scientific reports

OPEN

Development and validation of a risk prediction model for abdominal aortic aneurysm: a nationwide population-based cohort study

Hyung-jin Cho[1], Mi-hyeong Kim[1], Kyung-Jai Ko[2], Kang-woong Jun[3], Kyung-do Han[4] & Jeong-Kye Hwang[1]✉

Abdominal aortic aneurysm (AAA) is characterized by irreversible localized dilatation of the abdominal aorta. It poses a significant health risk. As AAA size tends to increase over time, there is a heightened risk of rupture, resulting in a substantially high mortality rate. Although AAA screening programs targeting specific demographics are available, there is room for improvement in terms of inclusivity and cost-effectiveness. This study aimed to develop a predictive model for AAA occurrence utilizing seven years of data from the Korean National Health Insurance Service database (NHIS). This study utilized NHIS data from 2009 to 2020. A total of 4,234,415 individuals who underwent health examinations in 2009 were identified. After applying exclusion criteria, a total of 3,937,535 individuals were selected. Of them, 70% were used for model development and 30% were used for validation. The mean follow-up duration was 10.11 ± 1.29 years, during which 2,836 cases of AAA were identified among 1,181,131 (2.4%) participants in the validation cohort. The model incorporated a set of 10 variables, encompassing age, sex, obesity, smoking, drinking, diabetes (DM), hypertension (HTN), dyslipidemia, chronic kidney disease (CKD), and cardiocerebrovascular disease (CVD). Evaluation of the model's predictive performance revealed an area under the curve (AUC) of 0.807 (95% CI: 0.80–0.81) when it was applied to the development cohort. The AUC remained high at 0.803 (95% CI: 0.79–0.81) when the model was applied to the validation cohort, indicating its effectiveness in forecasting AAA occurrence. A multivariable risk model for predicting the onset of AAA was successfully developed, showcasing an excellent performance with an AUC value of 0.807, surpassing traditional screening methods. This model has the potential to selectively identify high-risk patients from a slightly broader pool than current screening approaches. Priority should be given to proactive screening efforts targeting individuals at elevated risk for AAA, with the goal of reducing AAA-related mortality.

Keywords Aortic aneurysm, Abdominal, Nomograms

An abdominal aortic aneurysm (AAA) refers to an irreversible localized dilatation of the abdominal aorta. Globally, the prevalence of AAA is reported to be around 2–8%. In Korea, its prevalence is approximately 2.8%[1,2]. In general, the size of AAA tends to gradually increase over time. As the size increases, the risk of rupture also increases[3,4]. Once a rupture occurs, the mortality rate has been reported to be as high as 81%[5]. When treating intact AAA, the 30-day mortality rate has been reported to range from 1.16 to 3.27%. For cases with ruptured AAA, among patients receiving treatment in hospitals, the 30-day mortality rate has been reported to be in the range of 30.2–39.6%[6].

Due to reasons mentioned earlier, several countries including the United States, the United Kingdom, Sweden, Denmark, and others have implemented AAA screening programs. Numerous randomized controlled trials (RCTs) and observational follow-up studies have been conducted in this context[7–10]. These screening programs typically target men aged 65 and older with a history of smoking. While such screening programs do

[1]Division of Vascular and Transplant Surgery, Department of Surgery, Eunpyeong St. Mary's Hospital, College of Medicine, The Catholic University of Korea, 1021, Tongil-ro, Eunpyeong-gu, Seoul 03312, Korea. [2]Department of Surgery, Kangdong Sacred Heart Hospital, Seoul, Korea. [3]Division of Vascular and Transplant Surgery, Department of Surgery, Bucheon St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Bucheon, Gyeonggi-do, Korea. [4]Department of Statistics and Actuarial Science, Soongsil University, Seoul, Korea. ✉email: jjungyong@catholic.ac.kr

not significantly affect the overall all-cause mortality, studies have reported that such screening programs can reduce the rate of ruptured abdominal aortic aneurysms and decrease AAA-related mortality. Additionally, these screenings have been shown to be cost-effective in terms of healthcare resource utilization.

However, considering that these approaches have focused on men aged 65 and older with a history of smoking, there is a possibility of missing out on patients. According to Summers et al., there are still significant high-risk groups that fall outside the current guidelines who could greatly benefit from AAA screening[11]. By developing a model that can predict AAA occurrence based on basic screening results, it would be possible to expand screening to a wider population. This could help reduce chances of missing out on individuals who may be at risk and minimize unnecessary screenings, ultimately leading to improved cost-effectiveness.

Therefore, this study aimed to develop a predictive model for the presence of AAA using 11 years of data from the Korean National Health Insurance Service (NHIS) database and subsequently conduct validation. To the best of our knowledge, this is a novel approach.

Methods
This study was a parallel study to "Risk of various cancers in adults with abdominal aortic aneurysm" and "The risk of dementia in adults with abdominal aortic aneurysm" by Cho et al.[12,13] It showed similarities in protocol, patient group selection method, and statistical method.

Data source
The healthcare insurance system in Korea has been introduced in the two previous parallel studies. Data utilized in this research spanned from 2009 to 2020. They were gathered from the NHIS database.

Patients
The study initially enrolled 4,234,415 individuals aged 20 and above who underwent health examinations in 2009. Patients who had previously been diagnosed with AAA at the time of the health examination were excluded ($n = 2,409$). Individuals who lacked data for one or more variables required for this study at the time of the health examination were excluded from the analysis ($n = 284,471$). The AAA patient group was defined using diagnostic codes and procedure codes, similar to previous studies. (Appendix 1) Patients who were lost to follow-up within one year after the health examination were excluded. Likewise, those who developed AAA within one year were also excluded to establish a clear cause-and-effect relationship ($n = 10,431$). 70% of these patients were randomly assigned to a development cohort for model training, while the remaining 30% were allocated to a validation cohort. The allocation was performed using simple random sampling without stratification to preserve the natural distribution of baseline characteristics. (Fig. 1)

Data collection and definition
Demographic data were gathered from the NHIS database, encompassing age, sex, smoking habits, alcohol consumption, physical activity, waist circumference, body mass index (BMI), and income level. Information regarding underlying health conditions, including hypertension, diabetes mellitus (DM), dyslipidemia, chronic kidney disease (CKD), and a history of cardiocerebrovascular disease (CVD), was also collated. Definitions of variables were similar to those described in previous papers. They are summarized in Appendix 1.

