Osteoporosis Int (1993) 3:138–147
© 1993 European Foundation for Osteoporosis

Osteoporosis International

Original Article

The Assessment of Vertebral Deformity: A Method for Use in Population Studies and Clinical Trials

E. V. McCloskey[1], T. D. Spector[2], K. S. Eyres[1], E. D. Fern[1], N. O'Rourke[1], S. Vasikaran[1] and J. A. Kanis[1]

[1]WHO Collaborating Centre for Metabolic Bone Disease, University of Sheffield, Sheffield, and [2]Department of Environmental and Preventive Medicine, St Bartholomew's Hospital Medical College, London, UK

Abstract. The absence of specific criteria for the definition of vertebral fracture has major implications for assessing the apparent prevalence and incidence of vertebral deformity. Also, little is known of the effect of using different criteria for new vertebral fractures in clinical studies. We therefore developed radiological criteria for vertebral fracture in women for assessing both the prevalence and the incidence of vertebral osteoporosis in population and in prospective studies and compared these with several other published methods. Normal ranges for vertebral shape were obtained from radiographs in 100 women aged 45–50 years. These included ranges for the ratios of anterior/posterior, central/posterior and posterior/predicted posterior vertebral heights from T4 to L5. The predicted posterior height was calculated from adjacent vertebrae. In contrast to other methods, our definition of fracture required the fulfilment of two criteria at each vertebral site, and was associated with a lower apparent prevalence of fracture in the control women due to a lower false positive rate. The prevalence and incidence of vertebral deformity using different criteria were then compared in a series of women with skeletal metastases from breast cancer in whom radiographs were obtained 6 months apart. The prevalence of vertebral deformity and the specificity for deformity varied markedly with differing criteria. Using a cut-off of 3 standard deviations the prevalence of vertebral deformity in the women with breast cancer was 46%. Using other methods, the prevalences of deformity ranged from 33% to 74%. Over a 6-month interval 25% of patients with breast cancer sustained 61 deformities using our method, of which only 8% resulted from errors in reproducibility. The number of patients sustaining new deformities was increased twofold when assessed by other methods (45%–53%), but errors of reproducibility may have accounted for 21% of the new deformities. The magnitude and distribution of these errors have important implications for the apparent therapeutic efficacy of agents in clinical trials of osteoporosis. The rapid semi-automated technique for assessing vertebral deformities on lateral spine radiographs that we have developed has a high specificity, and reduces the impact of errors of reproducibility on estimates of prevalence and incidence. The method should prove a value in assessing vertebral deformity both in population studies and in prospective clinical trials.

Keywords: Incidence; Morphometry; Prevalence; Sensitivity; Specificity; Vertebral fracture

Correspondence and offprint requests to: Prof. J. A. Kanis, WHO Collaborating Centre for Metabolic Bone Disease, Department of Human Metabolism and Clinical Biochemistry, University of Sheffield, Sheffield S10 2RX, UK.

Introduction

Involutional osteoporosis is characterised by bone loss, disruption of trabecular architecture and an increased risk of fragility fracture. The wrist, hip and vertebrae are the most frequent sites of fracture [1]. In contrast to appendicular fractures [2,3] the epidemiology of vertebral fracture is not well documented for several reasons.

Firstly, the presence of a vertebral fracture depends upon the assessment of changes in vertebral shape, whereas the presence or absence of an appendicular fracture is usually obvious. In addition, there is uncertainty about the clinical significance or morbidity of vertebral osteoporosis.

Several techniques have been developed to quantitate vertebral deformities, usually on the basis of assessing the heights of vertebral bodies [4-16]. The strength of all these approaches is that, because of their objective nature, they reduce bias in the interpretation of radiographs. In addition, they can detect minor vertebral deformities, and this may increase the apparent frequency of vertebral events which means that fewer subjects or patients require to be studied to assess the natural history of osteoporosis or to assess the efficacy of interventions. A major disadvantage is that minor vertebral deformities may have less clinical relevance [17]. Furthermore, the smaller the clinical deformities that are measured, the more likely it is that errors of long-term reproducibility confound the data.

The assessment of sensitivity and specificity has been limited by the lack of any well-validated criteria for discriminating normal and deformed vertebrae. The lack of standardization probably contributes to the greater than 10-fold difference observed in the apparent prevalence of vertebral osteoporosis in the aging community [14,18–22].

It is also clear from clinical trials that the estimated incidence of new vertebral fractures in postmenopausal osteoporosis differs markedly (from 6 to 83 fractures/ 100 patient-years) [23–27]. As in the case of variable prevalence, it is likely that poor sensitivity and/or specificity account for a proportion of these differences.

For these reasons, we wished to develop a robust radiographic method for the assessment of vertebral deformity which would be suitable for epidemiological studies and could also be applied to study the incidence of fracture.

Methods

Patients and Normal Controls

The normal ranges of vertebral shape were obtained from spine radiographs in 100 normal women (aged 45–50 years) selected randomly from the age–sex register of a general practice population and invited for screening with a response rate of 79%. None had a history of back pain or osteoporotic fracture at vertebral or non-vertebral sites and the prevalence of vertebral deformity was presumed to be 0. The criteria developed for the presence of vertebral deformity derived from these controls were then applied to assess the prevalence of vertebral deformity in 163 women with skeletal metastases from breast cancer. These patients (mean age 59 years, range 30–75 years) were chosen since they were likely to have a high prevalence and incidence rate. Of these, 122 were studied again 6 months later to assess the incidence of vertebral deformity.

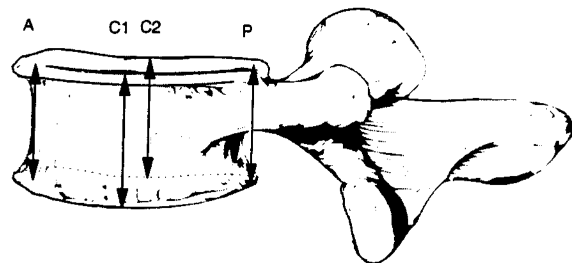

Fig. 1. Measurement of the anterior, central and posterior heights were made as shown. Points were not marked on the films prior to measurement. Anterior and posterior osteophytes were excluded from measurement.

