

Clinical performance of seven prescreening tools for osteoporosis in Iranian postmenopausal women

Abbas Keshtkar¹ · Omidreza Tabatabaie^{1,2} · Nassim Matin^{1,2} · Zahra Mohammadi¹ · Mehdi Ebrahimi¹ · Patricia Khashayar¹ · Mojgan Asadi¹

Received: 8 February 2015 / Accepted: 8 May 2015 / Published online: 17 May 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract This study was designed to evaluate seven prescreening osteoporosis models in postmenopausal Iranian women. This study was performed on 8644 postmenopausal women who have been referred for bone mineral densitometry (BMD) in BMD center of Shariati hospital in Tehran between 2001 and 2011. Diagnostic properties of seven prescreening instruments were evaluated. With regard to area under curve (AUC), these models have low accuracy ($AUC \leq 0.65$). Considering only femoral neck or total femur area, these models had low accuracy but were more sensitive. Except for three models with sensitivities of $\leq 65\%$ in both osteoporosis and fracture threshold, other models were around 70 % sensitive. However, these models were not considered clinically useful regarding

their positive predictive values (PPV) especially in BMDs ≤ 0.25 . With regard to different measures of diagnostic property, none of these models were good screening tools for osteoporosis or fracture threshold. Although some of them are sensitive, considering other measures such as PPV indicates that they are not completely useful for clinical use. Attempts should be made for developing newer prescreening methods and calibration of the existing models with regard to the studied population.

Keywords Osteoporosis · Prescreening tools · Model performance · Diagnostic property

Introduction

Osteoporotic fractures are serious complications of osteoporosis. Identifying those in risk of osteoporotic fractures is an important preventive strategy, for a better evaluation and on time use of prophylactic treatments. Dual X-ray absorptiometry (DXA) is the gold standard for identifying osteoporotic postmenopausal women [1, 2]. Since measuring bone mineral density (BMD) in all postmenopausal women is not a practical approach and DXA devices are not available in all regions of a country, prescreening tools should be first used to help identify high-risk patients. A number of different risk assessment tools and models have been defined and used to evaluate postmenopausal women. These scoring models are used in primary care centers for referring patients to a secondary care center or a BMD center.

Three regions are used for assessing osteoporosis: the femoral neck, the lumbar spine and total femur [3]. Osteoporosis is defined as a BMD of at least 2.5 standard deviations below the mean T-score for young adults of the same

✉ Nassim Matin
nsmatn@yahoo.com

Abbas Keshtkar
abkeshtkar@yahoo.com

Omidreza Tabatabaie
omidrezatabatabaie@hotmail.com

Zahra Mohammadi
mimina_13@yahoo.com

Mehdi Ebrahimi
m.ebrahimi@tums.ac.ir

Patricia Khashayar
patricia.kh@gmail.com

Mojgan Asadi
asadim@tums.ac.ir

¹ Osteoporosis Research Center (ORC), Endocrinology and Metabolism Clinical Sciences Research Institute (ECSI), Tehran University of Medical Sciences, Tehran, Iran

² MD-MPH, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran

race and sex according to World Health Organization [4]. For each standard deviation decrease in bone densitometry, the risk of osteoporotic fracture increases 1.5- to threefold or even more. Appropriate cutoff values should be defined for an effective prevention and treatment of fractures [5]. According to a large cohort study, only 6.4 % of osteoporotic fractures occurred in T-scores of -2.5 and less. Also, this study indicated that 82 % of postmenopausal women with fractures had T-scores more than -2.5 . The Study of Osteoporotic Fracture also showed that 54 % of hip fractures and 77 % of other osteoporotic fractures occurred in a T-score >-2.5 [6]. According to National Osteoporosis Foundation guidelines, pharmacological interventions should be started in women with T-scores of -2 or less [7, 8]. This cutoff point (-2 and more standard deviations below the mean T-score for young adults) is called the fracture threshold and is used for referring patients for bone densitometry.

Several risk factors are involved in developing osteoporosis, which ultimately leads to varying performance of prescreening models in different populations. In order to validate the use of a prescreening model in a study population, the sensitivity, specificity, positive predictive value (PPV) and negative predictive values (NPV), along with receiver operating characteristic (ROC) curve analysis should be calculated. These models have been further evaluated and calibrated in populations other than the study populations they were developed from; however, data are sparse if any, for Iranian postmenopausal women.

In this study, we evaluated seven prescreening models in Iranian postmenopausal women in order to define the most appropriate model.

Methods

The present study was a cross-sectional study in which we analyzed data obtained from 8644 postmenopausal women who had performed DXA at BMD center of Shariati Hospital in Tehran, Iran. These postmenopausal women had been referred by their family medicines, endocrinologists and nephrologists, between year 2001 and 2011. The percent of female to male, and participants with secondary osteoporosis to primary osteoporosis, who were referred to this center, were consistent throughout these years. Secondary osteoporosis, mainly due to prolonged corticosteroid usage, was <10 % prevalent among the study members. The study is approved by the ethics committee of Endocrinology and Metabolism Research Institute affiliated to Tehran University of Medical Sciences.

After obtaining an informed consent, participants were asked to fill a questionnaire. This questionnaire included demographic data as well as risk factors of

osteoporosis according to previous studies and previous suggested prescreening instruments. Bone densitometer measurements were taken using DXA machine (Lunar, 7164, GE, Madison, WI) at Shariati Hospital BMD center. We used osteoporosis definition (a BMD of at least 2.5 standard deviations below the mean for young adults of the same race and sex (T-score)) as the first outcome of interest. In order to start prophylactic maneuvers, the fracture threshold was also used. The fracture threshold is defined as a BMD of at least 2 standard deviations below the mean for young adults. (T-score ≤ -2) BMD results were assessed in three regions: the femoral neck, the total femur and the lumbar areas (L2-L4), selected based on previous studies [3, 9–11]. Diagnosis of osteoporosis was made based on T-score ≤ -2.5 in at least one of these regions. Those with T-score of ≤ -2 should be referred for DXA which was another outcome of interest in this study.

