scientific reports

Check for updates

OPEN

Sex differences in predicting dyslipidemia using polygenic risk score with fatty liver index and fibrotic nonalcoholic steatohepatitis index

Sei Kim & Hae Young Yoo✉

Non-alcoholic fatty liver disease (NAFLD) and non-alcoholic steatohepatitis (NASH) are recognized risk factors for dyslipidemia. Current prediction models that rely solely on dyslipidemia polygenic risk score (PRS) have certain limitations. We aimed to validate simple indexes for NAFLD and NASH as predictors of dyslipidemia using the PRS. This study utilized cohort data from an urban population-based dataset comprising 48,263 South Koreans. The incidence of dyslipidemia was higher in men than in women (32.4% and 27.8%; $p < 0.001$). The PRS model predicted dyslipidemia more accurately in men (AUROC [95% confidence intervals]: 0.645 [0.636–0.754]). Notably, integrating the fatty liver index (FLI) and fibrotic NASH index (FNI) with the PRS model resulted in the highest accuracy in diagnosing dyslipidemia, particularly in men (AUROC [95% confidence intervals]: 0.704 [0.698–0.711]). In conclusion, a predictive model combining the PRS with FLI and FNI was validated. This model offers more accurate predictive value for diagnosing dyslipidemia, particularly in East Asian men. Thus, our study has the clinical potential for identifying high-risk individuals and determining preventive measures for dyslipidemia in a sex-specific manner.

Keywords Dyslipidemia, Coronary artery disease, Polygenic risk score, Non-alcoholic fatty liver disease, Steatohepatitis

Dyslipidemia is characterized by elevated blood concentrations of total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), and triglycerides (TG), along with decreased high-density lipoprotein cholesterol (HDL-C) levels[1]. Dyslipidemia gradually develops over several years, resulting in coronary artery disease (CAD), which contributes to a substantial proportion of global mortality[2]. In the Asia-Pacific region, CAD and dyslipidemia have emerged as major health concerns with increasing incidence rates over the past few decades[3]. Dyslipidemia differs distinctly according to the sex of the patient, with previous studies from East Asia having indicated sex-based disparities in the age of onset, prevalence, and mortality of CAD[4,5]. In the last decade, the polygenic risk score (PRS) has been used to assess an individual's lifelong susceptibility to complex diseases, such as dyslipidemia and CAD, to support diagnosis and decision-making regarding treatments[6]. Additionally, the examination of sex differences in the context of CAD and its implications is currently the central focus of PRS research[7].

Non-alcoholic fatty liver disease (NAFLD) is the abnormal hepatic condition with excessive fat accumulation without excessive alcohol consumption and non-alcoholic steatohepatitis (NASH) is the inflamed liver or ballooning hepatocytes in addition to fat droplets. NAFLD is a recognized risk factor for dyslipidemia, similar to NASH, which can exacerbate the risk of dyslipidemia[8,9]. The fatty liver index (FLI) serves as a simple and highly predictive tool for NAFLD diagnosis, whereas the fibrotic NASH index (FNI) is an accurate, cost-effective, and non-invasive score for NASH prediction based on simple laboratory test results[10,11]. Because PRS implementation in clinical practice has limitations, a study on PRSs for CAD reported a significant improvement in classification when combined with conventional risk factors[12]. To the best of our knowledge, no study has evaluated the effectiveness of the PRS in combination with an index for these risk factors. Thus, the effect of the PRS on dyslipidemia beyond the FLI or FNI remains unknown and may be sex-specific, highlighting the need for further investigation.

Department of Nursing, Chung-Ang University, Seoul 06974, Republic of Korea. ✉email: hyoo@cau.ac.kr

Although predictive models have shown a certain degree of accuracy, most have been designed based on European populations, thereby raising concerns about their applicability to non-European populations, including Asian populations[13]. Hence, this study aimed to evaluate the performance of a PRS model derived from an Asian population in predicting dyslipidemia independently, in conjunction with other indexes in distinct male and female cohorts. We hypothesized that the development of a predictive PRS model incorporating the FLI and FNI would enhance the ability to predict dyslipidemia. Additionally, we expected that the outcomes would reveal sex-based disparities, indicating varying model effectiveness between men and women.