This study was conducted in accordance with the ethical standards of the Declaration of Helsinki and the regulations. The requirement for informed consent was waived because the study used de-identified data from the National Health Insurance Service (NHIS) database of Korea. This study was approved by the Institutional Review Board (IRB) of The Catholic University of Korea, Eunpyeong St. Mary's Hospital, Seoul, Korea (IRB approval number: PC23ZASI0143).

Statistical analysis
Continuous variables are presented as mean ± SD or 95% CIs, while categorical variables are expressed as numbers and percentages (%). For comparing characteristics between patient and control groups, Student's t-tests were employed for continuous variables and Chi-squared test or Fisher's exact tests were used for categorical variables. Incidence rates of AAA are presented per 1,000 person-years. To investigate hazard ratios (HR) of various variables on the occurrence of AAA, the Cox proportional hazard regression model was employed. Variables included factors associated with AAA based on the literature[14]. These variables were selected from data obtainable through health examination records.

Risk scores were allocated according to the HR for each risk factor identified in the final Cox hazard regression model. Each of the 10 variables (age, sex, obesity, smoking status, drinking, fasting glucose level, blood pressure, total cholesterol level, presence of CKD, and previous CVD) was assigned a score ranging from 0 to 100. Each variable was then mapped to a specific point by extending a line vertically along the score axis. To assess the performance of the model, calibration and discrimination were conducted. For calibration, predicted 5-years disease free survival was plotted against observed 5-years disease free survival to visually inspect the alignment[15]. (Appendix 2) For discrimination, receiver operating characteristic curves were generated and the area under curve was examined[16]. All statistical analyses were performed using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA) and the R Project for Statistical Computing version 3.3 (Vienna, Austria).

Results
Baseline characteristics according to presence of AAA
Initially, it was observed that distribution patterns of variables were not significantly different between the development cohort and the validation cohort. (Table 1) Average follow-up period was 10.11 ± 1.28 years, with
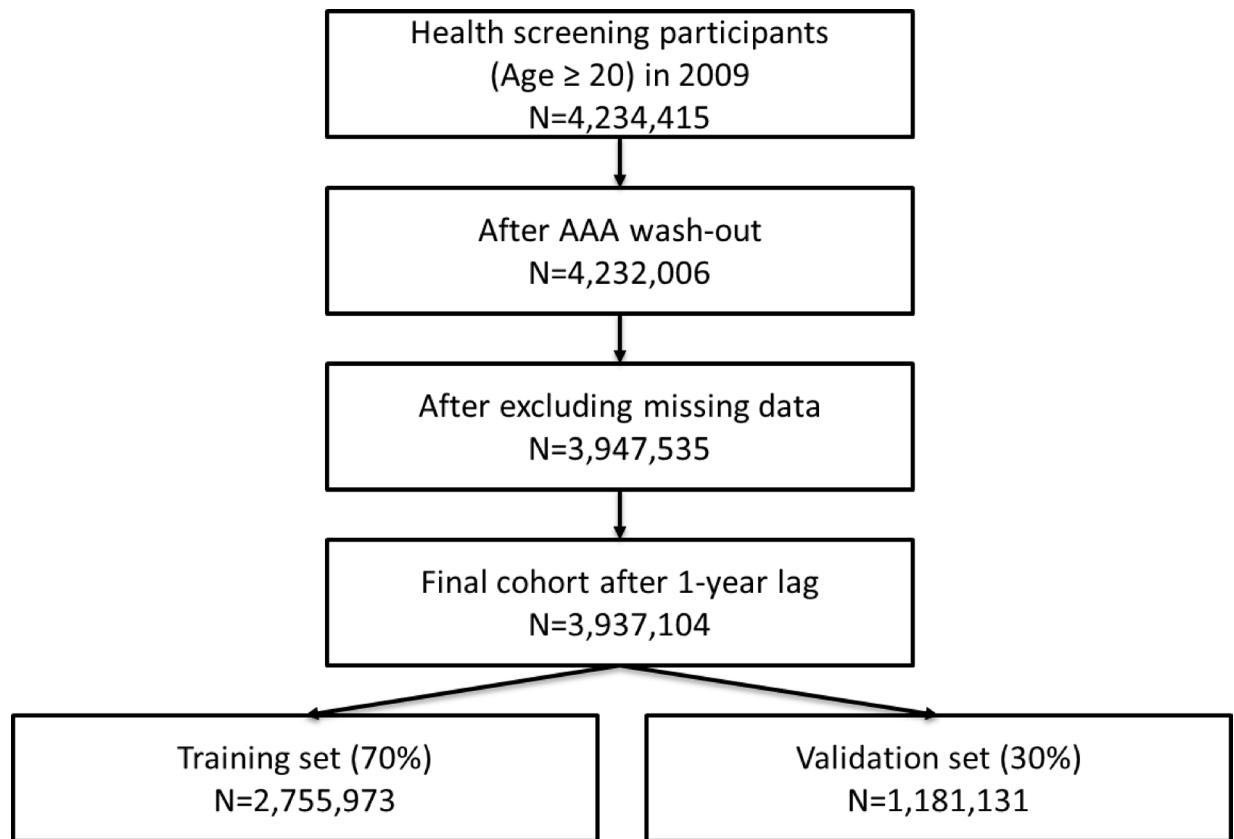
Fig. 1. Enrollment flow chart. AAA: Abdominal aortic aneurysm.

AAA occurring in 6,514 (2.36%) out of 2,755,973 participants in the development cohort. At baseline, the mean age was 47.22 ± 14.01 years. Male patients accounted for 54.56%. The AAA patient group was older (62.88 ± 11.3 years vs. 47.19 ± 14 years, p-value < 0.001), had a higher proportion of males (67.88% vs. 54.53%, p-value < 0.001), and higher BMI (24.25 ± 3.1 vs. 23.7 ± 3.22, p-value < 0.001) than the control group. Additionally, the smoking rate was higher (31.98% vs. 26.01%, p-value < 0.001), while the proportion of patients who consumed alcohol was comparatively lower (38.52% vs. 48.31%, p-value < 0.001) in the AAA patient group. Interestingly, the AAA group exhibited higher levels of physical activity (21.08% vs. 17.96%, p-value < 0.001). In terms of comorbidities, the AAA group had higher prevalances of hypertension (55.63% vs. 25.37%, p-value < 0.001), hyperlipidemia (32.7% vs. 17.34%, p-value < 0.001), CKD (15.32% vs. 6.91%, p-value < 0.001), CVD (7.52% vs. 1.87%, p-value < 0.001), and DM (12.48% vs. 8.66%, p-value < 0.001).