Vertebral Heights

Lateral radiographs of the thoracic and lumbar spine were obtained at a standard target-to-film distance of 105 cm. Thoracic films were centred on T9 and lumbar films on L3. The radiographs were positioned on a rear-illuminated digitizing tablet (Kontron). Using a fine cross-wire cursor, the anterior, posterior and central heights of each vertebra from T4 to L5 were measured by a trained technician without marking the films (Fig. 1). Where the projection was oblique, the central height was calculated as the mean of the two 'lateral' central heights (C1 and C2), except in cases of obvious central compression, where the minimum height (C3) between the superior and inferior concave borders was recorded. Anterior and posterior osteophytes were ignored. The data were captured automatically by a purpose-written program (XR2, © Applied Systems and Peripherals, UK) and transferred to a database for future analysis. The technique is rapid and all the measurements can be made and data recorded from a single spine in 4 min.

For sequential evaluation of the same patient, the program allowed recall of previous data and a schematic comparison of the current record with any previous film. Each vertebra was represented as a rectangle with the anterior, central and posterior heights of the rectangle scaled to the actual measurements for that vertebra. This enabled the confident assessment of vertebral levels so that previous deformities were identfied in the same vertebrae on both films. It ensured, for example, that a previous deformity recorded at T7 would not be recorded subsequently as a new deformity at T8 due to the incorrect identification of the vertebra.

Normal Ranges for Vertebral Morphometry

In the 100 normal women examined there was a progressive increase in posterior vertebral heights from the upper thoracic to mid-lumbar spine, with a subsequent

decrease in the lower lumbar spine (Table 1). From the ratios between posterior vertebral heights it is possible to predict in an individual the posterior vertebral height of one vertebra from the posterior height of adjacent vertebrae. For example, if the measured posterior height of T7 is 27.20 mm, then from the mean posterior heights (Table 1) the predicted posterior height of T5 would be (27.2 × 25.61)/26.82 = 25.97. If the actual height measured at T5 was 25 mm then the ratio of actual to predicted height would be 0.96. In order to minimize inaccuracies introduced by using a single adjacent vertebra, the 'mean predicted posterior' height (PP) was derived in a similar manner from the posterior heights of four adjacent vertebrae. For the most cephalic and caudal vertebrae, usually T4 and L5, the first four adjacent vertebrae below or above these levels

Table 1. Normal ranges for ratios of vertebral height and shape from T4 to L5: mean (SD) derived from measurements in 100 healthy women aged 45–50 years (the normal ranges for the posterior/predicted posterior (P/PP) at each level in women with metastatic breast cancer are also shown)

Vertebra	Controls				Breast cancer
	A/P ratio	C/P ratio	Posterior height (mm)	P/PP ratio	P/PP ratio
T4	0.961 (0.053)	0.944 (0.047)	24.88	1.000 (0.059)	1.000 (0.072)
T5	0.942 (0.048)	0.947 (0.044)	25.61	1.000 (0.056)	1.000 (0.060)
T6	0.920 (0.054)	0.946 (0.038)	26.34	1.000 (0.045)	1.000 (0.059)
T7	0.910 (0.055)	0.948 (0.044)	26.82	1.000 (0.048)	1.000 (0.049)
T8	0.924 (0.067)	0.958 (0.048)	27.11	1.000 (0.043)	1.000 (0.047)
T9	0.948 (0.068)	0.970 (0.049)	27.64	1.000 (0.045)	1.000 (0.052)
T10	0.961 (0.069)	0.968 (0.048)	29.21	1.000 (0.047)	1.000 (0.052)
T11	0.948 (0.065)	0.956 (0.051)	31.56	1.000 (0.052)	1.000 (0.056)
T12	0.940 (0.062)	0.944 (0.052)	34.16	1.000 (0.047)	1.000 (0.054)
L1	0.934 (0.055)	0.934 (0.043)	37.20	1.000 (0.046)	1.000 (0.051)
L2	0.970 (0.058)	0.927 (0.046)	38.03	1.000 (0.042)	1.000 (0.047)
L3	1.003 (0.061)	0.955 (0.039)	37.99	1.000 (0.048)	1.000 (0.050)
L4	1.041 (0.061)	1.002 (0.046)	36.86	1.000 (0.047)	1.000 (0.055)
L5	1.171 (0.089)	1.048 (0.067)	33.62	1.000 (0.085)	1.000 (0.084)

A, anterior; C, central; P, posterior.

Table 2. The reproducibility (cv%) of derived ratios of vertebral shape

Vertebra	Within same radiograph						Between paired radiographs		
	Intra-observer			Inter-observer			Intra-observer		
	A/P	C/P	P/PP	A/P	C/P	P/PP	A/P	C/P	P/PP
T4	3.45	2.70	4.03	4.28	2.62	3.40	4.58	5.08	5.29
T5	3.35	2.66	3.10	5.27	3.47	4.03	5.19	5.00	4.61
T6	3.14	1.31	2.01	3.06	2.91	2.78	5.07	4.56	4.92
T7	2.81	3.03	2.70	4.06	4.00	2.68	4.86	3.84	4.31
T8	2.69	2.46	2.67	3.20	2.24	1.45	4.01	5.29	3.94
T9	3.67	3.03	3.51	3.70	2.72	3.37	4.88	4.98	4.42
T10	1.96	2.24	3.03	3.58	2.10	3.33	4.66	4.38	4.17
T11	3.85	3.47	4.25	4.58	4.22	5.40	5.01	5.23	4.73
T12	5.56	3.19	3.53	6.54	3.19	4.27	5.05	5.42	4.49
L1	3.08	2.20	3.00	2.15	2.64	4.96	5.08	5.16	3.53
L2	3.16	2.35	2.76	2.99	2.74	1.41	3.88	4.99	3.14
L3	2.35	2.55	2.31	3.15	2.90	2.49	4.89	4.37	2.83
L4	2.15	2.43	2.37	2.88	3.03	2.20	5.22	4.30	4.25
L5	4.49	3.61	3.37	4.45	4.43	3.72	6.72	5.89	5.87
Mean	3.30	2.65	3.05	3.85	3.09	3.25	4.94	4.89	4.32
Overall	–	3.00	–	–	3.40	–	–	4.72	–

