CHAPTER V

5. DISCUSSION

5.1 Discussion of important finding

The objective of this present study is to determine the diagnostic performance of Quantitative Ultrasound (QUS) calcaneus measurement in the case finding of osteoporosis in postmenopausal women using Dual energy X-ray absorptiometry (DXA) as gold standard.

The interpretation of diagnostic test result depends on the ability of the test to distinguish disease from non-diseased subjects. When used appropriately, diagnostic tests can be of great assistance to the clinician. Tests can be helpful for **diagnosis**, i.e., to help establish or exclude the presence of disease in symptomatic persons. Besides, tests can also be helpful for **screening**, i.e., to identify risk factors for disease and to detect occult disease in asymptomatic persons. Identification of risk factors may allow early intervention to prevent disease occurrence, and early detection of occult disease may reduce disease morbidity and mortality through early treatment[35,36].

To assess the diagnostic performance we use the accuracy of the test which is its correspondence with the true value and has two separate components[37] (sensitivity and specificity). The sensitivity of the test is the proportion of correctly identified disease persons and specificity is proportion of correctly identified persons without the disease. In this study we found that QUS had a low sensitivity (39.25%) for predicting BMD-defined osteoporosis, but had a high specificity (91.71%).

In order to determine test sensitivity and specificity for osteoporosis, QUS measurement was compared against a "gold standard" (DXA) using the WHO criteria for osteoporosis: a value of BMD less than -2.5 SD below the average value of the peak bone mass of healthy adults[33]. By applying these concepts into practice, a number of problems have arisen. Firstly, young adults, used to calculate mean values and standard deviations which may or may not include populations that are randomly selected, were giving biased results. Also, reference data may exclude individuals with risk factors for bone disease. This will artifactually increase the mean value and reduce the standard deviation used to

compute threshold value[23]. And other problem is that normal ranges for DXA are available from many European countries where difference in mean BMD and the standard deviations used are relatively small[23].

Using a Western BMD reference, which is usually available with the bone densitometer machine, can result in misleading prevalence of osteoporosis in populations of Asian countries. This is most likely because Western women have a higher peak bone mass[28], and also because of the fact that cessation of menstruation in the Thai population in this study occurred at average age of 46±5 years before it occurs in North America.

Unlike sensitivity and specificity, positive and negative predictive values are not true indexes of validity because they depend on the relative proportions of diseased and disease-free persons being tested. In this study positive and negative predictive value (72.41% and 73.14%) were again relatively low for QUS as a predictor of BMD-defined osteoporosis. This might be because of a test with high specificity (few false positive among the disease-free) could have low positive predictive value if the ratio of disease-free to diseased subjects was high[36,37].

The likelihood ratio for a particular value of a diagnostic test is defined as the probability of that test result in people with disease divided by the probability of the result in people without disease. Likelihood expresses how many times more (or less) likely a test result is to be found in disease, compared with non diseased people[38].

In this study the positive likelihood ratio is 4.73 which means that an individual with disease is 4.73 times more likely to occur in than in one without it. A high likelihood ratio for a positive result has shown that the test provides useful information, as does a likelihood ratio close to zero for a negative result.

The major statistical attributes of any measurement, including the results of diagnostic tests, are reproducibility and validity. Reproducibility or precision refers to the ability of a test to yield the same results on retesting[39].

Test precision is a measure of a test's reproducibility when repeated on the same sample. An imprecise test is the one that yields widely varying results on repeated measurements (Figure 5.1.).

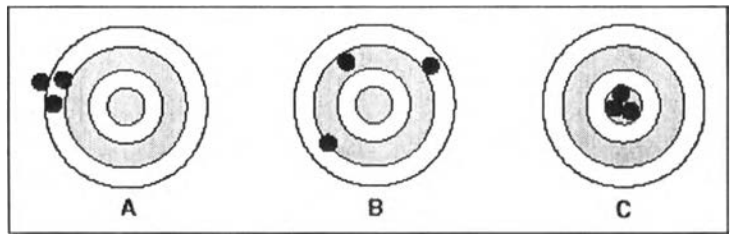

Figure 5.1 Relationship between accuracy and precision in diagnostic tests. The center of the target represents the true test result. Figure (A) represents a diagnostic test which is precise but inaccurate; on repeated measurement, the test yields very similar results, but all results are far from the true value. Figure (B) shows a test which is imprecise and inaccurate; repeated measurement yields widely different results, and the results are far from the true value. Figure (C) shows an ideal test, one that is both precise and accurate.

This study uses intraclass correlation coefficient (ICC) which is a reliability coefficient reflecting both degree of correspondence and agreement among rating[38]. In this study, we found that the ICC is .976 representing strong reliability to ensure reasonable validity.