Selection of variables

Among variables listed in Table 1 that showed significant distribution differences between the AAA group and the control group, a total of 12 variables were selected. When similar or overlapping variables were present (e.g., BMI vs. obesity, or diabetes vs. diabetes severity), the variable with greater clinical interpretability or statistical contribution was retained to avoid redundancy and overfitting in the model. (Tables 2 and 3) The Cox proportional hazard regression model was then utilized to examine the HR for the occurrence of AAA. In multivariate analysis, we ultimately selected 10 variables that were statistically significant, including age and sex, obesity, smoking, drinking, DM, HTN, dyslipidemia, CKD, and CVD. Old age [HR: 30.43 (95% CI: 26.48–34.97)], male sex [HR: 2.01 (95% CI: 1.88–2.16)], obesity [HR: 1.06 (95% CI: 1.01–1.11)], smoking [HR: 2.20 (95% CI: 2.05–2.36)], DM [HR: 0.64 (95% CI: 0.59–0.69)], HTN [HR: 2.04 (95% CI: 1.89–2.20)], dyslipidemia [HR: 1.56 (95% CI: 1.47–1.66)], CKD [HR: 1.41 (95% CI: 1.31–1.51)], CVD [HR: 1.50 (95% CI: 1.67–1.65)] were significant predictive factors for occurrence of AAA after adjusting for all 10 variables. A nomogram for risk scoring developed from the risk prediction model was constructed to estimate the five-year risk of AAA. (Fig. 2)

Validation of the prediction model

Average follow-up period was 10.11 ± 1.29 years, with AAA occurring in 2,836 (2.4%) out of 1,181,131 participants in the validation cohort. At baseline, the mean age was 47.24 ± 14.02 years. Male patients accounted for 54.58%. Additionally, after examining the area under curve (AUC) value for AAA occurrence prediction in the prediction model, when applied to the development cohort data, the AUC was found to be 0.807 (95% CI: 0.80–0.81). When applied to the validation cohort data, the AUC value was 0.803 (95% CI: 0.79–0.81). (Fig. 3) This suggests that the model is effective in predicting the occurrence of AAA.

	Development cohort ($n=2,755,973$)			Validation cohort ($n=1,181,131$)			
	Control ($n=2,749,459$)	AAA ($n=6,514$)	*P*-value	Control ($n=1,178,295$)	AAA ($n=2,836$)	*P*-value	*p*-value
Age (years)			<0.001			<0.001	>0.99
20–39	852,451 (31)	233 (3.58)		365,202 (30.99)	95 (3.35)		
40–64	1,540,291 (56.02)	3,033 (46.56)		659,514 (55.97)	1,363 (48.06)		
65≤	356,717 (12.97)	3,248 (49.86)		154,957 (13.12)	1,378 (48.59)		
Sex (male)	1,499,346 (54.53)	4,422 (67.88)	<0.001	642,713 (54.55)	1,899 (66.96)	<0.001	>0.99
Income level (1st quartile)	535,877 (19.49)	1,306 (20.05)	0.25	229,790 (19.5)	553 (19.5)	0.99	>0.99
Obesity (BMI>25)	896,723 (32.61)	2,582 (39.64)	<0.001	384,172 (32.6)	1,090 (38.43)	<0.001	>0.99
Smoking			<0.001			<0.001	>0.99
Never	1,640,878 (69.68)	3,105 (47.67)		702,867 (59.65)	1,367 (48.2)		
Former	393,372 (14.31)	1,326 (20.36)		168,886 (14.33)	558 (19.68)		
Current	715,209(26.01)	2,083 (31.98)		106,542 (26.02)	911 (32.12)		
Drinking (yes)	1,328,379 (48.31)	2,509 (38.52)	<0.001	569,823 (48.36)	1,139 (40.16)	<0.001	>0.99
Exercise (yes)	493,732 (17.96)	1,373 (21.08)	<0.001	211,776 (17.97)	574 (20.24)	0.001	>0.99
Diabetes mellitus (yes)	238,240 (8.66)	813 (12.48)	<0.001	102,195 (8.67)	368 (12.98)	<0.001	>0.99
Diabetes mellitus (3 levels)			<0.001			<0.001	>0.99
Normal	1,883,391 (68.5)	3,898 (59.84)		806,380 (68.44)	1,699 (59.91)		
Impaired fasting glucose	627,828 (22.83)	1,803 (27.68)		269,720 (22.89)	769 (27.12)		
Diabetes mellitus	238,240 (8.66)	813 (12.48)		102,195 (8.67)	368 (12.98)		
Hypertension (yes)	697,404 (25.37)	3,624 (55.63)	<0.001	298,660 (25.35)	1,572 (55.43)	<0.001	>0.99
Hypertension (3 levels)			<0.001			<0.001	>0.99
Normal	951,112 (34.59)	1,015 (15.58)		408,897 (34.7)	435 (15.34)		
Pre-Hypertension	1,100,943 (40.04)	1,875 (28.78)		470,738 (39.95)	829 (29.23)		
Hypertension	697,404 (25.37)	3,624 (55.63)		298,660 (25.35)	1,572 (55.43)		
Dyslipidemia (yes)	476,635 (17.34)	2,130 (32.7)	<0.001	203,517 (17.27)	925 (32.62)	<0.001	>0.99
Total cholesterol (3 levels)			<0.001			<0.001	>0.99
<200	1,480,637 (53.85)	2,556 (39.24)		634,808 (53.88)	1,094 (38.58)		
200–239	792,187 (28.81)	1,828 (28.06)		339,970 (28.85)	817 (28.81)		
240≤	476,635 (17.34)	2,130 (32.7)		203,517 (17.27)	925 (32.62)		
Chronic kidney disease (yes)	190,024 (6.91)	998 (15.32)	<0.001	81,389 (6.91)	488 (17.21)	<0.001	>0.99
Cardiocerebrovascular disease (yes)	51,358 (1.87)	490 (7.52)	<0.001	21,800 (1.85)	230 (8.11)	<0.001	>0.99
Age (years)	47.19±14	62.88±11.3	<0.001	47.2±14.01	62.75±11.27	<0.001	>0.99
BMI (kg/m²)	23.7±3.22	24.25±3.1	<0.001	23.7±3.23	24.2±3.1	<0.001	>0.99
Waist circumference (cm)	80.22±9.1	84.51±8.57	<0.001	80.24±9.1	84.28±8.56	<0.001	>0.99
Glucose (mmol/L)	97.29±23.81	99.99±23.64	<0.001	97.31±23.82	100.46±29.33	<0.001	>0.99
Total cholesterol (mmol/L)	195.05±36.8	200.38±39.56	<0.001	195±36.79	200.14±40.13	<0.001	>0.99
HDL-cholesterol (mmol/L)	56.09±27.87	52.31±31.15	<0.001	56.1±27.76	53.23±33.29	<0.001	>0.99
LDL-cholesterol (mmol/L)	113.59±38.63	120.3±42.97	<0.001	113.53±38.59	118.56±39.11	<0.001	>0.99

