

Individualization of osteoporosis risk

T. V. Nguyen

Received: 19 April 2007 / Accepted: 4 June 2007 / Published online: 5 July 2007
© International Osteoporosis Foundation and National Osteoporosis Foundation 2007

Anti-fracture therapy is often pursued in individuals with osteoporosis or low bone mineral density with or without a prior fracture [1–3], because they are at high risk of subsequent fracture [4, 5]. In recent years, there has been a proliferation of clinical prediction rules, including clinical risk indices and scores, for identifying individuals with osteoporosis. The main problem with these clinical prediction rules is that they categorize individuals into low-risk versus high-risk groups based on some arbitrary threshold. As a result, when this risk stratification-based approach is applied to an individual, its prognostic performance is often poor.

Among the clinical indices for predicting osteoporosis, the Osteoporosis Self-Assessment Tool (OST) [6] is perhaps the simplest score and has perhaps been the most widely studied. However, the clinical usefulness of this score in different populations has not been systematically studied. In this issue, Rud and colleagues [7] systematically summarize some key prognostic properties of the OST score across different Asian and Caucasian populations. They especially focus on the prognostic value of OST in ruling out osteoporosis (i.e., LR[−]). Their meta-analysis showed that the overall LR[−] was 0.19 (95% confidence interval: 0.17 and 0.21) with considerable heterogeneity between populations.

Although Rud and colleagues do not report other prognostic measures of the OST score in their paper, it is possible to estimate the LR⁺, sensitivity, specificity, and summary receiver operating characteristic curve (sROC)

from the data (Table 1). These results suggest that in both Asian and Caucasian populations, the OST score has good sensitivity, but poor specificity, and moderate discriminatory value (i.e., with sROC being between 0.74 and 0.79). The odds for OST positivity among women with osteoporosis is 11 times (in Asians) and 7.8 times (in Caucasians) higher than the odds for positivity among women without osteoporosis. This diagnostic odds ratio is much lower than the criterion required for an “adequate test”. However, these results are broadly consistent with prognostic values of the simple calculated osteoporosis risk estimation (SCORE), osteoporosis risk assessment instrument (ORAI) and National Osteoporosis Foundation guidelines, which were also found to have poor specificity and high sensitivity [8].

When the prevalence of osteoporosis is 17% (which is observed in Caucasian populations [7]) and given the observed LR[−] and LR⁺ in Table 1, it can be estimated that if an individual is classified as “low risk” by OST, the probability that the individual has no osteoporosis is 0.96; and if the individual is classified as “high risk”, the probability that the individual has osteoporosis is only ~0.25. In other words, the OST score is useful in ruling out the possibility of osteoporosis, but it performs badly in ruling in the possibility of osteoporosis. The best positive and negative predictive values are achieved when the sensitivity is around 70% and specificity is approximately 80% (Fig. 1). However, in all studies reviewed, the OST score has a sensitivity of >90% and specificity of <57%.

Several factors contribute to the relatively poor performance of OST. First, the prevalence of osteoporosis in the populations under review, and the relationships between age, weight and BMD is not constant across the population. For example, the prevalence of osteoporosis varied between 7 to 21%, with the LR[−] being between 0.02 and 0.49 among the Asian populations under review. Indeed, a

T. V. Nguyen (✉)
Bone and Mineral Research Program, Garvan Institute of Medical Research, St Vincent's Hospital,
384 Victoria Street Darlinghurst,
Sydney, NSW 2010, Australia
e-mail: t.nguyen@garvan.org.au

Table 1 Summary of diagnostic values for Asians and Caucasians

Diagnostic measures	Asians	Caucasians
Prevalence of osteoporosis (%)	11.2	16.7
Overall sensitivity	0.90 (0.88, 0.92)	0.94 (0.93, 0.95)
Overall specificity	0.57 (0.56, 0.58)	0.37 (0.36, 0.37)
Likelihood ratio positive (LR+)	2.06 (1.70, 2.51)	1.66 (1.51, 1.82)
Likelihood ratio negative (LR-)	0.19 (0.13, 0.28)	0.19 (0.17, 0.21)
Diagnostic odds ratio (DOR)	11.2 (7.3, 17.1)	7.83 (7.08, 8.67)
Summary AUC (mean, SE)	0.74 (0.03)	0.79 (0.03)

