Osteoporosis Int (1993) 3:265–270

Osteoporosis
International

Original Article

Revisable Criteria for Vertebral Deformity

K. M. Davies, R. R. Recker and R. P. Heaney

Center for Hard Tissue Research, Department of Medicine, Creighton University, Omaha, Nebraska, USA

Abstract. In order to study vertebral fractures in various study populations, we earlier prepared a database of vertebral dimensions derived from spinal radiographs of 191 normal women seen regularly over 25 years. In this report we have expanded the range of measurements to include vertebral levels T3 to L5. We report means and standard deviations on anterior and posterior heights, on wedge shape and on heights relative to adjacent vertebrae. When one or both of the latter two quantities are 'far' below the mean, a vertebra is called deformed. We also describe a more flexible way of expressing damage using the number of deformed vertebrae, the degree of deformity of individual vertebrae, or the total damage to the entire spine. In assessing damage we use criteria for deformity adjusted to the limits detected by an experienced diagnostician, replacing an earlier approach based on 95% probability limits of normal variation. The normal women from whom these variations are ascertained are a low-prevalence group with respect to vertebral deformity, with prevalence of 2.8%. When the criteria developed from these women were applied to a moderate-prevalence group (37%) the model had a sensitivity of 97%, a specificity of 89% and an accuracy of 92% as regards the identification of subjects with damaged vertebrae. When used epidemiologically for a moderate-prevalence group the model has a known overestimation of 15%. The model is compared with other schemes for identifying vertebral deformities.

Keywords: Osteoporosis; Vertebral deformity; Vertebral fracture

Introduction

The assessment of vertebral fractures in individuals and the determination of fracture prevalence and incidence in populations are key tools in the clinical study of osteoporosis. Vertebral fracture in osteoporosis is often atraumatic, and while the resulting deformity can assume a variety of shapes, it can usually be characterized by either wedging (anterior collapse of the vertebral body) or compression (crushing or pancaking of the body) or both. There is no gold standard of deformity. The best standard is the human expert who is called upon to make an anatomical judgment more subtle and complex than noticing an outright break in a long bone. Ordinarily, vertebral fracture is determined by the examination of lateral spine radiographs by an expert, often a clinician with some expertise in osteoporosis.

In clinical research it is generally considered desirable to employ an objective means of assessing vertebral fracture. Since there is no true gold standard, it seemed to us reasonable to fine-tune a system of radiographic assessment of vertebrae until it produced results comparable to those of an experienced clinician. Issues of (1) reliability and consistency in cross-sectional assessment and (2) the detection of serial change in longitudinal studies have prompted us to develop a morphometric system for determining vertebral deformity. The development of our system has been guided by two objectives:

1. To use a simplified, intuitive model of what a clinician considers in evaluating spinal radiographs.
2. To use explicit, empirically based standards.

In earlier work [1] we applied our model to a group with high prevalence of vertebral deformity and obtained nearly perfect agreement with a clinician's

---

Correspondence and offprint requests to: K. Michael Davies, PhD, Center for Hard Tissue Research, 601 No. 30th St., Suite 5740, Omaha, NE 68131, USA.

evaluations. Subsequent application of the model of a low-prevalence group and a moderate-prevalence group yielded far too many damage calls. This has been the experience of other workers using equivalent methods [2] and reflects the fact that, if 'normal' is taken as the 95% or 99% probability range for vertebral dimensions in normal women, any given lateral spine film set will offer 12 or 13 chances of exceeding these limits. Thus, even with the 99% limits, random chance alone predicts that about 12% of normal women will have one or more vertebrae outside these limits. Furthermore, there is a need to adjust criteria to populations in which vertebral deformities are low or moderate in number and degree.

In this paper we present an empirical method for refining our criteria in accordance with what would be detected as deformity by an experienced reader viewing films from a low-prevalence group. In addition, our results are brought into reasonable agreement with a clinician's judgment for a moderate-prevalence group. An advantage of this model is that the method of adjusting the criteria is intuitive, entirely explicit, and capable of further refinement.

