Exercise and incidence of myocardial infarction, stroke, hypertension, type 2 diabetes and site-specific cancers: prospective cohort study of 257 854 adults in South Korea

Youngwon Kim, Stephen Sharp, Semi Hwang, Sun Ha Jee

ABSTRACT

Objective The objective of this study was to examine the longitudinal associations of exercise frequency with the incidence of myocardial infarction, stroke, hypertension, type 2 diabetes and 10 different cancer outcomes.

Design A prospective cohort study.

Setting Physical examination data linked with the entire South Korean population’s health insurance system: from 2002 to 2015.

Participants 257 854 South Korean adults who provided up to 7 repeat measures of exercise (defined as exercises causing sweat) and confounders.

Primary outcome measures Each disease incidence was defined using both fatal and non-fatal health records (a median follow-up period of 13 years).

Results Compared with no exercise category, the middle categories of exercise frequency (3–4 or 5–6 times/week) showed the lowest risk of myocardial infarction (HR 0.79; 95% CI 0.70 to 0.90), stroke (HR 0.80; 95% CI 0.73 to 0.89), hypertension (HR 0.86; 95% CI 0.85 to 0.88), type 2 diabetes (HR 0.87; 95% CI 0.84 to 0.89), stomach (HR 0.87; 95% CI 0.79 to 0.96), lung (HR 0.80; 95% CI 0.71 to 0.91), liver (HR 0.85; 95% CI 0.75 to 0.98) and head and neck cancers (HR 0.76; 95% CI 0.63 to 0.93; for 1–2 times/week), exhibiting J-shaped associations. There was, in general, little evidence of effect modification by body mass index, smoking, alcohol consumption, family history of disease and sex in these associations.

Conclusions Moderate levels of sweat-inducing exercise showed the lowest risk of myocardial infarction, stroke, hypertension, type 2 diabetes, stomach, lung, liver and head and neck cancers. Public health and lifestyle interventions should, therefore, promote moderate levels of sweat-causing exercise as a behavioural prevention strategy for non-communicable diseases in a wider population of East Asians.

INTRODUCTION

Prevention and control of non-communicable diseases are a contemporary global public health priority. At present, 40 million deaths per year, which accounts for nearly 70% of total deaths globally, are attributable to non-communicable diseases. Moreover, the number of deaths due to non-communicable diseases, such as cardiovascular disease (CVD), hypertension, diabetes and cancer, has increased dramatically over the past few decades, although age-standardised CVD and cancer rates as well as systolic blood pressure levels have declined. However, trends in these disease traits have varied across different populations, particularly with less favourable changes observed in East Asian populations compared with Western populations. For example, the prevalence of diabetes has increased more rapidly, while the age-standardised prevalence of CVD and systolic blood pressure levels have fallen less steeply in East Asians in comparison with Westerners.

In addition, adults in East Asia tend to have higher prevalence of physical inactivity, which is one of the four target behaviours (including unhealthy diet, tobacco use and harmful use of alcohol) that have been set as the global focus to reduce the risk of non-communicable diseases. The beneficial impacts of increased physical activity on...
of over 500 maintaining national health examination programmes mandatory for all South Koreans to take part in the national South Korean population’s healthcare utilisation; it is which manages and maintains information on the entire NHIS is a single health insurance system in South Korea, between 2002 and 2003 made available by the NHIS. The total cholesterol levels, exhibiting different patterns of of South Korean adults with multiple repeated measures site-specific cancers, using a large-scale prospective cohort of incident non-communicable diseases, such as myocar- tionships between exercise frequency and various types of this research was to explore the dose–response rela- in understanding the relationships between physical activity and non-communicable diseases. Moreover, the dose–response relationship between physical activity and various non-communicable disease outcomes has remained unclear in East Asians. Therefore, the purpose of this research was to explore the dose–response relationships between exercise frequency and various types of incident non-communicable diseases, such as myocardial infarction, stroke, hypertension, type 2 diabetes and site-specific cancers, using a large-scale prospective cohort of South Korean adults with multiple repeated measures of exercise frequency and other risk markers.

