Potential Consequences of Contradictions in Bone Mineral Density Assessments in Injury Biomechanics

Randee L. Hunter, Zachary A. Haverfield, Yun-Seok Kang, Amanda M. Agnew

Abstract Previous work has demonstrated contradictory relationships between bone mineral density (BMD) and injury outcomes particularly in the thorax. This study evaluated the agreement between BMD quantified from dual energy x-ray absorptiometry (DXA) and quantitative computed tomography (QCT) and the subsequent implications for predicting rib fractures from experimental male PMHS tests. Areal BMD (aBMD) and volumetric BMD (vBMD) were obtained from the 2nd-4th lumbar vertebra and left femoral neck (Fem-N) of 83 male PMHS (24-102 years). Raw BMD values were normalized by a standard deviation score (SDS) in reference to aBMD and vBMD sample means per site and bone type. aBMD and vBMD SDS were not related in the hip (p=0.08-0.82) and only weakly related in the spine (p<0.001). Between 67.5-77.1% of PMHS demonstrated disagreement or mismatch between assessments with no clear trends of over- or under-prediction. Additionally, the relationship between BMD and number of rib fractures (NRF) was investigated in a subsample of whole-body PMHS experiments. Trabecular bone vBMD in the spine (p=0.04) and Fem-N (p=0.047) as well as the inferior cortex (p=0.01) predicted NRF whereas aBMD (p>0.15) did not. Overall, DXA and QCT do not provide similar representations of male bone quality, implying caution is necessary when categorizing PMHS injury risk using these measures.

Keywords CT, DXA, male bone quality, rib fracture, thorax,

I. INTRODUCTION

The ability to accurately characterize bone quality is crucial to understanding and mitigating injury risk. Skeletal assessment methods in injury biomechanics leverage the clinically available technologies of dual energy x-ray absorptiometry (DXA) and quantitative computed tomography (QCT) to evaluate bone mineral density (BMD). However, concordance, when two measurements of an individual's skeletal health places them within the same World Health Organization categories (normal, osteopenic, osteoporotic)[1], has historically been an issue with up to 40% discordance rates when using DXA [2]. This can be attributed to both physiologic discordance between skeletal sites, particularly the hip and spine due to differential rates of bone loss, as well as inherent limitations to the 2-dimensional technology of DXA (i.e., superimposition of structures, presence of hyperdense osteophytes, etc.) [3,4]. Additionally, [4] found that discrepancies between evaluations of BMD of the hip and spine persist across technologies (DXA vs. QCT). Similarly, [5] demonstrated weak relationships between DXA and QCT measurements of BMD at the hip, spine, and forearm sites suggesting differences in the aspects of bone quality captured by each technology. Although classically investigated as an issue in clinical assessments of bone quality and osteoporosis treatment, the concept of discordance will likely have tangible effects on the field of injury biomechanics.

Despite mitigation efforts, thoracic injury risk in motor vehicle crashes persists, particularly in elderly occupants [6] with rib fractures as a leading contributor to mortality [7]. Importantly, conflicting results have been reported for predicting thoracic injury, particularly rib fractures, from post-mortem human subject (PMHS) BMD values. No meaningful relationship was found between areal BMD (aBMD)[8] or volumetric BMD (vBMD)[9] and structural properties in isolated ribs tested to failure. However, some whole body experimental testing has demonstrated a weak association between measures of BMD and rib fractures such that the PMHS with the lowest BMD values experienced the highest number of rib fractures [10,11]. Yet it appears expectations about injury outcomes have not matched aBMD of PMHS in other studies [12,13]. Recent work by [14] demonstrated

R.L. Hunter (Randee.Hunter@osumc.edu) is an Assistant Professor, Z.A.Haverfield is a PhD student, Y-S. Kang is an Associate Professor, and A.M. Agnew is Professor in the Injury Biomechanics Research Center at The Ohio State University, Columbus, OH, U.S.A.

no clear trends between BMD and rib fractures during experimental near-side impact testing. Further complicating the evidence for BMD effects on susceptibility to fracture, multiple studies have demonstrated a fracture risk increase independent of aBMD T-scores [15-18]. The utilization of BMD assessments in experimental PMHS testing has been variable across the field [5] with ambiguous results in interpreting injury outcomes or normalization through PMHS selection criteria.