Results

Study population

A total of 48,263 individuals, comprising 17,064 men and 31,199 women, met all inclusion criteria and were included in the analysis (Supplementary Fig. S1). Supplementary Table S1 summarizes the characteristics of each group. Interestingly, both the FLI and FNI, along with systolic but not diastolic blood pressure, were significantly higher in men than in women ($p < 0.001$, $p < 0.001$, $p < 0.001$, and $p = 0.937$, respectively). Fasting glucose, TC, LDL-C, and TG levels showed sex-based differences among the metabolic profiles, whereas the liver function test results showed sex-based differences in alanine transaminase (ALT), aspartate transaminase (AST), alkaline phosphatase (ALP), gamma-glutamyl transferase (GGT), and platelet values (Supplementary Table S1).

PRS and dyslipidemia risk prediction

The prevalence of dyslipidemia was significantly higher in men than in women ($p < 0.001$). However, the PRS related to dyslipidemia showed no sex differences ($p = 0.952$) (Supplementary Table S1). In both male and female cohorts, the distribution of PRS among cases of dyslipidemia showed a rightward shift, indicating higher PRS than those of controls (both $p < 0.001$) (Supplementary Fig. S2). Supplementary Figure S3 shows the distribution of cases and controls according to PRS quartiles in the male and female cohorts (both $p < 0.001$). Regarding the risk of dyslipidemia based on PRS, the high PRS group had a higher odds ratio (OR) (up to 2.456) than that of the moderate PRS group in the male cohort only. In the female cohort, the top 1% PRS group showed a risk OR of 1.538 compared with the moderate PRS group (Table 1).

Correlation of PRS and risk factors

Supplementary Figure S4 shows how closely FLI and FNI variations in the male and female cohorts correlated with the PRS for dyslipidemia. For both men and women, the FLI and FNI showed a significant moderate correlation ($r = 0.425$, $p < 0.001$; $r = 0.44$, $p < 0.001$, respectively). In contrast, the FLI and PRS showed a very weak correlation in both male and female groups ($r = 0.089$, $p < 0.001$; $r = 0.084$, $p < 0.001$, respectively), as did the FNI and PRS ($r = 0.055$, $p < 0.001$; $r = 0.055$, $p < 0.001$, respectively). Glycated hemoglobin (HbA1c) levels and platelet counts were significantly correlated in both cohorts ($r = 0.997$, $p < 0.001$ and $r = 0.999$, $p < 0.001$, respectively) (Supplementary Tables S2 and S3).

Dyslipidemia incidence risk based on the FLI or FNI subgroup

As shown in Table 2, the participants were classified into three subgroups according to rule-out standard scores and diagnostic favorability in accordance with previous studies[10,11]. Individuals with lower FLI or FNI scores in rule-out zones were used as the reference group. The groups with higher FLI and FNI scores had a significantly higher risk of dyslipidemia than that of their respective reference groups in both male and female cohorts. A higher FLI indicated a higher OR of dyslipidemia incidence than a higher FNI in both men and women in each subgroup (Table 2).

			Model 1		Model 2	
	High PRS group	Reference group	OR	95% CI	OR	95% CI
Men	99–100%	45–55%	2.343	1.7063–3.2165	2.456	1.7802–3.3890
	98–100%		2.315	1.8294–2.9308	2.401	1.8881–3.0523
	95–100%		1.974	1.6667–2.3385	2.069	1.7409–2.4599
	90–100%		1.670	1.4508–1.9223	1.758	1.5222–2.0304
	80–100%		1.498	1.3241–1.6956	1.573	1.3862–1.7856
	55–100%		1.373	1.2269–1.5373	1.419	1.2646–1.5921
Women	99–100%	45–55%	1.566	1.2291–1.9952	1.538	1.2033–1.9669
	98–100%		1.362	1.1345–1.6357	1.351	1.1223–1.6266
	95–100%		1.409	1.2372–1.6054	1.391	1.2192–1.5871
	90–100%		1.417	1.2729–1.5781	1.412	1.2665–1.5741
	80–100%		1.304	1.1865–1.4332	1.296	1.1778–1.4259
	55–100%		1.190	1.0919–1.2972	1.180	1.0814–1.2872

Table 1. Stratification of risk using the PRS for dyslipidemia. Model 1 is unadjusted, and Model 2 is adjusted for age and BMI. OR, odds ratio; CI, confidence interval; BMI, body mass index.

