

Better Bone Density Reporting: T-score Report Versus Fracture Risk Report With Outcome Analysis

Lark Greenwald, MD; Karen Barajas, MW;
and Maria White-Greenwald, MD

Objective: To determine whether different classification systems for bone fracture risk result in different diagnoses and treatments.

Study Design: Assessment of 2 diagnostic classification systems by physicians who routinely read bone density scans, using a survey administered by an audience response system.

Methods: At a professional meeting, 103 rheumatologists were asked to read bone density results from 9 case studies in 2 formats: a standard format using T-scores and a format using absolute fracture risk. As physicians were shown each of the 18 scans, they diagnosed each patient with (1) normal bone, (2) osteopenia, (3) osteoporosis, or (4) unknown, and provided a treatment recommendation. Test scans included hip dual-energy x-ray absorptiometry (DEXA) (60%), spine DEXA (10%), phalangeal DEXA (10%), and calcaneal ultrasound (20%).

Results: Approximately 1850 anonymous responses were recorded. For women, diagnostic agreement between physicians was 17% for the T-score report and 89% for the absolute risk report ($P < .001$). For men, diagnostic agreement also was better with the absolute risk report (66% vs 10%; $P < .001$). With DEXA, overall diagnostic agreement improved from 30% to 68% with the absolute risk report ($P < .001$). With ultrasound, diagnostic agreement improved from 4% to 63% with the absolute risk report ($P < .001$). Bisphosphonates were more likely to be prescribed in cases of increasing age and risk if scans were reported in terms of absolute fracture risk.

Conclusion: A direct, risk-based bone density assessment may improve consistency and efficiency of diagnosis and treatment of osteoporosis.

(*Am J Manag Care* 2003;9:665-670)

Osteoporosis is a growing problem in the United States that poses significant challenges for physicians, healthcare professionals, and patients alike. According to the National Osteoporosis Foundation, about 25% of postmenopausal Caucasian women in the United States have osteoporosis, and an additional 54% have low bone density as measured at the hip, spine, or wrist.¹ The estimated cost to treat these patients is more than \$13.8 billion annually.²

Although techniques to measure bone density have improved significantly, the ability to clearly distinguish the need for treatment for osteoporosis is confounded

by the imprecise methods used to report bone mineral density. To date, the standard has been the T-score, which represents bone loss in standard deviations since young adult peak bone density. However, because the T-score was developed for epidemiologic purposes, it has limitations when applied to individual patient care, including the fact that it has no intrinsic meaning, making it difficult to understand for many patients and even some physicians. Moreover, the T-score cut-off of -2.5 was based on fracture data for 65- to 70-year-old women and does not account for large variations in fracture risk at different ages, despite the fact that age is the single most important predictor of fracture.³ Because T-scores do not account for age, they underestimate the risk in elderly patients and overestimate the risk in younger patients. A recent study showed that only 25% of subjects were consistently diagnosed when practitioners used the World Health Organization (WHO) criteria with a T-score of -2.5 relative to the manufacturer's young normative data.⁴

