Inter-Observer 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[1], Chad S. Mears[1], Colton M. Phippen[1], S. Taylor Brady[1], John G. Skedros[1]
[1]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 margins are parallel. Mather et al. (2013) [7] ("Mather method") made the two measurements (also 20 mm apart) in a location where the *outer* cortical margins are parallel. In view of our data showing high variation in the locations ("levels") of these two methods, we devised the third method ("Skedros method") which is based on vertical distances from a transverse tangent of a circle that is fit to the articular surface of the humeral head. This study examines how inter-observer variability of CI and MCCT measurements made using these three methods affects the strengths of correlations between these morphological measurements and ultimate fracture load (UFL) and energy absorbed to fracture, which were obtained by fracturing the same bones [1].

METHODS: Digitized true AP radiographs of 33 proximal humeri were obtained from a prior study (average age 59 years old; 15 males and 18 females) [1]. The bones ranged from robust cortices (i.e., "good quality") to osteoporotic (i.e., "poor quality") as established by radiographic inspection and DXA analysis [1]. The soft tissues were manually removed from the bones prior to being radiographed. Five trained observers analyzed the digitized radiographs and made CI and MCCT measurements in accordance with methods of: (1)Tingart et al. [2], (2) Mather et al. [7], and (3) Skedros et al. [1].

Tingart method: "The lateral and medial cortical thickness of the proximal humeral diaphysis was measured at two different levels. Level 1 was the most proximal level of the humeral diaphysis where the endosteal borders of the lateral and medial cortices were parallel to each other. Level 2 was 20 mm distal to level 1."

Mather method: "The first level was the most proximal point on the humerus where the outer medial and lateral cortical borders become parallel, as previously described [by Tingart et al. (2003)], [Note that this is an error; Tingart et al. used the endosteal (inner) cortical borders]. A perpendicular line was drawn from the medial outer cortex of the humerus to the lateral outer cortex of the humerus and measured with a digital caliper to provide the thickness of the entire bone (M1). At the same level, a measurement of the width of the intramedullary canal was obtained (M2)." **New method (Skedros method):** Our new employs a circle that is best fit to the curvature of the articular surface of the humeral head on AP radiographs (Fig. 1B). The circle was digitally placed and all subsequent measurements were made using the ImageJ program. The lower-most edge of the circle was defined as the proximal reference level (or "M1"; M = metaphysis = surgical neck region). Then seven successive "levels" (locations) were defined, and each was below (distal to) the M1 level, and each are separated by 10 mm. **Fracture data:** The fracture data were obtained from the same [1]. The bones were fractured in a simulated fall (Fig. 2). **Statistical analysis:** Pearson correlation coefficients (r values) were used and were stratified in accordance with Hinkle et al. (1979) [10].

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. **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). Nevertheless, there were no significant differences between r values representing the inter-observer range (i.e., proximal vs. distal data) within each method. But that there were instances where the r value stratification changed as a result of the inter-observer variation (e.g., one observer's data resulted in a "low correlation" vs. another observer's data resulted in a "moderate correlation"). For example, using our method there were three of four of our paired comparisons (proximal compared to distal) of MCCT vs. UFL where the r value changed from a "moderate" to a "low" correlation, and two of four comparisons of CI vs. UFL where the r values changed from "low" to "little if any" correlation. Using the Tingart and Mather MCCT vs. UFL comparisons none of the proximal vs. distal r-value stratifications changed in this way, but it is important to point out that *none* of these r values where greater than a "low" correlation (i.e., all r values were ≤ 0.3). Using Tingart and Mather CI vs. UFL data one of the r values changed from "low" to "little if any" correlation.

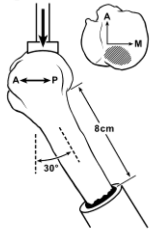
DISCUSSION: These results show that interpretations of the relationships of bone strength with radiographic morphometry can dramatically change as a result of variations that are inherent in the Tingart and Mather methods. These variations stem from the fact that these methods are based on the determining 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.

