

Development and validation of a risk-prediction nomogram for chronic low back pain using a national health examination survey

Jiwon Park

Seoul National University Bundang Hospital

Jung-Taek Kim

Ajou University School of Medicine and Graduate School of Medicine

Sang-Min Park (✉ psmini@naver.com)

Seoul National University Bundang Hospital <https://orcid.org/0000-0001-6171-3256>

Ohsang Kwon

Seoul National University Bundang Hospital

Hyoungmin Kim

Seoul National University Bundang Hospital

Ho-Joong Kim

Seoul National University Bundang Hospital

Jin S. Yeom

Seoul National University Bundang Hospital

Bong-Soon Chang

Seoul National University Bundang Hospital

Research article

Keywords: Chronic low back pain, KNHANES, Nomogram, Risk factors, Risk prediction model

Posted Date: October 30th, 2019

DOI: <https://doi.org/10.21203/rs.2.16547/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background Several prognostic factors for chronic low back pain (CLBP) have been reported. However, there is no study regarding the prediction of CLBP development in general population, using a risk prediction model. Based on this background, the aims of this study were: (1) to develop and validate a risk prediction model for CLBP (chronic low back pain) development in the general population, and (2) to create a nomogram which can help a person at risk of developing CLBP to receive appropriate counseling on risk modification.

Methods Data on CLBP development, demographics, socioeconomic history, and comorbid health condition of participants were obtained through a nationally representative health examination and survey from 2007 to 2009. Prediction models for CLBP development were derived for health survey on a random sample of 80% of the data and were validated in the remaining 20%. After developing the risk prediction model for CLBP development, this model was incorporated into a nomogram.

Results Data for 17,038 participants were finally analyzed, including 2,693 with CLBP and 14,345 without. The finally selected risk factors included age, gender, occupation, education level, mid-intensity physical activity, depressive symptom, and comorbidities. This model had good predictive performance in the validation dataset (concordance statistic = 0.7569, Hosmer-Lemeshow chi-square statistic = 12.10, $p = .278$). The findings indicated no significant differences between the observed probability and predicted probability according to our model.

Conclusions The risk prediction model, presented by a nomogram, which is a score-based prediction system, could be incorporated into the clinical setting. Thus, our prediction model with a nomogram can help a person at risk of developing CLBP to receive appropriate counseling on risk modification from primary physicians.

Background

Low back pain (LBP), one of the most common pain sources of musculoskeletal disorder and a major public health problem that influences the functional status and quality of life in elderly people, is reported in 7%–80% of the general population at least once in their lifetime [1–3]. Patients with LBP may recover spontaneously, but some LBP patients will develop chronic LBP (CLBP) [4]. Von Korff M et al. emphasized the prognostic factors for management and prevention of CLBP [5].

The risk factors of LBP are multifactorial and complex, and remain poorly understood [3, 6–13]. Many risk factors have been proposed, including high body mass index, family history, stress, anxiety, depression, low physical activity, long sitting time, low education level, occupation, and inadequacy of income [3, 7–11, 14–18]. Moreover, many studies aimed to determine the patient population that is likely to develop CLBP. The risk factors of CLBP are similar to LBP, such as age, gender, previous LBP episode, pain severity, and occupation [19–23]. However, these studies only evaluated the prognostic factors of patients who will develop CLBP from nonspecific LBP patients in hospital or workplace setting. In the

community-based setting, it is important to identify the risk factors of CLBP and also predict the probability of patients likely to develop CLBP. However, there is no study on the prediction probability of CLBP development in the general population using a risk prediction model.

Based on this background, the aims of this study were: (1) to develop and validate a risk prediction model for CLBP development, and (2) to create a shared decision-making nomogram that can help a patient at risk of developing CLBP to receive appropriate counseling on risk modification using a nationally representative sample of Korean adults.

Methods

Study participants and design

Data from version IV–1, 2, and 3 of the Korea National Health and Nutrition Examination Survey (KNHANES) performed in 2007, 2008, and 2009 were analyzed. This survey has been conducted annually since 1998 by the Korea Centers for Disease Control (KCDC). To evaluate the health and nutritional status of the general Korean population, a nationwide sampling method (clustered, multistage, stratified, and randomized) is used for proportional distribution according to geographic area, sex, and age. The survey participants are different every year, and are not monitored serially, resulting in a random sampling every year. The KNHANES evaluates three aspects: health questionnaires, health and physical examinations, and nutrition questionnaires that are administered by experienced interviewers, registered nurses, and laboratory technicians [24]. KNHANES IV–1 (2007), IV–2 (2008), and IV–3 (2009) examinations and health surveys were completed by 4594, 9744, and 10533 participants (total: 24,871 participants), respectively. The present analysis was confined to 17038 respondents aged 10–100 years who answered the chronic LBP examination survey and had no missing data regarding the demographics and health questionnaires (Fig. 1).

Definitions of chronic low back pain

Participants in this survey who answered “yes” to all three questionnaires were defined as having CLBP: (1) “Have you ever had LBP for the rest of your life?” (2) “Do you currently have LBP?” and (3) “Have you complained of LBP for more than 90 days during the past year?”

Description of demographics and health surveys

We analyzed the participants’ demographics and socioeconomic status, comorbidities, and lifestyle habits through health interviews and examinations. All participants were asked whether they had received diagnoses of major comorbidities by physicians, such as hypertension, diabetes mellitus, dyslipidemia, ischemic heart disease (myocardial infarction, angina), stroke, liver cirrhosis, asthma, chronic obstructive

pulmonary disease, arthritis, and chronic kidney disease major cancers (lung, stomach, liver, colon, breast, prostate or uterine cervical), were surveyed in all participants.