METHODS
Study design and participants
This study is based on data from the National Health Insurance Service—Health Screening (NHIS-HEALS) cohort dataset, which is a nationally representative random sample (stratified by 2 groups of sex (males and females), 18 groups of age ranges (less than 1 year, 1–4 years, every 5 years between 5 and 79 years, and more than 80 years), 3 groups of employment status (insured employees, self-employed individuals and medical aid beneficiaries) and 41 groups of income levels (upper 20% for insured employees, lower 20% for insured self-employed individuals and the lowest level for medical aid beneficiaries)) of over 500 000 South Korean adults aged 40–79 years between 2002 and 2003 made available by the NHIS. The NHIS is a single health insurance system in South Korea, which manages and maintains information on the entire South Korean population’s healthcare utilisation; it is mandatory for all South Koreans to take part in the national health insurance system. The NHIS is also responsible for maintaining national health examination programmes involving data from general health examinations of all insured employees, self-employed individuals and medical aid beneficiaries aged over 40 years; it is recommended for them to perform the health examination at least every 2 years. The health examination involves collection of information on body composition, blood profiles, blood pressure, self-reported lifestyles, self-reported physician-diagnosed disease and self-reported family history of disease.

The NHIS-HEALS cohort includes a wide variety of information collected between 2002 and 2015: health examination data and demographic and eligibility data (eg, inpatient and outpatient hospital records, medical bill, health insurance and medical aid beneficiaries). In the present analysis, we used health examination data collected between 2002 and 2008 to define the exercise frequency and all confounders. There was a change in the type of self-report methods in 2009; hence, health examination data collected in or after 2009 were not consid- ered in the analysis due to the inability to harmonise variables. However, we used full follow-up data accrued from 2002 until 2015.

Exposure
The primary exposure variable of this study was exercise frequency, assessed using questionnaires administered during the health examinations. The specific question asked was “How many times per week do you engage in exercise that causes sweating?” Participants were asked to choose only one of the following 5 possible answers: none, 1–2 times/week, 3–4 times/week, 5–6 times/week and almost every day.

Outcomes
In the present study, we evaluated 14 incident disease outcomes, namely, myocardial infarction; stroke, hyper- tension; type 2 diabetes mellitus; and stomach, colon, rectum, lung, liver, head and neck, pancreatic, kidney, gall bladder and oesophagus cancers. Participants’ inpatient and outpatient hospital records (ie, non-fatal status) and death records (ie, fatal status) obtained through linkage with Statistics Korea were both classified according to the International Classification of Disease (ICD)-10 codes to classify different incidence types (online supplementary table 1). Additionally, blood pressure (eg, systolic ≥140 mm Hg, diastolic ≥90 mm Hg) and fasting glucose levels (eg, ≥126 mg/dL), both of which were measured during physical examinations, were used in conjunction with physicians’ diagnosis information and ICD-10 codes to define incident hypertension and type 2 diabetes, respectively. Each incident disease outcome was defined as the first occurrence of either non-fatal or fatal respective disease cases. Incident disease cases were adjudicated using hospital and death records collected through 31 December 2015. The median follow-up was 13.0 years (IQR 10.2–11.3 years).

Other covariates
The following covariates were included as confounders in the analyses: sex, body mass index (weight in kilograms


Open access
of interaction terms in the fully adjusted models for each incident disease outcome. Visual inspections of log–log plots provided support for the assumptions of proportional hazards for all covariates. A sensitivity analysis where incident disease cases occurring during the first 2 years of follow-up were removed was performed to address reverse causality. Analyses were performed in Stata/SE V.14 (StataCorp LP, College Station, Texas, USA).

**Patient and public involvement**

Neither patients nor members of the public were involved in this study.

**RESULTS**

Of an initial sample of 512,190 individuals, 74,931 had missing data on at least one of the model covariates, and 179,405 had self-reported physician-diagnosed heart attack, stroke, hypertension (additionally, systolic ≥140 mm Hg or diastolic ≥90 mm Hg), diabetes (additionally, fasting glucose levels ≥126 mg/dL) or cancer at baseline, respectively. Excluding these individuals resulted in a final sample for analysis of 257,854 individuals (figure 1).

Individuals provided up to seven measures of exercise frequency and each confounder (ie, baseline plus six repeated measures). Participants’ characteristics at baseline are summarised in table 1. Online supplementary table 2 summarises participants’ characteristics at each repeat assessment. Individuals in the categories of 1–2, 3–4 or 5–6 times/week of exercise were slightly younger, but showed higher proportions of family history of disease

---

**Figure 1** A flow diagram. Note: ‘N’ indicates numbers of total observations (ie, participants who provided repeated measures are treated as separate observations) and ‘n’ indicates numbers of unique participants at baseline. Data without missingness and prevalence of major diseases were used to create final analysis datasets for different incident disease outcomes; while the number of unique participants at baseline is the same for all incident disease outcomes, the total number of observations varied due to the nature of time-updated covariate analyses (ie, censoring of subsequent time-updated covariates when an incident disease case occurs before the end date of repeated measures). Data were obtained from a prospective cohort, which has been established by linking physical examination data of over half a million South Korean adults (2002–2015) with the entire South Korean population’s health insurance system. NHIS, National Health Insurance Service.
Table 1  Characteristics of the participants at baseline

