2009

# Chest x-ray clues to osteoporosis : criteria, correlations, and consistency

Natalie Renee Simmons
*Yale University*

# CHEST X-RAY CLUES TO OSTEOPOROSIS: CRITERIA, CORRELATIONS, AND CONSISTENCY

A Thesis Submitted to the
Yale University School of Medicine
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Medicine

by

Natalie Renee Simmons

2009

CHEST X-RAY CLUES TO OSTEOPOROSIS:

CRITERIA, CORRELATIONS, AND CONSISTENCY

Natalie Renee Simmons, Cindy Miller, Michele Johnson, Lawrence Staib, Kevin Johnson, and Anne McBride Curtis. Department of Radiology, Yale University, School of Medicine, New Haven, CT.

The purpose of this study was to assess whether radiologists could accurately assess osteopenia on chest plain films. Two chest radiologists evaluated lateral chest films from 100 patients (80 female and 20 male), ranging in age from 16 to 86 years, for osteopenia and its associated findings. Intra- and interobserver agreement was determined using weighted kappa statistics, and accuracy was assessed by making comparisons to bone mineral density as measured by the non-invasive gold standard of dual-energy x-ray absorptiometry (DXA). Overall, radiologists were good at identifying signs of late, but not early, disease. Intraobserver consistency was substantial for fish vertebrae ($K_{w1}$=0.638; $K_{w2}$=0.0.712) with moderate interobserver agreement ($K_w$=0.45). Similarly for wedged vertebrae, intraobserver consistency was substantial to moderate ($K_{w1}$=0.654; $K_{w2}$=0.533) with substantial interobserver agreement ($K_w$=0.622). These radiographic signs correlated with true disease as shown by high specificity values. Therefore, this study indicates that if osteopenia is suspected (i.e., there is a wedge or fish vertebra) or its associated features are seen on a CXR, it is crucial for radiologists to comment on it. The literature suggests that referring physicians do not pay attention to such findings in radiology reports. Radiologists could effect change in clinical treatment by not burying these findings in the report body, but instead putting it in the impression, along with a recommendation that the finding be followed up with DXA. Because effective interventions for women with osteoporosis exist, the results of this study will contribute to a major change in the practice of chest radiology and improve women's health by preventing the devastating disability associated with osteoporosis.

## ACKNOWLEDGEMENTS

The author would like to thank her thesis advisor, Dr. Anne Curtis, for her outstanding mentorship and constant encouragement throughout this endeavor. This thesis would also not have been possible without Dr. Cindy Miller and Dr. Michele Johnson who graciously gave of their time and effort. Much appreciation goes to Dr. Larry Staib for his advice about research design and statistics as well as to Dr. Kevin Johnson who reviewed the final manuscript. Thanks to Doug Tabor and Matthew Zawalich of Yale-New Haven Hospital who helped in navigating the hospital's electronic resources. Finally, the author would like to thank her mother and best friend, Dr. Marsha T. Simmons for her unconditional support and example of life-long learning. Above all, the author expresses gratitude to the Great Physician, Jesus Christ, for the opportunity to help hurting people through the practice of medicine.

## TABLE OF CONTENTS

While it is likely that most women have some understanding of the sequelae of osteopenia, many do not know whether they are affected by it. Results from the National Osteoporosis Risk Assessment (1) showed that nearly one out of every two women tested had undetected low bone-mineral density (BMD), and 7% had undetected osteoporosis. In this same study, women with osteoporosis were four times more likely to have a fracture after one year, while the fracture rate in women with low BMD was twice that of women with normal BMD (1). Indeed, low BMD is the single best predictor of fracture risk in asymptomatic postmenopausal women (2). This is particularly significant since 20% of hip fractures lead to death within a year (3). Because effective interventions for women with osteopenia exist, the disability associated with osteopenia is no longer an inevitable result of aging, and is at least partially preventable (4). With prevention available, early detection becomes important.

GOLD STANDARD

It used to be that osteopenia was only diagnosed after a fracture had occurred (4). The World Health Organization (WHO) has established the following definitions: osteopenia is a value for BMD of 1-2.5 Standard Deviations (SDs) below the young-adult mean; osteoporosis is a value for BMD of 2.5 SDs or more below; and severe (established) osteoporosis is a value of 2.5 SDs or more below in the presence of one or more fragility fractures (5). This was done for epidemiological purposes to categorize patients who were at risk for the condition. However, they are now used by physicians as guidelines for treatment of the condition before the sequelae of its being present has resulted (4). BMD testing compares a patient's bone density to that of a young, normal, healthy 30-year-old adult with peak bone density. T-scores represent standard deviations above or below this normal value. If more than one body location (e.g., hip, spine, etc.) is tested, then the lowest T-score is used. Z-scores compare bone density to what is normal in someone of the same age and body size as the patient. Clinically, Z-scores are not used to diagnose osteopenia in postmenopausal women and men age 50 or older as well as perimenopausal women since low BMD is common in these patients. Z-scores are recommended for younger men, premenopausal women and children; however, diagnosis of low BMD in these patient populations is usually not based solely on BMD testing, but on other clinical features as well.

Dual-energy x-ray absorptiometry (DXA) of the hip and spine is currently the "gold standard" for measurement of BMD (6; 7). The hip and spine are the locations tested because patients have a higher risk of fracturing these bones, and fractures in these locations lead to longer recovery times and disability. However, different anatomical regions (e.g., femoral neck and trochanter) are often measured along with the hip and spine, and the DXA report usually provides individual T and Z-scores for each location. Other techniques sometimes used to assess bone density are pDXA (peripheral dual energy x-ray absorptiometry), QUS (quantitative ultrasound), QCT (quantitative computed tomography), pQCT (peripheral quantitative computed tomography). The peripheral tests (pDXA, QUS and pQCT) are only used for screening purposes but not for monitoring response to treatment.

PLAIN FILM FINDINGS

Standard radiographs have been reported to be sensitive enough to detect osteopenia after 25%-40% of bone density has been lost (8; 9). For most BMD tests, 1 SD difference in a T-score equals a 10-15% decrease in bone density (7). Thus, one would expect to be able to detect changes on plain films for patients who are currently classified by the WHO criteria has having osteoporosis (>25% bone density loss); it is likely that plain films would not detect patients with osteopenia.

Generalized osteopenia is most prominent in the axial skeleton (especially, the vertebral column, pelvis, ribs, and sternum). Several qualitative features of osteopenic bone have been noted (8; 9; 10). The first feature is increased radiolucency, reflecting a relative lucency of the central portion of the vertebral bodies compared with the subchondral bone at the superior and inferior endplates of the vertebrae. This radiolucency results from the spongy bone being composed of fewer plates which are each reduced in caliber. A second plain radiographic feature of osteopenia is the appearance of vertically oriented bars of increased radiodensity as individual trabeculae become thinner and eventually disappear. A “picture frame” or “empty box” appearance of the vertebral body occurs when the cortical margin is accentuated, by apparent loss of density of the remainder of the vertebra. Changes also occur in the shape of affected vertebrae. Wedge-shaped vertebrae reflect loss of height which is greater anteriorly than posteriorly. This change is common in the thoracic spine where normal kyphosis causes maximal pressure to be exerted on the anterior aspect of the vertebrae. The posterior aspect of the vertebrae is less apt to become reduced in height as it is supported by the paravertebral musculature and the posterior bony elements of the spine. A biconcave shape of the vertebral body results when the anterior and posterior aspects maintain their height, but the central portion is reduced in height; this may also be referred to as a "fish vertebra." Compression reflects loss of height of the entire vertebral body. Both wedging and compression are manifestations of vertebral fractures in a patient with osteoporosis and will lead to an accentuation of the normal thoracic kyphosis; it is generally not possible to distinguish acute from chronic fractures. It might be noted that not all vertebral bodies will be affected in the same way or some vertebrae may maintain normal density and shape while adjacent vertebral bodies demonstrate significant changes.

Osteopenia is most often due to osteoporosis, but may be seen in other diseases as well including osteomalacia, hyperparathyroidism, renal osteodystrophy, and neoplasia (8). In osteomalacia, a condition which occurs after the growth plates have closed, there is lack of mineralization of normal osteoid. Additional radiographic features in this disorder are coarsening of the trabeculae and pseudofractures. Hyperparathyroidism can be distinguished from simple osteopenia by the presence of subperiosteal resorption. Renal osteodystrophy reflects a combination of osteomalacia and hyperparathyroidism with their attendant radiographic features. An additional finding in renal osteodystrophy is the so called rugger jersey spine in which the vertebral endplates are markedly denser than the remainder of the vertebra. Malignancies including leukemia and multiple myeloma may have diffuse osteopenia in addition to focal lytic lesions; they may also demonstrate loss of height of vertebral bodies. In malignancies associated with elaboration of a parathyroid hormone like substance, osteopenia and features of hyperparathyroidism may be seen.

RELATIONSHIP BETWEEN OSTEOPENIA ON PLAIN FILMS AND BMD

Radiographic findings of osteopenia, such as those listed above, have been found by some researchers to be significantly related to BMD. Garton and colleagues (11) compared subjective estimates of vertebral body osteopenia on standardized lateral radiographs (performed of the 14 vertebrae between T4 and L5 with 2 exposures centered on T7 and L2) with BMD measurements of the anteroposterior (AP) lumbar spine and femoral neck in both men and women aged 52-90 years. Three radiologists assigned an osteopenia score to the radiographs based on the Saville method. The following is the 5-point Saville index which provides a semiquantitative method for describing osteopenia (11; 12; 13):

Grade 0: normal bone density.

Grade 1: minimal density loss; end plates stand out giving a stenciled effect.

Grade 2: vertical striation is more obvious; end plates are thinner.

Grade 3: more severe loss of bone density; end plates becoming less visible.

Grade 4: ghost-like vertebral bodies; density is no greater than soft tissue; no trabecular pattern is visible.

In Garton's study, BMD measurements were compared in individuals without significant vertebral deformity, those with mild (20-25%), or those with definite (> 25%) reductions of vertebral height. They found that intraobserver agreement was moderate to good and interobserver agreement was fair to moderate. There was overlap between gradings, but the BMD was indeed significantly related to visually estimated osteopenia. Further, the BMD measured at the hip and spine was related to vertebral deformity in women but not in men. Though it has also been utilized in other investigations, the Saville index has not been widely implemented because the radiographs are uncalibrated and interpretation is affected by interobserver variability (13).

McCullagh *et al.* (14) retrospectively measured the BMD in the lumbar spine of patients without vertebral fractures who had received a plain film diagnosis of osteopenia. They compared these patients to a group of one or more age and sex matched patients with one or more low impact vertebral fractures. They concluded that a radiological report of osteopenia on plain films was indeed a strong predictor of low bone density by BMD measurement; however, plain film estimation could not differentiate specifically between osteopenia and osteoporosis.

Ahmed *et al.* (15) also reported that radiographic evidence of osteopenia was a strong predictor of osteoporosis. They had reviewed 3530 referrals of women for bone density measurements of the spine and femur to determine the relationship between BMD measurements and the initial reason for referral to a BMD screening service. The highest proportion of women with osteoporosis in any group was in the group who had been diagnosed as osteopenic on plain films of the lumbar spine (n = 269).

However, other researchers have provided conflicting results. Williamson *et al.* (16) found that there was little ability to accurately diagnose osteopenia by chest film and proposed that it was unjustified to comment on the presence or absence of osteopenia on the basis of chest films. In their study, the estimated degree of bone density on 45 lateral chest films as read by nine radiologists was compared with DXA of only the lumbar spine taken within the same 6 month period. One potentially limiting factor of this study was that the radiologists reported an overall impression of bone density rather than referring to specific criteria which would have justified their impression. Furthermore, the bone density values were reported as grams of hydroxyapatite/square centimeter, rather than the more widely accepted T and Z-scores. Finally, it was suggested by McCullagh *et al.* (14) that this study was underpowered.

Jergas *et al*. (10) compared routine radiographs and PA DXA of the lumbar spine in the diagnosis of osteopenia (using a T-score of -2SD as the threshold for the diagnosis of osteopenia). They found a poor correlation between BMD, as measured by DXA, and a lumbar spine index (LSI). Additionally, in this study, radiographs of the lumbar spine (obtained in both the AP and lateral planes) were evaluated by nine observers in order to determine observer variation. The readers were not given specific criteria or training. Jergas *et al*. concluded that osteopenia can reliably be detected from lumbar spine radiographs by all readers only after a substantial amount of BMD is lost (i.e., 60% or more). Further, they noted that the most inconsistency between DXA and observers occurred in cases where the reduction in BMD was between 10% and 20%. This study also highlighted some of the caveats regarding spinal DXA. The authors suggested that DXA may not detect osteopenia in patients with overlying aortic calcification or with degenerative changes of the lumbar spine since the facet hypertrophy and disk space narrowing may artifactually increase bone density.

The latter study investigated the lumbar spine on plain films. This is a useful site for direct comparison with DXA, which also measures BMD at this location. However, women may not receive plain films of the lumbar spine unless they are symptomatic. Additionally, these x-rays involve a substantial amount of radiation exposure. Chest x-rays are performed on more people and more frequently per person. Thus, identification of osteopenia on these examinations would be an especially helpful tool.

## PURPOSE AND HYPOTHESES

The purpose of this investigation was to assess whether radiologists could accurately assess osteopenia on plain films of the chest. This possibility was tested by determining the degree to which radiologists' designation of "osteopenia" on chest plain films correlated with the bone mineral density status as measured by the gold standard of dual-energy x-ray absorptiometry as well as by determining the intra- and interobserver variability in this designation. If the designation of "osteopenia" on chest plain films is a valid indication of a patient's bone mineral density, then these designations should demonstrate a statistically significant correlation with the DXA classification. If there is consensus amongst radiologists as to which chest plain films show evidence of osteopenia, then there should be a low level of intra- and interobserver variability.

# METHODS

SUBJECT ELIGIBILITY

A retrospective cohort study was conducted on patients who had received both BMD testing by DXA and a lateral CXR within the same 12 month period using the Yale Diagnostic Radiology IDX Database Search (performed by Dough Tabor of Information Technology at Yale-New Haven Hospital). All radiographs and reports were obtained in the search from the electronic record known as PACS (picture archiving and communication system) which archives radiologic studies performed at Yale-New Haven Hospital. Patients had been referred for imaging at Yale-New Haven Hospital (a large teaching hospital) by a diverse number of attending physicians in the inpatient, outpatient, and emergency department settings between November 2002 and November 2008. Only lateral chest films were utilized since the lateral view provides the most information about the spine.

These patients were then assigned to one of three groups (normal, osteopenic, or osteoporotic) according to the results of their BMD by DXA. The DXA reports included separate T and Z-scores for each of several anatomical regions (i.e., lumbar spine, hip, femoral neck, and trochanter); however, DXA of the hip and spine is currently the "gold standard" for measurement of BMD (5; 6; 7). Therefore, if the T and Z-scores differed by anatomical region for a given patient, then the worst score (from either the lumbar spine or hip) was used to classify the patient into one of the three experimental groups as is done in clinical practice (5). All DXA reports had been generated by the same radiologist. From this patient pool, 100 radiographs were selected with the following distribution: 33 "normal," 33 "osteopenic," and 34 "osteoporotic." The films in each group were selected randomly by the principal investigator in one setting. All chest radiographs were standard lateral chest films taken in the erect position.

## DATA COLLECTION

Two experienced radiologists specializing in chest radiology on staff at Yale-New Haven Hospital were then asked to evaluate all 100 radiographs. The readers were blinded as to the patients' BMD by DXA reading. Readers were asked to use a Likert scale to evaluate the exams for the presence or absence of several radiographic findings associated with osteopenia, to form a conclusion regarding the presence or absence of osteopenia, and to comment on how often they reported the finding of osteopenia during their routine readings. Each reader's answers were scored according to the degree of certainty the observer felt about her response to each question. There were five possible grades of certainty for each answer, which ranged form "definitely normal" to "definitely abnormal," including an "uncertain" option. Readers recorded their answers on separate datasheets (Appendix A) for each radiograph.

Each reader evaluated the 100 exams twice, and the films were presented in different orders on separate worklists between the two readings. The worklists were generated by first randomly dividing the 100 radiographs into three smaller groups. These three groups were then combined in different orders to yield three distinctly different worklists. Each reader was assigned a different worklist. For the second reading, the order of presentation was again varied, by reversing the order of the films.

A non-chest radiologist also evaluated a subset of the selected cases following the above protocol in order to provide insights into how the results from the chest radiologists might be generalized to other radiologists.

## STATISTICAL ANALYSIS

Data were entered into a Microsoft Excel spreadsheet (Microsoft, 2003). Statistical tests were computed using Minitab® 15.1.20.0. (Minitab Inc., 2007) and MedCalc® 10.0.2.0 (Frank Schoonjans, 1993-2008) statistical software. Lawrence Staib, Ph.D. of the Yale University School of Medicine provided statistical advice; however, all statistical computations were done by Natalie Renee Simmons.

Initially, a Receiver Operating Characteristic (ROC) analysis was considered to determine the diagnostic test performance compared to the gold standard (17; 18; 19). However, because ROC's are not influenced by a reader's bias toward choosing a particular diagnosis (20), ROC's were not ideal for the present study which sought to explore these natural biases effecting inter- and intraobserver variability. As long as there is a difference in the difficulty with which early and advanced cases of osteopenia may be detected, the ROC curve will be influenced by the proportions of these extreme cases in the study sample (17). Additionally, in the current study, the cut-off levels for the gold standard had already been pre-determined based on the WHO clinical guidelines for T and Z-scores on DXA (5).

Percent agreement or disagreement was another potential statistical tool. This statistic represents the number of agreements divided by the total paired judgments multiplied by 100. However, this approach does not control for any observer agreement that may have occurred by chance (21).

Fleiss's and Cohen's Kappa statistics are useful for determining intra- and interobserver agreement (22; 23). Kappa (K) is a ratio of the proportion of times that the observers agree to the maximum proportion of times that the observers might possibly agree with both corrected for chance agreement. The kappa statistics for each value on the Likert scale (1-5) can be analyzed in addition to an overall observer evaluation in order to determine if the observers had difficulty with a particular value response.

Kappa statistics are useful when the variable classifications are nominal (e.g., osteopenia / no osteopenia); however, kappa does not take into account the degree of disagreement (i.e., all disagreement is treated equally as total disagreement). Therefore, when the categories are ordered (as in the Likert rating scales employed in the present study), it is preferable to use the weighted kappa statistic ($k_w$) (22; 24). Weights can be assigned in many different ways depending upon the study design. Linear weights are useful when the difference between the first and second category has the same importance as a difference between the second and third category. Quadratic weights are appropriate if the difference between the first and second category is less important than a difference between the second and third categories. In reality, one observer's conception of the magnitude of difference between any two answers may not coincide with those of other observers. Further, the magnitude of difference between various consecutive answers most likely does not fit a simple linear or quadratic formula. In the present study, five possible answers existed for all questions. Each answer represented a point on a continuum of possible opinions ranging from definitely absent to definitely present. A quadratic formula was applied since a difference between “definitely present” and “probably present” was less important than a difference between “probably present” and “uncertain.” This application was also chosen for simplicity and to facilitate study replication. The weights in the quadratic set were 1, 0.937, 0.750, 0.437 and 0 as calculated based on the following formula:

$$w_i = 1 - \frac{i^2}{(k-1)^2}$$

For both K and $K_w$, values range from -1 to +1. As shown in the chart below, $K_w$ equal to zero suggests that the observed agreement is equal to that expected by chance alone. $K_w$ equal to +1 indicates perfect agreement. $K_w$ less than 0 implies that the observed agreement is less than that expected by chance alone.

| **TABLE 1: STATISTICAL SIGNIFICANCE OF KAPPA** | |
|---|---|
| **Value of K** | **Interpretation** |
| K = 1 | perfect agreement |
| K = 0 | no agreement better than chance |
| K is negative | Agreement is worse than chance |

The relative significance of values between 0 and 1 has been interpreted differently depending upon the research study and author. Kramer and Feinstein (21) proposed that value of $K_w$ approaching +0.5 or +0.6 represent an "acceptable degree of agreement." The following chart shows an alternative interpretation by Altman (25):

| TABLE 2: QUALITATIVE INTERPRETATION OF KAPPA BY ALTMAN *ET AL.* | |
|---|---|
| **VALUE OF K** | **STRENGTH OF AGREEMENT** |
| < 0.20 | Poor |
| 0.21 - 0.40 | Fair |
| 0.41 - 0.60 | Moderate |
| 0.61 - 0.80 | Good |
| 0.81 - 1.00 | Very good |

Yet another more specific qualitative interpretation, which was proposed by Landis and Koch (26), was utilized in the present study and is shown below:

| TABLE 3: QUALITATIVE INTERPRETATION OF KAPPA BY LANDIS AND KOCH | |
|---|---|
| **vALUE OF k** | **Strength of Agreement** |
| <0 | Poor |
| 0-0.20 | Slight |
| 0.21-0.40 | Fair |
| 0.41-0.60 | Moderate |
| 0.61-0.80 | Substantial |
| 0.81-1.00 | Almost Perfect |

Another approach employed to interpret kappa statistics was to compare them to the null hypothesis. The null hypothesis ($H_0$) stated that any intra- and/or interobserver agreement was due to chance alone. The p-value provided the likelihood of obtaining the value of $K_w$ if $H_0$ were indeed true. If the p-value was less than or equal to a predetermined level of significance ($\alpha = 0.05$), then the null hypothesis was rejected. In order to determine the p-value, the standard error (a measurement of the precision of the estimated $K_w$) and a z-value (an approximate normal test statistic) were calculated (22). However, p-values are less important than the absolute value of $K_w$.

In order to use the standard normal z-distribution, N (total number of cases) must be greater than or equal to $2g^2$ where g equals the number of answer categories (27). In this study, the Likert scale contained 5 answer categories, so g equaled 5. Thus, the number of cases required to use the standard normal z-distribution was 50. In order to compare two $K_w$ values on a statistically robust basis, the number of cases must be greater than or equal to $3g^2$ (28), so the requisite number of cases was 75. The number of cases included in this study was 100. In some of the analyses, one or two cases could not be used to calculate a given $K_w$ because an observer had omitted an answer. However, in most analyses the number of cases was well above 75. The only exception was that the subset of cases examined by the non-chest radiologist totaled 38, so all analysis regarding that observer should be considered with caution.

Kappa statistics are well established in medical literature (29), and may be used to test rater independence as well as to quantify the level of agreement. However, some disagreement exists regarding the use of kappa to quantify actual levels of agreement. Thus, an important caveat, when evaluating the results of the present study, is that kappa statistics should not be viewed as the unequivocal standard for determining agreement. Kappa is referred to as a chance-corrected measure of agreement. However, because the expected agreement in the calculation of kappa is relevant only when the raters are independent, kappa could be viewed as not truly chance-corrected. Some critics have noted that observers are not independent because they are rating the same items. Thus, another statistic, Kendall's coefficient of concordance, was utilized to further explore the agreement relationships.

Because the variable classifications were ordinal, Kendall's coefficient of concordance (*W*) was used in addition to kappa statistics. Ordinal variables are categorical variables that have three or more possible levels with a natural ordering, such as strongly disagree, disagree, neutral, agree, and strongly agree. While kappa statistics represent absolute agreement among ratings and treat all misclassifications equally, Kendall's coefficient of concordance expresses the degree of association among the multiple ratings made by an observer using information about relative ratings and is sensitive to the seriousness of the misclassification. For example, osteopenia was rated on a 1-5 scale. The consequences of misclassifying “definitely present” (rating = 5) as “definitely absent” (rating=1) are more serious than misclassifying it as “probably present” (rating=4).

Kendall's coefficient of concordance can range from 0 to 1. The higher the value of Kendall's, then the stronger the measured association is. As with the kappa statistic, p-values were calculated to choose between the hypothesis that the ratings were associated with one another or the null hypothesis that there was no association. A chi-square statistic and the degrees of freedom were used to determine the p-value.

The purpose of studying a diagnostic test is to determine whether that test is useful in clinical practice. Therefore, causality and tests of statistical significance are less important. Instead, descriptive statistics with associated confidence intervals are used to assess test performance.

In order to compute test sensitivity and specificity, the data was dichotomized. Reader designations on the questionnaire Likert scale of 1 (definitely absent), 2 (probably absent), and 3 (uncertain) were considered "normal or non-osteopenic" while designations of 4 (probably present) and 5 (definitely present) were considered "osteopenic." Accuracy was assessed by comparing the chest x-ray findings with those of DXA, which is the non-invasive gold standard for bone mineral density. For the accuracy assessment, cases with scores less than -1, the WHO cut-off for osteopenia (5), were classified as having "presence of disease" while cases with scores greater than -1 were classified as having "absence of disease." DXA can be measured at several locations (lumbar spine, hip, femoral neck, and trochanter), and the DXA report provided separate T and Z-scores for each of these locations. Therefore, when making accuracy assessments at a given anatomical region, the specific T and Z-scores corresponding to that region were used.

