Skeletal Radiol (1986) 15:347–349

Skeletal Radiology

Observer variation in the detection of osteopenia *

David M. Epstein, M.D.[1], Murray K. Dalinka, M.D.[1], Frederick S. Kaplan, M.D.[2], Judith M. Aronchick, M.D.[1], David L. Marinelli, M.D.[1], and Harold L. Kundel, M.D.[1]

Departments of [1] Radiology and [2] Orthopedic Surgery, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania, USA

Abstract. In order to determine observer variation in the detection of osteopenia, 15 pairs of lateral chest radiographs obtained within two weeks of each other were reviewed separately by two radiologists and one orthopedist on three separate occasions. Intra- and interobserver variations were calculated for each individual film and film pairs using Kappa values. The individual observers were not able to give consistent readings on the same film on different days (average Kappa = 0.54). When the additional factors of repeat films (average Kappa = 0.47), or separate observers (average Kappa = 0.38) were analyzed, agreement was even worse. The identification of osteopenia from the lateral view of the thoracic spine is highly subjective and variable from film to film and observer to observer.

Key words: Osteopenia – Observer variability – Bone radiology

Osteopenia is defined as a paucity of normally mineralized bone. Osteoporosis, the major cause of osteopenia, has been defined as a clinical syndrome in which there is decreased bone mass for age, sex, and race, leading to structural bone failure manifested by fractures. It is the most prevalent bone disease in the world and a menacing health problem in the elderly. In its early stages, osteoporosis is a silent disorder, causing osteopenia but neither signs nor symptoms. Early detection may spare patients significant morbidity particularly from compression fractures of the spine. The radiologist may often suggest the presence of osteopenia from the lateral radiograph of the chest when the vertebral bodies appear radiolucent. However, in the absence of compression fracture a variety of technical, perceptual, and physiological factors may impair this assessment. These factors may make it difficult for a radiologist to be consistent with himself and with other radiologists on the presence of osteopenia from examination to examination. The following study was devised to assess the consistency of three observers both individually and with each other regarding the observation of osteopenia based on the lateral chest radiograph.

* Presented at the 71st Scientific Assembly of the Radiological Society of North America, Chicago, IL, November 17–22, 1985

Address reprint requests to: David M. Epstein, M.D., Department of Radiology, Hospital of the University of Pennsylvania, 3400 Spruce Street, Philadelphia, PA 19104, USA

Materials and methods

Paired lateral chest radiographs of 15 patients were selected based upon the following criteria: (1) two lateral films taken less than two weeks apart, (2) absence of compression fracture or disease overlying the thoracic spine, (3) films considered technically adequate for interpretation.

The radiographic technique in our automated chest room utilized Kronex 80L film (medium speed) with a high plus (fast screen). Radiographs are phototimed at 125 KV, 200 MA and normal density except for radiographs on large patients which are phototimed at 400 MA and +1 density.

Each set of 30 radiographs (15 pairs) was evaluated independently by two radiologists and one orthopedist specializing in metabolic bone disease. The name, age, and sex of each patient in the study was eliminated from each radiograph prior to interpretation to eliminate observer bias. There was no information provided to any of the readers regarding the distribution of normal or abnormal radiographs in the study. The readings were performed in a blind fashion on three separate occasions at two week intervals for a total of 270 observations. For the purpose of this study only the suspected presence or absence of osteopenia was recorded. The true presence of osteopenia was not known and accuracy (presence or absence) of osteopenia was not established by biopsy or other methods.

The result were analyzed for individual and interobserver agreement using the Kappa statistic which defines the proportion of agreement in excess of that which is to be expected

from chance alone [1, 3]. In its simplest form this can be calculated by the equation:

$$\text{Kappa} = \frac{\text{observed agreement minus expected agreement}}{\text{1 minus the expected agreement}}$$

A Kappa score of 0 indicates only chance agreement: up to 0.4 represents poor agreement; 0.41 to 0.59 indicates moderate agreement; 0.6 to 0.79 is regarded as a good rating. Above 0.8 is excellent observer consistency.

Results

Intraobserver variation (internal consistency) was assessed by calculating the Kappa value for each observer during the three sessions (Table 1). On the basis of 30 single radiographs interpreted three times by each observer, the average Kappa was 0.54 (0.49 to 0.64) indicating only fair agreement of each reader with himself. When the results were analyzed to determine if repeat films (paired radiographs) would influence consistency, the average Kappa decreased to 0.47. For each reader, the Kappa for paired films was lower than the Kappa for three readings of the same film. Hence, each observer was minimally more consistent when evaluating the same film than comparable films on the same patient.

Table 2 shows internal consistency for each reader of the paired radiographs for each session. Two of the readers became more consistent with themselves as the study progressed and by the third session, all three had achieved a good rating.

None of the 30 films was scored as showing osteopenia (positive) by all three observers on all three reading. However, 11 of the films were consistently considered normal and this included four matched pairs. Interobserver agreement for each session shown in Table 3 indicated moderate to poor agreement between each of the three readers for each session. Overall, the effect of multiple observers was an average Kappa of only 0.38 for a given film. There was no significant difference in agreement between the two radiologists with each other or either radiologist with the orthopedist.

