

Clinical performance of osteoporosis risk assessment tools in women aged 67 years and older

M. L. Gourlay · J. M. Powers · L.-Y. Lui ·
K. E. Ensrud ·
for the Study of Osteoporotic Fractures Research Group

Received: 11 September 2007 / Accepted: 20 November 2007 / Published online: 25 January 2008
© International Osteoporosis Foundation and National Osteoporosis Foundation 2008

Abstract

Summary Clinical performance of osteoporosis risk assessment tools was studied in women aged 67 years and older. Weight was as accurate as two of the tools to detect low bone density. Discriminatory ability was slightly better for the OST risk tool, which is based only on age and weight.

Introduction Screening performance of osteoporosis risk assessment tools has not been tested in a large, population-based US cohort.

Methods We conducted a diagnostic accuracy analysis of the Osteoporosis Self-assessment Tool (OST), Osteoporosis Risk Assessment Instrument (ORAI), Simple Calculated Osteoporosis Risk Estimation (SCORE), and individual risk

factors (age, weight or prior fracture) to identify low central (hip and lumbar spine) bone mineral density (BMD) in 7779 US women aged 67 years and older participating in the Study of Osteoporotic Fractures.

Results The OST had the greatest area under the receiver operating characteristic curve (AUC 0.76, 95% CI 0.74, 0.77). Weight had an AUC of 0.73 (95% CI 0.72, 0.75), which was \geq AUC values for the ORAI, SCORE, age or prior fracture. Using cut points from the development papers, the risk tools had sensitivities \geq 85% and specificities \leq 48%. When new cut points were set to achieve a likelihood ratio of negative 0.1–0.2, the tools ruled out fewer than 1/4 of women without low central BMD.

Conclusions Weight identified low central BMD as accurately as the ORAI and SCORE. The risk tools would be unlikely to show an advantage over simple weight cut points in an osteoporosis screening protocol for elderly women.

M. L. Gourlay
Department of Family Medicine, University of North Carolina,
Chapel Hill, NC, USA

J. M. Powers
Department of Biostatistics, University of North Carolina,
Chapel Hill, NC, USA

L.-Y. Lui
Research Institute, California Pacific Medical Center,
San Francisco, CA, USA

K. E. Ensrud
Department of Medicine and Division of Epidemiology,
University of Minnesota, VA Medical Center,
Minneapolis, MN, USA

M. L. Gourlay (✉)
Aycock Building, Manning Drive, CB #7595, UNC-Chapel Hill,
Chapel Hill, NC 27599-7595, USA
e-mail: margaret_gourlay@med.unc.edu

Keywords Bone density · Female · Mass screening ·
Osteoporosis · Postmenopause · Risk assessment

Introduction

Osteoporosis risk assessment tools have been developed to objectively select postmenopausal women who could benefit from central (hip and lumbar spine) bone mineral density testing. The best validated tools are the Osteoporosis Self-assessment Tool (OST) developed in an Asian study population [1], the Osteoporosis Risk Assessment Instrument (ORAI) from a population-based Canadian cohort [2], and the Simple Calculated Osteoporosis Risk Estimation (SCORE) from a study population recruited from US academic and community-based medical centers [3]. Despite multiple validation and comparative studies in

postmenopausal women [4–11], these tools have yet to be used in clinical practice in the US. The main methodological barriers to clinical use have been lack of validation in a large, population-based US database and varying performance of the original cut points among different study populations.

We evaluated the diagnostic accuracy of the OST, ORAI and SCORE to detect low bone density in white women aged 67 years and older from the Study of Osteoporotic Fractures (SOF) cohort. Our objective was to test whether the tools could identify low bone density accurately enough to be useful for clinical decision-making in elderly US white women.

Methods

Study population

The SOF inception cohort included 9704 ambulatory white women aged 65 years and older recruited between 1986 and 1988 from population-based listings at four US sites: Baltimore, Maryland; Minneapolis, Minnesota; the Monongahela Valley near Pittsburgh, Pennsylvania; Portland, Oregon [12]. Women with bilateral hip replacements were excluded. All participants provided informed consent, and the study was approved by the appropriate institutional review committees of all participating sites. The age range of the SOF cohort was appropriate for testing of the osteoporosis risk assessment tools, since the development cohorts of the tools included women aged 45 to 80+ years.

We conducted a secondary analysis of the SOF Online public database <http://sof.ucsf.edu/public/> that included 7779 SOF participants with technically adequate bone mineral density measurements and a complete set of variables to calculate the risk scores at the second follow-up visit (1/89–12/90; this was the earliest visit at which central [hip and lumbar spine] bone density testing was performed). The number of participants with a complete set of variables differed for each risk tool, e.g., N=7617 for OST, N=7679 for ORAI, N=7235 for SCORE. (Note: these N values are from our analysis of the SOF Online database. An investigator [LL] at the SOF Coordinating Center repeated the N calculations in the complete SOF database, which includes confidential extreme values for continuous variables that are not available online. This only yielded about 170 additional eligible participants without significant differences in the ROC curve analysis results [results available upon request]. Thus, we conducted all analyses on the SOF Online data only.) The secondary analysis protocol was reviewed and approved by the Institutional Review Board of the University of North Carolina.