When evaluating a diagnostic test, some cases with the disease will be correctly classified as positive (TP = True Positive) while other cases with the disease will be incorrectly classified as negative (FN = False Negative). Likewise, some cases without the disease will be correctly classified as negative (TN = True Negative) while others without the disease will be mistakenly classified as positive (FP = False Positive). Cases were identified as belonging to one of these four categories, and the tally for each group was then entered into a 2x2 table in order to calculate test characteristics such as sensitivity, specificity, positive and negative likelihood ratio, disease prevalence as well as positive and negative predictive power. The following 2x2 table displays the possible outcomes:

| **TABLE 4: POSSIBLE OUTCOMES OF A DIAGNOSTIC TEST** | | | | | |
|---|---|---|---|---|---|
| **Test** | **Disease** | | | | |
| | **Present** | **n** | **Absent** | **N** | **Total** |
| **Positive** | True Positive (TP) | *a* | False Positive (FP) | *C* | *a + c* |
| **Negative** | False Negative (FN) | *b* | True Negative (TN) | *d* | *b + d* |
| **Total** | | *a + b* | | *c + d* | |

Test sensitivity is the probability that a test result will be positive if the disease is truly present. Test specificity is the probability that a test result will be negative if the disease is truly not present. The positive likelihood ratio is a ratio between the probability of a positive test result given true disease and the probability of a positive test result given the absence of disease. The negative likelihood ratio is a ratio between the probability of a negative test result given true disease and the probability of a negative test result given disease absence. Positive predictive value is the probability that the disease is present when the test is positive while negative predictive value is the probability that the disease is not present when the test is negative. The above definitions can be calculated from the previous chart as displayed in the chart below:

| TABLE 5: CALCULATION OF TEST DESCRIPTIVE STATISTICS | | | |
|---|---|---|---|
| **Sensitivity** | a<br>a + b | **Specificity** | D<br>c + d |
| **Positive Likelihood Ratio** | Sensitivity<br>1 - Specificity | **Negative Likelihood Ratio** | 1 - Sensitivity<br>Specificity |
| **Positive Predictive Value** | a<br>a + c | **Negative Predictive Value** | d<br>b + d |

The sample size in the disease present and disease absent groups was equal and, therefore, did not reflect the real prevalence of the disease in the population as a whole. Positive and negative predictive values depend upon the prevalence of disease in the population studied. Therefore, likelihood ratios were calculated in lieu of predictive values. Because it is calculated from the sensitivity and specificity, the likelihood ratio (LR) is a stable operating test characteristic (i.e., not effected by disease prevalence).

A value of LR equal to 1.0 indicates that a test fails to change the opinion regarding disease probability from the pretest to posttest estimation. Positive likelihood ratios (LR+) are greater than 1.0, and as the number increases, so does the likelihood of the patient having the disease after a positive test result. Negative likelihood ratios (LR-) have values less than 1.0, and smaller numbers indicate a lower risk for disease. The following charts provide two common interpretations of likelihood ratios.

| TABLE 6: QUALITATIVE INTERPRETATION OF POSITIVE AND NEGATIVE LIKELIHOOD RATIOS | | |
|---|---|---|
| **Qualitative Strength** | **LR(+)** | **LR(-)** |
| Excellent | 10 | 0.1 |
| Very good | 6 | 0.2 |
| Fair | 2 | 0.5 |
| Useless | 1 | 1 |

| TABLE 7: ALTERNATIVE INTERPRETATION GUIDE FOR POSITIVE AND NEGATIVE LIKELIHOOD RATIOS | | | |
|---|---|---|---|
| | Poor-fair | Good | Excellent |
| Positive likelihood ratio | 2.1 - 5.0 | 5.1 - 10.0 | >10 |
| Negative likelihood ratio | 0.5 - 0.2 | 0.19 - 0.1 | <0.1 |

STUDY OVERSIGHT

This research study was approved by the Yale University School of Medicine Human Investigation Committee (HIC# 0810004330) who determined that signed consent from the patients was not necessary. There were no risks to the patients as a result of this study. Patients did not undergo any additional testing or questioning during this project. No changes or additions were made to the patient's medical record or within PACS. Although the readers may have identified signs of osteopenia on a patient's radiograph, which had not been noted in the report previously, this finding did not impact the patient's medical care since all participating patients had also already received a diagnosis based on the gold-standard of bone mineral density testing by DXA. A waiver of signed consent regarding the readers (physician-subjects) was also approved by the HIC, so only verbal consent from these physician-subjects was obtained and no personally identifiable information was recorded regarding them.

# RESULTS

## PATIENT PROFILE

Of the 100 patients in this study cohort, 80 were female and 20 were male. Their ages ranged from 16 to 86 years with a mean of 57.85 and median of 60 years as displayed in the following frequency histogram. No information regarding ethnicity was available.

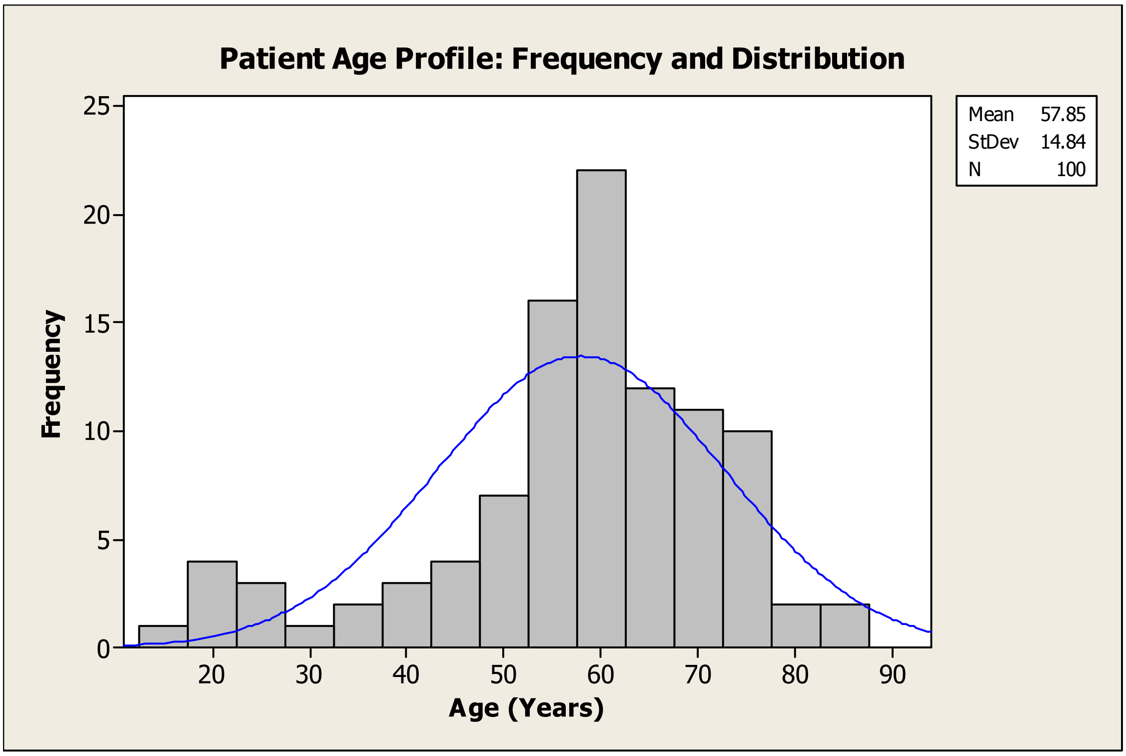


**Fig. 1: Frequency and distribution of ages of patients in the study sample.**

## INTRAOBSERVER AGREEMENT: FIRST CHEST RADIOLOGIST

The following table demonstrates the analysis of intraobserver agreement between two trials for the first chest radiologist as determined using weighted kappa statistics. The reader's overall assessment of the presence/absence of osteopenia yielded "substantial" agreement with a value of kappa ($K_w$ = 0.718) that was higher than that of any the specific radiologic findings associated with osteopenia. Of those specific findings, body shape (wedged) and end plates shape (fish vertebra) had the highest weighted kappa values of 0.654 and 0.638, respectively, which also corresponded to substantial intraobserver agreement. Both end plate definition compared to the vertebral body ($K_w$ = 0.470) and vertical striations ($K_w$ = 0.593) were associated with a moderate degree of intraobserver consistency. The least intraosbserver agreement was observed for the empty box appearance ($K_w$ = 0.162) and compression fractures ($K_w$ = -0.014).

**TABLE 8: INTRAOBSERVER CONSISTENCY FOR READER 1 BASED ON WEIGHTED KAPPA**

| Criteria | Chest Radiologist #1 | |
|---|---|---|
| | $K_w$ | Strength of Agreement |
| End Plates Definition (Compared to Vertebral Body) | 0.470 | Moderate |
| End Plates Shape (Fish Vertebra) | 0.638 | Substantial |
| Body Shape (Wedged) | 0.654 | Substantial |
| Compression Fractures | -0.014 | Poor |
| Vertical Striations | 0.593 | Moderate |
| Empty Box | 0.162 | Slight |
| Presence of Osteopenia | 0.718 | Substantial |

When the first chest radiologist's responses were analyzed using Kendall's Coefficient of Concordance, the highest statistically significant intraobserver agreement was, again, seen for the overall assessment of osteopenia ($W$ = 0.865317; p = 0.0000). As in the analysis using weighted kappa values, compression fractures had the lowest concordance ($W$ = 0.492230).

Because the p value was 0.5246, the association between the first and second reading was not greater than that expected by chance alone. The empty box appearance ($W$ = 0.605198) also did not reach statistical significance (p = 0.0758). For all other specific radiologic findings, $p < 0.05$ (0.0046, 0.0001, 0.0001, and 0.0006), so the null hypothesis was rejected thus supporting the hypothesis that the ratings for those criteria were significantly associated with one another. For each of these statistically significant associations, the concordance was high with $W$ = 0.712959 for end plates definition compared to vertebral body, $W$ = 0.820927 for fish vertebra, $W$ = 0.803161 for wedged body shape, and $W$ = 0.761598 for vertical striations.

**TABLE 9: INTRAOBSERVER CONSISTENCY FOR READER 1 BASED ON KENDALL'S COEFFICIENT**

| **Criteria** | **Chest Radiologist #1** | |
|---|---|---|
| | ***W*** | **P value** |
| End Plates Definition (Compared to Vertebral Body) | 0.712959 | 0.0046 |
| End Plates Shape (Fish Vertebra) | 0.820927 | 0.0001 |
| Body Shape (Wedged) | 0.803161 | 0.0001 |
| Compression Fractures | 0.492230 | 0.5246 |
| Vertical Striations | 0.761598 | 0.0006 |
| Empty Box | 0.605198 | 0.0758 |
| Presence of Osteopenia | 0.865317 | 0.0000 |

INTRAOBSERVER AGREEMENT: SECOND CHEST RADIOLOGIST

The following table also displays the analysis of intraobserver agreement for the second chest radiologist across two trials. As for the first reader, the overall assessment of osteopenia demonstrated “substantial” agreement ($K_w$ = 0.630). Likewise, the specific criterion of end plate shape (fish vertebra) had substantial agreement ($K_w$ = 0.712). Specific findings associated with “moderate” agreement included end plates definition compared to vertebral bodies ($K_w$ = 0.540), wedged body shape ($K_w$ = 0.533), and the empty box appearance ($K_w$ = 0.577). Vertical striations resulted in fair agreement ($K_w$ = 0.365). As with the first reader, the least agreement was seen for compression fractures ($K_w$ = 0.179).

**TABLE 10: INTRAOBSERVER CONSISTENCY FOR READER 2 BASED ON WEIGHTED KAPPA**

| **Criteria** | **Chest Radiologist #2** | |
|---|---|---|
| | **$K_w$** | **Strength of Agreement** |
| End Plates Definition (Compared to Vertebral Body) | 0.540 | Moderate |
| End Plates Shape (Fish Vertebra) | 0.712 | Substantial |
| Body Shape (Wedged) | 0.533 | Moderate |
| Compression Fractures | 0.179 | Slight |
| Vertical Striations | 0.365 | Fair |
| Empty Box | 0.577 | Moderate |
| Presence of Osteopenia | 0.630 | Substantial |

When the results for the second chest radiologist were analyzed using Kendall's Coefficient of Concordance, the highest degree of intraobserver agreement occurred for end plates shape/fish vertebra (W = 0.864783), the empty box appearance (W = 0.837777), and the presence of osteopenia (W = 0.808476). The overall assessment as well as each of the specific radiographic findings were statistically significant ($p < 0.05$) except for compression fractures (W = 0.597192; p = 0.0920). Thus, for all criteria except compression fractures, the null hypothesis was rejected and the data supported the hypothesis that the ratings across the two trials were significantly associated with one another.

**TABLE 11: INTRAOBSERVER CONSISTENCY FOR READER 2 BASED ON KENDALL'S COEFFICIENT**

| | **Chest Radiologist #2** | |
|---|---|---|
| **Criteria** | ***W*** | **P value** |
| End Plates Definition (Compared to Vertebral Body) | 0.773715 | 0.0004 |
| End Plates Shape (Fish Vertebra) | 0.864783 | 0.0000 |
| Body Shape (Wedged) | 0.793838 | 0.0002 |
| Compression Fractures | 0.597192 | 0.0920 |
| Vertical Striations | 0.713955 | 0.0034 |
| Empty Box | 0.837777 | 0.0002 |
| Presence of Osteopenia | 0.808476 | 0.0001 |

INTEROBSERVER AGREEMENT: BOTH CHEST RADIOLOGISTS

The consistency between the responses of both chest radiologists on the first read was substantial for the specific radiographic finding of a wedged vertebral body shape, which yielded the highest kappa ($K_w$ = 0.622) and

Kendall's coefficient (*W* = 0.780020; p = 0.0003). A moderate degree of agreement was found for end plate shape/fish vertebra which had the next highest kappa ($K_w$ = 0.45) and Kendall's coefficient (*W* = 0.724328; p = 0.0025). The consistency for the overall assessment of osteopenia had a $K_w$ of 0.383, which is only fair. Vertical striations ($K_w$ = 0.295) and the empty box appearance ($K_w$ = 0.387) were fair. However, only the presence of osteopenia (*W* = 0.688669) and vertical striations (*W* = 0.632159), but not the empty box appearance (*W* = 0.631779) showed statistically significant relationships ($p < 0.05$) based on Kendall's coefficients. The two chest radiologists demonstrated the least amount of agreement over end plate definition compared to the vertebral body ($K_w$ = 0.175; *W* = 0.600670) and compression fractures ($K_w$ = 0.141; *W* = 0.645192), which also did not meet the criteria for statistical significance ($p < 0.05$) using Kendall's coefficients.

**TABLE 12: INTEROBSERVER AGREEMENT FOR BOTH CHEST RADIOLOGISTS USING WEIGHTED KAPPA AND KENDALL'S**

| **Criteria** | **$K_w$** | **Strength of Agreement** | ***W*** | **P value** |
|---|---|---|---|---|
| End Plates Definition (Compared to Vertebral Body) | 0.175 | Slight | 0.600670 | 0.0882 |
| End Plates Shape (Fish Vertebra) | 0.45 | Moderate | 0.724328 | 0.0025 |
| Body Shape (Wedged) | 0.622 | Substantial | 0.780020 | 0.0003 |
| Compression Fractures | 0.141 | Slight | 0.645192 | 0.0286 |
| Vertical Striations | 0.295 | Fair | 0.632159 | 0.0389 |
| Empty Box | 0.387 | Fair | 0.631779 | 0.0579 |
| Presence of Osteopenia | 0.383 | Fair | 0.688669 | 0.0076 |

EFFECT OF OBSERVER BIAS ON OBSERVER AGREEMENT

It is possible that some of the interobserver disagreement was due to reader bias. An idea of possible reader bias can be obtained by examining the following frequency distribution graphs plotted for the overall assessment of osteopenia for each observer's first reading. As can be seen, the first chest radiologist was less "uncertain" (rating 3) than the second chest radiologists in terms of the frequency of response.

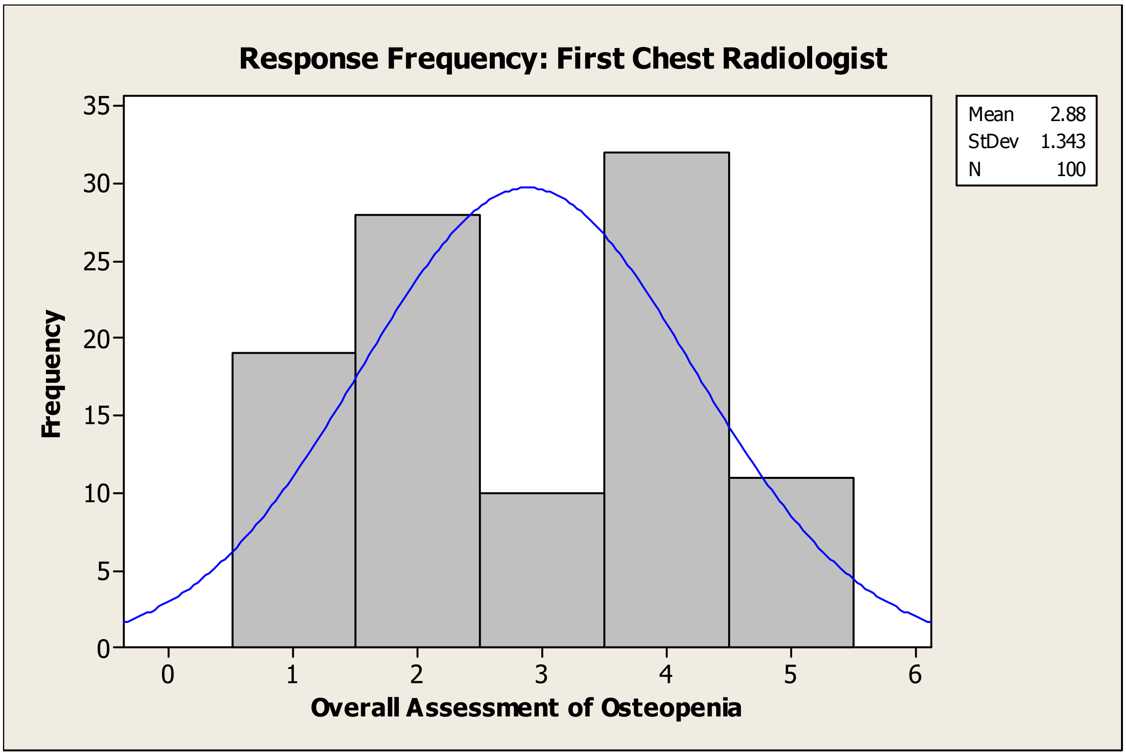


**Fig. 2: Response frequency for the first chest radiologist. Values on the x-axis correspond to ratings on the Likert scale for the overall assessment of osteopenia.**

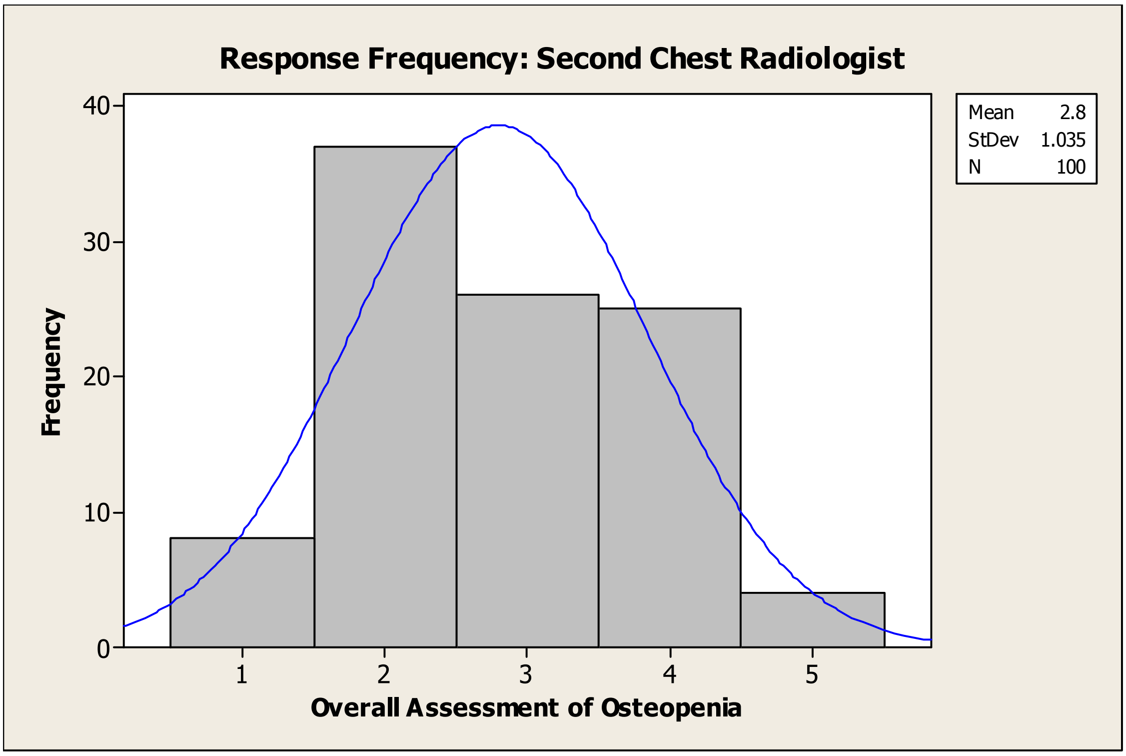


**Fig. 3: Response frequency for the second chest radiologist. Values on the x-axis correspond to ratings on the Likert scale for the overall assessment of osteopenia.**

In addition to reader bias, differences in the frequency and distribution of responses may reflect a disproportionate number of diseased and non-diseased test cases. While the study was designed to eliminate this possible confounder by assigning each group an equal number of cases, the non-chest radiologist only read a subset of randomly selected cases. Post-hoc analysis (which can be found in Appendix B) revealed that the prevalence of disease in this subset of cases was slightly lower than that of the total sample. This most likely accounts for the preponderance of “definitely absent” (rating 1) responses by the non-chest radiologist.

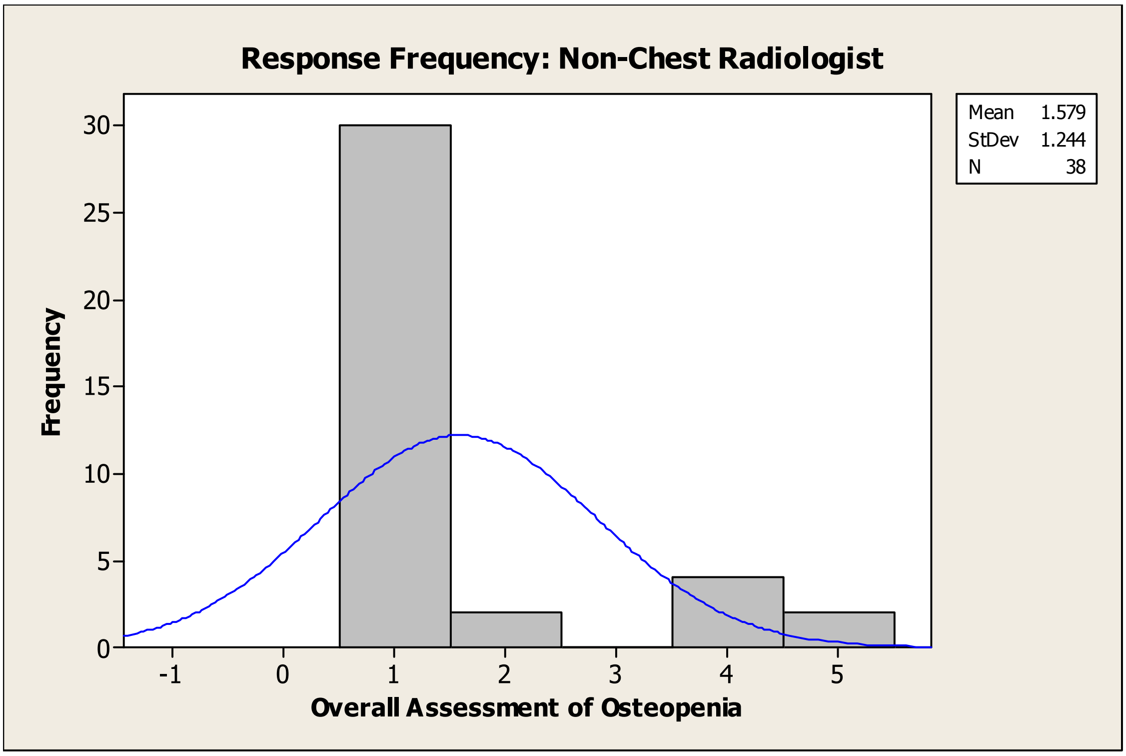


**Fig. 4: Response frequency for the non-chest radiologist. Values on the x-axis correspond to ratings on the Likert scale for the overall assessment of osteopenia.**

ACCURACY ASSESSMENT

As summarized in the following three charts grouped by individual reader, specificity was generally high while sensitivity was relatively low. Although DXA can be measured at multiple anatomic locations, the worst score of either the spine or hip is used clinically as the gold standard for osteoporosis diagnosis (4; 5). This was the same approach used in the present study to categorize cases into one of three experimental groups for the intra- and interobserver consistency assessments. However, accuracy assessments were made for each of the anatomic sites, using the T- and Z-scores corresponding to that specific site, to determine which cases would be classified as having true disease. Correlating the specific findings to the different DXA measurement sites showed some variability in sensitivity and specificity. For the overall assessment of osteopenia, no particular site appeared to be better except the lumbar spine for one reader.

For all three radiologists, the highest specificities were related to compression fractures. Specifically, the specificity for compression fractures, as rated by the first chest radiologist, was 100% when BMD was measured in the hip, femoral neck, and trochanter and 97.62% when measured in the lumbar spine. When the second chest radiologist rated compression fractures, the specificity was 100% for the trochanter comparison and similarly high in the lumbar spine (95.24%), hip (98.18%), and femoral neck (97.87%). Finally, there was 100% specificity across all DXA measurement locations as rated by the non-chest radiologist.