Materials and Methods

Human Subjects

The low-prevalence group itself is the source of our normal data and of the deformity and damage criteria. Lateral spine radiographs have been taken every 5 years over the past 25 years as a cohort of 191 normal, active, white, perimenopausal women employed in a wide range of jobs. These women have been described in detail elsewhere [3–5].

The moderate-prevalence group was a cohort of women recruited for a study of ultrasound transmission velocity and osteoporotic bone fragility [6]. The group consisted of 92 women aged 50 and above, on whom lateral spine radiographs were taken.

Radiographs

Standard lateral thoracic and lumbar spine radiographs were taken on each subject using a tube-to-film distance of 40 inches (100 cm). The X-ray tube was centred at about T7 for the thoracic films and at about L3 for the lumbar films. Backlit tracings were made on each set of radiographs. Anterior and posterior heights were determined by marking the tracings on a Bioquant digitizing tablet (R & M Biometrics, Nashville, TN).

Data Analysis

A set of measures was developed to express the shape and size of the vertebrae. The wedge variable, W, is defined as the difference between anterior and posterior height divided by the posterior height:

$$W_i = (A_i - P_i)/P_i \quad \text{for } i = \text{T1 to L5} \qquad (1)$$

and the $\{A_i\}$ and the $\{P_i\}$ are the anterior and the posterior heights of the vertebrae.

The crush variable, C_i, resembles the wedge variable except that it is based upon posterior heights of the adjacent vertebrae. The crush variables are defined by

$$C_i = (P_i - P_{i+1})/P_{i+1} \quad \text{for } i = \text{T1 to L3} \qquad (2)$$

except for L4 and L5 where:

$$C_{L4} = P_{L4}/(P_{L2} + P_{L3} + P_{L4}) \qquad (3)$$

$$C_{L5} = (P_{L5} - P_{L4})/P_{L4} \qquad (4)$$

A formula such as Eq. 2 can only be used for 16 of the 17 vertebrae: one vertebra must have some other measure of relative posterior height. Thus we chose to define C_{L4} differently. We chose to anchor the $\{C_i\}$ at L4 because it is an easy vertebra to identify and measure accurately. An alterative would be to consider the ratio of T1 to T2, T2 to T3, . . . , L4 to L5, going down the spine and then L5 to L4, . . . , T2 to T1, going up the spine. The difficulty is that T1 and T2 are usually unmeasurable, T3 is often so, and L5 is unmeasurable often enough to warrant giving L4 some reasonably sure means of being assessed. Further treatment of the W and C variables in our normal population is given in the Results section.

The physical content of these formulas can be seen by example. Consider defining a wedge deformity for T5. The average anterior height for T5 is about 2.18 cm, and the average posterior height is about 2.42 cm. Thus, the average ratio of anterior to posterior height is about 0.9. The average difference between anterior and posterior height, divided by posterior height, namely the average W for T5 in the normal population, is about −0.1 (see Table 1) with an SD of 0.05. If we take as our criterion for wedge deformity any W value more than 2 SD below the mean, then we are saying that deformity of T5 exists if:

$$W < -0.1 - (2 \times 0.05), \text{ or } W < -0.2 \qquad (5)$$

This is the same as saying that deformity of T5 is present when the anterior height of T5 is less than 0.8 of the posterior height. The ability to set the number of SD below the average, in this case 2, as the cutoff for normality (or the beginning of deformity) is the adjustable feature of the model (see below). Now, to generalize this and to express these relationships symbolically, the cutoff value, W_{ABNi}, the critical value for an abnormal wedge deformity at vertebra i, can be set as follows:

$$W_{ABNi} = W_{Xi} - (W_{Ei} \times W_{SDi}) \quad \text{for } i = \text{T1 to L5} \qquad (6)$$

where W_{Xi} signifies the mean value for W_i (−0.1 in the example), W_{Ei} the number of SD used as a criterion for abnormality (2 in the example) and W_{SDi} the SD for W_i (0.05 in the example).