Variables

Bone mineral density (BMD) of the femoral neck and lumbar spine was measured using dual energy X-ray absorptiometry (DXA, Hologic, Waltham, MA). T-scores ($[\text{BMD of participant} - \text{mean BMD of reference population}] / \text{SD of BMD of reference population}$) are the basis for the World Health Organization diagnostic criteria for osteoporosis [13]. Femoral neck T-scores were calculated using NHANES III bone density norms for non-Hispanic white women aged 20–29 years [14]. Lumbar spine T-scores were calculated using Hologic densitometer manufacturer norms for women aged 30 years [15].

The following were the published reference variables for the tools: femoral neck T-score ≤ -2.5 for the OST, femoral neck or lumbar spine T-score ≤ -2.0 for the ORAI, and femoral neck T-score ≤ -2.0 for the SCORE. We also tested the World Health Organization (WHO) diagnostic criteria for osteoporosis (femoral neck or lumbar spine T-score ≤ -2.5) as an alternative reference variable in the receiver operating characteristic (ROC) curve analysis.

The classification variables were OST, ORAI and SCORE risk scores calculated using published formulas based on these clinical factors: age, weight, estrogen use, race/ethnicity, nontraumatic fracture after age 45, rheumatoid arthritis (Table 1) [1–3]. At least four different formulas have been used to calculate the OST [16]; we used the simplest formula [$0.2 \times (\text{weight in kg} - \text{age})$ without truncation]. The OST was transformed to (-OST) for use in the receiver operating characteristic curve analysis. Age, weight, and prior fracture were also tested as classification variables, because they have been found to be among the strongest determinants of osteoporosis [17] and fractures [18] in published analyses.

Statistical analysis

We constructed separate receiver operating characteristic (ROC) curves and determined the area under the curve (AUC) and the corresponding 95% confidence intervals for each tool to identify the appropriate reference variables. Each point on the ROC curve marks the sensitivity and corresponding value of [1-specificity] for a given cut point on the risk tool. The AUC is an overall estimate of the accuracy of the risk score to identify patients with low bone density. This area could range from 1 for a perfect test, to 0.5 for a test that performs no better than chance. The risk tools under study showed AUC values of 0.79 to 0.81 in the development cohorts and 0.79 in a comparative study (ORAI and SCORE only) [7].

We conducted cut point testing to test risk score thresholds for each tool and individual risk factors (age, weight, prior fracture) in the SOF database. We first used

Table 1 Osteoporosis risk assessment tools used in the analysis

Risk assessment tool calculation	Characteristics of development cohort	Cut point for high-risk category (development cohort)
Osteoporosis Self-assessment Tool (OST) [1] $0.2 \times (\text{weight in kg} - \text{age})$	Population-based sample of 860 Asian women: Mean age 62.3 years 100% postmenopausal	≤ -1 to identify women with femoral neck osteoporosis (T-score ≤ -2.5)
Osteoporosis Risk Assessment Instrument (ORAI) [2]: Add subscores from the following three categories:	Population-based sample of 926 Canadian women: Mean age 62.8 years 88% postmenopausal 12% pre- or perimenopausal	≥ 9 to identify women with T-score ≤ -2.0 at either femoral neck or lumbar spine
Age, y		
≥ 75	+15	
65–74	+9	
55–64	+5	
45–54	0	
Weight, kg		
< 60	+9	
60–69	+3	
≥ 70	0	
Current estrogen use		
No	+2	
Yes	0	
Simple Calculated Osteoporosis Risk Estimation (SCORE) [3]: Add subscores for each of the following six factors:	Clinic-based sample of 1424 US women:	≥ 6 to identify women with femoral neck T-score ≤ -2.0
Nonblack race	+5	89% white, 6% African American, 3% Latino, 3% other
Rheumatoid arthritis	+4	
Nontraumatic fracture (wrist, rib, hip) after age 45 years	+4 for each type of fracture, maximum +12	
Age, y	+3 \times first digit of age	Mean age 61.5 y 90% postmenopausal 10% perimenopausal
No prior estrogen therapy	+1	
Weight, lb	$-1 \times (\text{pounds} \div 10)$, truncated to integer	50% from primary care and geriatric clinics 20% endocrinology 20% rheumatology 10% gynecology

the cut points from the development papers, which were set to achieve 90% sensitivity to detect the reference variables in the development cohorts. For age, a cut point of 65 years or older was reasonable because current evidence-based guidelines recommend routine screening in this age range of women [19]; however, because all of the women in our dataset were aged 67 years or older, the entire cohort was classified as “high-risk” by this criterion. For weight, we used the cut point of <127 pounds (<57.7 kg) from the National Osteoporosis Foundation clinical guidelines for osteoporosis management [20]. We calculated the sensitivity

(true positive fraction), 1-specificity (false positive fraction) and corresponding binomial 95% confidence intervals using the exact method [21] for each risk tool applied to the SOF cohort. We then adjusted the cut points to achieve 90% sensitivity for each tool to identify the published reference variable, since the original cut points did not extrapolate to 90% sensitivity in the SOF cohort.