The empty box appearance also had extremely high specificities across all readers and locations. Specifically, the specificity was 100% for the lumbar spine, femoral neck, and trochanter and 98.18% for the hip as rated by the first chest radiologist. The second chest radiologist’s specificity values were 100% in the lumbar spine, 95.24% in the hip, 97.30% in the femoral neck, and 96.77% in the trochanter. The specificity was 100% in all locations when rated by the non-chest radiologist. However, the empty box appearance was very rare as the radiologists only identified a maximum of 6 cases out of 100 with the finding.

| TABLE 13: SENSITIVITIES AND SPECIFICITIES FOR FIRST CHEST RADIOLOGIST | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **End Plates Definition (Compared to Vertebral Body)** | **End Plates Shape (Fish Vertebra)** | **Body Shape (Wedged)** | **Compression Fractures** | **Vertical Striations** | **Empty Box** | **Osteopenia?** |
| **L – Spine** | | | | | | | |
| Sensitivity (95% CI) | 20.37% | 34.38% | 31.03% | 1.75% | 18.97 | 10.34 | 51.72% |
| Specificity (95% CI) | 71.43% | 76.19% | 76.19% | 97.62% | 85.71 | 100.00 | 69.05% |
| **Hip** | | | | | | | |
| Sensitivity (95% CI) | 21.62% | 42.5% | 35.00% | 5.13% | 20.00 | 10.00 | 60.00% |
| Specificity (95% CI) | 74.55% | 78.18% | 76.36% | 100.00% | 87.27 | 98.18 | 70.91% |
| **Fem Neck** | | | | | | | |
| Sensitivity (95% CI) | 21.74% | 34.69% | 34.69% | 4.17% | 18.37 | 12.24 | 59.18% |
| Specificity (95% CI) | 71.74% | 74.47% | 76.60% | 100.00% | 85.11 | 100.00 | 72.34% |
| **Trochanter** | | | | | | | |
| Sensitivity (95% CI) | 27.78% | 35.90% | 35.90% | 5.26% | 20.51 | 10.26 | 51.28% |
| Specificity (95% CI) | 71.05% | 73.68% | 71.05% | 100.00% | 84.21 | 100.00 | 68.42% |

| TABLE 14: SENSITIVITIES AND SPECIFICITIES FOR SECOND CHEST RADIOLOGIST | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **End Plates Definition (Compared to Vertebral Body)** | **End Plates Shape (Fish Vertebra)** | **Body Shape (Wedged)** | **Compression Fractures** | **Vertical Striations** | **Empty Box** | **Osteopenia?** |
| **L – Spine** | | | | | | | |
| Sensitivity (95% CI) | 50.00% | 45.61% | 24.14% | 7.02% | 3.45% | 12.50% | 44.83% |
| Specificity (95% CI) | 76.19% | 69.05% | 85.71% | 95.24% | 97.62% | 100.00% | 92.86% |
| **Hip** | | | | | | | |
| Sensitivity (95% CI) | 45.00% | 52.50% | 27.50% | 7.69% | 7.50% | 9.09% | 37.50% |
| Specificity (95% CI) | 63.64% | 68.52% | 83.64% | 98.18% | 100.00% | 95.24% | 76.36% |
| **Fem Neck** | | | | | | | |
| Sensitivity (95% CI) | 42.86% | 46.94% | 24.49% | 8.33% | 4.08% | 12.82% | 34.69% |
| Specificity (95% CI) | 63.83% | 69.57% | 82.98% | 97.87% | 97.87% | 97.30% | 76.60% |
| **Trochanter** | | | | | | | |
| Sensitivity (95% CI) | 43.59% | 51.28% | 30.77% | 10.53% | 7.69% | 9.68% | 41.03% |
| Specificity (95% CI) | 63.16% | 67.57% | 84.21% | 100.00% | 100.00% | 96.77% | 78.95% |

| TABLE 15: SENSITIVITIES AND SPECIFICITIES FOR NON-CHEST RADIOLOGIST | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **End Plates Definition (Compared to Vertebral Body)** | **End Plates Shape (Fish Vertebra)** | **Body Shape (Wedged)** | **Compression Fractures** | **Vertical Striations** | **Empty Box** | **Osteopenia?** |
| **L – Spine** | | | | | | | |
| Sensitivity (95% CI) | 19.05% | 0.00% | 9.52% | 0.00% | 9.52% | 0.00% | 14.29% |
| Specificity (95% CI) | 81.25% | 94.12% | 88.24% | 100.00% | 82.35% | 100.00% | 82.35% |
| **Hip** | | | | | | | |
| Sensitivity (95% CI) | 20.00% | 0.00% | 13.33% | 0.00% | 13.33% | 0.00% | 20.00% |
| Specificity (95% CI) | 83.33% | 94.74% | 89.47% | 100.00% | 89.47% | 100.00% | 89.47% |
| **Fem Neck** | | | | | | | |
| Sensitivity (95% CI) | 22.22% | 0.00% | 11.11% | 0.00% | 16.67% | 0.00% | 16.67% |
| Specificity (95% CI) | 82.35% | 94.44% | 88.89% | 100.00% | 88.89% | 100.00% | 83.33% |
| **Trochanter** | | | | | | | |
| Sensitivity (95% CI) | 23.08% | 0.00% | 23.08% | 0.00% | 15.38% | 0.00% | 23.08% |
| Specificity (95% CI) | 78.57% | 93.33% | 100.00% | 100.00% | 86.67% | 100.00% | 93.33% |

As can be seen in the chart below, most of the highest positive likelihood ratios were related to an overall assessment of osteopenia. This was true for all three radiologists. For the first chest radiologist, the best +LRs regarding overall osteopenia were found in the hip (2.06), femoral neck (2.14), and lumbar spine (2.03). For the second chest radiologist, the best sites were in the lumbar spine (6.28) and trochanter (1.95). Finally, for the non-chest radiologist, the best sites for a +LR of overall osteopenia were the trochanter (3.46) and hip (1.90). It is important to note that the highest +LR in the entire study was the 6.28 found for the osteopenia assessment in the lumbar spine as rated by the second chest radiologist. Along with this very good +LR, the –LR of 0.59 approached a fair strength as well.

| TABLE 16: BEST POSITIVE LIKELIHOOD RATIOS | | | | |
|---|---|---|---|---|
| **Radiologist** | **Finding** | **Location** | **+LR** | **Strength** |
| First Chest | Empty Box | Hip | 5.50 | Very Good |
| Second Chest | Osteopenia? | Lumbar Spine | 6.28 | Very Good |
| First Chest | End Plate Shape (Fish Vertebra) | Hip | 1.95 | Fair |
| First Chest | Osteopenia? | Hip | 2.06 | Fair |
| First Chest | Osteopenia? | Femoral Neck | 2.14 | Fair |
| First Chest | Osteopenia? | Lumbar Spine | 2.03 | Fair |
| Second Chest | End Plates Definition | Lumbar Spine | 2.10 | Fair |
| Second Chest | Body Shape (Wedged) | Trochanter | 1.95 | Fair |
| Second Chest | Compression Fractures | Hip | 4.23 | Fair |
| Second Chest | Compression Fractures | Femoral Neck | 3.92 | Fair |
| Second Chest | Vertical Striations | Femoral Neck | 1.92 | Fair |
| Second Chest | Empty Box | Hip | 1.91 | Fair |
| Second Chest | Empty Box | Femoral Neck | 4.74 | Fair |
| Second Chest | Empty Box | Trochanter | 3.00 | Fair |
| Second Chest | Osteopenia? | Trochanter | 1.95 | Fair |
| Non-Chest | Osteopenia? | Hip | 1.90 | Fair |
| Non-Chest | Osteopenia? | Trochanter | 3.46 | Fair |

For further reference, Appendix B contains extensive charts which detail all descriptive statistics with 95% confidence intervals across all DXA measurement locations for each radiologist. Listings of the exact number of cases, true positives, false positives, true negatives, and false negatives can also be found in the appendix.

## DISCUSSION

Overall, radiologists were good at identifying signs of late, but not early, disease. Intraobserver consistency was substantial for fish vertebrae with moderate interobserver agreement. Similarly for wedged vertebrae, intraobserver consistency was substantial to moderate with substantial interobserver agreement. These radiographic signs correlated with true disease as shown by high specificity values. Therefore, this study indicates that if osteopenia is suspected (i.e., there is a wedge or fish vertebra) or its associated features are seen on a CXR, it is crucial for radiologists to comment on it.

Indeed, for the majority of radiographic findings, intraobserver agreement was greater than or equal to the $K_w = 0.5$ cutoff suggested by Kramer and Feinstein (21) as indicative of adequate reader consistency, but interobserver agreement was slightly less in general. Specificity was generally high across all readers and variables while sensitivity was relatively low. Many of the highest positive likelihood ratios (+LR) were related to an overall assessment of osteopenia. This was true for all three radiologists.

The highest specificities (at or near 100%) were related to compression fractures. Williamson *et al.* (16) similarly found a general overall correlation between BMD as determined by dual photon densitometry of vertebral bodies and lateral chest film assessment, but concluded that there was little ability to reliably diagnose osteoporosis in the absence of vertebral compression fractures.

INTERPRETATION OF INTEROBSERVER CONSISTENCY

Prior research (30) has suggested that radiologists may perceive an overall impression of a given chest film by applying a mental template rather than summing multiple decisions regarding specific criteria. In other words, radiologists may rely on a Gestalt rather than the presence or absence of specific findings. The results of this study showed that, whatever template the radiologists employed to come to their conclusions about the presence or absence of osteopenia, they were individually able to consistently evaluate lateral chest films for this finding.

However, the threshold or trigger point for positive or negative disease differed between the radiologists. The interobserver variability might reflect the inclusion of both osteopenia and osteoporosis cases in this study. It is possible that these readers were able to diagnose severe osteoporosis fairly confidently, but differed in their assessments of more subtle cases (i.e., mildly osteopenic patients). Future research could be done to sort this out by comparing the osteopenic and osteoporotic experimental groups directly.

A logical question is whether or not the findings of this study are applicable to radiologists who do not specialize in chest radiology, so the ratings from a non-chest radiologist were individually compared to the first and second chest radiologists on a subset of cases. Interestingly, only a fair level of agreement was seen regarding the overall assessment of osteopenia. Again, this may reflect differing thresholds for making a diagnosis of osteopenia. Of note, the degree of consistency between the non-chest radiologist and the two chest radiologists for the specific findings of wedged body shape and end plate shape/fish vertebra was much lower than the degree of agreement between the two chest radiologists. Differences in training and experience were less likely to have accounted for the variation between the two chest radiologists because they both routinely see chest films. This may have been a source of difference in the readings between the chest and non-chest radiologist. Unfortunately, the numbers were very small for this subset.

Training during the course of the research project is another potential source of interobserver variation as the radiologists may have become more adept at looking for osteopenia and its components with practice. This was minimized by using only the ratings from the first read for analysis of interobserver agreement. In theory, it could have had a negative impact on the intraobserver consistency; however, intraobserver consistency was already substantially high in this study.

It is possible that some of the observed interobserver disagreement was due to reader bias. Radiologists may tend to over- or under-call findings relative to other radiologists, while they are usually consistent within themselves. For example, prior research (30) has demonstrated that some radiologists (so-called "wet" readers) have a lower threshold for calling CHF, while other radiologists (“dry” readers) have a higher trigger point. Based upon the frequency distribution graphs for the overall assessment of osteopenia for each observer's first reading, the first chest radiologist was less "uncertain" than the second chest radiologists in terms of the frequency of response. This might reflect a more cautious approach or less experience on the part of the second reader compared to the first. Additionally, the non-chest radiologist exhibited more "definitely absent" ratings, which was reflective of the slightly lower prevalence of true disease in the subset of cases examined by the non-chest radiologist as compared to the total sample reviewed by the chest radiologists.

COMPARISON TO GOLD STANDARD

When evaluating the accuracy of a clinical test, it is important to compare the observers' findings with the true disease status (i.e., results from a gold standard test) (31). Because intra- and interobserver statistics do not make a comparison to the gold standard, these ratings may not necessarily have been correct, even though they were consistent. However, intra- and interobserver statistics are still expedient to determining the usefulness of a test. If a test is wonderfully accurate, but no one can reproduce it, then it is not really very good.

The higher the sensitivity and specificity of a test, then the greater its accuracy is. In an ideal world, the perfect test would have a sensitivity of 100% and specificity of 100% (i.e., minimal false negative and false positive results). However, this is rarely the case. No general agreement exists regarding what constitutes an acceptable level of sensitivity and specificity.

Often the acceptable levels vary depending upon the intent of the test, the prevalence of the condition in the test group, the availability of alternate methods for assessment as well as the cost versus benefits of testing (35).

Using different cut-off criteria for either the test or the gold standard will result in different values for sensitivity and specificity. Selecting a higher cut-off causes the false positives to decrease and the specificity to increase, but the true positives and sensitivity will decrease. Selecting a lower cut-off increases the true positives and sensitivity, but also increases the false positives and true negatives, so specificity decreases. In the present study, the cut-off values for the gold standard were set by the WHO guidelines (5), but the cut-offs for the reader response ratings could have been different. When the data was dichotomized, the Likert scale ratings of “definitely osteopenic” (value 5) and “probably osteopenic” (value 4) were combined. If, instead, value 4 had been added to values 1-3, then the specificity of the test may have been higher because only cases which the reader labeled as “definitely osteopenic” would have been counted as a “positive” test result. Although, the specificity values were already high in the present study, it would be interesting to evaluate the data using this alternative cut-off in the future for comparison.

Since different diagnostic tests for the same disease trade sensitivity for specificity and vice versa, physicians often use more than one test to make a diagnosis. A highly sensitive test is ideal to screen for a disease; a highly specific test is best to confirm (19). In other words, a highly sensitive test with a negative result is good at ruling-out a disease while a highly specific test with a positive result is good at ruling-in the disease. Given the low sensitivity of CXR identification of osteopenia and its associated features, the findings of this study suggests that this method should not be used in place of DXA as a screening method, and radiologists should be reluctant to comment on the absence of osteopenia. However, given the high specificity, the results of this study show that, if osteopenia is suspected (i.e., there is a wedge or fish vertebra) or its associated features are seen on a CXR (i.e., a positive test result), it is crucial to comment on it since the finding likely represents true disease.

It may be helpful to consider the example of another commonly used clinical test in order to understand these findings. Zaman *et al.* (36) reported the sensitivity and specificity of the nitrite dipstick test in diagnosing urinary tract infections (UTIs) in hospitalized inpatients to be 27% and 94%, respectively. The sensitivity is low, but the specificity is high. With its low sensitivity, a patient may indeed have a UTI, but test negative. On the other hand, given its high specificity such that so many people without the test have a negative test result, if a patient should test positive, it is likely that the patient does have a UTI. This illustrates how a highly specific test is most helpful to a clinician when the result is positive.

A diagnostic test is best used to supplement rather than substitute for clinical judgment, but calling attention to positive cases of osteopenia could help, for instance, in the setting of patients who have not received a DXA screening test. This might include women whose physicians are waiting till age 65 for DXA screening to be initiated based on the current practice guidelines (37) or women who do not have access to routine primary care. This is especially important in emergency room and inpatient units which treat a high percentage of the latter group of patients. Ghelback *et al.* (38) found that fewer than one-quarter of 65 hospitalized women were diagnosed with a vertebral fracture on their chest radiograph report and even fewer were treated for osteoporosis. Another group who could potentially be helped by the ability of chest plain films to suggest osteoporosis are women who do not meet their clinician's idea of "at-risk." For example, Neuner *et al.* (39) found that women younger than 50 as well as those 90 and older were less likely to be diagnosed with osteoporosis, whereas women with a prior hip or radial fracture or back pain were more likely to be diagnosed with osteoporosis. Further, clinicians have been shown to be less likely to diagnose osteoporosis in men and premenopausal women (39) despite the high prevalence of secondary causes of osteoporosis in men (40). Research has shown that primary care physicians may see postmenopausal women with atraumatic vertebral compression fractures as good candidates for osteoporosis treatment without BMD testing (41; 42). However, even for these patients who have been correctly identified as "at-risk," formal diagnostic testing has been shown to increase rates of treatment and to help in monitoring the effectiveness of treatment (43; 44).

DIFFERENT DXA MEASUREMENT SITES

The following is a discussion of the relationship of the individual sensitivities and specificities with varying DXA measurement sites. An important caveat is that even a small amount of error in a gold standard (which is the best, but by no means, the perfect test of true disease) can create the incorrect appearance of considerable error in the test that is being investigated (32). One possible source of error in the gold standard of DXA is that BMD can be measured at multiple locations; the lumbar spine, hip, femoral neck, and trochanter are common sites. Though not observed in the present study, past research has shown that BMD is not always consistent across locations. For example, Reinbold *et al.* (33) found that readings of the spine versus appendicular regions did not correlate well. Another study demonstrated discrepancies in the proportion of women with osteoporosis when comparisons were made between the spine and hip (34). Part of the variability may be due to confounding factors specific to individual patients at each site. For example, aortic calcification may result in a falsely elevated BMD as measured in the lumbar spine while increased abdominal adiposity may result in a falsely decreased BMD in this same region. Fibrosis secondary to degenerative joint disease may yield an increased BMD in multiple regions; although, the degree of sclerosis and formation of osteophytes may actually be diminished in elderly patients with osteoporosis (9). Compression fractures are a sequela of osteopenia, but they may contribute to an elevated BMD measurement by DXA secondary to compaction of trabeculae and callus formation. The area of greatest vulnerability in the hip is the femoral neck, but this is an anatomically small region, so the DXA reading is subject to fairly wide confidence intervals.

In the present study, correlating the specific findings to the different DXA measurement sites showed some variability in sensitivity and specificity values. For the overall assessment of osteopenia, no particular site appeared to be better except the lumbar spine for one reader.

## STUDY LIMITATIONS

The study's cohort was limited to one medical center, but images had been previously obtained using different machines. However, all instruments were presumably calibrated using the manufacturer's internal standard, and all testing was conducted by licensed technicians who had completed training by the manufacturer of the equipment they were using. Further, all DXA reports had been performed by the same radiologist.

## IMPLICATIONS

Once radiologists begin to include findings of osteopenia in their impressions, the effect of this practice could be assessed with regard to referral patterns for bone density testing. Neuner *et al.* (39) found that out of over 200 patients who had vertebral compression fractures noted on routine radiographs, only 38% were subsequently diagnosed by their primary care physician as having osteoporosis and only 32% received prescription medications for osteoporosis. Patients with vertebral compression fractures benefit from osteoporosis treatment and recent randomized controlled trials confirm that recurrent vertebral (46; 47) and hip (48) fractures can be prevented in these patients. However, patients with hip and radial (49-52) fractures are unlikely to be treated for osteoporosis.

If primary care physicians did not act upon a radiologist's report of a fracture, then it is unlikely that they will do so for other specific features or a general assessment of osteopenia. In the Neuner *et al.* (39) study, a lack of diagnosis and follow-up were associated with documentation of fracture in the "body" (rather than the "impression") section of the radiograph report. Thus, in order to effect a change in treatment based on radiographic reports, it will likely be necessary to educate practicing radiologists and their trainees to the significance of expressing their findings in a way that will indicate their importance. This will likely translate into less usage of such stock phrases as "the bones are unremarkable" and the inclusion of such findings as compressions fractures and fish shaped vertebral bodies as line items in the impression of a report. It will not only be important to remind radiologists of the significance of findings previously felt to be incidental, but to remind them that compression fractures of the vertebrae are not a "normal" accompaniment of aging. By including these findings in the impression of the report, there will be an increased chance that clinicians will note them and act upon them.

Because effective interventions for women with osteoporosis exist, the disability associated with osteoporosis is no longer an inevitable result of aging but rather a preventable disease (4). Therefore, it is expedient to the promotion of women's health to identify women who should receive a work-up for osteoporosis. The findings of this study will help achieve that goal by providing the impetus for a major change in the practice of chest radiology.

## REFERENCES

(1) Siris, E.S., Miller, P.D., Barrett-Connor, E., Faulkner, K.G., Wehren, L.E., *et al.* 2001. Identification and fracture outcomes of undiagnosed low bone mineral density in postmenopausal women: results from the National Osteoporosis Risk Assessment, *JAMA.* 286(22):2815-22.

(2) Lips, P. 1997. Epidemiology and predictors of fractures associated with osteoporosis. *Am J Med.* 103(2A):3S-11S.

(3) National Osteoporosis Foundation. 1998. Osteoporosis: review of the evidence for prevention, diagnosis, and treatment and cost-effectiveness analysis. *Osteoporos Int.* 8(suppl 4):S1-S88.

(4) Lentle, B.C., and Prior, J.C. 2003. Osteoporosis: what a clinician expects to learn from a patient's bone density examination. *Radiology.* 228(3):620-8.

(5) World Health Organization. 1994. Assessment of osteoporotic fracture risk and its role in screening for postmenopausal osteoporosis. In *WHO Technical Report Serices, no. 843.* Geneva, Switzerland: World Health Organization.

(6) Black, D.M., Cummings, S.R., Genant, H.K., Nevitt, M.C., Palermo, L., *et al.* 1992. Axial and appendicular bone density predict fractures in older women. *J Bone Miner Res.* 7:633-638.

(7) National Osteoporosis Foundation. BMD Testing: What the numbers mean. Accessed August 28, 2008 at http://www.nof.org/osteoporosis/bmdtest.htm#test_results.

(8) Resnick, D., and Niwayama, G. 1988. *Diagnosis of Bone and Joint Disorders.* Philadelphia: W. B. Saunders. 1939-2754.

(9) Novelline, R.A. 2004. *Squire's Fundamentals of Radiology.* Harvard University Press.

(10) Jergas, M., and Genant, H.K. 2001. Imaging of osteoporosis. In *Osteoporosis*. R. Marcus , D. Feldman D, and J.Kelsey, editors. New York: Academic Press. 411-431.

(11) Garton, M.J., Robertson, E.M., Gilbert, F.J., Gomersall, L., and Reid, D.M. 1994. Can radiologists detect osteopenia on plain radiographs? *Clin Radiol.* 49(2):118-22.

(12) Saville, P.D. 1967. A quantitative approach to simple radiographic diagnosis of osteoporosis: its application to the osteoporosis of rheumatoid arthritis. *Arthritis Rheum*. 10:416-422.

(13) Werning, J.W., Downey, N.M., Brinker, R.A., Khuder, S.A., Davis, W.J. *et al.* 2004. The impact of osteoporosis on patients with maxillofacial trauma. *Arch Otolaryngol Head Neck Surg*. 130(3):353-6.

(14) McCullagh, C.D., McCoy, K., Crawford, V.L., and Taggart, H. 2003. How reliable is a radiological report in osteoporosis in diagnosing low bone density? *Ulster Med J.* 72(1):34-7.

(15) Ahmed, A.I.H., Ilic, D., Blake, G.M., Rymer, J.M., and Fogelman, I. 1998. Review of 3530 referrals for bone density measurements of spine and femur: evidence that radiographic osteopenia predicts low bone mass. *Radiology*. 207:619-624.

(16) Williamson, M.R., Boyd, C.M., and Williamson, S.L. 1990. Osteoporosis: diagnosis by plain chest film versus dual photon bone densitometry. *Skeletal Radiol.* 19(1):27-30.

(17) Metz. C.E. 1978. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*. 8:283-298.

(18) Zweig, M.H., and Campbell, G. 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*. 39:561-577.

(19) Griner, P.F., Mayewski, R.J., Mushlin, A.I., and Greenland, P. 1981. Selection and interpretation of diagnostic tests and procedures. *Annals of Internal Medicine*. 94:555-600.

(20) Swets, J.A., Pickett, R.M., Whitehead, S.F., Getty, D.J., Schnur, J.A., *et al*. 1979. Assessment of diagnostic technologies. *Science*. 205(4408):753-759.

(21) Kramer, M.S., and Feinstein, A.R. 1981. Clinical biostatistics: LIV. The biostatistics of concordance. *Clinical and Pharmacological Therapy*. 29:111-123.

(22) Cohen, J. 1968. Wheighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull.* 70(4): 213-220.

(23) Fleiss, J.L., Cohen, J., and Everitt, B.S. 1969. Large sample standard errors of kappa and weighted kappa. *Psychol Bull.* 72(5):323-327.

(24) Cicchetti, D.V., Sharma Y., and Cotlier, E. 1982. Assessment of observer variability in the classification of human cataracts. *Yale Journal of Biology and Medicine.* 55:81-88.

(25) Altman, D.G. 1991. *Practical Statistics for Medical Research*. London: Chapman & Hall.

(26) Landis, J.R., and Koch, G. 1977. The measurement of observer agreement for categorical data. *Biometrics*. 33:159-174.

(27) Cicchetti, D.V., and Fleiss, J.L. 1977. Comparison of the null distributions of weighted kappa and the C ordinal statistic. *Applied Psychological Measurement*. 1:195-201.

(28) Fleiss, J.L., and Cicchetti, D.V. 1978. Inference about weighted kappa in the non-null case. *Applied Psychological Measurement*. 2(1):113-117.

(29) Rigby, A.S. 2000. Statistical methods in epidemiology. V. Towards an understanding of the kappa coefficient. *Disability and Rehabilitation.* 22(8):244–339.

(30) Sink, V.J., and Curtis, A.M. 1989. Intra- and interobserver variability in the radiological evaluation of pulmonary vascularity in patients at risk for congestive heart failure. Thesis (M.D.) - Yale University.

(31) Jekel, J.F., Katz, D.L., and Elmore, J.G. 2001. Epidemiology, Biostatistics, and Preventive Medicine. Philadelphia, Pennsylvania: W.B. Saunders Company. 417 pp.

(32) Greenber, R.A., and Jekel, J.F. 1969. Some problems in the determination of the false positive and false negative rates of tuberculin tests. *American Review of Respiratory Disease.* 100:645-650.

(33) Reinbold, W.D., Genant, H.K., Reiser, U. J., Harris, S.T., and Ettinger, B. 1986. Bone mineral content in early-postmenopausal osteoporotic women: comparison of measurement methods. *Radiology.* 160:469

(34) Arlot, M.E., Sornay-Rendu, E., Garnero, P., Vey-Marty, B., Delmas, P.D. 1997. Apparent pre- and postmenopausal bone loss evaluated by DXA at different sites in women: the OFELY cohort. *J Bone Miner Res.* 12:683-690.

(35) New York State Department of Health. Clinical practice guideline: report of the recommendations, communication Disorders, assessment and intervention for young children (age 0-3 years) (Appendix A). *Info for Providers.* Accessed March 22, 2009 at

http://www.health.state.ny.us/community/infants_children/early_intervention/disorders/appendix_a.htm.

(36) Zaman, Z., Borremans, A., Verhaegen, J., *et al*. 1998. Disappointing dipstick screening for urinary tract infection in hospital inpatients. *J Clin Pathol*. 51:471–2.

(37) Cummings, S.R., Bates, D., and Black, D.M. 2002. Clinical use of bone densitometry: scientific review. *JAMA.* 288:1889-1897.

(38) Ghelbach, S.H., Bigelow, C., Heimisdottir, M., May, S., Walker, M., *et al*. 2000. Recognition of vertebral fracture in a clinical setting. *Osteoporos Int.* 11(7):577–582.

(39) Neuner, J.M., Zimmer, J.K., and Hamel, M.B. 2003. Diagnosis and treatment of osteoporosis in patients with vertebral compression fractures. *J Am Geriatr Soc*. 51:483–491.

