KJP

Original Article

A prediction model of low back pain risk: a population based cohort study in Korea

David Mukasa[1] and Joohon Sung[1,2]

[1]Complex Diseases & Genome Epidemiology Branch, Division of Epidemiology, School of Public Health, Seoul National University, Seoul, Korea
[2]Department of Epidemiology, School of Public Health and Institute of Health and Environment, Seoul National University, Seoul, Korea

Background: Well-validated risk prediction models help to identify individuals at high risk of diseases and suggest preventive measures. A recent systematic review reported lack of validated prediction models for low back pain (LBP). We aimed to develop prediction models to estimate the 8-year risk of developing LBP and its recurrence.
Methods: A population based prospective cohort study using data from 435,968 participants in the National Health Insurance Service–National Sample Cohort enrolled from 2002 to 2010. We used Cox proportional hazards models.
Results: During median follow-up period of 8.4 years, there were 143,396 (32.9%) first onset LBP cases. The prediction model of first onset consisted of age, sex, income grade, alcohol consumption, physical exercise, body mass index (BMI), total cholesterol, blood pressure, and medical history of diseases. The model of 5-year recurrence risk was comprised of age, sex, income grade, BMI, length of prescription, and medical history of diseases. The Harrell's C-statistic was 0.812 (95% confidence interval [CI], 0.804-0.820) and 0.916 (95% CI, 0.907-0.924) in validation cohorts of LBP onset and recurrence models, respectively. Age, disc degeneration, and sex conferred the highest risk points for onset, whereas age, spondylolisthesis, and disc degeneration conferred the highest risk for recurrence.
Conclusions: LBP risk prediction models and simplified risk scores have been developed and validated using data from general medical practice. This study also offers an opportunity for external validation and updating of the models by incorporating other risk predictors in other settings, especially in this era of precision medicine.

Key Words: Big Data; Chronic Pain; Intervertebral Disc Degeneration; Low Back Pain; National Health Programs; Proportional Hazards Models; Recurrence; Risk Assessment; Risk Factors; Spondylolisthesis.

INTRODUCTION

Low back pain (LBP) is a condition characterized by pain, muscle tension, discomfort, or stiffness below the costal margin and above the inferior gluteal folds, with or with-out sciatica [1]. LBP is a common disorder causing disability, severe pain, and prolonged sick leave at personal and social expense [2]. This condition occurs in approximately 60%-80% of people at some points in their lives [3], with a potential childhood onset [4], and an estimated 6%-10% of

acute LBP patients experience repeated episodes [5]. The annual and point prevalence of LBP approximates 45% and 30%, respectively [6]. Lee et al. [7] reported a 17.1% total prevalence of LBP in Korea, whereas among hypertensive individuals, lifetime prevalence was 34.4% [8]. The reported prevalence of LBP varies substantially depending on the case definition used [9].

The course of LBP is highly variable [10], and the most appropriate description of LBP is based on its duration and the quality of symptoms that accompany the pain [11]. LBP is episodic or recurrent [12], with approximately 36% of individuals who experience an episode of LBP presenting with recurrence within one year [13]. Recurrence is associated with multiple treatments and work-related time loss, which are costly to the individuals and to the society [14]. LBP is the most important contributor to the Korean disability-adjusted life years and the disease burden is higher in females [15,16]. Well-validated risk prediction models help to identify individuals at high risk of diseases and suggestion of preventive measures.

1. Risk factors of LBP

Despite a high disease burden [17], the structural origin of most back pain episodes is unknown, with a poor correlation between symptoms and structural abnormalities [18,19], and often considered non-specific [20]. Risk factors associated with LBP include; female sex [21], older age [17], smoking [22], psychological stress [23], depression [24], education, occupation, income [25], high body weight [22], physical inactivity [26], coronary artery disease (CAD) [27], dyslipidemia [28], diabetes mellitus [29], disc degeneration (DD) [30], history of back injury [31], previous episodes [32], bone mineral density (BMD) disorders [33], spinal stenosis [34], and spondylolisthesis [30].

2. Aims of the study

A 2018 systematic review by McIntosh et al. [35] reported an absence of validated prediction models for LBP. Therefore, we undertook this study to develop and validate prediction models to estimate the 8-year risk of developing LBP as well as the 5-year recurrence risk.

MATERIALS AND METHODS

We developed and validated risk prediction equations based on guidelines stipulated by TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) protocols [36].

1. Study design and cohort description

A population based prospective study using data from National Health Insurance Service–National Sample Cohort was enrolled from 2002 to 2010. The data was comprised of members from different professions and having different demographic attributes, making it representative of the general Korean population. This computerized database contains longitudinal records of all claims data from anonymised patients, including diagnostic codes, treatment details, monthly insurance premiums, prescriptions, laboratory clinical results, physician visits, and demographic information. The database uses a disease diagnostic coding system based on the Korean Classification of Diseases, Sixth Revision, which is compatible with international classification of diseases, 10th revision, clinical modification (ICD-10-CM) [37]. The detailed description of this cohort profile has been published elsewhere [38].

2. Data extraction and candidate predictors

Participants were assigned randomly to the derivation and validation samples based on the split sample method using a ratio of 2:1 [39]. Based on literature reviews and established hypotheses, we extracted data on disease diagnosis, date of diagnosis, prescription, sex, age, insurance premium as a proxy for income grade (socioeconomic status), anthropometric measures, smoking status and alcohol consumption, physical activity, fasting blood glucose, total cholesterol, blood pressure measures, days seeking care (consultations), days of prescription (length of prescription), days of hospitalization and premorbid conditions (diabetes, CAD, hypertension [HTN] and DD, history of back and hip injury, spinal stenosis, BMD disorders, and spondylolisthesis). We used ICD-10-CM codes to extract premorbid conditions such as diabetes (E10-E14), CAD (I20-I25), HTN (I10-I15), intervertebral DD (IVDD) (M50-M518), disorders of bone density and structure (M80-M85), spinal stenosis (M480.0-M480.8), spondylolisthesis (M4310-M4318), and injury of back, hip, and thigh (S130-S139, S330-S3319, and S70-S79, respectively).

3. Assessment of covariates

Linear variables were assessed as both continuous and categorical predictors; however, to construct simplified risk scores, continuous predictors were categorized into clinically meaningful groups. Body mass index (BMI) was categorised as < 18.5 kg/m^2, ≥ 18.5 kg/m^2 to 24.9 kg/m^2, ≥ 25 kg/m^2 to 29.9 kg/m^2, and ≥ 30 kg/m^2; blood pressure or HTN status was categorised based on diastolic blood pressure (DBP) and systolic blood pressure (SBP) as SBP

< 120 mmHg and DBP < 80 mmHg, SBP ≥ 120-139 mmHg or DBP ≥ 80-89 mmHg, SBP ≥ 140-159 mmHg or DBP ≥ 90-99 mmHg, and SBP ≥ 160 mmHg or DBP ≥ 100 mmHg, or medical utilisation due to HTN. Fasting blood glucose was categorized as < 100 mg/dL, 100-125 mg/dL, and ≥ 126 mg/dL, or medical utilisation due to diabetes. Total cholesterol was categorised as < 200 mg/dL, 200-239 mg/dL, and ≥ 240 mg/dL. Smoking was grouped into those who had never smoked, former smokers, and current smokers, and alcohol consumption was categorized into rare drinkers (< 2 times/mo), moderate drinkers (2-3 times/mo), and heavy drinkers (≥ 4 times/mo). Physical activity levels were categorised based on frequency of physical activity per week into low (none), moderately active (1-3 times/wk), and very active (≥ 4 times/wk). Socioeconomic status was categorised based on monthly insurance premiums, on a scale of 100% to proxy income grade, into low (< 30%), medium (30%-60%), and high (> 60%). Baseline age was categorised as < 45 years, 45-54 years, 55-64 years, and ≥ 65 years. We imputed missing data using covariate values measured at the nearest time points.