Table 1. Clinical characteristics of study participants based on the occurrence of abdominal aortic aneurysm (AAA) in development and validation cohorts used for the predictive model.

Prediction model

The sum of total scores was obtained by combining scores of the 10 variables. It ranged from 0 to a maximum of 226 points. For example, a male (20 points) aged 65 or older (100 points) who smoked (23 points), had a normal weight (0 points), and had diabetes (0 points) with no other underlying conditions (0 points) would have a total score of 143 points. On the other hand, a female (0 points) aged 40 (57 points) who smoked (23 points), had a normal weight (0 points) who did not drink alcohol (12 points), had no DM (13 points) but had HTN (21 points), hyperlipidemia (13 points), and CKD (10 points) would have a total score of 149 points.(Table S1) In this case, it can be inferred that the probability of developing abdominal aortic aneurysm within 5 years is less than 0.5%.

In the patient group with the highest score range of 215 points or above in the prediction model, the incidence rate of AAA was confirmed to be 1.194 per 1000 person-years.(Fig. 4) Additionally, to assess the potential overfitting of the model, we compared the incidence in the validation cohort. It was observed that similar patterns were present in each interval, confirming the excellence of this predictive model.

		N	Event	Duration	Rate	Univariate model	P-value
Age	20–39	852,684	233	8,772,934.9	0.03	1 (Ref.)	<0.001
	40–64	1,543,324	3,033	15,764,115.73	0.19	7.25 (6.34–8.28)	
	65≤	359,965	3,248	3,337,159.96	0.97	37.285(32.64–42.58)	
Sex	Male	1,503,768	4,422	15,111,342.75	0.29	1.791(1.7–1.88)	<0.001
	Female	1,252,205	2,092	12,762,867.84	0.16	1 (Ref.)	
Income level (1st quartile)	No	2,218,790	5,208	22,459,933.36	0.23	1 (Ref.)	<0.001
	Yes	537,183	1,306	5,414,277.23	0.24	1.042 (0.98–1.10)	
Obesity (BMI > 25)	Yes	1,856,668	3,932	18,758,964.92	0.21	1 (Ref.)	<0.001
	No	899,305	2,582	9,115,245.68	0.28	1.351 (1.28–1.41)	
Smoking	Never	1,643,983	3,105	16,686,049.9	0.19	1 (Ref.)	<0.001
	Former	394,698	1,326	3,967,496.03	0.33	1.801 (1.68–1.92)	
	Current	717,292	2,083	7,220,664.67	0.29	1.555 (1.47–1.64)	
Drinking	No	1,425,085	4,005	14,356,278.92	0.28	1 (Ref.)	<0.001
	Yes	1,330,888	2,509	13,517,931.67	0.19	0.665 (0.63–0.69)	
Exercise	No	2,260,868	5,141	22,855,563.58	0.22	1 (Ref.)	<0.001
	Yes	495,105	1,373	5,018,647.02	0.27	1.216 (1.14–1.29)	
Diabetes mellitus (3 levels)	Normal	1,887,289	3,898	19,209,840.47	0.20	1 (Ref.)	<0.001
	Impaired fasting glucose	629,631	1,803	6,347,986.4	0.28	1.402 (1.32–1.48)	
	Diabetes mellitus	239,053	813	2,316,383.73	0.35	1.741 (1.61–1.87)	
Hypertension (3 levels)	Normal	952,127	1,015	9,744,813.14	0.10	1 (Ref.)	<0.001
	Pre-hypertension	1,102,818	1,875	11,215,353.96	0.17	1.607 (1.48–1.73)	
	Hypertension	701,028	3,624	6,914,043.5	0.52	5.056 (4.71–5.42)	
Total cholesterol	< 200	1,483,193	2,556	15,011,409.06	0.17	1 (Ref.)	<0.001
	200–239	794,015	1,828	8,056,784.65	0.23	1.332 (1.25–1.41)	
	240≤	478,765	2,130	4,806,016.88	0.44	2.604 (2.45–2.75)	
Chronic kidney disease	No	2,564,951	5,516	26,006,659.17	0.21	1 (Ref.)	<0.001
	Yes	191,022	998	1,867,551.43	0.53	2.529 (2.36–2.70)	
Cardiocerebrovascular disease	No	2,704,125	6,024	27,392,872.21	0.22	1 (Ref.)	<0.001
	Yes	51,848	490	481,338.38	1.02	4.681 (4.26–5.13)	

Table 2. Hazards ratios (95% CIs) for the occurrence of abdominal aortic aneurysm (Univariate model). Note: Rate: incidence rate per 1 000 person-years.

Discussion

Utilizing the NHIS database in Korea, we developed and validated a straightforward yet effective risk prediction model for the occurrence of AAA. To the best of our knowledge, this study developed the first model capable of predicting the likelihood of AAA occurrence after a prolonged period of 5 years using data obtained through long-term follow-up observations. Our model exhibited a strong performance with an AUC of 0.807 (95% CI: 0.80–0.81). Notably, older age, male sex, obesity, current smoking, non-drinking, absence of DM, presence of hypertension, hyperlipidemia, CKD, and CVD were identified as independent predictors of an increased risk of AAA.

In opinions of vascular and endovascular surgeons, prioritizing research on methods for predicting AAA growth is considered essential[17]. Therefore, various studies are being conducted to predict the growth of AAA. For example, approaches that include geometric perspectives from CT images, attempts to predict AAA growth using imaging modalities such as MRI to examine metabolism, and exploring associations with biomarkers have been reported[18–20]. Of course, while predicting the growth of AAA is important, it is even more crucial to first predict the presence of AAA.