<table>
<thead>
<tr>
<th>Variables</th>
<th>All (n=257854)</th>
<th>Exercise frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>None (n=148284)</td>
<td>1–2 times/week (n=62923)</td>
</tr>
<tr>
<td>Sex, %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men (%)</td>
<td>50.5</td>
<td>63.3</td>
</tr>
<tr>
<td>Women (%)</td>
<td>49.5</td>
<td>36.7</td>
</tr>
<tr>
<td>Age, years</td>
<td>50.7 (8.7)</td>
<td>49.0 (7.5)</td>
</tr>
<tr>
<td>Body mass index, kg/m2</td>
<td>23.5 (2.8)</td>
<td>23.6 (2.7)</td>
</tr>
<tr>
<td>Systolic blood pressure, mm Hg</td>
<td>116.8 (11.2)</td>
<td>117.1 (11.0)</td>
</tr>
<tr>
<td>Diastolic blood pressure, mm Hg</td>
<td>73.4 (7.9)</td>
<td>73.8 (7.8)</td>
</tr>
<tr>
<td>Fasting glucose levels, mg/dL</td>
<td>90.2 (12.3)</td>
<td>90.4 (12.2)</td>
</tr>
<tr>
<td>Total cholesterol, mg/dL</td>
<td>197.0 (36.7)</td>
<td>197.7 (36.2)</td>
</tr>
<tr>
<td>Family history of heart disease, stroke or hypertension, %</td>
<td>12.2</td>
<td>13.9</td>
</tr>
<tr>
<td>Family history of cancer, %</td>
<td>14.2</td>
<td>15.3</td>
</tr>
<tr>
<td>Family history of diabetes, %</td>
<td>6.3</td>
<td>7.2</td>
</tr>
<tr>
<td>Smoking status, %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never (%)</td>
<td>68.1</td>
<td>59.1</td>
</tr>
<tr>
<td>Previously (%)</td>
<td>8.4</td>
<td>12.0</td>
</tr>
<tr>
<td>Currently (%)</td>
<td>23.6</td>
<td>28.9</td>
</tr>
<tr>
<td>Alcohol consumption, %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never (%)</td>
<td>58.4</td>
<td>47.6</td>
</tr>
<tr>
<td>2–3 times/month (%)</td>
<td>16.4</td>
<td>21.3</td>
</tr>
<tr>
<td>1–2 times/week (%)</td>
<td>15.8</td>
<td>22.0</td>
</tr>
<tr>
<td>≥3 times/week (%)</td>
<td>9.5</td>
<td>9.1</td>
</tr>
<tr>
<td>Incident myocardial infarction, n (%)</td>
<td>3047 (1.2)</td>
<td>723 (1.1)</td>
</tr>
<tr>
<td>Incident stroke, n (%)</td>
<td>16134 (6.3)</td>
<td>3333 (5.3)</td>
</tr>
<tr>
<td>Incident hypertension, n (%)</td>
<td>120203 (46.6)</td>
<td>30623 (48.7)</td>
</tr>
<tr>
<td>Incident type 2 diabetes, n (%)</td>
<td>50459 (19.6)</td>
<td>10666 (6.5)</td>
</tr>
<tr>
<td>Incident stomach cancer, n (%)</td>
<td>4788 (1.9)</td>
<td>13226 (21.0)</td>
</tr>
<tr>
<td>Incident colon cancer, n (%)</td>
<td>2711 (1.1)</td>
<td>1160 (1.8)</td>
</tr>
<tr>
<td>Incident rectum cancer, n (%)</td>
<td>1494 (0.6)</td>
<td>692 (1.1)</td>
</tr>
<tr>
<td>Incident lung cancer, n (%)</td>
<td>3601 (1.4)</td>
<td>796 (1.3)</td>
</tr>
<tr>
<td>Incident liver cancer, n (%)</td>
<td>2620 (1.0)</td>
<td>680 (1.1)</td>
</tr>
<tr>
<td>Incident pancreas cancer, n (%)</td>
<td>864 (0.3)</td>
<td>205 (0.3)</td>
</tr>
</tbody>
</table>

Continued
and lower proportions of never smoking or drinking alcohol, compared with those in the categories of none or almost every day of exercise. Across the seven time points (online supplementary figure 1), the proportion of individuals who reported no exercise decreased while the proportion who reported 1–2 or 3–4 times/week of exercise increased; there were no noticeable changes for the categories of 5–6 times/week or almost every day of exercise.