Prescreening models

As mentioned above, the clinical performance of seven commonly used prescreening models is compared using a large cohort of postmenopausal women in this study. These models include: osteoporosis self-assessment tool (OST), osteoporosis risk assessment instrument (ORAI), simple calculated osteoporosis risk estimation (SCORE), age, body weight and no estrogen (ABONE), osteoporosis index of risk (OSIRIS), study of osteoporosis fracture–study utilizing risk factors, (SOF SURF) and national osteoporosis foundation (NOF). The characteristics of these models are described below and shown in details in Table 1.

The osteoporosis self-assessment tool (OST) was developed and validated in 1998. As one of the simple methods for evaluating the high-risk women for osteoporosis, it is calibrated for Asian, Caucasians, men and women. The score is calculated simply by subtracting age from weight divided by 5 and was designed in women for predicting femoral neck T-score of less than (and equal to) -2.5 . Based on this model, women with a score of ≤ 2 were considered to have osteoporosis in femoral neck area. OST was developed based on data of eight Asian countries [12, 13]. It was later validated for Caucasian European and US postmenopausal women [14]. A sensitivity of 91 %, specificity of 45 % and AUC of 0.79 was reported. Studies have reported sensitivity of 88–92 % in women and 82–85 % in men [15, 16].

Osteoporosis risk assessment instrument (ORAI) was developed and validated in 1376 women, from Canadian multicenter osteoporosis study (CAMOS) in 2000. Age, weight and use of hormone replacement therapy (HRT)

Table 1 Characteristics of the seven prescreening models that have been developed, evaluated and used in different populations

Prescreening tool	Objective	Number and characteristics of participants	Cutoff point	Predictors	Scoring formula	AUC (sensitivity–specificity)	Reference, year of study
ABONE	Spine, hip and femoral neck T-score ≤ -2.5	1610; White American postmenopausal women (mean age 62) having their first DEXA scan, was later validated by a Canadian database	≥ 2	Age, weight, estrogen use	Age > 65 years: 1 Weight < 63.5 kg: 1 Never used oral contraceptives or HRT more than 6 months: 1	0.72 (83–47 %)	[19, 20] 2000
ORAI	Femoral neck or lumbar spine T-score ≤ -2.0	1376; baseline data of the Canadian Multicentre Osteoporosis Study (CaMos), patients older than 45 years from 3 center	≥ 9	Age, weight, current estrogen use	Age ≥ 75 years: 15 Age 65–74 years: 9 Age 55–64 years: 5 Weight < 60 kg: 9 Weight 60–69 kg: 3 Not HRT user: 2	0.79 (91–46 %)	[17] 2000
OSTA	Asian women femoral neck T-score ≤ -2.5	860; developed in postmenopausal women of 8 Asian countries and later validated for Caucasian European and US postmenopausal women	< 2	Age, weight	0.2[weight (KG)–age (years)]	0.79 (91–45 %)	[12–16] 2001
SCORE	femoral neck T-score ≤ -2.0	1279; American postmenopausal women from 106 participating multispecialty centers	≥ 6	Age, weight, race, rheumatoid arthritis, estrogen use, fracture at age > 46 year	Non-black race: 5 Rheumatoid arthritis: 4, previous rib, wrist, or hip Fracture: 4 for each Never used HRT: 1 $3 \times \lfloor \text{age}/10 \rfloor$, truncated to integer $-1 \times \lfloor \text{weight in lb}/10 \rfloor$, truncated to integer	0.81 (89–50 %)	[18] 1998
OSIRIS	Lumbar spine, total hip and femoral neck ≤ -2.5	1303; retrospective study of postmenopausal women in Belgium referred for DEXA	< 1	Age, weight, current estrogen use, previous fracture	Age $\times (-2)$ and remove last digit Weight $\times 2$ and remove last digit If current HRT user: 2 If previous low impact fracture: -2	0.63–0.80 (78–51 %)	[21] 2002

Table 1 continued

Prescreening tool	Objective	Number and characteristics of participants	Cutoff point	Predictors	Scoring formula	AUC (sensitivity–specificity)	Reference, year of study
SOF SURF	Total hip T-score ≤ -2.0	10,000; the study of Osteoporotic fractures cohort was used for developing this model, the original study is available only in abstract form	≥ 3	Age, weight, smoking status, previous postmenopausal fracture	Weight 130–150 lb: 1 Weight <130 lb: 3 Current smoker: 1 Fracture after age 50: 1 0.2 points for each year above Age 65	0.75 Not reported	[22] 1998
NOF	T-score ≤ -2.0	2365; (Used for validation, obtained from the CAMOS database)	≥ 1	Age; weight; previous fracture at age >40 year; current smoker; parent had hip, wrist, or spine fracture age at ≥ 50 year	Age ≥ 65 year: 1 weight <57.6 kg: 1 Personal history of minimal trauma fracture >40 year: 1 Family history of fracture: 1 Current cigarette smoking: 1	0.60–0.70 (96–18 %)	[19, 23] 1999

Definitions, number of participants, scoring formula, cutoff point and performance indices are obtained from original studies. Reference numbers of the original studies are also shown in the last column

were the predictors in this model. BMD score at femoral neck or the lumbar spine was the objectives of this study. Based on ORAI, a score of ≥ 9 should be referred for further BMD study. This study also indicated that ORAI has a sensitivity of 94 % and a specificity of 41 % [17].