Intra-observer reproducibility was assessed by measuring 40 radiographs twice. Inter-observer reproducibility was assessed in 20 spine radiographs. Long-term reproducibility was derived from measurements of normal vertebrae in 100 paired radiographs obtained 6 months apart.
A, anterior; C, central; P, posterior; P/PP, posterior/predicted posterior.

were utilised to predict posterior height. In the mid-spine, two vertebrae on either side of the reference vertebra were used. At each vertebral level the actual posterior height (P) was expressed as a ratio to the mean predicted posterior height (PP). Similarly, within each vertebra the anterior (A) and central heights (C) were expressed as ratios to the posterior vertebral height. Normal ranges for the ratios of 'A/P', 'C/P' and 'P/PP' vertebral heights derived from normal women are shown in Table 1.

The short-term intra-observer and inter-observer reproducibility of these derived ratios is shown in Table 2. At each vertebral site the long-term reproducibility of the ratios in undeformed vertebrae measured on paired radiographs obtained 6 months apart ranged from 2.8% to 6.7% depending on the vertebral level measured. The long-term reproducibility errors of fractured vertebrae were greater. In 94 vertebrae the errors were 9.0%, 8.4% and 11.6% respectively for the A/P, C/P and P/PP ratios. For this reason we excluded previously deformed vertebrae from the assessment of incidence rates.

Vertebral Deformity

At each vertebral level the mean predicted posterior height (PP) was used to determine the presence or absence of a posterior vertebral deformity, an anterior deformity and a central deformity in that order.

After the most cephalic vertebra had been fully assessed, the algorithm then moved caudally, examining each vertebra in turn. Vertebrae with posterior deformities lying above the vertebra under examination were excluded. Lower vertebrae with reduced posterior heights were excluded from the calculation of the mean predicted posterior height where any of the predicted posterior heights were >3 SD below the maximum predicted height. The algorithm assumes that four adjacent posterior deformities would not occur.

The criteria for deformity were expressed as *z*-scores using the standard deviations at each vertebral site as shown in Table 1. For example, at a 3 SD cut-off the decrease in height from the normal mean necessary for detection of an anterior wedge deformity ranged from 14.4% to 26.7% depending on the vertebral level, 11.4% to 20.1% for central compression, and 12.6% to 25.5% for crush deformity, again depending on the vertebra under examination. In order to minimize the number of false positives (detecting a 'deformity' which does not actually exist) without substantially increasing the number of false negatives (true deformities which fail to be detected), we defined *two criteria* which must be both fulfilled in order to identify a vertebral deformity (Table 3). For example, a crush deformity would only be recorded if the P/PP ratio was below a specified cut-off value (mean P/PP − multiple of standard deviation) *and* the A/PP ratio also fell below the specified cut-off (mean A/P − multiple of standard deviations). For crush deformities, the A/PP ratio was used since the

Table 3. The classification of vertebral deformities and the criteria which needed to be fulfilled for each

Type of deformity	Criteria
Central collapse	C/P *and* C/PP < (mean C/P −3 SD)
Anterior wedge	A/P *and* A/PP < (mean A/P −3 SD)
Posterior wedge	P/PP < (mean P/PP −3 SD) *and* A/P > (mean A/P +3 SD)
Crush	P/PP < (mean P/PP −3 SD) *and* A/PP > (mean A/P −3 SD)

Note that two criteria must be fulfilled for each type of deformity. PP is the mean predicted posterior height of the vertebra under examination. The cut-offs were calculated from the normative data in Table 1.
A, anterior; C, central; P, posterior.

posterior height (P) was judged to be abnormally decreased and the mean P/PP ratio is unity.

Assessment of the Prevalence of Vertebral Deformity

Using the normal ranges for vertebral shape and the criteria in Table 3, we examined the prevalence of vertebral deformity in 163 women with breast cancer and skeletal metastases. The number of vertebral deformities at a range of cut-offs (2.0, 2.5, 3.0, 3.5 and 4.0 SD below the normal mean) were recorded. The prevalence of vertebral deformities found using this method was compared with the prevalence of deformities identified using the 'adjusted' algorithm of Melton et al. [10] adjusted to our normative data (hereafter referred to as the Melton method).

Assessment of the False Positive Rate for Prevalence of Vertebral Deformity

The number of deformities which are identified in the analysis will include a number of false positives (measurements fulfilling the criteria for deformity by chance). We used two methods for assessing the specificity of the criteria used. Firstly, we determined the apparent prevalence of deformity in a normal control population in which the presumed pervalence is zero. For this purpose, we applied the deformity criteria to the 100 young women used to derive our normative data. Secondly, because the derived ratios of vertebral shape were normally distributed (Fig. 2), we reasoned that errors of reproducibility (an apparently small or large vertebral ratio) would occur with equal frequency above and below any chosen cut-off value (e.g. >3 SD from the mean). Using this assumption, we also estimated the number of false positive deformities in the population of women with metastatic breast cancer by examining the number of measurements lying above the site and level-specific normal ranges (Table 1) at each of the five cut-offs described above.