Overall sensitivity, specificity, LR+, LR- and diagnostic odds ratio are shown in mean and 95% confidence interval. “Osteoporosis” here refers to the T-scores of femoral neck BMD ≤ -2.5 . LR+ is obtained as the true positive rate divided by false positive rate (i.e., sensitivity divided by one minus specificity). LR- is obtained as false negative rate divided by true negative rate (i.e., 1 minus sensitivity divided by specificity). DOR is obtained as the ratio of LR+ over LR-, which represents a compromise between the two likelihood ratio values; it is considered a good measure to compare the overall accuracy of a test evaluated in different studies

regression analysis of log DOR against the measure of threshold [9] among the Asian populations yielded a slope of 0.41 (with standard error=0.14, $p=0.014$), suggesting that the performance of OST varies with threshold across populations.

Second, the correlation between age or weight and BMD is moderate, and when these variables are dichotomized there is a loss of information. In the Dubbo Osteoporosis Epidemiology Study [10], for example, age and body weight—in their continuous measurements—collectively account for 36% of total variance in femoral neck BMD.

When BMD is dichotomized into osteoporosis or non-osteoporosis [11], and this dichotomous variable is modeled as a function of age and weight in their continuous scales in a logistic regression model, the pseudo R^2 value reduced to 24%. Furthermore, when the dichotomous variable is modeled as a function of OST category (i.e., “low risk” versus “high risk”) in the logistic regression model, the pseudo R^2 value was further reduced to 13%, a reduction of 23% from the R^2 value of the original continuous model. Therefore, the dichotomization of OST scores into a “low risk” and “high risk” category can result in a considerable