The simple calculated osteoporosis risk estimation (SCORE) score was developed in a study investigating 1279 American women in 1996. Age, weight and use of HRT were predictors of this model. Rheumatoid arthritis, race and personal history of fracture were also included. Osteoporosis was defined using a T-score of ≤ 2 . This study had a sensitivity of 93 % and a specificity of 43 % using a score of ≥ 6 for osteoporosis [18].

The ABONE model, using age, body weight and no estrogen as predictors, was developed in 1610 white women who had DEXA scan for the first time. This model had AUC of 0.72 when validated on 2365 women from the CAMOS database [19, 20].

The osteoporosis index of risk (OSIRIS) was developed using a retrospective database of 1303 postmenopausal women in 1999. Age, body weight, current HRT and history of previous low impact fracture were the predictor variables. AUC was 0.71. OSIRIS score of ≤ 1 had a sensitivity of 85 % and specificity of 39 % [21].

Another prescreening tool evaluated in this study is the study of osteoporosis fracture–study utilizing risk factors (SOF SURF) that has only been published as abstract. Age, weight, smoking status and previous postmenopausal fracture were identified as predictors of low bone mineral density. They defined the total hip score of ≤ -2.5 in BMD as osteoporosis. Based on this model, individuals with a score of ≥ 3 should be referred for DXA. The sensitivity and specificity of this model was not reported; however, it had AUC of 0.75 [22].

Cadarette et al. defined National Osteoporosis Foundation (NOF) score, using age, body weight, personal history of fracture after minimal trauma, family history of fracture and cigarette smoking in 1999 to incorporate screening guidelines recommended by NOF. The score was validated it on CAMOS database. Based on this model, a woman with NOF score of ≥ 1 should be referred for further imaging [19, 23].

These seven models are used to determine whether a patient has osteoporosis or not. Another tool, the Fracture Risk Assessment Tool (FRAX), developed by World Health Organization (WHO) task forces is also developed in 2008 to predict the ten year risk of fracture. FRAX integrates the clinical risk factors with BMD at femoral neck for this purpose and is derived from patient populations in North America, Latin America, Europe, Asia and Australia. Due to different outcome of interest, evaluating the performance of this model in Iranian population was not an objective of this study [24, 25].

Table 2 Participants are classified based on different important T-scores in femoral neck, total femur and lumbar areas

Data	L2–L4 number (%)	Femoral neck number (%)	Total femur number (%)	Either one of the regions number (%)
T-score > -1	2369 (28.68)	2937 (35.65)	3786 (45.96)	1540 (17.81) ^a
-2 < T-score ≤ -1	2255 (27.30)	2849 (34.58)	2575 (31.25)	2467 (28.54) ^b
-2.5 < T-score ≤ -2	1161 (14.06)	1113 (13.51)	878 (10.66)	1353 (15.65) ^c
T-score ≤ -2.5	2475 (29.96)	1340 (16.26)	1000 (12.14)	2917 (33.74) ^d

The last column represents those with at least one area meeting the criterion

^a This number represents the number of patients with no region with T-score of < -1

^b This number represents the number of patients with no region with T-score of < -2. Those with T-scores of more than -1 are also excluded. This number represents osteopenic participants without need to be referred for DXA

^c This number represents patients with at least one area with a T-score of < -2 and thus should be referred for DXA

^d This number represents patients with at least one area with a T-score of < -2.5 who have osteoporosis

Statistical analysis

Data are described and classified based on different cutoff points (-1, -2 and -2.5) for BMD scores in three regions. Predictors of different models were classified, and the scoring system of each model was formulated. Sensitivity, specificity, PPV, NPV and AUC are then calculated for each model. Data were analyzed using STATA 11.1.

Results

Participants of this study were postmenopausal women between 23 and 94 years of age (mean age 59 ± 8.7 years). Patients were 18–69 years at menopause (mean 47 ± 6.4). Body mass index was between 14 and 55.24 (mean 27.91 ± 0.06). Number and percent of patients with and without osteoporosis is shown in Table 2. Three regions of interest were first analyzed separately. Table 3 presents the performance of models based on the BMD T-score of ≤ -2.5 (osteoporosis) in the three regions of interest. For a better assessment of these models and the practical performance, these prescreening tools should also be tested using the fracture (referral) threshold (A BMD T-score of -2 and more standard deviations below the mean T-score for young adults). Results are shown in detail in Table 4. Tables 5 and 6 are allocated to the performance of models based on the three regions of interest together in both the osteoporosis threshold and the fracture threshold. The interpretation of these results is discussed below.

Discussion

Since mid-1990s, a number of studies were conducted to develop models that could help identify postmenopausal women at risk for osteoporosis or low bone mass who would then be referred for a DEXA scan for confirmation.

These prescreening tools were later evaluated and compared in different populations to further select the most appropriate model. In this study, we applied seven prescreening models in a large sample of Iranian postmenopausal women to evaluate their performance and choose the most appropriate tool for that population.