The age was categorized into age groups. Body mass index (BMI) was calculated as the body weight (kilogram) divided by the height (meter) squared and categorized into underweight ($<18.5 \text{ kg/m}^2$), normal weight ($18.5\text{--}24.9 \text{ kg/m}^2$), and obese ($\geq 25.0 \text{ kg/m}^2$). Patients were categorized as non/ex-smoker or current smoker based on the present smoking status. Alcohol consumption was categorized into none and ≥ 1 drink/month. Occupations were divided into five groups: unemployed (e.g., student, housewife), office workers (e.g., managers, professionals, and office workers), sales and services, machine fitting and simple labor (e.g., technicians, device and machine operators, and low-level laborers), and agriculture, forestry and fishery [25]. The household income level was categorized into 4 groups according to quartiles. The educational level was also categorized into four groups according to the degree of graduation: ≤ 6 years (elementary school), 7–9 years (middle school), 10–12 years (high school), and ≥ 13 years (university or college). Physical activity was defined as three categories. First, walk was defined as walking activity for 5 or more days per week at least 30 minutes. Moderate physical activity was defined as mid-intensity physical activity for 5 or more days per week for at least 30 minutes. High physical activity was defined as high-intensity physical activity for 3 or more days per week for at least 20 minutes. Depressive symptom was defined as individuals in this survey who felt sad or had a depressive symptom for 2 consecutive weeks during the past year.

Statistical analysis

Statistics were analyzed by Stata/MP 15.0 (StataCorp., 2017, Stata Statistical Software: Release 15; College Station, TX, USA; StataCorp LP). Continuous variables were presented as mean \pm standard deviation. Statistical significance was considered with two-tailed p-value <0.05 . Sampling weights were applied to the study population in order to represent the Korean population without bias.

General demographics and co-variables were evaluated between the participants with and without CLBP. Student's t-test was used to compare continuous variables, and chi-square test was used for categorical variables. Baseline demographics and co-variables listed in Table 1 were assessed as independent variables for the models.

To create a development and validation dataset from the entire datasets, we used the split-sample method [26]. A split sample with a 50% hold out results in models with suboptimal performance, that is, models with unstable and, on average, the same performance as that obtained with half the sample size [27]. Therefore, the development dataset for the prediction equation development was obtained by randomly selecting 80% of the entire datasets [28]. A logistic regression model was used for developing prediction equation. Only covariables with a p-value of <0.05 from univariate analysis were subsequently evaluated in the multiple logistic regression using backward stepwise selection with a 0.05 significant

level. Table 2 shows the list of variables, odds ratios, and regression coefficients that remained in the final prediction models after multiple logistic regression.

The developed model was validated by evaluating their performance with respect to discrimination and calibration using the C-statistics and Hosmer-Lemeshow chi-square statistics. The validation datasets for the prediction equation consisted of the remaining 20% of the full data after the development dataset was randomly selected. The area under the receiver operating characteristics curve (AUC), also called the C-statistic, for prediction model was measured for discrimination. Although it is controversial to determine the good value of the AUC, a value below 0.70 is often considered suboptimal, a value from 0.70 to 0.80 is considered good, and a value of 0.80 or above is excellent. To assess developed model calibration, Hosmer-Lemeshow chi-square statistics (HLS) was used to calculate how close the predicted risks were to the actual observed risks. To calculate the HLS, dataset was divided into 10 subgroups based on the predicted probabilities from the developed prediction model. Values exceeding 20 indicated a significant lack of calibration [29].

The risk of CLBP was predicted using a nomogram, which is a two-dimensional diagram designed to allow the approximate graphical computation of a mathematical function. The nomogram was generated using the independent risk factors analyzed in the multiple logistic regression analysis. To make the nomogram, we used the “nomolog” module of STATA to create the nomogram [30].

Results

Demographics of participants according to chronic low back pain

The baseline general characteristics of all participants are shown in Table 1. A total of 17,038 participants were finally analyzed, including 2,693 with CLBP and 14,345 without. The prevalence of CLBP was 15.8% in Korean subjects, with a prevalence of 11.8% in men and 24.5% in women. Of these, 80% of the participants (n=13,630) were randomly selected to the development dataset, with the remaining 20% of the participants (n=3,408) selected to the validation dataset. In the development dataset, participants with CLBP was 2,120, and without CLBP was 11,510. The mean age was 49.1 ± 16.6 years, and 5,776 participants (42.4%) were men. In the validation dataset, participants with CLBP was 573, and without CLBP was 2,835. The mean age was 49.3 ± 16.6 years, and 1,426 participants (41.8%) were men.

Risk factors for prediction model

Table 2 shows the estimated odds ratios (ORs) from multiple logistic regression in the development dataset. The finally selected risk factors included age, gender, occupation, education level, mid-intensity physical activity, depressive symptom, and comorbidities (stroke, ischemic heart disease, knee osteoarthritis, asthma, chronic obstructive pulmonary disease, cancer history). Based on this risk

analysis, the CLBP prediction equation were calculated, and the coefficients of the risk factors were developed. From this equation, the linear function of developing prediction probability was estimated.

Discrimination and calibration of prediction model

Our prediction model had good discrimination (AUC = 0.7518) and were well calibrated (HLS = 4.72, $p = .787$) in the development dataset. Moreover, this model also had good validation in the validation dataset (AUC = 0.7569, HLS = 12.10, $p = .278$). This indicated no significant differences between the observed probability and predicted probability according to our model. Figures 2 and 3 showed the discrimination and calibration plots for the CLBP prediction model.

Nomogram for prediction model

The risk factors that were found to predict CLBP in the development dataset were incorporated into the nomogram, as shown in Figure 4. The value of each risk factor is respectively loaded on each variable axis (the 1st-12th lines), and a line is drawn downwards to determine the number of points received for each variable (the 13th line). Then, the sum of these numbers is located on the total points axis (the last line), and a line is drawn upwards to the risk probability axis to determine the likelihood of CLBP.

Discussion

Our study developed and validated a clinical risk prediction model for CLBP development in the general population using a cross-sectional Korean population-based health survey. It is necessary to create a risk prediction model to identify common risk factors of CLBP and to modify these risk factors, especially that there are increasing needs not in research setting, but in clinical setting. To our knowledge, this is the first clinical risk prediction model for CLBP in the general population. We found that the prevalence of CLBP in the general population is 15.8% (11.8% in men and 24.5% in women). In the variables of demographics, medical histories, and socioeconomic status, we found significant risk factors, which influence the development of CLBP. Using these risk factors, we developed a clinical risk prediction model and nomogram to allow personalized CLBP prediction based on personal characteristics. Our final model had a good discrimination and calibration performance in the validation datasets, which demonstrated accurate prediction of CLBP in a new population with similar characteristics.