The notation for the adjustable feature of the model, W_{Ei}, is intended to suggest how far (Extent)W_i must be in units of SD from the mean in order to be called abnormal. These variables, W_{ABNi}, are level-specific (hence the subscript i) and use only the empirically derived values found in normal women for *that specific vertebral body*. As shown elsewhere [1], a single criterion for W applied to all vertebra (e.g. $A/P < 0.85$ or $W < -0.15$) will miss real deformities at some levels and overcall deformity at others.

The cutoff value, or critical value, for an abnormal compression deformity uses the same level-specific approach. However, it appears more complicated because it involves comparing a vertebra with both the one below it and the one above it. For example, we can focus on T8. We do this first by calculating the crush variables (C_{T7} and C_{T8}) for T7 and T8 according to Eq. 2:

$$C_{T7} = (P_{T7} - P_{T8})/P_{T8} \tag{7}$$

$$C_{T8} = (P_{T8} - P_{T9})/P_{T9} \tag{8}$$

If T8 is crushed, then P_{T8} will be smaller than expected. From Eq. 7 it follows that C_{T7} will be larger than expected, and from Eq. 8 that C_{T8} will be smaller than expected. Thus the cutoff value, or critical value (analogous to W_{ABNi}), for an abnormal compression deformity of T8 is actually two values: one to define how large C_{T7} might be, and one to define how small C_{T8} might be. These values, C_{Hi-1} and C_{Li}, are given as:

$$C_{Hi-1} = C_{Xi-1} + (C_{Ei-1} \times C_{SDi-1}) \tag{9}$$

and

$$C_{Li} = C_{Xi} - (C_{Ei} \times C_{SDi}) \tag{10}$$

In Eq. 9, C_{Xi-1} signifies the normal mean value of C for the vertebra above T8 (T7 in this case), C_{Ei-1} signifies the number of SD used as a criterion for abnormality and C_{SDi-1} signifies the SD for C_{Xi-1}. In Eq. 10, C_{Xi} signifies the normal mean value of C for T8 (the vertebra in question), C_{Ei} signifies the number of SD used as a criterion for abnormality and C_{SDi} signifies the SD for C_{Xi}. A compression deformity exists for T8 or any given vertebra, C_i, when both:

$$C_{i-1} > C_{Hi-1} \text{ and } C_i < C_{Li} \tag{11}$$

In the case of compression of T8 in our example, these formulas would read:

$$C_7 > C_{HT7} \text{ and } C_8 < C_{LT8} \tag{12}$$

The conjunction 'and' means that P_i is low compared with both its neighbors – an assumption suitable for a moderate-prevalence population. In a high-prevalence population the connector could be relaxed to an inclusive 'or.' As noted above for W_{Ei}, the notation for the adjustable features, C_{Ei-1} and C_E, of the model is intended to suggest how far (Extent)C_{i-1} and C_i must be (in units of SD from the mean) in order to be called abnormal. Further, C_{Hi-1} and C_{Li} are level-specific as in the case for W_{ABNi}.

Damage Values

The foregoing approach detects the presence of deformity but not its severity. In order to compare wedge and crush damage in the same vertebra, to compare damage between vertebrae and to compare damage between subjects we introduce the following damage variables.

If a wedge deformity is identified for a given vertebra, the degree of damage (W_D) can be expressed as:

$$W_{Di} = (W_{ABNi} - W_i)/W_{SDi} \tag{13}$$

in which the difference between the cutoff value and the wedge value for a given vertebra is divided by the SD. The variables W_{ABNi}, W_i and W_{SDi} are defined above.

If a crush deformity is identified for a given vertebra, the degree of damage (C_D) can be expressed as either:

$$C_{Di} = (C_{Li} - C_i)/C_{SDi} \tag{14}$$

or

$$C_{Di} = (C_{i-1} - C_{HI-1})/C_{SDi-1} \tag{15}$$

whichever is the larger quantity. This variable is analogous to W_D. The variables C_{Li}, C_i, C_{SDi}, C_{i-1}, C_{Hi-1} and C_{SDi-1} are defined above.