4. Outcomes and exclusion criteria

The primary outcome of interest was time to first diagnosis of LBP among LBP-free participants at baseline. We used four categories of LBP ICD-10-CM codes, including codes for LBP (M54.5, M54.50, M54.51, M54.52, M54.53, M54.54, M54.55, M54.56, M54.57, M54.58, and M54.59), codes for lumbago with sciatica (M54.4, M54.40, M54.41, M54.42 M54.43, M54.44, M54.45, M54.46, M54.47, M54.48, and M54.49), codes for sciatica (M54.3, M54.30, M54.31, M54.32, M54.33, M54.34, M54.35, M54.36, M54.37, M54.38, and M54.39) and unspecified dorsalgia (M54, M54.9, M54.90, M54.91, M54.92, M54.93, M54.94, M54.95, M54.96, M54.97, M54.98, and M54.99). The above categories have been used in case definitions of LBP in other studies [40-42]. The earliest recorded date for the LBP diagnosis was the index date for the diagnosis, and participants with history of LBP at baseline were excluded in the analysis of the primary outcome. Participants with no recorded LBP during follow-up were censored at the last recorded date, death, or the study end date (December 31, 2010). We defined person years at risk as the difference between the entry date and the right censoring date. The secondary outcome was time to LBP recurrence within 5-years following the index diagnosis in a consecutive cohort of LBP patients. A recurrence was defined as an episode occurring after at least 90 days from the index date of LBP diagnosis. For recurrence, we defined person years at risk as the difference between the initial date of LBP diagnosis and the right censoring date, and LBP patients who never experienced a recurrence were censored at the last recorded date, death, or the end of the 5-year follow up period.

5. Statistical analysis

1) Models derivation

We used the Cox proportional hazards model to assess associations between risk predictors and LBP. We first conducted two analyses: univariate analysis and fully adjusted analysis adjusting for age, sex, income grade, smoking status, physical activity, and alcohol consumption. We checked the Cox proportional hazards assumptions graphically and assessed the functional form of covariates for linearity using cumulative martingale residuals, Schoenfeld residue plots, and the Kolmogorov-type supremum test based on 1,000 simulation patterns. We used hierarchical cluster analysis to select the most representative variable for each cluster of correlated variables and assessed estimated coefficients for predictors in the univariate analysis to select representative predictors for model derivation. The models were fitted and variables retained if they were significant at $\alpha = 0.01$ using the backward selection procedure. To construct a risk score, the estimated β coefficient for each variable was multiplied by 100 and rounded to the nearest integer, with the total score obtained by a summation of the scores for each predictor. The same model derivation approach was used in the derivation of models for the primary and secondary outcomes.

2) Measures of predictive performance

We assessed calibration using the Hosmer–Lemeshow (H–L) type χ^2 statistic as extended for survival data by Nam and D'Agostino [43]. We calculated this statistic by dividing the data into 10 groups (deciles) based on the predicted probabilities, and the average predicted probabilities for the deciles were compared to the actual risk probabilities of LBP. The associated calibration graph was obtained. The model discrimination was evaluated based on Harrell's C-statistic, a modification of the area under the receiver-operating characteristic (ROC) curve adapted to survival data [44], and also calculated the positive predictive value (PPV), the negative predictive value (NPV), sensitivity and specificity, and accuracy based on optimal cutoff values determined by Youden's index. In addition, the brier score which simultaneously measures calibration and discrimination was calculated. All analyses were conducted using SAS ver. 9.4 (SAS Institute Inc., Cary, NC), and R v.3.5.2 (R Foundation, Vienna, Austria). This study was approved by the Institutional Review Board of the Seoul National University (No. E1811/002-009).

Table 1. Baseline Characteristics and Prevalence of Risk Factors in Cohort at Baseline (N = 435,968)

Covariate	Derivation cohort (n = 290,879)			Validation cohort (n = 145,089)		
	Without LBP (n = 195,315)	With LBP (n = 95,564)	P value	Without LBP (n = 97,257)	With LBP (n = 47,832)	P value
Follow-up (yr)	8.4 ± 1.2	5.2 ± 2.0	< 0.001	8.4 ± 1.2	5.2 ± 2.0	< 0.001
Height (cm)	164.7 ± 8.6	162.0 ± 9.0	< 0.001	164.7 ± 8.6	161.9 ± 9.0	< 0.001
Weight (kg)	62.9 ± 11.4	62.2 ± 10.9	< 0.001	62.9 ± 11.4	62.0 ± 10.8	< 0.001
BMI (kg/m^2)	23.1 ± 3.3	23.6 ± 3.2	< 0.001	23.1 ± 3.3	23.6 ± 3.3	< 0.001
SBP (mmHg)	121.8 ± 16.9	123.7 ± 17.2	< 0.001	121.8 ± 17.0	123.6 ± 17.3	< 0.001
DBP (mmHg)	77.0 ± 11.5	77.8 ± 11.4	< 0.001	77.0 ± 11.5	77.7 ± 11.4	< 0.001
Fasting blood glucose (mg/dL)	92.9 ± 28.6	94.9 ± 29.0	< 0.001	92.9 ± 28.6	94.8 ± 28.9	< 0.001
Total cholesterol (mg/dL)	189.9 ± 38.6	194.4 ± 38.5	< 0.001	189.8 ± 38.4	194.2 ± 38.4	< 0.001
Sex			< 0.001			< 0.001
M	110,325 (56.5)	43,025 (45.0)		54,967 (56.5)	21,362 (44.7)	
F	84,990 (43.5)	52,539 (55.0)		42,290 (43.5)	26,470 (55.3)	
Age (yr)			< 0.001			< 0.001
< 45	137,778 (70.5)	48,080 (50.3)		68,566 (70.5)	24,180 (50.5)	
45-54	32,963 (16.9)	21,452 (22.4)		16,330 (16.8)	10,861 (22.7)	
55-64	16,970 (8.7)	16,511 (17.3)		8,461 (8.7)	8,026 (16.9)	
≥ 65	7,604 (3.9)	9,521 (10.0)		3,900 (4.0)	4,765 (9.9)	
Income/Insurance premium (%)			< 0.001			< 0.001
Low (< 30)	28,157 (14.4)	14,545 (15.2)		13,816 (14.2)	7,408 (15.5)	
Medium (30-60)	70,623 (36.2)	34,175 (35.8)		35,253 (36.3)	16,982 (35.5)	
High (> 60)	96,535 (49.4)	46,844 (49.0)		48,188 (49.5)	23,442 (49.0)	
Physical activity (times/wk)			< 0.001			< 0.001
Low (none)	110,840 (56.8)	57,097 (59.8)		55,485 (57.0)	28,622 (59.8)	
Moderate (1-3)	72,319 (37.0)	32,749 (34.3)		35,674 (36.7)	16,350 (34.2)	
High (≥ 4)	12,156 (6.2)	5,718 (5.9)		6,098 (6.3)	2,860 (6.0)	
Smoking status			< 0.001			< 0.001
Never	120,959 (61.9)	67,877 (71.0)		60,189 (61.9)	34,035 (71.2)	
Former smoker	9,444 (4.9)	3,880 (4.1)		4,738 (4.9)	1,932 (4.0)	
Current smoker	64,912 (33.2)	23,807 (24.9)		32,330 (33.2)	11,865 (24.8)	
Alcohol consumption (times/wk)			< 0.001			< 0.001
Rarely (< 2)	90,872 (46.5)	52,473 (54.9)		45,327 (46.6)	25,976 (54.3)	
Moderate drinker (2-3)	86,181 (44.1)	35,873 (37.5)		42,745 (44.0)	18,125 (37.9)	
Heavy drinker (≥ 4)	18,262 (9.4)	7,218 (7.6)		9,185 (9.4)	3,731 (7.8)	
BMI (kg/m^2)			< 0.001			< 0.001
< 18.5	14,019 (7.2)	4,728 (4.9)		6,906 (7.1)	2,354 (4.9)	
18.5-24.9	128,129 (65.6)	59,985 (62.8)		64,085 (65.9)	30,216 (63.2)	
25.0-29.9	48,482 (24.8)	28,010 (29.3)		24,020 (24.7)	13,801 (28.9)	
≥ 30.0	4,685 (2.4)	2,841 (3.0)		2,246 (2.3)	1,461 (3.1)	
Total cholesterol (mg/dL)			< 0.001			< 0.001
< 200	120,099 (61.5)	54,205 (56.7)		59,908 (61.6)	27,255 (57.0)	
200-239	55,153 (28.2)	29,725 (31.1)		27,582 (28.4)	14,841 (31.0)	
≥ 240	20,063 (10.3)	11,634 (12.2)		9,767 (10.0)	57,36 (12.0)	
Fasting blood glucose/Diabetes (mg/dL)			< 0.001			< 0.001
< 100	125,157 (64.1)	58,422 (61.1)		62,381 (64.1)	29,355 (61.4)	
100-125	48,272 (24.7)	24,980 (26.2)		24,009 (24.7)	12,385 (25.9)	
≥ 126 or Tx	21,886 (11.2)	12,162 (12.7)		10,867 (11.2)	6,092 (12.7)	
Blood pressure/HTN (mmHg)			< 0.001			< 0.001
SBP < 120 and DBP < 80	76,923 (39.4)	34,366 (35.9)		38,441 (39.5)	17,341 (36.3)	
SBP 120-139 or DBP 80-89	98,598 (50.5)	49,866 (52.2)		48,715 (50.1)	24,829 (51.9)	
SBP 140-159 or DBP 90-99	17,676 (9.0)	10,110 (10.6)		9,021 (9.3)	4,994 (10.4)	
SBP ≥ 160 or DBP ≥ 100 or Rx	2,118 (1.1)	1,222 (1.3)		1,080 (1.1)	668 (1.4)	
Diagnosed IHD			< 0.001			< 0.001
No	193,983 (99.3)	94,223 (98.6)		96,543 (99.3)	47,159 (98.6)	
Yes	1,332 (0.7)	1,341 (1.4)		714 (0.7)	673 (1.4)	
Diagnosed IVDD			< 0.001			< 0.001
No	188,344 (96.4)	86,792 (90.8)		93,814 (96.5)	43,393 (90.7)	
Yes	6,971 (3.6)	8,772 (9.2)		3,443 (3.5)	4,439 (9.3)	
History of back injuries			< 0.001			< 0.001
No	194,695 (99.7)	94,824 (99.2)		96,918 (99.6)	47,444 (99.2)	
Yes	620 (0.3)	740 (0.8)		339 (0.4)	388 (0.8)	