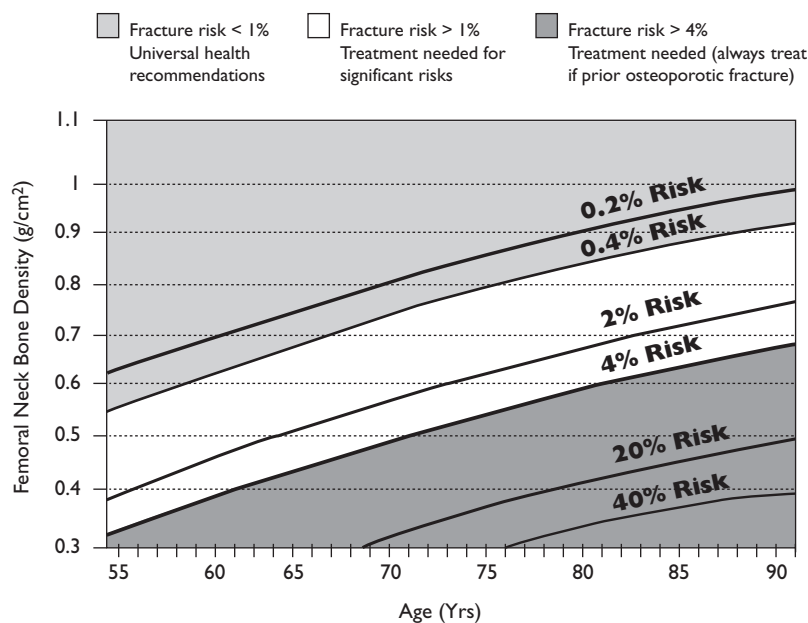
In a managed care environment, physicians must institute cost-containment measures while ensuring the accurate diagnosis and treatment of patients. The inability to consistently diagnose patients at risk for fractures leads to unnecessary costs for younger patients and, more importantly, may lead to mismanagement of older patients, who are more likely to develop fractures. The direct medical costs associated with hip fractures in men and women in the United States have been reported to be in excess of \$5 billion per year.⁵

From the Osteoporosis Medical Center, a Nonprofit Research Corporation, Palm Springs, Calif.

This project was supported by the Osteoporosis Medical Center. These data were presented at the 22nd Annual Meeting of the American Society for Bone and Mineral Research, Toronto, Canada, September 2000.

Address correspondence to: Maria White-Greenwald, MD, Osteoporosis Medical Center, 72855 Fred Waring Drive, Palm Desert, CA 92260.

Figure. Chart Used to Identify Patients' Absolute 4-Year Fracture Risk Based on Their Bone Density and Age



Bone density was measured using a Hologic QDR-2000 bone densitometer (Hologic Inc, Bedford, Mass)
Adapted from reference 8.

Variability in screening procedures has been recognized as a problem in other disciplines as well. In the 1980s, mammography was subject to scrutiny due to the lack of concordance in diagnosing breast cancer. To improve the accuracy in reading mammograms, the Food and Drug Administration (FDA) established the Mammography Quality Standards Act of 1992, which set forth quality standards for physicians.⁶ The FDA already has held discussions regarding the variability of diagnosis according to how bone density measurements are reported.⁷ Practitioners, however, often are leery of government intervention due to what are perceived as burdensome regulations that only serve to increase the cost of the procedures themselves. Therefore, a goal for rheumatologists should be to improve the diagnosis and treatment of osteoporosis, thereby obviating the need for government intervention.

To improve diagnosis and treatment of osteoporosis, an alternative method for reporting bone density measurements, based on a patient's future risk of fracture (absolute fracture risk), was compared with the traditional T-score report format. The new report uses a color graphic for presenting absolute fracture risk: the green zone represents normal bone, the yellow zone represents a point where preventive treatment should

be considered, and the red zone is generally considered to represent established osteoporosis, which requires more aggressive treatment (Figure).^{9,10} Fracture risk based on absolute risk can be determined because of prospective fracture databases including those of Hui et al,¹¹ Ross et al,¹² the Rotterdam study,⁸ the Study of Osteoporosis Fracture (SOF), and the National Osteoporosis Risk Assessment. These databases provide fracture statistics for different technologies including dual-energy x-ray absorptiometry (DEXA) and ultrasound, and exist to interpret machine output in terms of absolute fracture risk. Fracture risk prediction based on age and bone density was generated and validated in the Rotterdam study of 7046 persons age 55 years or older who were followed prospectively for approximately 3.8 years.⁸ Over the observation period, 110 hip fractures occurred in study subjects, and a model generated from these

events showed that the observed hip fractures correlated with age and bone density. The authors of the study published the positive sloped graph showing that low bone density and greater age were associated with the highest risk. The correlation of risk with advancing age independent of bone density also was confirmed using the SOF data.⁴ The Rotterdam and SOF databases were independently collected and validated, and each study yielded similar results—the classification of osteoporosis based on these models of fracture risk resulted in more consistency among physicians than the T-score report model.^{4,8}