We used likelihood ratios as clinically relevant measures to summarize diagnostic accuracy [22]. We focused on the likelihood ratio negative (LR-) because past analyses have shown the osteoporosis risk tools to be most useful for

ruling out low bone density [4, 16]. New cut points were set to achieve an LR- of 0.1–0.2 for the low-risk group, since LR- values in this range generate moderate shifts in pre- to post-test probability to rule out patients without disease [23]. The corresponding likelihood ratio positive (LR+) values and large sample 95% confidence intervals [24] were calculated for each classification variable.

For the OST tool, we constructed a flow diagram of test outcomes when the cut point for LR- 0.1–0.2 was applied. Since current guidelines encourage screening for women aged 65 and older, we retained a dichotomous cut point that would exclude fewer women than a strategy using multilevel likelihood ratios.

All analyses were performed using SAS 9.1 [25] and Stata SE 9.0 software [26], including the Stata diagnostic accuracy analysis programs BINSRN1, EMROC and AUCBS [27].

Results

Of the 7779 SOF participants in the analytical cohort, 65.1% were aged 67 to 74 years, 29.8 % weighed below 60 kg, and 42.7% were past or current users of hormone therapy (Table 2). Thirty-six percent reported a baseline history of some type of fracture since age 50 years.

The overall discriminatory ability of the tools as measured by the AUC ranged from 0.60 for history of any fracture to 0.76 for the OST when the original reference variables were used (Table 3). The AUC for each tool was slightly lower when WHO diagnosis of osteoporosis (T-score \leq -2.5 at either femoral neck or lumbar spine) was used as the reference variable, except for the ORAI.

LR- values ranged from 0.13 to 0.31 for the risk tools when cut points from the development papers were used, and from 0.72 to 0.96 for fracture-related variables (Table 4). The cut points required to achieve 90% sensitivity to identify low bone density were different from the cut points used in the development studies (Table 5). The cut points required to achieve LR- 0.1–0.2 (to rule out women without disease) were also different from the cut points used in the development studies, except for the SCORE (Table 6). The corresponding sensitivities (96% to 99%) of the tools were high, and specificities (6% to 22%) were low using the adjusted cut points. A weight <176 pounds (<80 kg) was required to achieve an LR- value of 0.1–0.2 to rule out women without low bone density.

Figure 1 shows a flow diagram of OST test outcomes when the cut point was set to create a low-risk category with LR- 0.1–0.2. Of women with a low-risk OST score (>1), 4.2% (57/1368) had femoral neck osteoporosis. Of women with a high-risk score (\leq 1), 24.1% (1505/6249)

Table 2 Risk variables and bone density measures in participants in the Study of Osteoporotic Fractures (n=7779)

Characteristics	No. (%)	Risk tools that include this variable
Age, y		OST
≥75	2714 (34.9)	ORAI
67–74	5065 (65.1)	SCORE
Weight, kg		OST
<60	2319 (29.8)	ORAI
60–69	2436 (31.3)	SCORE
≥70	3024 (38.9)	
Estrogen use		ORAI
Never	4399 (57.3)	SCORE
Past	2165 (28.2)	
Current	1115 (14.5)	
Race/ethnicity		SCORE
White	7756 (99.7)	
Rheumatoid arthritis	517 (6.8)	SCORE
History of any fracture after age 50	2786 (35.8)	–
History of fragility fracture after age 50		SCORE
Rib	346 (4.5)	
Wrist	849 (11.0)	
Hip	127 (1.64)	
FN T-score \leq -2.5	1562 (20.5)	OST
FN or LS T-score \leq -2.0	4502 (57.9)	ORAI
FN T-score \leq -2.0	3189 (41.9)	SCORE
FN femoral neck		
LS lumbar spine		

had osteoporosis. Considering all women without osteoporosis, 21.7% (1311/6055) had a low-risk score.

Discussion

We conducted the largest US population-based validation study of osteoporosis risk assessment tools to date. Overall discriminatory ability was best for the OST, which is based only on age and weight. When cut points were set to achieve LR- 0.1–0.2 (to rule out women without disease), the tools excluded fewer than 1/4 of women without low bone density (e.g., only 21.7% of women without osteoporosis had a low-risk OST score). Weight alone was at least as accurate as two of the three risk tools (ORAI and SCORE), suggesting these risk score calculations offer minimal advantage at the point of care for women aged 67 years and older.