Covariate	Derivation cohort (n = 290,879)			Validation cohort (n = 145,089)		
	Without LBP (n = 195,315)	With LBP (n = 95,564)	P value	Without LBP (n = 97,257)	With LBP (n = 47,832)	P value
BMD disorders			< 0.001			< 0.001
No	191,838 (98.2)	90,198 (94.4)		95,531 (98.2)	45,183 (94.5)	
Yes	3,477 (1.8)	5,366 (5.6)		1,726 (1.8)	2,649 (5.5)	
Spinal stenosis			< 0.001			< 0.001
No	194,495 (99.6)	93,948 (98.3)		96,829 (99.6)	46,996 (98.2)	
Yes	820 (0.4)	1,616 (1.7)		428 (0.4)	836 (1.8)	
Spondylolisthesis			< 0.001			< 0.001
No	195,251 (99.97)	95,435 (99.9)		97,221 (99.96)	47,769 (99.87)	
Yes	64 (0.03)	129 (0.1)		36 (0.04)	63 (0.13)	

Values are presented as mean ± standard deviation or number (%).
LBP: low back pain, BMI: body mass index, Tx: medication due to diabetes, HTN: hypertension, Rx: medication due to HTN, SBP: systolic blood pressure, DBP: diastolic blood pressure, IHD: ischemic heart disease, IVDD: intervertebral disc degeneration, BMD: bone mineral density.

RESULTS

1. Cohort baseline characteristics

The extracted data consisted of 502,342 participants. We excluded 66,374 participants who had experienced at least one episode of LBP before 1st January 2004. During the median follow-up of 8.4 years (2.0-8.9 yr), there were 95,564 (32.9%) and 47,832 (33.0%) newly diagnosed (first onset) LBP cases among 290,879 and 145,089 participants in the derivation and validation cohorts, respectively. The total number of person-years of follow-up was 3,205,271 years. The mean (standard deviation) of the covariates and the distribution of the baseline characteristics among cohorts are presented in **Table 1**. There were no discrepancies between the derivation and validation cohorts. However, there was a significant difference in baseline characteristics between those who developed and those who did not develop LBP (*P* < 0.001; **Table 1**). A total of 143,396 (32.9%) participants who experienced at least one LBP episode were randomly assigned to the derivation and validation cohorts at a ratio of 2:1. During a median follow-up of 2.9 years (1.7-5.0 yr), there were 61,494 (64.2%) and 30,668 (64.4%) cases of recurrent LBP among 95,763 and 47,633 participants in the derivation and validation cohorts, respectively (**Supplementary Table 1**).

2. Model derivation

The covariates assessed in the derivation of the risk prediction model for the primary outcome consisted of age, sex, income grade, smoking status, alcohol consumption, physical activity, BMI, fasting blood glucose/diabetes, blood pressure/HTN, total cholesterol, IVDD, history of back injury, spinal stenosis, history of BMD disorders, and spondylolisthesis. **Table 2** and **Supplementary Table 2**

presents the estimated coefficients and hazard ratios (HRs) for each covariate in the univariate and adjusted analyses of the primary and secondary outcomes, respectively. Based on univariate associations and after an assessment of multicollinearity, a total of 15 covariates that were significantly associated with LBP were assessed in the model derivation and retained if they were significant at α = 0.01. Newly diagnosed (first onset) LBP was predicted by 11 variables including DD, sex, age, socioeconomic status, physical exercise, alcohol consumption, BMI, total cholesterol, HTN, spinal stenosis, and BMD disorders. **Table 3** presents the estimated coefficients and HRs for predictors in the parsimonious model, which were retained from multivariate analysis at α = 0.01 using the backward selection procedure. In the right-most column of the table are the points associated with the presence of a given level of a risk factor (with the reference level assigned zero points). The 5-year LBP recurrence was predicted by seven variables including age, sex, income grade, and BMI, IVDD, spondylolisthesis, and total number of days of prescription (**Supplementary Table 3**).