Applying these tools in different populations has yielded different performance results. Some studies showed high sensitivities and good performance, while some others indicated lower indices than what was reported in earlier studies. In 2001, Cadarette et al. evaluated ABONE, ORAI, weight criterion (body weight <70 kg), SCORE and NOF models. The CaMos database (Canadian women) was used for this study, and these prescreening tools were applied to the data from 2365 postmenopausal women. Based on their evaluation, SCORE and ORAI were more appropriate (with sensitivities of 99 and 97 %, respectively) for detecting osteoporosis [19]. One study evaluated the use of ABONE, ORAI, OST, SCORE and weight criterion in 174 African-American women and reported lower sensitivities from what was reported earlier, ranging from 65 % for ORAI to 83 % for SCORE [1]. Evaluating the performance of ORAI, OST, OSIRIS and weight criterion in 665 Spanish postmenopausal women showed that ORAI, OST, OSIRIS and weight criterion would correctly identify only 45, 46, 37 and 70 % of women who would benefit from doing a DEXA scan, respectively. Calculated sensitivities were 58, 64, 69 and 83 % for OSIRIS, ORAI, OST and weight, respectively [2]. In one study, ORAI and SCORE were compared among three different ethnicities; Hispanics, non-Hispanic whites and African-American women. An overall sensitivity of 66 and 68 % was reported for SCORE and ORAI, respectively. Moreover, SCORE identified only 30 % of women who should have been referred for BMD [3]. A previous study of 5573 Iranian postmenopausal women showed a sensitivity of 71 % for OST and 74 % for both ORAI and weight criterion [9]. Another study with 341 Iranian postmenopausal women reported

Table 3 Different models are evaluated in femoral neck, total femur and the lumbar areas

Model	Sensitivity 95 % CI	Specificity 95 % CI	Positive predictive value	Negative predictive value	Correctly clas- sifies (%)	Area under curve
OSTA						
Femoral neck	87.2 % (85.3–89)	52.4 % (51.2–53.6)	26.1 % (24.8–27.5)	95.5 % (94.8–96.2)	58.06	0.698 (0.68–0.70)
Lumbar	75.2 % (73.4–76.9)	55.1 % (53.8–56.4)	41.8 % (40.3–43.3)	83.8 % (82.6–85)	61.11	0.651 (0.64–0.66)
Total femur	89 % (86.9–90.9)	50.9 % (49.7–52.1)	20.1 % (19–21.4)	97.1 % (96.5–97.6)	55.56	0.7 (0.68–0.71)
ABONE						
Femoral neck	84 % (81.9–85.9)	44.4 % (43.2–45.6)	22.7 % (21.5–23.9)	93.4 % (92.5–94.3)	50.81	0.642 (0.63–0.65)
Lumbar	73.5 % (71.7–75.2)	45.4 % (44.1–46.7)	36.5 % (35.2–37.9)	80 % (78.6–81.6)	53.81	0.594 (0.58–0.60)
Total femur	87.5 % (85.3–89.5)	43.5 % (42.4–44.7)	17.6 % (16.6–18.7)	96.2 % (95.5–96.8)	48.85	0.655 (0.64–0.66)
ORAI						
Femoral neck	87.8 % (85.9–89.6)	51.1 % (49.9–52.3)	25.7 % (24.4–27.1)	95.6 % (94.9–96.3)	57.02	0.695 (0.68–0.70)
Lumbar	75.3 % (73.6–77.1)	53.4 % (52.1–54.7)	41 % (39.5–42.4)	83.5 % (82.2–84.7)	60.01	0.644 (0.63–0.65)
Total femur	89.6 % (87.5–91.5)	49.6 % (48.4–50.7)	19.8 % (18.7–21)	97.2 % (96.6–97.7)	54.46	0.696 (0.68–0.70)
SCORE						
Femoral neck	85.4 % (83.4–87.3)	52 % (50.8–53.2)	25.6 % (24.3–26.9)	94.9 % (94.1–95.6)	57.39	0.687 (0.67–0.69)
Lumbar	73.4 % (71.6–75.1)	54.1 % (52.8–55.4)	40.7 % (39.2–42.2)	82.6 % (81.1–83.8)	59.88	0.637 (0.62–0.64)
Total femur	87.9 % (85.7–89.9)	50.6 % (49.4–51.8)	19.8 % (18.7–21.1)	96.8 % (96.2–97.3)	55.17	0.692 (0.68–0.70)
SOFSURF						
Femoral neck	56.9 % (54.1–59.6)	80.5 % (79.5–81.4)	36 % (33.9–38.1)	90.6 % (89.9–91.4)	76.65	0.687 (0.67–0.70)
Lumbar	44 % (42–46)	82.2 % (81.2–83.2)	51.5 % (49.3–53.7)	77.4 % (76.3–78.5)	70.76	0.631 (0.62–0.642)
Total femur	62.9 % (59.8–66)	79.6 % (78.7–80.6)	30.1 % (28.1–32.1)	93.9 % (93.3–94.5)	77.59	0.713 (0.69–0.72)
OSIRIS						
Femoral neck	72.1 % (69.6–74.5)	72 % (70.9–73.1)	33.2 % (31.5–35)	93 % (92.3–93.7)	72.03	0.721 (0.70–0.73)
Lumbar	56.9 % (54.9–58.9)	74.2 % (73–75.3)	48.7 % (46.7–50.4)	80 % (78.9–81.1)	68.98	0.655 (0.64–0.66)
Total femur	76.9 % (74.1–79.5)	70.7 % (69.6–71.8)	26.7 % (25.1–28.4)	95.7 % (95.1–96.2)	71.44	0.738 (0.72–0.75)
NOF						
Femoral neck	77.2 % (74.8–79.5)	57.7 % (56.5–58.9)	26.1 % (24.7–27.5)	92.6 % (92.1–93.7)	60.86	0.675 (0.66–0.68)
Lumbar	63.2 % (61.2–65.1)	58.6 % (57.3–59.9)	39.6 % (38–41.1)	78.8 % (77.5–80)	59.97	0.609 (0.59–0.62)
Total femur	81.9 % (79.4–84.3)	56.8 % (55.6–58)	20.9 % (19.6–22.2)	95.8 % (95.1–96.4)	59.86	0.694 (0.68–0.70)