Similar to past comparative studies, we found the OST to have discriminatory ability equivalent to that of more complicated osteoporosis risk tools [4, 6, 10]. Unlike past studies, we found weight to have an area under the curve

Table 3 Area under the receiver operating characteristic curves (AUC) for osteoporosis risk assessment tools to identify low bone mineral density (BMD) in women aged 67 years or older, n=7779

Risk tool	AUC (95% CI) using original reference variable for each tool	AUC (95% CI) using femoral neck or lumbar spine T-score ≤ 2.5 as reference variable
OST (transformed) ^a	0.76 (0.74, 0.77) ^a	0.72 (0.71, 0.73)
ORAI	0.70 (0.68, 0.71) ^b	0.70 (0.69, 0.71)
SCORE	0.73 (0.72, 0.74) ^c	0.71 (0.70, 0.72)
Age	0.64 (0.63, 0.66) ^a	0.58 (0.57, 0.59)
Weight (kg)	0.73 (0.72, 0.75) ^a	0.72 (0.71, 0.73)
Any fracture after age 50	0.60 (0.58, 0.61) ^a	0.58 (0.57, 0.59)

CI confidence interval

^a OST was transformed to (-OST) to construct ROC curves. Reference variable for OST: femoral neck T-score ≤ -2.5 ; same reference variable used for age, weight, and prior fracture.^b reference variable for ORAI: T-score ≤ -2.0 at either femoral neck or lumbar spine^c reference variable for SCORE: femoral neck T-score ≤ -2.0

equivalent to or greater than the ORAI and SCORE, especially if T-score ≤ -2.5 was used as the reference variable in the analysis (with femoral neck or lumbar spine T-score ≤ -2.5 as outcome, AUCs were: weight 0.72, ORAI 0.70, SCORE 0.71). This finding may be due to the lower mean bone density in the SOF cohort compared to that in cohorts from other analyses. For example, Cadarette et al. [7] compared a weight criterion (<70 kg) to the ORAI, SCORE and other risk tools (not including OST) in 2365 Canadian women aged 45 years and older. The AUC for the weight classification variable increased by 0.11 when a lower femoral neck T-score was used as the reference variable (AUC 0.68 for weight vs. T-score < -1 , AUC 0.79 for weight vs. T-score ≤ -2.5). The corresponding AUCs for the ORAI and SCORE increased by only 0.08 (AUC 0.72 for ORAI vs. T-score < -1 , AUC 0.80 for ORAI vs.

T-score ≤ -2.5 ; AUC 0.71 for SCORE vs. T-score < -1 , AUC 0.79 for SCORE vs. T-score ≤ -2.5). Of note, Dargent-Molina et al. found that weight was the strongest determinant of very low BMD (T-score ≤ -3.5) in 6958 women aged 75 years and older participating in the EPIDOS study [17]. Thus, body weight might be especially useful in identifying women in the lower ranges of bone density that are more common with advancing age.

In its 2002 evidence report on osteoporosis screening, the US Preventive Services Task Force (USPSTF) reviewed ten cross-sectional studies of osteoporosis risk assessment, including the development studies for the ORAI and SCORE risk assessment tools [19, 28]. Methodological limitations of these studies were noted, especially lack of validation and lack of generalizability due to a small number of subjects or nonrepresentative subjects. The Task

Table 4 Diagnostic performance of osteoporosis risk tools to identify low BMD using published cut points (development studies)

Risk tool	High-risk score cut point	No. of patients classified		% (95% CI) ^a		Likelihood ratio ^a	
		With high-risk score	With low-risk score	Sensitivity (TPF)	1-Specificity (FPF)	Neg	Pos
OST	≤ -1	4502	3115	0.85 (0.83, 0.87)	0.52 (0.51, 0.54)	0.31	1.64
ORAI	≥ 9	7679	0	1.0	1.0	–	1.00
SCORE	≥ 6	6927	308	0.99 (0.99, 0.99)	0.93 (0.93, 0.94)	0.13	1.06
Age, y ^b	≥ 65	7617	0	1.0	1.0	–	1.00
Weight	<127 lb (<57.7 kg)	1818	5799	0.47 (0.45, 0.50)	0.18 (0.17, 0.19)	0.64	1.61
Fracture history	Any fracture after age 50	2722	4850	0.51 (0.49, 0.54)	0.32 (0.31, 0.33)	0.72	1.60
	Fragility fracture after age 50	1209	6379	0.27 (0.25, 0.29)	0.13 (0.12, 0.14)	0.84	2.00
	Prior rib fracture	339	7122	0.07 (0.06, 0.09)	0.04 (0.03, 0.04)	0.96	1.90
	Prior wrist fracture	827	6757	0.18 (0.17, 0.20)	0.09 (0.08, 0.10)	0.90	2.10

CI confidence interval TPF true positive fraction FPF false positive fraction

^a Reference variable defined as femoral neck T-score ≤ -2.5 for OST, age, weight, and fracture variables; femoral neck or lumbar spine T-score ≤ -2.0 for ORAI; femoral neck T-score ≤ -2.0 for SCORE^b Because all women in the cohort were aged 67 years or older, all of them were classified as “high-risk” by age.