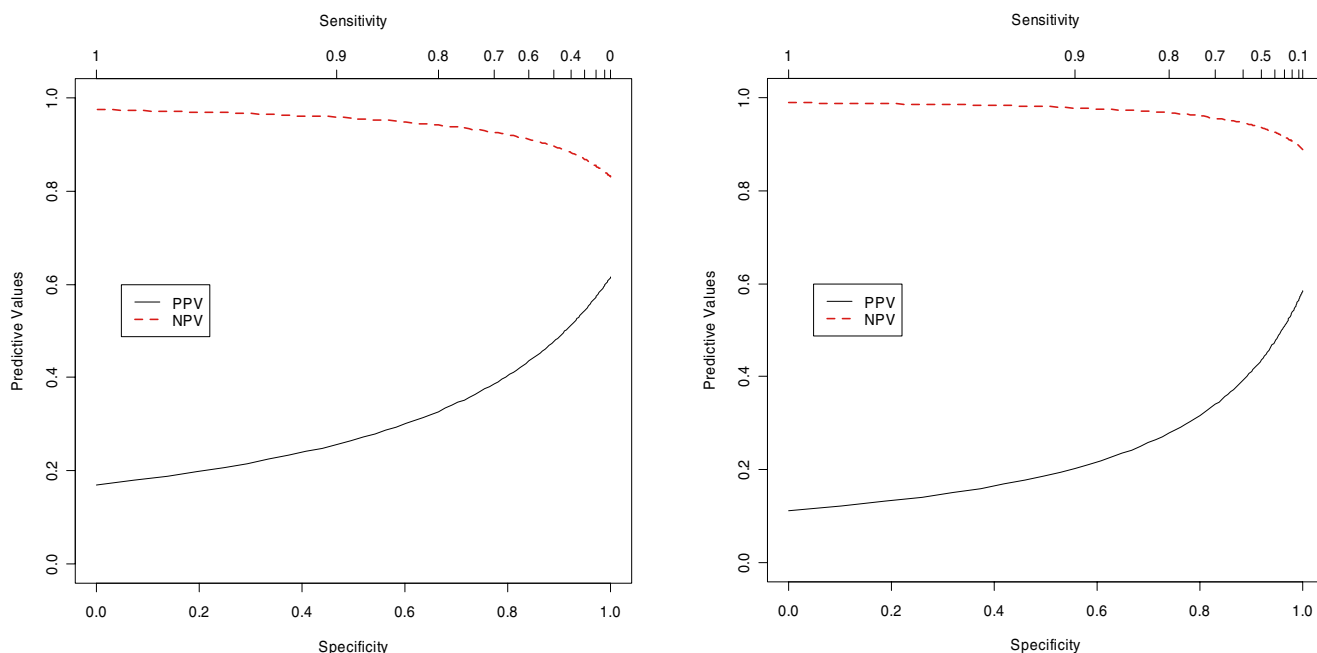


Fig. 1 Positive predictive value (PPV) and negative predictive value (NPV) as a function of sensitivity and specificity for Asian populations (left panel) and Caucasian population (right panel). The curves were generated according to the following relations: $PPV = \frac{1}{\frac{1}{DOR - Sensitivity \times (DOR - 1)} + \frac{1 - Prevalence}{Prevalence}}$ $NPV = \frac{1}{1 + \frac{Prevalence}{(1 - Prevalence) \times DOR - Sensitivity \times (DOR - 1)}}$ where DOR is the the diagnostic odds ratio, which was set at 11.2 in

Asians and 7.83 in Caucasians, and prevalence of osteoporosis was 11% in Asians and 17% in Caucasians. DOR is obtained as the ratio of LR+ over LR-, which reflects the odds of positive OST among subjects with osteoporosis relative to the odds of positive OST among subjects without osteoporosis

loss of information. With such a double loss of information (from the dichotomization of BMD and OST scores), the predictive power of discrimination of the OST score is thus expected to be low.

Third, BMD is measured with random error and random fluctuation within individuals. This variation, albeit small relative to the mean, can nevertheless result in a significant misclassification of osteoporosis versus non-osteoporosis [12]. An empirical analysis suggested that in individuals aged between 60 and 69 years, among whom the prevalence of osteoporosis is estimated to be 10.5%, the false positive rate could be as high as 12.4% [12]. Therefore, when correlating two imperfect and dichotomized measurements (OST scores and BMD), the discrimination and calibration can never be expected to be perfect.

Fourth, the derivation of the OST score was based on the concept of risk stratification, in which individuals were classified into subsets using some cut-off values, despite the fact that there is no threshold between high and low scores, but rather a progressive increase in the risk of osteoporosis with lowering OST scores. For example, the OST is commonly calculated by the following formula: $OST = 0.2 \times (\text{weight} - \text{age})$, and by which, individuals with OST scores being less than or equal to -1 were classified as “high risk”, and otherwise, a “low risk” classification was made. Thus, by the OST classification, a 50 year-old woman with 55 kg is classified as having the same risk of osteoporosis (e.g., high risk) as a 60 year-old woman with 65 kg. On the other hand, a 55 year-old woman with 50 kg is classified as “high risk”, whereas a woman with the same age but 51 kg is classified as “low risk”. It is therefore not surprising that when such a risk stratification is applied to an individual, the calibration and discrimination can not be expected to be high. Indeed, even if the OST score can perfectly resolve the population into two distinct groups of osteoporosis and non-osteoporosis by the cut-off value of -1 , the calibration is still not perfect, because the probability of osteoporosis associated with an OST score of -1 means exactly the same as the probability of osteoporosis associated with an OST score of -5 !

Prognosis is about imparting information of risk to an individual, and each individual is a unique case, in the sense that any two individuals are unlikely to have the same risk profile. Therefore, the risk of osteoporosis should ideally be individualized. One approach to increase the uniqueness of prediction is to consider the risk in its continuous scale, rather than in dichotomized scale based on a cut-off value. The uniqueness of prediction, or individualization of risk, can also be increased by considering multiple risk factors in a multivariable model, since the more factors considered, the greater likelihood of uniqueness of an individual is defined. Several studies in the cancer field have suggested that such a nomogram-based risk prediction performs better than risk-grouping approach [13, 14],

because the nomogram recognizes and can define the unique risk profile for an individual. In osteoporosis, it has recently been shown that a multivariable-based nomogram incorporating quantitative ultrasound measurement and the OST components (i.e., age and weight) could significantly increase the predictive value of osteoporosis [15].

An important weakness of all current prognostic models is that they are based on a single measurement, with the underlying, but not stated, assumption that the measurement does not change with time. Obviously, this assumption is not true in many measurements such as BMD and body weight that are known to decline or change with time, and that the rates of change varied substantially among individuals [16]. Therefore, an important perspective in clinical risk score development should take the time-varying nature of risk factors into account to achieve a better prediction of risk for an individual.

