Inter-Observable Variations When Using Popular Methods to Obtain Cortical Index and Mean Combined Cortical Thickness in Proximal Humerus Radiographs Can Result in Highly Variable Correlations with Fracture Strength

Tanner D. Langston, Chad S. Mears, Colton M. Phippen, S. Taylor Brady, John G. Skedros

University of Utah, Salt Lake City, Utah

Disclosures: No authors have any disclosures.

INTRODUCTION: Routine clinical radiographs are being used to predict bone quality of the proximal humerus [1-8]. Cortical index (CI) is the most common measurement and is obtained from standard anterior-posterior (AP) radiographs. CI is the difference between the outer (OD) and inner diameters (ID) divided by the OD [(OD-ID)/OD] (lower CI values = weaker bone). Mean combined cortical thickness (MCCT = OD-ID) also strongly correlates with bone quality and fracture load [1,2,5,7,8]. These CI and MCCT measurements are clinically relevant: (1) surgeons can evaluate radiographs of the fractured humerus and of the non-fractured side and use these measurements for pre-operative planning [9], (2) age-related changes in these simple measurements correlate with reduced bone quality and fracture strength [1,2], and (3) age-related changes in these measurements can result in complications of shoulder arthroplasty and fracture fixation [1]. However, various methods are used to obtain CI and MCCT. The three methods used in the present study are described below. The first method is the most popular and is that of Tingart et al. (2003) [2] (Fig. 1A). In the “Tingart method” CI and MCCT measurements are made at two levels that are 20 mm apart in the proximal diaphysis where the endosteal cortical borders are parallel. The generally reduced r values of CI and MCCT vs. UFL of the Tingart and Mather methods (when compared to our “D” levels) (see r values in the Table) can be attributed to the increased observer variations at each level of these two methods. The CI and MCCT data obtained using our method also showed much less variation at each of the D levels. Analysis of impact of inter-observer variation on correlation coefficients (Table): When applying our method, there were five of eight instances where the correlations of CI with UFL were significant (p<0.05). By contrast the Tingart and Mather methods had two (of eight) instances where the correlation of CI with UFL were significant. The results dealing with energy absorbed to fracture (N-m) can also be examined in the Table. With respect to correlations between MCCT and UFL obtained from data using our method, all of our correlations were significant. By contrast, one of the correlations was not significant when using the Tingart and Mather methods (but the correlations from the latter methods tended to be lower by 0.1 to 0.2 when compared to those obtained using from our method).

RESULTS: Magnitude of inter-observer variation: When using our new method, the average difference was 6.7 mm between the proximal and distal measurements (i.e., the range of the inter-observer difference in establishing each of the D1, D2, D3 and D4 levels. By contrast, the average difference was 31.6 mm between the most proximal and distal levels made when using the Tingart method and 26.9 mm when using the Mather method. The generally reduced r values of CI and MCCT vs. UFL of the Tingart and Mather methods (when compared to our “D” levels) (see r values in the Table) can be attributed to the increased observer variations at each level of these two methods. The CI and MCCT data obtained using our method also showed much less variation at each of the D levels.

DISCUSSION: These results show that interpretations of the relationships of bone strength with radiographic morphometry can dramatically change as a result of variations that are inherent in the Tingart and Mather methods. These variations stem from the fact that these methods are based on the determining parallelism along a 20 mm vertical distance of the proximal diaphysis (based either on the endosteal margins or periosteal margins of the cortex). We developed our novel proximal-referencing method because of the poor reliability that we had in establishing the proximal and distal levels of the 20 mm distance that is mandated by the Tingart and Mather methods (our unpublished data). Another important finding of the present study is that the correlations with fracture load data: (1) were typically reduced by r = 0.1 to 0.2 when the data are obtained using the Tingart and Mather methods, and (2) correlations from our method were much stronger with respect to UFL vs. CI.

SIGNIFICANCE: The Tingart et al. method can incur high inter-observer errors that can adversely influence the interpretation of fracture strength.