Development and validation of osteoporosis risk assessment model for Koreans (KORAM)

Sun Min Oh

Department of Medicine

The Graduate School, Yonsei University

Development and validation of osteoporosis risk assessment model for Koreans (KORAM)

Directed by Professor Hyeon Chang Kim

The Doctoral Dissertation
submitted to the Department of Medicine,
the Graduate School of Yonsei University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Sun Min Oh

June 2012

This certifies that the Doctoral Dissertation
of Sun Min Oh is approved.

Thesis Supervisor : Hyeon Chang Kim

Thesis Committee Member#1 : Byung-Ho Nam

Thesis Committee Member#2 : Yumie Rhee

Thesis Committee Member#3 : Seong-Hwan Moon

Thesis Committee Member#4 : Deog Young Kim

The Graduate School
Yonsei University

June 2012

ACKNOWLEDGEMENTS

I would like to express my appreciation for all the generous help given to me by the people around me when I first decided to study preventive medicine and to all of those that helped me finish this dissertation. Above all, I appreciate all the committee members: my advisor Dr. Hyeon Chang Kim who has excellently guided me throughout the research process, Dr. Byung-Ho Nam whose comments helped me to greatly improve this manuscript, Dr. Yumie Rhee who is an excellent role model as a successful female doctor, Dr. Seong-Hwan Moon who guided me to consider future steps of this research, and Dr. Deog Young Kim who taught me how to look at my research from new, interesting angles. In addition, I would like to acknowledge Dr. Dae Ryong Kang who provided me statistical advice and Katherine Stefani who helped to proofread and revise my English manuscript.

I would also like to express my appreciation to all of the professors at the Department of Preventive Medicine: Dr. Il Suh who inspired me to study epidemiology, Dr. Dong Chun Shin who encouraged me to study preventive medicine, Dr. Jaehoon Roh who considerably supported me, Dr. Chung Mo Nam who taught me statistics, and Dr. Changsoo Kim who made studying fun and enjoyable. In addition, I owe sincere and earnest thankfulness to Dr. Woojin Chung for his heartfelt support. I would also like to express appreciation for Dr.

Song Vogue Ahn, Dr. Kyoung Soo Ha, Dr. Hoo-Sun Chang, and Dr. Joo Young Lee who were all grateful seniors and advisors throughout my postgraduate career. I also would like to thank my colleagues in our department: Jeong Lim, Jung Ho, Eun, Jeongbae, Dong Phil, Hye Min, Won-Joon, Jaelim, Jungwoo, Sung-In, Roeul, HwanWoong, and my dear friends: Da Youn, Ji Young, Mi Ra, Mi Suk, Sangah, So Yeon for being supportive during this challenging time.

Through all of these people's encouragement, I was able to succeed in developing and executing my research proposal and develop a true passion for scientific research. Overall, I have learned that passion is necessary to successfully complete what one sets out to accomplish. From the deepest part of my heart, I appreciate everyone's encouragement.

All in all, I would be nowhere without the love and never ending support from my parents and my brother Seung Jae.

<TABLE OF CONTENTS>

ABSTRACT ····· 1

I. INTRODUCTION ····· 3

II. MATERIALS AND METHODS ····· 10
1. Study population ····· 10
A. Development dataset ····· 10
B. Validation dataset ····· 11
2. Measurements ····· 11
3. Statistical analysis ····· 13

III. RESULTS ····· 17
1. Baseline characteristics of postmenopausal women and men in the development and the validation dataset ····· 17
A. Postmenopausal women ····· 17
B. Men ····· 19
2. Development and validation of the Korean osteoporosis risk-assessment model (KORAM) ····· 21
A. Development of candidate models ····· 21
(A) Postmenopausal women ····· 21
(B) Men ····· 22
B. Model selection ····· 25
(A) Postmenopausal women ····· 25
(B) Men ····· 28
C. Performance of the final model ····· 31
(A) Postmenopausal women ····· 31
(B) Men ····· 35
D. Performance of KORAM-F by risk category ····· 39
(A) Postmenopausal women ····· 39
(B) Men ····· 43
E. Comparison of KORAMA with the current NHIC guidelines ··· 47

(A) Postmenopausal women ····· 47
(B) Men ····· 48

IV. DISCUSSION ····· 50

V. CONCLUSION ····· 56

REFERENCES ····· 57

ABSTRACT (IN KOREAN) ····· 64

LIST OF FIGURES

Figure 1. Comparison of AUCs for OSTA and the candidate models to predict osteoporosis in postmenopausal women 26

Figure 2. Comparison of AUCs for OSTA and the candidate models to predict osteoporosis in men 29

Figure 3. Calibration chart of KORAM-F to predict osteoporosis in postmenopausal women 34

Figure 4. Calibration chart of KORAM-M to predict osteoporosis in men 38

Figure 5. Distribution of BMD *T*-scores by KORAM-F scores 40

Figure 6. Distribution of BMD *T*-scores by KORAM-M scores 44

LIST OF TABLES

Table 1. Baseline characteristics of postmenopausal women in the development and the validation dataset 18

Table 2. Baseline characteristics of men in the development and the validation dataset 20

Table 3. Regression coefficients and index weights in the final multiple regression model in postmenopausal women 21

Table 4. Regression coefficients and index weight in the final multiple regression model in men 23

Table 5. Characteristics of OSTA and the candidate models in postmenopausal women 27

Table 6. Characteristics of OSTA and the candidate models in men.....30

Table 7. Discriminatory performance of OSTA and the candidate models to identify postmenopausal women with osteoporosis and low BMD 32

Table 8. Comparison of AUC of each model to identify postmenopausal women with osteoporosis and low BMD.....33

Table 9. Discriminatory performance of OSTA and the candidate models to identify men with osteoporosis and low BMD 36

Table 10. Comparison of AUC of each model to identify men with osteoporosis and low BMD.....37

Table 11. Performance of KORAM-F to predict osteoporosis according to its risk categories 41

Table 12. NRI of KORAM-F compared to OSTA in postmenopausal women.....41

Table 13. NRIs of KORAM-F compared to OSTA risk categories with the different cut-off values in postmenopausal women 42

Table 14. Performance of KORAM-M to predict osteoporosis according to risk categories 45

Table 15. NRI of KORAM-M compared to OSTA in men.....45

Table 16. NRIs of KORAM-M compared to OSTA risk categories with the different cut-off values in men.....46

Table 17. Comparison between the current Korean NHIC guidelines and KORAM-F 49

Table 18. Comparison between the current Korean NHIC guidelines and KORAM-M.....49

ABSTRACT

Development and validation of osteoporosis risk-assessment model for Koreans (KORAM)

Sun Min Oh

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Hyeon Chang Kim)

Objective: Currently, dual energy X-ray absorptiometry (DXA) is the gold standard for detecting osteoporosis. However, it is not recommended for screening purpose because of its high cost. Thus, several osteoporosis risk-assessment models have been developed for pre-screening without the added expenses, but their feasibility in the Korean population has not yet been proved. Therefore, this study aims to develop and validate a new model, named the Korean osteoporosis risk-assessment model for postmenopausal women (KORAM-F) and men (KORAM-M) using a nationally representative dataset.

Methods: Data from the 2009 Korean National Health and Nutrition Examination Survey (KNHANES) with 1,209 postmenopausal women and 1,340 men 50 years or older were used for the development of KORAM. For validation of KORAM, data including 1,046 postmenopausal women and 1,110 men from the 2010 KNHANES were used. Osteoporosis was defined as *T*-score ≤ -2.5 at either femoral neck or lumbar spine. To select variables for the model,

a 10-fold cross validation was performed. Performance of the candidate models and the Osteoporosis Self assessment Tool for Asian (OSTA) were compared by sensitivity, specificity, and area under the receiver operating characteristics curve (AUC). Model fitness was confirmed using the Hosmer-Lemeshow goodness-of-fit test. To compare KORAM with OSTA, a net reclassification improvement was further calculated.

Results: In the development dataset, the prevalence of osteoporosis was 33.9% in women and 8.1% in men. KORAM-F, which consists of age, weight, and hormone therapy, had a sensitivity of 91.2%, a specificity of 50.6%, and an AUC of 0.709 with the specific cut-off score of −9. KORAM-M, based on only age and weight, demonstrated a sensitivity of 90.8%, a specificity of 42.4%, and an AUC of 0.666 with the same cut-off. Additionally, the risk category with KORAM showed an improved reclassification over that of OSTA from 7.4% to 41.8% in women and up to 22.8% in men.

Conclusion: KORAM can be easily used as a pre-screening tool to decide who needs a DXA test to detect osteoporosis. Before using KORAM in clinical practice, the studies investigating cost-effectiveness and replicability in other dataset are necessary.

Key words : Osteoporosis, risk assessment, postmenopausal women, men

Development and validation of osteoporosis risk-assessment model for Koreans (KORAM)

Sun Min Oh

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Hyeon Chang Kim)

I. INTRODUCTION

Osteoporosis is a disease characterized by low bone mass and structural deterioration of bone tissue, leading to bone fragility and an increased susceptibility to fractures, especially of the hip, spine, and wrist.[1] Osteoporosis occurs primarily as a result of normal aging, but significant mortality, morbidity, and health care cost related to osteoporotic fracture is a growing health care burden in the aging population. In fact, osteoporosis affects an estimated 75 million people in Europe, the US, and Japan.[2] Worldwide, the number of osteoporotic fractures was estimated at 9.0 million and osteoporotic fractures accounted for 0.83% of the global burden of non-communicable disease.[3] Moreover, the incidence of osteoporotic fractures is projected to increase by 310% in men and 240% in women by 2050.[4]

In Korea, osteoporosis is also an important public health problem.

According to the 2009 Korean Health Statistics, among participants aged 50 years or older, the prevalence of osteoporosis was 38.7 % in women and 8.1 % in men. Furthermore, according to the National Health Insurance database, which includes almost all prescription and treatment claims with diagnostic codes in Korea, the number of patients being treated for osteoporosis had substantially increased by 26.7% in women and by 29.9% in men between 2005 and 2008.[5] However, osteoporosis is currently under-diagnosed and under-treated. Among people with osteoporosis, 30.3% of women and 6.4% of men were found to be diagnosed with osteoporosis by a physician.[6] Further, only 14.3% of women and 4.3% of men with osteoporosis were reported as being treated.[6]

Although men have a relatively lower risk of having osteoporosis compared to women, men with osteoporotic fracture demonstrated a higher rate of mortality and correspondingly a greater economic impact.[7-9] Thus, earlier intervention of osteoporosis is needed in men as well as women.[10] Accordingly, it is critical that further effort is given to finding more effective methods for prevention and early detection of osteoporosis for men and women.

At present, dual energy X-ray absorptiometry (DXA) is the most widely accepted method to diagnose osteoporosis and to monitor changes in bone density.[11] However, due to the relatively high cost and reimbursement

issue, DXA is not recommended to use for screening osteoporosis.[12] Rather, most guidelines limit the use of DXA to only for postmenopausal women 65 years of age or older and men 70 years of age or older, with the exception of younger adults with certain risk factors.[13] Therefore, several osteoporosis risk-assessment models have been developed for pre-screening without the added expenses. For Western postmenopausal women, various models are available as follows:

- Michaëlsson and colleagues suggested that Swedish postmenopausal women with their body weight under 70 kg have a probability of having osteoporosis with a sensitivity of 92% and a specificity of 45%.[14]
- Lydick and colleagues developed and validated the Simple Calculated Osteoporosis Risk Estimation (SCORE) in mostly white postmenopausal women.[15] With six variables (age, weight, race, fracture history, rheumatoid arthritis history, and estrogen use), SCORE showed a sensitivity of 90% and a specificity of 50% to predict low bone mineral density (BMD) at femoral neck (*T*-score ≤ -2.0).
- Black and colleagues introduced a simple and useful system based on data from the Study of Osteoporotic Fracture (SOF) with mostly white postmenopausal women.[16] SOFSURF used age, weight, smoking, and

history of fracture after age 50 to predict osteoporosis.

- Cadarette and colleagues developed and validated the Osteoporosis Risk Assessment Instrument (ORAI) in women aged 45 years or more, participated in the Canadian Multicentre Osteoporosis Study.[17] ORAI including three variables (age, weight, and estrogen usage) had a sensitivity of 93.3% and a specificity o 46.4% to predict low BMD at either femoral neck or lumbar spine.
- Sedrine and colleagues developed the Osteoporosis Index of Risk (OSIRIS) based on four variables (age, body weight, current hormone replacement therapy (HRT), and history of previous low impact fracture) for Belgian postmenopausal women.[18] The ability of OSIRIS to predict osteoporosis showed a sensitivity of 78.5% and a specificity of 51.4%.
- Salaffi and colleagues developed the Osteoporosis Prescreening Risk Assessment (OPERA) index to predict osteoporosis for women over 50 years of age in Italy.[19] OPERA with five variables (age, weight, history of previous low impact fracture, early menopause, and corticosteroid therapy) had a sensitivity of 88.1% and a specificity of 60.6%.
- Richy and colleagues developed the Osteoporosis Risk Assessment by Composite Linear Estimate (ORACLE) for the US postmenopausal

women aged 45 years and older.[20] With the five components (ultrasonometric bone profile index, age, body mass index (BMI), current use of HRT, and history of fracture after age 45 years) had a sensitivity of 90% and a specificity of 50%.