(40) McKinlay, S., Brambilla, D., and Posner, J. 1992. The normal menopausal transition. *Maturitas.* 14:103–115.

(41) Doube, A. 1999. Managing osteoporosis in older people with fractures: needs to be taken as seriously as coronary artery disease. *BMJ*. 317:477–478.

(42) Eddy, D.E., Johnston, C.C., Cummings, S.R. *et al*. 1998. Osteoporosis: review of the evidence for prevention, diagnosis and treatment and cost-effectiveness analysis. *Osteoporos Int.* 8(Suppl. 4):S1–S88.

(43) Rubin, S.M., and Cummings, S.R. 1992. Results of bone densitometry: effect on women's decisions about taking measures to prevent fractures. *Ann Intern Med.* 116:990–995.

(44) Fitt, N.S., Mitchell, S.L., Cranney, A. *et al*. 2001. Influence of bone densitometry results on the treatment of osteoporosis. *Can Med Assoc J.* 164:777–781.

(45) Akobeng, A.K. 2007. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatr.* 96(3):338-41.

(46) Ettinger, B., Black, D.M., Mitlak, B.H. *et al*. 1999. Reduction of vertebral fracture risk in postmenopausal women with osteoporosis treated with raloxifene: Results from a 3-year randomized clinical trial. *JAMA.* 282:637–645.

(47) Liberman, U., Weiss, S., Broll, J. *et al*. 1995. Effects of oral alendronate on bone mineral density and the incidence of fractures in postmenopausal osteoporosis. *N Engl J Med.* 333:1437–1444.

(48) Black, D.M., Cummings, S.R., Karpf, D.B. *et al*. 1996. Randomized trial of effect of alendronate on risk of fracture in women with existing vertebral fractures. *Lancet.* 348:1535–1541.

(49) Kamel, H.K., Hussain, M.S., Tariq, S. *et al*. 2000. Failure to diagnose and treat osteoporosis in elderly patients hospitalized with hip fracture. *Am J Med*. 109:326–328.

(50) Khan, S., De Geus, C., Holroyd, B. *et al*. 2001. Osteoporosis follow-up after wrist fractures following minor trauma. *Arch Intern Med*. 161:1309–1312.

(51) Bellantonio, S., and Fortinsky, R. 2001. How well are community-living women treated for osteoporosis after hip fracture? *J Am Geriatr Soc.* 49:1197–1204.

(52) Freedman, K., Kaplan, F., Bilker, W. *et al*. 2000. Treatment of osteoporosis: Are physicians missing an opportunity? *J Bone Joint Surg Am*. 82:1063–1070.

# APPENDIX A

Reader ID _______ Patient ID _______ DATE _______

Please Circle one of the following: First Read (Creation Date Triangle UP)
Second Read (Creation Date Triangle DOWN)

END PLATES: DEFINITION compared to vertebral body

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Definitely Normal | Probably Normal | Uncertain | Probably too Prominent | Definitely too Prominent |

END PLATES: SHAPE (FISH MOUTH)

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Definitely Normal | Probably Normal | Uncertain | Probably Concave | Definitely Concave |

BODY SHAPE (wedged)

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Definitely Normal | Probably Normal | Uncertain | Probably Wedged | Definitely Wedged |

COMPRESSION FRACTURES (decreased height anterior and posterior)

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Definitely Absent | Probably Absent | Uncertain | Probably Present | Definitely Present |

VERTICAL STRIATIONS

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Definitely Absent | Probably Absent | Uncertain | Probably Prominent | Definitely Prominent |

EMPTY BOX

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Definitely Absent | Probably Absent | Uncertain | Probably Present | Definitely Present |

OSTEOPENIA (OSTEOPOROSIS)

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Definitely Absent | Probably Absent | Uncertain | Probably Present | Definitely Present |

DO YOU ROUTINELY COMMENT ON SPINE ON CXR FOR OSTEOPENIA?

Never Almost Never Sometimes Almost Always

# APPENDIX B

INTRAOBSERVER AGREEMENT: FIRST CHEST RADIOLOGIST

**Intraobserver Agreement: First Radiologist, End Plates Definition (Compared to Vertebral Body), N=93**

```
Fleiss' Kappa Statistics
             Response      Kappa  SE Kappa         Z  P(vs > 0)
             1          0.258225  0.103695   2.49023     0.0064
             2         -0.148148  0.103695  -1.42869     0.9235
             3         -0.033333  0.103695  -0.32146     0.6261
             4          0.424149  0.103695   4.09034     0.0000
             5          0.489011  0.103695   4.71585     0.0000
             Overall    0.212198  0.069248   3.06432     0.0011

Cohen's Kappa Statistics
             Response      Kappa  SE Kappa         Z  P(vs > 0)
             1          0.275912  0.098354   2.80529     0.0025
             2         -0.148148  0.103695  -1.42869     0.9235
             3          0.000000  0.000000         *          *
             4          0.428132  0.101283   4.22710     0.0000
             5          0.491803  0.089307   5.50691     0.0000
             Overall    0.223382  0.066364   3.36600     0.0004

Kendall's Coefficient of Concordance
             Coef         Chi - Sq      DF     P
             0.712959     131.185       92     0.0046
```

| | First Read | | | | |
|---|---|---|---|---|---|
| **Second Read** | **1** | **2** | **3** | **4** | **5** |
| **1** | 43 | 10 | 6 | 6 | 1 |
| **2** | 7 | 0 | 0 | 5 | 0 |
| **3** | 0 | 0 | 0 | 0 | 0 |
| **4** | 2 | 2 | 0 | 9 | 1 |
| **5** | 0 | 0 | 0 | 0 | 1 |
| **Weighted Kappa** | 0.47 | | | | |
| **Standard error (Kw'=0)** | 0.097 | | | | |
| **Standard error (Kw'#0)** | 0.097 | | | | |

### Intraobserver Agreement: First Radiologist, End Plates Shape (Fish Vertebra), N=98

```
Fleiss' Kappa Statistics
       Response     Kappa  SE Kappa        Z  P(vs > 0)
           1         0.348693  0.101015  3.45188     0.0003
           2         0.034203  0.101015  0.33859     0.3675
           3         0.218085  0.101015  2.15893     0.0154
           4         0.442286  0.101015  4.37841     0.0000
           5         0.184789  0.101015  1.82932     0.0337
           Overall   0.279359  0.062014  4.50474     0.0000

Cohen's Kappa Statistics
       Response     Kappa  SE Kappa        Z  P(vs > 0)
           1         0.381833  0.090040  4.24073     0.0000
           2         0.068753  0.090799  0.75721     0.2245
           3         0.220159  0.097543  2.25706     0.0120
           4         0.447820  0.098629  4.54044     0.0000
           5         0.185273  0.100329  1.84666     0.0324
           Overall   0.300599  0.056966  5.27680     0.0000

Kendall's Coefficient of Concordance
       Coef  Chi - Sq  DF       P
       0.820927   159.260  97  0.0001
```

| | First Read | | | | |
|---|---|---|---|---|---|
| **Second Read** | **1** | **2** | **3** | **4** | **5** |
| **1** | 22 | 4 | 0 | 0 | 0 |
| **2** | 19 | 6 | 2 | 4 | 0 |
| **3** | 1 | 0 | 1 | 1 | 0 |
| **4** | 5 | 5 | 2 | 18 | 4 |
| **5** | 1 | 0 | 0 | 2 | 1 |
| **Weighted Kappa** | 0.638 | | | | |
| **Standard error (Kw'=0)** | 0.097 | | | | |
| **Standard error (Kw'#0)** | 0.068 | | | | |

## Intraobserver Agreement: First Radiologist, Body Shape (Wedged), N=100

```
Fleiss' Kappa Statistics
      Response      Kappa  SE Kappa         Z  P(vs > 0)
          1          0.532492  0.100000   5.32492     0.0000
          2          0.385561  0.100000   3.85561     0.0001
          3         -0.005025  0.100000  -0.05025     0.5200
          4          0.310470  0.100000   3.10470     0.0010
          5          0.385561  0.100000   3.85561     0.0001
          Overall    0.424874  0.067844   6.26248     0.0000

Cohen's Kappa Statistics
      Response     Kappa  SE Kappa        Z  P(vs > 0)
          1         0.537893  0.097294  5.52852     0.0000
          2         0.385561  0.100000  3.85561     0.0001
          3         0.000000  0.000000        *           *
          4         0.330986  0.088908  3.72278     0.0001
          5         0.389313  0.095224  4.08838     0.0000
          Overall   0.433198  0.063873  6.78214     0.0000

Kendall's Coefficient of Concordance
      Coef  Chi - Sq  DF       P
      0.803161   159.026  99  0.0001
```

| | First Read | | | | |
|---|---|---|---|---|---|
| **Second Read** | **1** | **2** | **3** | **4** | **5** |
| **1** | 59 | 3 | 1 | 11 | 0 |
| **2** | 4 | 3 | 0 | 0 | 0 |
| **3** | 0 | 0 | 0 | 0 | 0 |
| **4** | 0 | 1 | 0 | 7 | 2 |
| **5** | 1 | 0 | 0 | 5 | 3 |
| **Weighted Kappa** | 0.654 | | | | |
| **Standard error (Kw'=0)** | 0.098 | | | | |
| **Standard error (Kw'#0)** | 0.082 | | | | |

### Intraobserver Agreement: First Radiologist, Compression Fractures, N=99

```
Fleiss' Kappa Statistics
      Response       Kappa   SE Kappa          Z  P(vs > 0)
          1        -0.0206186  0.100504  -0.205152     0.5813
          3        -0.0050761  0.100504  -0.050507     0.5201
          4        -0.0102041  0.100504  -0.101529     0.5404
          5        -0.0050761  0.100504  -0.050507     0.5201
          Overall  -0.0140845  0.072941  -0.193095     0.5766

Cohen's Kappa Statistics
      Response       Kappa   SE Kappa          Z  P(vs > 0)
          1        -0.0153846  0.086588  -0.177676     0.5705
          3         0.0000000  0.000000          *          *
          4        -0.0102041  0.100504  -0.101529     0.5404
          5         0.0000000  0.000000          *          *
          Overall  -0.0102041  0.061020  -0.167225     0.5664

Kendall's Coefficient of Concordance
      Coef  Chi - Sq  DF       P
      0.492230   96.4770  98  0.5246
```

| | First Read | | | |
|---|---|---|---|---|
| **Second Read** | **1** | **3** | **4** | **5** |
| **1** | 95 | 1 | 1 | 1 |
| **3** | 0 | 0 | 0 | 0 |
| **4** | 1 | 0 | 0 | 0 |
| **5** | 0 | 0 | 0 | 0 |
| **Weighted Kappa** | -0.014 | | | |
| **Standard error (Kw'=0)** | 0.083 | | | |
| **Standard error (Kw'#0)** | 0.011 | | | |

## Intraobserver Agreement: First Radiologist, Vertical Striations, N=100

```
Fleiss' Kappa Statistics
      Response      Kappa  SE Kappa         Z  P(vs > 0)
           1          0.454365  0.100000   4.54365     0.0000
           2          0.184783  0.100000   1.84783     0.0323
           3         -0.020408  0.100000  -0.20408     0.5809
           4          0.252492  0.100000   2.52492     0.0058
           5         -0.041667  0.100000  -0.41667     0.6615
           Overall    0.294533  0.066796   4.40943     0.0000

Cohen's Kappa Statistics
      Response      Kappa   SE Kappa         Z  P(vs > 0)
           1          0.455446  0.0995086   4.57694     0.0000
           2          0.189189  0.0962784   1.96502     0.0247
           3         -0.015228  0.0861584  -0.17675     0.5701
           4          0.253112  0.0996551   2.53988     0.0055
           5         -0.030928  0.0856911  -0.36092     0.6409
           Overall    0.296703  0.0657789   4.51061     0.0000

Kendall's Coefficient of Concordance
      Coef  Chi - Sq  DF       P
      0.761598   150.796  99  0.0006
```

| | First Read | | | | |
|---|---|---|---|---|---|
| **Second Read** | **1** | **2** | **3** | **4** | **5** |
| **1** | 61 | 3 | 2 | 4 | 0 |
| **2** | 7 | 2 | 0 | 1 | 0 |
| **3** | 0 | 1 | 0 | 0 | 0 |
| **4** | 5 | 0 | 1 | 5 | 2 |
| **5** | 1 | 0 | 0 | 5 | 0 |
| **Weighted Kappa** | 0.593 | | | | |
| **Standard error (Kw'=0)** | 0.099 | | | | |
| **Standard error (Kw'#0)** | 0.094 | | | | |

## Intraobserver Agreement: First Radiologist, Empty Box, N=100

```
Fleiss' Kappa Statistics
       Response      Kappa  SE Kappa         Z  P(vs > 0)
            1         0.204244  0.100000   2.04244     0.0206
            2         0.218750  0.100000   2.18750     0.0144
            3        -0.015228  0.100000  -0.15228     0.5605
            4         0.290780  0.100000   2.90780     0.0018
            5        -0.015228  0.100000  -0.15228     0.5605
            Overall   0.199832  0.068496   2.91742     0.0018

Cohen's Kappa Statistics
       Response      Kappa   SE Kappa         Z  P(vs > 0)
            1         0.210526  0.0964753   2.18218     0.0145
            2         0.220779  0.0965679   2.28626     0.0111
            3        -0.013514  0.0941204  -0.14358     0.5571
            4         0.295775  0.0935958   3.16013     0.0008
            5        -0.013514  0.0941204  -0.14358     0.5571
            Overall   0.204688  0.0654082   3.12940     0.0009

Kendall's Coefficient of Concordance
       Coef  Chi - Sq  DF       P
       0.605198   119.829  99  0.0758
```

| | First Read | | | | |
|---|---|---|---|---|---|
| **Second Read** | **1** | **2** | **3** | **4** | **5** |
| **1** | 78 | 1 | 1 | 2 | 2 |
| **2** | 4 | 1 | 0 | 0 | 0 |
| **3** | 2 | 0 | 0 | 0 | 0 |
| **4** | 5 | 1 | 0 | 2 | 0 |
| **5** | 1 | 0 | 0 | 0 | 0 |
| **Weighted Kappa** | 0.162 | | | | |
| **Standard error (Kw'=0)** | 0.098 | | | | |
| **Standard error (Kw'#0)** | 0.132 | | | | |

### Intraobserver Agreement: First Radiologist, Presence of Osteopenia, N=100

```
Fleiss' Kappa Statistics
       Response     Kappa  SE Kappa        Z  P(vs > 0)
            1        0.649626  0.100000  6.49626     0.0000
            2        0.345489  0.100000  3.45489     0.0003
            3        0.023199  0.100000  0.23199     0.4083
            4        0.255952  0.100000  2.55952     0.0052
            5        0.405714  0.100000  4.05714     0.0000
            Overall  0.358429  0.053728  6.67120     0.0000

Cohen's Kappa Statistics
       Response     Kappa   SE Kappa        Z  P(vs > 0)
            1        0.649682  0.0999493  6.50011     0.0000
            2        0.348231  0.0990199  3.51678     0.0002
            3        0.024390  0.0992536  0.24574     0.4029
            4        0.261811  0.0980434  2.67036     0.0038
            5        0.406934  0.0990590  4.10800     0.0000
            Overall  0.360814  0.0531051  6.79433     0.0000

Kendall's Coefficient of Concordance
       Coef   Chi - Sq  DF       P
       0.865317   171.333  99  0.0000
```

| | First Read | | | | |
|---|---|---|---|---|---|
| **Second Read** | **1** | **2** | **3** | **4** | **5** |
| **1** | 14 | 4 | 1 | 1 | 0 |
| **2** | 5 | 17 | 5 | 7 | 0 |
| **3** | 0 | 2 | 1 | 5 | 0 |
| **4** | 0 | 4 | 2 | 13 | 5 |
| **5** | 0 | 1 | 1 | 6 | 6 |
| **Weighted Kappa** | 0.718 | | | | |
| **Standard error (Kw'=0)** | 0.1 | | | | |
| **Standard error (Kw'#0)** | 0.052 | | | | |

INTRAOBSERVER AGREEMENT: SECOND CHEST RADIOLOGIST

## Intraobserver Agreement: Second Radiologist, End Plates Definition (Compared to Vertebral Body), N=100

```
Fleiss' Kappa Statistics
      Response     Kappa  SE Kappa        Z  P(vs > 0)
          1         0.164256  0.100000  1.64256     0.0502
          2         0.064984  0.100000  0.64984     0.2579
          3         0.096556  0.100000  0.96556     0.1671
          4         0.269831  0.100000  2.69831     0.0035
          5         0.312715  0.100000  3.12715     0.0009
          Overall   0.157284  0.059434  2.64636     0.0041

Cohen's Kappa Statistics
      Response     Kappa   SE Kappa        Z  P(vs > 0)
          1         0.195545  0.0732569  2.66930     0.0038
          2         0.065421  0.0998908  0.65492     0.2563
          3         0.114555  0.0943265  1.21445     0.1123
          4         0.273768  0.0988063  2.77075     0.0028
          5         0.322034  0.0735093  4.38086     0.0000
          Overall   0.166894  0.0571811  2.91870     0.0018

Kendall's Coefficient of Concordance
      Coef  Chi - Sq  DF       P
       0.773715   153.195  99  0.0004
```

| | First Read | | | | |
|---|---|---|---|---|---|
| **Second Read** | **1** | **2** | **3** | **4** | **5** |
| **1** | 2 | 1 | 0 | 0 | 0 |
| **2** | 11 | 11 | 7 | 3 | 0 |
| **3** | 1 | 8 | 7 | 13 | 0 |
| **4** | 0 | 10 | 3 | 18 | 0 |
| **5** | 0 | 0 | 0 | 4 | 1 |
| **Weighted Kappa** | 0.54 | | | | |
| **Standard error (Kw'=0)** | 0.097 | | | | |
| **Standard error (Kw'#0)** | 0.067 | | | | |

### Intraobserver Agreement: Second Radiologist, End Plates Shape (Fish Vertebra), N=99

```
Fleiss' Kappa Statistics
      Response     Kappa  SE Kappa        Z  P(vs > 0)
          1         0.084874  0.100504  0.84449     0.1992
          2         0.139130  0.100504  1.38433     0.0831
          3         0.130173  0.100504  1.29521     0.0976
          4         0.515583  0.100504  5.12999     0.0000
          5         0.312500  0.100504  3.10934     0.0009
          Overall   0.262411  0.059599  4.40293     0.0000

Cohen's Kappa Statistics
      Response     Kappa  SE Kappa        Z  P(vs > 0)
          1         0.111746  0.087920  1.27100     0.1019
          2         0.153846  0.096423  1.59553     0.0553
          3         0.139918  0.095325  1.46780     0.0711
          4         0.516047  0.100303  5.14486     0.0000
          5         0.321918  0.073869  4.35796     0.0000
          Overall   0.271853  0.057048  4.76534     0.0000

Kendall's Coefficient of Concordance
      Coef  Chi - Sq  DF       P
      0.864783   169.497  98  0.0000
```

| | First Read | | | | |
|---|---|---|---|---|---|
| **Second Read** | **1** | **2** | **3** | **4** | **5** |
| **1** | 3 | 4 | 1 | 0 | 0 |
| **2** | 17 | 12 | 5 | 2 | 0 |
| **3** | 0 | 2 | 3 | 4 | 0 |
| **4** | 0 | 6 | 7 | 28 | 0 |
| **5** | 0 | 0 | 0 | 4 | 1 |
| **Weighted Kappa** | 0.712 | | | | |
| **Standard error (Kw'=0)** | 0.098 | | | | |
| **Standard error (Kw'#0)** | 0.044 | | | | |

### Intraobserver Agreement: Second Radiologist, Body Shape (Wedged), N=100

```
Fleiss' Kappa Statistics
      Response      Kappa  SE Kappa         Z  P(vs > 0)
          1          0.219922  0.100000   2.19922     0.0139
          2         -0.041667  0.100000  -0.41667     0.6615
          3          0.078341  0.100000   0.78341     0.2167
          4          0.369922  0.100000   3.69922     0.0001
          5          0.456522  0.100000   4.56522     0.0000
          Overall    0.213475  0.057930   3.68503     0.0001

Cohen's Kappa Statistics
      Response     Kappa  SE Kappa        Z  P(vs > 0)
          1         0.313863  0.075912  4.13457     0.0000
          2         0.005682  0.082290  0.06905     0.4725
          3         0.100450  0.080022  1.25527     0.1047
          4         0.375000  0.097313  3.85355     0.0001
          5         0.463087  0.091534  5.05919     0.0000
          Overall   0.262044  0.047932  5.46698     0.0000

Kendall's Coefficient of Concordance
      Coef  Chi - Sq  DF       P
      0.793838   157.180  99  0.0002
```

| | First Read | | | | |
|---|---|---|---|---|---|
| **Second Read** | **1** | **2** | **3** | **4** | **5** |
| **1** | 31 | 0 | 0 | 1 | 0 |
| **2** | 20 | 2 | 2 | 0 | 0 |
| **3** | 6 | 3 | 1 | 0 | 1 |
| **4** | 10 | 3 | 0 | 9 | 0 |
| **5** | 2 | 0 | 0 | 5 | 4 |
| **Weighted Kappa** | 0.533 | | | | |
| **Standard error (Kw'=0)** | 0.086 | | | | |
| **Standard error (Kw'#0)** | 0.077 | | | | |

**Intraobserver Agreement: Second Radiologist, Compression Fractures, N=99**

```
Fleiss' Kappa Statistics
      Response      Kappa  SE Kappa        Z  P(vs > 0)
          1         -0.086453  0.100504  -0.86020     0.8052
          2          0.022222  0.100504   0.22111     0.4125
          3          0.010291  0.100504   0.10240     0.4592
          4          0.221348  0.100504   2.20239     0.0138
          5         -0.010204  0.100504  -0.10153     0.5404
          Overall    0.009684  0.064399   0.15037     0.4402

Cohen's Kappa Statistics
      Response     Kappa   SE Kappa        Z  P(vs > 0)
          1         0.098361  0.0573090  1.71632     0.0431
          2         0.077998  0.0593867  1.31339     0.0945
          3         0.087805  0.0411820  2.13212     0.0165
          4         0.235099  0.0903951  2.60080     0.0047
          5         0.000000  0.0000000        *          *
          Overall   0.114894  0.0393687  2.91842     0.0018

Kendall's Coefficient of Concordance
      Coef  Chi - Sq  DF       P
      0.597192   117.050  98  0.0920
```

| | First Read | | | | |
|---|---|---|---|---|---|
| **Second Read** | **1** | **2** | **3** | **4** | **5** |
| **1** | 47 | 0 | 0 | 2 | 0 |
| **2** | 15 | 1 | 0 | 0 | 0 |
| **3** | 16 | 1 | 1 | 0 | 0 |
| **4** | 11 | 0 | 0 | 3 | 0 |
| **5** | 1 | 0 | 0 | 1 | 0 |
| **Weighted Kappa** | 0.179 | | | | |
| **Standard error (Kw'=0)** | 0.068 | | | | |
| **Standard error (Kw'#0)** | 0.095 | | | | |

## Intraobserver Agreement: Second Radiologist, Vertical Striations, N=100

```
Fleiss' Kappa Statistics
       Response      Kappa  SE Kappa         Z  P(vs > 0)
            1          0.205236  0.100000   2.05236     0.0201
            2         -0.123596  0.100000  -1.23596     0.8918
            3          0.169435  0.100000   1.69435     0.0451
            4          0.326599  0.100000   3.26599     0.0005
            5         -0.005025  0.100000  -0.05025     0.5200
            Overall    0.141368  0.066659   2.12076     0.0170

Cohen's Kappa Statistics
       Response      Kappa   SE Kappa          Z  P(vs > 0)
            1          0.280880  0.0772908   3.63406     0.0001
            2         -0.083744  0.0806574  -1.03827     0.8504
            3          0.193548  0.0875139   2.21163     0.0135
            4          0.342105  0.0753110   4.54257     0.0000
            5          0.000000  0.0000000         *          *
            Overall    0.192463  0.0535908   3.59134     0.0002

Kendall's Coefficient of Concordance
       Coef  Chi - Sq  DF       P
       0.713955   141.363  99  0.0034
```

| | First Read | | | | |
|---|---|---|---|---|---|
| **Second Read** | **1** | **2** | **3** | **4** | **5** |
| **1** | 52 | 1 | 1 | 0 | 0 |
| **2** | 15 | 0 | 1 | 0 | 1 |
| **3** | 14 | 2 | 4 | 0 | 0 |
| **4** | 3 | 2 | 2 | 2 | 0 |
| **5** | 0 | 0 | 0 | 0 | 0 |
| **Weighted Kappa** | 0.365 | | | | |
| **Standard error (Kw'=0)** | 0.082 | | | | |
| **Standard error (Kw'#0)** | 0.088 | | | | |

### Intraobserver Agreement: Second Radiologist, Empty Box, N=79

```
Fleiss' Kappa Statistics
      Response      Kappa  SE Kappa        Z  P(vs > 0)
           1         0.415995  0.112509   3.69744     0.0001
           2         0.165493  0.112509   1.47093     0.0707
           3         0.122222  0.112509   1.08633     0.1387
           4         0.192180  0.112509   1.70814     0.0438
           5        -0.019355  0.112509  -0.17203     0.5683
           Overall   0.260768  0.069884   3.73143     0.0001

Cohen's Kappa Statistics
      Response     Kappa  SE Kappa        Z  P(vs > 0)
           1        0.459913  0.094688  4.85712     0.0000
           2        0.188356  0.094615  1.99075     0.0233
           3        0.123613  0.111624  1.10741     0.1341
           4        0.231254  0.091015  2.54083     0.0055
           5        0.000000  0.000000        *          *
           Overall  0.296122  0.059405  4.98483     0.0000

Kendall's Coefficient of Concordance
      Coef  Chi - Sq  DF       P
      0.837777   130.693  78  0.0002
```

| | First Read | | | | |
|---|---|---|---|---|---|
| **Second Read** | **1** | **2** | **3** | **4** | **5** |
| **1** | 37 | 0 | 0 | 0 | 0 |
| **2** | 10 | 2 | 0 | 0 | 0 |
| **3** | 6 | 0 | 2 | 0 | 0 |
| **4** | 6 | 2 | 7 | 4 | 0 |
| **5** | 0 | 0 | 1 | 2 | 0 |
| **Weighted Kappa** | 0.577 | | | | |
| **Standard error (Kw'=0)** | 0.091 | | | | |
| **Standard error (Kw'#0)** | 0.078 | | | | |

