood urea nitrogen; LDL, low-density lipid; HDL, high-density lipid; HBsAg, hepatitis B surface antigen; Anti-HBs, anti-hepatitis B surface antibody; Anti-HCV, anti-hepatitis C antibody; *H. pylori* IgG Ab, hepatitis A immunoglobulin G antibody; VDRL, venereal disease research laboratory; HIV, human immunodeficiency virus, Hs-CRP, high sensitivity C-reactive protein; TSH, thyroid-stimulating hormone; AFP, alpha-fetoprotein; CA19-9, cancer antigen 19-9; CEA, carcinoembryonic antigen; CA-125, cancer antigen 125; PSA, prostate-specific antigen.

Table 2. Specific variables* collected at baseline and follow-up in the Health and Prevention Enhancement (H-PEACE) study

Specific items	List	
SNP	· SNP variation by genome-wide SNP arrays (Affymetrix platform of Axiom™ Customized Genome-Wide Human Assay) †	
Heart	· Coronary calcium score CT · Echocardiography, treadmill test,	· Coronary CT angiography · Holter monitoring
Stroke	· Lp (a) lipoprotein (a), homocysteine · Echocardiography, carotid Doppler	· Brain MRI, · Brain and carotid MRA
Kidney	· 24-h urinary uric acid, creatinine, urea nitrogen, protein, and microalbumin · 24-h urinary electrolytes including sodium, potassium, calcium, phosphorus, and magnesium · 24-h urinary citrate and oxalate	· Serum cystatin C · Dynamic kidney CT
Intestine	· Abdominal sonography · Colonoscopy with pathologic reports	· Abdomino-pelvic CT
Obesity	· 24-h recall diet questionnaire · Serum insulin levels	· Abdomen and thigh CT (for visceral fat) · Whole body DEXA
Dementia	· Apo-E genotyping · MR double time	· Brain MRI, brain MRA, hippocampus (non-contrast)
Prostate	· Free PSA · Uroflowmetry · Ultrasonography (for residual urine)	· Transrectal sonography · Prostate MRI (non-contrast)
Sex hormones	· Testosterone (males ≥ 60 years old) · Total and free E2, LH, FSH (postmenopausal women after 5 years from menopause)	
Cigarette smoking	· CO levels in exhalation, urine cotinine levels · Laryngoscopy, low-dose screening chest CT	
Bone	· Bone densitometry (males ≥ 50 years old and postmenopausal women) · Site-specific X-ray digital imaging (AP and Lat.) · Site-specific bone CT	
Dental	· Dental, peridental and periodontal examination	
Nutrition	· Plasma vitamin B6 profile (pyridoxal phosphate (PLP), pyridoxic acid (PA)) · Plasma vitamin A and E (HPLC);	· Plasma 25(OH)D; · Vitamin C, B12, and folate levels · Serum selenium, zinc, copper levels
Allergy	· *MAST* (Multiple *allergen* simultaneous *test*) · IgE (PRIST) · Skin prick test (inhalant)	· Methacholine bronchial c*hallenge test* · Histograms of Categorized Shapes (HCS) ear detection · *X-ray of PNS (paranasal* sinus)
Hepatitis	· HBV viral load, HBeAg (quantitation), HBeAb · HCV PCR	· Liver fibrosis scan
Heavy mineral	· Cd, Pb, Hg, As, Al, Se, Cu, Zn, Cr, Co, Mn, Mo	
Other	· HPV genotyping; · CLO test or UBT test for current *H. pylori* infection	

*The specific variables were selected for the health screenees according to the specific test or physician's request.

†GWAS genotyping was performed among the 17 455 cohort participants who provided informed consent and donated blood samples. We categorized the 17 455 participants with blood specimens as the Gene-environmental interaction and phenotype (GENIE) study (a specific subcohort derived from the H-PEACE study). SNP testing was performed for research purposes and was not a routinely performed test.

SNP, single nucleotide polymorphism; CT, computed tomography; MRI, magnetic resonance imaging; MRA, magnetic resonance angiogram; DEXA, dual-energy X-ray absorptiometry; PSA, prostate-specific antigen; LH, luteinizing hormone; FSH, follicle-stimulating hormone; HPLC, high-performance liquid chromatography; PRIST, paper radioimmunosorbent test, PCR, polymerase chain reaction; CLO, Campylobacter-like organism; UBT, urea breath test.

Table 3. Participation rate for the core and specific variables collected in the Health and Prevention Enhancement (H-PEACE) study at baseline

Variables	Study participants N	Participation rate %*
Core variables		
Most core variables	≥ 90 377	≥ 99
Electrocardiography	91 197	99.8
Estimated GFR	91 197	99.8
Stool parasite examination	82 493	90.3
H. pylori IgG Ab	81 142	88.8
Thyroid function test	77 313	84.6
CLO test or UBT test for current *H. pylori* infection	67 986	74.4
Gastroendoscopy with pathologic report	66 451	72.8
Specific variables		
Dental examination	43 232	47.3
Colonoscopy with pathologic reports	46 050	50.4
Abdomen CT with visceral fat CT	34 771	38.1
Abdomen sonography	33 446	36.6
Echocardiography	20 688	22.7
Coronary CT	12 846	14.1
Echocardiography, Treadmill test	9910	10.9
SNP from GWAS	6579	7.2
Bone densitometry (postmenopausal females)	26 376	99.4
E2 levels (postmenopausal females after 5 years from menopause)	4742	49.9
Bone densitometry (males ≥ 50 years)	9488	48.9
T levels (males ≥ 60 years)	6697	52.9
HPV genotyping (females)	12 999	31.8
Pap smear (females ≥ 35 years)	12 951	49.1
Blood collection for the GENIE study	17 455	19.1
SNP variations in the GWAS	6579	37.7†

*Participation rate of all 91 336 H-PEACE cohort members.

†Participation rate of all 17 455 GENIE (Gene-environmental interaction and phenotype) study members (a subcohort of the H-PEACE study).

GFR, glomerular filtration rate; CLO, Campylobacter-like organism; UBT, urea breath test; CT, computed tomography; GWAS, genome-wide association study; SNP, single nucleotide polymorphism.

Structured questionnaires

Self-reported questionnaires were used to obtain socio-demographic data, personal and familial medical history, health-related behaviors (such as smoking status, alcohol consumption, physical activity, dietary behavior), international prostate symptom score (IPSS), female reproductive factors, and psychological status. Physical activity levels were assessed using the validated Korean version of the International Physical Activity Questionnaire (IPAQ) short form and were classified into three categories: inactive, minimally active and health-enhancing physically active (HEPA).[20,21] Dietary habits were evaluated using the mini-dietary assessment index (MDAI), which was validated in Korean.[22] Psychological status was assessed by the Beck Depression Inventory (BDI) or Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR 16).[23]

Physical examinations

Blood pressure was measured using sphygmomanometers with patients in a seated position after a resting period. If systolic blood pressure was ≥140 mm Hg or diastolic blood pressure was ≥90 mm Hg after a rest period and two measurements, we recorded the values and calculated their averages. Height (cm), weight (kg), waist circumference (cm) and body fat composition (%) were measured by trained nurses with participants wearing a lightweight hospital gown and in bare feet. Height and weight were measured using digital scales in a standing position. Waist circumference was obtained by measuring the smallest natural waist circumference area, which is around the umbilicus using a non-stretch tape measure, without any pressure to the body surface during measurements. Percentage of body fat and visceral fat area were estimated with a multi-frequency bio-impedance analyzer with 8-point tactile electrodes (Inbody 720, Biospace co, Seoul, Korea). Comprehensive eye examinations (visual acuity, ocular tonometry, slit lamp test, fundus photography) and hearing tests were also

performed.

Laboratory tests

Blood samples from the antecubital vein were collected after at least ten hours of fasting. The blood parameters assessed included complete blood cell count, fasting blood glucose, glycated hemoglobin (HbA1c), uric acid, blood urea nitrogen, creatinine, total calcium, inorganic phosphorus, glucose, sodium, potassium, chloride, total CO2, total protein, albumin, total bilirubin, AST, ALT, gamma-GTP, total cholesterol, low-density lipid cholesterol (LDL-C), high-density lipid cholesterol (HDL-C), triglycerides, high sensitivity C-reactive protein (hs-CRP) concentration, prothrombin time, aPTT, HBsAg, anti-HBs, HAV Ab IgG, AFP, CA19-9, CEA and PSA in men and CA 125 in women. The laboratory medicine department at the SNUH has been certified by the Korean Society of Laboratory Medicine and participated in the College of American Pathologist's Survey/Proficiency Testing program. Urine tests were performed by stick test using spot urine. Semi-quantitative variables included pH, protein, glucose, ketone, bilirubin and blood. Stool samples were collected to conduct fecal occult blood tests and parasite assays.

Digital imaging and specific diagnostic variables

Chest X-ray was included as a core variable of the H-PEACE study, which included specific information on diagnostic variables (Table 1). First, the GENIE study collected donated blood samples from 17 455 recipients, and genome-wide SNP arrays from 6579 donated blood samples have already been analyzed using the Affymetrix platform (Axiom™ Customized Genome-Wide Human Assay) (Table 3). We plan to increase the available genetic information by 2020 using an SNP array. The PLINK program (Ver. 1.9) and R statistics (Ver. 3.3.0) were used for quality control procedures and data analysis. SNP genotype data combined with clinical data from the H-PEACE study were used to evaluate gene-environment interactions

and to define the related phenotypes.

In total, 12,846 participants underwent coronary CT with a coronary calcium score to assess coronary calcification. The calcium score of the coronary artery is a strong predictor of myocardial infarction and sudden cardiac death.[24] In addition, 34,771 participants underwent visceral fat CT to measure visceral adipose tissue (VAT) and subcutaneous adipose tissue (SAT). The detailed methods used to measure VAT area and SAT area on abdominal fat CT images have been described elsewhere.[17] This quantitative assessment of intra-abdominal adipose tissue is considered the gold standard for measuring the amount of visceral fat.[25] With advanced imaging techniques including echocardiography, brain MRI/MRA, carotid Doppler ultrasound, abdominal ultrasonography and abdominal CT, we elucidated the correlation of visceral obesity with vascular disease using both data in a complementary manner. The core variables of the H-PEACE study including the blood test results, questionnaire findings, and depression scores can contribute to determining the correlation between visceral obesity and metabolic phenotype.

We collected electrocardiogram (ECG) reports from 91,197 consecutive recipients to evaluate the incidence and risk factors of atrial fibrillation, a significant risk factor for stroke. The effects of a bundle branch block or atrioventricular block on stroke and cardiac disease were also analyzed. Furthermore, we collected Holter monitor results to confirm the ECG reports.

Human papillomavirus (HPV) is known to lead to cervical cancer in women.[26] Of the participants in the H-PEACE study, 12,951 women underwent both a liquid-based cervical cytology (SurePath LBC, Becton Dickinson, Franklin Lakes, NJ, USA) and an HPV genotyping test using an HPV DNA chip (MyHPV Chip, Biomedlab Co., Seoul, Korea) for cervical cancer screening.[27] We also included the results of gynecologic sonography, as well as VDRL and HIV tests.

The World Health Organization (WHO) considers *Helicobacter pylori* infection a class I

carcinogen for gastric cancer.[13] The diagnosis of *H. pylori* infection was based on the detection of serum *H. pylori* immunoglobulin G antibody using a kit (*H. pylori*-EIA-Well, Radim, Rome, Italy) that was previously validated in a nationwide Korean sero-epidemiologic study.[28] Furthermore, 66,451 recipients had available upper endoscopy data with pathology results. Serum pepsinogen data and eradication history were also collected.

A total of 46,050 participants in the H-PEACE study received a colonoscopy in our center. All colonoscopies were conducted by board certified endoscopists, and the average adenoma detection rate in recipients 50-70 years old was over 30%.[12] Endoscopes including the CF-H260 and CF-HQ290 series (Olympus, AIZU, Japan) and the EC-450HL5, EC-450WM5, and EC-590ZW series (Fujinon, Saitama, Japan) were used. Histological diagnoses at our center were determined according to the WHO classification of tumors of the digestive system.[29,30] All colonoscopy results and corresponding pathology reports were collected.

We calculated estimated glomerular filtration rate (GFR) using the chronic kidney disease epidemiology collaboration equation and the modification of diet in renal disease formula. We collected GFR data successively to identify the risk factors for chronic kidney disease. We also included spot urine sodium and creatinine ratio and 24-hour urine data including GFR.

Dental exams were conducted in 43,232 participants. Periodontitis is highly prevalent among adults and is one of the most common causes of teeth loss after 40 years.[31-33] Furthermore, periodontitis can induce systemic inflammation, as well as masticatory dysfunction and poor nutritional status.[32] This is the first cohort study to include a large number of dental exams to evaluate the progression of periodontitis. In this study, 43,232 participants received a dental exam; 33.1% had mild to moderate periodontitis, and 2.1% had severe periodontitis. We also included participants' colonoscopy and coronary CT results, including coronary calcification and colon polyps, to assess the correlation between periodontitis and systemic inflammation.

The presence of osteoporosis was determined using bone densitometry in 26,207 women and 9488 men. We measured bone mineral density (BMD) of the lumbar spine (L1-L4) and femur

using dual-energy X-ray absorptiometry (GE Medical, United Kingdom). Based on participants' lowest T scores, normal BMD, osteopenia, and osteoporosis were defined as T scores of −1.0 SD and above, −1.0 to -2.5 SDs, and -2.5 SDs and below, respectively. In this cohort, 26.6% of the participants who underwent bone densitometry were male. Of those, osteopenia was found in 24.75% and osteoporosis was found in 3.21%. The information collected in this cohort also included hormone levels, estrogen levels in women and testosterone levels in men.

FINDINGS TO DATE

The baseline characteristics are summarized in Table 4, and the morbidity rate of the major outcomes is shown in Table 5. The morbidity rates were calculated as the number of findings divided by the number of participants tested. The prevalence of major outcomes (mortality and cancer incidence rate) is provided in Table 6. The H-PEACE study has contributed to the research community by publishing >200 articles since the initiation of the data warehouse, HEALTH-WATCH®, as a prototype in 2003.[8] Our studies have primarily focused on cancer screening and identifying the risk factors and prognosis of metabolic disease.