Table 5 Diagnostic performance of osteoporosis risk tools using cut points to achieve 90% sensitivity to identify low BMD

Risk tool	High-risk score cut point	No. of patients classified		% (95% CI)*		Likelihood ratio*	
		With high-risk score	With low-risk score	Sensitivity (TPF)	1-Specificity (FPF)	Negative	Positive
OST	≤0	5494	2123	0.93 (0.91, 0.94)	0.67 (0.66, 0.68)	0.23	1.39
ORAI	≥12	5780	1899	0.86 (0.85, 0.87)	0.62 (0.60, 0.63)	0.37	1.39
SCORE	≥10	5656	1579	0.91 (0.89, 0.92)	0.69 (0.67, 0.71)	0.27	1.33
Age, y	≥69	6358	1259	0.89 (0.88, 0.91)	0.82 (0.81, 0.83)	0.58	1.09
Weight	<158 lb (<72 kg)	5464	2153	0.91 (0.90, 0.93)	0.67 (0.65, 0.68)	0.26	1.38

CI confidence interval

*Reference variable defined as femoral neck T-score ≤−2.5 for OST, age and weight; femoral neck or lumbar spine T-score ≤−2.0 for ORAI; femoral neck T-score ≤−2.0 for SCORE

Force concluded that further validation of existing risk instruments or development of new ones would be useful. Results from our study have better generalizability for US patients than most previous validation studies because our sample was large and was drawn from a US population-based cohort study. However, our results only apply to women aged 67 years and older, not to the full range of ages tested in the risk tool development cohorts (45 to 80 + years). The third National Health and Nutrition Examination Survey (NHANES III, 1988–1991) found a 4% prevalence of osteoporosis in women aged 50–59 years and a 20% prevalence in women aged 60–69 years [29], so that about a 10% prevalence of osteoporosis might be expected in the 50-to-66-year age range. Future studies should test performance of the risk tools in younger postmenopausal women, since excluding low-risk individuals is especially important in population groups with lower prevalence of disease. Future studies should also explore the economic implications of clinical use of the osteoporosis risk tools. Although a secondary analysis supported a possible economic benefit [30], the uncertain generalizability of these tools and lack of standard cut points across different

clinic populations (e.g., different ages of postmenopausal women, different ethnic groups) remain barriers to clinical use.

Since the Task Force report was published in 2002, secular change in medical policy has called into question the clinical role of the osteoporosis risk tools. Current research on risk assessment has focused on development and validation of the World Health Organization absolute fracture probability algorithm [31]. This algorithm determines 10-year absolute fracture risk for previously untreated patients aged 50 years and older based on the following risk factors: prior fracture, age, weight, family history, smoking, alcohol, corticosteroid use and other secondary causes of osteoporosis, and femoral neck bone mineral density [32]. Economic cost-effectiveness modeling will be used to determine reasonable treatment thresholds for different countries. Only patients with bone density T-scores ≤−2.5 are diagnosed with osteoporosis by current WHO criteria; in contrast, some patients with osteopenia plus other significant risk factors will have an absolute fracture risk above the treatment threshold according to the new algorithm. Since the WHO algorithm includes bone density measures, a

Table 6 Diagnostic performance of osteoporosis risk tools using cut points to achieve LR- 0.1–0.2 to identify low BMD