## Intraobserver Agreement: Second Radiologist, Presence of Osteopenia, N=100

```
Fleiss' Kappa Statistics
      Response     Kappa  SE Kappa        Z  P(vs > 0)
           1         0.476592  0.100000  4.76592     0.0000
           2         0.214454  0.100000  2.14454     0.0160
           3         0.322417  0.100000  3.22417     0.0006
           4         0.222571  0.100000  2.22571     0.0130
           5         0.424105  0.100000  4.24105     0.0000
           Overall   0.289544  0.056098  5.16142     0.0000

Cohen's Kappa Statistics
      Response     Kappa   SE Kappa        Z  P(vs > 0)
           1         0.477958  0.0984743  4.85363     0.0000
           2         0.218750  0.0987718  2.21470     0.0134
           3         0.325843  0.0985695  3.30572     0.0005
           4         0.225000  0.0992157  2.26779     0.0117
           5         0.429967  0.0913354  4.70757     0.0000
           Overall   0.292956  0.0552333  5.30397     0.0000

Kendall's Coefficient of Concordance
      Coef  Chi - Sq  DF       P
      0.808476   160.078  99  0.0001
```

| | **First Read** | | | | |
|---|---|---|---|---|---|
| **Second Read** | **1** | **2** | **3** | **4** | **5** |
| **1** | 5 | 6 | 0 | 0 | 0 |
| **2** | 3 | 16 | 9 | 2 | 0 |
| **3** | 0 | 4 | 11 | 5 | 0 |
| **4** | 0 | 11 | 6 | 12 | 1 |
| **5** | 0 | 0 | 0 | 6 | 3 |
| **Weighted Kappa** | 0.63 | | | | |
| **Standard error (Kw'=0)** | 0.098 | | | | |
| **Standard error (Kw'#0)** | 0.062 | | | | |

INTEROBSERVER AGREEMENT: CHEST RADIOLOGISTS

**Interobserver Agreement for Chest Radiologists: End Plates Definition (Compared to Vertebral Body), First Trial, N=96**

```
Fleiss' Kappa Statistics
Response       Kappa   SE Kappa          Z  P(vs > 0)
1          0.0142090  0.102062   0.139219     0.4446
2          0.0857143  0.102062   0.839825     0.2005
3         -0.0267380  0.102062  -0.261978     0.6033
4          0.0768031  0.102062   0.752514     0.2259
5         -0.0212766  0.102062  -0.208467     0.5826
Overall    0.0422158  0.060025   0.703300     0.2409

Cohen's Kappa Statistics
Response      Kappa   SE Kappa         Z  P(vs > 0)
1          0.165724  0.0667280   2.48357     0.0065
2          0.113300  0.0929143   1.21941     0.1113
3          0.000000  0.0883883   0.00000     0.5000
4          0.110220  0.0931435   1.18334     0.1183
5         -0.015873  0.0879157  -0.18055     0.5716
Overall    0.111970  0.0477222   2.34629     0.0095

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.600670   114.127  95  0.0882
```

| | **First Chest Radiologist** | | | | |
|---|---|---|---|---|---|
| **Second Chest Radiologist** | **1** | **2** | **3** | **4** | **5** |
| **1** | 12 | 2 | 0 | 0 | 0 |
| **2** | 11 | 6 | 4 | 7 | 0 |
| **3** | 12 | 1 | 1 | 2 | 0 |
| **4** | 18 | 5 | 1 | 10 | 2 |
| **5** | 0 | 0 | 0 | 1 | 0 |
| **Weighted Kappa** | 0.175 | | | | |
| **Standard error (Kw'=0)** | 0.083 | | | | |
| **Standard error (Kw'#0)** | 0.079 | | | | |

## Interobserver Agreement for Chest Radiologists: End Plates Shape (Fish Vertebra), First Trial, N=99

Fleiss' Kappa Statistics

| Response | Kappa | SE Kappa | Z | P(vs > 0) |
|---|---|---|---|---|
| 1 | 0.059276 | 0.100504 | 0.58979 | 0.2777 |
| 2 | 0.060127 | 0.100504 | 0.59825 | 0.2748 |
| 3 | 0.094431 | 0.100504 | 0.93958 | 0.1737 |
| 4 | 0.371429 | 0.100504 | 3.69567 | 0.0001 |
| 5 | 0.312500 | 0.100504 | 3.10934 | 0.0009 |
| Overall | 0.167309 | 0.058888 | 2.84112 | 0.0022 |

Cohen's Kappa Statistics

| Response | Kappa | SE Kappa | Z | P(vs > 0) |
|---|---|---|---|---|
| 1 | 0.135910 | 0.0821525 | 1.65436 | 0.0490 |
| 2 | 0.069549 | 0.0973610 | 0.71434 | 0.2375 |
| 3 | 0.122981 | 0.0827533 | 1.48611 | 0.0686 |
| 4 | 0.383675 | 0.0959766 | 3.99759 | 0.0000 |
| 5 | 0.321918 | 0.0738689 | 4.35796 | 0.0000 |
| Overall | 0.199677 | 0.0522675 | 3.82028 | 0.0001 |

Kendall's Coefficient of Concordance

| Coef | Chi - Sq | DF | P |
|---|---|---|---|
| 0.724328 | 141.968 | 98 | 0.0025 |

| | **First Chest Radiologist** | | | | |
|---|---|---|---|---|---|
| **Second Chest Radiologist** | **1** | **2** | **3** | **4** | **5** |
| **1** | 13 | 7 | 0 | 0 | 0 |
| **2** | 15 | 5 | 0 | 4 | 0 |
| **3** | 9 | 1 | 2 | 3 | 1 |
| **4** | 11 | 3 | 3 | 18 | 3 |
| **5** | 0 | 0 | 0 | 0 | 1 |
| **Weighted Kappa** | 0.45 | | | | |
| **Standard error (Kw'=0)** | 0.091 | | | | |
| **Standard error (Kw'#0)** | 0.074 | | | | |

**Interobserver Agreement for Chest Radiologists: Body Shape (Wedged), First Read, N=100**

```
Fleiss' Kappa Statistics
Response     Kappa  SE Kappa        Z  P(vs > 0)
1         0.483784  0.100000  4.83784     0.0000
2         0.063063  0.100000  0.63063     0.2641
3         0.489796  0.100000  4.89796     0.0000
4         0.480182  0.100000  4.80182     0.0000
5         0.368421  0.100000  3.68421     0.0001
Overall   0.415376  0.067755  6.13052     0.0000

Cohen's Kappa Statistics
Response     Kappa  SE Kappa        Z  P(vs > 0)
1         0.485228  0.099372  4.88295     0.0000
2         0.063401  0.099740  0.63566     0.2625
3         0.492386  0.086158  5.71489     0.0000
4         0.485531  0.096635  5.02438     0.0000
5         0.368421  0.100000  3.68421     0.0001
Overall   0.418041  0.066552  6.28143     0.0000

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.780020   154.444  99  0.0003
```

| | **First Chest Radiologist** | | | | |
|---|---|---|---|---|---|
| **Second Chest Radiologist** | **1** | **2** | **3** | **4** | **5** |
| **1** | 55 | 5 | 0 | 7 | 2 |
| **2** | 6 | 1 | 0 | 1 | 0 |
| **3** | 0 | 1 | 1 | 1 | 0 |
| **4** | 3 | 0 | 0 | 11 | 1 |
| **5** | 0 | 0 | 0 | 3 | 2 |
| **Weighted Kappa** | 0.622 | | | | |
| **Standard error (Kw'=0)** | 0.099 | | | | |
| **Standard error (Kw'#0)** | 0.088 | | | | |

**Interobserver Agreement for Chest Radiologists: Compression Fractures, First Trial, N=98**

```
Fleiss' Kappa Statistics
Response     Kappa  SE Kappa         Z  P(vs > 0)
1          0.28986  0.101015   2.86942     0.0021
2         -0.01031  0.101015  -0.10206     0.5406
3          1.00000  0.101015   9.89949     0.0000
4         -0.03704  0.101015  -0.36665     0.6431
5         -0.00513  0.101015  -0.05077     0.5202
Overall    0.21635  0.074055   2.92146     0.0017

Cohen's Kappa Statistics
Response     Kappa  SE Kappa         Z  P(vs > 0)
1          0.30125  0.086033   3.50156     0.0002
2          0.00000  0.000000         *          *
3          1.00000  0.101015   9.89949     0.0000
4         -0.01780  0.069361  -0.25669     0.6013
5          0.00000  0.000000         *          *
Overall    0.22767  0.056612   4.02157     0.0000

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.645192   125.167  97  0.0286
```

| | First Chest Radiologist | | | | |
|---|---|---|---|---|---|
| **Second Chest Radiologist** | **1** | **2** | **3** | **4** | **5** |
| **1** | 88 | 0 | 0 | 1 | 0 |
| **2** | 1 | 0 | 0 | 0 | 1 |
| **3** | 0 | 0 | 1 | 0 | 0 |
| **4** | 6 | 0 | 0 | 0 | 0 |
| **5** | 0 | 0 | 0 | 0 | 0 |
| **Weighted Kappa** | 0.141 | | | | |
| **Standard error (Kw'=0)** | 0.094 | | | | |
| **Standard error (Kw'#0)** | 0.1 | | | | |

**Interobserver Agreement for Chest Radiologists: Vertical Striations, First Read, N=100**

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1          0.216395  0.100000   2.16395     0.0152
2          0.134199  0.100000   1.34199     0.0898
3          0.134199  0.100000   1.34199     0.0898
4          0.164256  0.100000   1.64256     0.0502
5         -0.015228  0.100000  -0.15228     0.5605
Overall    0.172185  0.065389   2.63323     0.0042

Cohen's Kappa Statistics
Response      Kappa   SE Kappa         Z  P(vs > 0)
1          0.228029  0.0954905   2.38797     0.0085
2          0.134615  0.0995366   1.35242     0.0881
3          0.144487  0.0879832   1.64221     0.0503
4          0.207317  0.0609634   3.40068     0.0003
5         -0.013514  0.0941204  -0.14358     0.5571
Overall    0.188751  0.0567965   3.32328     0.0004

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.632159   125.167  99  0.0389
```

| | First Chest Radiologist | | | | |
|---|---|---|---|---|---|
| **Second Chest Radiologist** | **1** | **2** | **3** | **4** | **5** |
| **1** | 66 | 5 | 2 | 10 | 1 |
| **2** | 3 | 1 | 0 | 1 | 0 |
| **3** | 5 | 0 | 1 | 1 | 1 |
| **4** | 0 | 0 | 0 | 2 | 0 |
| **5** | 0 | 0 | 0 | 1 | 0 |
| **Weighted Kappa** | 0.295 | | | | |
| **Standard error (Kw'=0)** | 0.087 | | | | |
| **Standard error (Kw'#0)** | 0.115 | | | | |

### Interobserver Agreement for Chest Radiologists: Empty Box, First Read, N=79

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1          0.151258  0.112509   1.34441     0.0894
2         -0.039474  0.112509  -0.35085     0.6371
3          0.120594  0.112509   1.07186     0.1419
4          0.410887  0.112509   3.65205     0.0001
5         -0.006369  0.112509  -0.05661     0.5226
Overall    0.164463  0.075869   2.16772     0.0151

Cohen's Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1          0.189962  0.093648   2.02848     0.0213
2         -0.034934  0.105602  -0.33081     0.6296
3          0.162544  0.061487   2.64357     0.0041
4          0.414815  0.105346   3.93765     0.0000
5          0.000000  0.000000         *          *
Overall    0.192642  0.060331   3.19307     0.0007

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.631779   98.5575  78  0.0579
```

| | **First Chest Radiologist** | | | | |
|---|---|---|---|---|---|
| **Second Chest Radiologist** | **1** | **2** | **3** | **4** | **5** |
| **1** | 56 | 2 | 0 | 1 | 0 |
| **2** | 4 | 0 | 0 | 0 | 0 |
| **3** | 9 | 0 | 1 | 0 | 0 |
| **4** | 3 | 0 | 0 | 2 | 1 |
| **5** | 0 | 0 | 0 | 0 | 0 |
| **Weighted Kappa** | 0.387 | | | | |
| **Standard error (Kw'=0)** | 0.102 | | | | |
| **Standard error (Kw'#0)** | 0.153 | | | | |

**Interobserver Agreement for Chest Radiologists: Presence of Osteopenia, First Read, N=100**

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1          0.272104  0.100000   2.72104     0.0033
2         -0.071225  0.100000  -0.71225     0.7618
3         -0.084011  0.100000  -0.84011     0.7996
4         -0.006012  0.100000  -0.06012     0.5240
5          0.351351  0.100000   3.51351     0.0002
Overall    0.035540  0.054444   0.65278     0.2570

Cohen's Kappa Statistics
Response      Kappa   SE Kappa         Z  P(vs > 0)
1          0.290484  0.0888385   3.26980     0.0005
2         -0.061427  0.0979126  -0.62737     0.7348
3         -0.038961  0.0854482  -0.45596     0.6758
4          0.000000  0.0985318   0.00000     0.5000
5          0.362606  0.0868465   4.17526     0.0000
Overall    0.052932  0.0510222   1.03743     0.1498

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.688669   136.356  99  0.0076
```

| | First Chest Radiologist | | | | |
|---|---|---|---|---|---|
| **Second Chest Radiologist** | **1** | **2** | **3** | **4** | **5** |
| **1** | 5 | 2 | 1 | 0 | 0 |
| **2** | 8 | 9 | 5 | 13 | 2 |
| **3** | 6 | 8 | 2 | 10 | 0 |
| **4** | 0 | 9 | 2 | 8 | 6 |
| **5** | 0 | 0 | 0 | 1 | 3 |
| **Weighted Kappa** | 0.383 | | | | |
| **Standard error (Kw'=0)** | 0.096 | | | | |
| **Standard error (Kw'#0)** | 0.082 | | | | |

INTEROBSERVER AGREEMENT: NON-CHEST RADIOLOGIST

**Interobserver Agreement for Non-Chest Radiologist Compared to First Radiologist: End Plates Definition (Compared to Vertebral Body), N=34**

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1          0.284211  0.171499   1.65722     0.0487
2         -0.214286  0.171499  -1.24949     0.8943
3         -0.030303  0.171499  -0.17670     0.5701
4          0.059119  0.171499   0.34472     0.3652
5         -0.014925  0.171499  -0.08703     0.5347
Overall    0.079116  0.115217   0.68667     0.2461

Cohen's Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1          0.322259  0.152995   2.10634     0.0176
2         -0.207101  0.167992  -1.23280     0.8912
3          0.000000  0.000000         *          *
4          0.113744  0.141621   0.80316     0.2109
5          0.000000  0.000000         *          *
Overall    0.117486  0.101439   1.15820     0.1234

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.700559   46.2369  33  0.0629
```

| | First Chest Radiologist | | | | |
|---|---|---|---|---|---|
| **Non-Chest Radiologist** | **1** | **2** | **3** | **4** | **5** |
| **1** | 13 | 5 | 0 | 4 | 1 |
| **2** | 1 | 0 | 1 | 5 | 0 |
| **3** | 0 | 0 | 0 | 0 | 0 |
| **4** | 1 | 0 | 1 | 2 | 0 |
| **5** | 0 | 0 | 0 | 0 | 0 |
| **Weighted Kappa** | 0.256 | | | | |
| **Standard error (Kw'=0)** | 0.132 | | | | |
| **Standard error (Kw'#0)** | 0.129 | | | | |

**Interobserver Agreement for Non-Chest Radiologist Compared to Second Radiologist: End Plates Definition (Compared to Vertebral Body), N=37**

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa          Z  P(vs > 0)
1         -0.034161  0.164399  -0.20780     0.5823
2         -0.203828  0.164399  -1.23984     0.8925
3         -0.072464  0.164399  -0.44078     0.6703
4          0.040741  0.164399   0.24782     0.4021
5         -0.027778  0.164399  -0.16897     0.5671
Overall   -0.062019  0.099960  -0.62044     0.7325

Cohen's Kappa Statistics
Response      Kappa  SE Kappa          Z  P(vs > 0)
1          0.173697  0.092594   1.87591     0.0303
2         -0.175701  0.154257  -1.13901     0.8727
3          0.000000  0.000000          *          *
4          0.122034  0.128054   0.95299     0.1703
5          0.000000  0.000000          *          *
Overall    0.053881  0.069856   0.77132     0.2203

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.636298   45.8135  36  0.1266
```

| | **Non-Chest Radiologist** | | | | |
|---|---|---|---|---|---|
| **Second Chest Radiologist** | **1** | **2** | **3** | **4** | **5** |
| **1** | 5 | 0 | 0 | 0 | 0 |
| **2** | 8 | 1 | 0 | 2 | 1 |
| **3** | 3 | 1 | 0 | 0 | 1 |
| **4** | 7 | 5 | 0 | 3 | 0 |
| **5** | 0 | 0 | 0 | 0 | 0 |
| **Weighted Kappa** | 0.122 | | | | |
| **Standard error (Kw'=0)** | 0.121 | | | | |
| **Standard error (Kw'#0)** | 0.104 | | | | |

## Interobserver Agreement for Non-Chest Radiologist Compared to First Radiologist: End Plates Shape (Fish Vertebra), N=38

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1         -0.217949  0.162221  -1.34353     0.9104
2         -0.187500  0.162221  -1.15583     0.8761
3         -0.013333  0.162221  -0.08219     0.5328
4         -0.134328  0.162221  -0.82806     0.7962
5         -0.027027  0.162221  -0.16661     0.5662
Overall   -0.176636  0.111385  -1.58582     0.9436

Cohen's Kappa Statistics
Response       Kappa  SE Kappa           Z  P(vs > 0)
1          0.0330789  0.079918   0.413910     0.3395
2         -0.0961538  0.110736  -0.868313     0.8074
3          0.0000000  0.000000          *          *
4          0.0000000  0.000000          *          *
5         -0.0270270  0.162221  -0.166606     0.5662
Overall   -0.0096618  0.056798  -0.170109     0.5675

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.564866   41.8001  37  0.2702
```

| | First Chest Radiologist | | | | |
|---|---|---|---|---|---|
| **Non-Chest Radiologist** | **1** | **2** | **3** | **4** | **5** |
| **1** | 16 | 10 | 1 | 7 | 1 |
| **2** | 1 | 0 | 0 | 1 | 0 |
| **3** | 0 | 0 | 0 | 0 | 0 |
| **4** | 0 | 0 | 0 | 0 | 0 |
| **5** | 0 | 0 | 0 | 1 | 0 |
| **Weighted Kappa** | 0.142 | | | | |
| **Standard error (Kw'=0)** | 0.092 | | | | |
| **Standard error (Kw'#0)** | 0.119 | | | | |

**Interobserver Agreement for Non-Chest Radiologist Compared to Second Radiologist: End Plates Shape (Fish Vertebra), N=37**

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1         -0.523529  0.164399  -3.18451     0.9993
2         -0.116379  0.164399  -0.70791     0.7605
3         -0.072464  0.164399  -0.44078     0.6703
4         -0.193548  0.164399  -1.17731     0.8805
5         -0.013699  0.164399  -0.08333     0.5332
Overall   -0.286957  0.104904  -2.73542     0.9969

Cohen's Kappa Statistics
Response      Kappa   SE Kappa         Z  P(vs > 0)
1         0.0335821  0.0422465  0.794908     0.2133
2         0.0335821  0.0920961  0.364642     0.3577
3         0.0000000  0.0000000         *          *
4         0.0000000  0.0000000         *          *
5         0.0000000  0.0000000         *          *
Overall   0.0237467  0.0332670  0.713821     0.2377

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.577597   41.5870  36  0.2404
```

| | **Non-Chest Radiologist** | | | | |
|---|---|---|---|---|---|
| **Second Chest Radiologist** | **1** | **2** | **3** | **4** | **5** |
| **1** | 6 | 0 | 0 | 0 | 0 |
| **2** | 13 | 1 | 0 | 0 | 0 |
| **3** | 5 | 0 | 0 | 0 | 0 |
| **4** | 10 | 1 | 0 | 0 | 1 |
| **5** | 0 | 0 | 0 | 0 | 0 |
| **Weighted Kappa** | 0.089 | | | | |
| **Standard error (Kw'=0)** | 0.065 | | | | |
| **Standard error (Kw'#0)** | 0.075 | | | | |

**Interobserver Agreement for Non-Chest Radiologist Compared to First Radiologist: Body Shape (Wedged), N=38**

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1          0.318046  0.162221   1.96057     0.0250
2         -0.027027  0.162221  -0.16661     0.5662
4          0.208333  0.162221   1.28425     0.0995
5          0.652968  0.162221   4.02517     0.0000
Overall    0.289055  0.126112   2.29206     0.0110

Cohen's Kappa Statistics
Response     Kappa  SE Kappa        Z  P(vs > 0)
1         0.369004  0.125849  2.93211     0.0017
2         0.000000  0.000000        *          *
4         0.243781  0.133606  1.82462     0.0340
5         0.654545  0.152234  4.29959     0.0000
Overall   0.327434  0.100973  3.24279     0.0006

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.713390   52.7908  37  0.0446
```

| | First Chest Radiologist | | | |
|---|---|---|---|---|
| **Non-Chest Radiologist** | **1** | **2** | **4** | **5** |
| **1** | 25 | 1 | 7 | 1 |
| **2** | 0 | 0 | 0 | 0 |
| **4** | 0 | 1 | 2 | 0 |
| **5** | 0 | 0 | 0 | 1 |
| **Weighted Kappa** | 0.397 | | | |
| **Standard error (Kw'=0)** | 0.135 | | | |
| **Standard error (Kw'#0)** | 0.17 | | | |

## Interobserver Agreement for Non-Chest Radiologist Compared to Second Radiologist: Body Shape (Wedged), N=38

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa          Z  P(vs > 0)
1          0.474654  0.162221    2.92597     0.0017
2         -0.041096  0.162221   -0.25333     0.6000
4          0.441176  0.162221    2.71959     0.0033
5          0.652968  0.162221    4.02517     0.0000
Overall    0.424865  0.116929    3.63353     0.0001

Cohen's Kappa Statistics
Response     Kappa  SE Kappa         Z  P(vs > 0)
1         0.495575  0.140071  3.53803     0.0002
2         0.000000  0.000000        *          *
4         0.445255  0.155856  2.85683     0.0021
5         0.654545  0.152234  4.29959     0.0000
Overall   0.440000  0.102738  4.28273     0.0000

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.773230   57.2190  37  0.0180
```

| | Non-Chest Radiologist | | | |
|---|---|---|---|---|
| **Second Chest Radiologist** | **1** | **2** | **4** | **5** |
| **1** | 28 | 0 | 0 | 0 |
| **2** | 2 | 0 | 1 | 0 |
| **4** | 3 | 0 | 2 | 0 |
| **5** | 1 | 0 | 0 | 1 |
| **Weighted Kappa** | 0.547 | | | |
| **Standard error (Kw'=0)** | 0.149 | | | |
| **Standard error (Kw'#0)** | 0.189 | | | |

### Interobserver Agreement for Non-Chest Radiologist Compared to First Radiologist: Compression Fractures, N=37

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1          0.526652  0.164399   3.20350     0.0007
2         -0.072464  0.164399  -0.44078     0.6703
4         -0.013699  0.164399  -0.08333     0.5332
5         -0.013699  0.164399  -0.08333     0.5332
Overall    0.229167  0.131778   1.73903     0.0410

Cohen's Kappa Statistics
Response     Kappa  SE Kappa        Z  P(vs > 0)
1         0.535565  0.145593  3.67851     0.0001
2         0.000000  0.000000        *          *
4         0.000000  0.000000        *          *
5         0.000000  0.000000        *          *
Overall   0.257028  0.069873  3.67851     0.0001

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.778195   56.0301  36  0.0178
```

| | **First Chest Radiologist** | | | |
|---|---|---|---|---|
| **Non-Chest Radiologist** | **1** | **2** | **4** | **5** |
| **1** | 32 | 0 | 0 | 0 |
| **2** | 3 | 0 | 1 | 1 |
| **4** | 0 | 0 | 0 | 0 |
| **5** | 0 | 0 | 0 | 0 |
| **Weighted Kappa** | 0.519 | | | |
| **Standard error (Kw'=0)** | 0.144 | | | |
| **Standard error (Kw'#0)** | 0.092 | | | |

**Interobserver Agreement for Non-Chest Radiologist Compared to Second Radiologist: Compression Fractures, N=37**

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1          0.306250  0.164399   1.86285     0.0312
2          0.274510  0.164399   1.66978     0.0475
4         -0.057143  0.164399  -0.34759     0.6359
Overall    0.219880  0.128746   1.70785     0.0438

Cohen's Kappa Statistics
Response     Kappa  SE Kappa        Z  P(vs > 0)
1         0.306250  0.164399  1.86285     0.0312
2         0.301887  0.117708  2.56472     0.0052
4         0.000000  0.000000        *          *
Overall   0.238235  0.098620  2.41568     0.0079