339
340
341
342

Table 4. Baseline characteristics of the participants in the Health and Prevention Enhancement (H-PEACE) study

	Total (N=91 336)	Men (N=50 507)	Women (N=40 829)	p-value
	N (%)	N (%)	N (%)	
Age (years)				<0.001
< 35	16228 (17.8)	7666 (15.2)	8562 (21.0)	
35 – 44	26963 (29.5)	15139 (30.0)	11824 (29.0)	
45 – 54	27915 (30.6)	15776 (31.2)	12139 (29.7)	
55– 64	14937 (16.4)	8717 (17.3)	6220 (15.2)	
65 +	5293 (5.8)	3209 (6.4)	2084 (5.1)	
Sex				
Men	50509 (55.3)			
Women	40827 (44.7)			
Cigarette smoking status				<0.001
Non-smokers	48 521 (53.1)	11 779 (23.3)	36 742 (90.0)	
Ex-smokers	22 564 (24.7)	20 469 (40.5)	2095 (5.1)	
Current smokers	20 251 (22.2)	18 259 (36.2)	1992 (4.9)	
Regular exercise				<0.001
Inactive	36 343 (39.8)	17 843 (35.3)	18 500 (45.3)	
Minimally active	26 189 (28.7)	16 428 (32.5)	9761 (23.9)	
HEPA	28 804 (31.5)	16 236 (32.1)	12 568 (30.8)	
	N (%)	N (%)	N (%)	
Excess alcohol drinking (>140 g/week)	20 421 (22.3)	18 934 (37.5)	1469(3.6)	<0.001
Education ≥ high school graduate	78 821 (86.3)	45 538 (90.2)	33 283 (81.5)	<0.001
Hypertension	13 279 (14.5)	9003 (17.8)	4276 (10.5)	<0.001
Diabetes	4377 (4.8)	3181 (6.3)	1196 (2.9)	<0.001
Current HBsAg+	3810 (4.2)	2377 (4.7)	1433 (3.5)	<0.001
Current Anti-HCV+	887 (0.97)	492 (0.97)	395 (0.97)	0.946
	Mean (SD)	Mean (SD)	Mean (SD)	
Measurements				
Height (cm)	165.6 (11)	171.7 (5.6)	159.9 (5.2)	<0.001
Weight (kg)	64.3 (12.1)	71.5 (9.1)	55.0 (7.2)	<0.001
Body mass index (kg/m^2)	23.4 (12.1)	24.2 (2.6)	21.5 (2.8)	<0.001
Waist circumference (cm)	83.7 (9.3)	86.5 (7.1)	76.9 (7.5)	<0.001
Skeletal muscle mass (kg)	26.6 (6.0)	30.8 (3.7)	20.6 (2.4)	<0.001
Body fat mass (kg)	26.9 (5.9)	30.8 (3.6)	20.7 (2.3)	<0.001
Visceral fat area (cm^2)*	79.9 (24.2)	82.3 (23.6)	76.1 (24.7)	<0.001
SBP (mmHg)	115.4 (15.9)	118.0 (12.2)	111.2 (13.3)	<0.001
DBP (mmHg)	76.01 (11.9)	79.0 (9.45)	71.36 (9.56)	<0.001
Blood levels				
Fasting blood glucose (mg/dL)	97 (18.6)	100.2 (17.6)	90.7 (16.2)	<0.001
HbA1c (%)	5.56 (0.6)	5.6 (0.6)	5.4 (0.4)	<0.001
Creatinine (mg/dL)	0.84 (0.17)	0.93 (0.13)	0.68 (0.10)	<0.001
Albumin (g/dL)	4.62 (0.25)	4.64 (0.24)	4.58 (0.25)	0.002
Total cholesterol (mg/dL)	191.03 (33.02)	192.5 (33.1)	188.6 (32.8)	0.105
Triglycerides (mg/dL)	101.2 (64.2)	117.9 (68.8)	74.6 (44.9)	<0.001
LDL cholesterol (mg/dL)	119.8 (31.6)	126.3 (31.3)	112.6 (30.0)	<0.001
HDL cholesterol (mg/dL)	54.3 (13.2)	49.4 (9.5)	59.7 (12.6)	<0.001
Hs-CRP (mg/dL)	0.12 (0.26)	0.12 (0.26)	0.10 (0.26)	0.285

*Visceral fat area is defined here as the cross-sectional area of visceral fat found in the abdomen.

HEPA, health-enhancing physically active; HBsAg, hepatitis B surface antigen; Anti-HCV, anti-hepatitis C antibody; SBP, Systolic blood pressure; DBP, Diastolic blood pressure; LDL, low-density lipid; HDL, high-density lipid; Hs-CRP, high sensitivity C-reactive protein.

Table 5. Abnormal health findings detected through active follow-up with specific measurements in the Health and Prevention Enhancement (H-PEACE) study

Test	Definition	Participants tested N	Follow-up years Median (IQR)	Clinical finding* N	
Coronary calcification	· Coronary artery calcium score > 400 at coronary CT	12 846	5.6 (3.2-8.0)	805	
Atrial fibrillation	· ECG and/or ambulatory ECG	91 076	4.1 (2.1-6.5)	446	
Left bundle branch block	· ECG and/or ambulatory ECG	91 076	4.1 (2.1-6.5)	64	
Chronic kidney disease	· Estimated GFR† < 60 mL/min/1.73 m^2	91 197	4.0 (2.1-6.5)	5372	
Visceral obesity	· Men > 10000 cm^2, Women > 8000 cm^2 in abdomen visceral fat CT	34 771	5.0 (3.0-7.4)	23 126	
Fatty liver	· Abdominal sonography and/or abdominal CT	34 771	5.0 (3.0-7.4)	Mild	10 244
				Moderate	5963
				Severe	688
Abnormal thyroid function	· Thyroid hormone measurement	77 313	4.1 (2.2-6.6)	Hypothyroidism	363
				Hyperthyroidism	1260
H. Pylori infection	· *H. pylori* IgG Ab and/or CLO test or UBT test for current *H. pylori* infection	81 142	4.2 (2.3-6.7)	44 304	
High-risk HPV infection	· HPV 16, 18 positive finding in HPV genotyping	12 999	5.0 (3.1-7.2)	1462	
Parasite infection	· Stool examination	82 493	4.1 (2.2-6.6)	*Clonorchis sinensis*	1262
				Ascaris lumbricoides	124
				Trichuris trichiura	256
				Metagonimus yokogawai	412
Premalignant conditions of stomach	· Gastroendoscopy with pathological reports	66 451	4.2 (2.3-6.7)	Atrophic gastritis	23 896
				Intestinal metaplasia	10 187
Malignant gastric cancer	· Gastroendoscopy with pathological reports	66 451	4.2 (2.3-6.7)	299	
Squamous intraepithelial lesion (SIL)	· Pap smear and pathological report	12 951	5.0 (3.1-7.2)	Total SIL	375
				Low grade	315
				High grade	60
Premalignant colonic neoplasms	· Colonoscopy with pathological reports	46 746	4.8 (2.9-7.1)	Adenoma	18 656‡
				Serrated polyp	1178‡
				Carcinoid	150
Malignant colon cancer	· Colonoscopy with pathological reports	46 746	4.8 (2.9-7.1)	262	
Low bone density	· Bone densitometry	35 695	4.8 (2.9-7.2)	Osteopenia	10 601
				Osteoporosis	1653
Periodontitis	· Dental examination	43 232	4.9 (2.8-7.2)	Total periodontitis	15 204
				Mild to moderate	14 292
				Severe	912

*The number of findings among the participants who were at risk.

†The glomerular filtration rate was estimated by the chronic kidney disease epidemiology collaboration (CKD-EPI) equation and the modification of diet in renal disease (MDRD) formula.

‡Adenoma and serrated polyp detection rate were calculated in participants 50-75 years old

ECG: electrocardiogram; GFR, glomerular filtration rate; CT, computed tomography; CLO, Campylobacter-like organism; UBT, urea breath test; HPV, human papillomavirus.

Table 6. Major cancer incidence and outcomes based on record linkage to the Nationwide Death Certificate Database in the Health and Prevention Enhancement (H-PEACE) study*

	N of cancer incidence			N of all-cause death†		
	Total	Men	Women	Total	Men	Women
Total cancer	2814	1402	1412	207	150	57
Thyroid	956	390	566	3	3	0
Stomach	513	337	176	37	29	8
Breast	343	1	342	6	0	6
Colorectal	284	182	102	28	20	8
Lung	186	109	77	45	32	13
Prostate	127	127	0	3	3	0
Kidney	91	71	20	1	1	0
Liver	87	71	16	31	28	3
Gynecology	52	0	52	7	0	7
Pancreas	42	25	17	27	19	8
Brain	39	20	19	0	0	0
Hematologic	39	24	15	5	3	2
Biliary tract	32	26	6	11	9	2
Esophagus	23	19	4	3	3	0

*Median (interquartile range) follow-up years of the 91 336 participants=4.04 (2.13-6.5).
†Since the Nationwide Death Certificate Database does not provide information about cause of death, N of death includes death from not only cancer but also other causes. Among the 91 336 H-PEACE participants, the all-cause mortality rate was 969: 715 men, 254 women.

For subjects with 1-2 adenomas less than 10 mm in colonoscopy for colon cancer screening, they were classified as a low-risk group having low 5-year colon adenoma incidence rates and were recommended to take further colonoscopic screening test after 5 year. For subjects with advanced adenoma, multiple adenoma ≥ 3, or larger adenoma sized ≥10mm in baseline colonoscopy for colon cancer screening, they were classified as a high-risk group having higher incidence rates of advanced adenoma or higher recurrence rates of adenoma and were recommended to take colonoscopic surveillance at 3-year after initial polypectomy [12]. This strategy of colonoscopic surveillance has been reflected in the guidelines for colonoscopic surveillance [34].

We also examined the quality of colon cancer screening using metrics such as the adenoma detection rate [35]. Additionally, we evaluated the influence of image-enhanced endoscopy,

such as narrow band imaging and Fuginon Intelligent Color Enhancement, using our cohort data [10, 36, 37]. For gastric cancer screening, the effect of screening endoscopy remains controversial, but population-based screening has been undertaken in Korea and Japan. For participants with intestinal metaplasia in gastroendoscopy, we also classified them into high-risk group and recommended an annual endoscopic screening, based on study results that people with strong risk factors such as male and an older age can quickly find early-staged endoscopically-treatable gastric cancer by taking annual gastroendoscopic screening [11]. Gastric mucosa-associated lymphoid tissue lymphomas (MALToma) and suspicious MALToma lesions were also detected in screening endoscopies, and the prevalence of gastric MALToma was high in middle-aged women.[38] Active endoscopic screening for gastric cancer had the additional benefit of detecting early-stage MALToma.[38,39] Furthermore, we established that small subepithelial tumors that were incidentally found during upper endoscopy screening showed no size change in subsequent follow-up periods.[36]

Our research focus was not limited to cancer screening. We also examined metabolic diseases such as non-alcoholic fatty liver disease (NAFLD) and metabolic syndrome. We reported that metabolic syndrome was an independent risk factor of silent brain infarction using brain MRI data.[40] We also found an effect of body fat distribution on the incidence of NAFLD and reflux esophagitis using abdominal fat CT data.[9,14] We also identified a relation between NAFLD and the risk of coronary heart disease and arterial stiffness based on coronary CT and cardio-ankle vascular index, respectively.[16-19] The effect of physical activity on NAFLD was also demonstrated using data from the questionnaires.[41]

The list and number of publications can be requested on our homepage (http://en-healthcare.snuh.org/HPEACEstudy).

STRENGTHS AND LIMITATIONS

This study has several potential limitations as below. First, our study participants are composed of individuals who voluntarily visited our center and their data collection for the repeated measurements rely on participants' self-paid. Our system of enrolling cohort members includes a potential selection bias**.** Second, a self-record questionnaires are used to obtain the information of risk factors and past disease histories before the next visit of study participants to our center and this procedure leads to recall bias and response bias. Third, our active follow-up rates in assessing outcome variables and repeatedly measuring risk factors is not high, leading to selection bias. However, since active follow-up rates are increasing every year, at least 87% of follow-up is expected if all subjects are terminated for at least 10-year follow-up. Forth, in our cohort, we used to measure waist circumference as the definition of the smallest area around the belly button below the rib cage and above the hip bone, although the ideal way is to measure the midway between the top of the hip bone and below the rib cage. The former is usually called the natural waist, and we used this former definition for practical reason to reduce intra-individual measurement bias in waist circumference. In health check-up at our center, a lot of participants (about 100-120 health examinees) everyday visit to center for health examination and many health technicians and nurses take measurement of waist and hip circumferences. For each participants, there is a very little chance to take measurement of waist at the time of re-visit by same nurses met at the time of cohort enrollment. The intra-individual measurement bias in measuring waist circumference may be problematic at our center and thus we thought that it was necessary to use the most practical and easy waist measurement together with repeated nurse training. We did a small pilot study for measuring waist by the two methods for 10 men and 16 women. The ideal measuring method (midway measurement) has a limit in consuming time and effort due to difficulty in method itself. Despite of short time and smaller effort in measuring waist, the natural waist

measuring method showed an excellent agreement (intraclass correlation coefficients = 99% in men and 93% in women) with ideal method and there were no shift between obesity groups classified by each method. Fifth, there is no information of cancer histology subtypes for cohort participants ascertained as new cancer development. Individual cancer development in study participants was ascertained with data linkage to the nationwide cancer registry data. Korea cancer registry data includes ICD10 code, T-code (Topography) and M-code (Morphology in primary cancer sites). For example, in stomach cancer, 51.3% were adenocarcinomas, 22.1% were adenocarcinomas, and gastric cancer and histologic NOS were only 4.9% [42]. Future studies will attempt to merge the individual data associations of M-code and T-code. Finally, the definition of a disease status could be changed during the follow up periods, as new guidelines are declared. For instance, guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults opened at November 2017 [43] recommending a new definition for hypertension. Our manuscript was written before the declaration was done. We have the raw data of blood pressure and the definition can be applied to the study population. But for those who had already been under anti-hypertensive medication at the point of enrollment in our cohort, had followed the previous guidelines for indication of treatment. This makes it difficult to re-classify these patients according to the new guidelines. In the future follow up period, we will give a big attention to clearly declare the definition of disease we are applying according to the period of terms.