Overall, J-shaped associations were found between exercise frequency and incident myocardial infarction, stroke, hypertension and type 2 diabetes. HRs for these diseases were lowest in the middle categories of exercise frequency (eg, 3–4 or 5–6 times/week) (figure 2). There were no associations for the most frequent exercise category (eg, almost every day) with the incidence of myocardial infarction, stroke and type 2 diabetes. J-shaped associations were also found for incident stomach, lung, liver and head and neck cancers (figure 3). Higher exercise frequencies (eg, 1–2, 3–4 times/week and almost every day) were associated with lower hazards of incident stomach cancer. No statistical significance was observed for incident colon, rectum, pancreas, kidney, gallbladder and oesophagus cancers. Crude event rates per 100 000 person years in the middle categories of exercise frequency were relatively lower for incident rectum, and oesophagus cancers, but higher for incident pancreas, kidney and gallbladder cancers. Cox regression models with no adjustment for confounders (online supplementary figure 2) and a sensitivity analysis (online supplementary figure 3) in which incident cases occurring in the first 2 years of follow-up were removed both revealed nearly identical patterns of associations as the main analyses.

Figure 4 shows comparisons of results that showed statistical significance for multiplicative interaction terms between exercise frequency and each incident disease outcome. Strong J-shaped associations for incident hypertension were identified at each level of body mass index. J-shaped associations of exercise frequency with incident hypertension were strong only in the more favourable levels of smoking (eg, never, previously) and alcohol consumption (eg, never, 2–3 times/month, 1–2 times/week); no or weak associations were identified in the most harmful level of smoking (eg, current smokers) and alcohol consumption (eg, ≥3 times/week). J-shaped associations were evident at all levels of family history of CVD and sex for incident hypertension, and sex for incident type 2 diabetes. Exercise frequency was associated with incident lung cancer in non-obese individuals, but there was no evidence of association in obese individuals. All comparisons stratified by each potential effect modifier are presented in online supplementary figures 4 and 5.

**DISCUSSION**

This is the first investigation examining the prospective associations of exercise with various incident non-communicable disease outcomes using multiple repeated measures of covariates in East Asian populations. We identified J-shaped associations of sweat-inducing exercise with incident myocardial infarction, stroke, hypertension, type 2 diabetes, stomach, lung, liver and head and neck cancers, with the greatest benefits being observed in the middle categories of exercise frequency (eg, 3–4 or 5–6 times/week): 1–2 times/week for head and neck cancer. These findings provide two important clinical and public health implications. First, prevention and management of non-communicable diseases in East Asians may benefit considerably from employing an exercise promotion approach in the context of combined non-communicable disease prevention. Mechanism research indicates that CVD and type 2 diabetes have similar biological pathways relating to exercise, so an integrated prevention approach can be applied to control and manage...
these two diseases at a minimum. Moreover, regular participation in exercise can induce favourable changes in intermediate cardiometabolic risk markers, which are important predictors of typical non-communicable diseases. Hence, promoting exercise has great potential to act as an integrative behavioural strategy for preventing and controlling various non-communicable diseases simultaneously in East Asian populations.

Second, individuals who engage in exercise 3–4 or 5–6 times/week, rather than every day, may be able to reduce their risk of developing myocardial infarction, stroke, hypertension, type 2 diabetes, stomach, lung and liver cancers: 1–2 times/week for head and neck cancer. Similar J-shaped associations between high intensity exercise (eg, running) and CVD risk have also been reported in previous cohort studies of Western and Japanese adults. Nevertheless, the present study as well as previous research found that the risk of developing cardiovascular events in individuals who had the highest level of exercise was not noticeably higher compared with those who had the lowest level of exercise. No previous research in East Asians has found such J-shaped relationships between exercise or physical activity and other incident disease outcomes such as hypertension and different type of cancers. However, previous meta-analyses of cohort studies comprising predominantly Westerners found leisure-time physical activity to have curvilinear (but not J-shaped) associations with the incidence of type 2 diabetes, and linear associations with the incidence of hypertension and various site-specific cancers (liver, lung, head and neck, kidney, colon, rectal, bladder, gastric cardia, breast, endometrial, colorectal and prostate).
myeloid leukaemia, myeloma, oesophageal adenocarcinoma. 36 While additional research is needed to confirm the J-shaped associations of exercise with various incident diseases in other samples of East Asians, findings of this research provide a strong rationale for the development and implementation of public health policies and clinical trials aimed at promoting a moderate level of sweat-causing exercise to minimise the risk of myocardial infarction, stroke, hypertension, type 2 diabetes, stomach, lung, liver and head and neck cancers.