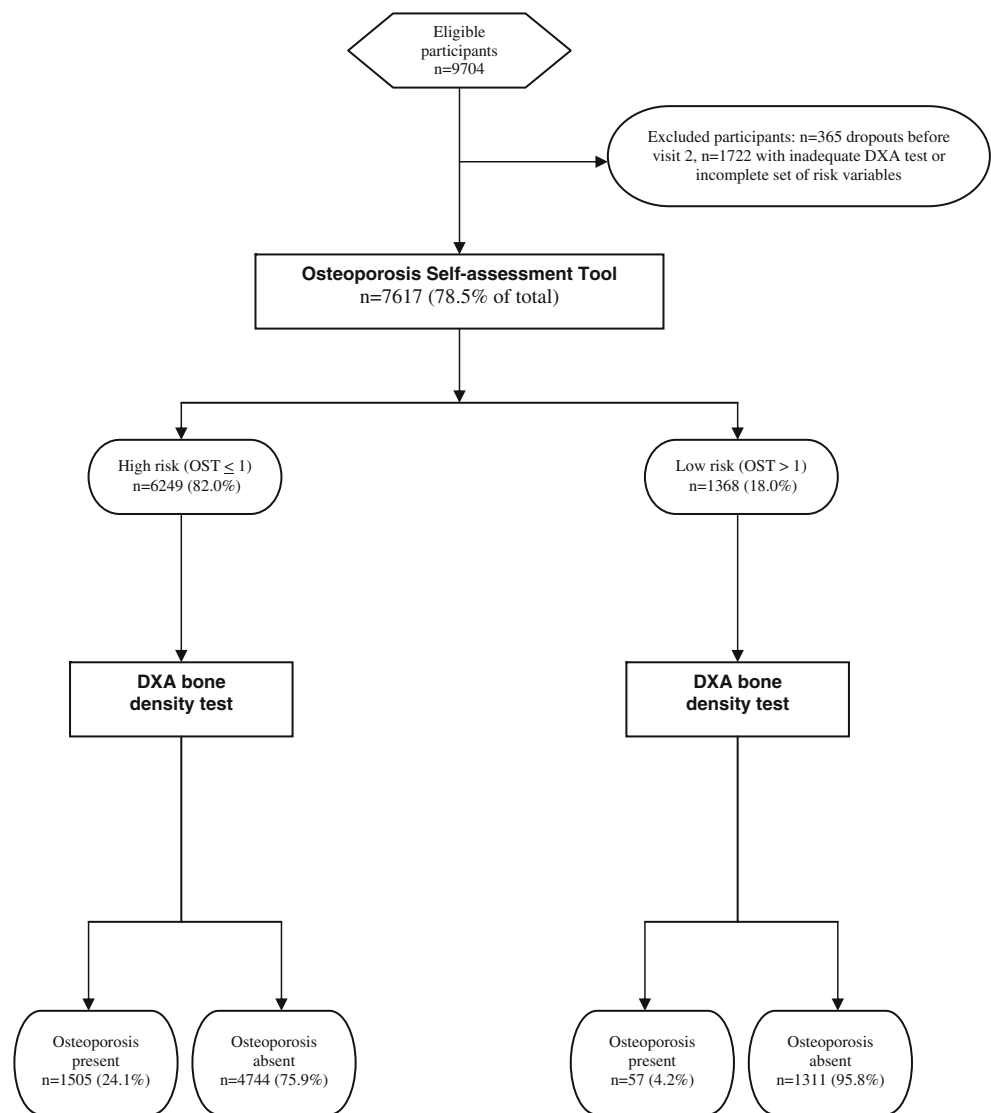
Risk tool	High-risk score cut point	No. of patients classified		% (95% CI) ^a		Likelihood ratio ^a	
		With high-risk score	With low-risk score	Sensitivity (TPF)	1-Specificity (FPF)	Negative	Positive
OST	≤1	6249	1368	0.96 (0.95, 0.97)	0.78 (0.77, 0.79)	0.17	1.23
ORAI	≥11	7414	265	0.99 (0.98, 0.99)	0.94 (0.93, 0.95)	0.20	1.05
SCORE	≥6	6927	308	0.99 (0.99, 0.99)	0.93 (0.93, 0.94)	0.13	1.06
Age, y ^b	–	–	–	–	–	–	–
Weight	<176 lb (<80 kg)	6609	1008	0.98 (0.97, 0.98)	0.84 (0.83, 0.85)	0.16	1.16

CI confidence interval

^aReference variable defined as femoral neck T-score ≤−2.5 for OST, age and weight; femoral neck or lumbar spine T-score ≤−2.0 for ORAI; femoral neck T-score ≤−2.0 for SCORE

^bLowest achievable LR- for age was 0.54

Fig. 1 Application of the Osteoporosis Self-assessment Tool (OST) to identify low BMD in the study sample of white women aged 67 years and older



standardized approach to selecting patients for bone density testing is still needed; however, such an approach will probably be dictated by upcoming policy changes. In October 2006, the Osteoporosis Work Group of the Physician Consortium for Performance Improvement reported central DXA screening for all women aged 65 and older as a Clinical Performance Measure [33]. Thus, one might anticipate little need for osteoporosis risk assessment tools in older women. However, pending legislative cuts are expected to decrease reimbursement for the performance of DXA bone density testing by over 70% by 2010 [34]. This reduction in reimbursement may lead to less testing, even in older women. These external factors are likely to influence physician decision-making much more than available risk tools. Finally, estrogen use has declined sharply after the Women's Health Initiative investigators reported adverse outcomes from estrogen therapy in postmenopausal women

in 2002 [35]. Thus, the estrogen variables in the ORAI and SCORE are less relevant for a greater number of women.

Our study had several limitations. Twenty percent (1925/9704) of the original SOF participants were excluded in this analysis due to drop-out between the baseline and second study visit, or technically inadequate or missing bone density measures or risk factor data. Results from the remaining sample may not be generalizable to all postmenopausal women in US primary care settings. We could only analyze women aged 67 and older in the SOF database; this limited our analysis of the ORAI tool, since an age of 67 years or older placed all participants in the high-risk category using the original cut points. As is true of most previous analyses of these tools, our sample comprised only white women; our results should not be extrapolated to nonwhite women, since the tools may have lower ability to detect low bone density in those groups [36]. Finally, the

tools were designed to identify women with risk factors for osteoporosis who may be candidates for bone density testing; the tools do not identify osteoporosis accurately enough to be used as the sole basis (without bone density testing) for treatment decisions.