To facilitate the necessary standardization of the use of bone quality assessments in the field of injury biomechanics suggested by [5], it is necessary to first investigate the contradictions in BMD classifications across technologies, and second, to quantify the effects of discrepancies on predicting actual injury outcomes. The purpose of this study is to evaluate the variance in standardized BMD scores between DXA and QCT and their relationship to thoracic injury outcomes in an experimental PMHS test series.

II. METHODS

Eighty-three male post-mortem human subjects (PMHS) were included for BMD analysis of imaging data curated in the Injury Biomechanics Research Center (Columbus, OH, USA). PMHS represented a wide range of ages and body sizes (Table I). Areal bone mineral density (aBMD, g/cm^2) was collected immediately following death using manufacturer protocols on a singular General Electric Lunar Prodigy DXA scanner (CV of 0.21%). Standard sites for bone quality assessment were included (Table II). Whole-body computed tomography (CT) scans were performed on each subject using emergency department clinical CT scanners which undergo daily quality control procedures (Siemens Edge and Siemens Force; combined CV of 1.92%) under consistent acquisition parameters (0.6 mm slice thickness, 120 kVp, and a reference 250 mAs.) An INTable^{TM} phantom with rods of known densities (0, 75, and 150 mg/cm^3) was included in each scan throughout the entirety of the body to facilitate site-specific Hounsfield unit (HU) to density calibration using scan-specific internal calibration to minimize effects of x-ray tube fluctuations across PMHS scans.

Volumes of interest (VOIs) for the lumbar spine (L2-L4) and femoral neck (left) were created in Osirix MD (v.12.0) using a combination of blunt (Fig. AI) and Hounsfield unit (HU) threshold segmentations to isolate bone voxels for BMD analysis from quantitative CT (QCT) scans. HU thresholds were held consistent from previous work [5] and were specific to bone type using 150-660 HU for trabecular (Tb), 661-3000 HU for cortical (Ct), and 150-3000 HU for total (Tb and Ct) bone. Lumbar spine VOIs consisted of five axial slices centered around the 50% location relative to the height of each lumbar vertebra. Femoral neck VOIs were defined in the coronal plane and included three slices defined from the coronal 50% (anterior to posterior) midpoint. Mean HU values for each VOI and bone type were used to calculate volumetric bone mineral density (vBMD) (Table II). Site-specific calibration curves were constructed using a custom, validated MATLAB code to quantify HU values from each INTable phantom rod which were subsequently plotted against their known densities (mg/cm^3) to obtain a linear fit equation. Raw aBMD and vBMD values for each site and bone type were transformed into standard deviation scores (SDS) to create a normalized distribution of the study sample (Eq. 1) to allow for direct comparisons between BMD values.

$$SDS = \frac{(BMD - \mu_{site})}{s_{site}} \quad (1)$$

where SDS is the normalized bone quality score, BMD is the calculated aBMD or vBMD value of a skeletal site, μ_{site} is the site-specific sample aBMD or vBMD mean, and s_{site} is the sample aBMD or vBMD standard deviation.