Another finding of this research is that associations of sweat-inducing exercise with hypertension were modified by body mass index, smoking, alcohol consumption, family history of CVD and sex: lung cancer by body mass index and type 2 diabetes by sex. Notably, exercise frequency was not associated with hypertension in individuals who are smokers or drinking alcohol ≥3 times/week (except for 3–4 times/week of exercise). This observation provides some evidence that the harmful impacts of smoking or binge drinking on hypertension may not be offset completely by exercise. This, in turn, appears to advocate for the need of implementing a combined hypertension prevention strategy targeting promotion of exercise in conjunction with smoking cessation and reductions in alcohol consumption in East Asians. 13

For lung cancer, the null associations in individuals with body mass index ≥25 may be indicative of potential residual confounding through reported bias in smoking behaviours. Nonetheless, there was little evidence for effect modification for other disease comparisons, highlighting the importance of promoting exercise for the prevention of various non-communicable diseases in individuals at different levels of body mass index, smoking, alcohol consumption, family history of disease and sex.

This study has several notable strengths. First, we used data from a large prospective cohort study in which exercise and other risk markers were assessed on multiple occasions (up to seven times). Nearly 84% and 5% of the

Figure 3  Associations of exercise frequency with various incident cancer outcomes. Cox regression models with age as the underlying timescale were adjusted for sex, body mass index, systolic blood pressure, fasting glucose levels, total cholesterol levels, family history of cancer, smoking status and alcohol consumption. Crude rates are per 100 000 person years. ‘N’ indicates numbers of total observations (ie, participants who provided repeated measures are treated as separate observations) and ‘n’ indicates numbers of unique participants at baseline. Data were obtained from a prospective cohort, which has been established by linking physical examination data of over half a million South Korean adults (2002–2015) with the entire South Korean population’s health insurance system.
Moreover, we examined the dose–response relationship using repeated measures of exposure and confounders. Compelling evidence indicates that the risk of regression dilution can be reduced by using repeated measures of exposure and confounders. Moreover, we examined the dose–response relationship of exercise frequency with a wide variety of specific types of incident non-communicable disease outcomes simultaneously using inpatient and outpatient diagnosis data as well as mortality data. The large sample size (n=257854) is another strength.

This study has some limitations. Findings of this study may not be generalisable to adult populations of other ethnic origins. Due to the observational nature of this research, no strong causal inference can be drawn about the exercise-incident disease relationships. In addition, the accuracy of hospital admission records is uncertain, although the accuracy of death records from Statistics Korea was found to be 92% in previous research. No information about medication use was available in the cohort data, so we could not use it as a potential confounder and another condition when defining disease status (eg, hypertension, type 2 diabetes) at both baseline and follow-up. Furthermore, no exercise duration was assessed; hence, inference was made purely based on exercise frequency. Moreover, ICD-10 codes for sex-specific cancers (eg, prostate and breast cancers) were masked due to the data management policy set forth by the NHIS, so it was not possible to examine such cancers in the present study. The lack of data on diet, which is another behavioural risk marker for non-communicable diseases, is another limitation. Moreover, a sizeable proportion (n=74931; 14.6%) of individuals were excluded due to the missing information on the covariates. Another limitation is that the measurement methods to assess the covariates were not standardised across the different medical institutes participating in the NHIS-HEALS cohort.

CONCLUSION

Individuals who engaged in sweat-inducing exercise around 3–6 times/week (as opposed to every day) generally had the lowest risk of developing myocardial infarction, stroke, hypertension, type 2 diabetes, stomach, lung, liver and head and neck cancers. These findings were generally applicable to different subpopulations as stratified by body mass index, smoking, alcohol consumption, family history of disease and sex. Public health and lifestyle interventions should promote a moderate level of sweat-inducing exercise as a behavioural strategy for prevention and control of non-communicable diseases in a wider population of East Asians.

Contributors YK designed this study, performed statistical analysis and drafted an initial version of the manuscript. SS, S-mH and SHJ all contributed to conceptualising the study idea and developing the analytical plans, and provided assistance with statistical analysis. All authors critically reviewed, approved the final version of the manuscript and agreed to be responsible for all facets of this work.

Funding This work was supported by the National R&D Program for Cancer Control, Ministry of Health & Welfare, Republic of Korea (Grant 1631020 to SHJ). The Korean Health Technology R&D Project, Ministry of Health & Welfare, Republic of Korea (Grant HI14C2686 to SHJ) and the Medical Research Council (Grant MC_UU_12015/1 to SS). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests None declared.
REFERENCES