	Index		Case/Control No. (%)	OR	95% CI	*p*-value
Men	FLI	< 30	1,336/7,020 (16.0/84.0)	1	Reference	
		30–60	2,274/3,487 (39.5/60.5)	4.541	4.143–4.978	< 0.001
		≥ 60	1,915/1,032 (65.0/35.0)	16.079	14.127–18.301	< 0.001
	FNI	≤ 0.10	1,617/6,535 (19.8/80.2)	1	Reference	
		0.10–0.30	2,849/4,168 (40.6/59.4)	2.633	2.444–2.837	< 0.001
		≥ 0.30	1,059/836 (55.9/44.1)	4.622	4.143–5.157	< 0.001
Women	FLI	< 30	5,359/19,096 (21.9/78.1)	1	Reference	
		30–60	2,478/2,886 (46.2/53.8)	3.646	3.381–3.932	< 0.001
		≥ 60	829/551 (60.1/39.9)	7.665	6.678–8.799	< 0.001
	FNI	≤ 0.10	4,673/16,591 (22.0/78.0)	1	Reference	
		0.10–0.30	3,187/5,022 (38.8/61.2)	2.070	1.965–2.192	< 0.001
		≥ 0.30	806/920 (46.7/53.3)	2.632	2.374–2.917	< 0.001

Table 2. Associations of clinical scores with incident dyslipidemia. OR, odds ratio; CI, confidence interval; BMI, body mass index; FLI, fatty liver index; FNI, fibrotic NASH index.

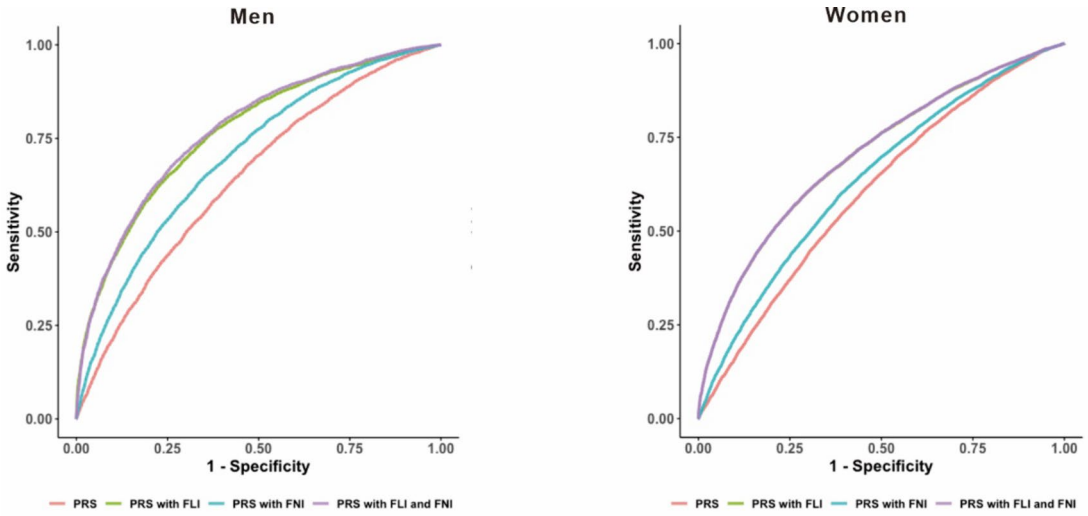

Fig. 1. Receiver Operating Characteristic Curve of the Prediction Model Based on the PRS.
PRS, polygenic risk score; FLI, fatty liver index; FNI, fibrotic NASH index; BMI, body mass index.

Performance of the PRS with the FLI and FNI

Using blood lipid concentrations as a reference, the PRS model demonstrated predictive performance for diagnosing dyslipidemia in both men and women, with an area under the receiver operating characteristic (AUROC) curve of 0.645 and 0.605 (95% confidence intervals [CIs], 0.636–0.654 and 0.598–0.612), respectively. The receiver operating characteristic (ROC) curves of the PRS model in men and women had sensitivities of 67.7% and 61.4%, specificities of 53.3% and 54.4%, and correct classification accuracies of 58.0% and 54.4%, respectively. The PRS model showed a significantly better performance in men than in women ($p < 0.001$). The combination of the FLI with the PRS model showed a significant increase in the AUROC, indicating better performance in both men and women (both $p < 0.001$). The PRS model beyond the FNI showed a significant improvement in the AUROC in both men and women (both $p < 0.001$). The highest AUROC value was observed when combining the FLI and FNI with the PRS model in both men and women within the same cohort (AUROC: 0.774 and 0.704, respectively) (Fig. 1 and Supplementary Table S4).