A major strength of the H-PEACE study is its structured and organized database, which not only includes data widely used in medical check-ups but also data from sophisticated high-quality advanced examinations, such as colonoscopy, brain MRI, abdominal CT and abdominal fat CT. Particularly in our study, comprehensive biomarkers related to NCDs, including cancer antigens and infection markers, were included, which have not been

included in the medical check-ups of other cohort profiles. It has only been about 10~15 years since the comprehensive health check-up program is actively performed in Korea and in the past decade, a large proportion of peoples participated in the check-up due the support from the affiliated company's welfare policy. This might result in the relatively young population enrollment in our study population. But since the health check-up program is getting more general, more elderly population is taking the health check-up. New enrollment is another future target study of our center. Since the enrolled cohort population is relatively young, it is possible studying the preclinical disease stage and its final long term outcomes. This might be the characteristics of our cohort that we could design a lot of prediction model for the non-communicable disease by using a data from the preclinical stages. As we obtained a large sample size, we could confirm disease statuses as well as numerous phenotypes of NCDs. Furthermore, the study protocols were conducted and monitored with intensive quality control procedures. Additional strengths include the active and passive follow-ups and the ability to obtain complete data, covering deaths and incidental cancer cases even among those who discontinued visiting our center. We believe that these data could contribute to active, effective prevention against the development of cancers and NCDs.

COLLABORATION

The genotype (SNP) data from the GENIE study and the epidemiologic data from the other cohorts are available to researchers after a quality control process has been completed. Potential collaborators can access the dataset after receiving approval from the SNUH Gangnam Center institutional review board. Applications can be submitted via our homepage. Further information is available at the H-PEACE study website (http://en-healthcare.snuh.org/HPEACEstudy).

ACKNOWLEDGEMENTS

The authors thank You Mi Kwon and Si Won Choi for their technical assistance.

COMPETING INTERESTS

J.E.L is an employee of DNALink, Inc.

FUNDING

This work was supported by the SK Telecom Research Fund through Seoul National University Hospital [34-2013-0390 and 34-2014-0210 to S.J.C.] and the Seoul National University Hospital Research Fund [04-2011-0700 to S.J.C.].

CONTRIBUTORS

Changhyun Lee (CL), Eun Kyung Choe (EKC), Ji Min Choi (JMC), Sue K. Park (SKP), Su Jin Chung (SJC), Min-Sun Kwak (MK), Jong-Eun Lee (JL), Joo Sung Kim (JSK) and Sang-Heon Cho (SC) were involved in the conception and design of the study.

CL, EKC, JMC, Yunji Hwang (YH), Young Lee (YL), MK, JL, SJC and Boram Park (BP) contributed to development of methods and data collection.

CL, EKC, JMC, YL, SKP, SC and BP were involved in data analysis and interpretation.

CL, EKC, JMC, SKP and SC drafted the manuscript.

All the authors have critically revised the article and approved the final version and the findings.

ETHICS APPROVAL

The H-PEACE study was approved by the ethics committee of Seoul National University Hospital.

DATA SHARING STATEMENT

Researchers can apply for the dataset after receiving approval from the SNUH Gangnam Center institutional review board. The H-PEACE study website provides information on the application process (http://en-healthcare.snuh.org/HPEACEstudy).

REFERENCES

1 Beaglehole R, Bonita R, Horton R, et al. Priority actions for the non-communicable disease crisis. *Lancet* 2011;377:1438-47.

2 Ministry of Health and Welfare of Korea, Korea Centers for Disease Control and Prevention. *2010 Korea Health statistics.* Seoul: Ministry of Health and Welfare of Korea, 2011.

3 Hunter DJ. Gene-environment interactions in human diseases. *Nat Rev Genet* 2005;6:287-98.

4 Renz H, von Mutius E, Brandtzaeg P, et al. Gene-environment interactions in chronic inflammatory disease. *Nat Immunol* 2011;12:273-7.

5 Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015;372:793-5.

6 Jameson JL, Longo DL. Precision medicine — personalized, problematic, and promising. *N Engl J Med* 2015;372:2229-34.

7 Elliott P, Peakman TC, UK Biobank. The UK biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol* 2008;37:234-44.

8 Park MJ, Kim DH, Lim SH, et al. Features of gastric neoplasm detected during the screening examination. *Gut Liver* 2007;1:33-9.

9 Kim D, Chung GE, Kwak MS, et al. Body fat distribution and risk of incident and regressed nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2016;14:132–8.e4.

10 Chung SJ, Kim D, Song JH, et al. Comparison of detection and miss rates of narrow band imaging, flexible spectral imaging chromoendoscopy and white light at screening colonoscopy: a randomised controlled back-to-back study. *Gut*

2014;63:785-91.

11 Chung SJ, Park MJ, Kang SJ, et al. Effect of annual endoscopic screening on clinicopathologic characteristics and treatment modality of gastric cancer in a high-incidence region of Korea. *Int J Cancer* 2012;131:2376-84.

12 Chung SJ, Kim YS, Yang SY, et al. Five-year risk for advanced colorectal neoplasia after initial colonoscopy according to the baseline risk stratification: a prospective study in 2452 asymptomatic Koreans. *Gut* 2011;60:1537-43.

13 Chung SJ, Lim SH, Choi J, et al. Helicobacter pylori serology inversely correlated with the risk and severity of reflux esophagitis in helicobacter pylori endemic area: a matched case-control study of 5,616 health check-up Koreans. *J Neurogastroenterol Motil* 2011;17:267-73.

14 Chung SJ, Kim D, Park MJ, et al. Metabolic syndrome and visceral obesity as risk factors for reflux oesophagitis: a cross-sectional case-control study of 7078 Koreans undergoing health check-ups. *Gut* 2008;57:1360-5.

15 Chung GE, Kim D, Kim W, et al. Non-alcoholic fatty liver disease across the spectrum of hypothyroidism. *J Hepatol* 2012;57:150-6.

16 Chung GE, Choi SY, Kim D, et al. Nonalcoholic fatty liver disease as a risk factor of arterial stiffness measured by the cardioankle vascular index. *Medicine (Baltimore)* 2015;94:e654.

17 Kim D, Choi SY, Park EH, et al. Nonalcoholic fatty liver disease is associated with coronary artery calcification. *Hepatology* 2012;56:605-13.

18 Kwak MS, Yim JY, Kim D, et al. Nonalcoholic fatty liver disease is associated with coronary artery calcium score in diabetes patients with higher HbA1c. *Diabetol Metab Syndr* 2015;7:28.

19 Choi SY, Kim D, Kim HJ, et al. The relation between non-alcoholic fatty liver disease

and the risk of coronary heart disease in Koreans. *Am J Gastroenterol* 2009;104:1953-60.

20 Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* 2003;35:1381-95.

21 Chun MY. Validity and reliability of Korean version of international physical activity questionnaire short form in the elderly. *Korean J Fam Med* 2012;33:144-51.

22 Kim WY, Cho MS, Lee HS. Development and validation of mini dietary assessment index for Koreans. Korean. *J Nutr* 2003;36:83-92.

23 Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. *Clin Psychol Rev* 1988;8:77-100.

24 Wong ND, Hsu JC, Detrano RC, et al. Coronary artery calcium evaluation by electron beam computed tomography and its relation to new cardiovascular events. *Am J Cardiol* 2000;86:495-8.

25 Seidell JC, Bakker CJ, van der Kooy K. Imaging techniques for measuring adipose-tissue distribution--a comparison between computed tomography and 1.5-T magnetic resonance. *Am J Clin Nutr* 1990;51:953-7.

26 Bosch FX, Manos MM, Muñoz N, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. *J Natl Cancer Inst* 1995;87:796-802.

27 Kim MJ, Kim JJ, Kim S. Type-specific prevalence of high-risk human papillomavirus by cervical cytology and age: data from the health check-ups of 7,014 Korean women. *Obstet Gynecol Sci* 2013;56:110-20.

28 Yim JY, Kim N, Choi SH, et al. Seroprevalence of Helicobacter pylori in South Korea. *Helicobacter* 2007;12:333-40.

29 Hamilton SR, Aaltonen LA, 2000. *WHO Classification of Tumours of the Digestive System*. 3rd ed. Geneva: IARC.

30 Bosman FT, Carneiro F, Hruban RH, et al., 2010. *WHO Classification of Tumours of the Digestive System*. 4th ed. Geneva: IARC.

31 Eke PI, Dye BA, Wei L, et al. Update on prevalence of periodontitis in adults in the United States: NHANES 2009 to 2012. *J Periodontol* 2015;86:611-22.

32 White DA, Tsakos G, Pitts NB, et al. Adult Dental Health survey 2009: Common oral health conditions and their impact on the population. *Br Dent J* 2012;213:567-72.

33 Murray H, Locker D, Kay EJ. Patterns of and reasons for tooth extractions in general dental practice in Ontario, Canada. *Commun Dent Oral Epidemiol* 1996;24:196-200.

34 Lieberman DA, Rex DK, Winawer SJ, et al. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology* 2012;143:844-57.

35 Kang HY, Kim D, Kim HJ, et al. The relationship between colonoscopy procedure order and adenoma detection rates: a prospective study. *J Clin Gastroenterol* 2015;49:683-9.

36 Song JH, Kim SG, Chung SJ, et al. Risk of progression for incidental small subepithelial tumors in the upper gastrointestinal tract. *Endoscopy* 2015;47:675-9.

37 Chung SJ, Kim D, Song JH, et al. Efficacy of computed virtual chromoendoscopy on colorectal cancer screening: a prospective, randomized, back-to-back trial of Fuji intelligent color enhancement versus conventional colonoscopy to compare adenoma miss rates. *Gastrointest Endosc* 2010;72:136-42.

38 Yang HJ, Lee C, Lim SH, et al. Clinical characteristics of primary gastric lymphoma detected during screening for gastric cancer in Korea. *J Gastroenterol Hepatol* 2016;31:1572-83.

39 Yang HJ, Lim SH, Lee C, et al. Management of suspicious mucosa-associated lymphoid tissue lymphoma in gastric biopsy specimens obtained during screening endoscopy. *J Korean Med Sci* 2016;31:1075-81.

40 Kwon HM, Kim BJ, Lee SH, et al. Metabolic syndrome as an independent risk factor of silent brain infarction in healthy people. *Stroke* 2006;37:466-70.

41 Kwak MS, Kim D, Chung GE, et al. Role of physical activity in nonalcoholic fatty liver disease in terms of visceral obesity and insulin resistance. *Liver Int* 2015;35:944-52.

42. Korea Ministry of Health and Welfare, Korea National Cancer Center. 2001 Annual Report of the Korea Central Cancer Registry. Korea National Cancer Center, 2003. [Available at file:///C:/Users/user/Downloads/2001cancer_regi_result.pdf]

43 Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol* 2017;1097:41517-8

FIGURE LEGENDS

Fig 1. Study flow diagram of the Health and Prevention Enhancement (H-PEACE) study

Fig 2. Regional distribution of the Health and Prevention Enhancement (H-PEACE) study in Korea

Fig 3. Flow diagram of health check-ups in the Health and Prevention Enhancement (H-PEACE) study

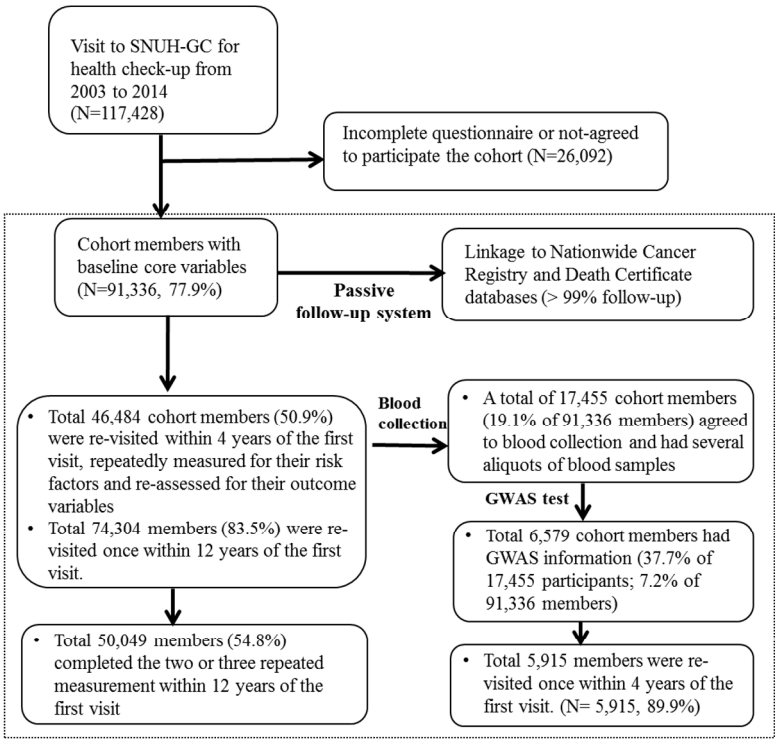
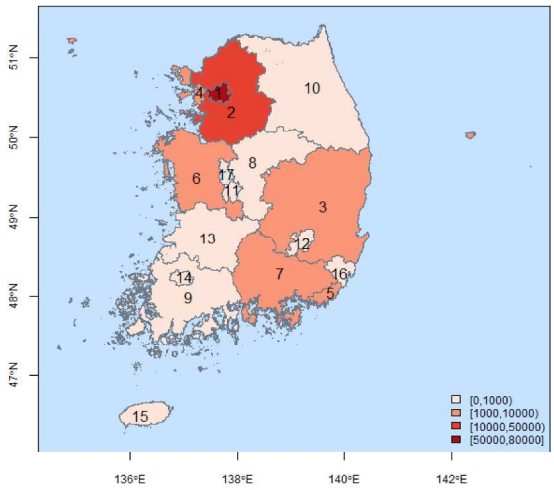