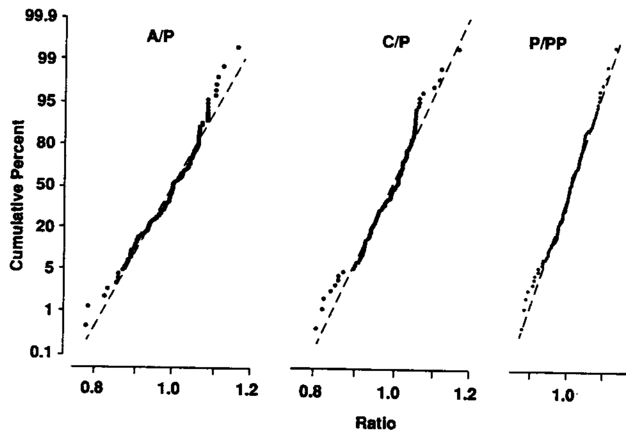

Fig. 2. Normal probability plots for the anterior/posterior (A/P), central/posterior (C/P) and posterior/predicted posterior (P/PP) heights of T8. The diagonal dotted line indicates the expected relationship assuming a Gaussian distribution. At this and at each level in the spine the three ratios showed a normal distribution.

Assessment of the Incidence of Vertebral Deformity

Follow-up radiographs in 122 of the patients with skeletal metastases were examined for deformities using the same criteria used to define prevalence, i.e. two estimates of point prevalence. The data from the two films were then compared to determine the number and type of new vertebral deformities. In addition, we measured the loss of vertebral height associated with each deformity (expressed as the percentage decrease in vertebral ratio from entry).

The point prevalence method was compared with two other methods. In the first of these a new deformity was judged to have occurred if any of the three ratios of vertebral shape (A/P, C/P and P/PP) decreased by more than 3 SD from the value at entry using the level-specific reproducibility (LSR) data on paired radiographs shown in Table 2 (hereinafter referred to as the LSR method). For example, a new anterior wedge fracture of T7 would be detected where the vertebral ratio A/P on a subsequent film has decreased by more than 14.58% (3 × 4.86). The other method that we tested was one used commonly in clinical trials [5,23,25,27] in which a new deformity was recorded if any of the three vertebral heights (A, C or P) decreased by 15% or more from that at entry (hereinafter referred to as the Riggs' method).

Irrespective of the method used, the analysis of the incidence of vertebral deformity was confined to those vertebrae which were defined as normal by the relevant algorithm on the initial radiograph.

Assessment of the False Positive Rate for the Incidence of Vertebral Deformity

In our point prevalence method, false positive deformities were defined in an identical manner to that used in the population study of prevalence. Thus a false positive deformity was scored if any of the vertebral ratios on the follow-up radiograph was more than a specified number of standard deviations above the mean for that site and level in a vertebra judged to be normal on the initial radiograph.

The number of false positive new deformities in the LSR method was studied using the same assumption as in the prevalence of deformities. Therefore, the number of false positives deformities was recorded as the number of patients and vertebrae showing increases in ratio of more than 3 SD from the value at entry. Although never stated in relevant publications, it was assumed that the 15% change in the Riggs' method was derived from estimates of reproducibility. Thus the false positive incidence rate was estimated as the number of patients and vertebrae showing increases in height of more than 15% from the value at entry.

Results

Normal Ranges for Vertebral Height and Shape

The ratio of actual to predicted posterior height (Table 1) was normally distributed with a mean of 1.00 (Fig. 2). The standard deviations differed markedly at different sites (a 2.5-fold range) suggesting the need to use site- and level-specific criteria for the assessment of vertebral deformity.

The standard deviations of the P/PP ratio were similar whether the predicted posterior height (PP) was derived from one adjacent or from four adjacent vertebrae (data not shown). Using four adjacent vertebrae in women with breast cancer, the P/PP ratios were unity and the SDs not markedly different from those measured in normal women. This finding suggests that there was little or no loss of accuracy in the estimation of PP due to the exclusion of adjacent vertebrae with posterior deformities.

False Positive Deformities

The prevalence of false positives at each of the cut-offs used in patients with breast cancer and normal controls is shown in Table 4. With our method using two criteria for vertebral deformity (Table 3), the false positive rate in the controls was high with a cut-off set at 2 SD and decreased with increasing stringency of cut-off (Table 4). At 3 SD the false positive rate in controls was 1%. With a similar cut-off the method of Melton gave a false positive rate of 12%.

In the women with breast cancer, 30 had one or more vertebra with ratios which lay above the normal range using a 2 SD cut-off. This false positive rate (18.4% of patients) decreased with increasing stringency of cut-off chosen as observed in the controls, and the 3 SD was 1%.

Table 4. The prevalence of false positive deformities at each of the cut-offs (expressed as the number of vertebrae and patients with values lying outside the normal range) in 100 normal subjects[a] and 163 patients with breast cancer[b]

Cut-off (SD)	Healthy subjects		Breast cancer	
	% subjects	% vertebrae	% patients	% vertebrae
2.0	20.0	2.1	18.4	1.8
2.5	5.0	0.5	7.4	0.6
3.0	1.0	0.0	1.2	0.0
3.5	0.0	0.0	0.6	0.0
4.0	0.0	0.0	0.6	0.0

False positives were scored as the number of vertebrae or patients with values lying [a] below the lower limit or [b] above the upper limit of the normal range.

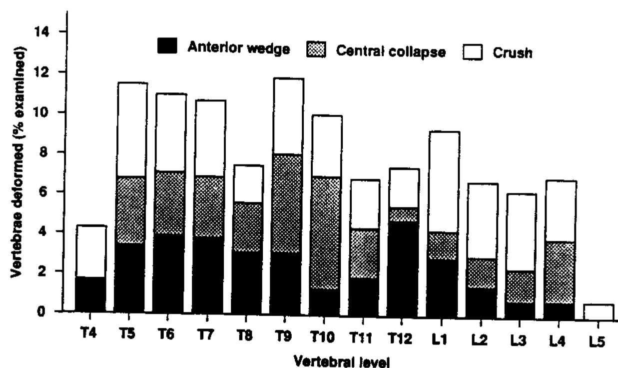

Fig. 3. The distribution of the various fracture types of the spines of 163 women with metastatic breast cancer. The number of fractures is expressed as a percentage of the total number of vertebrae examined at each level.