TABLE I
MALE SAMPLE SIZE DEMOGRAPHICS (N=83)

	Age (yrs)	Height (cm)	Weight (kg)	BMI (kg/cm^2)
Mean±SD	61.5±14.4	176.3±7.0	74.2±10.9	23.9±3.5
Range	24-102	160.0-190.5	52.2-100.8	17.0-32.8

Experimental Testing Subsample

To investigate the consequences of discordance in assessment methods of bone on predicting injury outcomes in the thorax during whole-body experimental PMHS testing, BMD SDS from a subsample (n=14) of the total 83 male PMHS were used (Table AI). This subsample was chosen as the remainder of the total sample were individuals for whom CTs were curated but were not subjected to comparable whole-body experimental testing. SDS values for these PMHS represent their position with respect to the distribution of the entire sample (Table AII) similar to the way in which DXA and QCT T or Z-scores are calculated in reference to normative populations[19]. Details of test conditions can be found in [20] and involved high-speed (ΔV of 56 km/h) rear-facing frontal-impact sled test scenario using two modern production seats (2018 Honda Odyssey with all-belt-to-seat or 2018 Honda Accord with fixed D-ring) in two recline positions (25 or 45 degrees). Number of rib fractures (NRF) were quantified during post-test dissection by anatomical injury experts (Hunter and Agnew, CAISS) and used as an indication of overall thoracic injury severity.

Data Analysis

All statistical tests were performed using Minitab 18 statistical software with an *a priori* α=0.05. Normality of BMD was assessed using Kolmogorov-Smirnoff tests. To investigate the relationship between aBMD and vBMD SDS values, linear regressions were used. In order to assess the level of agreement between DXA and QCT assessments, SDS values were evaluated using Bland-Altman plots representing paired differences for each individual in the sample [21]. The closer the paired differences between aBMD SDS and vBMD SDS, the more concordant (or matched) the assessment of the individual's bone quality with respect to the whole sample. In this study, discordance (or mismatched assessments) was defined as any paired differences for individuals falling outside of *maximum acceptable differences* (±0.44 standard deviations; SD) derived from literature as this value represents the mean difference in absolute values of the lumbar spine T-scores from the same individuals across two DXA machines [22]. Previous work has utilized 0.44 SD to investigate the average difference between vBMD values from two CT systems [23]. Lastly, to investigate the ability of SDS scores across both methods of assessment (DXA and QCT) to predict injury outcomes of the thorax (NRF), multiple regressions were performed to control for specific test conditions.

III. RESULTS

All values of aBMD and vBMD were normally distributed (p>0.05). Descriptive statistics for raw BMD values are provided in Table II for reference. SDS values were calculated for each skeletal site, bone type, and technology (DXA or QCT). There were no differences between the variances of aBMD SDS and vBMD SDS per site and bone type (p>0.05).

TABLE II
SITE-SPECIFIC BMD DESCRIPTIVE STATISTICS (N=83)

Site	Bone Type	Method	Mean ± std. dev (mg/cm^2)	Min (mg/cm^2)	Max (mg/cm^2)
L2-L4	Total	DXA	1306.3±184.9	891	1705
Fem-N	Total	DXA	1019.1±192.6	563	1577
Site	**Bone Type**	**Method**	**Mean ± std. dev (mg/cm^3)**	**Min (mg/cm^3)**	**Max (mg/cm^3)**
L2-L4	Total	QCT	306.9±32.6	240.7	383.1
L2-L4	Tb	QCT	224.0±29.2	169.8	297.2
Fem-N	Total	QCT	407.4±55.8	256.6	609.8
Fem-N	Tb	QCT	289.6±43.4	192.2	406.4
Inf Fem-N	Ct	QCT	999.7±73.1	871.8	1206.2
Sup Fem-N	Ct	QCT	846.2±75.2	682.3	1087.4

**All Fem-N values for the left femoral neck only. Tb= trabecular, Ct= cortical*

The relationship between aBMD and vBMD SDS varied between sites and bone type. Significant positive relationships were found between L2-L4 aBMD SDS and both total vBMD ($p<0.01$, $R^2=16.3\%$) and trabecular vBMD ($p<0.01$, $R^2=15.5\%$) (Fig. 1). Yet, a large number of individual values falling outside the 95% confidence intervals suggests the relative position of these individuals within the sample with respect to bone quality is dissimilar across technologies (DXA and QCT). Additionally, there were no significant relationships between aBMD SDS and vBMD SDS in the femoral neck for any bone type (Fig. 2). This implies that measures of BMD between DXA and QCT, obtained on identical anatomical sites, are capturing bone quality differently possibly due to re-ordering or mismatched assessments of individuals within the sample despite the initial findings here of similarities in distributions of SDS between technologies.