The outcome of interest is the T-score of -2.5 standard deviations and more below the mean T-score for young adults. (Osteoporosis)

Table 4 Different models are evaluated in femoral neck, total femur and the lumbar areas

Model	Sensitivity 95 %CI	Specificity 95 %CI	Positive predictive value	Negative predictive value	Correctly clas- sifies (%)	Area under curve
OSTA						
Femoral neck	81.4 % (79.8–82.9)	57.6 % (56.3–58.9)	44.9 % (43.4–46.4)	88 % (86.9–89)	64.71	0.695 (0.68–070)
Lumbar	69.3 % (67.7–70.8)	58.1 % (56.6/59.5)	56.7 % (55.2–58.1)	70.5 % (69–72)	63.04	0.647 (0.62–0.64)
Total femur	83.5 % (81.7–85.2)	54.7 % (53.5–56)	35.2 % (33.7–36.6)	91.9 % (90.9–92.7)	61.25	0.691 (0.68–0.70)
ABONE						
Femoral neck	78.6 % (76.9–80.2)	47.5 % (46.2–48.8)	38.8 % (37.5–40.2)	83.9 % (82.6–85.2)	56.77	0.695 (0.68–0.70)
Lumbar	69.2 % (67.7–70.7)	46.8 % (45.3–48.2)	50.6 % (49.2–52)	65.9 % (64.3–67.5)	56.66	0.637 (0.62–0.64)
Total femur	81 % (79.1–82.7)	45.9 % (44.6–47.1)	30.6 % (29.4–31.9)	89.1 % (88–90.1)	53.88 %	0.691 (0.68–0.70)
ORAI						
Femoral Neck	82.1 % (80.5–83.7)	56.2 % (54.9–57.5)	44.3 % (42.8–45.8)	88.1 % (87–89.2)	63.92	0.692 (0.68–0.7)
Lumbar	69.3 % (67.8–70.8)	56 % (54.5–57.4)	55.4 % (54–56.9)	69.8 % (68.2–71.3)	61.86	0.626 (0.61–0.63)
Total femur	83.7 % (81.9–85.4)	53.2 % (51.9–54.4)	34.5 % (33–35.9)	91.7 % (90.8–92.6)	60.11	0.684 (0.67–0.69)
SCORE						
Femoral neck	80.4 % (78.8–82)	57.1 % (55.8–58.4)	44.3 % (42.8–45.8)	87.3 % (86.2–88.4)	64.03	0.688 (0.67–0.69)
Lumbar	68.4 % (66.9–70)	57.1 % (55.1–58.6)	55.8 % (54.3–57.3)	69.6 % (68.1–71.1)	62.11	0.628 (0.61–0.63)
Total femur	82.6 % (80.7–84.3)	54.3 % (53–55.5)	34.7 % (33.3–36.1)	91.4 % (90.4–92.3)	60.71	0.684 (0.67–0.69)
SOF SURF						
Femoral neck	47.6 % (45.6–49.6)	83.8 % (82.8–84.7)	55.4 % (53.2–57.6)	79 % (78–80.1)	73	0.657 (0.646)
Lumbar	38 % (36.4–39.6)	84.1 % (83–85.2)	65.4 % (63.3–67.4)	63.2 % (61.9–64.4)	63.74	0.61 (0.60–0.62)
Total femur	52.3 % (49.9–54.6)	82.3 % (81.3–83.2)	46.4 % (44.3–48.6)	85.4 % (84.5–86.3)	75.46	0.673 (0.66–0.68)
OSIRIS						
Femoral neck	63.4 % (61.4–65.3)	76.9 % (75.7–78)	53.8 % (51.9–55.6)	83.2 % (82.2–84.2)	72.86	0.701 (0.69–0.71)
Lumbar	50.4 % (48.2–52)	76.8 % (75.6–78.1)	63.2 % (61.4–65)	66.2 % (64.9–67.5)	65.15	0.636 (0.62–0.64)
Total femur	67.9 % (65.7–70.1)	74.5 % (73.4–75.6)	43.9 % (42.1–45.8)	88.8 % (87.9–89.6)	73.02	0.712 (0.7–0.72)
NOF						
Femoral neck	70.3 % (68.4–72.1)	61.5 % (60.2–63.8)	43.6 % (42.1–45.2)	83 % (81.8–84.1)	64.12	0.66 (0.64–0.66)
Lumbar	58.7 % (57–60)	60.5 % (59.1–62)	54 % (52.4–55.6)	65 % (63.5–66.4)	59.72	0.596 (0.58–0.6)
Total femur	73.7 % (71.6–75.7)	59.6 % (58.4–60.9)	34.9 % (33.4–36.5)	88.5 % (87.5–89.5)	62.82	0.666 (0.65–0.67)

The outcome of interest is the T-score of -2 standard deviations and more below the mean T-score for young adults. (Should be referred for DXA)

Table 5 Performance of models at BMD cutoff score of -2.5 in three regions (the femoral neck, total femur and lumbar area)