3. Model performance evaluation

Table 4 and **Supplementary Table 4** present the model discrimination and calibration for the primary outcome and secondary outcomes, respectively. In the validation cohorts, the prediction models discriminated between patients with and without LBP or recurrence with Harrell's C-statistic of 0.812 (95% confidence interval [CI], 0.804-0.820) and 0.916 (95% CI, 0.907-0.924), for the newly diagnosed LBP and 5-year recurrence models, respectively. The calibration based on the Hosmer–Lemeshow (Nam and D'Agostino test) was ($\chi^2 = 7.6178$, $P = 0.3407$) and ($\chi^2 = 1.3981$, $P = 0.8752$) in the validation cohorts, respectively [43]. The Nam–D'Agostino test evaluates model calibration; a large

Table 2. HR, 95% CI, and β-Coefficients for Risk Predictors in the Univariate Analysis and Adjusted Analysis

Covariate	Unadjusted[a]			Adjusted[b]		
	β (SE)	HR (95% CI)	P value	β (SE)	HR (95% CI)	P value
Sex						
F	0.351 (0.005)	1.42 (1.41-1.44)	< 0.001	0.331 (0.007)	1.39 (1.37-1.41)	< 0.001
Age (yr)						
< 45	Reference	Reference		Reference	Reference	
45-54	0.501 (0.007)	1.65 (1.63-1.67)	< 0.001	0.505 (0.007)	1.66 (1.63-1.68)	< 0.001
55-64	0.805 (0.007)	2.24 (2.21-2.27)	< 0.001	0.810 (0.008)	2.25 (2.22-2.28)	< 0.001
≥ 65	1.07 (0.009)	2.91 (2.86-2.97)	< 0.001	1.061 (0.009)	2.89 (2.84-2.94)	< 0.001
Income/Insurance premium (%)						
High (> 60)	Reference	Reference		Reference	Reference	
Medium (30-60)	0.001 (0.006)	1.00 (0.99-1.01)	0.8166	0.054 (0.006)	1.06 (1.04-1.07)	< 0.001
Low (< 30)	0.065 (0.008)	1.07 (1.05-1.08)	< 0.001	0.029 (0.008)	1.03 (1.01-1.05)	< 0.001
Physical activity (times/wk)						
High (≥ 4)	Reference	Reference		Reference	Reference	
Moderate (1-3)	–0.028 (0.012)	0.97 (0.95-0.99)	< 0.001	–0.006 (0.012)	0.99 (0.97-1.02)	0.0012
Low (none)	0.074 (0.011)	1.08 (1.05-1.10)	< 0.001	0.005 (0.011)	1.01 (0.98-1.03)	0.0693
Smoking status						
Never	Reference	Reference		Reference	Reference	
Former	–0.248 (0.014)	0.78 (0.76-0.80)	< 0.001	–0.007 (0.014)	0.99 (0.97-1.02)	0.6060
Current smoker	–0.331 (0.006)	0.72 (0.71-0.73)	< 0.001	–0.016 (0.008)	0.98 (0.96-1.00)	0.0519
Alcohol consumption (times/mo)						
Rarely (< 2)	Reference	Reference		Reference	Reference	
Moderate (2-3)	–0.253 (0.006)	0.78 (0.77-0.79)	< 0.001	–0.031 (0.006)	0.97 (0.96-0.98)	< 0.001
Heavy (≥ 4)	–0.287 (0.010)	0.75 (0.74-0.77)	< 0.001	–0.007 (0.011)	0.99 (0.97-1.01)	0.5157
BMI (kg/m²)						
< 18.5	–0.278 (0.012)	0.76 (0.74-0.78)	< 0.001	–0.24452 (0.013)	0.783 (0.764-0.802)	< 0.001
18.5-24.9	Reference	Reference		Reference	Reference	
25.0-29.9	0.160 (0.006)	1.17 (1.16-1.19)	< 0.001	0.105 (0.006)	1.11 (1.10-1.12)	< 0.001
≥ 30.0	0.211 (0.016)	1.24 (1.20-1.27)	< 0.001	0.134 (0.016)	1.14 (1.11-1.18)	< 0.001
Total cholesterol (mg/dL)						
< 200	Reference	Reference		Reference	Reference	
200-239	0.141 (0.006)	1.15 (1.14-1.17)	< 0.001	0.035 (0.006)	1.04 (1.02-1.05)	< 0.001
≥ 240	0.202 (0.008)	1.22 (1.20-1.24)	< 0.001	0.007 (0.009)	1.01 (0.99-1.02)	0.4208
Fasting blood glucose (mg/dL)						
< 100	Reference	Reference		Reference	Reference	
100-125	0.082 (0.006)	1.09 (1.07-1.10)	< 0.001	0.008 (0.006)	1.01 (1.00-1.02)	0.1885
> 126 or Tx	0.139 (0.008)	1.15 (1.13-1.17)	< 0.001	–0.018 (0.008)	0.98 (0.97-0.99)	0.0318
Diagnosed HTN (mmHg)						
SBP < 120 and DBP < 80	Reference	Reference		Reference	Reference	
SBP 120-139 or DBP 80-89	0.110 (0.006)	1.12 (1.10-1.13)	< 0.001	0.028 (0.006)	1.03 (1.02-1.04)	< 0.001
SBP 140-159 or DBP 90-99	0.200 (0.009)	1.22 (1.20-1.24)	< 0.001	–0.010 (0.010)	0.99 (0.97-1.01)	0.2894
SBP ≥ 160 or DBP ≥ 100 or Rx	0.244 (0.023)	1.28 (1.22-1.34)	< 0.001	–0.043 (0.024)	0.96 (0.92-1.00)	0.0706
BMI (kg/m²)	0.037 (0.0008)	1.037 (1.036-1.039)	< 0.001	0.027 (0.00082)	1.027 (1.025-1.029)	< 0.001
Fasting blood glucose (mg/dL)	0.002 (0.00009)	1.002 (1.002-1.002)	< 0.001	–0.00004 (0.00009)	1.000 (1.000-1.000)	0.6896
Total cholesterol (mg/dL)	0.002 (0.00007)	1.002 (1.002-1.003)	< 0.001	0.0004 (0.00007)	1.000 (1.000-1.001)	< 0.001
SBP (mmHg)	0.006 (0.00015)	1.006 (1.005-1.006)	< 0.001	0.0005 (0.0002)	1.000 (1.000-1.001)	0.0048
DBP (mmHg)	0.005 (0.0002)	1.005 (1.004-1.005)	< 0.001	0.0006 (0.00024)	1.001 (1.000-1.001)	0.0081
Premorbid diseases						
History of diabetes	0.474 (0.012)	1.61 (1.57-1.65)	< 0.001	0.111 (0.013)	1.12 (1.09-1.15)	< 0.001
History of hypertension	0.530 (0.008)	1.70 (1.67-1.73)	< 0.001	0.101 (0.009)	1.11 (1.09-1.13)	< 0.001
Diagnosed IHD	0.542 (0.022)	1.72 (1.65-1.80)	< 0.001	0.177 (0.023)	1.19 (1.14-1.25)	< 0.001
Diagnosed IVDD	0.781 (0.009)	2.18 (2.14-2.22)	< 0.001	0.587 (0.009)	1.80 (1.77-1.83)	< 0.001
History of back injury	0.680 (0.030)	1.97 (1.86-2.09)	< 0.001	0.541 (0.030)	1.72 (1.62-1.82)	< 0.001
Spinal stenosis	1.07 (0.020)	2.93 (2.81-3.05)	< 0.001	0.591 (0.021)	1.81 (1.74-1.88)	< 0.001
BMD disorders	0.887 (0.012)	2.43 (2.37-2.48)	< 0.001	0.383 (0.012)	1.47 (1.43-1.50)	< 0.001
Spondylolisthesis	0.979 (0.072)	2.66 (2.31-3.07)	< 0.001	0.563 (0.072)	1.76 (1.52-2.02)	< 0.001

HR: hazard ratio, CI: confidence interval, SE: standard error, BMI: body mass index, Tx: medication due to diabetes, HTN: hypertension, SBP: systolic blood pressure, DBP: diastolic blood pressure, Rx: medication due to HTN, IHD: ischemic heart disease, IVDD: intervertebral disc degeneration, BMD: bone mineral density.
[a]Univariate analysis. [b]Fully adjusted accounting for age, sex, income grade, physical activity, smoking status, alcohol consumption, and each of the other risk factors (fasting blood glucose/diabetes, total cholesterol, blood pressure/HTN, prior history of [diabetes, HTN, IHD, IVDD, history of back injury, spinal stenosis, BMD disorders, spondylolisthesis]).