AAA screening is most widely conducted based on the U.S. Preventive Services Task Force guidelines[21]. Additionally, there are other guidelines such as the slightly more sensitive Society for Vascular Surgery (SVS) guideline and the SVS expanded criteria[22]. The SVS guideline significantly broadens the scope of patients who could potentially be included in screening programs. Nevertheless, about one-quarter to one-third of patients undergoing treatment for AAA could still potentially go unnoticed by any existing screening guidelines[23]. This aspect can also be verified through the model developed in this study. For instance, scores corresponding to the criteria of traditional guidelines targeting males over 60 who smoke can also be sufficiently obtained from individuals under 60, females, or non-smokers. Through this, one aspect of the model's excellence can be confirmed. Specifically, its ability to include a slightly more comprehensive range of patients while effectively identifying specific individuals within that group is confirmed.

The second excellence lies in the fact that variables included in this model are easily recognizable factors such as medical history, smoking history, age, and gender. Due to this simplicity, the general public can easily check their risk of developing AAA at any time. If a cutoff value is determined through subsequent research, strongly

		N	Event	Duration	Rate	Multivariate model	P-value	Final model	P-value
Age	20–39	852,684	233	8,772,934.9	0.03	1 (Ref.)	**< 0.001**	1 (Ref.)	**< 0.001**
	40–64	1,543,324	3,033	15,764,115.73	0.19	6.919 (6.04–7.91)		6.937 (6.06–7.93)	
	65≤	359,965	3,248	3,337,159.96	0.97	30.328 (26.38–34.85)		30.428 (26.47–34.96)	
Sex	Male	1,503,768	4,422	15,111,342.75	0.29	2.014 (1.87–2.15)	**< 0.001**	2.013 (1.87–2.15)	**< 0.001**
	Female	1,252,205	2,092	12,762,867.84	0.16	1 (Ref.)		1 (Ref.)	
Income level (1st quartile)	No	2,218,790	5,208	22,459,933.36	0.23	1 (Ref.)	**0.12**		
	Yes	537,183	1,306	5,414,277.23	0.24	1.048 (0.98–1.11)			
Obesity (BMI > 25)	Yes	1,856,668	3,932	18,758,964.92	0.21	1 (Ref.)	**0.029**	1 (Ref.)	**0.03**
	No	899,305	2,582	9,115,245.68	0.28	1.058 (1.00–1.11)		1.058 (1.00–1.11)	
Smoking	Never	1,643,983	3,105	16,686,049.9	0.19	1 (Ref.)	**< 0.001**	1 (Ref.)	**< 0.001**
	Former	394,698	1,326	3,967,496.03	0.33	1.442 (1.33–1.55)		1.442 (1.33–1.55)	
	Current	717,292	2,083	7,220,664.67	0.29	2.193 (2.04–2.35)		2.195 (2.04–2.35)	
Drinking	No	1,425,085	4,005	14,356,278.92	0.28	1 (Ref.)	**< 0.001**	1 (Ref.)	**< 0.001**
	Yes	1,330,888	2,509	13,517,931.67	0.19	0.662 (0.62–0.70)		0.662 (0.62–0.70)	
Exercise	No	2,260,868	5,141	22,855,563.58	0.22	1 (Ref.)	0.74		
	Yes	495,105	1,373	5,018,647.02	0.27	1.01 (0.95–1.07)			
Diabetes mellitus (3 levels)	Normal	1,887,289	3,898	19,209,840.47	0.20	1 (Ref.)	**< 0.001**	1 (Ref.)	**< 0.001**
	Impaired fasting glucose	629,631	1,803	6,347,986.4	0.28	0.919 (0.86–0.97)		0.919 (0.86–0.97)	
	Diabetes mellitus	239,053	813	2,316,383.73	0.35	0.637 (0.58–0.68)		0.637 (0.58–0.68)	
Hypertension (3 levels)	Normal	952,127	1,015	9,744,813.14	0.10	1 (Ref.)	**< 0.001**	1 (Ref.)	**< 0.001**
	Pre-hypertension	1,102,818	1,875	11,215,353.96	0.17	1.204 (1.11–1.3)		1.204 (1.11–1.30)	
	Hypertension	701,028	3,624	6,914,043.5	0.52	2.036 (1.88–2.19)		2.038 (1.89–2.19)	
Total cholesterol	< 200	1,483,193	2,556	15,011,409.06	0.17	1 (Ref.)	**< 0.001**	1 (Ref.)	**< 0.001**
	200–239	794,015	1,828	8,056,784.65	0.23	1.136 (1.06–1.20)		1.136 (1.06–1.20)	
	240≤	478,765	2,130	4,806,016.88	0.44	1.561 (1.47–1.65)		1.56 (1.46–1.65)	
Chronic kidney disease	No	2,564,951	5,516	26,006,659.17	0.21	1 (Ref.)	**< 0.001**	1 (Ref.)	**< 0.001**
	Yes	191,022	998	1,867,551.43	0.53	1.408 (1.31–1.50)		1.406 (1.31–1.50)	
Cardiocerebrovascular disease	No	2,704,125	6,024	27,392,872.21	0.22	1 (Ref.)	**< 0.001**	1 (Ref.)	**< 0.001**
	Yes	51,848	490	481,338.38	1.02	1.505 (1.36–1.65)		1.504 (1.36–1.65)	

Table 3. Hazards ratios (95% CIs) for the occurrence of abdominal aortic aneurysm (Multivariate model and final model). Note: Rate: incidence rate per 1 000 person-years.

recommending screening to individuals with risk levels above this value could help improve the screening rate, addressing one of the issues with the current screening method. Lastly, due to being developed with long-term data from a substantial number of patients, it is likely to have strength in terms of reliability. As the model developed in this study is an initial iteration, further research involving additional easily obtainable variables through subsequent validation could enhance the model's accuracy. This aligns with the initial goal of the model development, which is to expedite identification of AAA patients, enabling timely intervention before complications such as rupture, ultimately improving the survival rate.