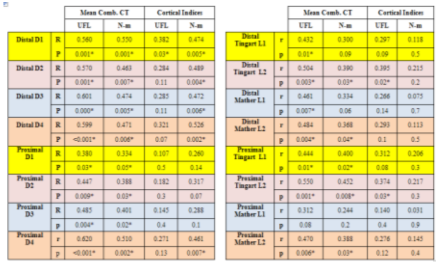
SIGNIFICANCE: The Tingart et al. method can incur high inter-observer errors that can adversely influence the interpretation of fracture strength.

REFERENCES: [1] Skedros et al. 2015 JOR (in press); [2] Tingart et al. 2003 JBJS 85:611-; [3] Hepp et al. 2009 Arch Orthop Trauma Surg 129; [4] Osterhoff et al. 2012 Arch Orthop Trauma Surg 132; [5] Namdari et al. 2012 JSES 21; [6] Giannotti et al. 2012 Clin Cases Miner Bone Met 9; [7] Mather et al. 2013 JSES 22; [8] Skedros et al. 2014 Biores Open Access 3; [9] Nho et al. 2007 JBJS 89; [10] Hinkle et al. 1979 *Applied Stat. for the Behavioral Sciences,* Rand McNally.

Figure 1A **Figure 1B**

Figure 2





		Mean Comb. CT		Cortical Indices	
		UFL	N-m	UFL	N-m
Distal D1	R	0.560	0.550	0.382	0.474
	P	0.001*	0.001*	0.03*	0.005*
Distal D2	R	0.570	0.463	0.284	0.489
	P	0.001*	0.007*	0.11	0.004*
Distal D3	R	0.601	0.474	0.285	0.472
	P	0.000*	0.005*	0.11	0.006*
Distal D4	R	0.599	0.471	0.321	0.526
	P	<0.001*	0.006*	0.07	0.002*
Proximal D1	R	0.380	0.334	0.107	0.260
	P	0.03*	0.05*	0.5	0.14
Proximal D2	R	0.447	0.388	0.182	0.317
	P	0.009*	0.03*	0.3	0.07
Proximal D3	R	0.485	0.401	0.145	0.288
	p	0.004*	0.02*	0.4	0.1
Proximal D4	r	0.620	0.510	0.271	0.461
	p	<0.001*	0.002*	0.13	0.007*

		Mean Comb. CT		Cortical Indices	
		UFL	N-m	UFL	N-m
Distal Tingart L1	r	0.432	0.300	0.297	0.118
	p	0.01*	0.09	0.09	0.5
Distal Tingart L2	r	0.504	0.390	0.395	0.215
	p	0.003*	0.03*	0.02*	0.2
Distal Mather L1	r	0.461	0.334	0.266	0.075
	p	0.007*	0.06	0.14	0.7
Distal Mather L2	r	0.484	0.368	0.293	0.113
	p	0.004*	0.04*	0.1	0.5
Proximal Tingart L1	r	0.444	0.400	0.312	0.206
	p	0.01*	0.02*	0.08	0.3
Proximal Tingart L2	r	0.550	0.452	0.374	0.217
	p	0.001*	0.008*	0.03*	0.3
Proximal Mather L1	r	0.312	0.244	0.140	0.031
	p	0.08	0.2	0.4	0.9
Proximal Mather L2	r	0.470	0.388	0.276	0.145
	p	0.006*	0.03*	0.12	0.4

Figure 1A: AP radiographs of proximal humeri from Tingart et al. [2]; **Figure 1B:** Ti = Tingart; Ma = Mather; L1 = Level 1; M1 = our method's first metaphyseal level at the lower-most (transversely tangential) edge of the best-fit circle.

Figure 2 (middle) Orientation of loaded humerus. Table at right = correlation coefficients

Table 1 The colors in the table represent the extremes of the proximal and distal data obtained by the five observers at each diaphyseal level of the 33 bones measured. Numbers marked with an asterisk represent a significant correlation ($p<0.05$). Mean Comb. CT = mean combined cortical thickness.

ORS 2016 Annual Meeting Poster No. 1655