Fig 1. Study flow diagram of the Health and Prevention Enhancement (H-PEACE) study

254x190mm (300 x 300 DPI)



	Location	N
1	Seoul (Capital, Metropolitan)	60,301
2	Gyeonggi-do	18,383
3	Gyeongsangbuk-do	1,564
4	Incheon (Metropolitan)	1,478
5	Busan (Metropolitan)	1,445
6	Chungcheongnam-do	1,400
7	Gyeongsangnam-do	1,278
8	Chungcheongbuk-do	769
9	Jeollanam-do	737
10	Gangwon-do	727
11	Daejeon (Metropolitan)	692
12	Daegu (Metropolitan)	686
13	Jeollabuk-do	580
14	Gwangju (Metropolitan)	488
15	Jeju-do	338
16	Ulsan (Metropolitan)	297
17	Sejong (Metropolitan)	20
	Total	91,183

Fig 2. Regional distribution of the Health and Prevention Enhancement (H-PEACE) study in Korea

254x190mm (300 x 300 DPI)

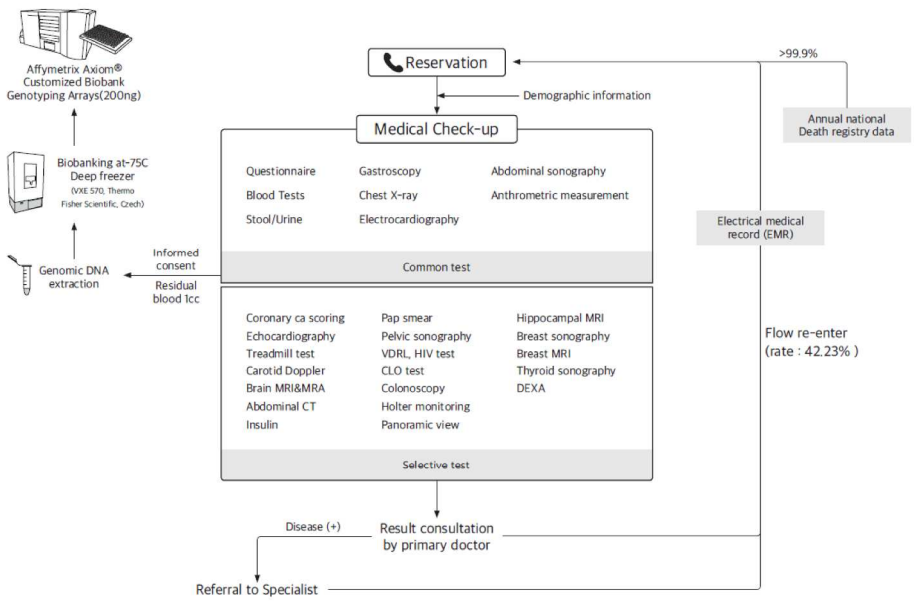


Fig 3. Flow diagram of health check-ups in the Health and Prevention Enhancement (H-PEACE) study

254x190mm (300 x 300 DPI)

STROBE (Strengthening The Reporting of OBservational Studies in Epidemiology) Checklist

A checklist of items that should be included in reports of observational studies. You must report the page number in your manuscript where you consider each of the items listed in this checklist. If you have not included this information, either revise your manuscript accordingly before submitting or note N/A.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Section and Item	Item No.	Recommendation	Reported on Page No.
Title and Abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/Rationale	2	Explain the scientific background and rationale for the investigation being reported	6
Objectives	3	State specific objectives, including any prespecified hypotheses	6-7
Methods			
Study Design	4	Present key elements of study design early in the paper	8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	8
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	8
		Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	N/A
		Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants	N/A
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed	N/A
		Case-control study—For matched studies, give matching criteria and the number of controls per case	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	9-15

Section and Item	Item No.	Recommendation	Reported on Page No.
Data Sources/ Measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	9-15
Bias	9	Describe any efforts to address potential sources of bias	31
Study Size	10	Explain how the study size was arrived at	9
Quantitative Variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	N/A
Statistical Methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	N/A
		(*b*) Describe any methods used to examine subgroups and interactions	N/A
		(*c*) Explain how missing data were addressed	N/A
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed	N/A
		Case-control study—If applicable, explain how matching of cases and controls was addressed	N/A
		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	22-24
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	Fig 1
Descriptive Data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 4
		(b) Indicate number of participants with missing data for each variable of interest	Fig 1
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	8
Outcome Data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	Table 5,6
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	N/A
		Cross-sectional study—Report numbers of outcome events or summary measures	N/A

Section and Item	Item No.	Recommendation	Reported on Page No.
Main Results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	N/A
		(*b*) Report category boundaries when continuous variables were categorized	Table 4,5
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other Analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	N/A
Discussion			
Key Results	18	Summarise key results with reference to study objectives	31
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	31
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	31
Generalisability	21	Discuss the generalisability (external validity) of the study results	31
Other Information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	32

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Once you have completed this checklist, please save a copy and upload it as part of your submission. DO NOT include this checklist as part of the main manuscript document. It must be uploaded as a separate file.

BMJ Open

Cohort profile: The Health and Prevention Enhancement (H-PEACE), a retrospective, population-based cohort study conducted at the Seoul National University Hospital Gangnam Center, Korea

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019327.R2
Article Type:	Cohort profile
Date Submitted by the Author:	12-Feb-2018
Complete List of Authors:	Lee, Changhyun; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Choe, Eun Kyung; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Choi, Ji Min; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Hwang, Yunji; Seoul National University College of Medicine, Department of Preventive Medicine; Seoul National University Graduate School, Department of Biomedical Science Lee, Young; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Park, Boram; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Chung, Su Jin ; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Kwak, Min-Sun ; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Lee, Jong-Eun ; DNA link Kim, Joo Sung ; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute Park, Sue K.; Seoul National University College of Medicine, Department of Preventive Medicine; Seoul National University Graduate School, Department of Biomedical Science Cho, Sang-Heon; Healthcare System Gangnam Center, Seoul National University Hospital, Healthcare Research Institute
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Patient-centred medicine
Keywords:	cohort, non-communicable disease, PREVENTIVE MEDICINE, EPIDEMIOLOGY

Cohort profile: The Health and Prevention Enhancement (H-PEACE), a retrospective, population-based cohort study conducted at the Seoul National University Hospital Gangnam Center, Korea

Short title: Cohort profile of the H-PEACE study

Changhyun Lee[1¶], Eun Kyung Choe[1¶], Ji Min Choi[1], Yunji Hwang[2,3,4], Young Lee[1], Boram Park[1], Su Jin Chung[1], Min-Sun Kwak[1], Jong-Eun Lee[5], Joo Sung Kim[1], Sue K. Park[2,3,4&*], Sang-Heon Cho[1&*]

[1]Healthcare Research Institute, Healthcare System Gangnam Center, Seoul National University Hospital, Seoul, Korea

[2]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea

[3]Department of Biomedical Science, Seoul National University Graduate School, Seoul, Korea

[4]Seoul National University Cancer Research Institute, Seoul, Korea

[5]DNA link, Inc., Seoul, Korea

[¶]Changhyun Lee and Eun Kyung Choe contributed equally to this paper.

[&]Sue Kyung Park and Sang-Heon Cho contributed equally to this paper.

***Corresponding authors**

Sang-Heon Cho

Gangnam Finance Center, Teheran-ro 152, Gangnam-gu, Seoul, 06236, Korea

Tel: +82-2-2072-2971; E-mail: shcho@snu.ac.kr

Word count: 4,207

ABSTRACT

Purpose: The Health and Prevention Enhancement (H-PEACE) study was designed to investigate the association of diagnostic imaging results, biomarkers and the pre-disease stage of non-communicable diseases (NCDs), such as malignancies and metabolic diseases, in an average-risk population in Korea.

Participants: This study enrolled a large-scale retrospective cohort at the Healthcare System Gangnam Center, Seoul National University Hospital, from October 2003 to December 2014.

Findings to date: The baseline and follow-up information collected in the pre-disease stage of NCDs allows for evaluation of an individual's potential NCD risk, which is necessary for establishing personalized prevention strategies. A total of 91,336 health examinees were included in the cohort, and we repeatedly measured and collected information for 50.9% (N=46,484) of the cohort members. All participants completed structured questionnaires (lifestyle, medical history, mini-dietary assessment index, sex-specific variables, and psychiatric assessment), doctors' physical examinations, laboratory blood and urine tests and digital chest X-ray imaging. For participants with available data, we also obtained information on specific diagnostic variables using advanced diagnostic tests, including coronary computed tomography (CT) for coronary calcium scores, colonoscopy, and brain magnetic resonance imaging (MRI). Furthermore, 17,455 of the participants who provided informed consent and donated blood samples were enrolled into the Gene-environmental interaction and phenotype (GENIE) study, a subcohort of the H-PEACE, from October 2013 and we analyzed genome-wide single nucleotide polymorphism (SNP) array data for 6,579 of these blood samples.

Future plans: The data obtained from this cohort will be used to facilitate advanced and accurate diagnostic techniques related to NCDs while considering various phenotypes. Potential collaborators can access the dataset after receiving approval from our institutional review board. Applications can be submitted on the study homepage (http://en-

healthcare.snuh.org/HPEACEstudy).

Keywords: epidemiology; cohort; non-communicable disease; preventive medicine

STRENGTHS AND LIMITATIONS OF THIS STUDY

- The strengths of the H-PEACE study include a large number of healthy subjects (N=91,336 healthy examinees, from the Healthcare System Gangnam Center, Seoul National University Hospital, between 2003 and 2014) and a structured and organized database.
- This study not only includes data widely used in medical check-ups but also data from sophisticated high-quality advanced examinations to investigate clinical effectiveness in predicting the pre-disease stage of non-communicable diseases, including malignancies and metabolic diseases, in an average-risk population in Korea.
- Another strength includes the active and passive follow-ups and the ability to obtain complete data, including deaths and incidental cancer cases, even among those who discontinued visiting our center.
- The data from this study will allow us to contribute to active, effective prevention of the development of cancers and non-communicable diseases.
- The major weakness of this cohort is that it may show selection bias because only subjects who voluntarily visited our center were included in the study.

INTRODUCTION

In recent decades, the prevalence of non-communicable diseases (NCDs), such as malignancy, metabolic disease, and cardiovascular disease, has rapidly increased in Korea.[1,2] To address this problem, a comprehensive approach that accounts for lifestyle, environmental factors and genetic variability is needed, as NCDs are known to be caused by both genetic and environmental factors.[3,4] Precision medicine is emerging as a potential solution.[5,6] This type of medicine categorizes individuals into different subgroups based on their susceptibility to disease and then focuses on individuals in whom interventions will be helpful.[7] One of the major components of efficient prevention and early detection of NCDs is thus identifying high-risk populations among those in a pre-disease or asymptomatic stage.

The Seoul National University Hospital (SNUH) Healthcare System Gangnam Center provides comprehensive medical check-ups and screening, and nearly 20,000 people visit this center each year. A data warehouse, HEALTH-WATCH®, was built as a prototype database for this cohort study within the healthcare research institute from the start of the center in 2003.[8] Using HEALTH-WATCH®, we have published many articles that have primarily been focused on cancer screening and metabolic diseases.[8-19] In 2013, we re-organized the cohorts to support our research and started collecting blood samples to analyze the genetic factors involved in NCDs. We then summarized and integrated our clinical and genetic data in the Health and Prevention Enhancement (H-PEACE) study.

The H-PEACE study was designed to investigate the pre-disease stage of NCDs, which helps us assess an individual's risk of NCDs and establish personalized prevention strategies. Through analyses of longitudinal data from comprehensive questionnaires and clinical and laboratory tests, the H-PEACE study can expand our knowledge of NCD prevention, as well as our knowledge of high-risk populations, to improve the early detection of NCDs. The findings can inform treatment priorities and therapeutic guidelines related to NCDs. Finally,

the H-PEACE study can contribute to enhancing public health and improving quality of life.

COHORT DESCRIPTION

Participants

The H-PEACE study collected data from 91,336 Koreans aged 45.5 ± 11.7 years (50,507 men and 40,829 women) who received a health check-up between October 2003 and December 2014 at the SNUH Gangnam Center (IRB No H-1311-031-531). We prospectively collected blood samples from 17 455 of those individuals (9396 men and 8059 women) from October 2013 to form the Gene-environmental interaction and phenotype (GENIE) study, a subcohort of the H-PEACE study. The informed consent and study protocols were approved (IRB No H-1103-127-357 & H-1505-047-671). Furthermore, we obtained data for genome-wide single nucleotide polymorphism (SNP) arrays using 6579 donated blood samples. The flowchart of the H-PEACE study subjects is illustrated in Fig 1, and the distribution of the participants in Korea in Fig 2.

Repeated measurement and active and passive follow-up

Annual health exams are mandatory for all workers under the Industrial and Safety Law in Korea. In the H-peace study, there are two follow-up systems (passive and active follow-up system) for tracking the development of the new diseases and death of the participants. We have annually updated the survival data by linking to National Death Certificates through requests to Statistics Korea. Under the passive follow-up systems, over 99% of participants are followed up their new cancer development and death annually.

Moreover, the study participants are followed up their new disease status by re-visit to center for repeated measurement and assessment system. This active follow-up system allows the subject to voluntarily visit to our center. The health checkup system in Korea has two systems in which the National Health Insurance Corporation (NHIC) pays for participants'

basic health examination fee once every two years or in which the health examination participants pays all expenses for their health examination fees. The latter program includes more precise health examination testing that individuals want, and all costs must be paid by individuals. We are under the latter system and the health examination participants have to pay for all the health checkups under their voluntary visit. The participants who have undergone health examinations at our center may receive health checkups afterwards at different centers that are under partial coverage of the NHIC without revisiting our center. Although the health check-up services conducted in our center includes a variety of tests that the subject wants, including the basic health check-up program conducted by the NHIC, we have to make a lot of effort in order for the subject to return to our center because the former health check-up system covered by the NHIC is free-paid for the basic health check-up program. We are using a number of ways to encourage our return to our center. To encourage participants, we provide reminder calls every year and send health care information about people's next health check-up date via phone and letter. So the eligibility for follow-up assessment depends on the participants' voluntary re-visit need and self-payment availability.