Table 3. HR, 95% CI, and β-Coefficients for Risk Predictors in the Parsimonious Prediction Models of LBP (Primary Outcome)

Covariate	β (SE)	HR (95% CI)	P value	Risk point[a]
Sex				
M	Reference	Reference		0
F	0.315 (0.008)	1.37 (1.35-1.39)	< 0.001	32
Age (yr)				
< 45	Reference	Reference		0
45-54	0.437 (0.009)	1.55 (1.52-1.58)	< 0.001	44
55-64	0.729 (0.010)	2.07 (2.04-2.11)	< 0.001	73
≥ 65	0.979 (0.012)	2.66 (2.60-2.72)	< 0.001	98
Income/Insurance premium (%)				
High (> 60)	Reference	Reference		0
Medium (30-60)	0.063 (0.008)	1.07 (1.05-1.08)	< 0.001	6
Low (< 30)	0.036 (0.010)	1.04 (1.02-1.06)	< 0.001	4
Physical activity (times/wk)				
High (≥ 4)	Reference	Reference		0
Moderate (1-3)	−0.007 (0.014)	0.99 (0.97-1.02)	0.0030	−1
Low (none)	0.017 (0.014)	1.02 (0.99-1.05)	0.0589	2
Alcohol consumption (times/mo)				
Rarely (< 2)	Reference	Reference		0
Moderate drinker (2-3)	−0.041 (0.008)	0.96 (0.95-0.97)	< 0.001	−4
Heavy drinker (≥ 4)	−0.021 (0.013)	0.98 (0.95-1.01)	0.1243	−2
BMI (kg/m^2)				
< 18.5	−0.231 (0.015)	0.79 (0.77-0.82)	< 0.001	−23
18.5-24.9	Reference	Reference		0
25.0-29.9	0.104 (0.007)	1.11 (1.09-1.13)	< 0.001	10
≥ 30	0.121 (0.019)	1.13 (1.09-1.17)	< 0.001	12
Blood pressure/HTN				
SBP < 120 and DBP < 80	Reference	Reference		0
SBP 120-139 or DBP 80-89	0.0004 (0.007)	1.00 (0.99-1.02)	0.9570	0
SBP 140-159 or DBP 90-99	−0.036 (0.012)	0.97 (0.94-0.99)	0.0031	−4
SBP ≥ 160 or DBP ≥ 100 or Rx	−0.082 (0.030)	0.92 (0.87-0.98)	0.0054	−8
Total cholesterol (mg/dL)				
< 200	Reference	Reference		0
200-239	0.014 (0.007)	1.01 (1.00-1.03)	0.0513	1
≥ 240	−0.029 (0.011)	0.97 (0.95-0.99)	0.0053	−3
Diagnosed IVDD				
No	Reference	Reference		0
Yes	0.532 (0.012)	1.70 (1.67-1.74)	< 0.001	53
Spinal stenosis				
No	Reference	Reference		0
Yes	0.207 (0.028)	1.23 (1.17-1.30)	< 0.001	21
History of BMD disorders				
No	Reference	Reference		0
Yes	0.254 (0.016)	1.29 (1.25-1.33)	< 0.001	25

HR: hazard ratio, CI: confidence interval, LBP: low back pain, SE: standard error, BMI: body mass index, HTN: hypertension, SBP: systolic blood pressure, DBP: diastolic blood pressure, Rx: medication due to HTN, IVDD: intervertebral disc degeneration, BMD: bone mineral density.
[a]Risk predictors retained in the models at α = 0.01 level of significance using backward selection procedure; The risk points were calculated by multiplying the estimated coefficients by 100 and rounding to the next integer.

P value (> 0.05) indicates a good match of predicted risk over observed risk. The Harrell's C-statistic corresponds to the area(s) under the time-dependent ROC curves, with a value of C = 0.5 corresponding to a non-informative prediction model, whereas C = 1.0 corresponds to perfect prediction. The brier score measures both discrimination and calibration with values ranging from 0.0 to 1.0, with lower values indicating better performance. In the validation cohorts, brier scores were 0.294 and 0.398 for the primary and secondary outcomes, respectively.

The Youden's J statistic suggested a risk probability of ≥ 0.795 and ≥ 0.430 as the optimal cutoff points to define high-risk individuals based on the derived prediction equations for the first onset of LBP and 5-year recurrence models, respectively. These thresholds showed an accuracy of 0.786, PPV of 0.825, and NPV of 0.777, and an accuracy of 0.694, PPV of 0.724, and NPV of 0.617 in the validation cohorts of first onset and 5-year recurrence models, respectively. The details of other model performance measures for the low back pain onset and recurrence prediction models and simplified risk scores are presented in **Table 4** and **Fig. 1**, **Supplementary Fig. 1**, as well as in

Table 4. Discrimination and Calibration of Models in Derivation and Validation Cohorts (Primary Outcome)

Model and risk score performance metrics	Derivation cohort	Validation cohort
Model based performance		
Brier score[a]	0.293	0.294
Nam–D'Agostino test[b]	$\chi^2 = 7.5930$, $P = 0.3031$	$\chi^2 = 7.6178$, $P = 0.3407$
Harrell's C statistic (95% CI)[c]	0.810 (0.805-0.816)	0.812 (0.804-0.820)
Sensitivity (95% CI)	0.442 (0.439-0.446)	0.443 (0.438-0.447)
Specificity (95% CI)	0.954 (0.953-0.955)	0.954 (0.953-0.956)
Positive predictive value (95% CI)	0.825 (0.822-0.828)	0.825 (0.821-0.830)
Negative predictive value (95% CI)	0.777 (0.776-0.778)	0.777 (0.776-0.779)
Accuracy (95% CI)	0.786 (0.784-0.787)	0.786 (0.784-0.788)
Risk score performance		
Sensitivity (95% CI)	0.209 (0.207-0.212)	0.211 (0.207-0.215)
Specificity (95% CI)	0.928 (0.927-0.929)	0.929 (0.928-0.931)
Positive predictive value (95% CI)	0.590 (0.585-0.595)	0.597 (0.590-0.604)
Negative predictive value (95% CI)	0.705 (0.704-0.705)	0.705 (0.704-0.706)
Accuracy (95% CI)	0.692 (0.690-0.693)	0.693 (0.690-0.695)

CI: confidence interval.