The advantages of using absolute fracture risk in diagnosis and treatment guidelines are increased agreement among physicians, inclusion of age when determining fracture risk, and uniform criteria to evaluate patients whose scans have been done with different technologies. Absolute fracture risk already has been shown to be easier to interpret for both patients and physicians and is associated with improved initial compliance and follow-up with treatment.¹³

The premise of our study was that physicians should be able to agree more than 90% of the time on a patient's diagnosis, given a technology as precise as bone density measurement. The purpose of our study was to compare

the T-score and absolute fracture risk reporting systems for agreement on diagnosis and treatment.

.....
MATERIALS AND METHODS

A total of 103 physicians participated in this study, evaluating 2 classification systems for interpreting bone density reports: a standard format using T-scores and an absolute fracture risk format. The physicians recorded approximately 1850 responses using an audience response system, and consensus about diagnosis with each of these classification systems was evaluated.

Case Selection

Bone density reports from 9 case studies were selected for use in this study. All of the cases represented actual patients, all of whom were ambulatory with no underlying concomitant disease and no prior therapy for osteoporosis. Two patients were male: a 67-year-old who had a history of vertebral fracture and an 87-year-old. The other 7 female patients were 10, 49, 55, 70, 72, 80, and 90 years old. The test scans included hip DEXA (60%), spine DEXA (10%), phalangeal DEXA (10%), and calcaneal ultrasound (20%). Of the scans, 90% were from female patients, 10% were from male patients, and 5% were from pediatric patients.

Participating Physicians

Physicians who attended the National American College of Rheumatology meeting in November 1999 and who identified themselves as specialists in osteoporosis (ie, those who had their own bone density measurement equipment or who routinely read bone mineral density scans) participated in the study. Approximately 30% of the physicians reported owning bone density measurement equipment, and the majority considered themselves "expert" or "very familiar" with reading bone density reports. None of the rheumatologists had previous experience with the absolute fracture risk report.

Bone Density Report Evaluation

A room for testing was available each afternoon for 3 consecutive days during the meeting. The test was conducted over a 2-hour period using an audience response system to ensure anonymity of answers. Physicians were given a brief explanation (10 minutes or less) of the fracture risk data before receiving the test scans. Eighteen patient scans were presented to each physician for diagnosis and prescription. Nine scans used the standard format with T-score criteria as defined by the WHO (ie, normal > -1.0; osteopenia ≤ -1.0 to -2.5;

osteoporosis ≤ -2.5), and the other 9 scans from the same patients used current absolute fracture risk. Ultrasound density reports were based on the SOF database, and the DEXA density reports were based on the Rotterdam database. The patient's absolute risk for fracture was illustrated using a simple chart (Figure). As physicians were shown each patient scan, they were asked to assign the diagnosis specifically to 1 of the following 4 categories: (1) normal bone, (2) osteopenia, (3) osteoporosis, or (4) unknown. For each scan, the physicians were asked to provide a treatment recommendation based on their diagnosis. After presentation of all of the scans, physicians were asked to indicate their format preference using a 5-point Likert scale (1 = very poor, 2 = poor, 3 = fair, 4 = good, 5 = very good).

Statistical Analysis

The statistical test to assess the proportion of inter-observer agreement was the κ statistic. Perfect agreement is indicated by a κ value of 1.0, and a value of 0 indicates the amount of agreement that would be expected to occur by chance alone. The following levels of agreement among observers were used to interpret the amount of agreement: κ < 0.2 = poor agreement; κ < 0.4 = fair agreement; κ < 0.6 = moderate agreement; κ < 0.8 = good agreement; and κ > 0.81 = excellent agreement. Cases were analyzed separately for DEXA and ultrasound and also were analyzed separately for men and postmenopausal women. A 2-way *t* test with a significance level of *P* < .05 was used to determine equality between κ values. The 5-point Likert scale used to evaluate preference was analyzed by using pairwise comparisons, and the likelihood of prescribing bisphosphonates was analyzed by using the Poisson regression. All statistical analyses were performed with Stata version 6.0 software (Stata Press, College Station, Tex).