In our model, age was a key predictor of CLBP development. At 80 years of age, the odds ratio was increased to 7.268 which is compared to the 10 years of age (reference age). In a previous study, LBP prevalence is the highest in individuals aged 45 to 64 years [31]. However, in our study, the risk of CLBP increases with aging up to the 9th decade of life. The increased risks in female gender is similar to a previous study [19]. However, our study revealed no significant difference in smokers and alcoholics compared with non-smokers and non-alcoholics [32]. Other prognostic factors, such as occupation,

education level, physical activity, depressive symptoms, and comorbidities, were associated with CLBP, which is similar to previous studies [3, 6–11, 19, 22].

Previous studies investigated the prognostic factors for developing CLBP from LBP in the workplace and general population. Heymans et al. reported a higher pain intensity of initial LBP, no clinically relevant change in pain intensity, and disability status in the first 3 months; higher score for kinesiophobia was most strongly related to CLBP in the workplace [23]. Similar studies in the general population also revealed that age, gender, height, health status, heavy work, chronic stress, low physical activity, smoking, and history of LBP were important predictors of CLBP [19, 31]. There are some differences between these studies and our study. First, these studies have the advantage of prospectively detecting developing CLBP, but they only analyzed individuals in the workplace or the general population with a small number of participants. Secondly, these studies only analyzed the prognostic factors that caused CLBP in patients with LBP. Although our study was not as prospective as in the previous studies, we analyzed the risk factors of patients with CLBP in nationwide representative populations with a relatively large population and analyzed the probability of developing CLBP. To calculate this, multiple logistic regression was used, and a nomogram was made using coefficients. Although it is more convenient for many people to use a web-based calculator than a nomogram in actual usage, our study could not provide web-based calculation services due to the absence of a funding source, web-developer, and servers.

As our risk prediction model is comprehensive and sophisticated, it is important that our study developed this model to modify the risks of CLBP. Previous studies have simply informed patients with several risk factors that they belong to a high-risk group, which does not give practical help in the clinical setting. Our model shows the scores according to the risk factors and predicted probability of developing LBP using a nomogram, which clearly presents and explains its risk factors and probabilities to patients and physicians. These are relevant for clinicians to recommend preventive measures or treatment strategies for their patients. Thus, it may be used not only in primary care but also in healthcare centers in the general population.

To the best of our knowledge, this is the first study investigating the predicted probability of CLBP using a nationwide Korean representative sample in the general population. The greatest strength of our study was the increased external validity of our findings by using the KNHANES data. However, there are some limitations in our study. First, this study was conducted through a national health and nutrition examination survey, which was designed as a cross-sectional. Therefore, the actual development of CLBP could not be analyzed in this study. To assess the development of CLBP, a large prospective cohort study in the general population is needed. Unlike other prediction models (i.e., prediction for surgical site infection), constructing a prediction model for CLBP in a cohort is actually difficult. Furthermore, CLBP is closely related to age, socioeconomic status, and comorbidities. These factors are well collected in our dataset, KNHANES, which is not a cohort dataset but a cross-sectional dataset. These datasets cannot assess actual CLBP development, but they may be useful for constructing a prediction model using risk factors. Second, KNHANES was designed to minimize the sampling errors by utilizing a clustered, multi-stage, and random sampling method. However, selection bias may exist because of missing data.

Participants from our raw data were selected to minimize selection bias, but missing data inevitably led to bias. Unlike other studies, such as cohort studies and clinical trials, imputation in missing values is impossible in our dataset. Therefore, we excluded participants with missing data, which is necessary for analysis. Third, the simple survey of CLBP used in this study did not evaluate the severity, source, or duration of CLBP, which would be possible with instruments of measuring pain on a scale (e.g., visual analog scale pain score). Fourth, this study also could not analyze some prognostic factors of CLBP, such as previous episodes of LBP, severity of pain, and disability. However, this study analyzed many other risk factors, which was not analyzed in other studies. Fifth, the prediction model of CLBP may be dependent on ethnicity. KNHANES is a health survey of the general Korean population. Therefore, this model is representative of the general population, but more representative of the general Korean population of Korea, and caution is warranted in extrapolating these findings to other ethnic populations.

Conclusion

We show that the risk of CLBP development can be reliably estimated in the general population. Our study developed a clinical risk prediction model to determine the probability of developing CLBP using a cross-sectional Korean population-based survey. Our prediction model showed good accuracy in the development and validation datasets. Moreover, our risk prediction model, presented by a nomogram, which is a score-based prediction system, could be incorporated in a clinical setting. Thus, our prediction model with a nomogram can help the person at risk of developing CLBP to receive appropriate counseling on risk modification from primary physicians.

List Of Abbreviations

AUC: Area under the receiver operating characteristics curve; CLBP: Chronic low back pain; KCDC: Korea Centers for Disease Control; KNHANES: Korea National Health and Nutrition Examination Survey; LBP: Low back pain; OR: Odds ratio

Declarations

Ethics approval and consent to participate

The IV-1, IV-2 and IV-3 versions of KNHANES were approved by the KCDC Institutional Review Board (approval no. 2007-02CON-04-P, 2008-04EXP-01-C, 2009-01CON-03-2C). Informed consent was obtained from all participants and a parent or guardian for participants under 16 years old when the surveys were conducted.

Consent for publication

Not applicable.

Availability of data and materials

The data used in this study were sourced from the KCDC (Korea Centers for Disease Control & Prevention) in Korea. Owing to the legal restrictions imposed by the Government of Korea related to the Personal Information Protection Act, the dataset cannot be made publicly available. Interested researchers can obtain the dataset through a formal application to Division of Health and Nutrition Survey, KCDC, Korea (<https://knhanes.cdc.go.kr/knhanes/eng/index.do>). Researchers are not allowed to carry raw data outside the KCDC. Interested researchers can access these data in the same manner as the authors. The authors did not have any special access privileges that other researchers would not have. The authors used 2007–2009 KNHANES data. All files are available from the KNHANES webpage (https://knhanes.cdc.go.kr/knhanes/sub03/sub03_02_02.do) and the name of datasets (HN07_ALL.SAV, HN08_ALL.SAV, HN09_ALL.SAV).