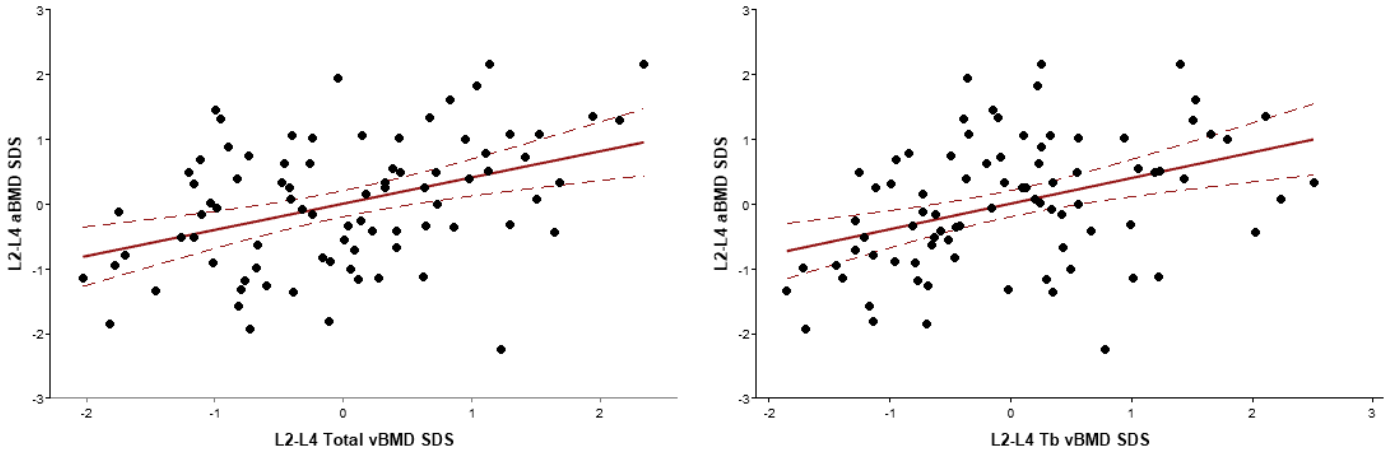

Fig. 1: Linear regressions for L2-L4 SDS between aBMD and vBMD for total bone (left: $p<0.01$, $R^2=16.3\%$) and trabecular bone (right; $p<0.01$, $R^2=15.5\%$). Regression lines (red) with 95% confidence interval (dotted red lines).