Discussion

This study examined the incidence of dyslipidemia by analyzing blood lipid concentrations and individuals' genetic susceptibility, while also considering comprehensive liver-related indexes. The PRS for dyslipidemia separately correlated with the FLI and FNI in middle-aged men and women. Our final model, which comprised a combination of the PRS, FLI, and FNI, was significantly effective in predicting dyslipidemia risk in East Asian male and female cohorts.

We found a substantial proportion of dyslipidemia cases in men, but not in women, similar to a meta-analysis that included the other ethnic populations[14]. However, genetic susceptibility to dyslipidemia risk based on the PRS showed no sex-based differences. Interestingly, the risk of dyslipidemia was significantly higher in the high PRS group than in the intermediate PRS group, especially in men. Furthermore, the predictive model based on the PRS showed better performance in men than in women. Similarly, Huang et al. showed that the risk of developing CAD based on genomic susceptibility is greater in men than in women[15]. This can be attributed to variations in genetic heritability between sexes and the impact of these variations on the association between PRS and the incidence of specific diseases such as cardiovascular diseases[16]. A previous meta-analysis identified only 12 lipid-related single-nucleotide polymorphisms (SNPs) with heterogeneous sex effects among more than 10,000 individuals, implying that genetic architecture differs greatly based on the individual's sex[17]. Moreover, lipid metabolism exhibits sexual dimorphism, which is likely influenced by hormonal and lipid metabolic differences that extend beyond genetic differences[18].

Our PRS, derived from a European population-based genome-wide association study, showed moderate predictive power in an East Asian population, although the effect sizes of SNPs, allele frequency, and linkage disequilibrium related to lipid profiles varied across races[19]. For Type 2 diabetes mellitus, the cross-racial PRS shows a good predictive value for risk classification[20]. Additionally, our PRS, based on the binominal incidence of dyslipidemia, has more predictive power than the PRS based on a single lipid index, such as TG or LDL-C levels, as previously reported[21]. However, Aulchenko et al. reported conflicting findings, showing that a PRS derived from TC levels alone has lower performance in males than in our male cohort[22]. Consequently, our PRS will primarily stratify individuals into risk groups for dyslipidemia, independent of single lipid concentrations.

The concurrent prevalence of dyslipidemia with NAFLD and NASH is 69% and 72%, respectively, as reported in a global cohort of >8 million people[23]. Incorporating the combined PRS into a risk model consisting of conventional clinical features improves its discriminative ability[24]. Herein, we found that each fatty liver-related index, FLI and FNI, contributed to a noteworthy modest improvement in the predictive performance of the PRS beyond the contribution of traditional risk factors, such as age and body mass index (BMI). Abnormal levels of TG, LDL-C, TC, and HDL-C are significantly correlated with GGT, a component of the FLI[25]. The FLI improved predictive performance more than the FNI using the PRS model, which may be explained by the fact that severe steatohepatitis is not associated with worse dyslipidemia in patients with NAFLD[26]. In the present study, the integration of both indexes with the PRS risk model provided the best dyslipidemia prediction ability.

Our PRS model combined with the FLI and FNI showed better dyslipidemia predictive performance in the male cohort than in the female cohort. NAFLD is more prevalent in women only after menopause, indicating sex differences among the major risk factors, with estrogen playing a protective role[27]. The genetic variants of patatin-like phospholipase domain-containing 3 (*PNPLA3*) are known to interact with estrogen receptor-α agonists, which contributes to progressive fatty liver diseases risk in menopausal women[28]. Additionally, NASH is an age- and sex-dimorphic disease, with both its prevalence and severity caused by obesity and an individual's lifestyle[29]. Once NAFLD is established, the risk of NASH progression is comparable between men and women[30]. Multiple pathogenic factors, including inflammation, fibrosis, the gut–liver axis, and insulin sensitivity, affect sex differences in NAFLD and NASH[31]. In addition, the livers of men and women have distinct metabolic characteristics that are modulated by sex-specific regulators[27]. FLI-defined NAFLD has a stronger association with dysmetabolic state in women than in men[32]. A recent study found that the ideal threshold values for visceral fat area in NAFLD were higher in lean and overweight/obese men than in women[33]. The hazard ratios clearly differed between male and female participants, indicating that a high FLI predicts the development of hypertension[34]. Moreover, a sex-specific NASH index might enhance the predictive ability tailoring predictions to the distinct metabolic profiles of men and women with regards to hepatic metabolism, bile acid synthesis, and cytochrome P-450 enzymes[35]. Future studies should consider premenopausal and postmenopausal women separately.