There is also the question of whether ultrasound measurement should use the same criteria as bone density measurements to define bone status. Because there are fundamental differences between QUS and DXA regarding what bone properties each technique actually measures, it may not be appropriate to apply the same criteria used with BMD when using QUS to diagnose osteoporosis. Ultrasound should have its own unique way to contribute to evaluation of bone status. It is hoped that as the results of prospective clinical studies become available to provide a more comprehensive understanding of QUS, specific unique clinical criteria can be established for ultrasound.

In this study, I have attempted to use the WHO definition of osteoporosis (BMD of less than -2.5 SD from young normal women) to evaluate whether QUS measurements can give the same discrimination as DXA. Apparently, few patients will have T-scores below -2.5, while most postmenopausal women will fall below this level with lateral DXA. This means that when using QUS measurement by WHO definition osteoporosis, the model of QUS may be accepted particularly the risk of false negatives. And when I used stiffness index of the instrument to find the optimal cut-off value from ROC curve, I got the higher sensitivity (from 39.25% to 77%).

A receiver operating characteristic (ROC) curve, which relates false-positive results to true-positive results at multiple cut-off points, can be constructed to help determining

the cut-off point that gives the optimum combination of sensitivity and specificity. If two tests are being compared, their results should be interpreted independently against the same gold standard. In this situation, ROC curves are used to determine which test is "better", but comparison of likelihood ratios at various levels of test results is a more powerful method[37,38,39]. The results of the ROC analysis presented in Figure 4.3. suggest that QUS measurements are poor at predicting osteoporosis in women as defined by the results of DXA measurements of femoral neck.

Sometimes, we wish to make a diagnosis on the basis of a continuous measurement. Often, there is no threshold above (or below) which disease definitely occurs. In these situations, we need to define a cut-off value ourselves, above (or below) which we believe that an individual has a very high chance of having the disease. These provide a way of assessing whether a particular type of test provides useful information, and can be used to compare two different tests, and to select the optimal cut-off value for a test. For a given test, we considered all cut-off points that give a unique pair of values for sensitivity and specificity, and plotted the sensitivity against 1 minus the specificity (thus comparing the probabilities of a positive test result in those with and without disease) and connected these points by lines[38].

By univariate analysis, it is found that being older, having longer duration from the post-menopausal period, and having lower BMI would associate with an increase risk of having osteoporosis. I also recategorized these variables to make it more statistically meaningful and used logistic regression to determine how many levels would be appropriate. As a result, I found that these variables can be appropriately categorized only in 2 categories to increase power of analysis.

It has recently been suggested that combining QUS measurements with information regarding clinical risk factors, such as previous fracture, maternal history of hip fracture, or low body weight, may provide a strategy for using QUS effectively in clinical practice[23,40]. In this study only age and BMI were associated with osteoporosis. The statistically significant of association between postmenopausal period and osteoporosis was a confounding effect by age and BMI confirmed by multiple logistic regressions (see appendix A.). Theoretically, both age and BMI should be taken into classifying subjects into groups. Due to small number of subjects who have high BMI, it is, however, more

practical to divide samples into only 2 groups according to age (<65, ≥65). This approach to use QUS alone would be more cost-effective than combining QUS and DXA measurements, as well as, another suggested strategy of using QUS to identify high-risk individuals who would then have additional assessment by DXA. The use of BMD combined with knowledge of clinical risk factors to assess fracture risk has been adopted in clinical guidelines on the prevention and treatment of osteoporosis[40,41].

5.2 Limitation of the study

This study has limitations especially from possible biases. Bias means: Are the diagnosis and test result determined independently from one another? Bias can affect both sensitivity and specificity. Biases in the assessment of diagnostic test can derive from subject selection and from methodology[38,39]. The former is from variation in making diagnosis, especially, when no definite criteria exist and the value of the test may depend strongly on the population on which it is used. The latter is mainly from different positive criteria we used or even the level of certainty the gold standard can be established. Apart from that, information bias from measures can also play a part. The results of the study, therefore, should be interpreted with caution.

5.3 Clinical implications.

The model of QUS as a "prescreening" modality may be acceptable assuming adequate education of clinicians and patients of its limitations, particularly the risk of false negatives. Suitable guidelines are needed to be established so that QUS can be used effectively in a clinical setting without the need for referring a large proportion of women for additional testing. However, it must be noted that at present, there is a lack of agreement on whether QUS can be used to monitor disease progression or treatment efficacy; therefore patients may still need to be monitored using DXA or other established bone densitometry techniques.

The goal of early detection is to find patients who have osteoporosis and offer them effective treatment to reduce their risk of fractures, which means making measurement of bone mass a routine part of preventive care for appropriate patients[40,41]. Better evidence is needed about the potential value of monitoring bone density, and better guidelines about testing frequency that take into account the bone density, treatment, age, and clinical conditions of each patient are needed[23].

The recommended routine screening for osteoporosis beginning at age 65 years. There is no recommendation about screening for women younger than 65 years. Therefore, if we use optimal cut-off from stiffness index and categorize age group of patient for diagnose osteoporosis in this population; QUS may be a suitable osteoporosis diagnostic test.