The current study is subject to several limitation. Firstly, due to the utilization of NHIS data, defining the presence or absence of a patient's diseases based on diagnosis codes and medication intake might have led to various errors and inconsistencies. This is because there may be undiagnosed asymptomatic AAA cases or patients who died from ruptured AAA without diagnosis compared to the group identified through screening. Secondly, in the selection of variables included in this model, efforts were made to incorporate well-known risk factors based on available information. However, due to the nature of the data, only variables that could be confirmed through health check-ups were included, while factors such as family history, transplantation status, or the presence of aneurysms in other peripheral arteries could not be incorporated. Thirdly, since the model was developed for the Korean population, it is necessary to assess its applicability and effectiveness when applied to different racial groups. Fourthly, although the cohort has a follow-up period exceeding 10 years, the model was not separately validated for short-term and long-term outcomes. Therefore, we could not assess potential differences in predictive performance or variable importance across different time horizons. Future studies with time-specific modeling may help address this limitation.

Further research is needed to establish an appropriate cut-off level for recommending screening. Subsequently, when conducting screening on the identified patient group, assessing improvements in aspects such as all-cause mortality, AAA-related mortality, and cost-effectiveness will be essential.

Conclusions

We developed a multivariable risk model capable of predicting the onset of AAA. In this study, the model demonstrated excellent performance with an AUC value of 0.807, surpassing traditional screening methods. It is
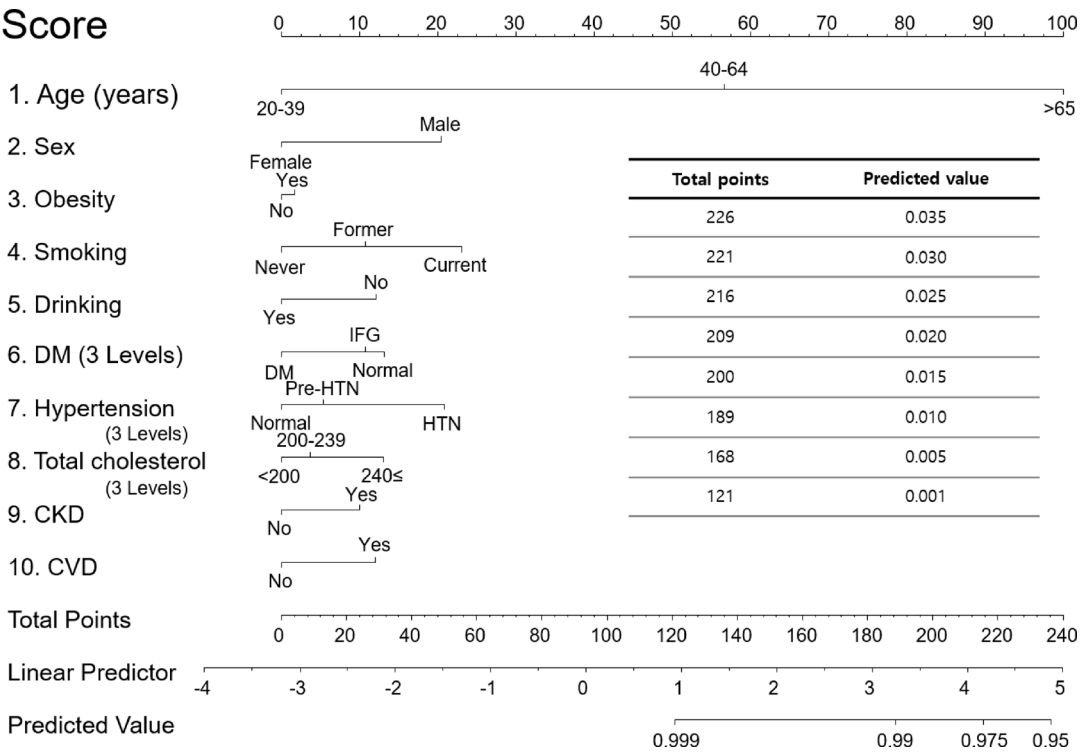
Fig. 2. A nomogram for predicting the 5-year probability of abdominal aortic aneurysm occurrence. Note: The 10 variables—age, sex, obesity, smoking status, drinking, presence of diabetes mellitus or hypertension, chronic kidney disease, cardiocerebrovascular disease, and total cholesterol level—were each assigned scores ranging from 0 to 100. The corresponding score for each variable can be determined by drawing a straight line to the scoring axis. The total score, calculated as the sum of the scores for all variables, ranges from 0 to 226 and is displayed at the bottom of the nomogram. DM: Diabetes mellitus, IFG: Impaired fasting glucose, HTN: Hypertension, CKD: Chronic kidney disease, CVD: Cardiocerebrovascular disease.