This study has several noteworthy findings. First, the discovery of the value of the FLI and FNI in predicting lipid disorders lies in their ability to be integrated with the PRS, beyond conventional risk factors. Second, it offers robust validity by encompassing a large sample derived from an East Asian population. Moreover, our approach used a sex-specific PRS to evaluate performance, thereby capturing the likelihood of differential genetic influences between men and women. However, this study has some limitations that require consideration. First, the use of case-control design limits the ability to determine causality between NAFLD and dyslipidemia, highlighting the necessity of a longitudinal study to improve our understanding of these relationships. Second, we did not consider sex chromosomes in the genetic information of the individuals. Third, although FLI and FNI are validated and non-invasive surrogate markers, they are not equivalent to imaging-based diagnostics, which offer higher accuracy for diagnosing NAFLD and NASH. Finally, the exclusive focus on the Korean population necessitates caution when generalizing these findings to diverse ethnic populations. Further studies are required for additional external validation of the model in cohorts of various races.

In conclusion, the findings of this study revealed significantly more dyslipidemia cases in men than in women within the Korean population. A considerable increase in the incidence risk of dyslipidemia was observed in individuals with a high PRS compared with those with an intermediate PRS, particularly in middle-aged men. Nonetheless, no significant differences in PRS were observed between male and female cohorts. For a more accurate prediction of the risk of dyslipidemia among East Asian individuals, NAFLD and NASH indexes should

be integrated with the PRS for dyslipidemia. Our model showed a significantly better predictive ability among East Asian men than among women. This study has significant implications for clinical nursing, as it implies that sex-based differential screening and guidance are necessary to support early screening and prevention of dyslipidemia. Moreover, this approach provides valuable evidence for identifying high-risk populations and could help tailor preventive guidelines and individualized nursing interventions to mitigate the disease, given the increasing prevalence of dyslipidemia in East Asia and globally.

Methods

Study population

The population-based cohort known as the KoGES_Health Examinee (HEXA) comprised urban-dwelling individuals recruited from the Korean Genome and Epidemiology Study (KoGES). These individuals were examined during the first follow-up period, between 2012 and 2014. A more detailed description of the HEXA cohort and KoGES has been previously reported[36]. This study was approved by the institutional review boards of Chung-Ang University (approval no.: 1041078-201908-HRBR-236-01) in accordance with the guidelines of the Declaration of Helsinki. All participants provided written informed consent for the study participation.

The disease status of the participants and their families was self-reported through interviews. Anthropometric and clinical data including height, weight, waist circumference, blood pressure, biochemical analysis of lipid profiles, and other biomarker levels, were analyzed. Biochemical traits including glycemic traits (fasting glucose and HbA1c), plasma lipid concentrations (TC, LDL-C, TG, and HDL-C), and liver enzymes (ALT, AST, ALP, GGT, and albumin) were measured[36].

Dyslipidemia cases were defined according to the 5th Korean Guidelines for the Management of Dyslipidemia by the Korean Society of Lipid and Atherosclerosis[1,37]. Individuals aged between 40 and 75 years were considered to have dyslipidemia if they met any of the following criteria: (1) TC levels ≥ 240 mg/dL, (2) LDL-C levels ≥ 160 mg/dL, (3) TG levels ≥ 200 mg/dL, or (4) HDL levels < 40 mg/dL. The control group comprised individuals who did not meet any of these criteria. Individuals who were receiving ongoing treatment for fatty liver disease, dyslipidemia, acute liver disease, or lung cancer; those who excessively consumed alcohol (men: ≥2 times per day, women: ≥1 time per day); and those with missing data were excluded (Fig. 1)[21,38].