- Weinstein developed the Age, Body size, No Estrogen (ABONE) to predict osteoporosis in the US postmenopausal women.[21]

For Asian women, the Osteoporosis Screening Tool for Asian (OSTA) was developed in community-dwelling participants who visited 21 clinics in eight Asian countries; Singapore, China, Hong Kong, Korea, Malaysia, Philippines, Taiwan, and Thailand, and validated in Japanese by OSTA research group.[22] OSTA based on only age and weight showed a sensitivity of 91% and a specificity of 45% in the original development dataset and it has been replicated in Korean postmenopausal women,[23] Filipino women,[24] and white women[25] and African American women.[26]

For men, relatively fewer osteoporosis risk-assessment models are available:

- Lynn and colleges developed and validated the Male Osteoporosis Screening Tool (MOST) for Hong Kong Chinese men, based on body weight and quantitative ultrasound index.[27] It provides a sensitivity of 94% and a specificity of 46%.

- Zimering and colleagues developed and validated Mscore with five variables (age, weight, gastrectomy, emphysema, and prior fracture) and a reduced Mscore with age and weight for Caucasian and African American men.[28] Mscore had a sensitivity of 88% and a specificity of 57% and $Mscore_{age\text{-}weight}$ showed a sensitivity of 100% and a specificity of 73%.
- Shepherd and colleagues developed and validated the Male Osteoporosis Risk Estimation Score (MORES) with three variables (age, weight, and history of chronic obstructive pulmonary disease) in the US men.[29] MORES had a sensitivity of 93% and a specificity of 59%.

In addition, OST, same as OSTA, has also been widely validated in American men,[30] African American men,[31] American and Hong Kong Chinese men,[32] Filipino men,[24] and Korean men.[33]

Currently, only OSTA is available for Korean postmenopausal women[22,23] and men 50 years or older.[33] However, the subjects used in the Korean OSTA studies were limited to only patients in each clinic. Meanwhile, from the second half of 2008, the Korea National Health and Nutrition Examination Survey (KNHANES) included DXA tests, which is the first nationwide BMD dataset in the Korean population. Therefore, this study aims to

use these datasets to develop and validate an osteoporosis risk-assessment model for Koreans.

II. MATERIALS AND METHODS

1. Study population

KNHANES, conducted by the Korea Centers for Disease Control and Prevention and the Ministry of Health and Welfare, is a nationwide survey to assess the health and nutritional status of a non-institutionalized representative sample of the Korean population. Participants were selected using a stratified, multi-stage clustered probability sampling design; for the 2009 survey (KNHANES IV-3), household units were selected using the 2005 census in Korea.[6] For the 2010 survey (KNHANES V-1), sampling was based on either the registered market value for apartment building complexes or by selecting registered residents information in the Korean government system.[34]

A. Development dataset

For development of the osteoporosis risk-assessment model, postmenopausal women and men 50 years or older, who participated in the KNHANES IV-3 were included. Among 3,509 participants (1,917 women and 1,592 men), 960 participants were excluded from the present analysis due to at least one of the following reasons: absence of BMD measurement (n = 298), previously diagnosed osteoporosis or treatment for osteoporosis (n = 507), missing blood tests (n = 343), and being in a bed-ridden state (n = 50). Finally, 2,549 participants (1,209 women and 1,340 men) were eligible for this study.

B. Validation dataset

For validation of the developed model, KNAHNES V-1 was used. Among 3,010 participants (1,657 postmenopausal women and 1,353 men aged 50 years or older), 616 participants were excluded in the same manner as the development dataset: absence of BMD measurement (n = 240), previously diagnosed osteoporosis or treatment for osteoporosis (n = 359), missing blood tests (n = 343), and being in a bed-ridden state (n = 17). Finally, data from 2,156 participants (1,046 women and 1,110 men) were used for validation of the developed model.

All participants in this survey signed an informed consent. This study was approved by the Institutional Review Board of Korea Centers for Diseases Control and Prevention (2009-01CON-03-2C, 2010-02CON-21-C) and Yonsei University Health System (4-2011-0222) and monitored by the Human Research Protection Center of Severance Hospital, Yonsei University Health System.

2. Measurements

The survey consists of a health interview survey, a health behavior survey, a nutrition survey, and a health examination survey. Household interview and self-reported questionnaires were used to assess health behavior, past or current history of disease, and family history. Smoking status was

categorized as either current smoker or non-smoker (past or never). Drinking status was classified into either current drinker or non-drinker (past or never). Regular exercise was defined as moderate-to-high intensity of physical activity at least three times per week.

Anthropometrics, blood tests, and BMD measurements were performed in specially equipped mobile examination centers by trained examiners. Body mass index (BMI) was calculated as an individual's weight in kilograms divided by their height in meters squared. Blood samples were obtained after a minimum fasting time of 8 hours and handled according to standard procedures. All samples were analyzed within 24 hours after arriving at the Central Testing Institute in Seoul, Korea. Serum 25-hydroxyvitamin D [25(OH)D] levels were measured using a γ counter (1470 Wizard, Perkin-Elmer Finland) with a RIA (Diasorin, Still Water, MN).[35] Low vitamin D was defined as serum 25(OH)D level of less than 20 ng/mL.[35] Serum parathyroid hormone (PTH) assay was performed using a chemiluminescence assay (Diasorin) for the measurement of intact PTH. Elevated PTH was operationally defined as serum PTH level of 80 pg/mL or over. Elevated alkaline phosphatase (ALP) was operationally defined as serum ALP level of 300 IU/L or over. BMD was measured using a QDR Discovery (formerly, the QDR 4500A) fan beam densitometer (Hologic, Inc., Bedford, MA, USA) at total femur, femoral neck, and L1~L4 spine. The results of DXA measurements were analyzed at the

Korean Society of Osteoporosis using the Hologic Discovery software (version 13.1).[36] The stability of the DXA measurements was maintained by daily calibration.[37] *T*-scores were calculated using sex-specific normal values for Japanese young adults; the referent mean (standard deviation) at the femoral neck and lumbar BMD was 0.803 (0.107) and 1.006 (0.115) for women and 0.846 (0.124) and 1.024 (0.120) for men, respectively. Osteoporosis was defined as *T*-score ≤ -2.5 and low BMD as *T*-score ≤ -2.0 at either the femoral neck or lumbar spine.

3. Statistical analyses

All analyses were performed separately in postmenopausal women and men. Potential risk factors for osteoporosis were selected based on previous studies and statistical investigation of the development dataset. Simple linear regression analyses were performed to detect variables that achieved borderline statistical significance ($p < 0.15$). Among the identified potential risk factors, covariates for the multiple linear regression model were selected by 10-fold cross validation. In detail, the development dataset was randomly divided into 10 subsamples. Nine subsamples were used to select significant covariates using stepwise addition and deletion ($p < 0.15$), and the remaining one subsample was used for validation. This process was repeated ten times with different subsamples to determine the optimal number of covariates.[38] Based on the result of the multiple linear regression analysis with the selected covariates, variables

which did not reach a statistical significance ($p \geq 0.05$) were excluded. Multicollinearity among the investigated variables was checked by computing a variance inflation factor. Then, a final multiple linear regression model with the selected covariates was computed.

BMD *T*-scores subject$_i = \beta_0 + \beta_1\chi_{1i} + \ldots\ldots + \beta_p\chi_{pi} + \varepsilon_i$

β_0 represents the intercept.

χ_p represents a risk factor.

β_p represents the regression coefficient of risk factor, χ_p.

ε_i represents the measurement error.

The regression coefficient (β_p) of each covariate was used to calculate its index weight. To standardize the effect of each variable which was a ratio using a coefficient for each covariate divided by the referent value, the absolute value of the coefficient for age (per 10 years) was calculated. Then, each ratio was multiplied by either 4 for women or 3 for men and rounded off as integers to simply discriminate the effect of each variable.

To develop a simple and effective model, three candidate models were tested: Model 1 included age and weight, Model 2 added behavioral factor(s), and Model 3 added behavioral factor(s) and blood test(s). In addition, these three models were compared to OSTA which was an available model in Korean population.

Correlations of the scores from the three candidate models and OSTA with actual BMD *T*-scores (lower values at either femoral neck or lumbar spine)

were evaluated using Spearman's correlation analyses. The goodness to fit of each model was assessed by using the Hosmer-Lemeshow test.[39] The ability of each model to discriminate those with osteoporosis from those without osteoporosis was compared by using an area under the receiver operating characteristics curves (AUC) with sensitivity on the y-axis and 1− specificity on the x-axis for all possible cut-off values.

Next, a cut-off score was chosen to yield 90% sensitivity or greater to detect those who with osteoporosis in each model.[15,17,22,25] Then, as applied the cut-off score in each model, the final model was selected based on sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio, and negative likelihood ratio with their exact binomial confidence intervals. Additionally, false negative (i.e. the number of missed cases which represents the number of undetected osteoporotic subjects per 1,000 subjects), false positive (i.e. the number of unnecessary DXA tests which represents the number of subjects without osteoporosis referred for DXA testing per 1,000 subjects), and AUC were compared. The model which showed the best performance using the fewest variables was selected and named the Korean Osteoporosis Risk-Assessment Model for postmenopausal women (KORAM-F) and for men (KORAM-M). Then, KORAM were validated using an independent dataset, KNAHNES V-1. Sensitivity analyses were performed with an outcome of low BMD with a *T*-score of –2.0 or less at either the femoral neck or lumbar spine).[17]

According to the scores of KORAM, three risk categories were operationally created: low, intermediate, and high-risk of having osteoporosis. Then, the net reclassification improvement (NRI)[40] was calculated to evaluate whether risk categories of KORAM provided an added value to discriminate participants with osteoporosis from those without osteoporosis, over those of OSTA. NRI was calculated by constructing 3 x 3 tables according to risk categories of KORAM and OSTA, separately in participants with or without osteoporosis. Any upward movement in categories for participants with osteoporosis implied an improved reclassification, and any downward movement indicated a poor reclassification. The interpretation was opposite for participants without osteoporosis.[40]

$$\text{NRI} = [p(\text{up} \mid D=1) - p(\text{down} \mid D=1)] - [p(\text{up} \mid D=0) - p(\text{down} \mid D=0)]$$

(D represents osteoporosis; 1 for osteoporosis, 0 for normal)

A p value less than 5% was considered significant. All statistical analyses were performed using the SAS software package (version 9.2.1; SAS Institute, Cary, NC, USA).

III. RESULTS

1. Baseline characteristics of postmenopausal women and men in the development and the validation dataset

Baseline characteristics of Korean postmenopausal women and men in the development dataset (KNHANES VI-3) and the validation dataset (KNHANES V-1) were summarized in Table 1 and Table 2, respectively.

A. Postmenopausal women

A total of 1,209 and 1,046 postmenopausal women in the development dataset and the validation dataset were eligible for the present study, respectively. Selected potential risk factors of osteoporosis were old age, low body weight, short height, currently smoking, diabetes, a low serum 25(OH)D, an elevated ALP, and an elevated PTH. Meanwhile, protective factors were currently drinking, regular exercise, and currently taking hormones. The mean age was 63.5 years in the development dataset and 62.3 years in the validation dataset. However, the percentage of participants who were 75 years or older was higher in the development dataset, 13.8%, compared to those in the validation dataset, 8.9%. The prevalence osteoporosis was 33.9% in the development dataset and 29.6% in the validation dataset (Table 1).

Table 1. Baseline characteristics of postmenopausal women in the development and the validation dataset

Variables	Development dataset (N = 1,209)	Validation dataset (N = 1,046)
Age (years)	63.5 ± 8.9	62.3 ± 8.2
50 - 54	239 (19.8)	216 (20.7)
55 - 59	215 (17.8)	240 (22.9)
60 - 64	239 (19.8)	201 (19.2)
65 - 69	193 (16.0)	164 (15.7)
70 - 74	156 (12.9)	132 (12.6)
≥ 75	167 (13.8)	93 (8.9)
Weight (kg)	57.1 ± 8.8	57.5 ± 8.6
≥ 60	413 (34.2)	363 (34.7)
55 - 59	296 (24.5)	258 (24.7)
50 - 54	258 (21.3)	235 (22.5)
< 50	242 (20.0)	190 (18.2)
Height (cm)	153.2 ± 5.9	153.6 ± 5.5
Body mass index (kg/m^2)	24.3 ± 3.2	24.3 ± 3.3
Current smoking	56 (4.6)	38 (3.6)
Current drinking	548 (45.3)	509 (48.7)
Regular exercise (≥ 3 time/week)	402 (33.3)	288 (27.5)
Hormone therapy	167 (13.8)	189 (18.1)
History of rheumatoid arthritis	58 (4.8)	43 (4.1)
Diabetes	194 (16.1)	143 (13.7)
Depression	41 (3.4)	36 (3.4)
Serum 25-hydroxyvitamin D (ng/mL)	17.5 [13.2, 22.2]	16.8 [13.2, 21.7]
Serum alkaline phosphatase (IU/L)	251 [211, 300]	249 [210, 302]
Serum parathyroid hormone (pg/mL)	64.8 [50.9, 82.2]	63.7 [52.1, 80.5]
BMD at femoral neck (g/cm^2)	0.63 ± 0.11	0.64 ± 0.11
T-score at femoral neck	−1.64 ± 1.06	−1.53 ± 0.98
T-score < – 1.0	901 (74.5)	763 (72.9)
T-score ≤ – 2.0	435 (36.0)	341 (32.6)
T-score ≤ – 2.5	250 (20.7)	155 (14.8)
BMD at lumbar spine (g/cm^2)	0.81 ± 0.14	0.82 ± 0.14
T-score at lumbar spine	−1.70 ± 1.25	−1.63 ± 1.21
T-score < – 1.0	889 (73.5)	741 (70.8)
T-score ≤ – 2.0	514 (42.5)	418 (40.0)
T-score ≤ – 2.5	319 (26.4)	252 (24.1)
Any of those		
Lower BMD at any of those (g/cm^2)*	0.63 ± 0.11	0.64 ± 0.10
Lower *T*-score at any of those*	−2.04 ± 1.07	−1.96 ± 1.03
T-score < – 1.0	1014 (83.9)	872 (83.4)
T-score ≤ – 2.0	633 (52.4)	514 (49.1)
T-score ≤ – 2.5	410 (33.9)	310 (29.6)

BMD: bone mineral density
Data are expressed as mean ± standard deviation or number (%) or median [interquartile range].
*Lower value at either femoral neck or lumbar spine *T*-score

B. Men

A total of 1,340 men in the development dataset and 1,110 men in the validation dataset were eligible for the present study. Selected potential risk factors of osteoporosis were old age, low body weight, short height, current smoking, diabetes, depression, a low serum 25(OH)D, an elevated ALP, and an elevated PTH. Meanwhile, protective factors were currently drinking and regular exercise. The mean age was 63.4 years in the development dataset and 63.5 years in the validation dataset. The prevalence osteoporosis was 8.1% in the development dataset and 8.2% in the validation dataset (Table 2).