anticipated that this model can selectively identify patients from a slightly more comprehensive pool compared to existing screening approaches. Moreover, efforts should be directed towards proactive screening of high-risk individuals for AAA, aiming to reduce AAA-related mortality. For further optimization of the model's performance, external validation is necessary, along with additional analysis and refinement using larger cohort data. The development of an appropriate surveillance program in response to these findings is also crucial.
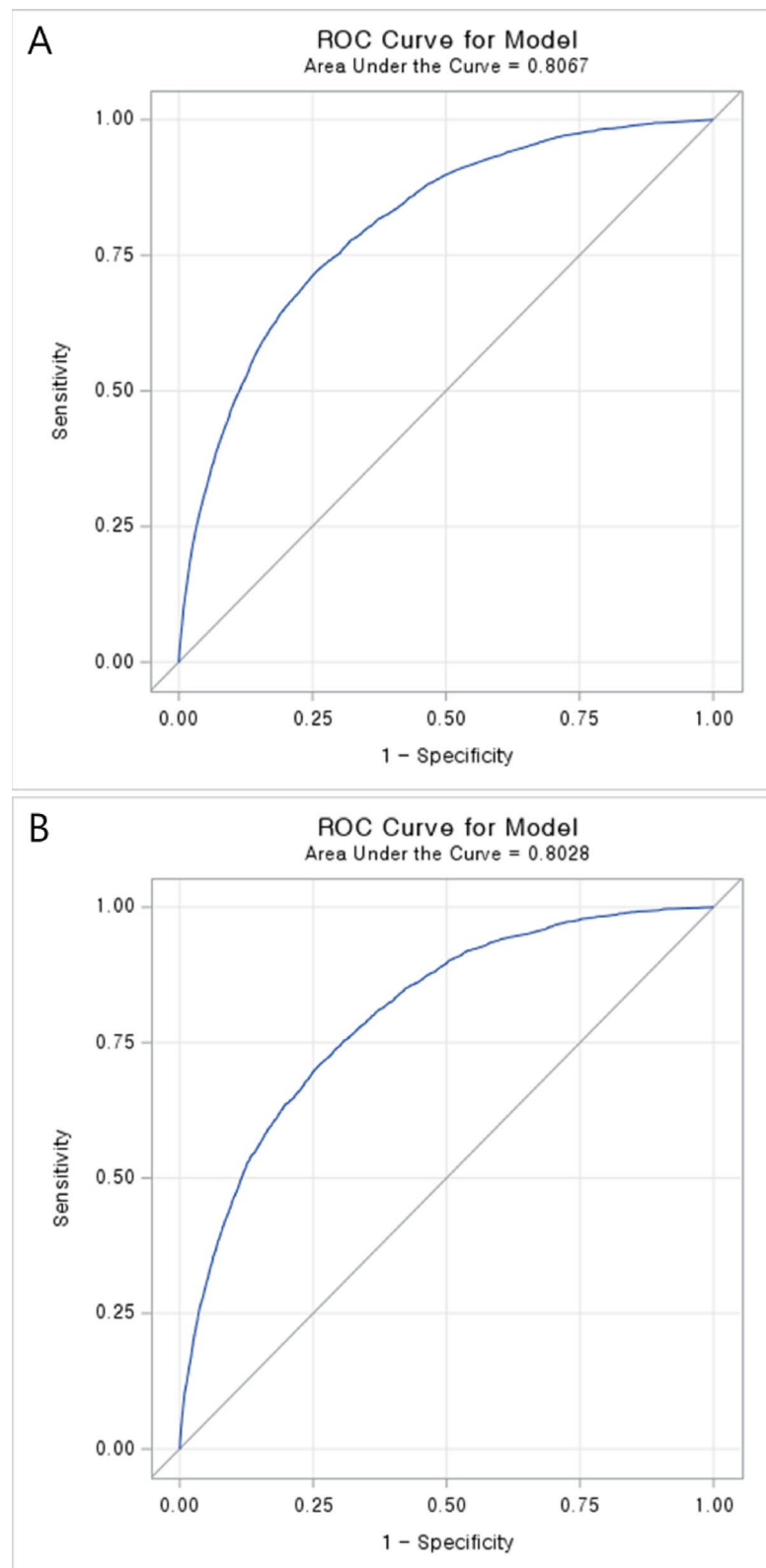
Fig. 3. The model's receiver operating characteristic curve (ROC) when using development and validation cohorts. Note: A ROC curve is using development cohort, and B ROC curve is using validation cohort.
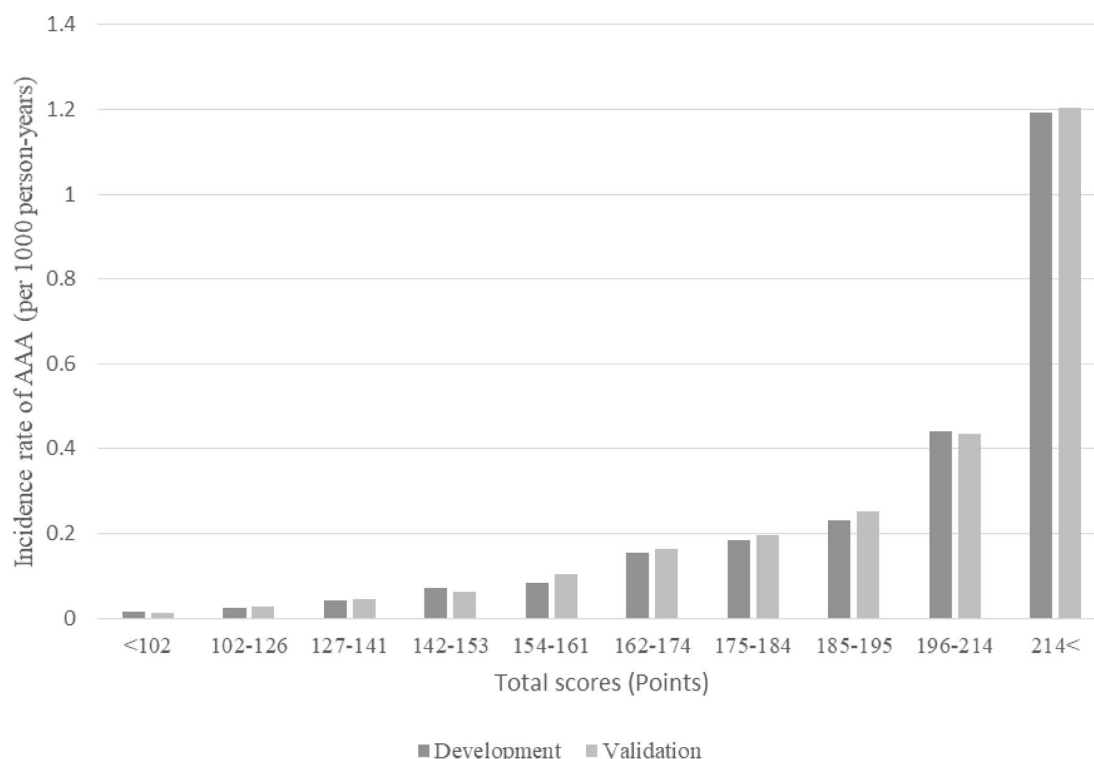
Fig. 4. Predicted incidence rate (per 1 000 person-years) using development and validation cohorts.

Data availability

The data that support the findings of this study are available from the National Health Insurance Service (NHIS) database of Korea. Restrictions apply to the availability of these data, which were used under license for the current study and are not publicly available. However, the data can be accessed through the NHIS upon reasonable request and institutional approval. Researchers interested in accessing the data may contact the corresponding author for guidance on how to obtain it from the NHIS.