Note that W_i is always more negative than W_{ABNi} when a deformity is called, and thus the quantity ($W_{ABNi} - W_i$) is always positive. The same applies to the quantity ($C_{Li} - C_i$) for a crush call. The quantity ($C_{i-1} - C_{Hi-1}$) is always positive when a deformity is called.

W_{Di} and C_{Di} are relative values and can be summed to measure the total deformity of a given vertebra; finally, the totals for several vertebrae of different sizes can be summed for an overall damage rating of a given spine.

Revision Procedure

The objective of the revision procedure was to find settings for the parameters $\{W_{ABNi}\}$ and $\{C_{ABNi}\}$ corresponding to the point where an experienced clinician begins to detect deformity in vertebrae. With $\{W_{Ei}\}$ and $\{C_{Ei}\}$ all set arbitrarily to the value 2, that is, the lower limits of normal set at 2 SD below the means, analyses of the radiographs from the low-prevalence group (normals) were performed. Damage values were calculated for the vertebrae and the sets of radiographs were ordered in descending fashion from those displaying vertebrae with the most damage. The radiographs were presented to the clinican (R.R.R.) in this order but without other prompting. The clinician called deformities as he saw them until he reached radiographs in which he no longer saw damage.

Results

Table 1 gives the mean values of vertebral heights determined from the normal group and the mean values of wedge and crush variables. For each woman in the group, vertebral height data from all her visits were averaged to damp the measurement errors inevitably introduced by small angulation, distance and parallax problems; these in turn were averaged to give the group means.

Table 1. Normal values for vertebral heights and vertebral shape parameters

Vertebra	Anterior	Posterior	Wedge	Crush	*n*
T3	2.105±0.162	2.252±0.177	−0.064±0.049	−0.033±0.062	136
T4	2.150±0.163	2.337±0.170	−0.079±0.051	−0.032±0.046	180
T5	2.180±0.162	2.419±0.170	−0.097±0.050	−0.030±0.043	188
T6	2.205±0.162	2.496±0.180	−0.115±0.053	−0.029±0.044	191
T7	2.290±0.152	2.572±0.174	−0.109±0.045	−0.009±0.044	191
T8	2.334±0.165	2.598±0.168	−0.100±0.053	−0.026±0.043	191
T9	2.457±0.175	2.671±0.176	−0.078±0.054	−0.048±0.042	191
T10	2.634±0.177	2.807±0.193	−0.060±0.051	−0.072±0.032	191
T11	2.788±0.200	3.021±0.218	−0.076±0.050	−0.073±0.064	191
T12	3.043±0.270	3.257±0.271	−0.067±0.055	−0.050±0.060	191
L1	3.280±0.226	3.453±0.241	−0.048±0.059	−0.011±0.047	191
L2	3.438±0.227	3.495±0.214	−0.014±0.061	−0.016±0.041	191
L3	3.557±0.225	3.554±0.240	0.002±0.062	0.022±0.045	191
L4	3.638±0.232	3.484±0.235	0.047±0.067	0.331±0.009	191
L5	3.605±0.271	3.195±0.280	0.134±0.097	0.082±0.069	191

Values are given in centimeters, expressed as mean ± SD.

Note that the maximum anterior height is found at L4 and the maximum posterior height at L3. The wedge parameter varies with vertebral level in such a way as to produce the S-shape of the spine: it is most negative in the mid-thoracic region, reaches zero at L3, and is very positive at L5. That is, a forward wedge shape in *normal* in the mid-thoracic area, and a backward wedge shape is *normal* in L4 and L5.

The means and SDs of the W and C parameters give us the central values and the dispersion of these values as found in normal women. The means and SDs also give us the natural units we need to judge deformity. 'Wedge' is not a bad name for a normal shape parameter, but because 'relative posterior height' is such a mouthful it is irresistible to call the C parameter a 'crush' even though it is only values beyond certain extremes which indicate an actual deformity.