Competing interests

The authors declare that they have no competing interests.

Funding

There was no external funding source used for this study.

Authors' contributions

JP and J-TK are equally contributed to the conception and design, acquisition of data, analysis and interpretation of data, and drafting/revisions of article. S-MP contributed to the conception and design, analysis and interpretation of data, drafting/revisions of article, as well as final approval of the article. HK and H-J Kim contributed to the acquisition of data and interpretation of data. OK conducted the statistical analyses. JSY and B-SC contributed to the conception and design, drafting/revisions of article, as well as final approval of the article. All authors contributed to and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Balague F, Mannion AF, Pellise F, Cedraschi C. Non-specific low back pain. *Lancet*. 2012;379(9814):482–91.

- 2.Hong JY, Song KS, Cho JH, Lee JH. An Updated Overview of Low Back Pain Management in Primary Care. *Asian Spine J.*2017;11(4):653–60.
- 3.Ganesan S, Acharya AS, Chauhan R, Acharya S. Prevalence and Risk Factors for Low Back Pain in 1,355 Young Adults: A Cross-Sectional Study. *Asian Spine J.*2017;11(4):610–7.
- 4.Frank JW, Brooker AS, DeMaio SE, Kerr MS, Maetzel A, Shannon HS, Sullivan TJ, Norman RW, Wells RP. Disability resulting from occupational low back pain. Part II: What do we know about secondary prevention? A review of the scientific evidence on prevention after disability begins. *Spine (Phila Pa 1976).*1996;21(24):2918–29.
- 5.Von Korff M, Miglioretti DL. A prognostic approach to defining chronic pain. *Pain.*2005;117(3):304–13.
- 6.Park SM, Kim HJ, Jang S, Kim H, Chang BS, Lee CK, Yeom JS. Depression is Closely Associated With Chronic Low Back Pain in Patients Over 50 Years of age: A Cross-sectional Study Using the Sixth Korea National Health and Nutrition Examination Survey (KNHANES VI–2). *Spine (Phila Pa 1976).*2018.
- 7.Park SM, Kim HJ, Jeong H, Kim H, Chang BS, Lee CK, Yeom JS. Longer sitting time and low physical activity are closely associated with chronic low back pain in population over 50 years of age: a cross-sectional study using the sixth korea national health and nutrition examination survey. *Spine J.*2018.
- 8.Kim W, Jin YS, Lee CS, Hwang CJ, Lee SY, Chung SG, Choi KH. Relationship between the type and amount of physical activity and low back pain in Koreans aged 50 years and older. *PM R.*2014;6(10):893–9.
- 9.Macfarlane GJ, Thomas E, Papageorgiou AC, Croft PR, Jayson MI, Silman AJ. Employment and physical work activities as predictors of future low back pain. *Spine (Phila Pa 1976).*1997;22(10):1143–9.
- 10.Papageorgiou AC, Macfarlane GJ, Thomas E, Croft PR, Jayson MI, Silman AJ. Psychosocial factors in the workplace—do they predict new episodes of low back pain? Evidence from the South Manchester Back Pain Study. *Spine (Phila Pa 1976).*1997;22(10):1137–42.
- 11.Webb R, Brammah T, Lunt M, Urwin M, Allison T, Symmons D. Prevalence and predictors of intense, chronic, and disabling neck and back pain in the UK general population. *Spine (Phila Pa 1976).*2003;28(11):1195–202.
- 12.Park SM, Kim GU, Kim HJ, Kim H, Chang BS, Lee CK, Yeom JS. Walking more than 90minutes/week was associated with a lower risk of self-reported low back pain in persons over 50years of age: a cross-sectional study using the Korean National Health and Nutrition Examination Surveys. *Spine J.*2019;19(5):846–52.
- 13.Park SM, Kim GU, Kim HJ, Kim H, Chang BS, Lee CK, Yeom JS. Low handgrip strength is closely associated with chronic low back pain among women aged 50 years or older: A cross-sectional study using a national health survey. *PLoS One.*2018;13(11):e0207759.

14. Shiri R, Coggon D, Falah-Hassani K. Exercise for the Prevention of Low Back Pain: Systematic Review and Meta-Analysis of Controlled Trials. *Am J Epidemiol.*2018;187(5):1093–101.
15. 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(1):38–42.
16. Hoogendoorn WE, van Poppel MN, Bongers PM, Koes BW, Bouter LM. Systematic review of psychosocial factors at work and private life as risk factors for back pain. *Spine (Phila Pa 1976).*2000;25(16):2114–25.
17. Parreira P, Maher CG, Steffens D, Hancock MJ, Ferreira ML. Risk factors for low back pain and sciatica: an umbrella review. *Spine J.*2018.
18. Parreira PCS, Maher CG, Ferreira ML, Machado GC, Blyth FM, Naganathan V, Waite LM, Seibel MJ, Handelsman D, Cumming RG. A longitudinal study of the influence of comorbidities and lifestyle factors on low back pain in older men. *Pain.*2017;158(8):1571–6.
19. Thomas E, Silman AJ, Croft PR, Papageorgiou AC, Jayson MI, Macfarlane GJ. Predicting who develops chronic low back pain in primary care: a prospective study. *Bmj.*1999;318(7199):1662–7.
20. van den Hoogen HJ, Koes BW, Deville W, van Eijk JT, Bouter LM. The prognosis of low back pain in general practice. *Spine (Phila Pa 1976).*1997;22(13):1515–21.
21. Kendall NA. Psychosocial approaches to the prevention of chronic pain: the low back paradigm. *Baillieres Best Pract Res Clin Rheumatol.*1999;13(3):545–54.
22. van der Windt DA, Kuijpers T, Jellema P, van der Heijden GJ, Bouter LM. Do psychological factors predict outcome in both low-back pain and shoulder pain? *Ann Rheum Dis.*2007;66(3):313–9.
23. Heymans MW, van Buuren S, Knol DL, Anema JR, van Mechelen W, de Vet HC. The prognosis of chronic low back pain is determined by changes in pain and disability in the initial period. *Spine J.*2010;10(10):847–56.
24. Kweon S, Kim Y, Jang MJ, Kim Y, Kim K, Choi S, Chun C, Khang YH, Oh K. Data resource profile: the Korea National Health and Nutrition Examination Survey (KNHANES). *Int J Epidemiol.*2014;43(1):69–77.
25. Punnett L, Pruss-Utun A, Nelson DI, Fingerhut MA, Leigh J, Tak S, Phillips S. Estimating the global burden of low back pain attributable to combined occupational exposures. *Am J Ind Med.*2005;48(6):459–69.
26. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol.*2016;69:245–7.

27. Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Stat Methods Med Res.* 2017;26(2):796–808.
28. Khor S, Lavalley D, Cizik AM, Bellabarba C, Chapman JR, Howe CR, Lu D, Mohit AA, Oskouian RJ, Roh JR, Shonnard N, Dagal A, Flum DR. Development and Validation of a Prediction Model for Pain and Functional Outcomes After Lumbar Spine Surgery. *JAMA Surg.* 2018;153(7):634–42.
29. D'Agostino RB, Sr., Grundy S, Sullivan LM, Wilson P. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA.* 2001;286(2):180–7.
30. Zlotnik A, Abaira V. Nomograms in Stata. 2015. <http://www.zlotnik.net/stata/nomograms/>. Accessed 12 Sept 2019
31. Kopec JA, Sayre EC, Esdaile JM. Predictors of back pain in a general population cohort. *Spine (Phila Pa 1976).* 2004;29(1):70–7; discussion 7–8.
32. Shiri R, Karppinen J, Leino-Arjas P, Solovieva S, Viikari-Juntura E. The association between smoking and low back pain: a meta-analysis. *Am J Med.* 2010;123(1):87 e7–35.

Tables

Table 1. Characteristics of the study population according to chronic low back pain

| Variables | Without CLBP (n=14,345) | CLBP (n=2,693) | p-Value |
|---|----------------------------|-------------------|---------|
| Age, year | 47.3 (16.1) | 59.1 (16.0) | <.001 |
| Age, n (%) | | | |
| 10-19 | 174 (1.2%) | 4 (0.1%) | <.001 |
| 20-29 | 1932 (13.5%) | 142 (5.3%) | |
| 30-39 | 3143 (21.9%) | 269 (10.0%) | |
| 40-49 | 3052 (21.3%) | 310 (11.5%) | |
| 50-59 | 2423 (16.9%) | 439 (16.3%) | |
| 60-69 | 2038 (14.2%) | 698 (25.9%) | |
| 70-79 | 1296 (9.0%) | 665 (24.7%) | |
| 80-89 | 275 (1.9%) | 161 (6.0%) | |
| ≥90 | 12 (0.1%) | 5 (0.2%) | |
| Gender, n (%) | | | |
| Male | 6444 (44.9%) | 758 (28.1%) | <.001 |
| Female | 7901 (55.1%) | 1935 (71.9%) | |
| Height, cm | 162.4 (9.1) | 157.3 (9.3) | <.001 |
| Weight, kg | 62.5 (11.5) | 59.0 (10.3) | <.001 |
| BMI, kg/m ² | 23.6 (3.4) | 23.8 (3.3) | .002 |
| Obesity, n (%) [*] | | | |
| Underweight (<18.5) | 683 (4.8%) | 106 (4.0%) | .001 |
| Normal (18.5 - 24.9) | 9080 (64.1%) | 1644 (61.6%) | |
| Obese (>25) | 4403 (31.1%) | 918 (34.4%) | |
| Smoking status, n (%) | | | |
| Non / Ex-smoker | 11018 (76.9%) | 2304 (85.7%) | <.001 |
| Current smoker | 3318 (23.1%) | 274 (14.3%) | |
| Alcohol consumption, n (%) | | | |
| None | 6453 (45.0%) | 1688 (62.7%) | <.001 |
| ≥1 drink/month | 7892 (55.0%) | 1005 (37.3%) | |
| Occupation, n (%) | | | |
| Unemployed (Student, housewife, etc.) | 5791 (40.4%) | 1375 (51.1%) | <.001 |
| Office work | 2891 (20.2%) | 187 (6.9%) | |
| Sales and services | 1893 (13.2%) | 221 (8.2%) | |
| Agriculture, forestry and fishery | 1102 (7.7%) | 524 (19.5%) | |
| Machine fitting and simple labor | 2668 (18.6%) | 386 (14.3%) | |
| Household income, n (%) [†] | | | |
| Low | 2572 (18.4%) | 976 (36.9%) | <.001 |
| Low-moderate | 3510 (25.1%) | 645 (24.4%) | |
| Moderate-high | 3881 (27.8%) | 554 (20.9%) | |
| High | 4008 (28.7%) | 471 (17.8%) | |
| Education level, n (%) [‡] | | | |
| ≤6 years | 3462 (24.1%) | 1503 (55.8%) | <.001 |
| 7-9 years | 1594 (11.1%) | 316 (11.7%) | |
| 10-12 years | 5289 (36.9%) | 563 (20.9%) | |
| ≥13 years | 4000 (27.9%) | 311 (11.5%) | |
| Physical activity, n (%) [§] | | | |
| Walk | 6570 (46.0%) | 1263 (47.1%) | .32 |
| Middle PA | 1903 (13.3%) | 514 (19.1%) | <.001 |
| High PA | 2276 (15.9%) | 408 (15.2%) | .34 |
| Depressive symptom, n (%) | 1987 (13.9%) | 698 (25.9%) | <.001 |
| Comorbidities, n (%) | | | |
| Hypertension | 2591 (18.1%) | 879 (32.6%) | <.001 |
| Dyslipidemia | 968 (6.7%) | 305 (11.3%) | <.001 |
| Stroke | 261 (1.8%) | 134 (5.0%) | <.001 |
| Ischemic heart disease | 258 (1.8%) | 133 (4.9%) | <.001 |

| | | | |
|------------------------|--------------|--------------|-------|
| Knee osteoarthritis | 2119 (14.8%) | 1052 (39.1%) | <.001 |
| Asthma | 512 (3.6%) | 215 (8.0%) | <.001 |
| COPD | 114 (0.8%) | 57 (2.1%) | <.001 |
| Diabetes | 1004 (7.0%) | 310 (11.5%) | <.001 |
| Chronic kidney disease | 48 (0.3%) | 17 (0.6%) | .022 |
| Liver cirrhosis | 23 (0.2%) | 10 (0.4%) | .022 |
| Cancer [¶] | 356 (2.5%) | 145 (5.4%) | <.001 |

Numeric parameters are expressed as mean and standard deviation in parentheses

Categorical parameters are expressed as counts and percentages in parentheses

CLBP; chronic low back pain, BMI; body mass index, PA; physical activity, COPD; chronic obstructive pulmonary disease

* Body mass index was categorized into underweight (<18.5 kg/m²), normal (18.5–24.9 kg/m²), and obese (≥25.0 kg/m²)

† Household income level was calculated by dividing the total household monthly income with the obtained levels then grouped into quartiles

‡ Educational level was divided into the following four groups: ≤6 years (elementary school), 7-9 years (middle school), 10-12 years (high school), and ≥13 years (college or university).