Table 2. Baseline characteristics of men in the development and the validation dataset

Variables	Development dataset (N = 1,340)	Validation dataset (N = 1,110)
Age (years)	63.4 ± 8.9	63.5 ± 8.3
50 - 54	278 (20.8)	193 (17.4)
55 - 59	228 (17.0)	212 (19.1)
60 - 64	253 (18.9)	223 (20.1)
65 - 69	229 (17.1)	197 (17.8)
70 - 74	190 (14.2)	161 (14.5)
≥ 75	162 (12.1)	124 (11.2)
Weight (kg)	66.1 ± 9.7	66.1 ± 9.3
< 50	175 (13.1)	131 (11.8)
50 - 54	185 (13.8)	154 (13.9)
55 - 59	229 (17.1)	232 (20.9)
50 - 54	300 (22.4)	224 (20.2)
65 - 69	215 (16.0)	187 (16.9)
≥ 70	236 (17.6)	182 (16.4)
Height (cm)	166.6 ± 5.9	166.6 ± 5.8
Body mass index (kg/m^2)	23.8 ± 3.0	23.7 ± 2.8
Current smoking	452 (33.7)	317 (28.6)
Current drinking	1001 (74.7)	842 (75.9)
Regular exercise (≥ 3 time/week)	480 (35.8)	353 (31.8)
History of rheumatoid arthritis	11 (0.8)	18 (1.6)
Diabetes	249 (18.6)	208 (18.7)
Depression	20 (1.5)	12 (1.1)
Serum 25-hydroxyvitamin D (ng/mL)	20.4 [15.7, 31.2]	20.2 [15.8, 25.0]
Serum alkaline phosphatase (IU/L)	230 [194, 276]	232 [195, 278]
Serum parathyroid hormone (pg/mL)	64.7 [50.9, 79.3]	62.6 [50.1, 79.0]
BMD at femoral neck (g/cm^2)	0.75 ± 0.12	0.75 ± 0.12
T-score at femoral neck	−0.79 ± 0.98	−0.81 ± 0.94
T-score < – 1.0	568 (42.4)	479 (43.2)
T-score < – 2.0	133 (9.9)	112 (10.1)
T-score < – 2.5	44 (3.3)	35 (3.2)
BMD at lumbar spine (g/cm^2)	0.94 ± 0.16	0.94 ± 0.15
T-score at lumbar spine	−0.70 ± 1.30	−0.67 ± 1.28
T-score < – 1.0	561 (41.9)	471 (42.4)
T-score < – 2.0	199 (14.9)	153 (13.8)
T-score < – 2.5	91 (6.8)	73 (6.6)
Any site[†]		
Lower BMD at any of those (g/cm^2)	0.75 ± 0.12	0.74 ± 0.12
Lower *T*-score at any of those	−1.14 ± 1.03	−1.14 ± 0.99
T-score < – 1.0	747 (55.8)	630 (56.8)
T-score < – 2.0	255 (19.0)	210 (18.9)
T-score < – 2.5	109 (8.1)	91 (8.2)

BMD: bone mineral density

Data are expressed as mean ± standard deviation or number (%) or median [interquartile range].

*Lower value at either femoral neck or lumbar spine *T*-score

2. Development and validation of the Korean Osteoporosis Risk-Assessment Model (KORAM)

A. Development of candidate models

(A) Postmenopausal women

According to 10-fold cross validation, seven variables associated with BMD *T*-scores were selected: age, weight, HRT, regular exercise, current smoking, current drinking, an and elevated serum ALP levels. After adjusted covariates, age, weight, HRT, and an elevated serum ALP levels had significant and independent associations with BMD *T*-scores. No significant multi-collinearity was observed among those variables. Regression coefficient, standard error, and index weight of each variable in the final multiple regression model were demonstrated in Table 3.

Table 3. Regression coefficients and index weights in the final multiple regression model in postmenopausal women

Variables	Regression coefficient	Standard error	*P*-value	Index weight
Intercept	−0.852	0.262	0.001	−
Age (10 years)	−0.516	0.028	<.001	−4
Weight (10 kg)	0.430	0.027	<.001	3
No HRT	−0.358	0.069	<.001	−3
Elevated ALP*	−0.234	0.054	<.001	−2

HRT: hormone therapy

*Elevated ALP: serum alkaline phosphatase ≥ 300 IU/L

Based on the selected variables and their index weights, three candidate models were developed as follows:

- Model 1 = [(age in years/10) x (−4) + (weight in kilogram/10) x 3]
- Model 2 = [(age in years/10) x (−4) + (weight in kilogram/10) x 3 + (if no HRT) x (−3)]
- Model 3 = [(age in years/10) x (−4) + (weight in kilogram/10) x 3 + (if no HRT) x (−3) + (if elevated ALP) x (−2)]

Ranges of scores in Model 1, Model 2, and Model 3 in postmenopausal women were from −23 to 9 (median −8), from −26 to 9 (median -10), and from −27 to 9 (median −11), respectively.

(B) Men

In men, age, weight, regular exercise, a low serum 25(OH)D, and an elevated serum ALP levels were selected by 10-fold cross validation. After adjusted covariates, all variables had significant and independent associations with BMD *T*-scores. No significant multicollinearity was observed among those variables. Regression coefficient, standard error, and index weight of each variable in the final multiple regression model were demonstrated in Table 4.

Table 4. Regression coefficients and index weight in the final multiple regression model in men

Variables	Regression coefficient	Standard error	P-value	Index weight
Intercept	-2.847	0.299	<.0001	−
Age (10 years)	-0.161	0.029	<.0001	-3
Weight (10 kg)	0.441	0.027	<.0001	8
No regular exercise	-0.112	0.051	0.027	-2
Low vitamin D*	-0.119	0.049	0.015	-2
Elevated ALP†	-0.320	0.067	<.0001	-6

*Low vitamin D: serum 25(OH)D < 20 ng/mL, †Elevated ALP: serum alkaline phosphatase ≥ 300 IU/L

Based on the selected variables and their index weights, three candidate models were developed as follows:

- Model 1 = [(age in years/10) x (-3) + (weight in kilogram/10) x 8]
- Model 2 = [(age in years/10) x (-3) + (weight in kilogram/10) x 8 + (if no regular exercise) x (-2)]
- Model 3 = [(age in years/10) x (-3) + (weight in kilogram/10) x 8 + (if no regular exercise) x (-2) + (if low vitamin D) x (-2) + (if elevated ALP) x (-6)]

Ranges of scores in Model 1, Model 2, and Model 3 in men were from 6 to 61 (median 34), from 4 to 61 (median 33), and from −2 to 59 (median 31), respectively.

The range of scores in the candidate models were different by sex; −27

to 9 in postmenopausal women and −2 to 59 in men. Therefore, the scores for the models used for men were adjusted by subtracting 45 from the scores for the women's models:

- Model 1 = [(age in years/10) x (-3) + (weight in kilogram/10) x 8 – 45]
- Model 2 = [(age in years/10) x (-3) + (weight in kilogram/10) x 8 + (if no regular exercise) x (-2) – 45]
- Model 3 = [(age in years/10) x (-3) + (weight in kilogram/10) x 8 + (if no regular exercise) x (-2) + (if low vitamin D) x (-2) + (if elevated ALP) x (-6) – 45]

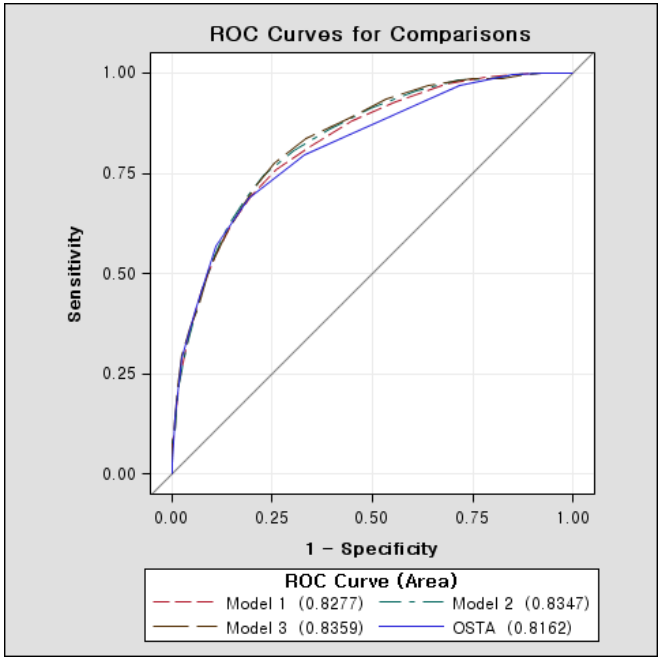
Ranges of the final Model 1, Model 2, and Model 3 in men were from −39 to 16 (median −11), from −41 to 16 (median −12), and from −47 to 14 (median −14), respectively.

B. Model selection

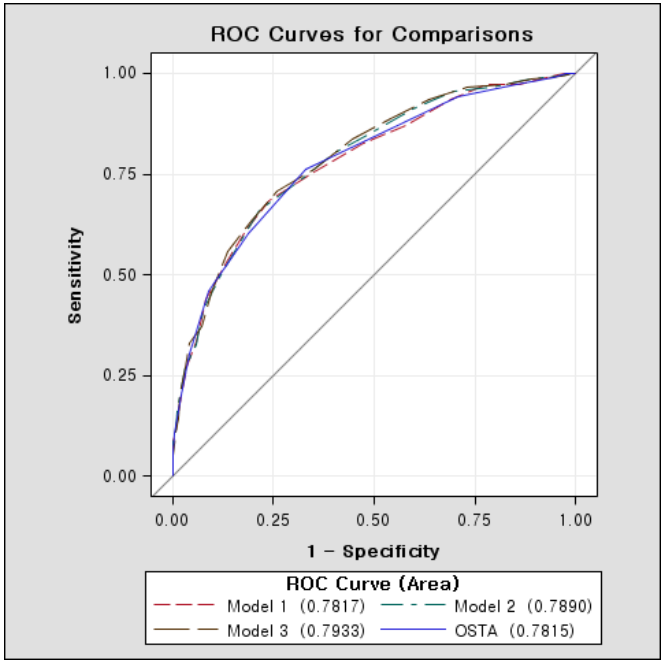
To compare the performance of OSTA and the candidate models to predict osteoporosis, model fitness using the Hosmer-Lemeshow goodness-of-fit test, Spearman's correlation between the predicted scores and actual BMD *T*-scores, and discrimination of models using AUC were calculated.

(A) Postmenopausal women

In the development dataset, all models for postmenopausal women showed reasonable model fitness and the scores in all models had a significantly positive correlation with BMD *T*-scores. Figure 1 showed AUCs of all models. In the development dataset, AUC of Model 2 demonstrated a significantly higher value compared to those of OSTA and Model 1, but comparable to that of Model 3. However in the validation dataset, AUC of each model was not significantly different. For sensitivity analysis, model fitness, correlation, and AUC were also evaluated with the outcome of low BMD (Table 5).



A. Development dataset



B. Validation dataset

Figure 1. Comparison of AUCs for OSTA and the candidate models to predict osteoporosis in postmenopausal women

OSTA: osteoporosis self-assessment tool for Asians, Model 1 included age and weight, Model 2 added HRT, and Model 3 added HRT and an elevated ALP.