To interpret the normal values of the crush parameter, we begin at L4 and go upward: typically the posterior height of L3 is greater than that of L4, but above that the posterior heights diminish successively. The mean crush parameter value for L4, the anchoring vertebra mentioned above, differs from the others because it is a fraction of a sum rather than a relative difference.

Revised Criteria

We set the damage ratings of the low-prevalence group's radiographs at zero at that point where the clinician no longer detected deformity, by asking the question 'What values of E could we choose that would reset the ABN values so that our set of normal radiographs would have zero calls and therefore no damage?' The answer lay in seeing how far (in terms of SDs) the W and C values departed from the means at the point where the clinician stopped seeing deformities. The results we obtained for the particular vertebrae in question, regardless of which they were, we generalize to all vertebral levels as:

$$W_{Ei} = 2.5 \tag{16}$$

and

$$C_{Ei} = 4.05 \tag{17}$$

The moderate prevalence group was analyzed with these settings for W_{Ei} and C_{Ei} and the results compared with the clinician's calls, which were taken as the 'gold standard'. In addition, even though the low-prevalence group was used to derive these settings, the criteria with the revised settings were used to re-analyze the data from the low-prevalence group as a cross-check of the model, and the results were also compared with the clinician's calls. The results are presented in terms of subjects and of vertebrae in Table 2. According to the clinician the prevalence of vertebral deformity in terms of subjects in the low-prevalence group was 2.8% (4 of 141). These women had one or two deformed vertebrae each. In the moderate-prevalence group 37% of the subjects (34 of 92) had an average of four deformed vertebrae each. Definitions for sensitivity, specificity and accuracy are given in the notes to Table 2. Our model is very *sensitive* (97%) in identifying *patients* with deformed vertebrae. However, it is less sensitive (74%) in identifying *vertebrae* that are deformed. This is due to the relatively low prevalence of deformity in the set of vertebrae (134/1099 = 12.2%) as compared with the set of people (34/92 = 37%). Nevertheless, the model has very high accuracy (96%) and specificity (99%) in identifying deformed vertebrae.

Discussion

The chief practical interest of the present work is the clear-cut way it shifts the criteria for deformity from the 'limits of natural variation' of other models, including the earlier version of our own model, to the 'limits of diagnostic perception' of an experienced clinician. The goal is for the algorithm to present results consistent with the human reader, who remains the 'gold standard.' Our revised model does not obtain perfect

Table 2. Prevalence data using revisable vertebral deformity criteria

	Moderate-prevalence group	95% confidence limits	Low-prevalence group	95% confidence limits
By subject				
n	92		141	
Mean age (SD)	67.3 (10.1)		40.3 (2.86)	
Subjects with deformity	34		4	
Sensitivity	97.1%	(100.0, 93.6)	100.0%	(39.8, 100.0)
Specificity	89.7%	(83.7, 95.9)	96.4%	(90.0, 98.7)
Accuracy	92.4%	(87.0, 97.8)	96.5%	(91.3, 98.7)
By vertebra				
n	1099		1692	
Deformed vertebrae	134		6	
Sensitivity	73.9%	(71.3, 76.5)	100.0%	(54.1, 100.0)
Specificity	99.3%	(98.8, 99.8)	99.6%	(99.3, 99.9)
Accuracy	96.2%	(95.0, 97.3)	99.6%	(99.3, 99.9)

The numbers of calls of deformities are those made by the clinician; these are taken to be the standard. Sensitivity is the percentage of correct positive calls made by the model compared with all the clinician's positive calls. Specificity is the percentage of correct negative calls made by the model compared with all the clinician's negative calls. Accuracy is the percentage of true calls compared with all calls.

agreement with the clinician's calls but there is a known level of disagreement. Furthermore, the model could be used to determine the perception limits of different clinicians to explain disagreements between them in assessing deformities in the same subjects.