§ Physical activity was defined as three categories. First, walk was defined as walking activity for 5 or more days per week at least 30 minutes. Middle physical activity was defined as mid-intensity physical activity for 5 or more days per week at least 30 minutes. High physical activity was defined as high-intensity physical activity for 3 or more days per week at least 20 minutes.

|| Depressive symptom was defined as individuals in this survey who felt sad or depressive symptom for two consecutive weeks during the past one year

¶ History of major cancer: stomach, liver, colon, breast, uterine cervical, prostate or lung cancer

Table 2. Chronic low back pain risk prediction model using multiple logistic regression in development dataset

| Variables | Coefficient | Odds ratio | 95% CI | | p-Value |
|---------------------------------------|-------------|------------|--------|--------|---------|
| Age group | | | | | |
| 10-19 | reference | | | | |
| 20-29 | 0.8090 | 2.246 | 0.810 | 6.224 | .120 |
| 30-39 | 0.9342 | 2.545 | 0.926 | 6.996 | .070 |
| 40-49 | 0.9534 | 2.595 | 0.945 | 7.122 | .064 |
| 50-59 | 1.1288 | 3.092 | 1.125 | 8.501 | .029 |
| 60-69 | 1.5562 | 4.741 | 1.722 | 13.049 | .003 |
| 70-79 | 1.8467 | 6.339 | 2.297 | 17.494 | <.001 |
| 80-89 | 1.9834 | 7.268 | 2.580 | 20.473 | <.001 |
| ≥90 | 1.2191 | 3.384 | 0.625 | 18.334 | .157 |
| Gender | | | | | |
| Male | reference | | | | |
| Female | 0.6023 | 1.826 | 1.623 | 2.055 | <.001 |
| Occupation | | | | | |
| Unemployed (Student, housewife, etc.) | reference | | | | |
| Office work | -0.2717 | 0.762 | 0.620 | 0.937 | .010 |
| Sales and services | -0.1310 | 0.877 | 0.730 | 1.054 | .161 |
| Agriculture, forestry and fishery | 0.5694 | 1.767 | 1.516 | 2.060 | <.001 |
| Machine fitting and simple labor | 0.0195 | 1.002 | 0.859 | 1.169 | .980 |
| Education level* | | | | | |
| ≤6 years | reference | | | | |
| 7-9 years | -0.1874 | 0.829 | 0.700 | 0.981 | .029 |
| 10-12 years | -0.3344 | 0.716 | 0.607 | 0.844 | <.001 |
| ≥13 years | -0.4863 | 0.615 | 0.499 | 0.757 | <.001 |
| Middle PA [†] | 0.4069 | 1.502 | 1.315 | 1.716 | <.001 |
| Depressive symptom [‡] | 0.4907 | 1.633 | 1.448 | 1.843 | <.001 |
| Comorbidities | | | | | |
| Stroke | 0.4657 | 1.593 | 1.223 | 2.075 | .001 |
| Ischemic heart disease | 0.4271 | 1.533 | 1.184 | 1.984 | .001 |
| Knee osteoarthritis | 0.5326 | 1.703 | 1.514 | 1.916 | <.001 |
| Asthma | 0.2921 | 1.339 | 1.090 | 1.646 | .005 |
| COPD | 0.6031 | 1.828 | 1.228 | 2.720 | .003 |
| Cancer [§] | 0.3648 | 1.440 | 1.133 | 1.831 | .003 |

OR, Odds ratio; 95% CI, 95% confidence interval, PA; physical activity, COPD; chronic obstructive pulmonary disease

* Educational level was divided into the following four groups: ≤6 years (elementary school), 7-9 years (middle school), 10-12 years (high school), and ≥13 years (college or university).

† Levels of sitting time were divided into 4 categories using quartiles: <5, 5 - 7, 8 - 10, >10 hours/days

‡ Middle physical activity was defined as mid-intensity physical activity for 5 or more days per week at least 30 minutes.

§ Depressive symptom was defined as individuals in this survey who felt sad or depressive symptom for two consecutive weeks during the past one year

§ History of major cancer: stomach, liver, colon, breast, uterine cervical, prostate or lung cancer

Risk prediction model was fully adjusted by all co-variables listed in the table.