Table 5. Characteristics of OSTA and the candidate models in postmenopausal women

Scores in models	*H-L*	Spearman's correlation		ROC			
	p	ρ	p	AUC (SE)	Statistical difference		
Osteoporosis ($T_{FN} \leq -2.5$ or $T_{LS} \leq -2.5$)							
Development dataset							
OSTA	0.541	0.635	<.001	0.816 (0.013)	(ref)		
Model 1	0.773	0.650	<.001	0.828 (0.012)	0.001	(ref)	
Model 2	0.791	0.660	<.001	0.835 (0.012)	<.001	0.017	(ref)
Model 3	0.504	0.665	<.001	0.836 (0.012)	<.001	0.041	0.638
Validation dataset							
OSTA	0.404	0.541	<.001	0.782 (0.016)	(ref)		
Model 1	0.382	0.543	<.001	0.782 (0.016)	0.953	(ref)	
Model 2	0.665	0.551	<.001	0.789 (0.015)	0.153	0.086	(ref)
Model 3	0.510	0.569	<.001	0.793 (0.015)	0.061	0.039	0.208
Low BMD ($T_{FN} \leq -2.0$ or $T_{LS} \leq -2.0$)							
Development dataset							
OSTA	0.032	0.635	<.001	0.795 (0.012)	(ref)		
Model 1	0.222	0.650	<.001	0.809 (0.012)	<.001	(ref)	
Model 2	0.568	0.660	<.001	0.814 (0.012)	<.001	0.145	(ref)
Model 3	0.165	0.665	<.001	0.820 (0.012)	<.001	0.020	0.047
Validation dataset							
OSTA	0.280	0.541	<.001	0.757 (0.014)	(ref)		
Model 1	0.935	0.543	<.001	0.763 (0.015)	0.196	(ref)	
Model 2	0.865	0.551	<.001	0.773 (0.014)	0.006	0.022	(ref)
Model 3	0.922	0.569	<.001	0.783 (0.014)	<.001	<.001	0.002

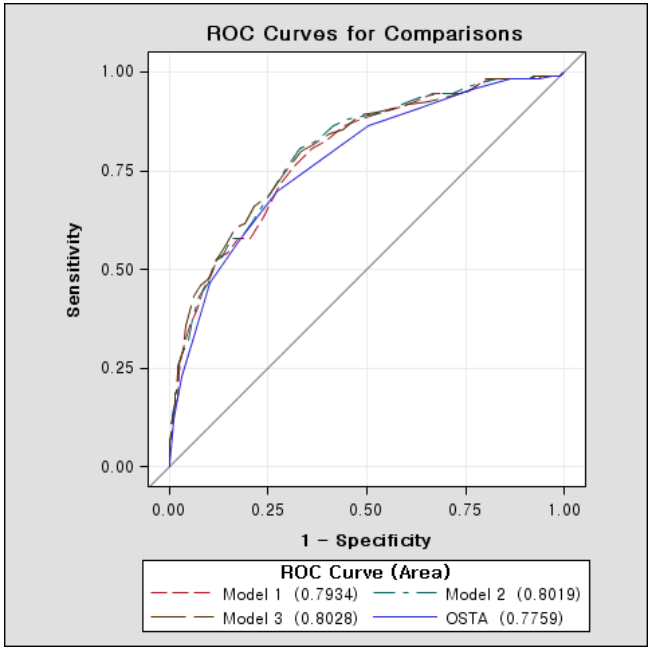
OSTA: osteoporosis self-assessment tool for Asians, BMD: bone mineral density, *H-L*: *H*osmer-*L*emeshow goodness-of-fit test, ROC: receiver operation characteristics, AUC: area under the curve, SE: standard error

T_{FN}: *T*-score at femoral neck, T_{LS}: *T*-score at lumbar spine

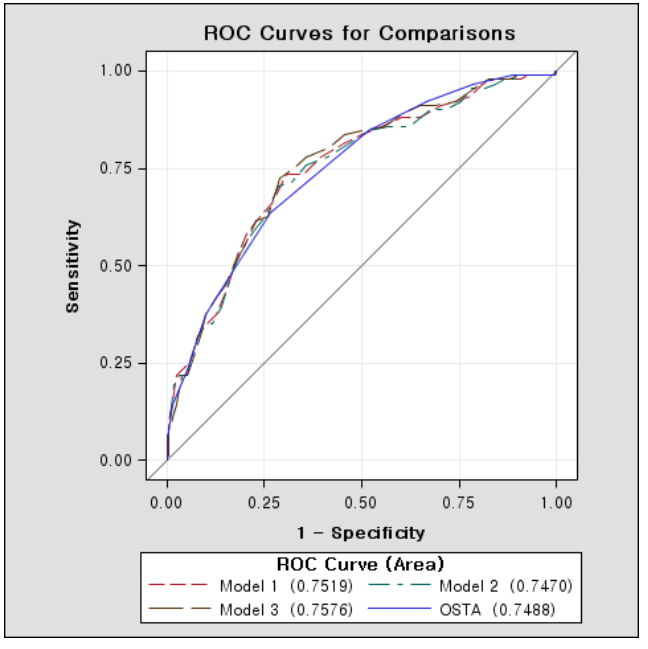
Model 1 included age and weight, Model 2 added HRT, and Model 3 added HRT and an elevated ALP.

(B) Men

All models showed reasonable goodness of fit and the scores in all models had a significantly positive correlation with BMD *T*-scores in the development dataset. Figure 2 showed AUCs of all models. Model 1 and Model 2 demonstrated significantly higher values of AUC compared to that of OSTA. Additionally Model 2 showed a significantly higher value than that of Model 1, but comparable to that of Model 3. However in the validation dataset, AUC of each model was not significantly different. For sensitivity analysis, model fitness, correlation, and AUC were also evaluated with the outcome of low BMD (Table 6).



A. Development dataset



B. Validation dataset

Figure 2. Comparison of AUCs for OSTA and the candidate models to predict osteoporosis in men

OSTA: osteoporosis self-assessment tool for Asians, Model 1 included age and weight, Model 2 added regular exercise, and Model 3 added regular exercise, a low vitamin D, and an elevated ALP.

Table 6. Characteristics of OSTA and the candidate models in men

	H-L	Spearman's correlation		ROC			
	p	ρ	p	AUC (SE)	Statistical difference		
Osteoporosis ($T_{FN} \leq -2.5$ or $T_{LS} \leq -2.5$)							
Development dataset							
OSTA	0.969	0.320	<.001	0.776 (0.024)	(ref)		
Model 1	0.776	0.480	<.001	0.793 (0.023)	0.049	(ref)	
Model 2	0.717	0.482	<.001	0.802 (0.022)	0.003	<.001	(ref)
Model 3	0.630	0.494	<.001	0.803 (0.023)	0.014	0.142	0.875
Validation dataset							
OSTA	0.989	0.277	<.001	0.749 (0.026)	(ref)		
Model 1	0.683	0.442	<.001	0.752 (0.027)	0.767	(ref)	
Model 2	0.915	0.442	<.001	0.747 (0.028)	0.869	0.135	(ref)
Model 3	0.609	0.457	<.001	0.758 (0.026)	0.481	0.514	0.213
Low BMD ($T_{FN} \leq -2.0$ or $T_{LS} \leq -2.0$)							
Development dataset							
OSTA	0.751	0.320	<.001	0.756 (0.017)	(ref)		
Model 1	0.361	0.480	<.001	0.766 (0.012)	0.095	(ref)	
Model 2	0.269	0.482	<.001	0.767 (0.017)	0.083	0.791	(ref)
Model 3	0.639	0.494	<.001	0.769 (0.017)	0.087	0.607	0.654
Validation dataset							
OSTA	0.450	0.277	<.001	0.725 (0.020)	(ref)		
Model 1	0.563	0.442	<.001	0.743 (0.020)	0.014	(ref)	
Model 2	0.837	0.442	<.001	0.742 (0.020)	0.022	0.760	(ref)
Model 3	0.897	0.457	<.001	0.749 (0.019)	0.011	0.339	0.252

OSTA: osteoporosis self-assessment tool for Asians, BMD: bone mineral density, H-L: Hosmer-Lemeshow goodness-of-fit test, ROC: receiver operation curve, AUC: area under the curve, SE: standard error

T_{FN}: *T*-score at femoral neck, T_{LS}: *T*-score at lumbar spine

Model 1 included age and weight, Model 2 added regular exercise, and Model 3 added regular exercise, a low vitamin D, and an elevated ALP.

C. Performance of the final model

To compare model performance, sensitivity, specificity, PPV, NPV, false negative, false positive, positive likelihood ratio, negative likelihood ratio, and AUC of OSTA and the candidate models with the selected cut-off scores to predict osteoporosis.

(A) Postmenopausal women

In case of OSTA, the predefined cut-off score of −1 showed a relatively low sensitivity, 79.5%. Thereby another cut-off score of 0 to yield 90% or greater sensitivity was used as a reference. Compared with other models, Model 2 showed an improved specificity, PPV, false negative, false positive, positive likelihood ratio, and negative likelihood ratio (Table 7). Additionally, AUC of Model 2 was significantly higher than those of OSTA and Model 1, but comparable to that of Model 3 (Table 8). Similar findings were shown in the validation dataset, and in sensitivity analyses in both development and validation dataset (Table 7 and Table 8). Therefore, Model 2 with cut-off score of −9 was finally selected and named the Korean Osteoporosis Risk-Assessment Model for postmenopausal women (KORAM-F).

Calibration chart of KORAM-F displayed the observed and expected risks for deciles of predicted risk in the development and the validation dataset (Figure 3).

Table 7. Discriminatory performance of OSTA and the candidate models to identify postmenopausal women with osteoporosis and low BMD

	Sensitivity	Specificity	PPV	NPV	NMC	NUDT	LR(+)	LR(−)
	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)			(95% CI)	(95% CI)
Osteoporosis ($T_{FN} \leq -2.5$ or $T_{LS} \leq -2.5$)								
Development dataset								
OSTA (≤ −1)	79.5 (75.3, 83.3)	66.8 (63.5, 70.1)	55.2 (51.1, 59.2)	86.4 (83.5, 89.0)	205	332	2.40 (2.15, 2.68)	0.31 (0.25, 0.37)
OSTA (≤ 0)	96.8 (94.6, 98.3)	28.3 (25.2, 31.6)	40.9 (37.8, 44.1)	94.6 (90.9, 97.1)	32	718	1.35 (1.29, 1.42)	0.11 (0.06, 0.19)
Model 1 (≤ −6)	92.7 (89.7, 95.0)	44.1 (40.6, 47.6)	46.0 (42.5, 49.4)	92.2 (89.0, 94.6)	74	560	1.66 (1.55, 1.77)	0.17 (0.12, 0.24)
Model 2 (≤ −9)	91.2 (88.1, 93.8)	50.6 (47.0, 54.1)	48.6 (45.1, 52.2)	91.8 (88.9, 94.2)	88	495	1.85 (1.71, 1.99)	0.17 (0.13, 0.24)
Model 3 (≤ −9)	93.4 (90.6, 95.6)	46.4 (42.9, 50.0)	47.2 (43.7, 50.7)	93.2 (90.3, 95.5)	66	536	1.74 (1.63, 1.87)	0.14 (0.10, 0.21)
Validation dataset								
OSTA (≤ −1)	76.1 (71.0, 80.8)	67.1 (63.6, 70.5)	49.4 (44.8, 54.0)	87.0 (83.9, 89.6)	239	329	2.32 (2.05, 2.61)	0.36 (0.29, 0.44)
OSTA (≤ 0)	94.2 (91.0, 96.5)	29.2 (26.0, 32.6)	35.9 (32.6, 393)	92.3 (88.1, 95.4)	59	708	1.33 (1.26, 1.40)	0.20 (0.13, 0.32)
Model 1 (≤ −6)	87.1 (82.9, 90.6)	42.1 (38.5, 45.8)	38.8 (35.2, 42.5)	88.6 (84.8, 91.7)	129	579	1.50 (1.40, 1.62)	0.31 (0.23, 0.41)
Model 2 (≤ −9)	84.8 (80.4, 88.6)	51.6 (48.0, 55.3)	42.5 (38.6, 46.5)	89.0 (85.6, 91.8)	152	484	1.75 (1.61, 1.92)	0.29 (0.22, 0.39)
Model 3 (≤ −9)	88.7 (84.7, 92.0)	46.1 (42.4, 49.7)	40.9 (37.2, 44.8)	90.6 (87.2, 93.4)	113	540	1.64 (1.52, 1.78)	0.25 (0.18, 0.34)
Low BMD ($T_{FN} \leq -2.0$ or $T_{LS} \leq -2.0$)								
Development dataset								
OSTA (≤ −1)	71.4 (67.7, 74.9)	75.9 (72.2, 79.3)	76.5 (72.9, 79.8)	70.7 (67.0, 74.3)	286	242	2.96 (0.54, 3.45)	0.38 (0.33, 0.43)
OSTA (≤ 0)	93.7 (91.5, 95.5)	34.6 (30.7, 38.6)	61.1 (58.0, 64.2)	83.3 (77.9, 87.8)	64	655	1.43 (1.34, 1.52)	0.18 (0.13, 0.25)
Model 1 (≤ −6)	87.7 (84.9, 90.1)	52.8 (48.6, 56.9)	67.1 (63.8, 70.3)	79.6 (75.2, 83.5)	124	473	1.86 (1.70, 2.03)	0.23 (0.19, 0.29)
Model 2 (≤ −9)	85.2 (82.1, 87.8)	60.1 (55.9, 64.1)	70.1 (66.7, 73.3)	78.6 (74.5, 82.4)	149	340	2.13 (1.92, 2.37)	0.25 (0.20, 0.30)
Model 3 (≤ −9)	88.0 (85.2, 90.4)	55.9 (51.7, 60.0)	68.7 (65.4, 71.9)	80.9 (76.7, 84.7)	121	441	1.20 (1.81, 2.20)	0.21 (0.17, 0.27)
Validation dataset								
OSTA (≤ −1)	66.0 (61.7, 70.1)	73.9 (69.9, 77.6)	70.9 (66.6, 75.0)	69.2 (65.2, 73.0)	341	262	2.52 (2.16, 2.95)	0.46 (0.40, 0.52)
OSTA (≤ 0)	90.9 (88.0, 93.2)	35.0 (30.9, 39.2)	57.4 (54.0, 60.9)	79.8 (74.1, 84.8)	92	651	1.40 (1.31, 1.50)	0.26 (0.19, 0.35)
Model 1 (≤ −6)	82.9 (79.3, 86.0)	49.3 (44.9, 53.6)	61.2 (57.5, 64.8)	74.9 (70.0, 79.3)	172	508	1.63 (1.49, 1.79)	0.35 (0.28, 0.43)
Model 2 (≤ −9)	79.2 (75.4, 82.6)	60.2 (55.9, 64.3)	65.8 (61.9, 69.5)	74.9 (70.6, 79.0)	209	399	1.99 (1.77, 2.23)	0.35 (0.29, 0.42)
Model 3 (≤ −9)	84.1 (80.6, 87.1)	54.9 (50.6, 59.2)	64.3 (60.5, 67.9)	78.1 (73.5, 82.2)	160	452	1.86 (1.68, 2.06)	0.29 (0.23, 0.36)

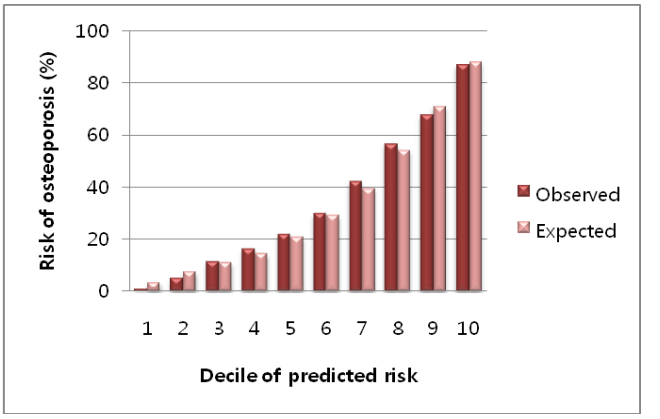
OSTA: osteoporosis self-assessment tool for Asians, BMD: bone mineral density, PPV: positive predictive value, NPV: negative predictive value
NMC: the number of missed cases which represents the number of undetected osteoporotic subjects (i.e. false negatives) per 1,000 subjects
NUDT: the number of unnecessary DXA tests which represents the number of subjects referred for DXA testing (i.e. false positives) per 1,000 subjects
LR(+): positive likelihood ratio, LR(−): negative likelihood ratio, T_{FN}: *T*-score at femoral neck, T_{LS}: *T*-score at lumbar spine
Model 1 included age and weight, Model 2 added HRT, and Model 3 added HRT and an elevated ALP.