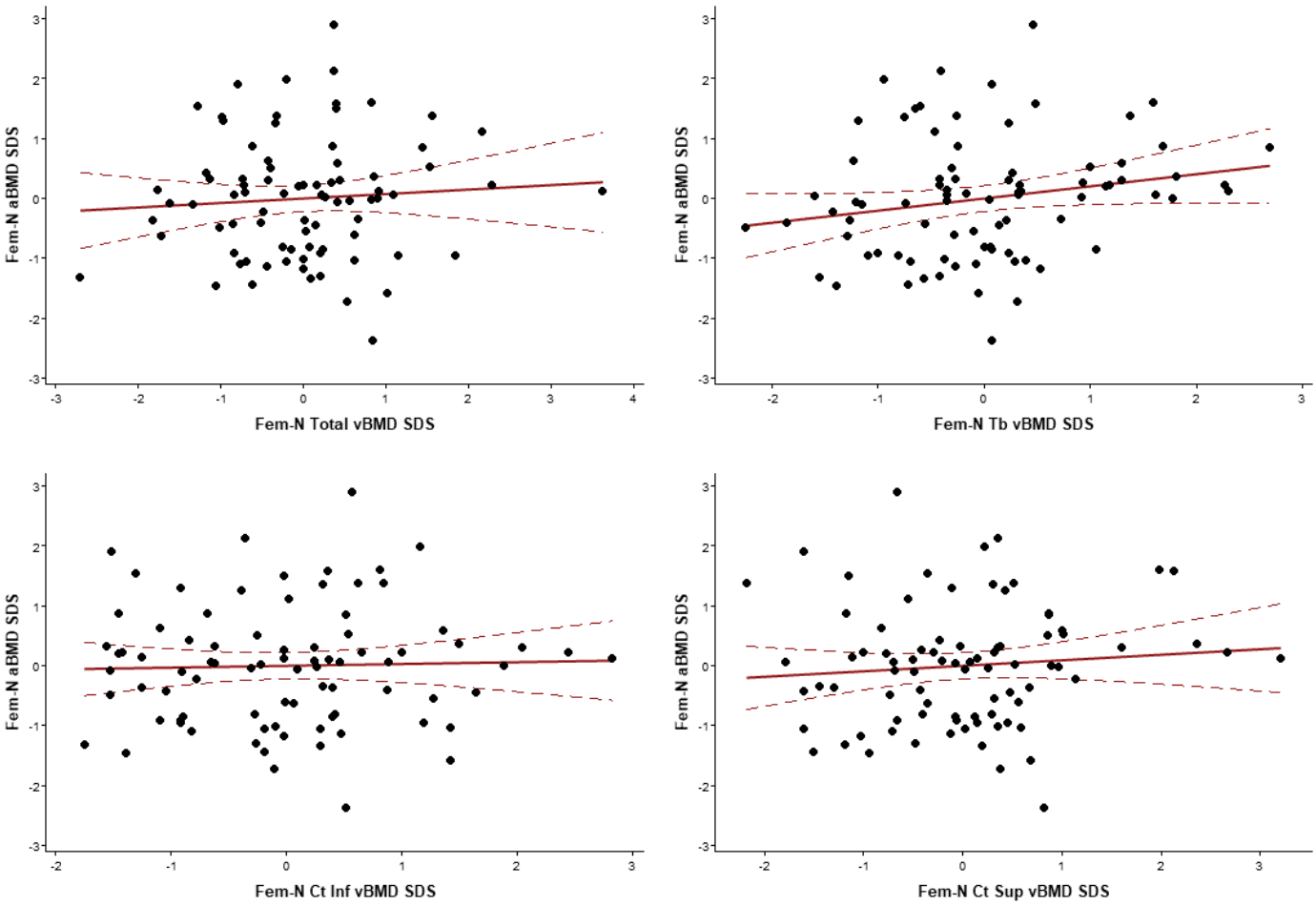

Fig. 2: Linear regressions for Fem-N SDS between aBMD and vBMD for total bone (top left; p=0.59), trabecular bone (top right; p=0.08), and cortical bone (inf. bottom left; p=0.82; sup. bottom right, p=0.46). Regression lines (red) with 95% confidence interval (dotted red lines).