.....
RESULTS

A total of 103 physicians responded to all questions for a total of approximately 1850 responses. Diagnostic agreement scores are summarized in **Table 1**. The use of bone density reports based on absolute fracture risk criteria instead of simple T-score criteria greatly improved the concordance of expert opinion regarding the presence of osteoporosis in the individual patient for both DEXA and ultrasound scans. When given cases involving postmenopausal women, physicians achieved a κ score of 89% (indicating excellent concordance) when reading bone mineral density results in terms of absolute fracture risk. Overall diagnostic agreement for DEXA scans improved from 30%

Table 1. Concordance Among Physician Specialists Diagnosing Osteoporosis, Reported as κ Scores*

Report	Women	Men	DEXA	Calcaneal Ultrasound
T-score	0.17	0.10	0.30	0.04
Absolute fracture risk	0.89	0.66	0.68	0.63

* $P < .001$ between reports in all categories ($\kappa < 0.2$ = poor; $\kappa < 0.4$ = fair; $\kappa < 0.6$ = moderate; $\kappa < 0.8$ = good; $\kappa > .81$ = excellent). DEXA indicates dual-energy x-ray absorptiometry.

to 68% when the absolute fracture risk report was used ($P < .001$); for ultrasound scans, agreement improved from 4% to 63% with the absolute fracture risk report ($P < .001$). Cases involving men resulted in less consistency in diagnosis using either report format. For all categories (ie, women, men, DEXA, ultrasound), the T-score resulted in very inconsistent diagnosis. The best concordance using the T-score was for DEXA scans; however, the κ score was 30%, indicating only fair concordance. All other κ scores associated with T-reports were less than 20%, indicating poor concordance.

Treatment with bisphosphonates correlated far better with the absolute fracture risk report than with the T-score report ($P < .0001$). **Table 2** presents sample cases that highlight the differences in bisphosphonate prescriptions after the absolute fracture risk reports versus the T-score reports.

Physicians in this study rated the absolute fracture risk report significantly higher than the T-score report on the 5-point Likert scale (4.03 vs 3.19; $P < .001$); 72% rated the absolute fracture risk report as good or excellent compared with 44% for the T-score report ($P < .001$).

DISCUSSION

The results of this study suggest that the use of an absolute fracture risk report offers a significant advantage over a T-score report regarding consistency in diagnosis and treatment of patients at risk for osteoporosis. In women, consensus in diagnosis improved from poor with the T-score report to excellent with the absolute fracture risk report; in men, consensus improved from poor to moderate ($P < .001$ for both categories). The T-score report produced fair or poor results in all groups studied. These results are consistent with those of a previous study, in which sensitivity and specificity were found to be superior when an

absolute fracture risk format based on the SOF database was used compared with the traditional T-score format.⁴ Additionally, we found that bisphosphonates were more likely to be prescribed appropriately (ie, more therapy was prescribed for women at higher risk for fracture) when scans were presented in terms of absolute fracture risk rather than a T-score report.

Our results suggest that although there have been significant technological advances in measuring bone density, there are inherent limitations to using the T-score as a criterion for diagnosis and treatment of individuals at risk for osteoporosis.

Although T-scores are useful for epidemiologic studies, they are much less useful for individual patient management. The T-score does not account for age, an important risk factor for new fractures.³ When only the T-score is used for diagnosis, age is ignored; and 30-year-old bone looks equivalent to 75-year-old bone at the same bone mineral density, despite the logarithmic escalation of hip fracture risk with age. Because T-scores do not account for age, undertreatment of older patients and inappropriate treatment of younger patients are likely to occur, as evidenced by the results of our study. Ultimately, the consequences are mismanagement of the older patient and unnecessary medical costs for the younger patient. In contrast to T-score reports, the absolute fracture risk graphs take both age and bone mineral density into consideration.