Table 8. Comparison of AUC of each model to identify postmenopausal women with osteoporosis and low BMD

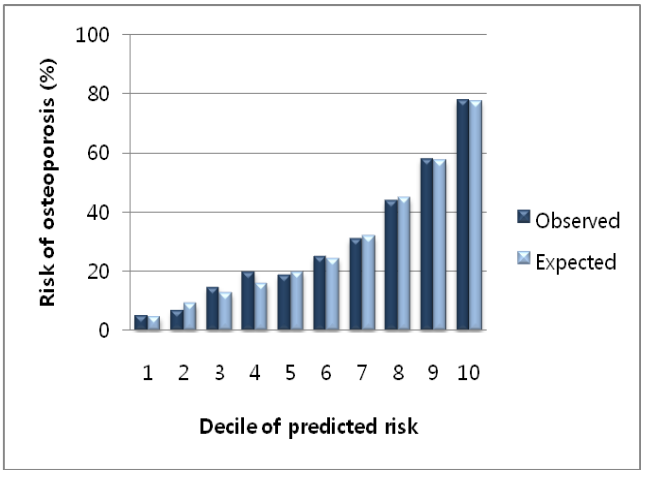
	ROC			
	AUC (SE)	Statistical difference		
Osteoporosis ($T_{FN} \leq -2.5$ or $T_{LS} \leq -2.5$)				
Development dataset				
OSTA (≤ 0)	0.626 (0.009)	(ref)		
Model 1 (≤ −6)	0.684 (0.011)	<.001	(ref)	
Model 2 (≤ −9)	0.709 (0.011)	<.001	<.001	(ref)
Model 3 (≤ −9)	0.699 (0.011)	<.001	0.025	0.056
Validation dataset				
OSTA (≤ 0)	0.617 (0.011)	(ref)		
Model 1 (≤ −6)	0.646 (0.013)	0.002	(ref)	
Model 2 (≤ −9)	0.682 (0.014)	<.001	<.001	(ref)
Model 3 (≤ −9)	0.674 (0.013)	<.001	0.003	0.220
Low BMD ($T_{FN} \leq -2.0$ or $T_{LS} \leq -2.0$)				
Development dataset				
OSTA (≤ 0)	0.641 (0.011)	(ref)		
Model 1 (≤ −6)	0.702 (0.012)	<.001	(ref)	
Model 2 (≤ −9)	0.726 (0.012)	<.001	<.001	(ref)
Model 3 (≤ −9)	0.720 (0.012)	<.001	0.027	0.214
Validation dataset				
OSTA (≤ 0)	0.629 (0.012)	(ref)		
Model 1 (≤ −6)	0.661 (0.014)	0.001	(ref)	
Model 2 (≤ −9)	0.697 (0.014)	<.001	<.001	(ref)
Model 3 (≤ −9)	0.695 (0.014)	<.001	0.005	0.769

AUC: area under the curve, BMD: bone mineral density, ROC: receiver operation characteristics, SE: standard error, T_{FN}: *T*-score at femoral neck, T_{LS}: *T*-score at lumbar spine, OSTA: osteoporosis self-assessment tool for Asians

Model 1 included age and weight, Model 2 added HRT, and Model 3 added HRT and an elevated ALP.



A. Development dataset



B. Validation dataset

Figure 3. Calibration chart of KORAM-F to predict osteoporosis in postmenopausal women

KORAM-F: the Korean Osteoporosis Risk-Assessment Model for postmenopausal women.

(B) Men

In men, the predefined cut-off score of 0 of OSTA showed a relatively low sensitivity, 86.2%. Thereby another cut-off score of 1 to yield 90% or greater sensitivity was used as a reference. Model 1 showed an improved specificity, PPV, NPV comparable to OSTA, but comparable to Model 2 and Model 3 (Table 9). Additionally, AUC of Model 1 was significantly higher than that of OSTA, but comparable to those of Model 2 and Model 3 (Table 10). Therefore, Model 1, only based on age and weight, with cut-off score of −9 was finally selected and named the Korean Osteoporosis Risk-Assessment Model for Men (KORAM-M).

Calibration chart of KORAM-M displayed the observed and expected risks for deciles of predicted risk in the development and the validation dataset (Figure 4).

Table 9. Discriminatory performance of OSTA and the candidate models to identify men with osteoporosis and low BMD

	Sensitivity	Specificity	PPV	NPV	NMC	NUDT	LR(+)	LR(−)
	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)			(95% CI)	(95% CI)
Osteoporosis ($T_{FN} \leq -2.5$ or $T_{LS} \leq -2.5$)								
Development dataset								
OSTA (≤ 0)	86.2 (78.3, 92.1)	49.7 (46.9, 52.6)	13.2 (10.8, 15.9)	97.6 (96.1, 98.7)	138	503	1.72 (1.56, 1.88)	0.28 (0.17, 0.44)
OSTA (≤ 1)	90.8 (83.8, 95.5)	36.9 (34.2, 39.7)	11.3 (9.3, 13.6)	97.8 (96.1, 99.0)	92	632	1.44 (1.34, 1.55)	0.25 (0.14, 0.45)
Model 1 (≤ −9)	90.8 (83.8, 95.5)	42.4 (39.6, 45.2)	12.3 (10.1, 14.7)	98.1 (96.6, 99.1)	92	576	1.58 (1.46, 1.70)	0.22 (0.12, 0.39)
Model 2 (≤ −10)	91.7 (84.9, 96.2)	41.2 (38.4, 44.0)	12.1 (10.0, 14.6)	98.3 (96.7, 99.2)	83	589	1.56 (1.45, 1.68)	0.20 (0.11, 0.38)
Model 3 (≤ −12)	90.8 (83.8, 95.5)	43.3 (40.5, 46.1)	12.4 (10.2, 14.9)	98.2 (96.6, 99.1)	92	567	1.60 (1.48, 1.73)	0.21 (0.12, 0.38)
Validation dataset								
OSTA (≤ 0)	84.6 (75.5, 91.3)	48.4 (45.3, 51.5)	12.8 (10.2, 15.7)	97.2 (95.4, 98.5)	154	517	1.64 (1.47, 1.82)	0.32 (0.20, 0.52)
OSTA (≤ 1)	92.3 (84.8, 96.9)	33.2 (30.3, 36.2)	11.0 (8.9, 13.4)	98.0 (95.9, 99.2)	77	669	1.38 (1.28, 1.49)	0.23 (0.11, 0.48)
Model 1 (≤ −9)	87.9 (79.4, 93.8)	39.7 (36.7, 42.8)	11.5 (9.3, 14.1)	97.4 (95.3, 98.7)	121	603	1.46 (1.33, 1.60)	0.30 (0.17, 0.53)
Model 2 (≤ −10)	85.7 (76.8, 92.2)	37.9 (34.9, 40.9)	11.0 (8.8, 13.5)	96.7 (94.5, 98.3)	143	622	1.38 (1.25, 1.52)	0.38 (0.23, 0.63)
Model 3 (≤ −12)	89.0 (80.7, 94.6)	39.5 (36.4, 42.5)	11.6 (9.3, 14.2)	97.6 (95.6, 98.8)	110	606	1.47 (1.35, 1.60)	0.28 (0.15, 0.50)
Low BMD ($T_{FN} \leq -2.0$ or $T_{LS} \leq -2.0$)								
Development dataset								
OSTA (≤ 0)	80.0 (74.6, 84.7)	53.1 (50.1, 56.1)	28.6 (25.3, 32.1)	91.9 (89.4, 93.9)	200	470	1.71 (1.56, 1.86)	0.38 (0.29, 0.48)
OSTA (≤ 1)	88.2 (83.6, 91.9)	40.0 (37.1, 43.0)	25.7 (22.8, 28.7)	93.5 (90.9, 95.6)	118	600	1.47 (1.38, 1.57)	0.29 (0.21, 0.41)
Model 1 (≤ −9)	85.5 (80.6, 89.6)	45.6 (42.6, 48.6)	27.0 (24.0, 30.2)	93.1 (90.5, 95.1)	146	544	1.57 (1.46, 1.69)	0.32 (0.23, 0.43)
Model 2 (≤ −10)	85.5 (80.6, 89.6)	44.2 (41.2, 47.2)	26.5 (23.5, 29.6)	92.8 (90.3, 94.9)	146	559	1.53 (1.42, 1.65)	0.33 (0.24, 0.45)
Model 3 (≤ −12)	85.1 (80.1, 89.2)	46.5 (43.5, 49.6)	27.2 (24.2, 30.5)	93.0 (90.5, 95.0)	149	535	1.59 (1.48, 1.71)	0.32 (0.24, 0.43)
Validation dataset								
OSTA (≤ 0)	77.1 (70.1, 82.6)	51.0 (47.7, 54.3)	26.9 (23.4, 30.1)	90.5 (88.0, 92.9)	229	490	1.57, 1.43, 1.74)	0.45 (0.35, 0.58)
OSTA (≤ 1)	86.7 (81.3, 91.0)	35.2 (32.1, 384)	23.8 (20.8, 27.0)	91.9 (88.5, 94.5)	134	648	1.34 (1.25, 1.44)	0.38 (0.27, 0.54)
Model 1 (≤ −9)	84.8 (79.2, 89.3)	42.7 (39.4, 46.0)	25.7 (22.4, 29.1)	92.3 (89.3, 94.7)	153	574	1.48 (1.36, 1.60)	0.36 (0.26, 0.50)
Model 2 (≤ −10)	86.2 (80.8, 90.6)	41.1 (37.9, 44.4)	25.5 (22.3, 28.8)	92.7 (89.7, 95.1)	139	589	1.46 (1.36, 1.58)	0.34 (0.24, 0.48)
Model 3 (≤ −12)	86.7 (81.3, 91.0)	42.7 (39.4, 46.0)	26.1 (22.9, 29.5)	93.2 (90.3, 95.4)	134	574	1.50 (1.39, 1.63)	0.32 (0.23, 0.46)

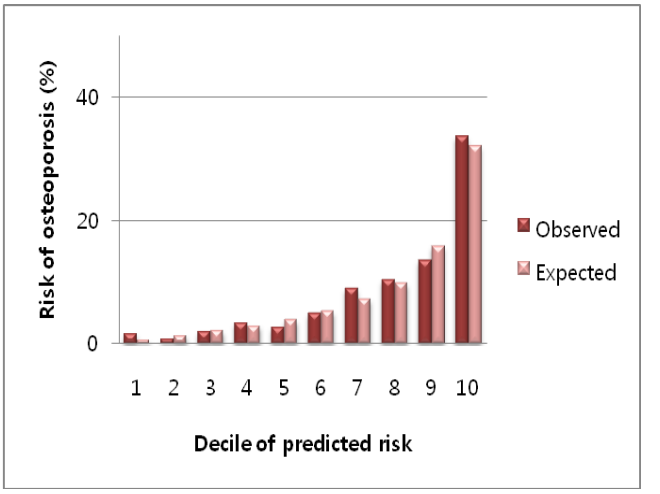
OSTA: osteoporosis self-assessment tool for Asians, BMD: bone mineral density, PPV: positive predictive value, NPV: negative predictive value
NMC: the number of missed cases which represents the number of undetected osteoporotic subjects (i.e. false negatives) per 1,000 subjects
NUDT: the number of unnecessary DXA tests which represents the number of subjects referred for DXA testing (i.e. false positives) per 1,000 subjects
LR(+): positive likelihood ratio, LR(−): negative likelihood ratio, T_{FN}: *T*-score at femoral neck, T_{LS}: *T*-score at lumbar spine
Model 1 included age and weight, Model 2 added regular exercise, and Model 3 added regular exercise, a low vitamin D and an elevated ALP.