Osteoporosis as operationally defined by BMD is one of many biologic, genetic and environmental risk factors that affect the likelihood of fracture in an individual. While osteoporosis is the best predictor of fracture risk, it can not account for all fractures in the elderly population. Indeed, in individuals aged 60+ years, 55 and 74% of fracture cases occurred in non-osteoporotic women and men, respectively [17]. Thus, from a clinical point of view, the prediction of osteoporosis is not as useful as the prediction of fracture risk, and to this end, the OST does not perform well [18].

The goal of a clinical risk score or any clinical prediction rule is to suggest a prognosis or therapeutic action for an individual. However, due to its modest predictive properties, the OST score is far from suitable for clinical use on a regular basis, but it may be used for ruling out BMD scans or for counselling purposes. Nevertheless, there is room for further refinement of the score by making use of its continuous measurements and other clinical measurements to better characterize the risk of osteoporosis and/or fracture for an individual.

References

1. National Osteoporosis Foundation (2003) National Osteoporosis Foundation physician's guide to prevention and treatment of osteoporosis. National Osteoporosis Foundation, Washington, DC
2. Sambrook PN, Seeman E, Phillips SR, Ebeling PR (2002) Preventing osteoporosis: outcomes of the Australian fracture prevention summit. *Med J Aust* 176(Suppl):S1–S16
3. Seeman E, Eisman JA (2004) 7: Treatment of osteoporosis: why, whom, when and how to treat. The single most important consideration is the individual's absolute risk of fracture. *Med J Aust* 180:298–303
4. Marshall D, Johnell O, Wedel H (1996) Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures. *BMJ* 312:1254–1259

5. Nguyen ND, Pongchaiyakul C, Center JR, Eisman JA, Nguyen TV (2005) Identification of high-risk individuals for hip fracture: a 14-year prospective study. *J Bone Miner Res* 20:1921–1928
6. Koh LK, Sedrine WB, Torralba TP, Kung A, Fujiwara S, Chan SP, Huang QR, Rajatanavin R, Tsai KS, Park HM, Reginster JY (2001) A simple tool to identify Asian women at increased risk of osteoporosis. *Osteoporos Int* 12:699–705
7. 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. DOI [10.1007/s00198-006-0319-3](https://doi.org/10.1007/s00198-006-0319-3)
8. Mauck KF, Cuddihy MT, Atkinson EJ, 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
9. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F (1994) Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 120:667–676
10. Nguyen ND, Ahlborg HG, Center JR, Eisman JA, Nguyen TV (2007) Residual lifetime risk of fractures in women and men. *J Bone Miner Res* 22(6):781–788
11. Kanis JA, Melton LJ 3rd, Christiansen C, Johnston CC, Khaltaev N (1994) The diagnosis of osteoporosis. *J Bone Miner Res* 9:1137–1141
12. Nguyen TV, Pocock N, Eisman JA (2000) Interpretation of bone mineral density measurement and its change. *J Clin Densitom* 3:107–119
13. Kattan MW (2002) Nomograms: introduction. *Semin Urol Oncol* 20:79–81
14. Kattan MW, Leung DH, Brennan MF (2002) Postoperative nomogram for 12-year sarcoma-specific death. *J Clin Oncol* 20:791–796
15. Pongchaiyakul C, Panichkul S, Songpatanasilp T, Nguyen TV (2007) A nomogram for predicting osteoporosis risk based on age, weight and quantitative ultrasound measurement. *Osteoporos Int* 18:525–531
16. Nguyen TV, Sambrook PN, Eisman JA (1998) Bone loss, physical activity, and weight change in elderly women: the Dubbo Osteoporosis Epidemiology Study. *J Bone Miner Res* 13:1458–1467
17. Nguyen ND, Eisman JA, Center JR, Nguyen TV (2007) Risk factors for fracture in nonosteoporotic men and women. *J Clin Endocrinol Metab* 92:955–962
18. Nguyen TV, Center JR, Pocock NA, Eisman JA (2004) Limited utility of clinical indices for the prediction of symptomatic fracture risk in postmenopausal women. *Osteoporos Int* 15:49–55