[a]Measures both discrimination and calibration; lower values indicate higher accuracy. [b]A modification of Hosmer–Lemeshow test suited for survival data; measure of calibration that is specific to censored survival data (lower χ^2 and higher P values) indicate better calibration. [c]A measure of discrimination for which higher values indicate better discrimination.
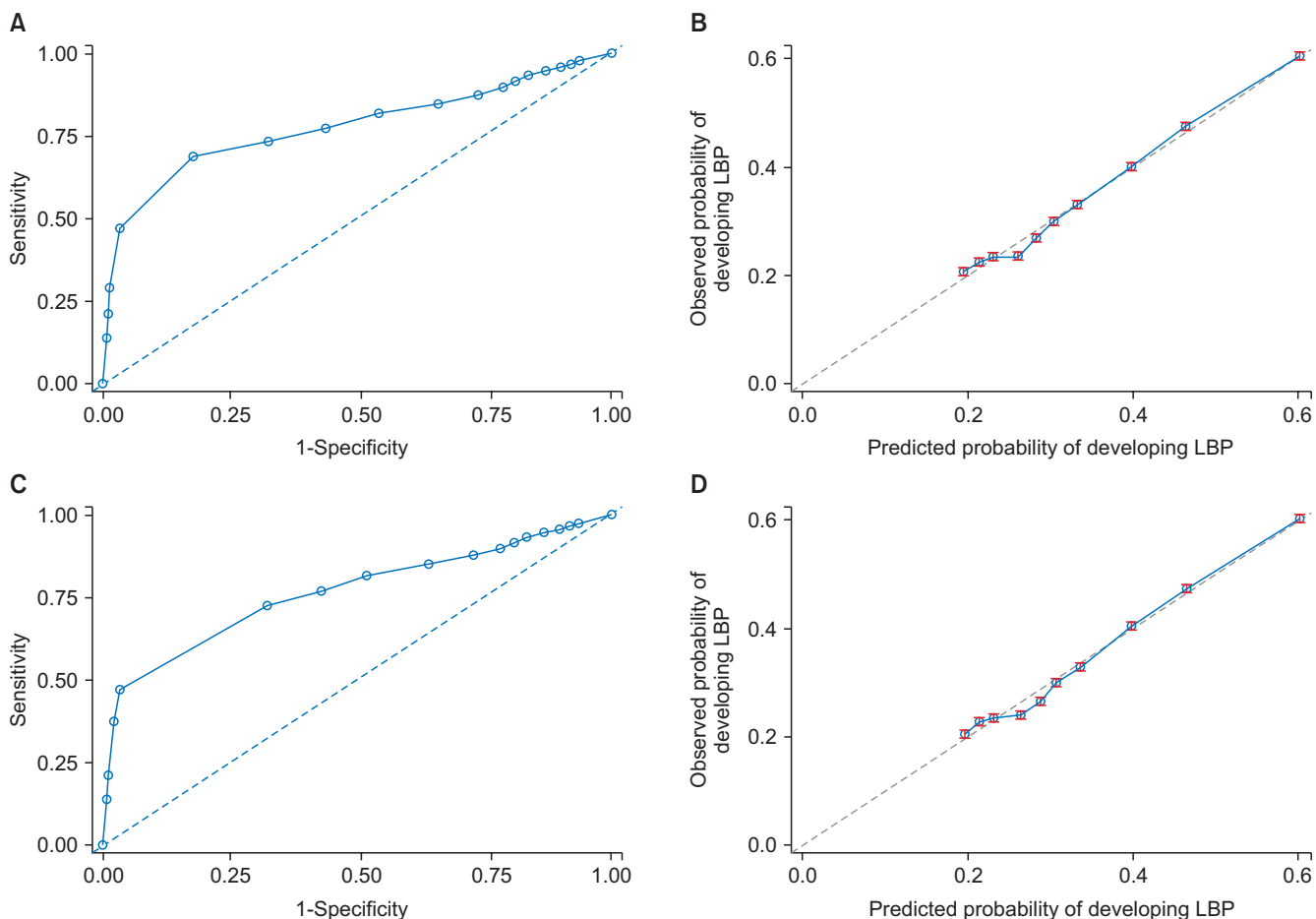

Fig. 1. Discrimination and calibration plots in the derivation and validation cohorts for low back pain (LBP) primary outcome model. (A, C) Discrimination. (B, D) Calibration.

Supplementary Table 4 [43].

4. Prediction equations

Basing on the parsimonious models, individualized probability of developing LBP within the years of follow-up (t = 8 yr), or its recurrence (t = 5 yr) for an individual with covariate values $\chi = (\chi_1, \ldots, \chi_K)$ for K risk factors can be estimated using the following equation:

$$P(\text{LBP}) = 1 - S_o(t)^{\exp[(fx)]}, \text{ where}$$

$$f(x) = \sum_i \beta_i x_i$$

$S_o(t)$ is the baseline survival probability at time (t) for an individual with all covariates equivalent to zero (0), and the β_i are the estimated coefficients from the Cox proportional hazard models. Using the estimated coefficients β_i and survival probabilities $S_o(t)$, personalized probabilities of developing LBP or its recurrence can be calculated.

5. Simplified risk score for the primary outcome

The observed minimum and maximum sum of risk points were –33 and 249, respectively. The median risk score was 36, while the 25th and 75th percentiles were 7 and 75, respectively. The Youden's J statistic suggested a risk score of ≥ 103 as the optimal cutoff point to define high-risk individuals based on the simplified risk score. This threshold showed an accuracy of 0.693, NPV of 0.705, and PPV of 0.597 in the validation cohort.

6. Simplified risk score for the secondary outcome

The observed minimum and maximum sum of risk points were –14 and 120, respectively. The median risk score was 23, while the 25th and 75th percentiles were 7 and 45, respectively. The Youden's J statistic suggested a risk score of ≥ 4 as the optimal cutoff point to define high-risk individuals based on the 5-year recurrence simplified risk score. This threshold showed an accuracy of 0.628, PPV of 0.646, and NPV of 0.499 in the validation cohort.

7. Practical application of the risk score for first onset LBP

Based on **Table 5**, the following example illustrates how LBP risk can be estimated using the simplified points system.

Case: A 52-year-old female with an income grade of 30%-60%, who is a moderate drinker (2-3 times/mo) with low physical exercise (none), obese (BMI > 30 kg/m^2), with normal blood pressure (SBP < 120 mmHg and DBP < 80 mmHg) and normal total cholesterol (< 200 mg/dL), without spinal stenosis but with IVDD and BMD disorders can have her probability of LBP estimated, based on the point system, as follows:

$$P(\text{LBP}) = 1 - S_o(t)^{\exp[(score/100)]}$$
$$= 1 - (0.6334)^{\exp[(170/100)]} \qquad = 0.918$$

The $S_o(t)$ is the baseline survival probability at time (t = 8 yr) for an individual with all covariates equivalent to zero, which was estimated by Cox regression analysis.

8. Practical application of the 5-year LBP recurrence risk score

Based on **Supplementary Table 5**, the following example illustrates how the 5-year LBP recurrence risk can be estimated using the simplified points system.

Case: A 36-year-old female with income grade of 30%-60%, low BMI (< 18.5 kg/m^2), without a history of IVDD but with a previous diagnosis of spondylolisthesis and who receive low back treatment for more than 8 days during the initial LBP episode can have her probability of 5-year LBP recurrence estimated, based on the point system, as follows:

$$P(\text{Five yr LBP recurrene}) = 1 - S_o(t)^{\exp[(score/100)]}$$
$$= 1 - (0.2704)^{\exp[(36/100)]} \qquad = 0.847$$

The $S_o(t)$ is the baseline survival probability at time (t = 5 yr) for an individual with all covariates equivalent to zero, which was estimated by Cox regression analysis. The beta coefficient was set to an integer by multiplying by 100, and also was tabulated for both outcomes (**Table 3** and **Supplementary Table 3**). Thus, in an actual calculation, the sum of risk scores should be divided by 100 to give an overall risk estimate. The examples shown above are thus divided by 100 to give the 8-year risk and 5-year recurrence risk, respectively.

DISCUSSION

This study was based on a large, representative Korean population with data obtained from a well-established national cohort [38]. The prevalence of LBP in this study was comparable with previous studies [6-9,45]. Risk prediction equations based on general medical practice data are easily implemented in medical practice [46], and we believe our results are applicable to the general Korean popula-

Table 5. Calculated Score for a Hypothetical Example of a LBP Risk Profile

Risk factor	Value	Point
Sex	F	32
Age (yr)	45-54	44
Insurance premium/Income grade (%)	Medium (30-60)	6
Physical activity	Low (none)	2
Alcohol consumption (times/mo)	Moderate drinker	–4
BMI (kg/m^2)	≥ 30	12
Blood pressure/HTN (mmHg)	SBP < 120 and DBP < 80	0
Total cholesterol (mg/dL)	< 200	0
Diagnosed IVDD	Yes	53
Spinal stenosis	No	0
BMD disorders	Yes	25
Total point		170
Estimate of risk		0.918

$S_0(t)$ = 0.6334.
LBP: low back pain, BMI: body mass index, HTN: hypertension, IVDD: intervertebral disc degeneration, BMD: bone mineral density.

tion. LBP is predictable and individuals can reduce their risks by modification of lifestyle risk factors and managing associated premorbidities. These equations were derived from a variety of candidate predictors, including demographics, anthropometrics, premorbid conditions, and several clinical measurements which individuals and clinicians are likely to know, which makes them easily applicable.