To investigate discordance in assessment techniques, the level of agreement of normalized measures of bone quality was performed using Bland-Altman plots. The paired differences per site between aBMD SDS and vBMD SDS (DXA-QCT) were plotted against their paired averages. The maximum acceptable difference range (±0.44 SD) or acceptable bias within which paired assessments do not result in a meaningful difference in categorizing bone quality was established based on literature values [22,23]. Fig. 3 demonstrates the entire sample for L2-L4 with similar 95% upper (ULA) and lower (LLA) limits of agreement at ±2.14 SDS (total) and ±2.16 (trabecular). The femoral neck vBMD sites, when compared to aBMD SDS, have wider limits of agreement (ranging from ±2.48 to ±2.73) than the lumbar spine suggesting more error in categorizing PMHS bone quality between technologies at this anatomical site (Fig. 4). PMHS with paired differences within the ±0.44 maximum acceptable differences thresholds represent general agreement in their bone quality assessment. Table III demonstrates the summary of PMHS who fall outside of these limits per site and bone type. For L2-L4, only 21/83 (25.3%) and 27/83 (32.5%) of paired differences are within the maximum acceptable differences range for total and trabecular bone types, respectively. Between 22.9% (total; 19/83) and 30.1% (superior cortex; 25/83) PMHS were within these limits for Fem-N. Fem-N Tb (21/83) had 25.3% of PMHS within the ±0.44 limits and only 19/83 (22.9%) for Fem-N Ct Inf. Since the averages of paired differences (x-axis) are not close to zero for any site, there is no consistent over- or under-prediction between methods suggesting they are producing highly variable results across the sample. Furthermore, there were no noticeable patterns for individuals consistently falling above or below the 0.44 thresholds per anatomical location. For example, of the 21 PMHS who were within the ±0.44 limits (matched assessments) for L2-L4 Total SDS, 6 were above the limit for L2-L4 Tb SDS and 6 were below the maximum acceptable limit. For the Fem-N, even within the cortical bone type, of the 19 PMHS who had matched assessments (or fell within ±0.44 SD) for the inferior cortex, 8 PMHS were under -0.44 SD and 3 PMHS were above

0.44 SD further supporting a clear lack of over- or under-prediction between DXA and QCT within an anatomical site. Lastly, the discordance between assessments was not influenced by body size (BMI) as there were no significant relationships between SDS differences and BMI at any site (p>0.30; Fig. AII).

TABLE III
FREQUENCY OF N=83 PMHS WITH RESPECT TO MAXIMUM ACCEPTABLE DIFFERENCES (±0.44 SD) THRESHOLDS AND 95% LIMITS OF AGREEMENT

Site	Bone Type	Below LLA	Below -0.44 SD	Within ±0.44 SD	Above 0.44 SD	Above ULA
L2-L4 SDS	Total	1	30	21	29	2
	Tb	5	20	27	30	1
Fem-N SDS	Total	3	31	19	28	2
	Tb	0	34	21	25	3
	Inf Ct	2	29	19	31	2
	Sup Ct	2	27	25	25	4