Model At -2.5 T-score	Sensitivity	Specificity	PPV	NPV	Correctly classifies (%)	AUC
OSTA	76.2 % (74.6–77.7)	57.4 % (56–58.7)	47.8 % (46.3–49.3)	82.5 % (81.2–83.6)	63.73	0.668 (0.65–0.67)
ABONE	74.3 % (72.7–75.9)	47.1 % (45.8–48.4)	41.7 % (40.4–43.1)	78.3 % (76.9–79.6)	56.29	0.607 (0.59–0.61)
ORAI	76.4 % (74.8–77.9 %)	55.6 % (54.3–56.9)	46.8 % (45.4–48.3)	82.1 % (80.9–83.3)	62.64	0.66 (0.65–0.67)
SCORE	74.7 % (73–76.2)	56.3 % (55–57.7)	46.7 % (45.2–48.2)	81.3 % (80–82.5)	62.55	0.655 (0.64–0.66)
SOFSURF	44 % (42.4–45.9)	83.9 % (82.9–82.8)	58.3 % (56.2–60.4)	74.5 % (73.4–75.6)	70.39	0.64 (0.63–0.65)
OSIRIS	57.9 % (56.1–59.8)	76.4 % (75.2–77.5)	55.7 % (53.9–57.5)	78 % (76.9–79.1)	70.14	0.672 (0.66–0.68)
NOF	64.7 % (62.9–66.5)	60.4 % (59.1–61.7)	45.6 % (44–47.1)	77 % (75.7–78.2)	61.78	0.626 (0.61–0.63)

Osteoporosis diagnosis is made when at least one of these three regions has a T-score of <-2.5

Table 6 Performance of models at BMD cutoff score of -2 in three regions (the femoral neck, total femur and lumbar areas)

Model At -2 T-score	Sensitivity	Specificity	PPV	NPV	Correctly classifies (%)	AUC
OSTA	70 % (68.6–71.4)	61.8 % (60.3–63.2)	64.3 % (62.9–65.7)	67.7 % (62.9–69.2)	65.86	0.66 (0.64–0.67)
ABONE	70 % (68.6–71.4)	49.5 % (48.1–51)	57.5 % (56.2–58.9)	62.9 % (61.2–64.5)	59.66	0.598 (0.58–0.6)
ORAI	70.4 % (69–71.8)	59.7 % (58.2–61.2)	63.2 % (61.8–64.6)	67.2 % (65.7–68.7)	65	0.651 (0.64–0.66)
SCORE	69.5 % (68–70.9)	60.9 % (54.4–62.4)	63.6 % (62.2–65)	67 % (65.4–68.1)	65.16	0.652 (0.64–0.66)
SOFSURF	37.5 % (36.1–39)	86.2 % (85.1–87.2)	72.8 % (70.9–74.7)	58.4 % (57.1–59.6)	62.06	0.62 (0.61–0.63)
OSIRIS	50.8 % (49.3–52.3)	80.1 % (78.9–81.3)	71.5 % (69.9–73.2)	62.3 % (61–63.6)	65.57	0.654 (0.64–0.66)
NOF	59.7 % (58.2–61.2)	63.3 % (61.8–64.7)	61.5 % (60–63)	61.4 % (60–63)	61.48	0.615 (0.6–0.62)

This is the fracture threshold. Anyone with a T-score of <-2 in at least one of these regions is candidate for further evaluation with DEXA or pharmacological intervention

a sensitivity of only 48 % for NOF but 70 and 87 % for ORAI and SCORE, respectively [26].

Before interpreting the results of different studies and comparing them with our results, some points must be noted. First, in the studies mentioned above, sensitivity was addressed as the sole index of interest for evaluating the performance of prescreening models. Although sensitivity is an important index, clinicians do not approach their patients merely based on this index; rather, they would like to know the probability of disease in an individual with a positive test result or positive predictive value, which is itself a factor of prevalence. Another point of strength in a model is the diagnostic accuracy, a concept interpreted

based on the ROC analysis and AUC. In a model with AUC of <0.5 , the chance of correctly identifying a patient is even less than flipping a coin and thus considered useless. A good screening tool should have AUC of more than 0.7. Before comparing models regarding their sensitivities and specificities in different settings, AUCs should be reported and only models with moderate-to-high accuracy (AUC of more than 0.7) should be considered [27, 28]. In fact, in 2010, The US Preventive Services Task Force reviewed the performance of 14 prescreening instruments and indicated that they are only modest predictors of osteoporosis. The AUC of these models ranged from as low as 0.13 to as high as 0.87 [29].

Another important point for interpreting the results of different studies is the exact description of osteoporosis, both in studies that evaluate the performance of these tools and in the studies that developed them. For example, ORAI was developed to correctly classify women with T-score ≤ -2.0 in femoral neck or lumbar spine area, while OSTA considered a femoral neck T-score ≤ -2.5 as the objective of the study. While a T-score ≤ -2.5 is the threshold for defining osteoporosis based on WHO [5], some tools prefer to identify at-risk patients sooner and consider a T-score of -2 as the appropriate cutoff for a prescreening model. The National Osteoporosis Risk Assessment study in 149529 white postmenopausal women reported that 82 % of women with fractures had T-scores more than -2.5 . Also 67 % of them had a T-score > -2 [6]. Thus, a lower threshold provides a more secure and appropriate border for prophylactic strategies and helps assign screening intervals. ORAI, SCORE, SOFSURF and NOF have used -2 as the appropriate cutoff for detecting osteoporosis. Another important criterion is the skeletal sites of osteoporosis. Studies have shown that T-scores of the lumbar spine and total hip might be discordant in at least one diagnostic class in about 40 % patients and BMD of one anatomical location cannot predict the density in the sites [30, 31]. The International Society for Clinical Densitometry (ISCD) emphasizes that WHO classification should be considered in three regions in postmenopausal women: the lumbar spine, total hip and femoral neck. AT-score below the determined cutoff in any one of these sites is sufficient for the diagnosis [32]. This definition was only used in ABONE and OSIRIS with T-score < -2.5 as the cutoff.