Prevalence of Vertebral Deformities

The apparent prevalence and types of vertebral deformity in patients with breast cancer at each of the cut-offs are shown in Table 5. As expected, the number of patients with deformity and the number of deformed vertebrae decreased with increasing stringency of the cut-off level. The method of Melton gave a prevalence of deformity intermediate to that utilizing a 2 SD or 2.5 SD cut-off. The proportions of the various types of vertebral deformities were similar and remained constant over the range of cut-offs. In contrast, the Melton method resulted in a higher apparent prevalence of crush and wedge deformities (Table 5).

In patients with breast cancer, the distribution of the various types of deformity within the spine using the 3 SD cut-off are shown in Fig. 3. No posterior wedge deformities were observed. Anterior wedge deformities were most frequently observed in the mid-thoracic spine, whereas fractures in the lower spine were predominantly central compression or crush fractures. A similar distribution was observed using the other cut-offs in our algorithm.

Incidence of Vertebral Deformities

Since the false positive rate was minimal at 3 SD, we decided to use the 3 SD cut-off in the analysis of incidence. The number of patients with breast cancer sustaining deformities and the number of newly deformed vertebrae over the 6-month period of follow-up are shown in Table 6. Using the point prevalence method with a 3 SD cut-off, 30 of the 122 patients (24.6%) sustained 61 new vertebral deformities (Table 6). The majority of newly deformed vertebrae had a decrease in vertebral ratio of more than 15%, and almost half involved decreases of more than 25% (Fig. 4). Ten newly deformed vertebrae had decreases in ratio of less than 15%. However, 6 of these 10 vertebrae were assessed in further radiographs at 12 months from the initial radiographs (data not shown). Of these, 5 (83%) were now deformed by more than 15%.

Using the LSR method (based on reproducibility at each vertebral site) and the Riggs' method, the incidence of new vertebral fractures was increased 2-fold over that determined by our method (Table 6). Thus, 64 (53%) and 56 (45%) of the patients were respectively

Table 5. The prevalence of vertebral deformities in 163 patients with skeletal metastases and breast cancer

Cut-off (SD)	Patients with deformity		Vertebrae with deformity		Type of fracture		
	n	(%)	*n*	(%)	Anterior wedge (%)	Central compression (%)	Crush (%)
2.0	120	73.6	295	14.7	28.1	38.0	33.9
2.5	88	53.9	201	9.9	31.3	33.3	35.4
3.0	75	46.0	153	7.6	30.1	34.0	35.9
3.5	63	38.6	123	6.1	23.6	40.6	35.8
4.0	54	33.1	94	4.7	26.5	39.3	34.2
Melton	103	63.2	238	11.8	40.3	19.7	40.0

A total of 2011 vertebrae were measured in the patients. The cut-off is defined as a number of standard deviations (SD) below the normal mean, with two criteria being fulfilled for each deformity (see text). The prevalence is also shown in the same population using the 'adjusted' algorithm of Melton et al. [10].

Table 6. The incidence of new and false positive deformities on radiographs obtained in a 6-month interval in 122 women with breast cancer

Method	Patients with new deformities	No. of new deformities	Patients with false positive deformities	No. of false positive deformities	FRP/ vertebra[a] (%)
Riggs	56	128	19	27	2.3
LSR	64	127	38	48	3.9
McCloskey	30	61	5	5	0.4

The rightmost column gives the number of false positive events as a proportion of the number of paired vertebrae compared on the radiographs.

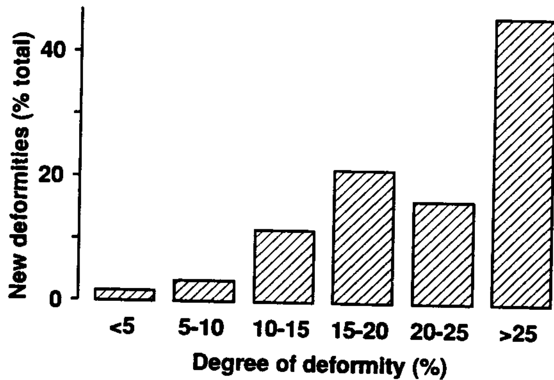

Fig. 4. The degree of vertebral collaspe associated with the identification of a new vertebral deformity. The collapse is expressed as the percentage decrease in vertebral ratio between paired radiographs obtained 6 months apart. The vertebral height decreased by more than 15% in the majority and by more than 25% in approximately half the vertebrae.

judged to have sustained new vertebral deformities in the 6 months between radiographs.

The false positive rates were higher using these two methods compared with our point prevalence method. Using the Riggs' method, 128 vertebrae showed decreases in vertebral height of more than 15%, but 27 vertebrae showed increases of more than 15%. This suggests that 21% of the new deformities may represent errors of reproducibility. Similarly, 56 patients showed apparent new deformities, but 19 patients showed increases in vertebral height. The false positive rates for patients and vertebrae sustaining new deformities using our point prevalence method were substantially lower (16% and 8% respectively).

The incidence of new vertebral deformities varied according to the number of deformities present on the initial radiograph (Table 7). Thus the rate of deformity (expressed as a fracture rate/100 patient-years) was only 58 in patients with no initial deformities, but was 224 in those with more than 2 deformities at entry.

Discussion

The problems associated with the definition and classification of vertebral deformity have recently attracted much attention and discussion [9–12,14–16,28]. The absence of a 'gold standard' for defining vertebral deformity has resulted in the development of several arbitrary methods, the majority of which involve quantitative or semi-quantitative vertebral morphometry [8–12,14–16]. They have been used in population studies, and in studies of the incidence of vertebral fracture in clinical trials [5,23,25–27].