Genotyping

Peripheral blood samples from all the participants were genotyped using the Korea Biobank Array (Affymetrix, Santa Clara, CA, USA), referred to as the "Korean Chip" by the National Bank of Korea. The Korean Chip is a customized array optimized specifically for the Korean population, with more than 600,000 variants selected for genome-wide tagging among 833,000 markers[39]. We used genome-wide data from the baseline recruitment, and the sample quality control genotyping call rate was set at < 99%. For quality control of single nucleotide polymorphisms (SNPs), SNPs with a Hardy–Weinberg equilibrium failure *P*-value (1E-6) or a low SNP call rate (< 95%) were excluded. Imputed variants (minor allele frequency < 0.05) were also excluded. Only the SNPs found in HapMap3 were included. Consequently, we analyzed 5,407,270 variants from chromosomes 1 to 22 after genotyping and purification. Only the SNPs found at the HapMap3 site were included.

PRS calculation

We calculated the PRS, a weighted sum of risk alleles, related to dyslipidemia using the "auto" mode of the PRS-continuous shrinkage method (version released on November 3, 2022), which infers the posterior effect sizes of specific SNPs based on high-dimensional Bayesian regression and continuous shrinkage[40]. We used genome-wide association summary statistics from Willer et al. and an external linkage disequilibrium reference panel from the same ancestry, the East Asian LD score panels from the 1,000 Genome Project Phase 3[41,42]. We recalculated the PRS z-scores for subsequent analyses.

Liver-related index calculation

The FLI was calculated using the simple algorithm developed by Bedogni et al. (Equation [Eq] 1)[10]. FLIs of 30 and 60 were used as cutoff scores for NAFLD in the original study[43]. The FNI was calculated as reported by Tavaglioni et al. (Eq. 2)[11]. An FNI score ≤ 0.10 is considered a rule-out zone for fibrotic NASH and has been validated in individuals with metabolic disorders[11].

$$FLI = \frac{e^{(-15.745+0.953\times ln\,TG[mg/dL]+0.139\times BMI[kg/m^2]+0.718\times ln\,GGT[IU/dL]+0.053\times W\bullet C[cm])}}{1+e^{(-15.745+0.953\times ln\,TG[mg/dL]+0.139\times BMI[kg/m^2]+0.718\times ln\,GGT[IU/dL]+0.053\times W\bullet C[cm])}}\times 100 \quad (1)$$

$$FNI = \frac{e^{(-10.33+2.54\times ln\,AST[IU/L]+3.86\times HbA1c[\%]-1.66\times ln\,HDL-C[mg/dL])}}{1+e^{(-10.33+2.54\times ln\,AST[IU/L]+3.86\times HbA1c[\%]-1.66\times ln\,HDL-C[mg/dL])}} \quad (2)$$

Statistical analysis

Baseline characteristics of the HEXA cohort were described for the entire group based on sex using numbers with percentages for categorical variables and means with standard deviations or medians with interquartile ranges (25th – 75th percentiles) for continuous variables, as appropriate. A Kolmogorov–Smirnov test was performed to examine the normal distribution of each continuous variables. Student's *t*-test or Chi-squared test was used to compare the aforementioned variables between male and female participants and to determine the difference in the incidence percentage of cases and controls between the male and female cohorts. Furthermore, we conducted a two-sample Kolmogorov–Smirnov test on the PRS distribution in the case and control groups. The association between the PRS and dyslipidemia prevalence was evaluated using a logistic regression model with adjusted ORs

and 95% CIs. We compared the granular top PRS groups with the intermediate PRS group (45–55%) in terms of ORs, as described in a previous study[7]. Additionally, we compared the high FLI and FNI groups with their respective disease-rule-out zones. Correlation coefficients between the PRS and risk factors were measured using Pearson's correlation analysis to examine their linear relationships. ROC analysis was conducted to evaluate the performance of the PRS prediction model for case and control discrimination. This analysis was conducted separately for male and female cohorts, both in the presence and absence of the FLI and FNI. The AUROC was calculated using a non-parametric method. We also calculated sensitivity, specificity, accuracy, positive and negative predictive values, and positive and negative likelihood ratios. The difference between the prediction models was tested by the DeLong method, which compares AUROCs. General information on age and BMI were included as covariates in all analyses. Statistical analyses were performed using R software version 4.3.0 (R Project for Statistical Computing, Vienna, Austria) and Python version 3.11.3 (Python Software Foundation). Statistical tests were two-sided with an $\alpha = 0.05$.