Conclusions

Our analysis does not support use of the osteoporosis risk assessment tools in US women aged 67 and older. The tools showed fair ability to rule out women without osteoporosis. However, weight alone showed comparable performance, indicating that these tools have little or no advantage over simple weight assessment in osteoporosis screening protocols. An accurate diagnosis of osteoporosis by current WHO criteria can only be made using direct measurement of bone mineral density. As financial limitations on screening increase, primary care physicians should direct screening efforts toward older patients with lower body weight to ensure that the highest-risk patients are selected for diagnostic bone density testing.

Acknowledgments We acknowledge Philip D. Sloane, MD, MPH for his comments on an earlier draft of the manuscript.

Financial disclosures None reported.

Funding/Support The project described was funded by Grant Number K23RR024685 from the National Center for Research Resources and by the University of North Carolina Translational Science Program. The Study of Osteoporotic Fractures (SOF) is supported by the National Institutes of Health (Public Health Service research grants AG05407, AR35582, AG05394, AR35584, AR35583, R01 AG005407, R01 AG027576-22, 2 R01 AG005394-22A1, and 2 R01 AG027574-22A1). The content is solely the responsibility of the authors and does not necessarily reflect the official views of the funding agencies.

Role of the sponsor The funding bodies had no role in data extraction and analyses, in the writing of the manuscript, or in the decision to submit the manuscript for publication.

References

- Koh L, Ben Sedrine W, Torralba T, Kung A, Fujiwara S, Chan S, Huang Q, Rajatanavin R, Tsai K, Park H, Reginster J (2001) A simple tool to identify Asian women at increased risk of osteoporosis. *Osteoporos Int* 12:699–705
- Cadarette S, Jaglal S, Kreiger N, McIsaac W, Darlington G, Tu J (2000) Development and validation of the Osteoporosis Risk Assessment Instrument to facilitate selection of women for bone densitometry. *CMAJ* 162:1289–1294
- 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 Man Care* 4:37–48
- Richy F, Gourlay M, Ross P, Sen S, Radican L, De Ceulaer F, Ben Sedrine W, Ethgen O, Bruyere O, Reginster J-Y (2004) Validation and comparative evaluation of the Osteoporosis Self-assessment Tool (OST) in a Caucasian population from Belgium. *QJM* 97:39–46
- Ben Sedrine W, Devogelear JP, Kaufman J-M, Goemaere S, Depresseux G, Zegels B, Deroisy R, Reginster J-Y (2001) Evaluation of the Simple Calculated Osteoporosis Risk Estimation (SCORE) in a sample of white women from Belgium. *Bone* 29:374–380
- Geusens P, Hochberg M, van der Voort D, Pols H, van der Klift M, Siris E, Melton M, Turpin J, Byrnes C, Ross P (2002) Performance of risk indices for identifying low bone density in postmenopausal women. *Mayo Clin Proc* 77:629–637
- Cadarette S, Jaglal S, Murray T, McIsaac W, Joseph L, Brown J (2001) Evaluation of decision rules for referring women for bone densitometry by dual-energy x-ray absorptiometry. *JAMA* 286:57–63
- Cadarette S, McIsaac W, Hawker GA, Jaakkimainen L, Culbert A, Zarifa G, Ola E, Jaglal S (2004) The validity of decision rules for selecting women with primary osteoporosis for bone mineral density testing. *Osteoporos Int* 15:361–366
- Rud B, Jensen J, Mosekilde L, Nielsen S, Hilden J, Abrahamsen B (2005) Performance of four clinical screening tools to select peri- and early postmenopausal women for dual X-ray absorptiometry. *Osteoporos Int* 16:764–772
- Gourlay M, Miller W, Richy F, Garrett J, Hanson L, Reginster J-Y (2005) Performance of osteoporosis risk assessment tools in postmenopausal women aged 45 to 64 years. *Osteoporos Int* 16:921–927
- Mauck K, Cuddihy M-T, Atkinson E, Melton LJ, 3rd (2005) Use of clinical prediction rules in detecting osteoporosis in a population-based sample of postmenopausal women. *Arch Intern Med* 165:530–536
- Cummings S, Black D, Nevitt M, Browner W, Cauley J, Genant H, Mascioli S, Scott J, Seeley D, Steiger P, Vogt T (1990) Appendicular bone density and age predict hip fracture in women. The Study of Osteoporotic Fractures Research Group. *JAMA* 263:665–668
- World Health Organization. Assessment of osteoporotic fracture risk and its role in screening for postmenopausal osteoporosis. WHO Technical report series no. 843. Geneva: World Health Organization 1994
- Looker A, Wahner H, Dunn W, Calvo M, Harris T, Heyse S, Johnston C Jr, Lindsay R (1998) Updated data on proximal femur bone mineral levels of US adults. *Osteoporos Int* 8:468–489
- Hologic, Inc. Bone mineral density international normative data tables for men and women, report from manufacturer modified 11/4/91 and 2/1/97
- 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* 18:1177–1187
- Dargent-Molina P, Poitiers F, Breart G (2002) In elderly women weight is the best predictor of a very low bone mineral density: evidence from the EPIDOS Study. *Osteoporos Int* 11:881–888
- Eisman J, Johnell O, De Laet C, Johansson H, Oden A, Delmas P, Eisman J, Fujiwara S, Garnero P, Kroger H, McCloskey E, Mellstrom D, Melton LJ, 3rd, Pols H, Reeve J, Silman A, Tenenhouse A (2004) A meta-analysis of previous fracture and subsequent fracture risk. *Bone* 35:375–382