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.643508   46.3325  36  0.1161
```

| | Non-Chest Radiologist | | |
|---|---|---|---|
| **Second Chest Radiologist** | **1** | **2** | **4** |
| **1** | 29 | 3 | 0 |
| **2** | 0 | 1 | 0 |
| **4** | 3 | 1 | 0 |
| **Weighted Kappa** | 0.182 | | |
| **Standard error (Kw'=0)** | 0.134 | | |
| **Standard error (Kw'#0)** | 0.17 | | |

## Interobserver Agreement for Non-Chest Radiologist Compared to First Radiologist: Vertical Striations, N=38

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa          Z  P(vs > 0)
1         -0.136590  0.162221  -0.841999     0.8001
2         -0.085714  0.162221  -0.528378     0.7014
3         -0.013333  0.162221  -0.082192     0.5328
4          0.117745  0.162221   0.725827     0.2340
5         -0.013333  0.162221  -0.082192     0.5328
Overall   -0.047794  0.115301  -0.414515     0.6608

Cohen's Kappa Statistics
Response      Kappa  SE Kappa          Z  P(vs > 0)
1         -0.135458  0.161756  -0.837423     0.7988
2         -0.085714  0.162221  -0.528378     0.7014
3          0.000000  0.000000          *          *
4          0.119205  0.160932   0.740718     0.2294
5          0.000000  0.000000          *          *
Overall   -0.045872  0.114093  -0.402054     0.6562

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.445483   32.9657  37  0.6587
```

| | **First Chest Radiologist** | | | | |
|---|---|---|---|---|---|
| **Non-Chest Radiologist** | **1** | **2** | **3** | **4** | **5** |
| **1** | 22 | 3 | 1 | 4 | 0 |
| **2** | 3 | 0 | 0 | 0 | 0 |
| **3** | 0 | 0 | 0 | 0 | 0 |
| **4** | 3 | 0 | 0 | 1 | 0 |
| **5** | 1 | 0 | 0 | 0 | 0 |
| **Weighted Kappa** | -0.023 | | | | |
| **Standard error (Kw'=0)** | 0.162 | | | | |
| **Standard error (Kw'#0)** | 0.171 | | | | |

**Interobserver Agreement for Non-Chest Radiologist Compared to Second Radiologist: Vertical Striations, N=38**

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1          0.309091  0.162221   1.90536     0.0284
2         -0.041096  0.162221  -0.25333     0.6000
3         -0.013333  0.162221  -0.08219     0.5328
4          0.357746  0.162221   2.20530     0.0137
5         -0.013333  0.162221  -0.08219     0.5328
Overall    0.231214  0.113342   2.03997     0.0207

Cohen's Kappa Statistics
Response     Kappa  SE Kappa        Z  P(vs > 0)
1         0.344828  0.122555  2.81366     0.0024
2         0.000000  0.000000        *          *
3         0.000000  0.000000        *          *
4         0.373626  0.126455  2.95461     0.0016
5         0.000000  0.000000        *          *
Overall   0.261111  0.079278  3.29363     0.0005

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.702700   51.9998  37  0.0519
```

| | Non-Chest Radiologist | | | | |
|---|---|---|---|---|---|
| **Second Chest Radiologist** | **1** | **2** | **3** | **4** | **5** |
| **1** | 30 | 3 | 0 | 2 | 1 |
| **2** | 0 | 0 | 0 | 0 | 0 |
| **3** | 0 | 0 | 0 | 1 | 0 |
| **4** | 0 | 0 | 0 | 1 | 0 |
| **5** | 0 | 0 | 0 | 0 | 0 |
| **Weighted Kappa** | 0.397 | | | | |
| **Standard error (Kw'=0)** | 0.122 | | | | |
| **Standard error (Kw'#0)** | 0.214 | | | | |

### Interobserver Agreement for Non-Chest Radiologist Compared to First Radiologist: Empty Box, N=38

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1          0.441176  0.162221   2.71959     0.0033
2          0.472222  0.162221   2.91097     0.0018
3         -0.013333  0.162221  -0.08219     0.5328
4         -0.027027  0.162221  -0.16661     0.5662
5         -0.013333  0.162221  -0.08219     0.5328
Overall    0.327434  0.112646   2.90676     0.0018

Cohen's Kappa Statistics
Response     Kappa  SE Kappa        Z  P(vs > 0)
1         0.445255  0.155856  2.85683     0.0021
2         0.472222  0.162221  2.91097     0.0018
3         0.000000  0.000000        *          *
4         0.000000  0.000000        *          *
5         0.000000  0.000000        *          *
Overall   0.333333  0.101421  3.28663     0.0005

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.733450   54.2753  37  0.0332
```

| | First Chest Radiologist | | | | |
|---|---|---|---|---|---|
| **Non-Chest Radiologist** | **1** | **2** | **3** | **4** | **5** |
| **1** | 32 | 1 | 0 | 2 | 0 |
| **2** | 1 | 1 | 0 | 0 | 0 |
| **3** | 0 | 0 | 0 | 0 | 1 |
| **4** | 0 | 0 | 0 | 0 | 0 |
| **5** | 0 | 0 | 0 | 0 | 0 |
| **Weighted Kappa** | 0.392 | | | | |
| **Standard error (Kw'=0)** | 0.11 | | | | |
| **Standard error (Kw'#0)** | 0.226 | | | | |

## Interobserver Agreement for Non-Chest Radiologist Compared to Second Radiologist: Empty Box, N=37

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1          0.274510  0.164399   1.66978     0.0475
2         -0.042254  0.164399  -0.25702     0.6014
3         -0.013699  0.164399  -0.08333     0.5332
4         -0.027778  0.164399  -0.16897     0.5671
Overall    0.116945  0.118923   0.98337     0.1627

Cohen's Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1          0.274510  0.164399   1.66978     0.0475
2         -0.037383  0.154257  -0.24234     0.5957
3          0.000000  0.000000         *          *
4          0.000000  0.000000         *          *
Overall    0.123223  0.103212   1.19388     0.1163

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.651984   46.9429  36  0.1047
```

| | **Non-Chest Radiologist** | | | |
|---|---|---|---|---|
| **Second Chest Radiologist** | **1** | **2** | **3** | **4** |
| **1** | 32 | 2 | 0 | 0 |
| **2** | 1 | 0 | 0 | 0 |
| **3** | 0 | 0 | 0 | 0 |
| **4** | 1 | 0 | 1 | 0 |
| **Weighted Kappa** | 0.446 | | | |
| **Standard error (Kw'=0)** | 0.139 | | | |
| **Standard error (Kw'#0)** | 0.3 | | | |

**Interobserver Agreement for Non-Chest Radiologist Compared to First Radiologist: Presence of Osteopenia, N=38**

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1         -0.106029  0.162221  -0.65361     0.7433
2         -0.206349  0.162221  -1.27202     0.8983
3         -0.027027  0.162221  -0.16661     0.5662
4          0.050000  0.162221   0.30822     0.3790
5          0.276190  0.162221   1.70255     0.0443
Overall   -0.042744  0.099540  -0.42942     0.6662

Cohen's Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1          0.152866  0.086201   1.77337     0.0381
2         -0.097778  0.105431  -0.92741     0.8231
3          0.000000  0.000000         *          *
4          0.109375  0.130531   0.83792     0.2010
5          0.283019  0.151439   1.86886     0.0308
Overall    0.098540  0.063075   1.56228     0.0591

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.702147   51.9589  37  0.0523
```

| | First Chest Radiologist | | | | |
|---|---|---|---|---|---|
| **Non-Chest Radiologist** | **1** | **2** | **3** | **4** | **5** |
| **1** | 9 | 10 | 1 | 8 | 2 |
| **2** | 0 | 0 | 1 | 1 | 0 |
| **3** | 0 | 0 | 0 | 0 | 0 |
| **4** | 0 | 1 | 0 | 2 | 1 |
| **5** | 0 | 0 | 0 | 1 | 1 |
| **Weighted Kappa** | 0.297 | | | | |
| **Standard error (Kw'=0)** | 0.114 | | | | |
| **Standard error (Kw'#0)** | 0.118 | | | | |

**Interobserver Agreement for Non-Chest Radiologist Compared to Second Radiologist: Presence of Osteopenia, N=38**

```
Fleiss' Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1         -0.511364  0.162221  -3.15226     0.9992
2         -0.192982  0.162221  -1.18962     0.8829
3         -0.134328  0.162221  -0.82806     0.7962
4          0.164835  0.162221   1.01611     0.1548
5         -0.041096  0.162221  -0.25333     0.6000
Overall   -0.213940  0.093948  -2.27723     0.9886

Cohen's Kappa Statistics
Response      Kappa  SE Kappa         Z  P(vs > 0)
1          0.029197  0.038913   0.75031     0.2265
2          0.012232  0.079535   0.15380     0.4389
3          0.000000  0.000000         *          *
4          0.189573  0.144849   1.30877     0.0953
5         -0.036364  0.152234  -0.23887     0.5944
Overall    0.044207  0.037607   1.17549     0.1199

Kendall's Coefficient of Concordance
    Coef  Chi - Sq  DF       P
0.656678   48.5941  37  0.0961
```

| | Non-Chest Radiologist | | | | |
|---|---|---|---|---|---|
| **Second Chest Radiologist** | **1** | **2** | **3** | **4** | **5** |
| **1** | 2 | 0 | 0 | 0 | 0 |
| **2** | 15 | 1 | 0 | 0 | 1 |
| **3** | 7 | 1 | 0 | 1 | 0 |
| **4** | 6 | 0 | 0 | 2 | 1 |
| **5** | 0 | 0 | 0 | 1 | 0 |
| **Weighted Kappa** | 0.234 | | | | |
| **Standard error (Kw'=0)** | 0.102 | | | | |
| **Standard error (Kw'#0)** | 0.114 | | | | |

ACCURACY ASSESSMENT

**First Chest Radiologist, End Plates Definition, CXR Compared to DXA (All Locations)**

| | Count |
|---|---|
| normal | 73 |
| osteopenia | 23 |
| N= | 96 |

| Impression | Count |
|---|---|
| False Negative | 49 |
| False Positive | 9 |
| True Negative | 24 |
| True Positive | 14 |
| N= | 96 |

| Lumbar | Count |
|---|---|
| False Negative | 43 |
| False Positive | 12 |
| True Negative | 30 |
| True Positive | 11 |
| N= | 96 |

| Hip | Count |
|---|---|
| False Negative | 29 |
| False Positive | 14 |
| True Negative | 41 |
| True Positive | 8 |
| N= | 92 |

| FemNeck | Count |
|---|---|
| False Negative | 36 |
| False Positive | 13 |
| True Negative | 33 |
| True Positive | 10 |
| N= | 92 |

| Trochanter | Count |
|---|---|
| False Negative | 26 |
| False Positive | 11 |
| True Negative | 27 |
| True Positive | 10 |
| N= | 74 |

**First Chest Radiologist, End Plates Definition, CXR Compared to DXA (Lumbar Spine)**

| | | |
|---|---|---|
| Sensitivity (95% CI) | 20.37% | (10.65% to 33.53%) |
| Specificity (95% CI) | 71.43% | (55.41% to 84.27%) |
| Positive Likelihood Ratio (95% CI) | 0.71 | (0.35 to 1.45) |
| Negative Likelihood Ratio (95% CI) | 1.11 | (0.88 to 1.41) |
| Disease prevalence (95% CI) | 56.25% | (45.75% to 66.36%) |
| Positive Predictive Value (95% CI) | 47.83% | (26.85% to 69.39%) |
| Negative Predictive Value (95% CI) | 41.10% | (29.71% to 53.23%) |

**First Chest Radiologist, End Plates Definition, CXR Compared to DXA (Hip)**

| | | |
|---|---|---|
| Sensitivity (95% CI) | 21.62% | 9.86% to 38.22% |
| Specificity (95% CI) | 74.55% | 60.99% to 85.32% |
| Positive Likelihood Ratio (95% CI) | 0.85 | 0.40 to 1.82 |
| Negative Likelihood Ratio (95% CI) | 1.05 | 0.84 to 1.32 |
| Disease prevalence (95% CI) | 40.22% | 30.12% to 50.96% |
| Positive Predictive Value (95% CI) | 36.36% | 17.24% to 59.33% |
| Negative Predictive Value (95% CI) | 58.57% | 46.17% to 70.23% |

**First Chest Radiologist, End Plates Definition, CXR Compared to DXA (Femoral Neck)**

| | | |
|---|---|---|
| Sensitivity (95% CI) | 21.74% | 10.97% to 36.37% |
| Specificity (95% CI) | 71.74% | 56.54% to 84.00% |
| Positive Likelihood Ratio (95% CI) | 0.77 | 0.38 to 1.57 |
| Negative Likelihood Ratio (95% CI) | 1.09 | 0.86 to 1.38 |
| Disease prevalence (95% CI) | 50.00% | 39.39% to 60.61% |
| Positive Predictive Value (95% CI) | 43.48% | 23.22% to 65.49% |
| Negative Predictive Value (95% CI) | 47.83% | 35.65% to 60.20% |

**First Chest Radiologist, End Plates Definition, CXR Compared to DXA (Trochanter)**

| | | |
|---|---|---|
| Sensitivity (95% CI) | 27.78% | 14.22% to 45.19% |
| Specificity (95% CI) | 71.05% | 54.09% to 84.56% |
| Positive Likelihood Ratio (95% CI) | 0.96 | 0.46 to 1.98 |
| Negative Likelihood Ratio (95% CI) | 1.02 | 0.76 to 1.35 |
| Disease prevalence (95% CI) | 48.65% | 36.85% to 60.56% |
| Positive Predictive Value (95% CI) | 47.62% | 25.75% to 70.19% |
| Negative Predictive Value (95% CI) | 50.94% | 36.84% to 64.93% |

### First Chest Radiologist, End Plates Shape (Fish Vertebra), CXR Compared to DXA (All Locations)

```
     R1FQ2  Count           Lumbar_1  Count             Hip_1  Count
    normal     70     False Negative     38    False Negative     23
osteopenia     30     False Positive     10    False Positive     12
        N=    100      True Negative     32     True Negative     43
                       True Positive     20     True Positive     17
                                  N=    100                N=     95
                                                           *=      4

     FemNeck_1  Count      Trochanter_1  Count
False Negative     32    False Negative     25
False Positive     12    False Positive     10
 True Negative     35     True Negative     28
 True Positive     17     True Positive     14
            N=     96                N=     77
            *=      3                *=     17
```

### First Chest Radiologist, End Plates Shape (Fish Vertebra), CXR Compared to DXA (Lumbar Spine)

| | | |
|---|---|---|
| Sensitivity (95% CI) | 34.48% | 22.50% to 48.12% |
| Specificity (95% CI) | 76.19% | 60.55% to 87.93% |
| Positive Likelihood Ratio (95% CI) | 1.45 | 0.76 to 2.77 |
| Negative Likelihood Ratio (95% CI) | 0.86 | 0.67 to 1.11 |
| Disease prevalence (95% CI) | 58.00% | 47.71% to 67.80% |
| Positive Predictive Value (95% CI) | 66.67% | 47.19% to 82.69% |
| Negative Predictive Value (95% CI) | 45.71% | 33.75% to 58.06% |

### First Chest Radiologist, End Plates Shape (Fish Vertebra), CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity (95% CI) | 42.50% | 27.05% to 59.11% |
| Specificity (95% CI) | 78.18% | 64.99% to 88.17% |
| Positive Likelihood Ratio (95% CI) | 1.95 | 1.05 to 3.61 |
| Negative Likelihood Ratio (95% CI) | 0.74 | 0.54 to 0.99 |
| Disease prevalence (95% CI) | 42.11% | 32.04% to 52.67% |
| Positive Predictive Value (95% CI) | 58.62% | 38.94% to 76.46% |
| Negative Predictive Value (95% CI) | 65.15% | 52.42% to 76.47% |

### First Chest Radiologist, End Plates Shape (Fish Vertebra), CXR Compared to DXA (FemNeck)

| | | |
|---|---|---|
| Sensitivity (95% CI) | 34.69% | 21.68% to 49.64% |
| Specificity (95% CI) | 74.47% | 59.65% to 86.04% |
| Positive Likelihood Ratio (95% CI) | 1.36 | 0.73 to 2.53 |
| Negative Likelihood Ratio (95% CI) | 0.88 | 0.67 to 1.14 |
| Disease prevalence (95% CI) | 51.04% | 40.63% to 61.39% |
| Positive Predictive Value (95% CI) | 58.62% | 38.94% to 76.46% |
| Negative Predictive Value (95% CI) | 52.24% | 39.68% to 64.60% |

### First Chest Radiologist, End Plates Shape (Fish Vertebra), CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity (95% CI) | 35.90% | 21.22% to 52.82% |
| Specificity (95% CI) | 73.68% | 56.90% to 86.58% |
| Positive Likelihood Ratio (95% CI) | 1.36 | 0.69 to 2.69 |
| Negative Likelihood Ratio (95% CI) | 0.87 | 0.64 to 1.18 |
| Disease prevalence (95% CI) | 50.65% | 39.01% to 62.24% |
| Positive Predictive Value (95% CI) | 58.33% | 36.66% to 77.86% |
| Negative Predictive Value (95% CI) | 52.83% | 38.64% to 66.69% |

### First Chest Radiologist, Body Shape (Wedged), CXR Compared to DXA (All Locations)

```
     R1FQ3  Count          Lumbar_2  Count            Hip_2  Count
    normal     72    False Negative     40   False Negative     26
osteopenia     28    False Positive     10   False Positive     13
        N=    100     True Negative     32    True Negative     42
                      True Positive     18    True Positive     14
                                 N=    100               N=     95
                                                         *=      4

     FemNeck_2  Count      Trochanter_2  Count
False Negative     32    False Negative     25
False Positive     11    False Positive     11
 True Negative     36     True Negative     27
 True Positive     17     True Positive     14
            N=     96                N=     77
            *=      4                *=     20
```

### First Chest Radiologist, Body Shape (Wedged), CXR Compared to DXA (Lumbar Spine)

| | | |
|---|---|---|
| Sensitivity | 31.03% | 19.55% to 44.55% |
| Specificity | 76.19% | 60.55% to 87.93% |
| Positive Likelihood Ratio | 1.30 | 0.67 to 2.53 |
| Negative Likelihood Ratio | 0.91 | 0.71 to 1.15 |
| Disease prevalence | 58.00% | 47.71% to 67.80% |
| Positive Predictive Value | 64.29% | 44.07% to 81.33% |
| Negative Predictive Value | 44.44% | 32.73% to 56.63% |

### First Chest Radiologist, Body Shape (Wedged), CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 35.00% | 20.64% to 51.68% |
| Specificity | 76.36% | 62.98% to 86.76% |
| Positive Likelihood Ratio | 1.48 | 0.78 to 2.80 |
| Negative Likelihood Ratio | 0.85 | 0.65 to 1.12 |
| Disease prevalence | 42.11% | 32.04% to 52.67% |
| Positive Predictive Value | 51.85% | 31.96% to 71.32% |
| Negative Predictive Value | 61.76% | 49.18% to 73.29% |

### First Chest Radiologist, Body Shape (Wedged), CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 34.69% | 21.68% to 49.64% |
| Specificity | 76.60% | 61.97% to 87.68% |
| Positive Likelihood Ratio | 1.48 | 0.78 to 2.82 |
| Negative Likelihood Ratio | 0.85 | 0.66 to 1.10 |
| Disease prevalence | 51.04% | 40.63% to 61.39% |
| Positive Predictive Value | 60.71% | 40.58% to 78.47% |
| Negative Predictive Value | 52.94% | 40.45% to 65.17% |

### First Chest Radiologist, Body Shape (Wedged), CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 35.90% | 21.22% to 52.82% |
| Specificity | 71.05% | 54.09% to 84.56% |
| Positive Likelihood Ratio | 1.24 | 0.65 to 2.38 |
| Negative Likelihood Ratio | 0.90 | 0.66 to 1.23 |
| Disease prevalence | 50.65% | 39.01% to 62.24% |
| Positive Predictive Value | 56.00% | 34.94% to 75.57% |
| Negative Predictive Value | 51.92% | 37.63% to 65.98% |

### First Chest Radiologist, Compression Fractures, CXR Compared to DXA (All Locations)

```
     R1FQ4  Count          Lumbar_3  Count            Hip_3  Count
    normal     97    False Negative     56   False Negative     37
osteopenia      2    False Positive      1    True Negative     55
        N=     99     True Negative     41    True Positive      2
                      True Positive      1               N=     94
                                 N=     99               *=      5

     FemNeck_3  Count      Trochanter_3  Count
False Negative     46    False Negative     36
 True Negative     47     True Negative     38
 True Positive      2     True Positive      2
            N=     95                N=     76
            *=      4                *=     23
```

### First Chest Radiologist, Compression Fractures, CXR Compared to DXA (Lumbar)

| | | |
|---|---|---|
| Sensitivity | 1.75% | 0.29% to 9.43% |
| Specificity | 97.62% | 87.39% to 99.60% |
| Positive Likelihood Ratio | 0.74 | 0.05 to 11.45 |
| Negative Likelihood Ratio | 1.01 | 0.95 to 1.07 |
| Disease prevalence | 57.58% | 47.23% to 67.45% |
| Positive Predictive Value | 50.00% | 8.17% to 91.83% |
| Negative Predictive Value | 42.27% | 32.30% to 52.72% |

### First Chest Radiologist, Compression Fractures, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 5.13% | 0.78% to 17.36% |
| Specificity | 100.00% | 93.45% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 0.95 | 0.88 to 1.02 |
| Disease prevalence | 41.49% | 31.42% to 52.12% |
| Positive Predictive Value | 100.00% | 19.29% to 100.00% |
| Negative Predictive Value | 59.78% | 49.04% to 69.88% |

### First Chest Radiologist, Compression Fractures, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 4.17% | 0.63% to 14.28% |
| Specificity | 100.00% | 92.38% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 0.96 | 0.90 to 1.02 |
| Disease prevalence | 50.53% | 40.07% to 60.95% |
| Positive Predictive Value | 100.00% | 19.29% to 100.00% |
| Negative Predictive Value | 50.54% | 39.97% to 61.07% |

### First Chest Radiologist, Compression Fractures, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 5.26% | 0.80% to 17.78% |
| Specificity | 100.00% | 90.66% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 0.95 | 0.88 to 1.02 |
| Disease prevalence | 50.00% | 38.30% to 61.70% |
| Positive Predictive Value | 100.00% | 19.29% to 100.00% |
| Negative Predictive Value | 51.35% | 39.44% to 63.15% |

**First Chest Radiologist, Vertical Striations, CXR Compared to DXA (All Locations)**

```
     R1FQ5  Count          Lumbar_4  Count           Hip_4  Count
    normal     83    False Negative     47    False Negative     32
osteopenia     17    False Positive      6    False Positive      7
        N=    100     True Negative     36     True Negative     48
                      True Positive     11     True Positive      8
                                 N=    100                N=     95
                                                          *=      3

     FemNeck_4  Count      Trochanter_4  Count
False Negative     40    False Negative     31
False Positive      7    False Positive      6
 True Negative     40     True Negative     32
 True Positive      9     True Positive      8
            N=     96                N=     77
            *=      3                *=     20
```

**First Chest Radiologist, Vertical Striations, CXR Compared to DXA (Lumbar)**

| | | |
|---|---|---|
| Sensitivity | 18.97% | 9.88% to 31.41% |
| Specificity | 85.71% | 71.45% to 94.54% |
| Positive Likelihood Ratio | 1.33 | 0.53 to 3.30 |
| Negative Likelihood Ratio | 0.95 | 0.79 to 1.13 |
| Disease prevalence | 58.00% | 47.71% to 67.80% |
| Positive Predictive Value | 64.71% | 38.35% to 85.70% |
| Negative Predictive Value | 43.37% | 32.53% to 54.71% |

**First Chest Radiologist, Vertical Striations, CXR Compared to DXA (Hip)**

| | | |
|---|---|---|
| Sensitivity | 20.00% | 9.08% to 35.65% |
| Specificity | 87.27% | 75.51% to 94.70% |
| Positive Likelihood Ratio | 1.57 | 0.62 to 3.98 |
| Negative Likelihood Ratio | 0.92 | 0.76 to 1.10 |
| Disease prevalence | 42.11% | 32.04% to 52.67% |
| Positive Predictive Value | 53.33% | 26.65% to 78.66% |
| Negative Predictive Value | 60.00% | 48.44% to 70.80% |

**First Chest Radiologist, Vertical Striations, CXR Compared to DXA (Femoral Neck)**

| | | |
|---|---|---|
| Sensitivity | 18.37% | 8.78% to 32.03% |
| Specificity | 85.11% | 71.69% to 93.77% |
| Positive Likelihood Ratio | 1.23 | 0.50 to 3.04 |
| Negative Likelihood Ratio | 0.96 | 0.80 to 1.15 |
| Disease prevalence | 51.04% | 40.63% to 61.39% |
| Positive Predictive Value | 56.25% | 29.92% to 80.17% |
| Negative Predictive Value | 50.00% | 38.61% to 61.39% |

**First Chest Radiologist, Vertical Striations, CXR Compared to DXA (Trochanter)**

| | | |
|---|---|---|
| Sensitivity | 20.51% | 9.32% to 36.47% |
| Specificity | 84.21% | 68.74% to 93.94% |
| Positive Likelihood Ratio | 1.30 | 0.50 to 3.39 |
| Negative Likelihood Ratio | 0.94 | 0.76 to 1.17 |
| Disease prevalence | 50.65% | 39.01% to 62.24% |
| Positive Predictive Value | 57.14% | 28.92% to 82.24% |
| Negative Predictive Value | 50.79% | 37.89% to 63.62% |

### First Chest Radiologist, Empty Box, CXR Compared to DXA (All Locations)

```
     R1FQ6  Count          Lumbar_5  Count            Hip_5  Count
    normal     94    False Negative     52   False Negative     36
osteopenia      6     True Negative     42   False Positive      1
        N=    100     True Positive      6    True Negative     54
                                 N=    100    True Positive      4
                                                         N=     95
                                                         *=      4

     FemNeck_5  Count      Trochanter_5  Count
False Negative     43    False Negative     35
 True Negative     47     True Negative     38
 True Positive      6     True Positive      4
            N=     96                N=     77
            *=      4                *=     21
```

### First Chest Radiologist, Empty Box, CXR Compared to DXA (Lumbar)

| | | |
|---|---|---|
| Sensitivity | 10.34% | 3.92% to 21.18% |
| Specificity | 100.00% | 91.51% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 0.90 | 0.82 to 0.98 |
| Disease prevalence | 58.00% | 47.71% to 67.80% |
| Positive Predictive Value | 100.00% | 54.05% to 100.00% |
| Negative Predictive Value | 44.68% | 34.42% to 55.29% |