The method we have developed shares several features with some previously published methods. Firstly, it is semi-automated and can be undertaken by a trained technician very rapidly (4 min). The data presented in this report were from measurements made exclusively by one technician. The method is therefore suitable for large-scale studies, and does not rely on the skills and potential subjectivity of a trained radiologist [24,29]. Secondly, like other recent methods, it takes account of the different vertebral heights and anatomical vertebral shape at different vertebral levels. At the cut-off that we adopted (>3 SD difference from the predicted height) the decrease in height from the normal mean necessary for detection of a deformity ranged from 14.4% to 26.7% for anterior wedge deformity, 11.4% to 20.1% for central compression and 12.6% to 25.5% for crush deformity, depending on the vertebral level. Thus, taking a 15% or 20% decrease in vertebral dimensions as a cut-off may lie anywhere between 1 and 3 SDs from the mean depending on the vertebral level. Thirdly, the methods predicts vertebral

Table 7. The effect of the prevalence of deformities at entry to the study on the subsequent incidence of new deformities in patients with metastatic breast cancer

Vertebral deformities at entry	No. of patients	Patients with new deformities (%)	No. of new deformities	Fracture incidence (rate/100 patient-years)
0	68	14.7	18	54
1–2	37	35.1	30	162
>2	17	52.9	19	224

heights and ratios for the individual patient rather than comparing them exclusively to a reference population [7].

The present study also highlights a problem intrinsic to all these methods including our own, namely the trade-offs between sensitivity and specificity. The more sensitive the technique, the more likely a vertebral deformity is to be artifactual. We have assessed this on the reasonable assumption that normal vertebrae do not grow in height in mature women. Accepting this assumption, the false positive rate was assessed indirectly by the number of normal vertebrae which appear to grow using any algorithm and any cut-off. We have shown that the distribution of vertebral height ratios is Gaussian, which validates in part the assumption that the incidence of apparently enlarged vertebrae on sequential assessment is likely to be an accurate index of the false positive rate.

It is clear that the specificity of most methods of assessing vertebral deformity is poor, with up to 70% of healthy controls appearing to have vertebral deformities depending on the chosen criteria. The adjusted algorithm of Melton, which we have compared in this study, takes into account the normal variations in vertebral shape before using a 15% difference in height to define a fracture. However, as we have demonstrated, the specificity remains relatively poor (12% false positive rate) and even higher false positive rates have been calculated by others [15].

There are several reasons for the poor specificity of methods for defining vertebral deformity. It is possible that a proportion of false positive deformities may result from reversible deformations of vertebral shape being incorrectly classified as pathological deformities [30]. However, the most important reason for the poor specificity of previously published methods relates to the large number of sites measured within each spine (42 sites if all vertebrae from T4 and L5 are included), which increases the statistical likelihood of detecting an 'abnormality'. Thus, in 100 patients a total of 1400 vertebrae would be measured, representing 4200 derived ratios of vertebral shape. At 2 SD or 95% confidence limits, 105 sites (2.5%) would be expected to lie below the normal range. If it is assumed that these events occur independently, then a deformity is likely to be detected in all patients, giving an apparent prevalence of deformity of 100%. At 3 SD or 99% confidence limits, the corresponding apparent frequency would be 21%. In reality, these events may not be truly independent and the apparent prevalence would probably be less, but still have a major impact on the apparent prevalence of fracture, particularly when the true prevalence is low. The effect is compounded when the measured incidence of vertebral deformity includes fractures which were scored as deformed on the initial radiographs, since, as we have shown, the reproducibility errors are greater. Thus, studies of incidence should be confined to vertebrae considered to be normal at the initial examination unless account is taken of the differences in reproducibility.

Table 8. The effect of different criteria for vertebral deformity on the apparent false positive rate in 100 healthy subjects

No. of criteria	No. of adjacent vertebrae examined	Cut-off point	False positive rate (% of patients)
1[a]	1–2	15%	12
1[b]	1–2	3 SD	9
1	4	3 SD	9
2	4	3 SD	1

[a]The method of Melton et al. [10].
[b]Analogous to the method of Eastell et al. [14].

The algorithm which we have developed has a much lower false positive rate even compared with methods using similar cut-offs or decreases in vertebral ratios to define deformity [10,14,15]. The use of two criteria for each deformity reduced the false positive rate in the control population from 9% to 1% (Table 8), identical to that observed in our patients with metastatic breast cancer (Table 4). Reasons for this include the use of two criteria to define the presence of a deformity (Table 8), and in the case of incidence rate, the avoidance of previously deformed vertebrae.

Our findings have a number of important implications for the study of vertebral fracture. First, in the case of population studies of prevalence, a false positive rate of the magnitude we have shown with other methods will have a major impact on the apparent prevalence of vertebral fracture in the community [3,31]. This effect will be proportionately greater the lower the true prevalence. Thus, for example, in describing prevalence by age, a significant false positive rate will distort the slope describing prevalence with age [32].

Second, differences in the criteria for the presence of deformity affect the types of deformities that are detected. It is probable that the higher apparent prevalence of crush fractures (posterior deformities) using the Melton method compared with our own (Table 5) reflects not only the cut-off but also that the criteria for posterior deformity are fulfilled by comparing the posterior height with the vertebra above *or* below. There are, therefore, two chances for detecting a posterior deformity, but only one chance for detecting an anterior or central deformity. The higher prevalence of anterior wedge deformities using the Melton method probably reflects the fact that the 15% cut-off lies within the normal range of vertebral shape as discussed above.

Specificity can be improved by increasing the degree of deformity necessary for detection of an abnormality, for example using 4 SD or a 25% decrease in height. The likely disadvantage of this approach is a decrease in sensitivity [15,31]. The true magnitude of this effect is impossible to quantify in the absence of a gold standard for fracture. A number of studies have examined this issue by qualitatively defining fractured and non-fractured populations, on the basis of either reading by a trained observer [11,15] or attendance at a specialist clinic for osteoporosis [14]. It has been suggested that

the specificity can be improved by using differing cut-offs for anterior, central and posterior sites within vertebrae [11] by calibrating the method to agree with fracture detection by a trained observer, but this seems at odds with the purpose of vertebral morphometry.