Data availability

The data supporting the conclusions of this study are available from the National Biobank of Korea upon request under the data access and sharing policy of the National Institute of Health, Republic of Korea (https://biobank.nih.go.kr/eng/cmm/main/mainPage.do). The code for PRScs is available to the public at the following website: https://github.com/getian107/PRScs.

Received: 12 December 2024; Accepted: 3 March 2025
Published online: 06 March 2025

References

1. Arnett, D. K. et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: executive summary: a report of the American college of cardiology/american heart association task force on clinical practice guidelines. *Circulation* **140**, e563–e595 (2019).
2. Khan, M. A. et al. Global epidemiology of ischemic heart disease: results from the global burden of disease study. *Cureus* **12**, e9349 (2020).
3. Sze, R. et al. The burden of cardiovascular disease in Asia from 2025 to 2050: a forecast analysis for East Asia, South Asia, South-East Asia, central Asia, and high-income Asia Pacific regions. *Lancet Reg. Health West. Pac.* **49**, 101138 (2024).
4. Mehndiratta, M. M., Khan, M., Mehndiratta, P. & Wasay, M. Stroke in Asia: geographical variations and Temporal trends. *J. Neurol. Neurosurg. Psychiatry*. **85**, 1308–1312 (2014).
5. Qu, Y. et al. Gender-related differences in clinical characteristics and outcomes of premature coronary artery disease: insight from the FOCUS registry. *J. Interv. Cardiol.* 6762089 (2019). (2019).
6. Lewis, C. M. & Vassos, E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med.* **12**, 44 (2020).
7. Abraham, G., Rutten-Jacobs, L. & Inouye, M. Risk prediction using polygenic risk scores for prevention of stroke and other cardiovascular diseases. *Stroke* **52**, 2983–2991 (2021).
8. Duell, P. B. et al. Nonalcoholic fatty liver disease and cardiovascular risk: A scientific statement from the American heart association. *Arterioscler. Thromb. Vasc Biol.* **42**, 168–185 (2022).
9. Sheka, A. C. et al. *Nonalcoholic Steatohepatitis JAMA* **323**, 1175 (2020).
10. Bedogni, G. et al. The fatty liver index: a simple and accurate predictor of hepatic steatosis in the general population. *BMC Gastroenterol.* **6**, 33 (2006).
11. Tavaglione, F. et al. Development and validation of a score for fibrotic nonalcoholic steatohepatitis. *Clin. Gastroenterol. Hepatol.* **21** (e1), 1523–1532 (2023).
12. Elliott, J. et al. Predictive accuracy of a polygenic risk score-Enhanced prediction model vs a clinical risk score for coronary artery disease. *JAMA* **323**, 636–645 (2020).
13. Fitipaldi, H. & Franks, P. W. Ethnic, gender and other sociodemographic biases in genome-wide association studies for the most burdensome non-communicable diseases: 2005–2022. *Hum. Mole Genet.* **32**, 520–532 (2022).
14. Riazi, K. et al. The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis. *Lancet Gastroenterol. Hepatol.* **7**, 851–861 (2022).
15. Huang, Y. et al. Sexual differences in genetic predisposition of coronary artery disease. *Circ. Genom Precis Med.* **14**, e003147 (2021).
16. Vaura, F., Palmu, J., Aittokallio, J., Kauko, A. & Niiranen, T. Genetic, molecular, and cellular determinants of sex-specific cardiovascular traits. *Circ. Res.* **130**, 611–631 (2022).
17. Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
18. Sakkers, T. R. et al. Sex differences in the genetic and molecular mechanisms of coronary artery disease. *Atherosclerosis* **384**, 117279 (2023).
19. Graham, S. E. et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679 (2021).
20. Liu, J. et al. Polygenic risk score, lifestyles, and type 2 diabetes risk: a prospective Chinese cohort study. *Nutrients* **15**, 2144 (2023).
21. Tam, C. H. T. et al. Development of genome-wide polygenic risk scores for lipid traits and clinical applications for dyslipidemia, subclinical atherosclerosis, and diabetes cardiovascular complications among East Asians. *Genome Med.* **13**, 29 (2021).
22. Aulchenko, Y. S. et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat. Genet.* **41**, 47–55 (2009).
23. Younossi, Z. M. et al. Global epidemiology of nonalcoholic fatty liver disease-meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* **64**, 73–84 (2016).
24. Gao, X. R. et al. Explainable machine learning aggregates polygenic risk scores and electronic health records for Alzheimer's disease prediction. *Sci. Rep.* **13**, 450 (2023).
25. Kathak, R. R. et al. The association between elevated lipid profile and liver enzymes: a study on Bangladeshi adults. *Sci. Rep.* **12**, 1711 (2022).
26. Bril, F. et al. Hepatic steatosis and insulin resistance, but not steatohepatitis, promote atherogenic dyslipidemia in NAFLD. *J. Clin. Endocrinol. Metab.* **101**, 644–652 (2016).
27. DiStefano, J. K. NAFLD and NASH in Postmenopausal Women: Implications for Diagnosis and Treatment. *Endocrinology* 161, (2020).
28. Cherubini, A. et al. Interaction between Estrogen receptor-α and PNPLA3 p.I148M variant drives fatty liver disease susceptibility in women. *Nat. Med.* **29**, 2643–2655 (2023).
29. Hashimoto, E. & Tokushige, K. Prevalence, gender, ethnic variations, and prognosis of NASH. *J. Gastroenterol.* **46** (Suppl 1), 63–69 (2011).