Figures

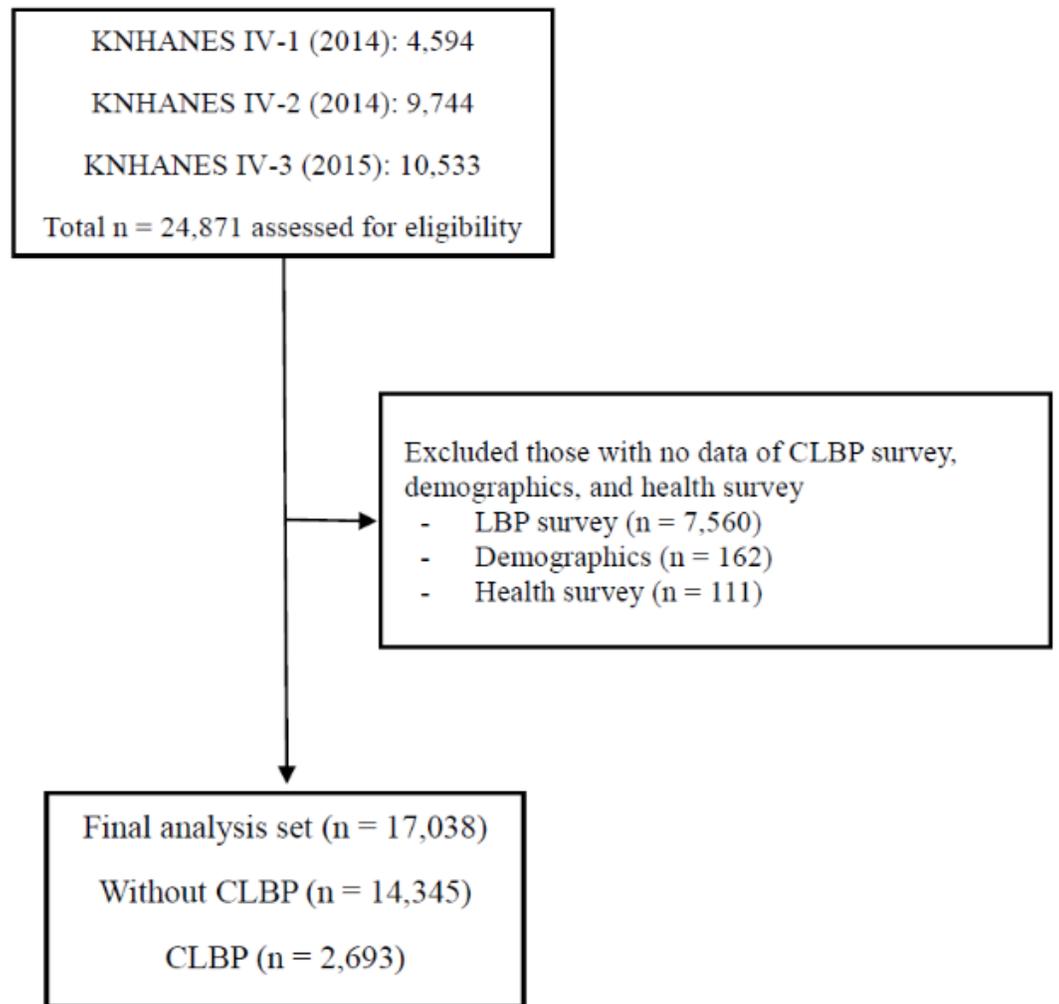


Figure 1

Flow diagram of the inclusion and exclusion of participants from the 2007, 2008 and 2009 Korea National Health and Nutrition Examination Surveys (KNHANES IV-1, IV-2 and IV-3). CLBP, chronic low back pain.

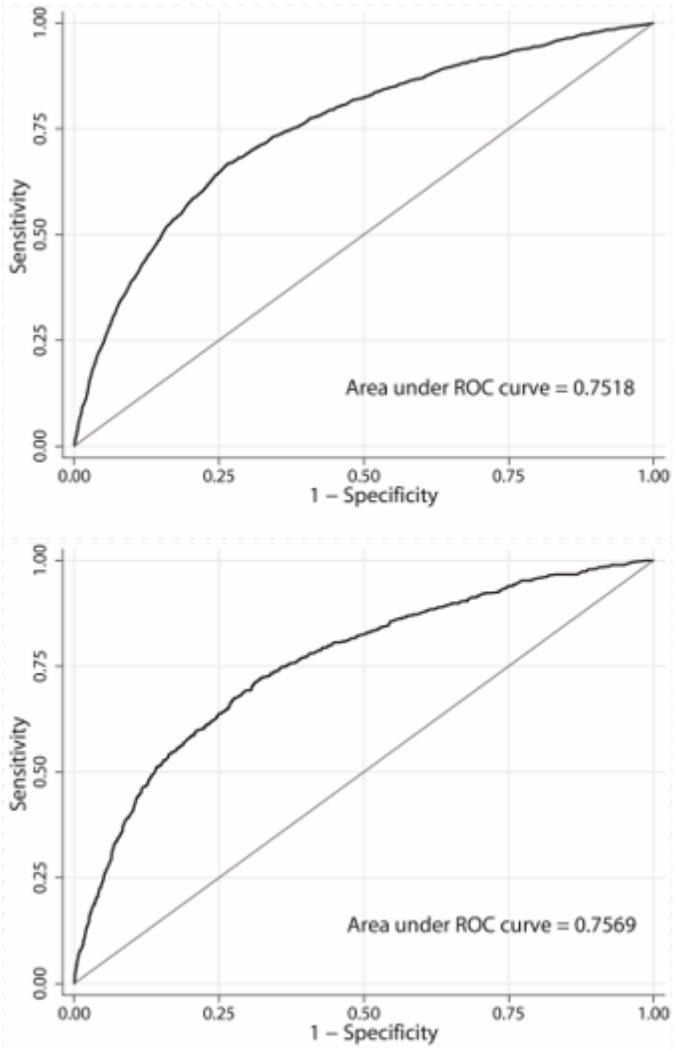


Figure 2

Discrimination plots in the (A) development and (B) validation datasets.

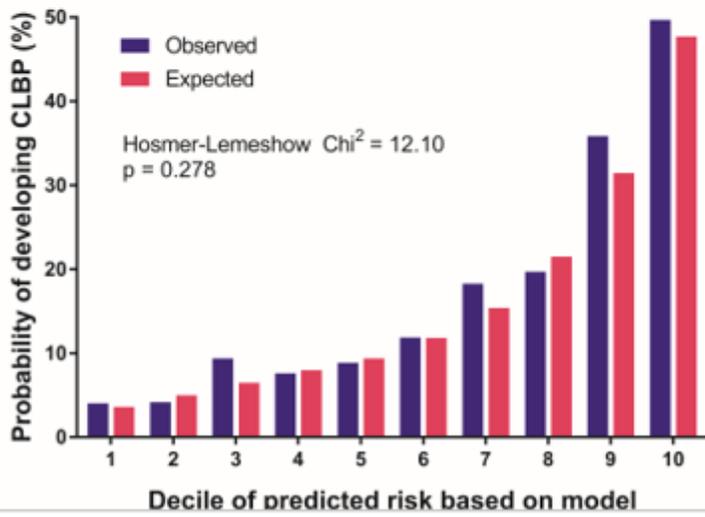
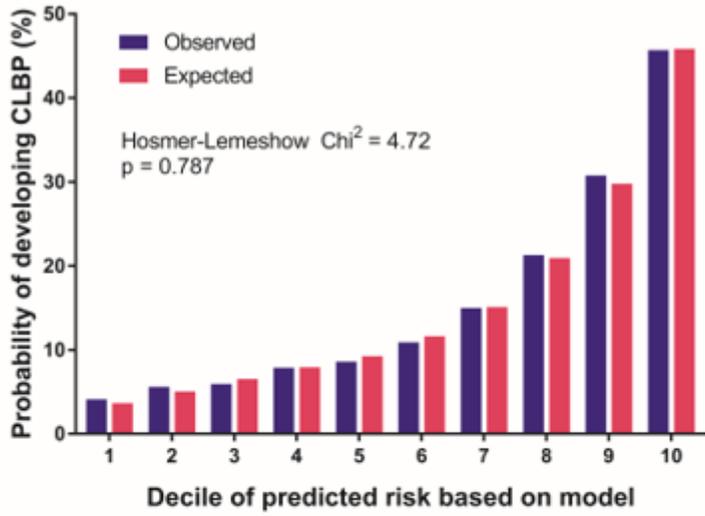