The trade-off between sensitivity and specificity is difficult to assess by these methods in the absence of a true gold standard. It is perhaps more appropriate to examine the clinical correlates utilizing different methods. We have shown that the presence of vertebral deformity correlates more closely with the presence of back pain using our method than those of Eastell or Melton [33]. We have also shown in a population survey that women with vertebral fracture have a significantly lower bone mineral density of the lumbar spine than those without [34]. The use of other algorithms discriminates cases from controls less adequately [35]. These observations suggest that our algorithm does not sacrifice sensitivity for the increase in specificity.

The effects of poor specificity have implications for studies of fracture incidence. We have shown that the apparent incidence of fracture, determined prospectively, is critically dependent on the method used. It appears that our method is sensitive to minor decreases in vertebral height in vertebrae which subsequently undergo more marked deformation. The two methods which used a decrease in vertebral height or ratio between radiographs are prone to 'noise' generated by errors of reproducibility (Table 6). The magnitude of this 'noise' needs to be put into context. For example, in the patients with breast cancer, 1223 normal vertebra in 122 patients at entry were re-examined 6 months later. This represent 3669 sites with paired measurements. Using the Riggs' method, 27 vertebrae showed apparent increases in height of more than 15%, suggesting a false positive rate of only 0.7% in terms of sites but 2.2% in terms of vertebrae. Moreover, these events occurred in 19 patients, suggesting that up to one third of patients with new deformities may have been misclassified due to errors in reproducibility.

The high false positive rates generated by these methods are likely to obscure true differences between treatment groups and have a major impact on the outcome of clinical trials where vertebral fracture is a primary endpoint. These potential effects are demonstrated in Table 9. If one assumes that previous estimates of fracture frequency using the Riggs' method in clinical trials in osteoporosis (approximately 60–80/100 patient-years) were subject to the same degree of error, then the true rate of fracture may be somewhat lower. In Table 9 we have assumed a true number of vertebral deformities as 60/100 patient-years in the control group, with a 50%, 33% or 20% reduction in a treatment group. As the false positive rate increases to the levels seen with the Riggs' method the apparent efficacy decreases by more than one third. Thus, a treatment may be rejected if the degree of effect is not deemed to be clinically significant. In addition, it is unlikely that the false positive fractures will be exactly evenly distributed between a control and treatment group. Even

Table 9. The effect of increasing false positive rate (FPR) on the apparent efficacy of a bone-active drug

FRP[a] (%)	No. of deformities in placebo wing	Apparent therapeutic efficacy (%)		
0	60	50	33	20
0.5[b]	67	45	30	18
1.0	74	41	27	16
1.5	81	37	25	15
2.0	88	34	23	14
2.5[c]	95	32	21	13
3.0	102	29	20	12
3.5	109	28	18	11
4.0[d]	116	26	17	10
2.5[e]	91 (99)	24 (38)	13 (28)	4 (20)
2.5[f]	88 (102)	18 (43)	7 (33)	−2 (25)

The true number of vertebral deformities is 60 in the placebo wing and the active agent has a true efficacy of 50%, 33% or 20%. The apparent efficacy decreases as the FPR increases. The bottom two rows show the effect of an uneven distribution of false positive deformities.
[a]Number of false positive deformities as a percentage of the vertebrae examined.
[b]Current method (0.29%).
[c]Riggs' method.
[d]LSR method.
[e]55% of false positives in test wing or, in parentheses, the placebo wing.
[f]60% of false positives in test wing or, in parentheses, the placebo wing.

small changes in the distribution of the false positive rate can have dramatic effects on the trial outcome. A 10% maldistribution of false positives with a false positive rate of 2.5% would yield an apparent efficacy of 7%–33% where the true efficacy was 33% (Table 9). Indeed, it is tempting to speculate that the recent studies showing small effects of sodium fluoride in fracture were in reality large effects [23,27]. Conversely, where basal rates of fracture are low, a high and uneven distribution of false positives may give variable increases or decreases in vertebral fracture rates as reported in one of the recent trials of etidronate [29]. The errors will be compounded if vertebrae which are fractured at entry to the studies are included in the analysis of new or 'progressing' fractures due to the poorer reproducibility in these vertebrae.

In conclusion, the technique that we have developed for the assessment of vertebral deformities is robust and rapid, and has minimal effects on sensitivity whilst maximizing specificity. Such goals are necessary for the study of fracture incidence, particularly in clinical trials, where sensitivity and specificity have a major impact on the power to detect differences between treatment groups. The method is able to detect minor vertebral deformities which subsequently progress and there is a close relationship between the existence of deformities and the subsequent rate of deformity in breast cancer, as recently shown for osteoporosis [36]. Substantial refinements will prove difficult in the absence of a gold standard. However, the relationship between the occurrence of new vertebral deformities, their magnitude and

association with clinical symptoms requires much more study [17]. The lack of an association between back pain and the presence of one or more vertebral fractures defined by less specific criteria has recently been demonstrated [37]. This agrees with our conclusion that more stringent criteria for vertebral fracture, such as the method we have proposed, should be used in epidemiological and clinical studies.

Acknowledgements. We are grateful to the Medical Research Council and to Rhone Poulenc Rorer for Programme Support. Our studies in breast cancer were supported by the Breast Cancer Research Trust and by Huhtamaki Oy Leiras. We would wish to thank Mrs Linda Reaney for her extensive technical assistance in the quantitation of the radiographs.