### First Chest Radiologist, Empty Box, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 10.00% | 2.85% to 23.68% |
| Specificity | 98.18% | 90.24% to 99.70% |
| Positive Likelihood Ratio | 5.50 | 0.64 to 47.37 |
| Negative Likelihood Ratio | 0.92 | 0.82 to 1.02 |
| Disease prevalence | 42.11% | 32.04% to 52.67% |
| Positive Predictive Value | 80.00% | 28.81% to 96.70% |
| Negative Predictive Value | 60.00% | 49.13% to 70.19% |

### First Chest Radiologist, Empty Box, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 12.24% | 4.66% to 24.78% |
| Specificity | 100.00% | 92.38% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 0.88 | 0.79 to 0.97 |
| Disease prevalence | 51.04% | 40.63% to 61.39% |
| Positive Predictive Value | 100.00% | 54.05% to 100.00% |
| Negative Predictive Value | 52.22% | 41.43% to 62.87% |

### First Chest Radiologist, Empty Box, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 10.26% | 2.93% to 24.24% |
| Specificity | 100.00% | 90.66% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 0.90 | 0.81 to 1.00 |
| Disease prevalence | 50.65% | 39.01% to 62.24% |
| Positive Predictive Value | 100.00% | 40.23% to 100.00% |
| Negative Predictive Value | 52.05% | 40.04% to 63.90% |

### First Chest Radiologist, Presence of Osteopenia, First Reading, CXR Compared to DXA (All Locations)

| R1FQ7 | Count |
|---|---|
| normal | 57 |
| osteopenia | 43 |
| N= | 100 |

| Impression | Count |
|---|---|
| False Negative | 33 |
| False Positive | 9 |
| True Negative | 24 |
| True Positive | 34 |
| N= | 100 |

| Status Lumbar | Count |
|---|---|
| False Negative | 28 |
| False Positive | 13 |
| True Negative | 29 |
| True Positive | 30 |
| N= | 100 |

| Status Hip | Count |
|---|---|
| False Negative | 16 |
| False Positive | 16 |
| True Negative | 39 |
| True Positive | 24 |
| N= | 95 |
| *= | 2 |

| Status FemNeck | Count |
|---|---|
| False Negative | 20 |
| False Positive | 13 |
| True Negative | 34 |
| True Positive | 29 |
| N= | 96 |
| *= | 3 |

| Status Trochanter | Count |
|---|---|
| False Negative | 19 |
| False Positive | 12 |
| True Negative | 26 |
| True Positive | 20 |
| N= | 77 |
| *= | 13 |

### First Chest Radiologist, Presence of Osteopenia, First Reading, CXR Compared to DXA (Lumbar Spine)

| | | |
|---|---|---|
| Sensitivity | 51.72% | 38.22% to 65.05% |
| Specificity | 69.05% | 52.91% to 82.36% |
| Positive Likelihood Ratio | 1.67 | 1.00 to 2.80 |
| Negative Likelihood Ratio | 0.70 | 0.50 to 0.98 |
| Disease prevalence | 58.00% | 47.71% to 67.80% |
| Positive Predictive Value | 69.77% | 53.87% to 82.80% |
| Negative Predictive Value | 50.88% | 37.29% to 64.37% |

### First Chest Radiologist, Presence of Osteopenia, First Reading, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 60.00% | 43.33% to 75.12% |
| Specificity | 70.91% | 57.10% to 82.36% |
| Positive Likelihood Ratio | 2.06 | 1.27 to 3.35 |
| Negative Likelihood Ratio | 0.56 | 0.37 to 0.85 |
| Disease prevalence | 42.11% | 32.04% to 52.67% |
| Positive Predictive Value | 60.00% | 43.33% to 75.12% |
| Negative Predictive Value | 70.91% | 57.10% to 82.36% |

### First Chest Radiologist, Presence of Osteopenia, First Reading, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 59.18% | 44.21% to 73.00% |
| Specificity | 72.34% | 57.36% to 84.36% |
| Positive Likelihood Ratio | 2.14 | 1.28 to 3.59 |
| Negative Likelihood Ratio | 0.56 | 0.39 to 0.83 |
| Disease prevalence | 51.04% | 40.63% to 61.39% |
| Positive Predictive Value | 69.05% | 52.91% to 82.36% |
| Negative Predictive Value | 62.96% | 48.74% to 75.70% |

### First Chest Radiologist, Presence of Osteopenia, First Reading, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 51.28% | 34.79% to 67.58% |
| Specificity | 68.42% | 51.35% to 82.48% |
| Positive Likelihood Ratio | 1.62 | 0.93 to 2.84 |
| Negative Likelihood Ratio | 0.71 | 0.48 to 1.05 |
| Disease prevalence | 50.65% | 39.01% to 62.24% |
| Positive Predictive Value | 62.50% | 43.70% to 78.88% |
| Negative Predictive Value | 57.78% | 42.15% to 72.34% |

### First Chest Radiologist, Presence of Osteopenia, Second Reading

| R1SQ7 | Count |
|---|---|
| normal | 62 |
| osteopenia | 38 |
| N= | 100 |

| Impression | Count |
|---|---|
| False Negative | 37 |
| False Positive | 8 |
| True Negative | 25 |
| True Positive | 30 |
| N= | 100 |

| Lumbar | Count |
|---|---|
| False Negative | 30 |
| False Positive | 10 |
| True Negative | 32 |
| True Positive | 28 |
| N= | 100 |

| Hip | Count |
|---|---|
| False Negative | 19 |
| False Positive | 15 |
| True Negative | 40 |
| True Positive | 21 |
| N= | 95 |

| FemNeck | Count |
|---|---|
| False Negative | 24 |
| False Positive | 12 |
| True Negative | 35 |
| True Positive | 25 |
| N= | 96 |

| Trochanter | Count |
|---|---|
| False Negative | 23 |
| False Positive | 10 |
| True Negative | 28 |
| True Positive | 16 |
| N= | 77 |

### First Chest Radiologist, Presence of Osteopenia, Second Reading, CXR Compared to DXA (Lumbar Spine)

| | | |
|---|---|---|
| Sensitivity | 48.28% | 34.95% to 61.78% |
| Specificity | 76.19% | 60.55% to 87.93% |
| Positive Likelihood Ratio | 2.03 | 1.11 to 3.71 |
| Negative Likelihood Ratio | 0.68 | 0.50 to 0.92 |
| Disease prevalence | 58.00% | 47.71% to 67.80% |
| Positive Predictive Value | 73.68% | 56.90% to 86.58% |
| Negative Predictive Value | 51.61% | 38.57% to 64.50% |

### First Chest Radiologist, Presence of Osteopenia, Second Reading, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 52.50% | 36.13% to 68.48% |
| Specificity | 72.73% | 59.04% to 83.85% |
| Positive Likelihood Ratio | 1.93 | 1.14 to 3.25 |
| Negative Likelihood Ratio | 0.65 | 0.45 to 0.94 |
| Disease prevalence | 42.11% | 32.04% to 52.67% |
| Positive Predictive Value | 58.33% | 40.76% to 74.47% |
| Negative Predictive Value | 67.80% | 54.36% to 79.37% |

### First Chest Radiologist, Presence of Osteopenia, Second Reading, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 51.02% | 36.34% to 65.57% |
| Specificity | 74.47% | 59.65% to 86.04% |
| Positive Likelihood Ratio | 2.00 | 1.14 to 3.50 |
| Negative Likelihood Ratio | 0.66 | 0.47 to 0.92 |
| Disease prevalence | 51.04% | 40.63% to 61.39% |
| Positive Predictive Value | 67.57% | 50.21% to 81.97% |
| Negative Predictive Value | 59.32% | 45.75% to 71.93% |

### First Chest Radiologist, Presence of Osteopenia, Second Reading, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 41.03% | 25.58% to 57.90% |
| Specificity | 73.68% | 56.90% to 86.58% |
| Positive Likelihood Ratio | 1.56 | 0.81 to 2.99 |
| Negative Likelihood Ratio | 0.80 | 0.58 to 1.11 |
| Disease prevalence | 50.65% | 39.01% to 62.24% |
| Positive Predictive Value | 61.54% | 40.58% to 79.75% |
| Negative Predictive Value | 54.90% | 40.34% to 68.87% |

### Second Chest Radiologist, End Plates Definition, First Reading, CXR Compared to DXA (All Locations)

| R3FQ1 | Count |
|---|---|
| normal | 61 |
| osteopenia | 39 |
| N= | 100 |

| Lumbar | Count |
|---|---|
| False Negative | 29 |
| False Positive | 10 |
| True Negative | 32 |
| True Positive | 29 |
| N= | 100 |

| Hip | Count |
|---|---|
| False Negative | 22 |
| False Positive | 20 |
| True Negative | 35 |
| True Positive | 18 |
| N= | 95 |

| FemNeck | Count |
|---|---|
| False Negative | 28 |
| False Positive | 17 |
| True Negative | 30 |
| True Positive | 21 |
| N= | 96 |

| Trochanter | Count |
|---|---|
| False Negative | 22 |
| False Positive | 14 |
| True Negative | 24 |
| True Positive | 17 |
| N= | 77 |

### Second Chest Radiologist, End Plates Definition, First Reading, CXR Compared to DXA (Lumbar)

| | | |
|---|---|---|
| Sensitivity | 50.00% | 36.58% to 63.42% |
| Specificity | 76.19% | 60.55% to 87.93% |
| Positive Likelihood Ratio | 2.10 | 1.15 to 3.82 |
| Negative Likelihood Ratio | 0.66 | 0.48 to 0.89 |
| Disease prevalence | 58.00% | 47.71% to 67.80% |
| Positive Predictive Value | 74.36% | 57.87% to 86.94% |
| Negative Predictive Value | 52.46% | 39.27% to 65.40% |

### Second Chest Radiologist, End Plates Definition, First Reading, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 45.00% | 29.27% to 61.51% |
| Specificity | 63.64% | 49.56% to 76.18% |
| Positive Likelihood Ratio | 1.24 | 0.76 to 2.02 |
| Negative Likelihood Ratio | 0.86 | 0.61 to 1.22 |
| Disease prevalence | 42.11% | 32.04% to 52.67% |
| Positive Predictive Value | 47.37% | 30.99% to 64.18% |
| Negative Predictive Value | 61.40% | 47.58% to 74.00% |

### Second Chest Radiologist, End Plates Definition, First Reading, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 42.86% | 28.83% to 57.79% |
| Specificity | 63.83% | 48.52% to 77.32% |
| Positive Likelihood Ratio | 1.18 | 0.72 to 1.95 |
| Negative Likelihood Ratio | 0.90 | 0.65 to 1.24 |
| Disease prevalence | 51.04% | 40.63% to 61.39% |
| Positive Predictive Value | 55.26% | 38.30% to 71.37% |
| Negative Predictive Value | 51.72% | 38.22% to 65.05% |

### Second Chest Radiologist, End Plates Definition, First Reading, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 43.59% | 27.82% to 60.38% |
| Specificity | 63.16% | 46.00% to 78.17% |
| Positive Likelihood Ratio | 1.18 | 0.68 to 2.05 |
| Negative Likelihood Ratio | 0.89 | 0.62 to 1.29 |
| Disease prevalence | 50.65% | 39.01% to 62.24% |
| Positive Predictive Value | 54.84% | 36.04% to 72.67% |
| Negative Predictive Value | 52.17% | 36.95% to 67.11% |

### Second Chest Radiologist, End Plates Shape (Fish Vertebra), First Reading, CXR Compared to DXA (All Locations)

```
     R3FQ2  Count          Lumbar_1  Count            Hip_1  Count
    normal     60    False Negative     31    False Negative     19
osteopenia     39    False Positive     13    False Positive     17
        N=     99     True Negative     29     True Negative     37
                      True Positive     26     True Positive     21
                                 N=     99                N=     94

     FemNeck_1  Count      Trochanter_1  Count
False Negative     26    False Negative     19
False Positive     14    False Positive     12
 True Negative     32     True Negative     25
 True Positive     23     True Positive     20
            N=     95                N=     76
```

### Second Chest Radiologist, End Plates Shape (Fish Vertebra), First Reading, CXR Compared to DXA (Lumbar)

| | | |
|---|---|---|
| Sensitivity | 45.61% | 32.36% to 59.34% |
| Specificity | 69.05% | 52.91% to 82.36% |
| Positive Likelihood Ratio | 1.47 | 0.86 to 2.51 |
| Negative Likelihood Ratio | 0.79 | 0.58 to 1.08 |
| Disease prevalence | 57.58% | 47.23% to 67.45% |
| Positive Predictive Value | 66.67% | 49.78% to 80.90% |
| Negative Predictive Value | 48.33% | 35.23% to 61.60% |

### Second Chest Radiologist, End Plates Shape (Fish Vertebra), First Reading, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 52.50% | 36.13% to 68.48% |
| Specificity | 68.52% | 54.45% to 80.47% |
| Positive Likelihood Ratio | 1.67 | 1.02 to 2.73 |
| Negative Likelihood Ratio | 0.69 | 0.48 to 1.01 |
| Disease prevalence | 42.55% | 32.41% to 53.18% |
| Positive Predictive Value | 55.26% | 38.30% to 71.37% |
| Negative Predictive Value | 66.07% | 52.19% to 78.18% |

### Second Chest Radiologist, End Plates Shape (Fish Vertebra), First Reading, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 46.94% | 32.54% to 61.72% |
| Specificity | 69.57% | 54.24% to 82.25% |
| Positive Likelihood Ratio | 1.54 | 0.91 to 2.62 |
| Negative Likelihood Ratio | 0.76 | 0.55 to 1.06 |
| Disease prevalence | 51.58% | 41.10% to 61.96% |
| Positive Predictive Value | 62.16% | 44.76% to 77.53% |
| Negative Predictive Value | 55.17% | 41.54% to 68.25% |

### Second Chest Radiologist, End Plates Shape (Fish Vertebra), First Reading, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 51.28% | 34.79% to 67.58% |
| Specificity | 67.57% | 50.21% to 81.97% |
| Positive Likelihood Ratio | 1.58 | 0.91 to 2.76 |
| Negative Likelihood Ratio | 0.72 | 0.49 to 1.07 |
| Disease prevalence | 51.32% | 39.57% to 62.96% |
| Positive Predictive Value | 62.50% | 43.70% to 78.88% |
| Negative Predictive Value | 56.82% | 41.04% to 71.64% |

**Second Chest Radiologist, Body Shape (Wedged), First Reading, CXR Compared to DXA (All Locations)**

```
            Count           Lumbar_2  Count             Hip_2  Count
    normal     80    False Negative     44    False Negative     29
osteopenia     20    False Positive      6    False Positive      9
        N=    100     True Negative     36     True Negative     46
                      True Positive     14     True Positive     11
                                 N=    100                N=     95

     FemNeck_2  Count       Trochanter_2  Count
False Negative     37    False Negative     27
False Positive      8    False Positive      6
 True Negative     39     True Negative     32
 True Positive     12     True Positive     12
            N=     96                N=     77
```

**Second Chest Radiologist, Body Shape (Wedged), First Reading, CXR Compared to DXA (Lumbar Spine)**

| | | |
|---|---|---|
| Sensitivity | 24.14% | 13.88% to 37.17% |
| Specificity | 85.71% | 71.45% to 94.54% |
| Positive Likelihood Ratio | 1.69 | 0.71 to 4.03 |
| Negative Likelihood Ratio | 0.89 | 0.73 to 1.07 |
| Disease prevalence | 58.00% | 47.71% to 67.80% |
| Positive Predictive Value | 70.00% | 45.73% to 88.03% |
| Negative Predictive Value | 45.00% | 33.85% to 56.53% |

**Second Chest Radiologist, Body Shape (Wedged), First Reading, CXR Compared to DXA (Hip)**

| | | |
|---|---|---|
| Sensitivity | 27.50% | 14.62% to 43.89% |
| Specificity | 83.64% | 71.19% to 92.22% |
| Positive Likelihood Ratio | 1.68 | 0.77 to 3.67 |
| Negative Likelihood Ratio | 0.87 | 0.69 to 1.08 |
| Disease prevalence | 42.11% | 32.04% to 52.67% |
| Positive Predictive Value | 55.00% | 31.55% to 76.90% |
| Negative Predictive Value | 61.33% | 49.38% to 72.36% |

**Second Chest Radiologist, Body Shape (Wedged), First Reading, CXR Compared to DXA (Femoral Neck)**

| | | |
|---|---|---|
| Sensitivity | 24.49% | 13.36% to 38.87% |
| Specificity | 82.98% | 69.18% to 92.33% |
| Positive Likelihood Ratio | 1.44 | 0.65 to 3.20 |
| Negative Likelihood Ratio | 0.91 | 0.74 to 1.12 |
| Disease prevalence | 51.04% | 40.63% to 61.39% |
| Positive Predictive Value | 60.00% | 36.07% to 80.83% |
| Negative Predictive Value | 51.32% | 39.57% to 62.96% |

**Second Chest Radiologist, Body Shape (Wedged), First Reading, CXR Compared to DXA (Trochanter)**

| | | |
|---|---|---|
| Sensitivity | 30.77% | 17.04% to 47.57% |
| Specificity | 84.21% | 68.74% to 93.94% |
| Positive Likelihood Ratio | 1.95 | 0.81 to 4.66 |
| Negative Likelihood Ratio | 0.82 | 0.64 to 1.06 |
| Disease prevalence | 50.65% | 39.01% to 62.24% |
| Positive Predictive Value | 66.67% | 41.01% to 86.58% |
| Negative Predictive Value | 54.24% | 40.76% to 67.28% |

### Second Chest Radiologist, Compression Fractures, First Reading, CXR Compared to DXA (All Locations)

```
           Count             Lumbar  Count              Hip  Count
    normal    93    False Negative     53    False Negative     36
osteopenia     6    False Positive      2    False Positive      1
        N=    99     True Negative     40     True Negative     54
                     True Positive      4     True Positive      3
                                N=     99                N=     94

      FemNeck  Count       Trochanter  Count
False Negative     44    False Negative     34
False Positive      1     True Negative     38
 True Negative     46     True Positive      4
 True Positive      4                N=     76
            N=     95
```

### Second Chest Radiologist, Compression Fractures, First Reading, CXR Compared to DXA (Lumbar Spine)

| | | |
|---|---|---|
| Sensitivity | 7.02% | 1.99% to 17.02% |
| Specificity | 95.24% | 83.80% to 99.28% |
| Positive Likelihood Ratio | 1.47 | 0.28 to 7.67 |
| Negative Likelihood Ratio | 0.98 | 0.88 to 1.08 |
| Disease prevalence | 57.58% | 47.23% to 67.45% |
| Positive Predictive Value | 66.67% | 22.68% to 94.67% |
| Negative Predictive Value | 43.01% | 32.79% to 53.69% |

### Second Chest Radiologist, Compression Fractures, First Reading, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 7.69% | 1.70% to 20.89% |
| Specificity | 98.18% | 90.24% to 99.70% |
| Positive Likelihood Ratio | 4.23 | 0.46 to 39.18 |
| Negative Likelihood Ratio | 0.94 | 0.85 to 1.04 |
| Disease prevalence | 41.49% | 31.42% to 52.12% |
| Positive Predictive Value | 75.00% | 20.34% to 95.88% |
| Negative Predictive Value | 60.00% | 49.13% to 70.19% |

### Second Chest Radiologist, Compression Fractures, First Reading, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 8.33% | 2.37% to 20.00% |
| Specificity | 97.87% | 88.66% to 99.64% |
| Positive Likelihood Ratio | 3.92 | 0.45 to 33.76 |
| Negative Likelihood Ratio | 0.94 | 0.85 to 1.03 |
| Disease prevalence | 50.53% | 40.07% to 60.95% |
| Positive Predictive Value | 80.00% | 28.81% to 96.70% |
| Negative Predictive Value | 51.11% | 40.35% to 61.80% |

### Second Chest Radiologist, Compression Fractures, First Reading, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 10.53% | 3.01% to 24.82% |
| Specificity | 100.00% | 90.66% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 0.89 | 0.80 to 1.00 |
| Disease prevalence | 50.00% | 38.30% to 61.70% |
| Positive Predictive Value | 100.00% | 40.23% to 100.00% |
| Negative Predictive Value | 52.78% | 40.65% to 64.67% |

### Second Chest Radiologist, Vertical Striations, First Reading, CXR Compared to DXA (All Locations)

| | Count |
|---|---|
| normal | 97 |
| osteopenia | 3 |
| N= | 100 |

| Lumbar | Count |
|---|---|
| False Negative | 56 |
| False Positive | 1 |
| True Negative | 41 |
| True Positive | 2 |
| N= | 100 |

| Hip | Count |
|---|---|
| False Negative | 37 |
| True Negative | 55 |
| True Positive | 3 |
| N= | 95 |

| FemNeck | Count |
|---|---|
| False Negative | 47 |
| False Positive | 1 |
| True Negative | 46 |
| True Positive | 2 |
| N= | 96 |

| Trochanter | Count |
|---|---|
| False Negative | 36 |
| True Negative | 38 |
| True Positive | 3 |
| N= | 77 |

### Second Chest Radiologist, Vertical Striations, First Reading, CXR Compared to DXA (Lumbar Spine)

| | | |
|---|---|---|
| Sensitivity | 3.45% | 0.52% to 11.93% |
| Specificity | 97.62% | 87.39% to 99.60% |
| Positive Likelihood Ratio | 1.45 | 0.14 to 15.45 |
| Negative Likelihood Ratio | 0.99 | 0.92 to 1.06 |
| Disease prevalence | 58.00% | 47.71% to 67.80% |
| Positive Predictive Value | 66.67% | 11.55% to 94.53% |
| Negative Predictive Value | 42.27% | 32.30% to 52.72% |

### Second Chest Radiologist, Vertical Striations, First Reading, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 7.50% | 1.66% to 20.41% |
| Specificity | 100.00% | 93.45% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 0.93 | 0.85 to 1.01 |
| Disease prevalence | 42.11% | 32.04% to 52.67% |
| Positive Predictive Value | 100.00% | 30.48% to 100.00% |
| Negative Predictive Value | 59.78% | 49.04% to 69.88% |

### Second Chest Radiologist, Vertical Striations, First Reading, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 4.08% | 0.62% to 14.01% |
| Specificity | 97.87% | 88.66% to 99.64% |
| Positive Likelihood Ratio | 1.92 | 0.18 to 20.46 |
| Negative Likelihood Ratio | 0.98 | 0.91 to 1.05 |
| Disease prevalence | 51.04% | 40.63% to 61.39% |
| Positive Predictive Value | 66.67% | 11.55% to 94.53% |
| Negative Predictive Value | 49.46% | 38.93% to 60.03% |

### Second Chest Radiologist, Vertical Striations, First Reading, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 7.69% | 1.70% to 20.89% |
| Specificity | 100.00% | 90.66% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 0.92 | 0.84 to 1.01 |
| Disease prevalence | 50.65% | 39.01% to 62.24% |
| Positive Predictive Value | 100.00% | 30.48% to 100.00% |
| Negative Predictive Value | 51.35% | 39.44% to 63.15% |

### Second Chest Radiologist, Empty Box, First Reading, CXR Compared to DXA (All Locations)

| | Count |
|---|---|
| normal | 73 |
| osteopenia | 6 |
| N= | 79 |

| Lumbar | Count |
|---|---|
| False Negative | 42 |
| True Negative | 31 |
| True Positive | 6 |
| N= | 79 |

| Hip | Count |
|---|---|
| False Negative | 30 |
| False Positive | 2 |
| True Negative | 40 |
| True Positive | 3 |
| N= | 75 |

| FemNeck | Count |
|---|---|
| False Negative | 34 |
| False Positive | 1 |
| True Negative | 36 |
| True Positive | 5 |
| N= | 76 |

| Trochanter | Count |
|---|---|
| False Negative | 28 |
| False Positive | 1 |
| True Negative | 30 |
| True Positive | 3 |
| N= | 62 |

### Second Chest Radiologist, Empty Box, First Reading, CXR Compared to DXA (Lumbar Spine)

| | | |
|---|---|---|
| Sensitivity | 12.50% | 4.76% to 25.26% |
| Specificity | 100.00% | 88.68% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 0.88 | 0.79 to 0.97 |
| Disease prevalence | 60.76% | 49.13% to 71.56% |
| Positive Predictive Value | 100.00% | 54.05% to 100.00% |
| Negative Predictive Value | 42.47% | 30.97% to 54.59% |

### Second Chest Radiologist, Empty Box, First Reading, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 9.09% | 2.02% to 24.36% |
| Specificity | 95.24% | 83.80% to 99.28% |
| Positive Likelihood Ratio | 1.91 | 0.34 to 10.77 |
| Negative Likelihood Ratio | 0.95 | 0.84 to 1.08 |
| Disease prevalence | 44.00% | 32.55% to 55.94% |
| Positive Predictive Value | 60.00% | 15.40% to 93.51% |
| Negative Predictive Value | 57.14% | 44.75% to 68.91% |

### Second Chest Radiologist, Empty Box, First Reading, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 12.82% | 4.34% to 27.44% |
| Specificity | 97.30% | 85.79% to 99.55% |
| Positive Likelihood Ratio | 4.74 | 0.58 to 38.71 |
| Negative Likelihood Ratio | 0.90 | 0.79 to 1.02 |
| Disease prevalence | 51.32% | 39.57% to 62.96% |
| Positive Predictive Value | 83.33% | 36.10% to 97.24% |
| Negative Predictive Value | 51.43% | 39.17% to 63.56% |

### Second Chest Radiologist, Empty Box, First Reading, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 9.68% | 2.15% to 25.78% |
| Specificity | 96.77% | 83.24% to 99.46% |
| Positive Likelihood Ratio | 3.00 | 0.33 to 27.29 |
| Negative Likelihood Ratio | 0.93 | 0.82 to 1.06 |
| Disease prevalence | 50.00% | 37.03% to 62.97% |
| Positive Predictive Value | 75.00% | 20.34% to 95.88% |
| Negative Predictive Value | 51.72% | 38.22% to 65.05% |