In addition to age, the T-score fails to account for the risk factor of a history of fracture. Patients with a prior osteoporotic-related vertebral fracture have more than a 5-fold risk of another vertebral fracture.¹⁴ Using the absolute fracture risk report, all patients with a prior osteoporotic fracture are considered to be in the "red zone" and require pharmacologic treatment regardless of bone density.

Another disadvantage to using the T-score criterion for patient management is that the number of patients identified as having osteoporosis varies widely depending on the technology used (eg, DEXA, ultrasound, computed tomography).¹⁵ Prevalence of postmenopausal osteoporosis according to the -2.5 T-score cut-off is 18% by central DEXA scan, 6% by the SAHARA Clinical Bone Sonometer (Hologic Inc, Bedford, Mass), and 90% by quantitative computed tomography. Such large discrepancies across technologies occur not only because variable amounts of trabecular and cortical bone are being evaluated, but because the intrinsic differences inherent in the technologies themselves further magnify the discrepancies. Using absolute fracture risk rates based

on prospective trials offers a uniform platform for all technologies and sites, providing a single standard easily interpreted by a physician to measure osteoporosis and fracture risk. Databases from these prospective trials include fracture statistics for different technologies including DEXA, ultrasound, wrist, and phalangeal radiography. These databases are published and readily available. (See <http://www.doctorgreenwald.com/graphics/>).

It should be noted that a distinction exists among diagnosis of osteoporosis, risk assessment, and intervention thresholds. In terms of diagnosis, osteoporosis is defined in terms of bone mineral density. Therefore, the T-score, as an appropriate measure of bone mineral density, should result in 100% agreement on each case if physicians used only the T-score system for diagnosis. However, agreement in diagnosis among physicians was poor because in addition to the T-score physicians used their own judgment in each case. For example, a 10-year-old child with a T-score of -2.4 was judged normal by approximately one third of the physicians, was considered osteopenic by another one third, and was considered osteoporotic by the remaining one third. Therefore, this patient would receive 3 different diagnoses from 3 physicians, all of whom routinely read bone density reports. When the fracture risk report was used with age and bone density weighted, the physicians were more consistent in the diagnosis. However, even with the absolute fracture risk report, we did not find 100% agreement among physicians because there are factors in addition to age and previous fracture involved in evaluating and treating patients. For example, the use of cortico-steroids contributes to fracture risk independently of bone mineral density and may contribute to a decision to intervene with bisphosphonates. Conversely, other risk factors may not necessarily be useful when considering intervention thresholds. For example, a propensity to falls may contribute to risk, but may not contribute to a decision whether to intervene with a bisphosphonate.

High-quality bone density testing and interpretations are of little value unless the information they contain can be conveyed clearly to clinicians. Minimizing the variability in bone mineral density interpretation should be a priority to maximize the accuracy of bone mineral density tests. Ultimately, such efforts will reduce the unnecessary costs associated with faulty interpretation of these tests. Further studies should be

Table 2. Percentage of Bisphosphonate Prescriptions With Absolute Fracture Risk Reports Versus T-Score Reports

Age (y) and Sex	Absolute Fracture Risk (%)	Bisphosphonate Prescriptions (%)	T-Score	Bisphosphonate Prescriptions (%)
10 Female	0	0	-2.4	14
49 Female	0.02	0	-2.0	18
55 Female	0.1	4	-2.3	33
70 Female	1	2	-2.0	5
87 Male	2	90	-1.7	6

performed to evaluate the cost effectiveness of using the absolute fracture risk report versus the T-score report.