References

1. Riggs BL, Melton LJ. Involutional osteoporosis. N Engl J Med 1986;314:1676–86.
2. Spector TD, Cooper C, Fenton Lewis A. Trends in admissions for hip fracture in England and Wales, 1968–85. BMJ 1990;300:1173–4.
3. Kanis JA, Pitt FA. Epidemiology of osteoporosis. Bone 1992; 13Suppl 1:S7–S15.
4. Jensen KK, Tougaard L. A simple x-ray method for monitoring progress of osteoporosis. Lancet 1981;2:19–20.
5. Riggs BL, Seeman E, Hodgson SF, Taves DR, O'Fallon WM. Effect of the fluoride/calcium regimen on vertebral fracture occurrence in postmenopausal osteoporosis. N Engl J Med 1982;306:446–50.
6. Kleerekoper M, Parfitt AM, Ellis J. Measurement of vertebral fracture rates in osteoporosis. In: Christiansen C, Arnaud CD, Nordin BEC, Parfitt AM, Peck WA, Riggs BL, editors. Proceedings of the Copenhagen International Symposium on Osteoporosis. Copenhagen: Glostrup Hospital, 1984:103–9.
7. Gallagher JC, Hedlund LR, Stoner S, Meeger C. Vertebral morphometry: normative date. Bone Miner 1988;4:189–96.
8. Hedlund LR, Gallagher JC. Vertebral morphometry in diagnosis of spinal fractures. Bone Miner 1988;5:59–67.
9. Minne HW, Leidig C, Wuster CHR. A newly developed spine deformity index (SDI) to quantitate vertebral crush fractures in patients with osteoporosis. Bone Miner 1988;3:335–49.
10. Melton LJ, Kan SH, Frye MA, Wahner HW, O'Fallon WM, Riggs BL. Epidemiology of vertebral fractures in women. Am J Epidemiol 1989;129:1000–11.
11. Davies KM, Recker RR, Heaney RP. Normal vertebral dimensions and normal variation in serial measurements of vertebrae. J Bone Miner Res 1989;4:341–9.
12. Raymakers, JA, Kapelle JW, van Beresteijn ECH, Duursma SA. Assessment of osteoporotic spine deformity: a new method. Skeletal Radiol 1990;19:91–7.
13. Nelson D, Peterson E, Tilley B, et al. Measurement of vertebral area on spine x-rays in osteoporosis: reliability of digitizing techniques. J Bone Miner Res 1990;5:707–16.
14. Eastell R, Cedel SL, Wahner HW, Riggs BL, Melton LJ. Classification of vertebral fractures. J Bone Miner Res 1991;6:207–15.
15. Smith-Bindman R, Cummings SR, Steiger P, Genant HK. A comparison of morphometric definitions of vertebral fracture. J Bone Miner Res 1991;6:25–34.
16. Black DM, Cummings SR, Stone K, Hudes E, Palermo L, Steiger P. A new approach to defining normal vertebral dimensions. J Bone Miner Res 1991;6:882–92.
17. Leidig G, Minne HW, Sauer P, et al. A study of complaints and their relation to vertebral destruction in patients with osteoporosis. Bone Miner 1990;8:217–29.
18. Benger U, Johnell O, Redlund-Johnell I. Changes in incidence and prevalence of vertebral fractures during 30 years. Calcif Tissue Int 1988;42:293–6.
19. Harma M, Heliovaara M, Aromaa A, Knekt P. Thoracic spine compression fractures in Finland. Clin Orthop 1986;205:188–94.
20. Pogrund H, Makin M, Robin G, Menczel J, Steinberg R. Osteoporosis in patients with fractured femoral neck in Jerusalem. Clin Orthop 1977;124:165–72.
21. Ross PD, Wasnich RD, Vogel JM. Detection of prefracture spinal osteoporosis using bone mineral absorptiometry. J Bone Miner Res 1988;3:1–11.
22. Jensen GF, Christiansen C, Boesen J, Hegedus V. Epidemiology of postmenopausal spinal and long bone fractures. Clin Orthop 1982;166:75–81.
23. Riggs BL, Hodgson SF, O'Fallon WM, et al. Effect of fluoride treatment on the fracture rate in postmenopausal women with osteoporosis. N Engl J Med 1990;322:802–9.
24. Storm T, Thamsborg G, Steiniche T, Genant HK, Sorensen OH. Effect of intermittent cyclical etidronate therapy on bone mass and fracture rate in women with postmenopausal osteoporosis. N Engl J Med 1990;322:1265–71.
25. Gallagher JC, Goldgar D. Treatment of postmenopausal osteoporosis with high doses of synthetic calcitriol. Ann Intern Med 1990;113:649–55.
26. Tilyard M, Spears GFS, Thomson J, Dovey S. Treatment of postmenopausal osteoporosis with calcitriol or calcium. N Engl J Med 1992;326:357–62.
27. Kleerekoper M, Peterson EL, Nelson D, et al. A randomised trial of sodium fluoride as a treatment for postmenopausal osteoporosis. Osteoporosis Int 1991;1:155–61.
28. Sauer P, Leidig G, Minne HW, et al. Spine 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.
29. Watts NB, Harris ST, Genant HK, et al. Intermittent cyclical etidronate treatment of postmenopausal osteoporosis. N Engl J Med 1990;323:73–9.
30. Kleerekoper M, Nelson DA. Vertebral fracture or vertebral deformity. Calcif Tissue Int 1992;50:5–6.
31. Kanis JA, Caulin F, Russell RGG. Problems in the design of clinical trials in osteoporosis. In: St J Dixon A, Russell RGG, Stamp TCB, editors. Osteoporosis: a multidisciplinary problem. New York: Grune and Stratton, 1983:205–22.
32. Kanis JA, McCloskey EV. Epidemiology of vertebral osteoporosis. Bone 1992;13:S1–S10.
33. Kanis JA, McCloskey EV. Detection of vertebral osteoporosis. Rorer Foundation 1993 (in press).
34. Spector TD, McCloskey EV, Mastoswamy I, et al. Epidemiology of vertebral fractures in the general population: the Chingford Study. J Bone Miner Res 1991;6(Suppl 1):S277.
35. McCloskey EV, Spector TD, Eyres KS, et al. Relationship between criteria for vertebral deformity and bone mass in a population study. J Bone Miner Res 1992;7(Suppl 1):S137.
36. Ross PD, Davis JW, Epstein RS, Wasnich RD. Pre-existing fractures and bone mass predict vertebral fracture incidence in women. Ann Intern Med 1991;114:919–23.
37. Ettinger B, Black DM, Nevitt MC, et al. Contribution of vertebral deformities to chronic back pain and disability. J Bone Miner Res 1992;7:449–56.

Received for publication 19 May
Accepted in revised form 14 December 1992