**Second Chest Radiologist, Presence of Osteopenia, First Reading, CXR Compared to DXA (All Locations)**

| | Count |
|---|---|
| normal | 71 |
| osteopenia | 29 |
| N= | 100 |

| Impression | Count |
|---|---|
| False Negative | 39 |
| False Positive | 1 |
| True Negative | 32 |
| True Positive | 28 |
| N= | 100 |

| Lumbar | Count |
|---|---|
| False Negative | 32 |
| False Positive | 3 |
| True Negative | 39 |
| True Positive | 26 |
| N= | 100 |

| Hip | Count |
|---|---|
| False Negative | 25 |
| False Positive | 13 |
| True Negative | 42 |
| True Positive | 15 |
| N= | 95 |

| FemNeck | Count |
|---|---|
| False Negative | 32 |
| False Positive | 11 |
| True Negative | 36 |
| True Positive | 17 |
| N= | 96 |

| Trochanter | Count |
|---|---|
| False Negative | 23 |
| False Positive | 8 |
| True Negative | 30 |
| True Positive | 16 |
| N= | 77 |

**Second Chest Radiologist, Presence of Osteopenia, First Reading, CXR Compared to DXA (Lumbar Spine)**

| | | |
|---|---|---|
| Sensitivity | 44.83% | 31.75% to 58.46% |
| Specificity | 92.86% | 80.49% to 98.42% |
| Positive Likelihood Ratio | 6.28 | 2.03 to 19.37 |
| Negative Likelihood Ratio | 0.59 | 0.46 to 0.76 |
| Disease prevalence | 58.00% | 47.71% to 67.80% |
| Positive Predictive Value | 89.66% | 72.62% to 97.69% |
| Negative Predictive Value | 54.93% | 42.66% to 66.77% |

**Second Chest Radiologist, Presence of Osteopenia, First Reading, CXR Compared to DXA (Hip)**

| | | |
|---|---|---|
| Sensitivity | 37.50% | 22.74% to 54.20% |
| Specificity | 76.36% | 62.98% to 86.76% |
| Positive Likelihood Ratio | 1.59 | 0.85 to 2.95 |
| Negative Likelihood Ratio | 0.82 | 0.62 to 1.08 |
| Disease prevalence | 42.11% | 32.04% to 52.67% |
| Positive Predictive Value | 53.57% | 33.88% to 72.47% |
| Negative Predictive Value | 62.69% | 50.01% to 74.20% |

**Second Chest Radiologist, Presence of Osteopenia, First Reading, CXR Compared to DXA (Femoral Neck)**

| | | |
|---|---|---|
| Sensitivity | 34.69% | 21.68% to 49.64% |
| Specificity | 76.60% | 61.97% to 87.68% |
| Positive Likelihood Ratio | 1.48 | 0.78 to 2.82 |
| Negative Likelihood Ratio | 0.85 | 0.66 to 1.10 |
| Disease prevalence | 51.04% | 40.63% to 61.39% |
| Positive Predictive Value | 60.71% | 40.58% to 78.47% |
| Negative Predictive Value | 52.94% | 40.45% to 65.17% |

**Second Chest Radiologist, Presence of Osteopenia, First Reading, CXR Compared to DXA (Trochanter)**

| | | |
|---|---|---|
| Sensitivity | 41.03% | 25.58% to 57.90% |
| Specificity | 78.95% | 62.67% to 90.42% |
| Positive Likelihood Ratio | 1.95 | 0.95 to 4.01 |
| Negative Likelihood Ratio | 0.75 | 0.55 to 1.02 |
| Disease prevalence | 50.65% | 39.01% to 62.24% |
| Positive Predictive Value | 66.67% | 44.68% to 84.33% |
| Negative Predictive Value | 56.60% | 42.28% to 70.16% |

### Second Chest Radiologist, Presence of Osteopenia, Second Reading, CXR Compared to DXA (All Locations)

| | Count |
|---|---|
| normal | 61 |
| osteopenia | 39 |
| N= | 100 |

| Impression | Count |
|---|---|
| False Negative | 34 |
| False Positive | 6 |
| True Negative | 27 |
| True Positive | 33 |
| N= | 100 |

| Lumbar | Count |
|---|---|
| False Negative | 29 |
| False Positive | 10 |
| True Negative | 32 |
| True Positive | 29 |
| N= | 100 |

| Hip | Count |
|---|---|
| False Negative | 21 |
| False Positive | 18 |
| True Negative | 37 |
| True Positive | 19 |
| N= | 95 |

| FemNeck | Count |
|---|---|
| False Negative | 26 |
| False Positive | 15 |
| True Negative | 32 |
| True Positive | 23 |
| N= | 96 |

| Trochanter | Count |
|---|---|
| False Negative | 18 |
| False Positive | 11 |
| True Negative | 27 |
| True Positive | 21 |
| N= | 77 |

### Second Chest Radiologist, Presence of Osteopenia, Second Reading, CXR Compared to DXA (Lumbar)

| | | |
|---|---|---|
| Sensitivity | 50.00% | 36.58% to 63.42% |
| Specificity | 76.19% | 60.55% to 87.93% |
| Positive Likelihood Ratio | 2.10 | 1.15 to 3.82 |
| Negative Likelihood Ratio | 0.66 | 0.48 to 0.89 |
| Disease prevalence | 58.00% | 47.71% to 67.80% |
| Positive Predictive Value | 74.36% | 57.87% to 86.94% |
| Negative Predictive Value | 52.46% | 39.27% to 65.40% |

### Second Chest Radiologist, Presence of Osteopenia, Second Reading, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 47.50% | 31.52% to 63.87% |
| Specificity | 67.27% | 53.29% to 79.31% |
| Positive Likelihood Ratio | 1.45 | 0.88 to 2.39 |
| Negative Likelihood Ratio | 0.78 | 0.55 to 1.10 |
| Disease prevalence | 42.11% | 32.04% to 52.67% |
| Positive Predictive Value | 51.35% | 34.41% to 68.07% |
| Negative Predictive Value | 63.79% | 50.12% to 76.00% |

### Second Chest Radiologist, Presence of Osteopenia, Second Reading, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 46.94% | 32.54% to 61.72% |
| Specificity | 68.09% | 52.88% to 80.90% |
| Positive Likelihood Ratio | 1.47 | 0.88 to 2.46 |
| Negative Likelihood Ratio | 0.78 | 0.56 to 1.08 |
| Disease prevalence | 51.04% | 40.63% to 61.39% |
| Positive Predictive Value | 60.53% | 43.39% to 75.95% |
| Negative Predictive Value | 55.17% | 41.54% to 68.25% |

### Second Chest Radiologist, Presence of Osteopenia, Second Reading, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 53.85% | 37.19% to 69.90% |
| Specificity | 71.05% | 54.09% to 84.56% |
| Positive Likelihood Ratio | 1.86 | 1.04 to 3.31 |
| Negative Likelihood Ratio | 0.65 | 0.44 to 0.96 |
| Disease prevalence | 50.65% | 39.01% to 62.24% |
| Positive Predictive Value | 65.62% | 46.81% to 81.41% |
| Negative Predictive Value | 60.00% | 44.33% to 74.29% |

### Non-Chest Radiologist, End Plates Definition, First Reading, CXR Compared to DXA (All Locations)

| | Count |
|---|---|
| normal | 30 |
| osteopenia | 7 |
| N= | 37 |

| Lumbar | Count |
|---|---|
| False Negative | 17 |
| False Positive | 3 |
| True Negative | 13 |
| True Positive | 4 |
| N= | 37 |

| Hip | Count |
|---|---|
| False Negative | 12 |
| False Positive | 3 |
| True Negative | 15 |
| True Positive | 3 |
| N= | 33 |

| FemNeck_1 | Count |
|---|---|
| False Negative | 14 |
| False Positive | 3 |
| True Negative | 14 |
| True Positive | 4 |
| N= | 35 |

| Trochanter_1 | Count |
|---|---|
| False Negative | 10 |
| False Positive | 3 |
| True Negative | 11 |
| True Positive | 3 |
| N= | 27 |

### Non-Chest Radiologist, End Plates Definition, First Reading, CXR Compared to DXA (Lumbar Spine)

| | | |
|---|---|---|
| Sensitivity | 19.05% | 5.56% to 41.92% |
| Specificity | 81.25% | 54.34% to 95.73% |
| Positive Likelihood Ratio | 1.02 | 0.26 to 3.91 |
| Negative Likelihood Ratio | 1.00 | 0.73 to 1.36 |
| Disease prevalence | 56.76% | 39.49% to 72.89% |
| Positive Predictive Value | 57.14% | 18.75% to 89.58% |
| Negative Predictive Value | 43.33% | 25.48% to 62.56% |

### Non-Chest Radiologist, End Plates Definition, First Reading, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 20.00% | 4.57% to 48.09% |
| Specificity | 83.33% | 58.56% to 96.23% |
| Positive Likelihood Ratio | 1.20 | 0.28 to 5.10 |
| Negative Likelihood Ratio | 0.96 | 0.69 to 1.33 |
| Disease prevalence | 45.45% | 28.12% to 63.64% |
| Positive Predictive Value | 50.00% | 12.42% to 87.58% |
| Negative Predictive Value | 55.56% | 35.34% to 74.50% |

### Non-Chest Radiologist, End Plates Definition, First Reading, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 22.22% | 6.55% to 47.64% |
| Specificity | 82.35% | 56.55% to 95.99% |
| Positive Likelihood Ratio | 1.26 | 0.33 to 4.82 |
| Negative Likelihood Ratio | 0.94 | 0.68 to 1.31 |
| Disease prevalence | 51.43% | 34.00% to 68.61% |
| Positive Predictive Value | 57.14% | 18.75% to 89.58% |
| Negative Predictive Value | 50.00% | 30.66% to 69.34% |

### Non-Chest Radiologist, End Plates Definition, First Reading, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 23.08% | 5.31% to 53.80% |
| Specificity | 78.57% | 49.21% to 95.09% |
| Positive Likelihood Ratio | 1.08 | 0.26 to 4.42 |
| Negative Likelihood Ratio | 0.98 | 0.65 to 1.47 |
| Disease prevalence | 48.15% | 28.68% to 68.04% |
| Positive Predictive Value | 50.00% | 12.42% to 87.58% |
| Negative Predictive Value | 52.38% | 29.81% to 74.25% |

### Non-Chest Radiologist, End Plates Shape (Fish Vertebra), First Reading, CXR Compared to DXA (All Locations)

```
            Count          Lumbar_2  Count           Hip_2  Count
    normal     37    False Negative     21    False Negative     15
osteopenia      1    False Positive      1    False Positive      1
        N=     38     True Negative     16     True Negative     18
                                N=     38                N=     34

     FemNeck_2  Count      Trochanter_2  Count
False Negative     18    False Negative     13
False Positive      1    False Positive      1
 True Negative     17     True Negative     14
            N=     36                N=     28
```

### Non-Chest Radiologist, End Plates Shape (Fish Vertebra), First Reading, CXR Compared to DXA (Lumbar Spine)

| | | |
|---|---|---|
| Sensitivity | 0.00% | 0.00% to 16.25% |
| Specificity | 94.12% | 71.24% to 99.02% |
| Positive Likelihood Ratio | 0.00 | |
| Negative Likelihood Ratio | 1.06 | 0.94 to 1.20 |
| Disease prevalence | 55.26% | 38.30% to 71.37% |
| Positive Predictive Value | 0.00% | 0.00% to 83.45% |
| Negative Predictive Value | 43.24% | 27.11% to 60.51% |

### Non-Chest Radiologist, End Plates Shape (Fish Vertebra), First Reading, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 0.00% | 0.00% to 21.97% |
| Specificity | 94.74% | 73.90% to 99.12% |
| Positive Likelihood Ratio | 0.00 | |
| Negative Likelihood Ratio | 1.06 | 0.95 to 1.17 |
| Disease prevalence | 44.12% | 27.20% to 62.11% |
| Positive Predictive Value | 0.00% | 0.00% to 83.45% |
| Negative Predictive Value | 54.55% | 36.36% to 71.88% |

### Non-Chest Radiologist, End Plates Shape (Fish Vertebra), First Reading, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 0.00% | 0.00% to 18.68% |
| Specificity | 94.44% | 72.63% to 99.07% |
| Positive Likelihood Ratio | 0.00 | |
| Negative Likelihood Ratio | 1.06 | 0.95 to 1.18 |
| Disease prevalence | 50.00% | 32.93% to 67.07% |
| Positive Predictive Value | 0.00% | 0.00% to 83.45% |
| Negative Predictive Value | 48.57% | 31.39% to 66.00% |

### Non-Chest Radiologist, End Plates Shape (Fish Vertebra), First Reading, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 0.00% | 0.00% to 24.88% |
| Specificity | 93.33% | 67.98% to 98.89% |
| Positive Likelihood Ratio | 0.00 | |
| Negative Likelihood Ratio | 1.07 | 0.94 to 1.23 |
| Disease prevalence | 46.43% | 27.53% to 66.12% |
| Positive Predictive Value | 0.00% | 0.00% to 83.45% |
| Negative Predictive Value | 51.85% | 31.96% to 71.32% |

### Non-Chest Radiologist, Body Shape (Wedged), First Reading, CXR Compared to DXA (All Locations)

```
            Count            Lumbar  Count               Hip  Count
    normal     34    False Negative     19    False Negative     13
osteopenia      4    False Positive      2    False Positive      2
        N=     38     True Negative     15     True Negative     17
                      True Positive      2     True Positive      2
                                 N=     38                N=     34

      FemNeck  Count       Trochanter  Count
False Negative     16    False Negative     10
False Positive      2     True Negative     15
 True Negative     16     True Positive      3
 True Positive      2                N=     28
            N=     36
```

### Non-Chest Radiologist, Body Shape (Wedged), First Reading, CXR Compared to DXA (Lumbar Spine)

| | | |
|---|---|---|
| Sensitivity | 9.52% | 1.45% to 30.42% |
| Specificity | 88.24% | 63.52% to 98.20% |
| Positive Likelihood Ratio | 0.81 | 0.13 to 5.16 |
| Negative Likelihood Ratio | 1.03 | 0.82 to 1.28 |
| Disease prevalence | 55.26% | 38.30% to 71.37% |
| Positive Predictive Value | 50.00% | 8.30% to 91.70% |
| Negative Predictive Value | 44.12% | 27.20% to 62.11% |

### Non-Chest Radiologist, Body Shape (Wedged), First Reading, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 13.33% | 2.05% to 40.49% |
| Specificity | 89.47% | 66.82% to 98.39% |
| Positive Likelihood Ratio | 1.27 | 0.20 to 7.97 |
| Negative Likelihood Ratio | 0.97 | 0.75 to 1.25 |
| Disease prevalence | 44.12% | 27.20% to 62.11% |
| Positive Predictive Value | 50.00% | 8.30% to 91.70% |
| Negative Predictive Value | 56.67% | 37.44% to 74.52% |

### Non-Chest Radiologist, Body Shape (Wedged), First Reading, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 11.11% | 1.70% to 34.75% |
| Specificity | 88.89% | 65.25% to 98.30% |
| Positive Likelihood Ratio | 1.00 | 0.16 to 6.35 |
| Negative Likelihood Ratio | 1.00 | 0.79 to 1.26 |
| Disease prevalence | 50.00% | 32.93% to 67.07% |
| Positive Predictive Value | 50.00% | 8.30% to 91.70% |
| Negative Predictive Value | 50.00% | 31.90% to 68.10% |

### Non-Chest Radiologist, Body Shape (Wedged), First Reading, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 23.08% | 5.31% to 53.80% |
| Specificity | 100.00% | 78.03% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 0.77 | 0.57 to 1.04 |
| Disease prevalence | 46.43% | 27.53% to 66.12% |
| Positive Predictive Value | 100.00% | 30.48% to 100.00% |
| Negative Predictive Value | 60.00% | 38.68% to 78.84% |

### Non-Chest Radiologist, Compression Fractures, First Reading, CXR Compared to DXA (All Locations)

```
         Count          Lumbar_1  Count            Hip_1  Count
normal     37    False Negative     21    False Negative     14
    N=     37     True Negative     16     True Negative     19
                             N=     37                N=     33

     FemNeck_1  Count      Trochanter_1  Count
False Negative     17    False Negative     12
 True Negative     18     True Negative     15
            N=     35                N=     27
```

### Non-Chest Radiologist, Compression Fractures, First Reading, CXR Compared to DXA (Lumbar Spine)

| | | |
|---|---|---|
| Sensitivity | 0.00% | 0.00% to 16.25% |
| Specificity | 100.00% | 79.24% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 1.00 | 1.00 to 1.00 |
| Disease prevalence | 56.76% | 39.49% to 72.89% |
| Positive Predictive Value | | |
| Negative Predictive Value | 43.24% | 27.11% to 60.51% |

### Non-Chest Radiologist, Compression Fractures, First Reading, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 0.00% | 0.00% to 23.34% |
| Specificity | 100.00% | 82.20% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 1.00 | 1.00 to 1.00 |
| Disease prevalence | 42.42% | 25.49% to 60.78% |
| Positive Predictive Value | | |
| Negative Predictive Value | 57.58% | 39.22% to 74.51% |

### Non-Chest Radiologist, Compression Fractures, First Reading, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 0.00% | 0.00% to 19.67% |
| Specificity | 100.00% | 81.32% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 1.00 | 1.00 to 1.00 |
| Disease prevalence | 48.57% | 31.39% to 66.00% |
| Positive Predictive Value | | |
| Negative Predictive Value | 51.43% | 34.00% to 68.61% |

### Non-Chest Radiologist, Compression Fractures, First Reading, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 0.00% | 0.00% to 26.65% |
| Specificity | 100.00% | 78.03% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 1.00 | 1.00 to 1.00 |
| Disease prevalence | 44.44% | 25.50% to 64.66% |
| Positive Predictive Value | | |
| Negative Predictive Value | 55.56% | 35.34% to 74.50% |

### Non-Chest Radiologist, Vertical Striations, First Reading, CXR Compared to DXA (All Locations)

```
            Count          Lumbar_2  Count             Hip_2  Count
    normal     33    False Negative     19    False Negative     13
osteopenia      5    False Positive      3    False Positive      2
        N=     38     True Negative     14     True Negative     17
                      True Positive      2     True Positive      2
                                 N=     38                N=     34

     FemNeck_2  Count      Trochanter_2  Count
False Negative     15    False Negative     11
False Positive      2    False Positive      2
 True Negative     16     True Negative     13
 True Positive      3     True Positive      2
            N=     36                N=     28
```

### Non-Chest Radiologist, Vertical Striations, First Reading, CXR Compared to DXA (Lumbar Spine)

| | | |
|---|---|---|
| Sensitivity | 9.52% | 1.45% to 30.42% |
| Specificity | 82.35% | 56.55% to 95.99% |
| Positive Likelihood Ratio | 0.54 | 0.10 to 2.87 |
| Negative Likelihood Ratio | 1.10 | 0.85 to 1.43 |
| Disease prevalence | 55.26% | 38.30% to 71.37% |
| Positive Predictive Value | 40.00% | 6.49% to 84.60% |
| Negative Predictive Value | 42.42% | 25.49% to 60.78% |

### Non-Chest Radiologist, Vertical Striations, First Reading, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 13.33% | 2.05% to 40.49% |
| Specificity | 89.47% | 66.82% to 98.39% |
| Positive Likelihood Ratio | 1.27 | 0.20 to 7.97 |
| Negative Likelihood Ratio | 0.97 | 0.75 to 1.25 |
| Disease prevalence | 44.12% | 27.20% to 62.11% |
| Positive Predictive Value | 50.00% | 8.30% to 91.70% |
| Negative Predictive Value | 56.67% | 37.44% to 74.52% |

### Non-Chest Radiologist, Vertical Striations, First Reading, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 16.67% | 3.77% to 41.44% |
| Specificity | 88.89% | 65.25% to 98.30% |
| Positive Likelihood Ratio | 1.50 | 0.28 to 7.93 |
| Negative Likelihood Ratio | 0.94 | 0.72 to 1.22 |
| Disease prevalence | 50.00% | 32.93% to 67.07% |
| Positive Predictive Value | 60.00% | 15.40% to 93.51% |
| Negative Predictive Value | 51.61% | 33.07% to 69.83% |

### Non-Chest Radiologist, Vertical Striations, First Reading, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 15.38% | 2.37% to 45.46% |
| Specificity | 86.67% | 59.51% to 97.95% |
| Positive Likelihood Ratio | 1.15 | 0.19 to 7.08 |
| Negative Likelihood Ratio | 0.98 | 0.72 to 1.32 |
| Disease prevalence | 46.43% | 27.53% to 66.12% |
| Positive Predictive Value | 50.00% | 8.30% to 91.70% |
| Negative Predictive Value | 54.17% | 32.84% to 74.42% |

### Non-Chest Radiologist, Empty Box, First Reading, CXR Compared to DXA (All Locations)

```
        Count           Lumbar  Count              Hip  Count
normal     38    False Negative    21    False Negative    15
    N=     38     True Negative    17     True Negative    19
                             N=    38                N=    34

       FemNeck  Count        Trochanter  Count
False Negative     18    False Negative     13
 True Negative     18     True Negative     15
            N=     36                N=     28
```

### Non-Chest Radiologist, Empty Box, First Reading, CXR Compared to DXA (Lumbar)

| | | |
|---|---|---|
| Sensitivity | 0.00% | 0.00% to 16.25% |
| Specificity | 100.00% | 80.33% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 1.00 | 1.00 to 1.00 |
| Disease prevalence | 55.26% | 38.30% to 71.37% |
| Positive Predictive Value | | |
| Negative Predictive Value | 44.74% | 28.63% to 61.70% |

### Non-Chest Radiologist, Empty Box, First Reading, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 0.00% | 0.00% to 21.97% |
| Specificity | 100.00% | 82.20% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 1.00 | 1.00 to 1.00 |
| Disease prevalence | 44.12% | 27.20% to 62.11% |
| Positive Predictive Value | | |
| Negative Predictive Value | 55.88% | 37.89% to 72.80% |

### Non-Chest Radiologist, Empty Box, First Reading, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 0.00% | 0.00% to 18.68% |
| Specificity | 100.00% | 81.32% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 1.00 | 1.00 to 1.00 |
| Disease prevalence | 50.00% | 32.93% to 67.07% |
| Positive Predictive Value | | |
| Negative Predictive Value | 50.00% | 32.93% to 67.07% |

### Non-Chest Radiologist, Empty Box, First Reading, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 0.00% | 0.00% to 24.88% |
| Specificity | 100.00% | 78.03% to 100.00% |
| Positive Likelihood Ratio | | |
| Negative Likelihood Ratio | 1.00 | 1.00 to 1.00 |
| Disease prevalence | 46.43% | 27.53% to 66.12% |
| Positive Predictive Value | | |
| Negative Predictive Value | 53.57% | 33.88% to 72.47% |

### Non-Chest Radiologist, Presence of Osteopenia, First Reading, CXR Compared to DXA (All Locations)

| | Count |
|---|---|
| normal | 32 |
| osteopenia | 6 |
| N= | 38 |

| Impression | Count |
|---|---|
| False Negative | 21 |
| False Positive | 2 |
| True Negative | 11 |
| True Positive | 4 |
| N= | 38 |

| Lumbar | Count |
|---|---|
| False Negative | 18 |
| False Positive | 3 |
| True Negative | 14 |
| True Positive | 3 |
| N= | 38 |

| Hip | Count |
|---|---|
| False Negative | 12 |
| False Positive | 2 |
| True Negative | 17 |
| True Positive | 3 |
| N= | 34 |

| FemNeck | Count |
|---|---|
| False Negative | 15 |
| False Positive | 3 |
| True Negative | 15 |
| True Positive | 3 |
| N= | 36 |

| Trochanter | Count |
|---|---|
| False Negative | 10 |
| False Positive | 1 |
| True Negative | 14 |
| True Positive | 3 |
| N= | 28 |

### Non-Chest Radiologist, Presence of Osteopenia, First Reading, CXR Compared to DXA (Lumbar Spine)

| | | |
|---|---|---|
| Sensitivity | 14.29% | 3.22% to 36.37% |
| Specificity | 82.35% | 56.55% to 95.99% |
| Positive Likelihood Ratio | 0.81 | 0.19 to 3.51 |
| Negative Likelihood Ratio | 1.04 | 0.79 to 1.38 |
| Disease prevalence | 55.26% | 38.30% to 71.37% |
| Positive Predictive Value | 50.00% | 12.42% to 87.58% |
| Negative Predictive Value | 43.75% | 26.38% to 62.33% |

### Non-Chest Radiologist, Presence of Osteopenia, First Reading, CXR Compared to DXA (Hip)

| | | |
|---|---|---|
| Sensitivity | 20.00% | 4.57% to 48.09% |
| Specificity | 89.47% | 66.82% to 98.39% |
| Positive Likelihood Ratio | 1.90 | 0.36 to 9.95 |
| Negative Likelihood Ratio | 0.89 | 0.66 to 1.20 |
| Disease prevalence | 44.12% | 27.20% to 62.11% |
| Positive Predictive Value | 60.00% | 15.40% to 93.51% |
| Negative Predictive Value | 58.62% | 38.94% to 76.46% |

### Non-Chest Radiologist, Presence of Osteopenia, First Reading, CXR Compared to DXA (Femoral Neck)

| | | |
|---|---|---|
| Sensitivity | 16.67% | 3.77% to 41.44% |
| Specificity | 83.33% | 58.56% to 96.23% |
| Positive Likelihood Ratio | 1.00 | 0.23 to 4.31 |
| Negative Likelihood Ratio | 1.00 | 0.75 to 1.34 |
| Disease prevalence | 50.00% | 32.93% to 67.07% |
| Positive Predictive Value | 50.00% | 12.42% to 87.58% |
| Negative Predictive Value | 50.00% | 31.31% to 68.69% |

### Non-Chest Radiologist, Presence of Osteopenia, First Reading, CXR Compared to DXA (Trochanter)

| | | |
|---|---|---|
| Sensitivity | 23.08% | 5.31% to 53.80% |
| Specificity | 93.33% | 67.98% to 98.89% |
| Positive Likelihood Ratio | 3.46 | 0.41 to 29.36 |
| Negative Likelihood Ratio | 0.82 | 0.59 to 1.14 |
| Disease prevalence | 46.43% | 27.53% to 66.12% |
| Positive Predictive Value | 75.00% | 20.34% to 95.88% |
| Negative Predictive Value | 58.33% | 36.66% to 77.86% |