Figure 3

Calibration plots in the (A) development and (B) validation datasets.

Nomogram for predicting chronic low back pain

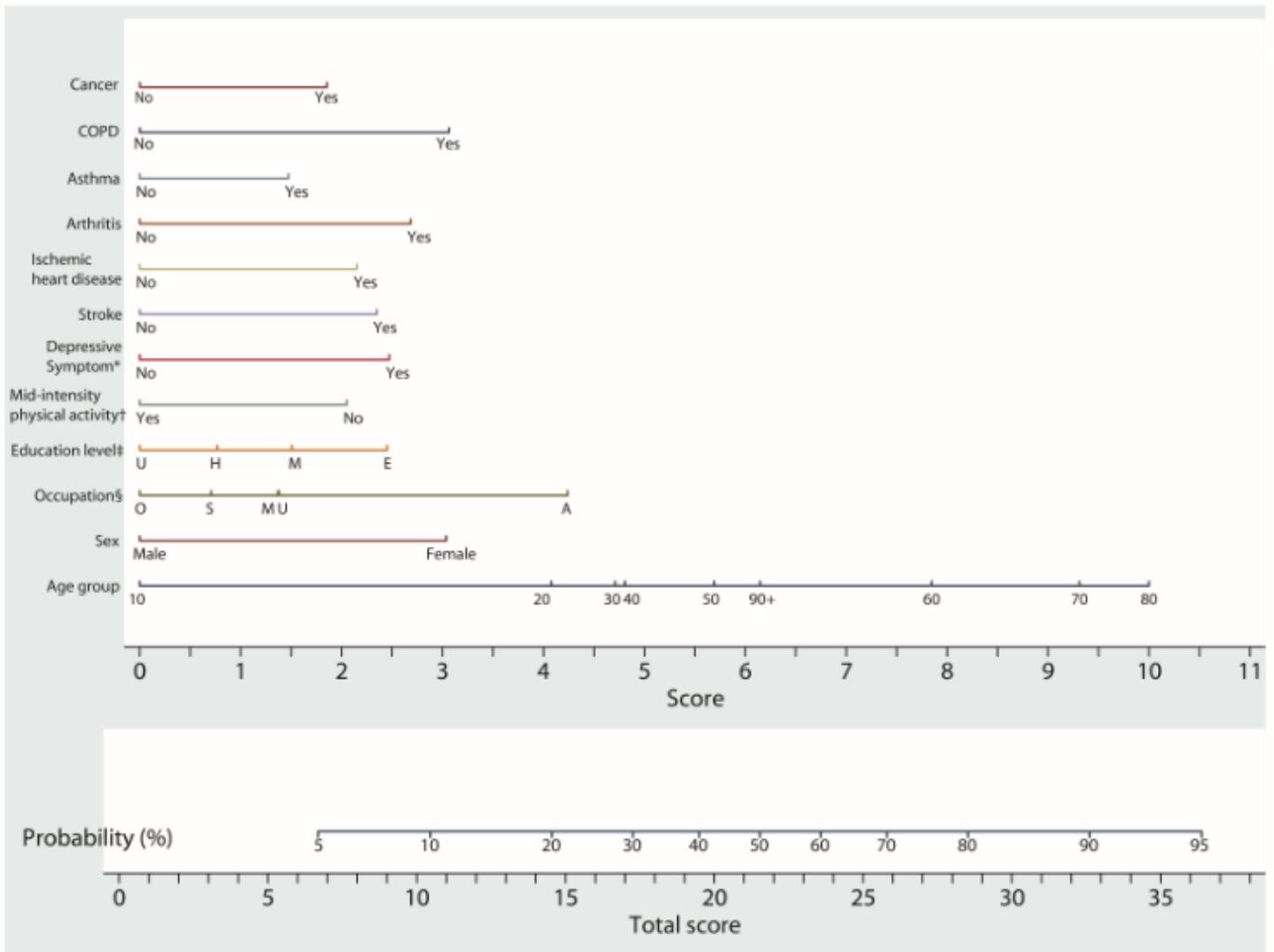


Figure 4

Nomogram for predicting chronic low back pain based on the risk factors. The value of each risk factor is respectively loaded on each variable axis (the 1st-12th lines), and a line is drawn downwards to determine the number of points received for each variable (the 13th line). Then, the sum of these numbers is located on the total points axis (the last line), and a line is drawn upwards to the risk probability axis to determine the likelihood of chronic low back pain. COPD, chronic obstructive pulmonary disease. *Depressive symptom was defined as individuals in this survey who felt sad or had a depressive symptom for two consecutive weeks during the past year. †Mid-intensity physical activity was defined as mid-intensity physical activity for 5 or more days per week for at least 30 minutes. ‡U: ≥ 13 years (college or university), H: 10-12 years (high school), M: 7-9 years (middle school), E: ≤ 6 years (elementary school). §O: Office worker, S: Sales and services, U: Unemployed (student, housewife, etc.), M: Machine fitting and simple labor, A: Agriculture, forestry and fishery.