Under this active follow-up system, we did a repeated measurement of risk factors and an assessment of outcome variables for 46,484 individuals (50.9%) of total 91,336 cohort members participated in the baseline health check-up once within 4 years from the first visit and 74,304 individuals (83.5%) were actively followed up once within 10 years from the first visit. Moreover, 50,049 participants (54.8%) completed the two or three repeated measurements (Fig 1). The median follow-up was 4.04 years (interquartile range [IQR] 2.1 – 6.5).

Of 91,336 cohort members, a total of 17,455 members (19.1%) agreed to blood collection and had several aliquots of blood specimens. Of 17,455 members, 6,579 participants were included in the GENIE study with GWAS information (37.7%) and of them,

5,915 participants (89.9%) completed at least 1st follow-up test within 4 years of the first visit. The median follow-up of the GENIE study was 5.23 years (IQR 2.7 – 8.0).

Data collection

The H-PEACE study consists of core and specific variables (Tables 1 and 2). The core variables were tests performed in all of the enrolled participants at the regular health check-ups. The specific variables were the tests selectively performed in the participants upon request or based on the recommendation of their medical provider according to the participant's symptoms or disease risk factors. These tests were sophisticated and expensive tests that are rarely utilized in other cohort profiles. In Table 3, we describe the representative test for each variable and its participation rate. The tests performed for the covariables and specific variables are described below. All data were collected as part of the comprehensive health check-ups. The flow diagram of health check-ups in our study is illustrated in Fig 3. All study protocols are available upon request at our homepage (http://en-healthcare.snuh.org/HPEACEstudy).

Table 1. Core variables collected at baseline and follow-up in the Health and Prevention Enhancement (H-PEACE) study from 2003 to 2014

Item	List	
Self-record questionnaires	· Demographic and lifestyles questionnaire: socioeconomic variables, including education, household income, marital status, and occupation; past medical history of diabetes, hypertension, dyslipidemia, and cancer; family history of diabetes, hypertension, dyslipidemia, and cancer in first relatives; alcohol consumption and cigarette smoking · Physical activity: Korea-validated version of the International Physical Activity Questionnaire (IPAQ) short form · Dietary factors: Mini-Dietary Assessment Index (MDAI) · Sex-specific questionnaire: International Prostate Symptom score (IPSS) and the International Index of Erectile Function (IIEF-5) for men; reproductive factors for women, including menstrual history (age at menarche, length and regularity of menstrual cycle), menopausal status and age, pregnancy history (including the number/methods of pregnancies and abortions), and breastfeeding · Psychiatric assessment questionnaire: Beck Depression Inventory (BDI) and Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR 16)	
Anthropometric measurements and physical examination	· Systolic and diastolic blood pressure · Height, weight, and waist circumference · Body composition using a multi-frequency bio-impedance analyzer · Physical examination by physician · Eye examination, including visual acuity, ocular tonometry, and slit lamp test by Ophthalmologist · Fundus photography (persons ≥ 35 years old) by Ophthalmologist · Hearing test (persons ≥ 50 years old) by otolaryngologist	
Laboratory blood tests	· Complete blood cell count (WBC, RBC, and platelet count, hemoglobin, hematocrit); · Coagulation profiles (PT, PT-INR, aPTT); · Diabetes profiles (fasting glucose, Hemoglobin A1c); · Liver function profiles (protein/albumin, AST/ALT/Gamma-GTP, total bilirubin); · Kidney function profiles (BUN, creatinine); · Electrolytes (calcium, phosphate, sodium, potassium, chloride, total CO_2); · Lipid profiles (total cholesterol, triglycerides, LDL/HDL cholesterol levels); · Uric acid	
Urine tests	· Protein, glucose, ketone, bilirubin, blood, and pH	
Stool exams	· Occult blood in stool, parasite assay	
Infections and inflammation markers	· HBsAg, Anti-HBs, Anti-HCV, · HAV Ab IgG (peoples < 50 years old) · *H. pylori* IgG Ab	· VDRL (Rapid Plasma Reagin Card Test) · HIV Ag/Ab combo test; · Hs-CRP
Hormones	· Thyroid function tests (TSH, T3, total-T4, Free-T4)	
Tumor markers	· AFP, CEA, · CA19-9	· CA-125 (all females); · PSA (males ≥ 35 years old)
Other clinical tests	· Chest X-ray (digital imaging); · Pulmonary function test; · Electrocardiography; · Gastroendoscopy with pathological reports	· Pap smear and gynecological examination (females ≥ 35 years old) · Mammography (females ≥ 35 years old)

WBC, white blood cell; RBC, red blood cell; PT, prothrombin time; INR, International Normalized Ratio; aPTT, activated partial thromboplastin time; AST, aspartate transaminase; ALT, alanine transaminase; GTP, gamma-glutamyl transferase; BUN, blood urea nitrogen; LDL, low-density lipid; HDL, high-density lipid; HBsAg, hepatitis B surface antigen; Anti-HBs, anti-hepatitis B surface antibody; Anti-HCV, anti-hepatitis C antibody; *H. pylori* IgG Ab, hepatitis A immunoglobulin G antibody; VDRL, venereal disease research laboratory; HIV, human immunodeficiency virus, Hs-CRP, high sensitivity C-reactive protein; TSH, thyroid-stimulating hormone; AFP, alpha-fetoprotein; CA19-9, cancer antigen 19-9; CEA, carcinoembryonic antigen; CA-125, cancer antigen 125; PSA, prostate-specific antigen.

Table 2. Specific variables* collected at baseline and follow-up in the Health and Prevention Enhancement (H-PEACE) study

Specific items	List	
SNP	· SNP variation by genome-wide SNP arrays (Affymetrix platform of Axiom™ Customized Genome-Wide Human Assay) †	
Heart	· Coronary calcium score CT · Echocardiography, treadmill test,	· Coronary CT angiography · Holter monitoring
Stroke	· Lp (a) lipoprotein (a), homocysteine · Echocardiography, carotid Doppler	· Brain MRI, · Brain and carotid MRA
Kidney	· 24-h urinary uric acid, creatinine, urea nitrogen, protein, and microalbumin · 24-h urinary electrolytes including sodium, potassium, calcium, phosphorus, and magnesium · 24-h urinary citrate and oxalate	· Serum cystatin C · Dynamic kidney CT
Intestine	· Abdominal sonography · Colonoscopy with pathologic reports	· Abdomino-pelvic CT
Obesity	· 24-h recall diet questionnaire · Serum insulin levels	· Abdomen and thigh CT (for visceral fat) · Whole body DEXA
Dementia	· Apo-E genotyping · MR double time	· Brain MRI, brain MRA, hippocampus (non-contrast)
Prostate	· Free PSA · Uroflowmetry · Ultrasonography (for residual urine)	· Transrectal sonography · Prostate MRI (non-contrast)
Sex hormones	· Testosterone (males ≥ 60 years old) · Total and free E2, LH, FSH (postmenopausal women after 5 years from menopause)	
Cigarette smoking	· CO levels in exhalation, urine cotinine levels · Laryngoscopy, low-dose screening chest CT	
Bone	· Bone densitometry (males ≥ 50 years old and postmenopausal women) · Site-specific X-ray digital imaging (AP and Lat.) · Site-specific bone CT	
Dental	· Dental, peridental and periodontal examination	
Nutrition	· Plasma vitamin B6 profile (pyridoxal phosphate (PLP), pyridoxic acid (PA)) · Plasma vitamin A and E (HPLC);	· Plasma 25(OH)D; · Vitamin C, B12, and folate levels · Serum selenium, zinc, copper levels
Allergy	· *MAST* (Multiple *allergen* simultaneous *test*) · IgE (PRIST) · Skin prick test (inhalant)	· Methacholine bronchial c*hallenge test* · Histograms of Categorized Shapes (HCS) ear detection · *X-ray of PNS (paranasal* sinus)
Hepatitis	· HBV viral load, HBeAg (quantitation), HBeAb · HCV PCR	· Liver fibrosis scan
Heavy mineral	· Cd, Pb, Hg, As, Al, Se, Cu, Zn, Cr, Co, Mn, Mo	
Other	· HPV genotyping; · CLO test or UBT test for current *H. pylori* infection	

*The specific variables were selected for the health screenees according to the specific test or physician's request.

†GWAS genotyping was performed among the 17 455 cohort participants who provided informed consent and donated blood samples. We categorized the 17 455 participants with blood specimens as the Gene-environmental interaction and phenotype (GENIE) study (a specific subcohort derived from the H-PEACE study). SNP testing was performed for research purposes and was not a routinely performed test.

SNP, single nucleotide polymorphism; CT, computed tomography; MRI, magnetic resonance imaging; MRA, magnetic resonance angiogram; DEXA, dual-energy X-ray absorptiometry; PSA, prostate-specific antigen; LH, luteinizing hormone; FSH, follicle-stimulating hormone; HPLC, high-performance liquid chromatography; PRIST, paper radioimmunosorbent test, PCR, polymerase chain reaction; CLO, Campylobacter-like organism; UBT, urea breath test.

Table 3. Participation rate for the core and specific variables collected in the Health and Prevention Enhancement (H-PEACE) study at baseline

Variables	Study participants N	Participation rate %*
Core variables		
Most core variables	≥ 90 377	≥ 99
Electrocardiography	91 197	99.8
Estimated GFR	91 197	99.8
Stool parasite examination	82 493	90.3
H. pylori IgG Ab	81 142	88.8
Thyroid function test	77 313	84.6
CLO test or UBT test for current *H. pylori* infection	67 986	74.4
Gastroendoscopy with pathologic report	66 451	72.8
Specific variables		
Dental examination	43 232	47.3
Colonoscopy with pathologic reports	46 050	50.4
Abdomen CT with visceral fat CT	34 771	38.1
Abdomen sonography	33 446	36.6
Echocardiography	20 688	22.7
Coronary CT	12 846	14.1
Echocardiography, Treadmill test	9910	10.9
SNP from GWAS	6579	7.2
Bone densitometry (postmenopausal females)	26 376	99.4
E2 levels (postmenopausal females after 5 years from menopause)	4742	49.9
Bone densitometry (males ≥ 50 years)	9488	48.9
T levels (males ≥ 60 years)	6697	52.9
HPV genotyping (females)	12 999	31.8
Pap smear (females ≥ 35 years)	12 951	49.1
Blood collection for the GENIE study	17 455	19.1
SNP variations in the GWAS	6579	37.7†

*Participation rate of all 91 336 H-PEACE cohort members.

†Participation rate of all 17 455 GENIE (Gene-environmental interaction and phenotype) study members (a subcohort of the H-PEACE study).

GFR, glomerular filtration rate; CLO, Campylobacter-like organism; UBT, urea breath test; CT, computed tomography; GWAS, genome-wide association study; SNP, single nucleotide polymorphism.

Structured questionnaires

Self-reported questionnaires were used to obtain socio-demographic data, personal and familial medical history, health-related behaviors (such as smoking status, alcohol consumption, physical activity, dietary behavior), international prostate symptom score (IPSS), female reproductive factors, and psychological status. Physical activity levels were assessed using the validated Korean version of the International Physical Activity Questionnaire (IPAQ) short form and were classified into three categories: inactive, minimally active and health-enhancing physically active (HEPA).[20,21] Dietary habits were evaluated using the mini-dietary assessment index (MDAI), which was validated in Korean.[22] Psychological status was assessed by the Beck Depression Inventory (BDI) or Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR 16).[23]

Physical examinations

Blood pressure was measured using sphygmomanometers with patients in a seated position after a resting period. If systolic blood pressure was ≥140 mm Hg or diastolic blood pressure was ≥90 mm Hg after a rest period and two measurements, we recorded the values and calculated their averages. Height (cm), weight (kg), waist circumference (cm) and body fat composition (%) were measured by trained nurses with participants wearing a lightweight hospital gown and in bare feet. Height and weight were measured using digital scales in a standing position. Waist circumference was obtained by measuring the smallest natural waist circumference area, which is around the umbilicus using a non-stretch tape measure, without any pressure to the body surface during measurements. Percentage of body fat and visceral fat area were estimated with a multi-frequency bio-impedance analyzer with 8-point tactile electrodes (Inbody 720, Biospace co, Seoul, Korea). Comprehensive eye examinations (visual acuity, ocular tonometry, slit lamp test, fundus photography) and hearing tests were also

performed.

Laboratory tests

Blood samples from the antecubital vein were collected after at least ten hours of fasting. The blood parameters assessed included complete blood cell count, fasting blood glucose, glycated hemoglobin (HbA1c), uric acid, blood urea nitrogen, creatinine, total calcium, inorganic phosphorus, glucose, sodium, potassium, chloride, total CO2, total protein, albumin, total bilirubin, AST, ALT, gamma-GTP, total cholesterol, low-density lipid cholesterol (LDL-C), high-density lipid cholesterol (HDL-C), triglycerides, high sensitivity C-reactive protein (hs-CRP) concentration, prothrombin time, aPTT, HBsAg, anti-HBs, HAV Ab IgG, AFP, CA19-9, CEA and PSA in men and CA 125 in women. The laboratory medicine department at the SNUH has been certified by the Korean Society of Laboratory Medicine and participated in the College of American Pathologist's Survey/Proficiency Testing program. Urine tests were performed by stick test using spot urine. Semi-quantitative variables included pH, protein, glucose, ketone, bilirubin and blood. Stool samples were collected to conduct fecal occult blood tests and parasite assays.

Digital imaging and specific diagnostic variables

Chest X-ray was included as a core variable of the H-PEACE study, which included specific information on diagnostic variables (Table 1). First, the GENIE study collected donated blood samples from 17 455 recipients, and genome-wide SNP arrays from 6579 donated blood samples have already been analyzed using the Affymetrix platform (Axiom™ Customized Genome-Wide Human Assay) (Table 3). We plan to increase the available genetic information by 2020 using an SNP array. The PLINK program (Ver. 1.9) and R statistics (Ver. 3.3.0) were used for quality control procedures and data analysis. SNP genotype data combined with clinical data from the H-PEACE study were used to evaluate gene-environment interactions

and to define the related phenotypes.