The prediction model of first onset LBP consisted of age, sex, and income grade, alcohol consumption, physical exercise, BMI, total cholesterol, blood pressure, BMD disorders, DD, and spinal stenosis. The model of the 5-year recurrence risk was comprised of age, sex, income grade, BMI, spondylolisthesis, DD, and days of prescription. Based on the simplified risk scores, age, DD, and sex conferred the highest risk points for LBP onset, with maximum possible risk points of 98, 53, and 32, respectively, whereas age, spondylolisthesis, and DD conferred the highest risk for recurrence with 51, 34, and 16 risk points, respectively. Low BMI, moderate physical activity, moderate alcohol consumption, and high blood pressure or antihypertensive medication were inversely associated with LBP onset in the multivariate analysis ($P < 0.01$; **Table 3**). However, most risk factors associated with complex diseases have weaker associations. Therefore, although some risk factors showed weak associations, they may significantly contribute to LBP pathogenesis.

The equations showed excellent calibration with good agreement between observed and predicted risk, which is extremely important with respect to making decisions in clinical practice. Furthermore, the models showed good discrimination abilities with Harrell's C-statistic of 0.812 (95% CI, 0.804-0.820) and 0.916 (95% CI, 0.907-0.924) in val-

idation cohorts of LBP onset and 5-year recurrence models, respectively. The equations may be useful in informing clinicians and patients about LBP risks, the prognosis of initial episodes and prevention strategies. Knowledge of personalized risk can motivate individuals to reduce their risks through appropriate interventions, thereby promoting population health and reducing societal and personal costs. The equations can be used when a clinician counsels individuals after a routine check-up by providing information regarding their risk profile and giving the precise probability of LBP or its recurrence. This will motivate lifestyle modifications and promote adherence to the treatment of some premorbidities which are predictive of LBP. Reducing risk factors associated with metabolic syndrome (MetS), proper and regular therapy for individuals with MetS and management of BMD disorders can reduce LBP risk. Furthermore, lifestyle modification can reduce LBP risk conferred through MetS components by prevention of complications from obesity, HTN, and dyslipidemia. The equations will also improve self-awareness regarding overall health status because some predictors in the models are also predictive of other health outcomes. In addition, since some modifiable risk predictors are predictive of recurrence, chronicity, and disability, the models may be somewhat useful in the motivation of individuals with a recent onset of LBP to adjust their lifestyle and reduce the risk of developing the chronic form or recurrence and associated disability. This can subsequently reduce personal and societal costs associated with LBP.

There is a lack of prospective studies attempting to derive and validate LBP risk models [35], especially using routinely collected data. Previous studies have developed prediction models from occupation cohorts [42], among acute LBP patients in relation to developing chronic LBP (CLBP) [47], and based on pain trajectories [48], among others. These studies were comprised of few participants, fewer cases, considered ergonomics and occupation related variables, and did not incorporate routinely collected medical data. Here, we have developed and validated prediction equations and simplified risk scores to estimate future risk of LBP and its recurrence among apparently healthy individuals at baseline, in a large cohort using data from general medical practice. This makes our prediction equations more applicable to the general population, and able to distinguish individuals at risk in medical practice compared to these algorithms. In addition, the equations performed well in terms of discrimination and calibration. However, the derived equations cannot be a substitute for clinical expertise, but rather augment precision in clinical decision making. We believe that knowledge of personalized risk as well as the general health status of a patient with respect to LBP risk, as well as expert knowledge from

clinicians will create a much more comprehensive picture than either one alone. The information for predictors in the derived equations can easily be obtained in clinical practice, and the points system is simple to use.

This study has the strengths of representativeness, duration of follow-up, adequate sample size, and lack of recall, respondent, and selection bias. The ICD-10-CM diagnostic codes in the Korean National Health Insurance database were evaluated and found to have good concordance with the actual health status of the individuals, based on medical charts and reports [49]. This study is based on a wide range of risk predictors that can be applied in medical practice, and which individuals are likely to know. However, our study is limited because we did not incorporate psychosocial factors, genetics, and ergonomics related variables because these are not routinely collected in general practice. In addition, a low medical care-seeking behavior has been reported among LBP patients [50], with care-seeking more common in women, individuals with poor general health, and those with more disabling or more painful episodes [50]. Therefore, it is possible that some individuals did not seek medical services for LBP, and therefore were possibly missed in some cases. Nevertheless, the LBP prevalence in this study was comparable with a previous study conducted in Korea [8]. However, we used the same underlying population for model derivation and validation; thus, careful considerations are necessary in generalizing these results to other populations.

We have developed and validated risk prediction equations and simplified risk scores to estimate LBP risk in a nationwide sample cohort using data from general medical practice. The models showed good discrimination in identifying individuals at risk of developing LBP and its recurrence. To our knowledge, this study is the first nationwide cohort study that has attempted to derive and validate LBP risk prediction models using routinely collected health data. These models will improve individual decision-making, especially motivation for lifestyle modifications, guide physicians in practice, and define groups at high risk for LBP. We recommend further studies to validate and update these prediction models using cohorts from other populations and to incorporate other predictors in other settings.

CONFLICT OF INTEREST

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

FUNDING

ORCID

David Mukasa, https://orcid.org/0000-0001-6046-2715
Joohon Sung, https://orcid.org/0000-0001-9948-0160

SUPPLEMENTARY MATERIALS

Supplementary materials can be found *via* https://doi.org/10.3344/kjp.2020.33.2.153.

REFERENCES

1. Koes BW, van Tulder MW, Thomas S. Diagnosis and treatment of low back pain. BMJ 2006; 332: 1430-4.
2. Lambeek LC, Bosmans JE, Van Royen BJ, Van Tulder MW, Van Mechelen W, Anema JR. Effect of integrated care for sick listed patients with chronic low back pain: economic evaluation alongside a randomised controlled trial. BMJ 2010; 341: c6414.
3. Andersson GB. Epidemiological features of chronic low-back pain. Lancet 1999; 354: 581-5.
4. Swain MS, Henschke N, Kamper SJ, Gobina I, Ottová-Jordan V, Maher CG. An international survey of pain in adolescents. BMC Public Health 2014; 14: 447.
5. Deyo RA, Mirza SK, Martin BI. Back pain prevalence and visit rates: estimates from U.S. national surveys, 2002. Spine (Phila Pa 1976) 2006; 31: 2724-7.
6. Manchikanti L, Singh V, Falco FJ, Benyamin RM, Hirsch JA. Epidemiology of low back pain in adults. Neuromodulation 2014; 17 Suppl 2: 3-10.
7. Lee S, Nam CM, Yoon DH, Kim KN, Yi S, Shin DA, et al. Association between low-back pain and lumbar spine bone density: a population-based cross-sectional study. J Neurosurg Spine 2013; 19: 307-13.
8. Bae YH, Shin JS, Lee J, Kim MR, Park KB, Cho JH, et al. Association between Hypertension and the Prevalence of Low Back Pain and Osteoarthritis in Koreans: a cross-sectional study. PLoS One 2015; 10: e0138790.
9. Ozguler A, Leclerc A, Landre MF, Pietri-Taleb F, Niedhammer I. Individual and occupational determinants of low back pain according to various definitions of low back pain. J Epidemiol Community Health 2000; 54: 215-20.
10. Jones GT, Macfarlane GJ. Epidemiology of low back pain in children and adolescents. Arch Dis Child 2005; 90: 312-6.
11. Kamran M, Eeric T. Epidemiology and natural history of low

back pain. Semin Spine Surg 2008; 20: 87-92.

12. Von Korff M. Studying the natural history of back pain. Spine (Phila Pa 1976) 1994; 19(18 Suppl): 2041S-6S.

13. Marras WS, Ferguson SA, Burr D, Schabo P, Maronitis A. Low back pain recurrence in occupational environments. Spine (Phila Pa 1976) 2007; 32: 2387-97.