Table 10. Comparison of AUC of each model to identify men with osteoporosis and low BMD

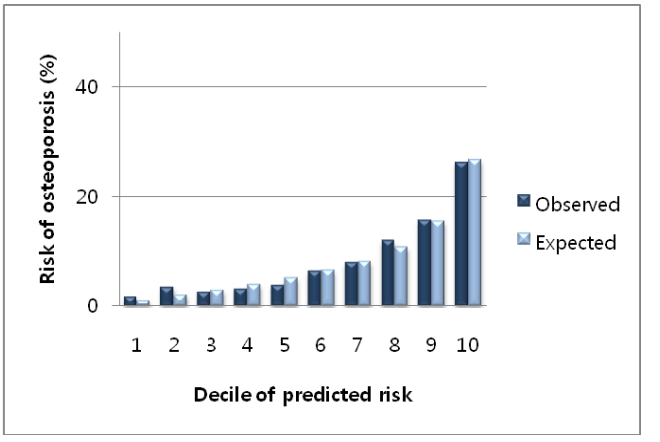
	ROC			
	AUC (SE)	Statistical difference		
Osteoporosis ($T_{FN} \leq -2.5$ or $T_{LS} \leq -2.5$)				
Development dataset				
OSTA (≤ 1)	0.639 (0.016)	(ref)		
Model 1 (≤ −9)	0.666 (0.016)	<.001	(ref)	
Model 2 (≤ −10)	0.665 (0.015)	0.004	0.780	(ref)
Model 3 (≤ −12)	0.671 (0.016)	<.001	0.559	0.311
Validation dataset				
OSTA (≤ 1)	0.627 (0.016)	(ref)		
Model 1 (≤ −9)	0.638 (0.019)	0.365	(ref)	
Model 2 (≤ −10)	0.618 (0.020)	0.506	0.017	(ref)
Model 3 (≤ −12)	0.642 (0.018)	0.272	0.703	0.017
Low BMD ($T_{FN} \leq -2.0$ or $T_{LS} \leq -2.0$)				
Development dataset				
OSTA (≤ 1)	0.642 (0.013)	(ref)		
Model 1 (≤ −9)	0.656 (0.013)	0.091	(ref)	
Model 2 (≤ −10)	0.648 (0.013)	0.434	0.142	(ref)
Model 3 (≤ −12)	0.658 (0.014)	0.071	0.749	0.160
Validation dataset				
OSTA (≤ 1)	0.609 (0.014)	(ref)		
Model 1 (≤ −9)	0.637 (0.015)	0.004	(ref)	
Model 2 (≤ −10)	0.636 (0.015)	0.009	0.931	(ref)
Model 3 (≤ −12)	0.647 (0.014)	0.001	0.287	0.218

AUC: area under the curve, BMD: bone mineral density, ROC: receiver operation characteristics, SE: standard error, T_{FN}: *T*-score at femoral neck, T_{LS}: *T*-score at lumbar spine, OSTA: osteoporosis self-assessment tool for Asians

Model 1 included age and weight, Model 2 added regular exercise, and Model 3 added regular exercise, a low vitamin D, and an elevated ALP.



A. Development dataset



B. Validation dataset

Figure 4. Calibration chart of KORAM-M to predict osteoporosis in men

KORAM-M: the Korean Osteoporosis Risk-Assessment Model for Men.

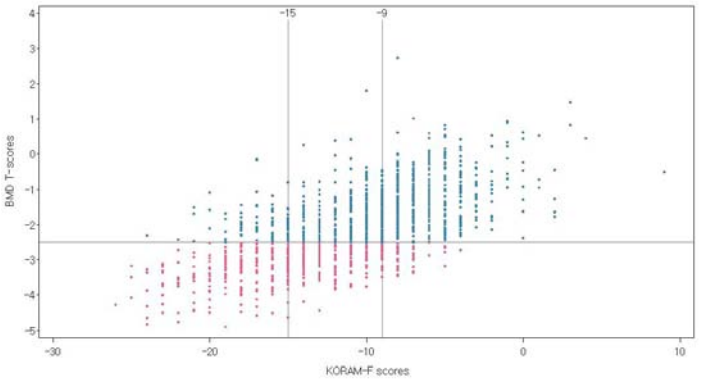
D. Performance of KORAM by risk category

(A) Postmenopausal women

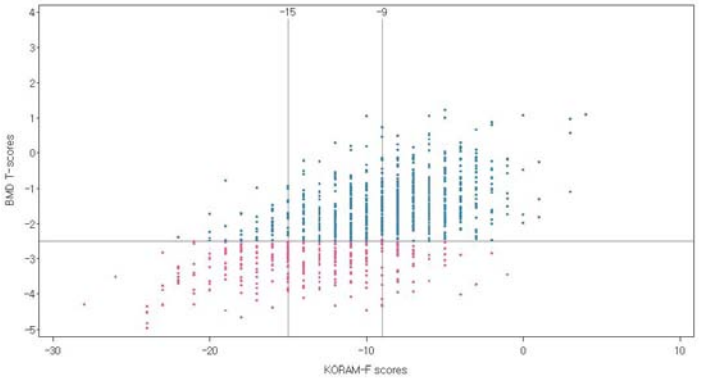
To define clinical implications of KORAM-F, three risk categories were created with cut-off values of −9 and −15; > −9 for low risk, −15 - −9 for intermediate risk, and < −15 for high-risk group. Scatter plots of distribution of BMD *T*-scores by KORAM-F scores were illustrated in Figure 5.

In the development dataset, 36.4%, 43.1%, and 20.5% were classified into low, intermediate, and high risk categories, respectively. Among women in the low, intermediate, and high-risk category, 8.2%, 35.7%, and 75.8% had osteoporosis, respectively. In the validation dataset, the percentage of women in the low, intermediate, and high-risk category was 40.5%, 45.4%, and 13.8%, respectively. The prevalence of osteoporosis was 11.0%, 34.1%, and 70.1% in the low, intermediate, and high-risk category (Table 11).

Table 12 demonstrated the NRI of KORAM-F compared to OSTA. The net improvement was 28.5% (117 out of 410) for participants with osteoporosis, −21.2% (169 out of 799) for those without osteoporosis, and 7.4% (95% CI 1.1 - 13.6), overall. Further, the NRIs of KORAM-F compared to OSTA showed an improved classification from 7.4% to 41.7% across the different cut-off scores of OSTA; −4 and −1, −3 and 0, −4 and 0, and −5 and 0 (Table 13).



A. Development dataset



B. Validation dataset

Figure 5. Distribution of BMD *T*-scores by KORAM-F scores

KORAM-F: the Korean Osteoporosis Risk-Assessment Model for postmenopausal women.

Table 11. Performance of KORAM-F to predict osteoporosis according to its risk categories

Risk category	Development dataset		Validation dataset	
	Total N (column %)	Osteoporosis N (row %)	Total N (column %)	Osteoporosis N (row %)
High (< −15)	248 (20.5)	188 (75.8)	144 (13.8)	101 (70.1)
Intermediate (−15 - −9)	521 (43.1)	186 (35.7)	475 (45.4)	162 (34.1)
Low (> −9)	440 (36.4)	36 (8.2)	427 (40.8)	47 (11.0)
Total	1209 (100.0)	410 (33.9)	1046 (100.0)	310 (29.6)

KORAM-F: the Korean Osteoporosis Risk-Assessment Model for postmenopausal women

Table 12. NRI of KORAM-F compared to OSTA in postmenopausal women

OSTA risk category	KORAM-F risk category			Total
	Low (> −9)	Intermediate (−15 - −9)	High (< −15)	
Participants having osteoporosis				
Low (> −1)	35	49	0	84
Intermediate (−4 - −1)	1	136	70	207
High (< −4)	0	1	118	119
Participants without having osteoporosis				
Low (> −1)	394	140	0	534
Intermediate (−4 - −1)	10	195	39	244
High (< −4)	0	0	21	21

NRI: Net Reclassification Improvement
KORAM-F: the Korean Osteoporosis Risk-Assessment Model for postmenopausal women
The net improvement was 28.5% (117/410) for participants with osteoporosis, −21.2% (169 out of 799) for those without osteoporosis, and 7.4% overall.

Table 13. NRIs of KORAM-F compared to OSTA risk categories with the different cut-off values in postmenopausal women

OSTA risk category	Total	Osteoporosis	Participants with osteoporosis, %		Participants without osteoporosis, %		NRI (95% CI), %
	N (column %)	N (row %)	Up	Down	Up	Down	
Cut-off: −4, −1							
High (< −4)	215 (17.8)	168 (78.1)	0.0	0.2	0.0	0.0	
Intermediate (−4 - −1)	376 (31.1)	158 (42.0)	17.1	0.2	4.9	1.3	7.4 (1.1, 13.6)
Low (> −1)	618 (51.1)	84 (13.6)	12.0	0.0	17.5	0.0	
Cut-off: −3, 0							
High (< −3)	320 (26.5)	233 (72.8)	0.0	0.7	0.0	0.1	
Intermediate (−3 - 0)	650 (53.8)	164 (25.2)	5.6	5.6	1.8	22.3	19.9 (15.1, 24.7)
Low (> 0)	239 (19.8)	13 (5.4)	0.0	0.0	0.0	0.0	
Cut-off: −4, 0							
High (< −4)	215 (17.8)	168 (78.1)	0.0	0.1	0.0	0.0	
Intermediate (−4 - 0)	755 (62.4)	229 (30.3)	17.1	5.6	4.9	22.3	28.6 (22.7, 34.5)
Low (> 0)	239 (19.8)	13 (5.4)	0.0	0.0	0.0	0.0	
Cut-off: −5, −0							
High (< −5)	140 (11.6)	119 (85.0)	0.0	0.0	0.0	0.0	
Intermediate (−5 - 0)	830 (68.7)	278 (33.5)	31.7	5.6	6.6	22.3	41.7 (34.8, 48.7)
Low (> 0)	239 (19.8)	13 (5.4)	0.0	0.0	0.0	0.0	

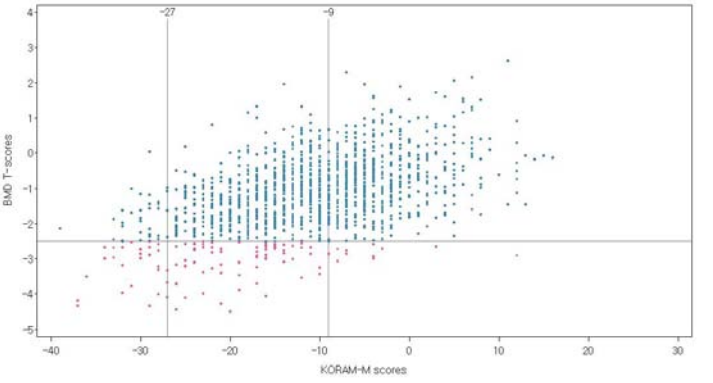
NRI: Net Reclassification Improvement, KORAM-F: the Korean Osteoporosis Risk-Assessment Model for postmenopausal women, OSTA: Osteoporosis Self-assessment Tool for Asians

(B) Men

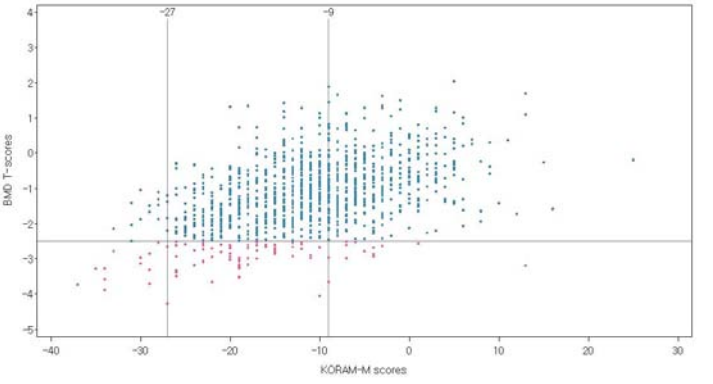
To define clinical implications of KORAM-M, three risk categories were created with the cut-off values of −9 and −27; > −9 for low risk, −27 - −9 for intermediate risk, and < −27 for high-risk group. Scatter plots of distribution of BMD *T*-scores by KORAM-M scores were illustrated in Figure 6.

In the development dataset, 39.7%, 56.1%, and 4.2% were classified into low, intermediate, and high risk categories, respectively. Among men in the low, intermediate, and high-risk category, 1.9%, 9.7%, and 46.4% had osteoporosis, respectively. In the validation dataset, the percentage of men in the low, intermediate, and high-risk category was 37.5%, 60.4%, and 2.2%, respectively. The prevalence of osteoporosis was 2.6%, 10.1%, and 50.0% in the low, intermediate, and high-risk category (Table 14).

Table 15 demonstrated the NRI of KORAM-M compared to OSTA. The net improvement was 0.9% (1 out of 109) for participants with osteoporosis, 6.3% (78 out of 1231) for those without osteoporosis, and 7.3% (95% CI 0.5 – 14.0), overall. Further, the NRIs of KORAM-M compared to OSTA showed an improved classification up to 22.8% across the different cut-off scores of OSTA; −3 and 1, −4 and 1, −5 and 1, and −6 and 1 (Table 16).



A. Development dataset



B. Validation dataset

Figure 6. Distribution of BMD *T*-scores by KORAM-M scores

KORAM-M: the Korean Osteoporosis Risk-Assessment Model for Men.

Table 14. Performance of KORAM-M to predict osteoporosis according to risk categories

Risk category	Development dataset		Validation dataset	
	Total N (column %)	Osteoporosis N (row %)	Total N (column %)	Osteoporosis N (row %)
High (< −27)	56 (4.2)	26 (46.4)	24 (2.2)	12 (50.0)
Intermediate (−27 - −9)	752 (56.1)	73 (9.7)	670 (60.4)	68 (10.1)
Low (> −9)	532 (39.7)	10 (1.9)	416 (37.5)	11 (2.6)
Total	1340 (100.0)	109 (8.1)	1110 (100.0)	91 (8.2)

KORAM-M: the Korean Osteoporosis Risk-Assessment Model for Men.