In total, 12,846 participants underwent coronary CT with a coronary calcium score to assess coronary calcification. The calcium score of the coronary artery is a strong predictor of myocardial infarction and sudden cardiac death.[24] In addition, 34,771 participants underwent visceral fat CT to measure visceral adipose tissue (VAT) and subcutaneous adipose tissue (SAT). The detailed methods used to measure VAT area and SAT area on abdominal fat CT images have been described elsewhere.[17] This quantitative assessment of intra-abdominal adipose tissue is considered the gold standard for measuring the amount of visceral fat.[25] With advanced imaging techniques including echocardiography, brain MRI/MRA, carotid Doppler ultrasound, abdominal ultrasonography and abdominal CT, we elucidated the correlation of visceral obesity with vascular disease using both data in a complementary manner. The core variables of the H-PEACE study including the blood test results, questionnaire findings, and depression scores can contribute to determining the correlation between visceral obesity and metabolic phenotype.

We collected electrocardiogram (ECG) reports from 91,197 consecutive recipients to evaluate the incidence and risk factors of atrial fibrillation, a significant risk factor for stroke. The effects of a bundle branch block or atrioventricular block on stroke and cardiac disease were also analyzed. Furthermore, we collected Holter monitor results to confirm the ECG reports.

Human papillomavirus (HPV) is known to lead to cervical cancer in women.[26] Of the participants in the H-PEACE study, 12,951 women underwent both a liquid-based cervical cytology (SurePath LBC, Becton Dickinson, Franklin Lakes, NJ, USA) and an HPV genotyping test using an HPV DNA chip (MyHPV Chip, Biomedlab Co., Seoul, Korea) for cervical cancer screening.[27] We also included the results of gynecologic sonography, as well as VDRL and HIV tests.

The World Health Organization (WHO) considers *Helicobacter pylori* infection a class I

carcinogen for gastric cancer.[13] The diagnosis of *H. pylori* infection was based on the detection of serum *H. pylori* immunoglobulin G antibody using a kit (*H. pylori*-EIA-Well, Radim, Rome, Italy) that was previously validated in a nationwide Korean sero-epidemiologic study.[28] Furthermore, 66,451 recipients had available upper endoscopy data with pathology results. Serum pepsinogen data and eradication history were also collected.

A total of 46,050 participants in the H-PEACE study received a colonoscopy in our center. All colonoscopies were conducted by board certified endoscopists, and the average adenoma detection rate in recipients 50-70 years old was over 30%.[12] Endoscopes including the CF-H260 and CF-HQ290 series (Olympus, AIZU, Japan) and the EC-450HL5, EC-450WM5, and EC-590ZW series (Fujinon, Saitama, Japan) were used. Histological diagnoses at our center were determined according to the WHO classification of tumors of the digestive system.[29,30] All colonoscopy results and corresponding pathology reports were collected.

We calculated estimated glomerular filtration rate (GFR) using the chronic kidney disease epidemiology collaboration equation and the modification of diet in renal disease formula. We collected GFR data successively to identify the risk factors for chronic kidney disease. We also included spot urine sodium and creatinine ratio and 24-hour urine data including GFR.

Dental exams were conducted in 43,232 participants. Periodontitis is highly prevalent among adults and is one of the most common causes of teeth loss after 40 years.[31-33] Furthermore, periodontitis can induce systemic inflammation, as well as masticatory dysfunction and poor nutritional status.[32] This is the first cohort study to include a large number of dental exams to evaluate the progression of periodontitis. In this study, 43,232 participants received a dental exam; 33.1% had mild to moderate periodontitis, and 2.1% had severe periodontitis. We also included participants' colonoscopy and coronary CT results, including coronary calcification and colon polyps, to assess the correlation between periodontitis and systemic inflammation.

The presence of osteoporosis was determined using bone densitometry in 26,207 women and 9488 men. We measured bone mineral density (BMD) of the lumbar spine (L1-L4) and femur

using dual-energy X-ray absorptiometry (GE Medical, United Kingdom). Based on participants' lowest T scores, normal BMD, osteopenia, and osteoporosis were defined as T scores of −1.0 SD and above, −1.0 to -2.5 SDs, and -2.5 SDs and below, respectively. In this cohort, 26.6% of the participants who underwent bone densitometry were male. Of those, osteopenia was found in 24.75% and osteoporosis was found in 3.21%. The information collected in this cohort also included hormone levels, estrogen levels in women and testosterone levels in men.

FINDINGS TO DATE

The baseline characteristics are summarized in Table 4, and the morbidity rate of the major outcomes is shown in Table 5. The morbidity rates were calculated as the number of findings divided by the number of participants tested. The prevalence of major outcomes (mortality and cancer incidence rate) is provided in Table 6. The H-PEACE study has contributed to the research community by publishing >200 articles since the initiation of the data warehouse, HEALTH-WATCH®, as a prototype in 2003.[8] Our studies have primarily focused on cancer screening and identifying the risk factors and prognosis of metabolic disease.

336
337
338
339

Table 4. Baseline characteristics of the participants in the Health and Prevention Enhancement (H-PEACE) study

	Total (N=91 336)	Men (N=50 507)	Women (N=40 829)	p-value
	N (%)	N (%)	N (%)	
Age (years)				<0.001
< 35	16228 (17.8)	7666 (15.2)	8562 (21.0)	
35 – 44	26963 (29.5)	15139 (30.0)	11824 (29.0)	
45 – 54	27915 (30.6)	15776 (31.2)	12139 (29.7)	
55– 64	14937 (16.4)	8717 (17.3)	6220 (15.2)	
65 +	5293 (5.8)	3209 (6.4)	2084 (5.1)	
Sex				
Men	50509 (55.3)			
Women	40827 (44.7)			
Cigarette smoking status				<0.001
Non-smokers	48 521 (53.1)	11 779 (23.3)	36 742 (90.0)	
Ex-smokers	22 564 (24.7)	20 469 (40.5)	2095 (5.1)	
Current smokers	20 251 (22.2)	18 259 (36.2)	1992 (4.9)	
Regular exercise				<0.001
Inactive	36 343 (39.8)	17 843 (35.3)	18 500 (45.3)	
Minimally active	26 189 (28.7)	16 428 (32.5)	9761 (23.9)	
HEPA	28 804 (31.5)	16 236 (32.1)	12 568 (30.8)	
	N (%)	N (%)	N (%)	
Excess alcohol drinking (>140 g/week)	20 421 (22.3)	18 934 (37.5)	1469(3.6)	<0.001
Education ≥ high school graduate	78 821 (86.3)	45 538 (90.2)	33 283 (81.5)	<0.001
Hypertension	13 279 (14.5)	9003 (17.8)	4276 (10.5)	<0.001
Diabetes	4377 (4.8)	3181 (6.3)	1196 (2.9)	<0.001
Current HBsAg+	3810 (4.2)	2377 (4.7)	1433 (3.5)	<0.001
Current Anti-HCV+	887 (0.97)	492 (0.97)	395 (0.97)	0.946
	Mean (SD)	Mean (SD)	Mean (SD)	
Measurements				
Height (cm)	165.6 (11)	171.7 (5.6)	159.9 (5.2)	<0.001
Weight (kg)	64.3 (12.1)	71.5 (9.1)	55.0 (7.2)	<0.001
Body mass index (kg/m^2)	23.4 (12.1)	24.2 (2.6)	21.5 (2.8)	<0.001
Waist circumference (cm)	83.7 (9.3)	86.5 (7.1)	76.9 (7.5)	<0.001
Skeletal muscle mass (kg)	26.6 (6.0)	30.8 (3.7)	20.6 (2.4)	<0.001
Body fat mass (kg)	26.9 (5.9)	30.8 (3.6)	20.7 (2.3)	<0.001
Visceral fat area (cm^2)*	79.9 (24.2)	82.3 (23.6)	76.1 (24.7)	<0.001
SBP (mmHg)	115.4 (15.9)	118.0 (12.2)	111.2 (13.3)	<0.001
DBP (mmHg)	76.01 (11.9)	79.0 (9.45)	71.36 (9.56)	<0.001
Blood levels				
Fasting blood glucose (mg/dL)	97 (18.6)	100.2 (17.6)	90.7 (16.2)	<0.001
HbA1c (%)	5.56 (0.6)	5.6 (0.6)	5.4 (0.4)	<0.001
Creatinine (mg/dL)	0.84 (0.17)	0.93 (0.13)	0.68 (0.10)	<0.001
Albumin (g/dL)	4.62 (0.25)	4.64 (0.24)	4.58 (0.25)	0.002
Total cholesterol (mg/dL)	191.03 (33.02)	192.5 (33.1)	188.6 (32.8)	0.105
Triglycerides (mg/dL)	101.2 (64.2)	117.9 (68.8)	74.6 (44.9)	<0.001
LDL cholesterol (mg/dL)	119.8 (31.6)	126.3 (31.3)	112.6 (30.0)	<0.001
HDL cholesterol (mg/dL)	54.3 (13.2)	49.4 (9.5)	59.7 (12.6)	<0.001
Hs-CRP (mg/dL)	0.12 (0.26)	0.12 (0.26)	0.10 (0.26)	0.285

*Visceral fat area is defined here as the cross-sectional area of visceral fat found in the abdomen.

HEPA, health-enhancing physically active; HBsAg, hepatitis B surface antigen; Anti-HCV, anti-hepatitis C antibody; SBP, Systolic blood pressure; DBP, Diastolic blood pressure; LDL, low-density lipid; HDL, high-density lipid; Hs-CRP, high sensitivity C-reactive protein.

Table 5. Morbidity rates after active follow-up with specific measurements in the Health and Prevention Enhancement (H-PEACE) study

Test	Definition	Participants tested N	Follow-up years Median (IQR)	Clinical finding* N	
Coronary calcification	· Coronary artery calcium score > 400 at coronary CT	12 846	5.6 (3.2-8.0)	805	
Atrial fibrillation	· ECG and/or ambulatory ECG	91 076	4.1 (2.1-6.5)	446	
Left bundle branch block	· ECG and/or ambulatory ECG	91 076	4.1 (2.1-6.5)	64	
Chronic kidney disease	· Estimated GFR† < 60 mL/min/1.73 m^2	91 197	4.0 (2.1-6.5)	5372	
Visceral obesity	· Men > 10000 cm^2, Women > 8000 cm^2 in abdomen visceral fat CT	34 771	5.0 (3.0-7.4)	23 126	
Fatty liver	· Abdominal sonography and/or abdominal CT	34 771	5.0 (3.0-7.4)	Mild	10 244
				Moderate	5963
				Severe	688
Abnormal thyroid function	· Thyroid hormone measurement	77 313	4.1 (2.2-6.6)	Hypothyroidism	363
				Hyperthyroidism	1260
H. Pylori infection	· *H. pylori* IgG Ab and/or CLO test or UBT test for current *H. pylori* infection	81 142	4.2 (2.3-6.7)	44 304	
High-risk HPV infection	· HPV 16, 18 positive finding in HPV genotyping	12 999	5.0 (3.1-7.2)	1462	
Parasite infection	· Stool examination	82 493	4.1 (2.2-6.6)	*Clonorchis sinensis*	1262
				Ascaris lumbricoides	124
				Trichuris trichiura	256
				Metagonimus yokogawai	412
Premalignant conditions of stomach	· Gastroendoscopy with pathological reports	66 451	4.2 (2.3-6.7)	Atrophic gastritis	23 896
				Intestinal metaplasia	10 187
Malignant gastric cancer	· Gastroendoscopy with pathological reports	66 451	4.2 (2.3-6.7)	299	
Squamous intraepithelial lesion (SIL)	· Pap smear and pathological report	12 951	5.0 (3.1-7.2)	Total SIL	375
				Low grade	315
				High grade	60
Premalignant colonic neoplasms	· Colonoscopy with pathological reports	46 746	4.8 (2.9-7.1)	Adenoma	18 656‡
				Serrated polyp	1178‡
				Carcinoid	150
Malignant colon cancer	· Colonoscopy with pathological reports	46 746	4.8 (2.9-7.1)	262	
Low bone density	· Bone densitometry	35 695	4.8 (2.9-7.2)	Osteopenia	
				Male	2322
				Female	8279
				Osteoporosis	
				Male	305
				Female	1348
Periodontitis	· Dental examination	43 232	4.9 (2.8-7.2)	Total periodontitis	15 204
				Mild to moderate	14 292
				Severe	912

*The number of findings among the participants who were at risk.

[†]The glomerular filtration rate was estimated by the chronic kidney disease epidemiology collaboration (CKD-EPI) equation and the modification of diet in renal disease (MDRD) formula.
[‡]Adenoma and serrated polyp detection rate were calculated in participants 50-75 years old
ECG: electrocardiogram; GFR, glomerular filtration rate; CT, computed tomography; CLO, Campylobacter-like organism; UBT, urea breath test; HPV, human papillomavirus.

Table 6. Major cancer incidence and outcomes based on record linkage to the Nationwide Death Certificate Database in the Health and Prevention Enhancement (H-PEACE) study*

	N of cancer incidence			N of all-cause death†		
	Total	**Men**	**Women**	**Total**	**Men**	**Women**
Total cancer	2814	1402	1412	207	150	57
Thyroid	956	390	566	3	3	0
Stomach	513	337	176	37	29	8
Breast	343	1	342	6	0	6
Colorectal	284	182	102	28	20	8
Lung	186	109	77	45	32	13
Prostate	127	127	0	3	3	0
Kidney	91	71	20	1	1	0
Liver	87	71	16	31	28	3
Gynecology	52	0	52	7	0	7
Pancreas	42	25	17	27	19	8
Brain	39	20	19	0	0	0
Hematologic	39	24	15	5	3	2
Biliary tract	32	26	6	11	9	2
Esophagus	23	19	4	3	3	0

*Median (interquartile range) follow-up years of the 91 336 participants=4.04 (2.13-6.5).
†Since the Nationwide Death Certificate Database does not provide information about cause of death, N of death includes death from not only cancer but also other causes. Among the 91 336 H-PEACE participants, the all-cause mortality rate was 969: 715 men, 254 women.