The principal theoretical interest of this work is that we present continuous measures of damage in addition to counts of damaged vertebrae. These measures allow us to compare damage due to wedging, damage due to crushing, damage in different regions of the spine, and damage in different individuals or in the same individual at different times.

Our earlier work [1] was based on measurements for T7 through L4. The results for vertebral and for wedge parameters in Table 1 do not differ from earlier results for those vertebrae. The crush parameters do differ because they are now based on L4 rather than T7. W and C, means and standard deviations, remain level-specific in our model, as do the criteria for deformity.

The new features of the model, the variables W_E and C_E, were given values we determined by the limits of an expert's perception of deformity.

It might seem that the crush criterion in this model is too stringent because two conditions must be met to call a deformity so that it would be harder for it to detect adjacent crush deformities. We do not think changing this to achieve an increase in sensitivity (at the expense of a decrease in specificity) is desirable. The adjacent deformities found by the clinician in the low-prevalence group were wedge deformities, and in the moderate-prevalence group the single subject with the false negative call had no adjacent deformities.

Our model takes a straightforward approach to normalization by dividing differences between heights by one of the heights, creating a normalized ratio. In a study of morphometric definitions of vertebral deformity, Smith-Bindman et al. [7] found that parameters based on ratios produced the best sensitivity and specificity for two arbitrarily set levels of 15% and 20% of deformity. These authors acknowledge the lack of a gold standard and the problem of reconciliation with clinician readings. We turn the problem around by using the clinician as the gold standard to set the limits of recognizable deformity.

In a recent study on classification of vertebral fractures, Eastell and others [8] used wedge and crush parameters similar to ours and, in addition, a biconcavity parameter in which the difference between middle height of the vertebra and posterior height is divided by posterior height. Data were obtained from a group of normal individuals and applied to a moderate-prevalence group. When the criteria for abnormality were set at 2.5 SD for all three parameters, they found 37% of the women had vertebral fractures of some type. When the criteria were set at 3 SD for all three parameters, the figure was 21%. Their analysis used numbers derived from the normals to set the criteria for deformity arbitrarily in terms of SD, while our analysis adjusted the criteria for deformity in terms of SD to match what a clinician recognizes as radiographic deformity. Their approach did not permit a way to set criteria at different values of SD for different parameters. In contrast, our method sets different SDs for different parameters in order to fit the gold standard of the experienced clinician's reading.

Minne and others [9] have developed a different approach to detecting and quantifying vertebral deformities, and they have recently [2] compared their approach with the early version of our method, and with other published methods [10,11]. Minne's relative height value method divides the anterior, middle and posterior heights of T5 through L5 by those of T4. The cutoffs for fracture are set equal to the lowest values of the height ratios from his normative group after the lowest 3% have been discarded. If any of the three ratios falls below its cutoff, then a deformity for the vertebra is called. This gives a count of deformed vertebra. In addition, Minne and co-workers quantify damage to the entire spine by summing the amounts by which a subject ratios fall below the cutoffs. Since all the ratios have T4 in the denominator, they become larger as one progresses down the spine. In addition, this strategy results in the cutoff values falling farther and farther below the mean as one progresses down the spine. Thus it underemphasizes deformities in the upper spine compared with the lower spine.

The chief advantage of Minne's method is that it does not depend on ratios of adjacent posterior heights and thus would theoretically be better able to detect a consecutive series of crush deformities. An objection that the method depends on T4 not being damaged can be set because it can substitute an undamaged vertebra if T4 is deformed. Two contrasting features of the Minne method and our own are:

1. The Minne cutoffs are set in a fashion which depends on the extremes of a normative group. By contrast, we use the limits of a clinician's perception to set our cutoffs in terms of our normative group's means and SDs.
2. The Minne method requires increasing departure of cutoff values from the means at lower spine levels. However, a short person does not have vertebral heights that are all proportionately short. The range of variability in individuals over the spine is remarkable. When we took the vertebral heights of individuals in our normative group as a set of Z-scores compared with the group means, the average difference between the largest and the smallest Z-scores in the individual is 2!