References

1. Marcaccio, C. L. & Schermerhorn, M. L. Epidemiology of abdominal aortic aneurysms. *Semin. Vasc. Surg.* **34** (1), 29–37. https://doi.org/10.1053/j.semvascsurg.2021.02.004 (2021).
2. An, S. Y., Hwang, W., Sun, B. J. & Park, J-H. Prevalence and prediction of aneurysmal dilatation of the abdominal aorta in koreans: results of screening during transthoracic echocardiographic examination. *J. Cardiovasc. Imaging.* **28** (4), 257–264 (2020).
3. Sakalihasan, N., Limet, R. & Defawe, O. D. Abdominal aortic aneurysm 2005;365.
4. Oliver-Williams, C. et al. Lessons learned about prevalence and growth rates of abdominal aortic aneurysms from a 25-year ultrasound population screening programme. *Br. J. Surg.* **105** (1), 68–74. https://doi.org/10.1002/bjs.10715 (2017).
5. Reimerink, J. J., Van Der Laan, M. J., Koelemay, M. J., Balm, R. & Legemate, D. A. Systematic review and meta-analysis of population-based mortality from ruptured abdominal aortic aneurysm. *Br. J. Surg.* **100** (11), 1405–1413. https://doi.org/10.1002/bjs.9235 (2013).
6. Schmitz-Rixen, T., Böckler, D., Vogl, T. J. & Grundmann, R. T. Endovascular and open repair of abdominal aortic aneurysm. *Deutsches Ärzteblatt Int.* https://doi.org/10.3238/arztebl.2020.0813 (2020).
7. Søgaard, R. & Lindholt, J. S. Cost-effectiveness of population-based vascular disease screening and intervention in men from the Viborg vascular (VIVA) trial. *Br. J. Surg.* **105** (10), 1283–1293. https://doi.org/10.1002/bjs.10872 (2018).
8. Glover, M. J., Kim, L. G., Sweeting, M. J., Thompson, S. G. & Buxton, M. J. Cost-effectiveness of the National health service abdominal aortic aneurysm screening programme in England. *Br. J. Surg.* **101** (8), 976–982. https://doi.org/10.1002/bjs.9528 (2014).
9. US Preventive Services Task Force et al. Screening for abdominal aortic aneurysm: US preventive services task force recommendation statement. *JAMA* **322** (22), 2211. https://doi.org/10.1001/jama.2019.18928 (2019).
10. Wanhainen, A. et al. Outcome of the Swedish nationwide abdominal aortic aneurysm screening program. *Circulation* **134** (16), 1141–1148. https://doi.org/10.1161/CIRCULATIONAHA.116.022305 (2016).
11. Summers, K. L., Kerut, E. K., Sheahan, C. M. & Sheahan, M. G. Evaluating the prevalence of abdominal aortic aneurysms in the united States through a National screening database. *J. Vasc. Surg.* **73** (1), 61–68. https://doi.org/10.1016/j.jvs.2020.03.046 (2021).
12. Cho, H. et al. The risk of dementia in adults with abdominal aortic aneurysm. *Sci. Rep.* **12** (1), 1228. https://doi.org/10.1038/s41598-022-05191-1 (2022).
13. Cho, H. et al. Risk of various cancers in adults with abdominal aortic aneurysms. *J. Vasc. Surg.* **77** (1), 80–88e2. https://doi.org/10.1016/j.jvs.2022.03.896 (2023).
14. Musto, L., Smith, A., Pepper, C., Bujkiewicz, S. & Bown, M. Risk factor-targeted abdominal aortic aneurysm screening: systematic review of risk prediction for abdominal aortic aneurysm. *Br. J. Surg.* **111** (9), znae239. https://doi.org/10.1093/bjs/znae239 (2024).

15. Bonnett, L. J., Snell, K. I. E., Collins, G. S. & Riley, R. D. Guide to presenting clinical prediction models for use in clinical settings. *BMJ* l737. https://doi.org/10.1136/bmj.l737 (2019).
16. Lee, Y., Bang, H. & Kim, D. J. How to Establish clinical prediction models. *Endocrinol. Metab.* **31** (1), 38. https://doi.org/10.3803/EnM.2016.31.1.38 (2016).
17. Lee, R., Jones, A., Cassimjee, I. & Handa, A. International opinion on priorities in research for small abdominal aortic aneurysms and the potential path for research to impact clinical management. *Int. J. Cardiol.* **245**, 253–255. https://doi.org/10.1016/j.ijcard.2017.06.058 (2017).
18. Vermeulen, J. J. M. et al. A systematic review summarizing local vascular characteristics of aneurysm wall to predict for progression and rupture risk of abdominal aortic aneurysms. *J. Vasc. Surg.* **77** (1), 288–298e2. https://doi.org/10.1016/j.jvs.2022.07.008 (2023).
19. Forneris, A. et al. AI-powered assessment of biomarkers for growth prediction of abdominal aortic aneurysms. *JVS-Vascular Sci.* **4**, 100119. https://doi.org/10.1016/j.jvssci.2023.100119 (2023).
20. Chandrashekar, A. et al. Prediction of abdominal aortic aneurysm growth using geometric assessment of computerized tomography images acquired during the aneurysm surveillance period. *Ann. Surg.* **277** (1), e175–e183. https://doi.org/10.1097/SLA.0000000000004711 (2023).
21. U.S. Preventive Services Task Force. Screening for abdominal aortic aneurysm: recommendation statement. *Ann. Intern. Med.* **142** (3), 198–202. https://doi.org/10.7326/0003-4819-142-3-200502010-00011 (2005).
22. Chaikof, E. L. et al. The society for vascular surgery practice guidelines on the care of patients with an abdominal aortic aneurysm. *J. Vasc. Surg.* **67** (1), 2–77e2. https://doi.org/10.1016/j.jvs.2017.10.044 (2018).
23. Carnevale, M. L., Koleilat, I., Lipsitz, E. C., Friedmann, P. & Indes, J. E. Extended screening guidelines for the diagnosis of abdominal aortic aneurysm. *J. Vasc. Surg.* **72** (6), 1917–1926. https://doi.org/10.1016/j.jvs.2020.03.047 (2020).

Acknowledgements

Author contributions

Hyungjin Cho: Conception and design, analysis and interpretation, Writing the manuscript, Critical revision, Approval of the manuscript, Agreement to be accountable Mi-hyeong Kim: Conception and design, Critical revision, Approval of the manuscript, Agreement to be accountable Kyung-Jai Ko: Conception and design, Critical revision, Approval of the manuscript, Agreement to be accountable Kang-woong Jun: Conception and design, Critical revision, Approval of the manuscript, Agreement to be accountable Kyung-do Han: Conception and design, analysis and interpretation, Data collection, Critical revision, Approval of the manuscript, Agreement to be accountable, Statistical Analysis Jeong-Kye Hwang: Conception and design, Critical revision, Approval of the manuscript, Agreement to be accountable.

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/s41598-025-11956-1.

Correspondence and requests for materials should be addressed to J.-K.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.