Table 15. NRI of KORAM-M compared to OSTA in men

OSTA risk category	KORAM-M risk category			Total
	Low (> -9)	Intermediate (-27 − -9)	High (< -27)	
Participants having osteoporosis				
Low (> 1)	9	1	0	10
Intermediate(−4 - 1)	1	67	6	74
High (< −4)	0	5	20	25
Participants without having osteoporosis				
Low (> 1)	427	27	0	454
Intermediate(−4 - 1)	95	638	4	737
High (< −4)	0	14	26	40

KORAM-M: the Korean Osteoporosis Risk-Assessment Model for Men
The net improvement for event was 0.9% (1/109) and for those with non-event was 6.3% (78/1231), and 7.3% overall.

Table 16. NRIs of KORAM-M compared to OSTA risk categories with the different cut-off values in men

OSTA risk category	Total	Osteoporosis	Participants with osteoporosis, %		Participants without osteoporosis, %		NRI (95% CI), %
	N (column %)	N (row %)	Up	Down	Up	Down	
Cut−off: −3, 1							
High (< −3)	117 (8.7)	37 (31.6)	0.0	11.9	0.0	4.1	
Intermediate (1 - −3)	759 (56.6)	62 (8.2)	1.8	0.9	0.0	7.7	−0.5 (−8.2, 7.2)
Low (> 1)	464 (34.6)	10 (2.2)	0.9	0.0	2.2	0.0	
Cut−off: −4, 1							
High (< −4)	65 (4.9)	25 (38.5)	0.0	4.6	0.0	1.1	
Intermediate (1 - −4)	811 (60.5)	74 (9.1)	5.5	0.9	0.3	7.7	7.3 (0.5, 14.0)
Low (> 1)	464 (34.6)	10 (2.2)	0.9	0.0	2.2	0.0	
Cut−off: −5, 1							
High (< −5)	27 (2.0)	13 (48.1)	0.0	0.0	0.0	0.0	
Intermediate (1 - −5)	849 (63.4)	86 (10.1)	11.9	0.9	1.3	7.7	16.2 (8.9, 23.4)
Low (> 1)	464 (34.6)	10 (2.2)	0.9	0.0	2.2	0.0	
Cut−off: −6, 1							
High (< −6)	10 (0.7)	5 (50.0)	0.0	0.0	0.0	0.0	
Intermediate (1 - −6)	866 (64.6)	94 (10.9)	19.3	0.9	2.0	7.7	22.8 (13.9, 31.6)
Low (> 1)	464 (34.6)	10 (2.2)	0.9	0.0	2.2	0.0	

NRI: Net Reclassification Improvement, KORAM-M: the Korean Osteoporosis Risk-Assessment Model for men,
OSTA: Osteoporosis Self-assessment Tool for Asians

E. Comparison of KORAM with the current Korean NHIC guidelines

Currently the Korean National Health Insurance Corporation (NHIC) guidelines for screening osteoporosis using DXA limited to women 65 years of age or older and men 70 years of age and older with exceptions being younger women with low body weight (BMI < 18.5), early menopause (before 40 years old), surgical menopause, and past history or family history of non-traumatic fracture, etc. Meanwhile, as KORAM was developed using each individual's age, weight, and HRT status, it may provide a more efficient targeting of DXA tests. Thus, we compared the performance of KORAM with the current NHIC guidelines.

(A) Postmenopausal women

According to the current NHIC guidelines, 64.1% of women in the development dataset would be recommended to measure BMD by DXA. With the NHIC guidelines, 85.1% of osteoporosis cases could be detected. The number of missed cases, i.e. false negative, was estimated to 149 per 1,000 subjects. And the number of unnecessary DXA tests, i.e. false positive, was estimated to 534 per 1,000 subjects. Compared to the NHIC guidelines, KORAM-F showed a better performance, overall. With KORAM-F, 91.2% of patients with osteoporosis could be detected. Moreover, the number of missed cases and unnecessary DXA tests per 1,000 subjects would reduce to 88 and 397, respectively (Table 17).

(B) Men

In men, the number of recommended DXA tests of KORAM-M would be more than double than that of the NHIC guidelines; 60.3% vs. 27.8% of men, respectively. With KORAM-M, 90.8% of men with osteoporosis could be detected, but only 60.0% of men with osteoporosis would be detected with the NHIC guidelines. Accordingly, the number of missed cases per 1,000 subjects of KORAM-M was estimated to 92, which would be lower than that of the NHIC guidelines, 391. However, the number of unnecessary DXA tests per 1,000 subjects of KORAM-M would be 576, which was higher than that of NHIC guidelines, 307 (Table 18).

Table 17. Comparison between the current Korean NHIC guidelines and KORAM-F

	NHIC guideline	**KORAM-F**
No. of subjects recommended for DXA testing	775/1,209 (64.1%)	769/1,209 (63.6%)
Sensitivity	349/410 (85.1%)	374/410 (91.2%)
Specificity	373/799 (46.7%)	404/799 (50.6%)
PPV	349/775 (45.0%)	374 /769 (48.6%)
NPV	373/434 (85.9%)	404/440 (91.8%)
No. of missed cases	61/410 (149 per 1,000 subjects)	36/410 (88 per 1,000 subjects)
No. of unnecessary DXA tests	426/799 (534 per 1,000 subjects)	395/799 (495 per 1,000 subjects)

NHIC: the National Health Insurance Corporation, KORAM-F: the Korean Osteoporosis Risk-Assessment Model for postmenopausal women, DXA: dual-energy X-ray absorptiometry, PPV: positive predictive value, NPV: negative predictive value

Table 18. Comparison between the current Korean NHIC guidelines and KORAM-M

	NHIC guideline	**KORAM-M**
No. of subjects recommended for DXA testing	373/1,340 (27.8%)	808/1,340 (60.3%)
Sensitivity	66/110 (60.0%)	99/109 (90.8%)
Specificity	924/1231 (75.1%)	522/1231 (42.4%)
PPV	66/373 (17.7%)	99/808 (12.3%)
NPV	924/967 (95.6%)	522/532 (98.1%)
No. of missed cases	43/110 (391 per 1,000 subjects)	10/109 (92 per 1,000 subjects)
No. of unnecessary DXA tests	307/1231 (250 per 1,000 subjects)	709/1231 (576 per 1,000 subjects)

NHIC: the National Health Insurance Corporation, KORAM-M: KORAM-F: the Korean Osteoporosis Risk-Assessment Model for men, DXA: dual-energy X-ray absorptiometry, PPV: positive predictive value, NPV: negative predictive value

IV. DISCUSSION

In this study, the Korean osteoporosis risk-assessment model for postmenopausal women (KORAM-F) and men (KORAM-M) were developed and validated in Korean population based on a nationally representative BMD and health examination dataset.

- KORAM-F = [(age in years/10) x (−4) + (weight in kilogram/10) x 3 + (if no HRT) x (−3)].
- KORAM-M = [(age in years/10) x (−3) + (weight in kilogram/10) x 8 – 45].

Sensitivity, specificity, and AUC of KORAM-F with the specific cut-off score of− 9 were 91.2%, 50.6%, and 0.709 in the development dataset and 84.8%, 51.6%, and 0.682 in the validation dataset, respectively. Sensitivity, specificity, and AUC of KORAM-M with the same cut-off were 90.8%, 42.4%, and 0.666 in the development dataset and 87.9%, 39.7%, and 0.638 in the validation dataset, respectively.

To develop KORAM, we investigated clinically or statistically significant factors to be associated with BMD; age, body weight, height, current smoking, currently drinking, regular exercise, hormone therapy, diabetes, depression, a low serum 25(OH)D, an elevated ALP, and an elevated PTH. Of those, four variables in women (age, weight, HRT, and an elevated ALP) and

five variables in men (age, weight, regular exercise, an elevated ALP, and a low 25(OH)D were selected as potential components of the risk-assessment models. To form a simple and an effective model, we first defined a baseline model with age and weight, which were core variables in the previously developed models.[15-19,21,22,28,29] Then, we assessed an incremental effect of adding HRT status (only for women) and regular exercise (only for men) to predict osteoporosis. Currently taking HRT is a protective factor for osteoporosis and it has been applied in previous models including SCORE,[15] ORAI,[17] OSIRIS,[18] ORACLE,[20] and ABONE.[21] Meanwhile, the effect of regular exercise on BMD is still controversial. In the present study, a positive association between regular exercise and BMD was shown in men, but not in women. Although only a few studies are available, our findings are consistent with previous studies that found the effects of exercise on BMD to be more prominent in men.[41,42] Therefore, we assessed regular exercise as a component of osteoporosis risk-assessment model in men, even though it has never been used in previous models. Additionally, in this study, serum ALP levels were negatively correlated with BMD in both men and women, and vitamin D levels were positively correlated to BMD only in men. Vitamin D deficiency is known to exacerbate osteoporosis[43] and osteoporotic fracture,[44] and ALP is also related to an increased risk of osteoporotic fracture.[45] Thus, we evaluated whether invasive laboratory tests further improved the prediction of osteoporosis. In the present study population, Sensitivity, specificity, PPV, NPV, false positive, false

negative, positive likelihood ratio, negative likelihood ratio, and AUC were compared to evaluated performance of each model. In postmenopausal women, the model with age, weight, and HRT showed better performance than the baseline model. However, further addition of an elevated ALP did not improve the performance. In men, the baseline model was comparable to other models adding regular exercise, and laboratory tests. Since a low serum 25(OH)D and an elevated ALP did not provide an additive value to predict osteoporosis, KORAM-F was developed based on age, weight, and HRT and KORAM-M was established with age and weight. Therefore, KORAM-F and KORAM-M can be easily used in a primary care setting for pre-screening to decide whether to test DXA as well as in the general population for self-screening purposes.

In the present study, we evaluated both linear and logistic regression models to calculate an index weight of each variable to make a scoring system to predict osteoporosis. However, when we compared the performance of the models, the scoring system using the weights from linear regression analyses slightly outperformed those from logistic regression analyses (data not shown).

Additionally, all candidate models were compared with OSTA which has been extensively validated in many countries.[22-26,31-33] The previously defined cut-off score of -1 for postmenopausal women[22,23] and 0 for men[33] demonstrated a relatively low sensitivity in this study population. Therefore,

new cut-off scores which yielded 90% sensitivity or greater to detect those with osteoporosis were applied; 0 for postmenopausal women and 1 for men. Overall, both KORAM-F and KORAM-M showed a better performance to detect osteoporosis than OSTA.

Then, we classified KORAM into three risk categories and calculated NRI to check whether KORAM provides an improved or worsened risk classification when compared with OSTA. Based on KORAM-F, the probability of having osteoporosis was about 10% in the low risk category and more than 70% in the high risk category (Table 11). KORAM-F demonstrated an improved net reclassification from 7.4% to 41.8% over OSTA across the different cut-off scores. Interestingly, from 13.3% to 41.8% of women in the intermediate risk category of OSTA were reclassified into the high or low risk category of KORAM-F (Table 13). In KORAM-M, the probability of having osteoporosis was less than 5% in the low risk category and more than 45% in the high risk category (Table 14). NRIs of KORAM-M compared to OSTA were up to 22.8% across the different cut-off scores of OSTA. In men, 8.6% to 24.1% in the intermediate risk category of OSTA were reclassified into the high or low risk category of KORAM-M (Table 16).

Currently, the Korean NHIC reimburses BMD measurements by DXA for women 65 years of age or older and men 70 years of age or older with a few

exceptions. When compared with the Korean NHIC guidelines, KORAM-F provided a more efficient targeting of DXA tests because it considers each individual's age, weight, and HRT status (Table 17). Meanwhile, KORAM-M showed an increased sensitivity by 30%, but the number of target population to test DXA was more than double compared to that of the current guidelines (Table 18). Osteoporosis itself presents no specific symptoms, but the burden of osteoporotic fracture continues to grow in men as well as women with aging. Therefore, potential benefits and costs of testing for men under 70 years old with DXA should be investigated. Still no clinical trial is available yet evaluating the effectiveness of screening osteoporosis or any potential risks that can result from screening.[12] In false negative cases, diagnosis of osteoporosis can be overlooked and further treatment can be delayed. Conversely, false positive cases can lead to unnecessary DXA tests, unnecessary exposure to radiation, and increased health care costs. On the other hand, even if participants who underwent DXA test did not have osteoporosis (false positive cases), the result of the DXA test will still check the status of their bone health and help to estimate a proper interval for BMD testing. According to a recent study in white postmenopausal women 65 years of age or older, the estimated BMD testing interval for 10% of participants to develop osteoporosis was over 15 years in women with normal BMD or mild osteopenia (*T*-scores ≥ -1.5).[46] However, it would be shortened to 5 years in women with moderate osteopenia, BMD *T*-scores between -1.5 and -1.99 , and to even 1 year in women with

severe osteopenia, BMD *T*-scores between −2.0 and −2.49.[46]

This study has several limitations. First, the cost-effectiveness of KORAM is not considered in the present study. Although more than 90% of people with osteoporosis could be detected by KORAM, approximately 50% of could be underwent unnecessary DXA test. Therefore, to use of KORAM in clinical practice, its potential benefits and costs should be further evaluated. Second, KORAM is limited to estimating clinically important disease because KNHANES is a cross-sectional study. Considering that the ultimate goal for improving bone health is to prevent osteoporotic fracture, BMD scores provides only a marginal benefit to predict osteoporotic fracture.[38,47] Thus, well-designed cohort studies for estimating risk of osteoporotic fracture should be needed in Korean population. Third, to calculate sex-specific *T*-scores at femoral neck and lumbar spine, site and sex-specific referent means and standard deviations were adopted from the Japanese data because currently no Korean referent data is available. Thus, KORAM should be further adjusted when Korean referent BMD data is available.