For subjects with 1-2 adenomas less than 10 mm in colonoscopy for colon cancer screening, they were classified as a low-risk group having low 5-year colon adenoma incidence rates and were recommended to take further colonoscopic screening test after 5 year. For subjects with advanced adenoma, multiple adenoma ≥ 3, or larger adenoma sized ≥10mm in baseline colonoscopy for colon cancer screening, they were classified as a high-risk group having higher incidence rates of advanced adenoma or higher recurrence rates of adenoma and were recommended to take colonoscopic surveillance at 3-year after initial polypectomy [12]. This strategy of colonoscopic surveillance has been reflected in the guidelines for colonoscopic surveillance [34]. In our study, the colonic adenoma incidence rate was 37%. Our center performed a previous study on the prevalence and risks of colorectal adenoma in asymptomatic Koreans aged 40-49 years undergoing screening colonoscopies.[35] In that

study, the prevalence of adenoma was 22.2% in the 40-49 year age group and 32.8% in the 50-59 year age group. This finding is quite consistent with that in our cohort. In the paper, we concluded that the prevalence of adenoma in subjects aged 40–49 years was higher than that in previous studies. Male sex and current smoking habits were among the risk factors.

We also examined the quality of colon cancer screening using metrics such as the adenoma detection rate [36]. Additionally, we evaluated the influence of image-enhanced endoscopy, such as narrow band imaging and Fuginon Intelligent Color Enhancement, using our cohort data [10, 37, 38]. For gastric cancer screening, the effect of screening endoscopy remains controversial, but population-based screening has been undertaken in Korea and Japan. For participants with intestinal metaplasia in gastroendoscopy, we also classified them into high-risk group and recommended an annual endoscopic screening, based on study results that people with strong risk factors such as male and an older age can quickly find early-staged endoscopically-treatable gastric cancer by taking annual gastroendoscopic screening [11]. In our cohort, 54% had *H. pylori* infection, and 36% had atrophic gastritis. In a nationwide multicenter study in Korea performed by our center for the prevalence of *H. pylori* infection, the seropositive rate of *H. pylori* was 59.6%.[39] This rate is quite consistent with our study result and consequently explains the relatively high percentage of atrophic gastritis that we observed. Gastric mucosa-associated lymphoid tissue lymphomas (MALToma) and suspicious MALToma lesions were also detected in screening endoscopies, and the prevalence of gastric MALToma was high in middle-aged women.[40] Active endoscopic screening for gastric cancer had the additional benefit of detecting early-stage MALToma.[40, 41] Furthermore, we established that small subepithelial tumors that were incidentally found during upper endoscopy screening showed no size change in subsequent follow-up periods.[37]

Our research focus was not limited to cancer screening. We also examined metabolic

diseases such as non-alcoholic fatty liver disease (NAFLD) and metabolic syndrome. Consistent with our results, the population prevalence of NAFLD in Korea is greater than 25%, like that in many Western countries.[19, 42-44] Although subjects in this cohort are relatively young and lean, lean NAFLD (or non-obese NAFLD) is more common in Asians (including Koreans) than in Western populations. Asians generally have a higher percentage of visceral fat and body fat than people of other races of the same age, sex, and even BMI.[45-47] As visceral obesity is a major risk factor for lean NAFLD, this might have affected the higher prevalence of NAFLD in this cohort, comprised of relatively lean subjects.[45] In addition, the term "fatty liver" in this manuscript includes not only NAFLD but also alcoholic fatty liver disease. The proportion of subjects who drink alcohol heavily in this cohort is as high as 37.5% in men, which might affect the high prevalence of fatty livers that we observe.

We reported that metabolic syndrome was an independent risk factor of silent brain infarction using brain MRI data (OR, 2.18; 95% CI, 1.38-5.82, p=0.001).[48] The prevalences of metabolic syndrome and its components were higher in subjects with silent brain infarctions than in those without. This is the first study to demonstrate an association between metabolic syndrome and silent brain infarction. This finding might help identify healthy people at increased risk for developing silent brain infarction. We also found an effect of body fat distribution on the incidence of NAFLD and reflux esophagitis using abdominal fat CT data.[9,14] We identified a relationship between NAFLD and the risks of coronary heart disease and arterial stiffness based on coronary CT and the cardio-ankle vascular index (CAVI), respectively.[16-19] To elucidate the relationship between NAFLD and the risk of coronary heart disease, we used the coronary artery calcification score as measured by coronary CT. This measurement showed that patients with NAFLD are at increased risk for coronary atherosclerosis (OR, 1.28, 95% CI, 1.04-1.59, p=0.023).[17] In the study on the

association between NAFLD and arterial stiffness, we used CAVI, a new measurement of arterial stiffness. In an age-, sex-, and BMI-adjusted model, NAFLD was associated with a 42% increase in the risk for arterial stiffness, and this increased according to the severity of NAFLD [16]. The effect of physical activity on NAFLD was also demonstrated using data from the questionnaires.[49]

The list and number of publications can be requested on our homepage (http://en-healthcare.snuh.org/HPEACEstudy).

STRENGTHS AND LIMITATIONS

This study has several potential limitations as below. First, our study participants are composed of individuals who voluntarily visited our center and their data collection for the repeated measurements rely on participants' self-paid. Our system of enrolling cohort members includes a potential selection bias**.** Second, self-recorded questionnaires are used to obtain the information of risk factors and past disease histories before the next visit of study participants to our center and this procedure leads to recall bias and response bias. Third, our active follow-up rates in assessing outcome variables and repeatedly measuring risk factors is not high, leading to selection bias. However, since active follow-up rates are increasing every year, at least 87% of follow-up is expected if all subjects are terminated for at least 10-year follow-up. Forth, in our cohort, we used to measure waist circumference as the definition of the smallest area around the belly button below the rib cage and above the hip bone, although the ideal way is to measure the midway between the top of the hip bone and below the rib cage. The former is usually called the natural waist, and we used this former definition for practical reason to reduce intra-individual measurement bias in waist circumference. In health check-up at our center, a lot of participants (about 100-120 health examinees) everyday visit to center for health examination and many health technicians and nurses take measurement of waist and hip circumferences. For each participants, there is a very little chance to take measurement of waist at the time of re-visit by same nurses met at the time of cohort enrollment. The intra-individual measurement bias in measuring waist circumference may be problematic at our center and thus we thought that it was necessary to use the most practical and easy waist measurement together with repeated nurse training. We did a small pilot study for measuring waist by the two methods for 10 men and 16 women. The ideal measuring method (midway measurement) has a limit in consuming time and effort due to difficulty in method itself. Despite of short time and smaller effort in measuring waist, the natural waist

measuring method showed an excellent agreement (intraclass correlation coefficients = 99% in men and 93% in women) with ideal method and there were no shift between obesity groups classified by each method. Fifth, there is no information of cancer histology subtypes for cohort participants ascertained as new cancer development. Individual cancer development in study participants was ascertained with data linkage to the nationwide cancer registry data. Korea cancer registry data includes ICD10 code, T-code (Topography) and M-code (Morphology in primary cancer sites). For example, among the stomach cancer types, 51.3% were adenocarcinomas, 22.1% were tubular adenocarcinomas, and only 4.9% were histologic NOS [50]. Future studies will attempt to merge the individual data associations of M-code and T-code. Finally, the definition of a disease status could be changed during the follow up periods, as new guidelines are declared. For instance, guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults opened at November 2017 [51] recommending a new definition for hypertension. Our manuscript was written before the declaration was done. We have the raw data of blood pressure and the definition can be applied to the study population. But for those who had already been under anti-hypertensive medication at the point of enrollment in our cohort, had followed the previous guidelines for indication of treatment. This makes it difficult to re-classify these patients according to the new guidelines. In the future follow up period, we will give a big attention to clearly declare the definition of disease we are applying according to the period of terms.

A major strength of the H-PEACE study is its structured and organized database, which not only includes data widely used in medical check-ups but also data from sophisticated high-quality advanced examinations, such as colonoscopy, brain MRI, abdominal CT and abdominal fat CT. Particularly in our study, comprehensive biomarkers related to NCDs, including cancer antigens and infection markers, were included, which have not been

included in the medical check-ups of other cohort profiles. It has only been about 10~15 years since the comprehensive health check-up program is actively performed in Korea and in the past decade, a large proportion of people participated in the check-up due the support from the affiliated company's welfare policy. This might result in the relatively young population enrollment in our study population. But since the health check-up program is getting more general, more elderly population is taking the health check-up. New enrollment is another future target study of our center. Since the enrolled cohort population is relatively young, it is possible studying the preclinical disease stage and its final long term outcomes. This might be the characteristics of our cohort that we could design a lot of prediction models for the non-communicable disease by using a data from the preclinical stages. As we obtained a large sample size, we could confirm disease statuses as well as numerous phenotypes of NCDs. Furthermore, the study protocols were conducted and monitored with intensive quality control procedures. Additional strengths include the active and passive follow-ups and the ability to obtain complete data, covering deaths and incidental cancer cases even among those who discontinued visiting our center. We believe that these data could contribute to active, effective prevention against the development of cancers and NCDs.

COLLABORATION

The genotype (SNP) data from the GENIE study and the epidemiologic data from the other cohorts are available to researchers after a quality control process has been completed. Potential collaborators can access the dataset after receiving approval from the SNUH Gangnam Center institutional review board. Applications can be submitted via our homepage. Further information is available at the H-PEACE study website (http://en-healthcare.snuh.org/HPEACEstudy).

ACKNOWLEDGEMENTS

The authors thank You Mi Kwon and Si Won Choi for their technical assistance and all the contributions of the H-PEACE study investigators

COMPETING INTERESTS

J.E.L is an employee of DNALink, Inc.

FUNDING

This work was supported by the SK Telecom Research Fund through Seoul National University Hospital [34-2013-0390 and 34-2014-0210 to S.J.C.] and the Seoul National University Hospital Research Fund [04-2011-0700 to S.J.C.].

CONTRIBUTORS

Changhyun Lee (CL), Eun Kyung Choe (EKC), Ji Min Choi (JMC), Sue K. Park (SKP), Su Jin Chung (SJC), Min-Sun Kwak (MK), Jong-Eun Lee (JL), Joo Sung Kim (JSK) and Sang-Heon Cho (SC) were involved in the conception and design of the study.

CL, EKC, JMC, Yunji Hwang (YH), Young Lee (YL), MK, JL, SJC and Boram Park (BP) contributed to development of methods and data collection.

CL, EKC, JMC, YL, SKP, SC and BP were involved in data analysis and interpretation.

CL, EKC, JMC, SKP and SC drafted the manuscript.

All the authors have critically revised the article and approved the final version and the findings.

ETHICS APPROVAL

The H-PEACE study was approved by the ethics committee of Seoul National University Hospital.

DATA SHARING STATEMENT

Researchers can apply for the dataset after receiving approval from the SNUH Gangnam Center institutional review board. The H-PEACE study website provides information on the application process (http://en-healthcare.snuh.org/HPEACEstudy).

REFERENCES

1 Beaglehole R, Bonita R, Horton R, et al. Priority actions for the non-communicable disease crisis. *Lancet* 2011;377:1438-47.

2 Ministry of Health and Welfare of Korea, Korea Centers for Disease Control and Prevention. *2010 Korea Health statistics.* Seoul: Ministry of Health and Welfare of Korea 2011.

3 Hunter DJ. Gene-environment interactions in human diseases. *Nat Rev Genet* 2005;6:287-98.

4 Renz H, von Mutius E, Brandtzaeg P, et al. Gene-environment interactions in chronic inflammatory disease. *Nat Immunol* 2011;12:273-7.

5 Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015;372:793-5.

6 Jameson JL, Longo DL. Precision medicine — personalized, problematic, and promising. *N Engl J Med* 2015;372:2229-34.

7 Elliott P, Peakman TC, UK Biobank. The UK biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol* 2008;37:234-44.

8 Park MJ, Kim DH, Lim SH, et al. Features of gastric neoplasm detected during the screening examination. *Gut Liver* 2007;1:33-9.

9 Kim D, Chung GE, Kwak MS, et al. Body fat distribution and risk of incident and regressed nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2016;14:132–8.e4.

10 Chung SJ, Kim D, Song JH, et al. Comparison of detection and miss rates of narrow band imaging, flexible spectral imaging chromoendoscopy and white light at screening colonoscopy: a randomised controlled back-to-back study. *Gut*

2014;63:785-91.

11 Chung SJ, Park MJ, Kang SJ, et al. Effect of annual endoscopic screening on clinicopathologic characteristics and treatment modality of gastric cancer in a high-incidence region of Korea. *Int J Cancer* 2012;131:2376-84.

12 Chung SJ, Kim YS, Yang SY, et al. Five-year risk for advanced colorectal neoplasia after initial colonoscopy according to the baseline risk stratification: a prospective study in 2452 asymptomatic Koreans. *Gut* 2011;60:1537-43.

13 Chung SJ, Lim SH, Choi J, et al. Helicobacter pylori serology inversely correlated with the risk and severity of reflux esophagitis in helicobacter pylori endemic area: a matched case-control study of 5,616 health check-up Koreans. *J Neurogastroenterol Motil* 2011;17:267-73.

14 Chung SJ, Kim D, Park MJ, et al. Metabolic syndrome and visceral obesity as risk factors for reflux oesophagitis: a cross-sectional case-control study of 7078 Koreans undergoing health check-ups. *Gut* 2008;57:1360-5.

15 Chung GE, Kim D, Kim W, et al. Non-alcoholic fatty liver disease across the spectrum of hypothyroidism. *J Hepatol* 2012;57:150-6.

16 Chung GE, Choi SY, Kim D, et al. Nonalcoholic fatty liver disease as a risk factor of arterial stiffness measured by the cardioankle vascular index. *Medicine (Baltimore)* 2015;94:e654.

17 Kim D, Choi SY, Park EH, et al. Nonalcoholic fatty liver disease is associated with coronary artery calcification. *Hepatology* 2012;56:605-13.

18 Kwak MS, Yim JY, Kim D, et al. Nonalcoholic fatty liver disease is associated with coronary artery calcium score in diabetes patients with higher HbA1c. *Diabetol Metab Syndr* 2015;7:28.