30. Balakrishnan, M. et al. Women have a lower risk of nonalcoholic fatty liver disease but a higher risk of progression vs men: a systematic review and meta-analysis. *Clin. Gastroenterol. Hepatol.* **19**, 61–71e15 (2021).
31. Burra, P. et al. Clinical impact of sexual dimorphism in non-alcoholic fatty liver disease (NAFLD) and non-alcoholic steatohepatitis (NASH). *Liver Int.* **41**, 1713–1733 (2021).
32. Fresneda, S. et al. Sex and age differences in the association of fatty liver index-defined non-alcoholic fatty liver disease with cardiometabolic risk factors: a cross-sectional study. *Biol. Sex. Differ.* **13**, 64 (2022).
33. Lee, S., Kim, K. W. & Lee, J. Sex-specific cutoff values of visceral fat area for lean vs. overweight/obese nonalcoholic fatty liver disease in Asians. *J. Clin. Transl Hepatol.* **10**, 595–599 (2022).
34. Higashiura, Y. et al. Elevated fatty liver index is independently associated with new onset of hypertension during a 10-year period in both male and female subjects. *J. Am. Heart Assoc.* **10**, e021430 (2021).
35. Jeong, S., Kim, H. J. & Han, H. W. Sex-separated NAFLD/NASH scores May improve predictive performance. *Clin. Gastroenterol. Hepatol.* **21**, 855–856 (2023).
36. Kim, Y. & Han, B. G. KoGES group. Cohort profile: the Korean genome and epidemiology study (KoGES) consortium. *Int. J. Epidemiol.* **46**, e20 (2017).
37. Rhee, E. J. et al. 2018 Guidelines for the management of dyslipidemia in Korea. *J. Lipid Atheroscler.* **8**, 78–131 (2019).
38. Alkhouri, N. et al. The prevalence of alcoholic and nonalcoholic fatty liver disease in adolescents and young adults in the united States: analysis of the NHANES database. *BMC Gastroenterol.* **22**, 366 (2022).
39. Moon, S. et al. The Korea biobank array: design and identification of coding variants associated with blood biochemical traits. *Sci. Rep.* **9**, 1382 (2019).
40. Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A. & Smoller, J. W. Polygenic prediction via bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
41. Willer, C. J. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
42. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
43. Wu, J. et al. Population-specific cut-off points of fatty liver index: a study based on the National health and nutrition examination survey data. *BMC Gastroenterol.* **22**, 265 (2022).

Acknowledgements

This study was conducted with bioresources from the National Biobank of Korea, the Korea Disease Control and Prevention Agency, Republic of Korea (KBN-2019-066).

Author contributions

SK and HYY conceptualized and designed the study. SK analyzed and interpreted the data. SK and HYY wrote the first draft of the manuscript and critically revised it for intellectual content. All the authors have read and approved the final manuscript.

Funding

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (NRF-2022R1F1A1068307).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-92766-3.

Correspondence and requests for materials should be addressed to H.Y.Y.