19. USPSTF (2002) Screening for osteoporosis in postmenopausal women: recommendations and rationale. *Ann Intern Med* 137:526–528
20. National Osteoporosis Foundation. Physician's guide to prevention and treatment of osteoporosis. 1999, updated in 2003. <http://www.nof.org/>. Accessed November 10, 2007.
21. StataCorp. Confidence intervals for means, proportions, and counts. In: *Stata Statistical Software: Release 9.0 User's Guide, Volume 1*. College Station, TX: Stata Corporation;2005
22. Deeks J, Altman D (2004) Diagnostic tests 4: likelihood ratios. *BMJ* 329:168–169
23. Jaeschke R, Guyatt G, Lijmer J (2002) Diagnostic tests. In: Guyatt G, Drummond R (eds) *Users' guides to the medical literature: a manual for evidence-based clinical practice*. AMA Press, Chicago, pp 121–139
24. Simel D, Samsa G, Matchar D (1991) Likelihood ratios with confidence: sample size estimation from diagnostic test studies. *J Clin Epidemiol* 44:763–770
25. SAS Institute Inc (2004) *SAS/STAT 9.1 User's Guide*. Cary, NC: SAS Institute Inc. In:
26. StataCorp (2005). *Stata Statistical Software: Release SE 9.0*. College Station, TX: Stata Corporation. In:
27. Fred Hutchinson Cancer Research Center, Department of Biostatistics, University of Washington. The statistical evaluation of medical tests for classification and prediction: datasets and Stata programs used in the text. <http://www.fhcr.org/science/labs/pepe/book/#progs>, accessed November 10, 2007.
28. Nelson H, Helfand M, Woolf S, Allan J (2002) Screening for postmenopausal osteoporosis: a review of the evidence for the US Preventive Services Task Force. *Ann Intern Med* 137:529–541
29. Looker A, Johnston C Jr, Wahner H, Dunn W, Calvo M, Harris T, Heyse S, Lindsay R (1995) Prevalence of low femoral bone density in older US women from NHANES III. *J Bone Miner Res* 10:796–802
30. Richey F, Ethgen O, Bruyere O, Mawet A, Reginster J-Y (2004) Primary prevention of osteoporosis: mass screening scenario or prescreening with questionnaires? An economic perspective. *J Bone Miner Res* 19:1955–1960
31. Kanis J, Oden A, Johnell O, Johansson H, DeLaet C, Brown J, Burckhardt P, Cooper C, Christiansen C, Cummings S, Eisman J, Fujiwara S, Gluer C, Goltzman D, Hans D, Krieg M-A, LaCroix A, McCloskey E, Mellstrom D, Melton LJ, 3rd, Pols H, Reeve J, Sanders K, Schott A, Silman A, Torgerson D, van Staa T, Watts N, Yoshimura N (2007) The use of clinical risk factors enhances the performance of BMD in the prediction of hip and osteoporotic fractures in men and women. *Osteoporos Int*:Feb 24; [Epub ahead of print]
32. Dawson-Hughes, B. Development of the WHO fracture risk assessment paradigm. National Osteoporosis Foundation 7th International Symposium on Osteoporosis: Translating Research into Clinical Practice, Washington, DC, April 18–22, 2007. In:
33. Physician Consortium for Performance Improvement Osteoporosis Work Group. Osteoporosis Physician Performance Measurement Set. <http://www.ama-assn.org/ama/pub/category/4837.html>. October 2006. Accessed November 10, 2007.
34. International Society for Clinical Densitometry. Medicare Finalizes Cuts to DXA and VFA. <http://www.iscd.org/Visitors/positions/ProposedCutstoDXA.cfm>, accessed November 10, 2007.
35. Udell J, Fischer M, Brookhart M, Solomon D, Choudhry N (2006) Effect of the Women's Health Initiative on osteoporosis therapy and expenditure in Medicaid. *J Bone Miner Res* 21:765–771
36. Cass A, Shepherd A, Carlson C (2006) Osteoporosis risk assessment and ethnicity: validation and comparison of 2 clinical risk stratification instruments. *J Gen Intern Med* 21:630–635