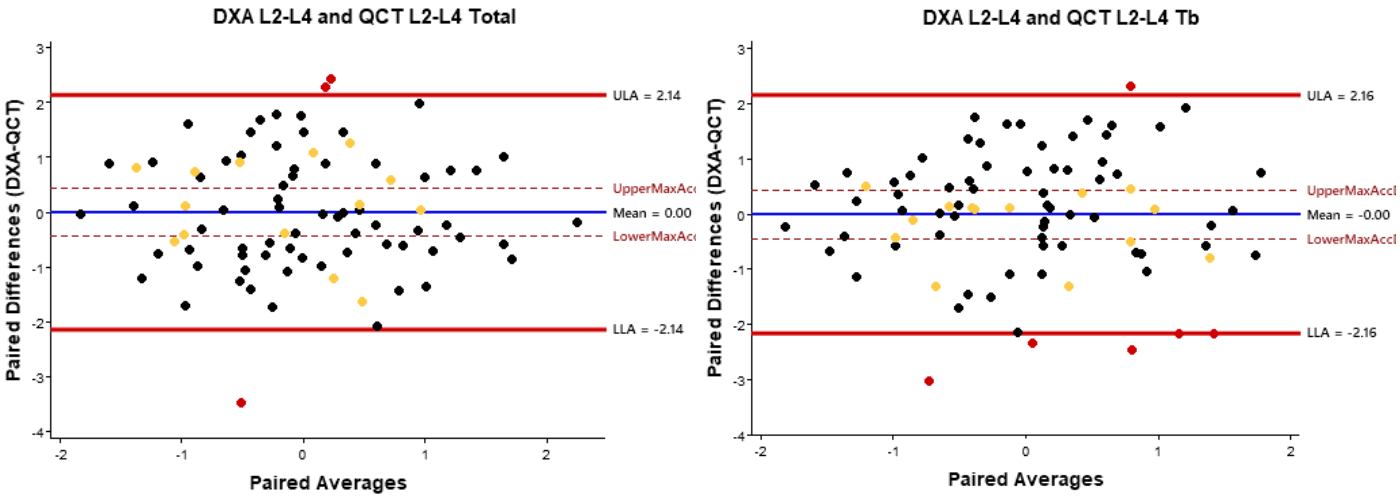

Fig. 3: Bland-Altman plots for L2-L4 aBMD SDS compared to L2-L4 vBMD total SDS (left) and QCT L2-L4 trabecular SDS (right). Maximum acceptable differences indicated in solid red line and set at 0.44SD (UpperMaxAccDiff) and -0.44SD (LowerMaxAccDiff). Yellow dots represent n=14 subsample of experimentally tested PMHS. Red dots are PMHS that fall outside the ULA (upper limit of agreement) or LLA (lower limit of agreement).