With regard to the points above, considering -2.5 as the cutoff for T-score in three sites of interest, none of these models fulfill the criteria for an accurate model (AUC is somewhere between 0.6 and 0.7 for all of them as shown in Table 5). The least appropriate model based on AUC is ABONE, correctly classifying only 56 % of patients, and the less sensitive model is SOFSURF, only 44 % sensitive. OSIRIS, OSTA and ORAI are considered more appropriate models with an AUC of approximately 0.66. Although OSIRIS correctly classifies 70 % of patients, OSTA and ORAI are considered more suitable because of their higher sensitivity (76 % in OSTA and ORAI vs, 58 % in OSIRIS). When the data are analyzed in depth for different skeletal sites, OSIRIS (AUC ≥ 0.7) is considered as a good model in total femur and femoral neck areas and has sensitivity of about 70 % in these regions. For predicting osteoporosis at lumbar spine, OSTA has the highest AUC (65 %) and the highest sensitivity (75 %). Although OSTA, ORAI and SCORE are not good models regarding their AUCs, their sensitivities for the three regions are 75–90 % (shown in Table 3).

When considering the T-score of -2 and less in three skeletal sites, again none of the models has an AUC value > 0.7 ; however, OSTA, ORAI, SCORE and OSIRIS have values of around 0.65. Adding other indices, OSTA and ORAI are considered more appropriate and SOFSURF with AUC = 0.62 and a sensitivity of only 37 % is the least appropriate tool (Table 6). With regard to different sites of osteoporosis, OSIRIS has AUC of 0.7 in total femur and femoral neck with sensitivities of 67 and 63 %, respectively. However, OSTA, ORAI and SCORE with a slightly lower AUC values have sensitivities of about 82 % and thus are considered more appropriate. In terms of lumbar spine, all AUCs are < 0.65 and again none of them is appropriate. OSTA, ABONE and ORAI with AUC values of around 0.63 have sensitivity of 69 %. In the lumbar spine, SOFSURF has the lowest sensitivity and AUC (Table 4).

From a clinician's point of view, when PPV is considered, at T-score of -2.5 , although SOFSURF and OSIRIS have higher PPVs (and specificities), they are not appropriate for screening due to their low sensitivity. OSTA, followed closely by ORAI and SCORE, should be considered more appropriate with quite acceptable PPVs of around 47 % and sensitivities of around 75 %. At T-score of -2 , OSIRIS has the highest PPV (71 %) in the expense of the lowest sensitivity (50 %). After classifying models based on AUCs and sensitivities, OSTA and ORAI are considered the most appropriate models, correctly classifying around 65 % of patients, with sensitivities of 70 % and PPVs of 64 % (Tables 5, 6).

Combining all of the above criteria, although none of these prescreening tools is considered "good" due to the low AUC, OSTA and ORAI (followed closely by SCORE) are more appropriate than others. This statement should be interpreted with caution. The generalizability of these findings depends on the characteristics of the studied population; although a very large sample of Iranian postmenopausal women was included here, they might not be representative of all Iranian women. The database belongs to a tertiary center, an academic hospital in the capital of Iran. Also, we previously showed that based on ORAI and SCORE, BMD orders of insured Iranians are appropriate only half of the times. Sixty percentage of BMDs ordered based on OSTA and OSIRIS were inappropriate. This percent increased up to 67 and 90 % for NOF and ABONE, respectively, necessitating further analyses [33].

In conclusion, in order to accept the use of these models among Iranian postmenopausal women, first they should be evaluated in a national sample from different provinces and different BMD centers. If results are still unsatisfactory, either newer models should be designed and tested or preformed models should be calibrated.

Conflict of interest The authors declare that they have no conflict of interest.