The data in Table 2 show the results for our model when applied to the moderate-prevalence group. We are disinclined to alter the E values to make the vertebral sensitivity (74%) for the moderate-prevalence group look better, since our model is a tool for scientific investigation rather than patient management. However, even in a research setting we are following a conservative policy in identifying subjects with deformities. We do so in order to avoid errors in labeling subjects as osteoporotic, with the attendant changes in attitude and behavior [12]. However, the point of our approach is that the criteria are explicitly adjusted. In a given application it may be desirable to improve the model's sensitivity by adjusting the E values downward.

Note also that the rule 'the higher the prevalence, the better the sensitivity' is not really violated by the 100% sensitivities for subjects and vertebrae in the low-prevalence group. In these cases the 95% confidence intervals calculated exactly for the binomial distribution on small sets (4 people out of 4 called positive, 6 vertebrae out of 6) are very broad.

One cause of low sensitivity is the difficulty the model has in detecting a series of adjacent crush fractures in an individual. This could be improved by relaxing the logical 'and' in the crush criterion to the logical 'or'. Minne's system is better here because it can detect every one of a series of adjacent crush fractures.

In conclusion, our vertebral deformity model, with its adjustable cutoffs, produces results in good agreement with a clinician's calls in low- and moderate-prevalence groups. As is common with tests, the higher the prevalence in a test population, the better the sensitivity. The test is feasible in an epidemiological study of a low-prevalence group because the excellent specificity means one need only scan the low number of positive calls to distinguish true from false. Scanning is not necessary in a moderate-prevalence group provided one is willing to accept a 15% overestimate of prevalence. In fact with a given level of prevalence, one can adjust the sensitivity and specificity to acceptable levels with due regard for subject labeling problems.

Acknowledgements. This work was supported in part by NIH grants AR 07912 and AR 39221.

References

1. Davies KM, Recker RR, Heaney RP. Normal vertebral dimension and normal variation in serial measurement of vertebrae. J Bone Miner Res 1989;4:341–9.
2. Sauer P, Leidig G, Minne HW, et al. Spinal deformity index (SDI) versus other objective procedures of vertebral fracture identification in patients with osteoporosis: a comparative study. J Bone Miner Res 1991;6:227–38.
3. Heaney RP, Saville PD, Recker RR. Calcium absorption as a function of calcium intake. J Lab Clin Med 1975;85:881–890.
4. Heaney RP, Recker RR, Saville PD. Calcium balance and calcium requirements in middle-aged women. Am J Clin Nutr 1977;30:1603–11.
5. Heaney RP, Recker RR. Distribution of calcium absorption in middle-aged women. Am J Clin Nutr 1986;43:299–305.
6. Heaney RP, Avioli LV, Chesnut CH, III, et al. Osteoporotic bone fragility. JAMA 1989;261:2986–90.
7. Smith-Bindman R, Cummings SR, Steiger P, et al. A comparison of morphometric definitions of vertebral fracture. J Bone Miner Res 1991;6:25–34.
8. Eastell R, Cedel SL, Wahner HW, et al. Classification of vertebral fractures. J Bone Miner Res 1991;6:207–15.
9. Minne HW, Leidig G, Wuster C, et al. A newly developed spine deformity index (SDI) to quantitate vertebral crush fractures in patients with osteoporosis. Bone Miner 1988;3:335–49.
10. Hedlund LR, Gallagher JC. Vertebral morphometry in diagnosis of spinal osteoporosis. Bone Miner 1988;5:59–67.
11. Melton LJ III, Kan SH, Frye KM, et al. Epidemiology of vertebral fractures in women. Am J Epidemiol 1989;129:1000–11.
12. Sackett DL, Haynes RB, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. Boston: Little, Brown, 1985:152–3.

Received for publication 20 December 1991
Accepted in revised form 8 February 1993