19 Choi SY, Kim D, Kim HJ, et al. The relation between non-alcoholic fatty liver disease

and the risk of coronary heart disease in Koreans. *Am J Gastroenterol* 2009;104:1953-60.

20 Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* 2003;35:1381-95.

21 Chun MY. Validity and reliability of Korean version of international physical activity questionnaire short form in the elderly. *Korean J Fam Med* 2012;33:144-51.

22 Kim WY, Cho MS, Lee HS. Development and validation of mini dietary assessment index for Koreans. Korean. *J Nutr* 2003;36:83-92.

23 Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. *Clin Psychol Rev* 1988;8:77-100.

24 Wong ND, Hsu JC, Detrano RC, et al. Coronary artery calcium evaluation by electron beam computed tomography and its relation to new cardiovascular events. *Am J Cardiol* 2000;86:495-8.

25 Seidell JC, Bakker CJ, van der Kooy K. Imaging techniques for measuring adipose-tissue distribution--a comparison between computed tomography and 1.5-T magnetic resonance. *Am J Clin Nutr* 1990;51:953-7.

26 Bosch FX, Manos MM, Muñoz N, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. *J Natl Cancer Inst* 1995;87:796-802.

27 Kim MJ, Kim JJ, Kim S. Type-specific prevalence of high-risk human papillomavirus by cervical cytology and age: data from the health check-ups of 7,014 Korean women. *Obstet Gynecol Sci* 2013;56:110-20.

28 Yim JY, Kim N, Choi SH, et al. Seroprevalence of Helicobacter pylori in South Korea. *Helicobacter* 2007;12:333-40.

29 Hamilton SR, Aaltonen LA. *WHO Classification of Tumours of the Digestive System*. 3rd ed. Geneva: IARC 2000.

30 Bosman FT, Carneiro F, Hruban RH, et al. *WHO Classification of Tumours of the Digestive System*. 4th ed. Geneva: IARC 2010.

31 Eke PI, Dye BA, Wei L, et al. Update on prevalence of periodontitis in adults in the United States: NHANES 2009 to 2012. *J Periodontol* 2015;86:611-22.

32 White DA, Tsakos G, Pitts NB, et al. Adult Dental Health survey 2009: Common oral health conditions and their impact on the population. *Br Dent J* 2012;213:567-72.

33 Murray H, Locker D, Kay EJ. Patterns of and reasons for tooth extractions in general dental practice in Ontario, Canada. *Commun Dent Oral Epidemiol* 1996;24:196-200.

34 Lieberman DA, Rex DK, Winawer SJ, et al. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology* 2012;143:844-57.

35 Chung SJ, Kim YS, Yang SY, et al. Prevalence and risk of colorectal adenoma in asymptomatic Koreans aged 40-49 years undergoing screening colonoscopy. *J Gastroenterol Hepatol* 2010;25:519-25.

36 Kang HY, Kim D, Kim HJ, et al. The relationship between colonoscopy procedure order and adenoma detection rates: a prospective study. *J Clin Gastroenterol* 2015;49:683-9.

37 Song JH, Kim SG, Chung SJ, et al. Risk of progression for incidental small subepithelial tumors in the upper gastrointestinal tract. *Endoscopy* 2015;47:675-9.

38 Chung SJ, Kim D, Song JH, et al. Efficacy of computed virtual chromoendoscopy on colorectal cancer screening: a prospective, randomized, back-to-back trial of Fuji intelligent color enhancement versus conventional colonoscopy to compare adenoma miss rates. *Gastrointest Endosc* 2010;72:136-42.

636 39 Lim SH, Kwon JW, Kim N, et al. Prevalence and risk factors of Helicobacter pylori
637 infection in Korea: nationwide multicenter study over 13 years. *BMC Gastroenterol*
638 2013;24;13:104. doi: 10.1186/1471-230X-13-104
639 40 Yang HJ, Lee C, Lim SH, et al. Clinical characteristics of primary gastric lymphoma
640 detected during screening for gastric cancer in Korea. *J Gastroenterol Hepatol*
641 2016;31:1572-83.
642 41 Yang HJ, Lim SH, Lee C, et al. Management of suspicious mucosa-associated
643 lymphoid tissue lymphoma in gastric biopsy specimens obtained during screening
644 endoscopy. *J Korean Med Sci* 2016;31:1075-81.
645 42 Fan JG, Kim SU, Wong VW. New trends on obesity and NAFLD in Asia. *J Hepatol*
646 2017;67:862-73.
647 43 Bae JC, Cho YK, Lee WY, et al. Impact of nonalcoholic fatty liver disease on insulin
648 resistance in relation to HbA1c levels in nondiabetic subjects. *Am J Gastroenterol*
649 2010;105:2389-95.
650 44 Chang Y, Jung HS, Yun KE, et al. Cohort study of non-alcoholic fatty liver disease,
651 NAFLD fibrosis score, and the risk of incident diabetes in a Korean population. *Am J*
652 *Gastroenterol* 2013;108:1861-8.
653 45 Kim D, Kim WR. Nonobese fatty liver disease. *Clin Gastroenterol Hepatol*
654 2017;15:474-85.
655 46 Alberti KG, Zimmet P, Shaw J, et al. The metabolic syndrome--a new worldwide
656 definition. *Lancet* 2005;366:1059-62.
657 47 Nazare JA, Smith JD, Borel AL, et al. Ethnic influences on the relations between
658 abdominal subcutaneous and visceral adiposity, liver fat, and cardiometabolic risk
659 profile: the International Study of Prediction of Intra-Abdominal Adiposity and Its

Relationship With Cardiometabolic Risk/Intra-Abdominal Adiposity. *Am J Clin Nutr* 2012;96:714-26.

48 Kwon HM, Kim BJ, Lee SH, et al. Metabolic syndrome as an independent risk factor of silent brain infarction in healthy people. *Stroke* 2006;37:466-70.

49 Kwak MS, Kim D, Chung GE, et al. Role of physical activity in nonalcoholic fatty liver disease in terms of visceral obesity and insulin resistance. *Liver Int* 2015;35:944-52.

50 Korea Ministry of Health and Welfare, Korea National Cancer Center. 2001 Annual Report of the Korea Central Cancer Registry. Korea National Cancer Center 2003. https://www.e-crt.org/upload/pdf/crt-36-19.pdf

51 Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol* 2017;1097:41519-1

FIGURE LEGENDS

Fig 1. Study flow diagram of the Health and Prevention Enhancement (H-PEACE) study

Fig 2. Regional distribution of the Health and Prevention Enhancement (H-PEACE) study in Korea

Fig 3. Flow diagram of health check-ups in the Health and Prevention Enhancement (H-PEACE) study

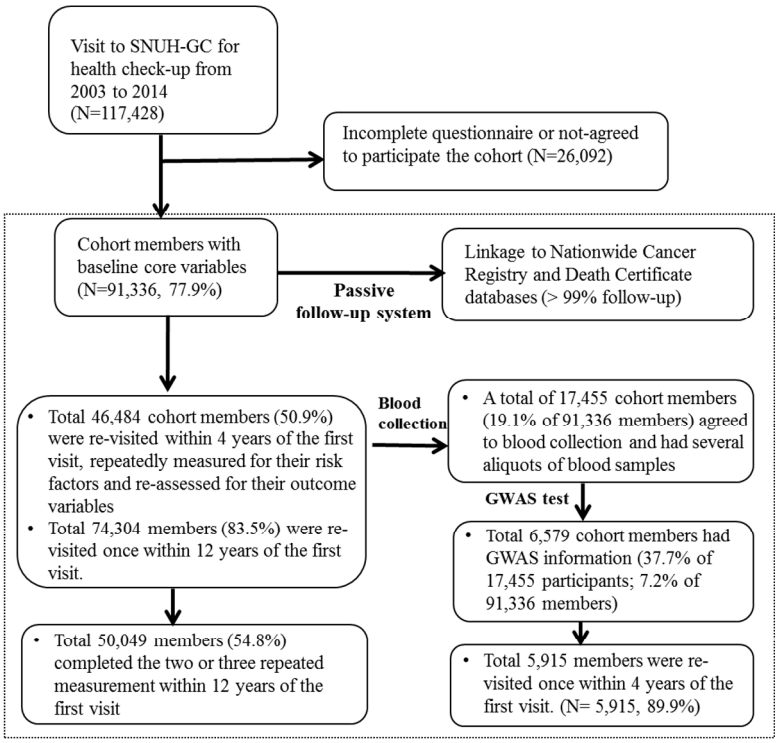
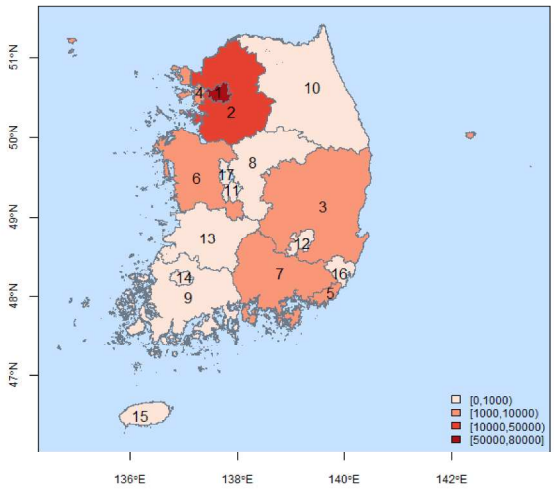


Fig 1. Study flow diagram of the Health and Prevention Enhancement (H-PEACE) study

254x190mm (300 x 300 DPI)



	Location	N
1	Seoul (Capital, Metropolitan)	60,301
2	Gyeonggi-do	18,383
3	Gyeongsangbuk-do	1,564
4	Incheon (Metropolitan)	1,478
5	Busan (Metropolitan)	1,445
6	Chungcheongnam-do	1,400
7	Gyeongsangnam-do	1,278
8	Chungcheongbuk-do	769
9	Jeollanam-do	737
10	Gangwon-do	727
11	Daejeon (Metropolitan)	692
12	Daegu (Metropolitan)	686
13	Jeollabuk-do	580
14	Gwangju (Metropolitan)	488
15	Jeju-do	338
16	Ulsan (Metropolitan)	297
17	Sejong (Metropolitan)	20
	Total	91,183

Fig 2. Regional distribution of the Health and Prevention Enhancement (H-PEACE) study in Korea

254x190mm (300 x 300 DPI)

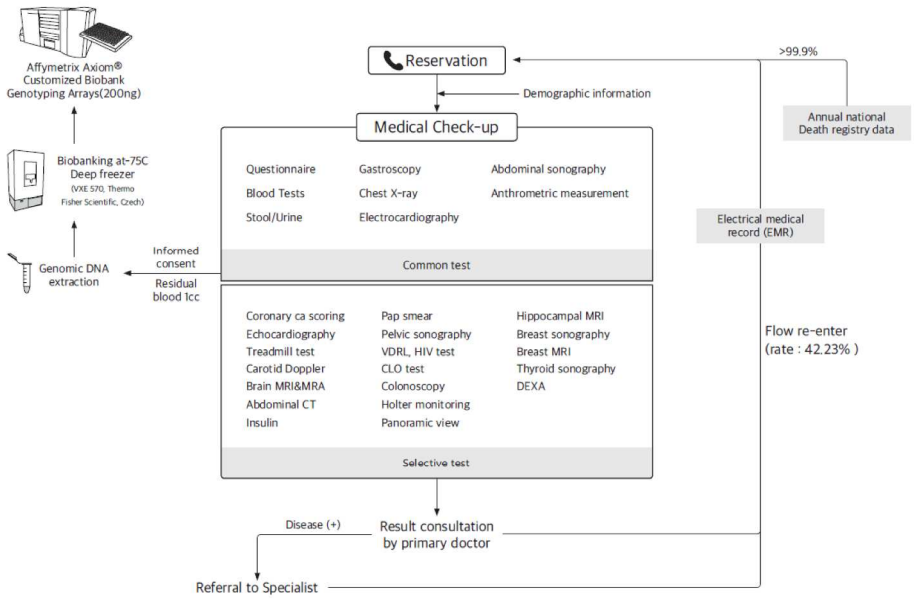


Fig 3. Flow diagram of health check-ups in the Health and Prevention Enhancement (H-PEACE) study

254x190mm (300 x 300 DPI)

STROBE (Strengthening The Reporting of OBservational Studies in Epidemiology) Checklist

A checklist of items that should be included in reports of observational studies. You must report the page number in your manuscript where you consider each of the items listed in this checklist. If you have not included this information, either revise your manuscript accordingly before submitting or note N/A.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Section and Item	Item No.	Recommendation	Reported on Page No.
Title and Abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/Rationale	2	Explain the scientific background and rationale for the investigation being reported	6
Objectives	3	State specific objectives, including any prespecified hypotheses	6-7
Methods			
Study Design	4	Present key elements of study design early in the paper	8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	8
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	8
		Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	N/A
		Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants	N/A
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed	N/A
		Case-control study—For matched studies, give matching criteria and the number of controls per case	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	9-15

Section and Item	Item No.	Recommendation	Reported on Page No.
Data Sources/ Measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	9-15
Bias	9	Describe any efforts to address potential sources of bias	31
Study Size	10	Explain how the study size was arrived at	9
Quantitative Variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	N/A
Statistical Methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	N/A
		(*b*) Describe any methods used to examine subgroups and interactions	N/A
		(*c*) Explain how missing data were addressed	N/A
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	N/A N/A N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	22-24
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	Fig 1
Descriptive Data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 4
		(b) Indicate number of participants with missing data for each variable of interest	Fig 1
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	8
Outcome Data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	Table 5,6
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	N/A
		Cross-sectional study—Report numbers of outcome events or summary measures	N/A

Section and Item	Item No.	Recommendation	Reported on Page No.
Main Results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	N/A
		(*b*) Report category boundaries when continuous variables were categorized	Table 4,5
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other Analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	N/A
Discussion			
Key Results	18	Summarise key results with reference to study objectives	31
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	31
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	31
Generalisability	21	Discuss the generalisability (external validity) of the study results	31
Other Information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	32

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Once you have completed this checklist, please save a copy and upload it as part of your submission. DO NOT include this checklist as part of the main manuscript document. It must be uploaded as a separate file.