Although this study did not address patient compliance with treatment, a previous study by Greenwald et al showed that initial compliance and physician follow-up were greater for patients who received their bone mineral density results in terms of absolute fracture risk compared with patients who received a standard T-score report.¹³ One limitation of both of these studies, however, is that neither addressed the issue of long-term patient compliance and the cost effectiveness of treatment. Although studies have shown improved compliance when patients understand their bone mineral density results, there are no long-term, prospective studies that have evaluated whether using absolute fracture risk to report bone mineral density results can actually prevent fractures and improve treatment effectiveness. However, it would be reasonable to expect that greater patient understanding and consistency in diagnosis and treatment by physicians would improve patient compliance, thereby reducing the economic burden of osteoporosis and improving the overall medical care for patients at risk for osteoporosis.

In conclusion, our data underscore the importance of developing a report format that physicians can use to accurately diagnosis osteoporosis and thereby direct appropriate interventions. Using absolute fracture risk in diagnosis and treatment guidelines confers several advantages compared with T-score report, including increased agreement among physicians, the inclusion of age as a risk factor, and uniform guideline criteria to evaluate patients whose bone mineral density scans are done with different technologies.

REFERENCES

1. **National Osteoporosis Foundation.** Osteoporosis: review of the evidence for prevention, diagnosis, and treatment and cost effective analysis. *Osteoporosis Int.* 1998;8:S1-S88.

CLINICAL

2. **Lindsay R.** The burden of osteoporosis: Cost. *Am J Med.* 1995; 98(2A):9S-16S.
3. **Kanis J, Johnell O, Oden A, Jonsson B, De Laet C, Dawson A.** Risk of hip fracture according to the World Health Organization criteria for osteopenia and osteoporosis. *Bone.* 2000;27:585-590.
4. **Lu Y, Genant HK, Shepherd J, et al.** Classification of osteoporosis based on bone mineral densities. *J Bone Miner Res.* 2001;16:901-910.
5. **Ray NF, Chan JK, Thamer M, Melton LJ III.** Medical expenditures for the treatment of +osteoporotic fractures in the United States in 1995: report from the National Osteoporosis Foundation. *J Bone Miner Res.* 1997;12:24-35.
6. **Medicare program; Medicare coverage of screening mammography—HCFA.** Final rule. *Fed Regist.* September 30, 1994;59(189): 49826-49834.
7. **Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health, Office of Device Education.** Meeting of the Radiological Devices Panel, Monday, May 17, 1999. Copies of the report can be obtained from Miller Reporting Company, 135 8th Street SE, Washington, DC, 2003 (202-546-6666).
8. **De Laet C, VanHorst B, Burger H.** Hip fracture prediction in elderly men and women: validation of the Rotterdam study. *J Bone Miner Res.* 1998;13:1587-1593.
9. **Greenwald L, Barajas K, White-Greenwald M.** New bone density report form: improved diagnosis. *Arth Rheum.* 2000; 43: S199.
10. **Greenwald L, Barajas K, Greenwald MW.** The who, why, and how to use bone density reports: DEXA and ultrasound. *J Musculoskel Dis.* 2002;19:452-459..
11. **Hui SL, Slemenda CW, Johnston CC.** Baseline measurement of bone mass predicts fracture in white women. *Ann Intern Med.* 1989;111:355-361.
12. **Ross P, Huang C, Davis J, et al.** Predicting vertebral deformity using bone densitometry at various skeletal sites and calcaneus ultrasound. *Bone.* 1995;16:325-332.
13. **Greenwald B, Bardwell A, Malinak J, Rude R, Silverman SL, White-Greenwald M.** New bone density report format influences patient compliance in filling prescriptions for osteoporosis. *Clin J Womens Health* 2002;2:13-18.
14. **Ross PD, Davis JW, Epstein RS, et al.** Pre-existing fractures and bone mass predict vertebral fracture incidence in women. *Ann Intern Med.* 1991;114:919-923.
15. **Faulkner KG, Von Stetten E, Miller P.** Discordance in patient classification using T-scores. *J Clin Densitom.* 1999;2:343-350.