V. CONCLUSION

To our knowledge, this is the first introduction of osteoporosis risk-assessment model to be developed for Korean postmenopausal women and men using a nationally representative dataset that includes BMD measurements and other relevant risk factors of osteoporosis. This study suggests that KORAM is a useful pre-screening tool for screening osteoporosis by DXA in the Korean population. Since KORAM is easy to calculate with simple variables, it can be used in either a primary care setting or general use as a self-screening tool. However, prior to using KORAM in these setting, its cost-effectiveness, especially compared to the current NHIC guidelines, should be investigated. In addition, replication studies using other Korean BMD datasets are recommended. Finally, further adjustment of KORAM using Korean referent BMD data is necessary.

REFERENCES

1. World Health Organization. Prevention and management of osteoporosis. Switzerland: WHO Press, 2003.

2. Who are candidates for prevention and treatment for osteoporosis? Osteoporos Int 1997; 7:1-6.

3. Johnell O, Kanis JA. An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. Osteoporos Int 2006; 17:1726-33.

4. Gullberg B, Johnell O, Kanis JA. World-wide projections for hip fracture. Osteoporos Int 1997; 7:407-13.

5. Choi HJ, Shin CS, Ha Y, Jang S, Park CM, Yoon H, et al. Burden of osteoporosis in adults in Korea: a national health insurance database study. J Bone Miner Metab 2012; 30: 54-8.

6. MOHW and KCDC. Korea Health Statistics 2009: Korean National Health and Nutrition Examination Survey (KNHANES IV-3). Osong, Republic of Korea: Ministry of Health and Social Welfare and Korea Centers for Disease Control & Prevention, 2010.

7. Kiebzak GM, Beinart GA, Perser K, Ambrose CG, Siff SJ, Heggeness MH. Undertreatment of osteoporosis in men with hip fracture. Arch Intern Med 2002; 162:2217-22.

8. Haney EM, Bliziotes MM. Male osteoporosis: new insights in an understudied disease. Curr Opin Rheumatol 2008; 20:423-8.

9. Seeman E. Unresolved issues in osteoporosis in men. Rev Endocr Metab Disord 2001; 2:45-64.

10. Chang KP, Center JR, Nguyen TV, Eisman JA. Incidence of hip and other osteoporotic fractures in elderly men and women: Dubbo Osteoporosis Epidemiology Study. J Bone Miner Res 2004; 19:532-6.

11. Krieg MA, Barkmann R, Gonnelli S, Stewart A, Bauer DC, Del Rio Barquero L, et al. Quantitative ultrasound in the management of osteoporosis: the 2007 ISCD Official Positions. J Clin Densitom 2008; 11:163-87.

12. Nelson HD, Haney EM, Dana T, Bougatsos C, Chou R. Screening for osteoporosis: an update for the U.S. Preventive Services Task Force. Ann Intern Med 2010; 153: 99-111.

13. Lim LS, Hoeksema LJ, Sherin K. Screening for osteoporosis in the adult U.S. population: ACPM position statement on preventive practice. Am J Prev Med 2009; 36:366-75.

14. Michalsson K, Bergstrm R, Mallmin H, Holmberg L, Wolk A, Ljunghall S. Screening for osteopenia and osteoporosis: selection by body composition. Osteoporos Int 1996; 6: 120-6.

15. Lydick E, Cook K, Turpin J, Melton M, Stine R, Byrnes C. Development and validation of a simple questionnaire to facilitate identification of women likely to have low bone density. Am J Manag Care 1998; 4: 37-48.

16. Black DM, Palermo L, Pearson J, Abbott T, Johnell O. SOFSURF: A simple, useful risk factor system can identify the large majority of women with

osteoporosis [abstract]. Bone 1998; 23: S605.

17. Cadarette SM, Jaglal SB, Kreiger N, McIsaac WJ, Darlington GA, Tu JV. Development and validation of the Osteoporosis Risk Assessment Instrument to facilitate selection of women for bone densitometry CMAJ 2000; 162:1289-94.

18. Sedrine WB, Chevallier T, Zegels B, Kvasz A, Micheletti MC, Gelas B, et al. Development and assessment of the Osteoporosis Index of Risk (OSIRIS) to facilitate selection of women for bone densitometry. Gynecol Endocrinol 2002; 16: 245-50.

19. Salaffi F, Silveri F, Stancati A, Grassi W. Development and validation of the osteoporosis prescreening risk assessment (OPERA) tool to facilitate identification of women likely to have low bone density. Clin rheumatol 2005; 24: 203-11.

20. Richy F, Deceulaer F, Ethgen O, Bruyre O, Reginster J. Development and validation of the ORACLE score to predict risk of osteoporosis. Mayo Clin Proc 2004; 79: 1402-8.

21. Weinstein L, Ullery B. Identification of at-risk women for osteoporosis screening. Am J Obstet Gynecol 2000; 183: 547-9.

22. Koh LK, Sedrine WB, Torralba TP, Kung A, Fujiwara S, Chan SP, et al. A simple tool to identify asian women at increased risk of osteoporosis. Osteoporos Int 2001; 12: 699-705.

23. Park HM, Ben Sedrine W, Reginster JY, Ross PD. Korean experience with OSTA risk index for osteoporosis: a validation study. J Clin Densitom 2003;

6:247-50.

24. Li-Yu JT, Llamado LJ, Torralba TP. Validation of OSTA among Filipinos. Osteoporos Int 2005; 16: 1789-93.

25. Rud B, Hilden J, Hyldstrup L, Hróbjartsson A. The Osteoporosis Self-Assessment Tool versus alternative tests for selecting postmenopausal women for bone mineral density assessment: a comparative systematic review of accuracy. Osteoporos Int 2009; 20: 599-607.

26. Wallace LS, Ballard JE, Holiday D, Turner LW, Keenum AJ, Pearman CM. Evaluation of decision rules for identifying low bone density in postmenopausal African-American women. J Natl Med Assoc 2004; 96: 290-6.

27. Lynn HS, Lau EM, Wong SY, Hong AW. An osteoporosis screening tool for Chinese men. Osteoporos Int 2005; 16: 829-34.

28. Zimering MB, Shin J, Shah J, Wininger E, Engelhart C. Validation of a novel risk estimation tool for predicting low bone density in Caucasian and African American men veterans. J Clin Densitom 2007; 10: 289-97.

29. Shepard AJ, Cass AR, Carlson CA, Ray L. Development and internal validation of the Male Osteoporosis Risk Estimation Score. Ann Fam Med 2007; 5:540-6.

30. Adler RA, Tran MT, Petkov VI. Performance of the Osteoporosis Self-assessment Screening Tool for osteoporosis in American men. Mayo Clin Proc 2003; 78: 723-7.

31. Sinnott B, Kukreja S, Barengolts E. Utility of screening tools for the

prediction of low bone mass in African American men. Osteoporos Int 2006; 17: 684-92.

32. Lynn HS, Woo J, Leung PC, Barrett-Connor EL, Nevitt MC, Cauley JA, et al.; Osteoporotic Fractures in Men (MrOS) Study. An evaluation of osteoporosis screening tools for the osteoporotic fractures in men (MrOS) study. Osteoporos Int 2008; 19: 1087-92.

33. Lee CH, Oh SW, Seung JH, Do HJ, Lym YL, Choi JK, et al. Evaluation of osteoporosis self-assessment too usefulness as a screening test for osteoporosis in Korean men. Korean J Fam Med. 2009; 30: 944-50.

34. MOHW and KCDC. Korea Health Statistics 2010: Korean National Health and Nutrition Examination Survey (KNHANES V-1). Osong, Republic of Korea: Ministry of Health and Social Welfare and Korea Centers for Disease Control & Prevention, 2012.

35. Choi HS, Oh HJ, Choi H, Choi WH, Kim JG, Kim KM et al. Vitamin D insufficiency in Korea−a greater threat to younger generation: the Korea National Health and Nutrition Examination Survey (KNHANES) 2008. J Clin Endocrinol Metab 2011; 96:643-51.

36. Hong S, Oh HJ, Choi H, Kim JG, Lim SK, Kim EK, et al. Characteristics of body fat, body fat percentage and other body composition for Koreans from KNHANES IV. J Korean Med Sci 2011; 26:1599-605.

37. KCDC. Education and quality control of bone mineral density in the Fourth National Health and Nutrition Examination Survey (KNHANES IV-3), Osong,

Republic of Korea: Korea Centers for Disease Control & Prevention, 2009.

38. Robbins J, Aragaki AK, Kooperberg C, Watts N, Wactawski-Wende J, Jackson RD, et al. Factors associated with 5-year risk of hip fracture in postmenopausal women. JAMA 2007; 298:2389-98.

39. Hosmer D, Lemeshow S. Applied logistic regression. 2nd ed. New York (NY): Wiley; 2000.

40. Pencina MJ, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Statistics in medicine. 2008; 27: 157-72; discussion 207.

41. Maddalozzo GF, Snow CM. High intensity resistance training: effects on bone in older men and women. Calcif Tissue Int 2000; 66:399-404.

42. Forwood MR, Burr DB. Physical activity and bone mass: exercises in futility? Bone Miner 1993; 21:89-112.

43. Holick MF. Sunlight and vitamin D for bone health and prevention of autoimmune diseases, cancers, and cardiovascular disease. Am J Clin Nutr 2004; 80:1678S-88S.

44. van Schoor NM, Visser M, Pluijm SM, Kuchuk N, smit JH, Lips P. Vitamin D deficiency as a risk factor for osteoporotic fractures. Bone 2008; 42:260-6.

45. Ross PD, Kress BC, Parson RE, Wasnich RD, Armour KA, Mizrahi IA. Serum bone alkaline phosphatase and calcaneus bone density predict fractures: a prospective study. Osteoporosis Int 2000; 11:76-82.

46. Gourlay ML, Preisser JS, Lui L, Cauley JA, Ensrud, KE; Study of

Osteoporosis Fractures Research Group. BMD screening in older women: initial measurement and testing interval. J Bone Miner Res 2012; 27: 743-746.

47. Leslie WD, Morin S, Lix LM, Johansson H, Oden A, McCloskey E, et al. Manitoba Bone Density Program. Fracture risk assessment without bone density measurement in routine clinical practice. Osteoporos Int 2012; 23: 75-85.

ABSTRACT (IN KOREAN)

한국인의 골다공증 위험 예측 모델 개발

<지도교수 김현창>

연세대학교 대학원 의학과

오 선 민

서론: 골다공증의 진단은 dual energy X-ray absorptiometry (DXA)로 측정한 골밀도를 기준으로 하지만, 직간접 비용이 높기 때문에 일반 인구 전체를 대상으로 하는 선별검사로는 추천되지 않는다. 따라서 간단한 변수들을 이용해 추가비용 없이 골다공증 고위험군을 선별할 수 있는 위험예측모델들이 여러 나라에서 개발되었는데 한국인을 대상으로는 그 연구가 미미하다. 따라서 이 연구에서는 한국인을 대표할 수 있는 자료를 이용하여 한국인의 골다공증 위험예측모델을 개발하고 검증하고자 하였다.

방법: 2009년 국민건강영양조사 참여자 중 골밀도 및 혈액측정 자료가 없거나, 골다공증을 진단받았거나 치료 중인 경우, 대부분 시간을 침대에 누워서 보내는 경우를 제외한 폐경후 여성 1,209명과 50세 이상 남성 1,340명을 대상으로 한국인의 골다공증 위험 예측 모델 The Korean Osteoporosis Risk-Assessment Model for postmenopausal women (KORAM-F) and men (KORAM-M)을 개발하였다. 개발된 모델은 2010년 국민건강영양조사 참여자 중 같은 기준으로 선정한 여성 1,046명과 남성 1,110명의 자료에서 검증하였다. 골다공증은 허리뼈 또는 넓적

다리뼈 골밀도 *T*-score −2.5 이하로 정의하였다. 모델에 사용한 변수는 10-fold cross validation 방법으로 선정하였다. 골다공증을 예측하는데 있어 모델들 간의 차이는 민감도, 특이도, area under the receive operating characteristics curve (AUC)로 비교하였다. 모델 적합도는 Hosmer-Leme show goodness-of-fit test로 확인하였다. KORAM의 위험도 분류 기준과 기존 모델 Osteoporosis Self assessment Tool for Asian (OSTA)의 위험도 분류 기준의 정확도를 비교하기 위해 Net Re-classification Improvement (NRI)를 계산하였다.

결과: 모델을 개발한 2009년 데이터에서 여자의 33.9%, 남자의 8.1%가 골다공증 진단 기준에 부합하였다. KORAM-F는 나이, 체중, 호르몬 대체요법 유무로 계산되는데, 점수 −9를 기준으로 골다공증을 선별하는 민감도는 91.2%, 특이도는 50.6%, AUC는 0.709였다. KORAM-M는 나이와 체중으로 계산되며, 같은 점수를 기준으로 민감도는 90.8%, 특이도는 42.4%, AUC는 0.666로 조사되었다. 또한, 기존의 OSTA의 위험도 범주를 KORAM의 위험도 범주로 재분류하면 NRI는 여자에서 7.4% ~ 41.8%, 남자에서 22.8% 정도까지 향상되는 것으로 조사되었다.

고찰: KORAM은 한국인에서 골다공증 위험을 선별하는 유용한 방법으로 보인다. 추후 KORAM을 임상 진료에서 사용하기 위해서는 비용-효과 분석이 뒷받침되어야 하며, 국민건강영양조사 이외의 다른 자료에서도 KORAM이 유용한지 평가해 볼 필요가 있다.

핵심되는 말: 골다공증, 위험도 예측, 폐경후 여성, 남성