Table 1. Intraobserver variation in kappa values

	Film	Film pair
Reader 1	0.5	0.42
Reader 2	0.49	0.38
Reader 3	0.64	0.61

Table 2. Intraobserver variation in kappa values

	Session		
	1	2	3
Reader 1	0.25	0.58	0.68
Reader 2	0	0.52	0.67
Reader 3	0.61	0.47	0.61

Table 3. Interobserver variation by session in kappa values for 30 films

		Session		
		1	2	3
Observer	1 versus 2	0.14	0.45	0.42
	2 versus 3	0.15	0.55	0.42
	1 versus 3	0.36	0.41	0.54

Discussion

Osteopenia is a generic term meaning too little bone. Osteoporosis is a form of osteopenia in which there is decreased density of normally mineralized bone compared to that of age and sex matched controls [4]. Osteoporosis can affect both cortical and trabecular bone, but to different degrees. The axial skeleton (vertebral bodies) is composed largely of the metabolically active trabecular bone which has a higher turnover rate than cortical bone and is usually more severely affected in osteoporosis.

Because the spine is composed of metabolically active trabecular bone, it has been thought that the vertebral bodies might provide the best radiologic evidence for osteopenia and hence early osteoporosis [6]. The normal vertebral bodies have a dense network of interlacing osseous trabeculae which provide radiopacity when contrasted to the adjacent intervertebral discs. When osteopenia exists, the vertebral bodies appear radiolucent and the vertical striations become more prominent owing to the early resorption of small horizontal trabeculae. Later, resorption of the vertical striations renders a hollow appearance to the vertebral body with biconcavity from pressure of the intervertebral disc on the weakened end plate and finally compression fracture [6, 7].

Our study suggests that the early radiologic assessment of osteopenia based on observations of the thoracic spine is highly subjective and may be misleading since radiologists at best have only fair consistency with themselves and each other regarding its presence. The reasons for this inconsistency are not entirely clear from our study. It is known

that the overall density of a film is related to the film screen combination, film processing, overlying soft tissues, and patient positioning [6]. However, all the films in this study were obtained with a standardized technique and yet the Kappa was only 0.54 for multiple readings of the same film. The alterations produced by body habitus and overlying soft tissues on the radiographic density should have been negated by obtaining paired radiographs of the same patient within two weeks and yet the Kappa decreased to 0.47. It was more difficult to consistently appreciate osteopenia in an individual patient than for a given single film.

Other factors in addition to film technique affect our perception of density on a radiograph. Our perception of density is influenced by the surrounding background structures (background contrast effects) [2]. The eye is challenged with structures of different optical density and as a result of lateral inhibition, the true shade of the object in question appears altered.

One should realize that any study which attempts to assess observer variation is a contrived situation [1]. Clinical information or external visual clues may be suppressed and case selection may influence results. In general, films in a sample should show a wide range of severity, but this study was devised to test our perception of osteopenia prior to the development of compression fractures. Hence, patients with multiple compression fractures and biconcave vertebral bodies due to severe osteoporosis were excluded from the sample. Perhaps the radiographic abnormalities necessary to diagnose early osteopenia are too subtle for the eye to discriminate consistently. This leads to a more subjective judgement which lowers agreement.

It is said that experienced observers have greater perception and constistency [1]. All three observers in our study were considered experienced and yet still scored relatively low ratings. Two of the three readers seemed to improve as the study progessed in sessions two and three. Perhaps familiarity with the study design improved their results or each reader independently developed his own internal criteria for osteopenia.

For all three readers, it was easier to agree on the absence of osteopenia rather than its presence. Eleven films consistently were called negative while no film was consistently positive. The effect of multiple observers should be an increase in consistency [1]. However, our study revealed the opposite. The lowest average Kappa (0.38) was for interobserver agreement between multiple readers. This clearly depicts the inherent difficulty in identifying osteopenia from plain film radiography of the thoracic spine.

This study was concerned with concordance or reproducibility of observations whereas accuracy is concerned with truth. The actual presence of osteopenia in our study was not determined by biopsy or other methods. There are more objective methods which may be used to identify the early presence of osteopenia in the axial skeleton [5, 6]. Briefly, dual photon absorptiometry using a modified rectilinear scanner can measure bone mineral content with an accuracy of 4 to 10%. Quantitative computed tomography (CT) employing single or dual energy and a phantom can make direct density measurements of trabecular bone in the vertebral bodies with precision and accuracy. It may permit the reliable early evaluation of osteopenia and enable effective monitoring of its treatments.

References

1. Cockshott WP, Park WM (1983) Observer variation in skeletal radiology. Skeletal Radiol 10:86
2. Daffner RH (1983) Visual illusions affecting perception of the roentgen image. CRC Crit Rev Diagn Imaging 20:79
3. Fleiss JL (1981) Statistical methods for rates and proportions. John Wiley and Sons, New York, p 227
4. Kaplan FS (1983) Osteoporosis. Clin Symp 35:2
5. Richardson ML, Genant HK, Cann CE, Ettinger B, Gordan GS, Kolb FO, Reiser UJ (1985) Assessment of metabolic bone disease by quantitative computed tomography. Clin Orthop 185:224
6. Schneider R (1984) Radiologic methods of evaluating generalized osteopenia. Orthop Clin North Am 15:631
7. Siegelman SS (1970) The radiology of osteoporosis. In: Osteoporosis. Grune and Stratton, New York, p 68