References

- Wallace LS, Ballard JE, Holiday D, Turner LW, Keenum AJ, Pearman CM (2004) Evaluation of decision rules for identifying low bone density in postmenopausal African-American women. *J Natl Med Assoc* 96(3):290
- Martinez-Aguila D, Gomez-Vaquero C, Rozadilla A, Romera M, Narvaez J, Nolla JM (2007) Decision rules for selecting women for bone mineral density testing: application in postmenopausal women referred to a bone densitometry unit. *J Rheumatol* 34(6):1307–1312
- Cass AR, Shepherd AJ, Carlson CA (2006) Osteoporosis risk assessment and ethnicity: validation and comparison of 2 clinical risk stratification instruments. *J Gen Intern Med* 21(6):630–635
- Organization WH (1994) Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: report of a WHO study group [meeting held in Rome from 22 to 25 June 1992]
- Kanis JA, Melton LJ, Christiansen C, Johnston CC, Khaltaev N (1994) The diagnosis of osteoporosis. *J Bone Miner Res* 9(8):1137–1141
- Siris ES, Chen Y-T, Abbott TA, Barrett-Connor E, Miller PD, Wehren LE, Berger ML (2004) Bone mineral density thresholds for pharmacological intervention to prevent fractures. *Arch Intern Med* 164(10):1108–1112
- Cummings SR, Black DM, Thompson DE, Applegate WB, Barrett-Connor E, Musliner TA, Palermo L, Prineas R, Rubin SM, Scott JC (1998) Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures: results from the Fracture Intervention Trial. *JAMA* 280(24):2077–2082
- Black DM, Thompson DE, Bauer DC, Ensrud K, Musliner T, Hochberg MC, Nevitt MC, Suryawanshi S, Cummings SR (2000) Fracture risk reduction with alendronate in women with osteoporosis: the fracture intervention trial. *J Clin Endocrinol Metab* 85(11):4118–4124
- Dabbaghmanesh M, Sabet R, Aria A, Omrani GR (2007) Performance of osteoporosis risk assessment tools in Iranian postmenopausal women. *Int J Endocrinol Metab* 5(1):26–32
- Fujiwara S, Masunari N, Suzuki G, Ross PD (2001) Performance of osteoporosis risk indices in a Japanese population. *Curr Ther Res* 62(8):586–594
- Richy F, Gourlay M, Ross P, Sen S, Radican L, De Ceulaer F, Sedrine WB, Ethgen O, Bruyère O, Reginster J-Y (2004) Validation and comparative evaluation of the osteoporosis self-assessment tool (OST) in a Caucasian population from Belgium. *QJM* 97(1):39–46
- Skedros JG, Sybrowsky CL, Stoddard GJ (2007) The osteoporosis self-assessment screening tool: a useful tool for the orthopaedic surgeon. *J Bone Joint Surg* 89(4):765–772
- Koh L, Sedrine WB, Torralba T, Kung A, Fujiwara S, Chan S, Huang Q, Rajatanavin R, Tsai K, Park H (2001) A simple tool to identify Asian women at increased risk of osteoporosis. *Osteoporos Int* 12(8):699–705
- Rud B, Hilden J, Hyldstrup L, Hrobjartsson A (2007) Performance of the osteoporosis self-assessment tool in ruling out low bone mineral density in postmenopausal women: a systematic review. *Osteoporos Int* 18(9):1177–1187
- Gourlay M, Powers J, Lui L-Y, Ensrud K (2008) Clinical performance of osteoporosis risk assessment tools in women aged 67 years and older. *Osteoporos Int* 19(8):1175–1183
- Adler RA, Tran MT, Petkov VI (2003) Performance of the osteoporosis self-assessment screening tool for osteoporosis in American men. In: Mayo clinic proceedings (Vol. 6, pp 723–727). Elsevier
- Cadarette SM, Jaglal SB, Kreiger N, McIsaac WJ, Darlington GA, Tu JV (2000) Development and validation of the Osteoporosis Risk Assessment Instrument to facilitate selection of women for bone densitometry. *Can Med Assoc J* 162(9):1289–1294
- Lydick E, Cook K, Turpin J, Melton M, Stine R, Byrnes C (1998) Development and validation of a simple questionnaire to facilitate identification of women likely to have low bone density. *Am J Manag Care* 4(1):37–48
- Cadarette SM, Jaglal SB, Murray TM, McIsaac WJ, Joseph L, Brown JP (2001) Evaluation of decision rules for referring women for bone densitometry by dual-energy X-ray absorptiometry. *JAMA* 286(1):57–63
- Weinstein L, Ullery B (2000) Identification of at-risk women for osteoporosis screening. *Am J Obstet Gynecol* 183(3):547–549
- Sedrine WB, Chevallier T, Zegels B, Kvasz A, Micheletti M, Gelas B, Reginster J-Y (2002) Development and assessment of the Osteoporosis Index of Risk (OSIRIS) to facilitate selection of women for bone densitometry. *Gynecol Endocrinol* 16(3):245–250
- Black D, Palermo L, Abbott T, Johnell O (1998) SOFSURF: a simple, useful risk factor system can identify the large majority of women with osteoporosis. *Bone* 23(Suppl 5):S605
- Foundation NO (1999) Physician's guide to prevention and treatment of osteoporosis. Belle Mead NJ Excerpta Medica Inc
- Kanis J, McCloskey E, Johansson H, Oden A, Ström O, Borgström F (2010) Development and use of FRAX® in osteoporosis. *Osteoporos Int* 21(2):407–413
- Frax W (2008) Fracture risk assessment tool. World Health Organization
- Mottaghi P, Karimifar M, Salesi M, Mehrabi A (2010) Osteoporosis screening tools in Iranian postmenopausal women. *Iran Red Crescent Med J* 12(2):122–126
- Akobeng AK (2007) Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatr* 96(3):338–341
- Fischer JE, Bachmann LM, Jaeschke R (2003) A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med* 29(7):1043–1051
- Nelson HD, Haney EM, Dana T, Bougatsos C, Chou R (2010) Screening for osteoporosis: an update for the U.S. Preventive Services Task Force. *Ann Intern Med* 153(2):99–111. doi:10.7326/0003-4819-153-2-201007200-00262
- Mounach A, Abayi DM, Ghazi M, Ghozani I, Nouijai A, Achemlal L, Bezza A, El Maghraoui A (2009) Discordance between hip and spine bone mineral density measurement using DXA: prevalence and risk factors. In: Seminars in arthritis and rheumatism (Vol. 6, pp 467–471). Elsevier
- Looker AC, Statistics NCFH (2012) Osteoporosis or low bone mass at the femur neck or lumbar spine in older adults, United States, 2005–2008. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics Hyattsville, MD
- Hamdy RC, Petak SM, Lenchik L (2002) Which central dual X-ray absorptiometry skeletal sites and regions of interest should be used to determine the diagnosis of osteoporosis? *J Clin Densitom* 5(3):s11–s17
- Keshtkar A, Avestai A, Khashayar P, Heshmat R, Arabali A, Abdollahi N, Ebrahimi M, Larijani B (2011) Appropriateness of the BMD orders among insured iranians. In: Osteoporosis international, 2011. Springer London Ltd 236 Grays inn rd, 6th floor, London WC1X 8HL, England, pp S692–S693