14. Stanton TR, Latimer J, Maher CG, Hancock M. Definitions of recurrence of an episode of low back pain: a systematic review. Spine (Phila Pa 1976) 2009; 34: E316-22.

15. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015; 386: 743-800.

16. Kim YE, Park H, Jo MW, Oh IH, Go DS, Jung J, et al. Trends and patterns of burden of disease and injuries in Korea using disability-adjusted life years. J Korean Med Sci 2019; 34(Suppl 1): e75.

17. Hoy D, March L, Brooks P, Blyth F, Woolf A, Bain C, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014; 73: 968-74.

18. Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, Ross JS. Magnetic resonance imaging of the lumbar spine in people without back pain. N Engl J Med 1994; 331: 69-73.

19. Kim HS, Jung KH, Park IH, Ryu JK, Sun KJ, Lim KJ, et al. Diagnosis and treatment of sacral asymlocation in back pain patients: clinical application of prolotherapy. Korean J Pain 2007; 20: 130-7.

20. Chou L, Brady SR, Urquhart DM, Teichtahl AJ, Cicuttini FM, Pasco JA, et al. The association between obesity and low back pain and disability is affected by mood disorders: a population-based, cross-sectional study of men. Medicine (Baltimore) 2016; 95: e3367.

21. Fillingim RB, King CD, Ribeiro-Dasilva MC, Rahim-Williams B, Riley JL 3rd. Sex, gender, and pain: a review of recent clinical and experimental findings. J Pain 2009; 10: 447-85.

22. Shiri R, Karppinen J, Leino-Arjas P, Solovieva S, Varonen H, Kalso E, et al. Cardiovascular and lifestyle risk factors in lumbar radicular pain or clinically defined sciatica: a systematic review. Eur Spine J 2007; 16: 2043-54.

23. Paanalahti K, Holm LW, Magnusson C, Carroll L, Nordin M, Skillgate E. The sex-specific interrelationship between spinal pain and psychological distress across time in the general population. Results from the Stockholm public health study. Spine J 2014; 14: 1928-35.

24. Pinheiro MB, Ferreira ML, Refshauge K, Ordoñana JR, Machado GC, Prado LR, et al. Symptoms of depression and risk of new episodes of low back pain: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2015; 67: 1591-603.

25. Suman A, Bostick GP, Schaafsma FG, Anema JR, Gross DP. Associations between measures of socio-economic status, beliefs about back pain, and exposure to a mass media campaign to improve back beliefs. BMC Public Health 2017; 17: 504.

26. Heneweer H, Picavet HS, Staes F, Kiers H, Vanhees L. Physical fitness, rather than self-reported physical activities, is more strongly associated with low back pain: evidence from a working population. Eur Spine J 2012; 21: 1265-72.

27. Fernandez M, Ordoñana JR, Hartvigsen J, Ferreira ML, Refshauge KM, Sánchez-Romera JF, et al. Is chronic low back pain associated with the prevalence of coronary heart disease when genetic susceptibility is considered? A co-twin control study of Spanish twins. PLoS One 2016; 11: e0155194.

28. Leino-Arjas P, Kaila-Kangas L, Solovieva S, Riihimäki H, Kirjonen J, Reunanen A. Serum lipids and low back pain: an association? A follow-up study of a working population sample. Spine (Phila Pa 1976) 2006; 31: 1032-7.

29. Hassoon A, Bydon M, Kerezoudis P, Maloney PR, Rinaldo L, Yeh HC. Chronic low-back pain in adult with diabetes: NHANES 2009-2010. J Diabetes Complications 2017; 31: 38-42.

30. Hancock MJ, Maher CM, Petocz P, Lin CW, Steffens D, Luque-Suarez A, et al. Risk factors for a recurrence of low back pain. Spine J 2015; 15: 2360-8.

31. Nolet PS, Kristman VL, Côté P, Carroll LJ, Cassidy JD. The association between a lifetime history of low back injury in a motor vehicle collision and future low back pain: a population-based cohort study. Eur Spine J 2018; 27: 136-44.

32. Machado GC, Maher CG, Ferreira PH, Latimer J, Koes BW, Steffens D, et al. Can recurrence after an acute episode of low back pain be predicted? Phys Ther 2017; 97: 889-95.

33. Manabe T, Takasugi S, Iwamoto Y. Positive relationship between bone mineral density and low back pain in middle-aged women. Eur Spine J 2003; 12: 596-601.

34. Visuri T, Ulaska J, Eskelin M, Pulkkinen P. Narrowing of lumbar spinal canal predicts chronic low back pain more accurately than intervertebral disc degeneration: a magnetic resonance imaging study in young Finnish male conscripts. Mil Med 2005; 170: 926-30.

35. McIntosh G, Steenstra I, Hogg-Johnson S, Carter T, Hall H. Lack of prognostic model validation in low back pain prediction studies: a systematic review. Clin J Pain 2018; 34: 748-54.

36. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015; 162: 55-63.

37. Song SO, Jung CH, Song YD, Park CY, Kwon HS, Cha BS, et al. Background and data configuration process of a nationwide population-based study using the Korean National Health Insurance System. Diabetes Metab J 2014; 38: 395-403.

38. Lee J, Lee JS, Park SH, Shin SA, Kim K. Cohort profile: the National Health Insurance Service-National Sample Cohort (NHIS-NSC), South Korea. Int J Epidemiol 2017; 46: e15.

39. Dobbin KK, Simon RM. Optimally splitting cases for training and testing high dimensional classifiers. BMC Med Genomics 2011; 4: 31.

40. MacGregor AJ, Dougherty AL, Mayo JA, Rauh MJ, Galarneau MR. Occupational correlates of low back pain among U.S. Marines following combat deployment. Mil Med 2012; 177: 845-9.

41. Jöud A, Petersson IF, Englund M. Low back pain: epidemiology of consultations. Arthritis Care Res (Hoboken) 2012; 64: 1084-8.

42. Bosman LC, Dijkstra L, Oling CI, Heymans MW, Twisk JW, Roelen CA. Prediction models to identify workers at risk of sick leave due to low-back pain in the Dutch construction industry. Scand J Work Environ Health 2018; 44: 156-62.

43. D'Agostino RB, Nam BH. Evaluation of the performance of survival analysis models: discrimination and calibration measures. Handb Stat 2003; 23: 1-25.

44. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Stat Med 2004; 23: 2109-23.

45. Ha IH, Lee J, Kim MR, Kim H, Shin JS. The association between the history of cardiovascular diseases and chronic low back pain in South Koreans: a cross-sectional study. PLoS One 2014; 9: e93671.

46. Hippisley-Cox J, Coupland C. Development and validation of QMortality risk prediction algorithm to estimate short term risk of death and assess frailty: cohort study. BMJ 2017; 358: j4208.

47. Traeger AC, Henschke N, Hübscher M, Williams CM, Kamper SJ, Maher CG, et al. Estimating the risk of chronic pain: development and validation of a prognostic model (PICKUP) for patients with acute low back pain. PLoS Med 2016; 13: e1002019.

48. Panken G, Hoekstra T, Verhagen A, van Tulder M, Twisk J, Heymans MW. Predicting chronic low-back pain based on pain trajectories in patients in an occupational setting: an exploratory analysis. Scand J Work Environ Health 2016; 42: 520-7.

49. Park BJ, Sung J, Park K, Seo SW, Kim SH. Studying on diagnosis accuracy for health insurance claims data in Korea. Seoul, Seoul National University. 2003, pp 17-29.

50. Ferreira ML, Machado G, Latimer J, Maher C, Ferreira PH, Smeets RJ. Factors defining care-seeking in low back pain--a meta-analysis of population based surveys. Eur J Pain 2010; 14: 747.e1-7.