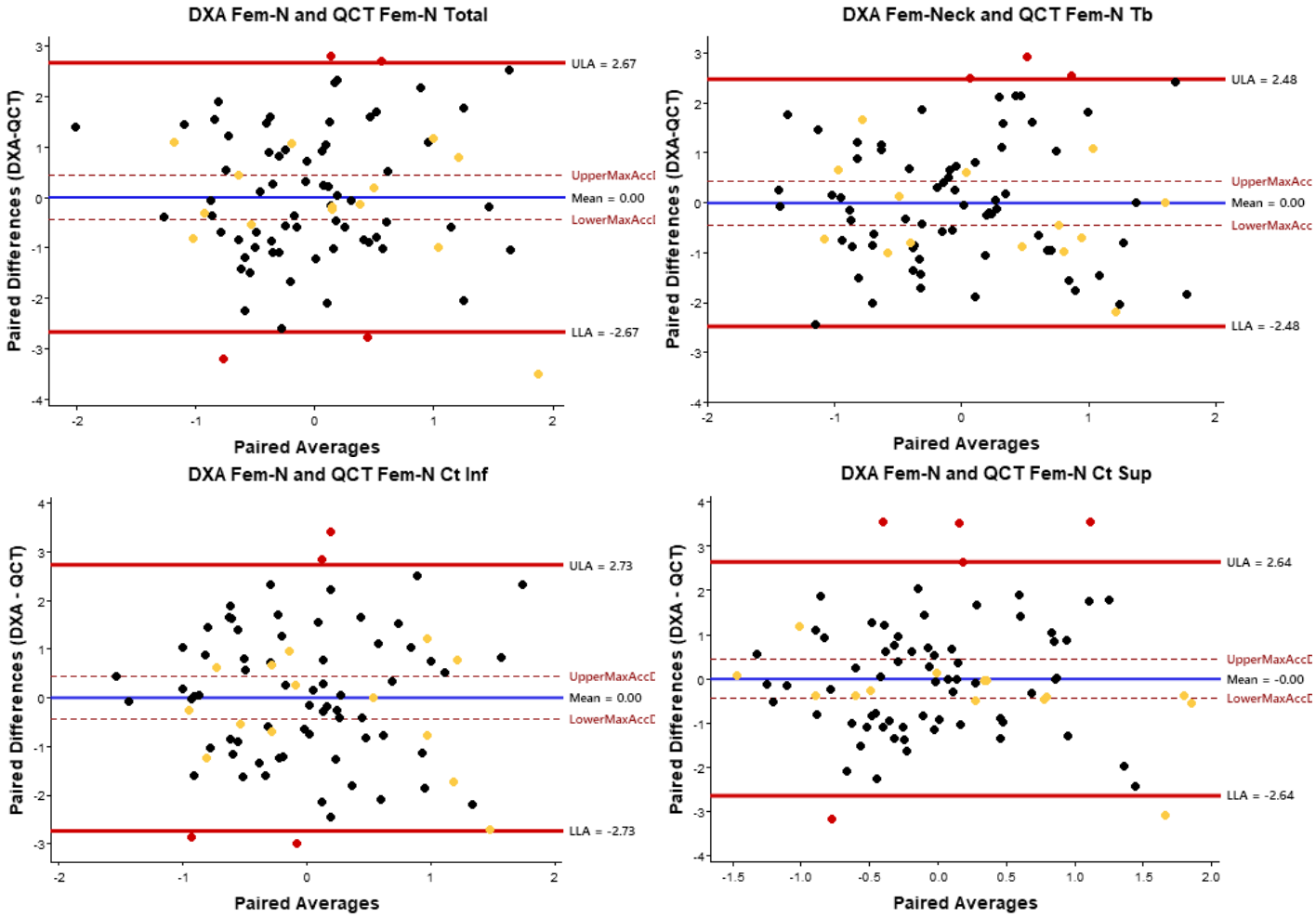

Fig. 4: Bland-Altman plots for femoral neck aBMD SDS compared to femoral neck vBMD total SDS (top left), trabecular (top right), inferior cortex (bottom left), and superior cortex (bottom right). Maximum acceptable differences indicated in solid red line and set at 0.44SD (UpperMaxAccDiff) and -0.44SD (LowerMaxAccDiff). Yellow dots represent n=14 subsample of experimentally tested PMHS. Red dots are PMHS that fall outside the ULA (upper limit of agreement) or LLA (lower limit of agreement).

The implications of the mismatched assessment of bone quality across technologies was investigated with respect to NRF in the subsample of experimentally tested whole-body PMHS (n=14). SDS vales per site and bone type are reported in Table AII. Overall thoracic injury severity as determined by NRF ranged from as low as 0 to as high as 42 rib fractures in the sample (Table IV). The paired differences in SDS between DXA and QCT per anatomical site are demonstrated in Table IV. Fig. 3 and 4 shows the position of these PMHS (yellow dots) within the entire sample with no clear trend of systematic over or under-prediction of their SDS. For L2-L4, 9/14 total vBMD SDS (64.2%) and 4/14 trabecular vBMD SDS (28.6%) were outside the ±0.44 maximum acceptable differences range (Fig.3; Table IV). In the femoral neck, discordance or mismatched assessments between methods was generally higher with 12/14 (85.7%) of PMHS outside the maximum acceptable differences for Fem-N Tb vBMD SDS, 11/14 (78.6%) for Fem-N Ct Inf vBMD SDS, 8/14 (57.1%) for Fem-N Total vBMD SDS, and 5/14 (35.7%) for Fem-N Ct Sup vBMD SDS (Fig.4; Table IV). PMHS21, the only individual with 0 NRF, was also the only individual who consistently fell below the -0.44 SD maximum acceptable differences; the remainder of PMHS demonstrated no observable trend in SDS mismatched assessments (Table IV).

Multiple regressions, controlling for test condition, demonstrated the inability of DXA aBMD SDS to predict NRF for either L2-L4 (p=0.13) or Fem-N (p=0.20) (Table V, AIII). In this subsample, QCT vBMD SDS values for some sites and bone type were able to predict NRF. L2-L4 Tb vBMD SDS (p=0.047) significantly predicted the injury outcome of NRF (R^2=48.3%) (Fig 5, Table V, AIV). The strongest predictor of NRF in this subsample was the inferior cortex of the femoral neck (p=0.01, R^2=59.9%) (Fig 5, Table V, AIV). This may suggest the femoral neck as a preferred anatomical site for more accurate assessment of bone quality when using QCT.

TABLE IV

PAIRED DIFFERENCES IN SDS PER SITE (DXA-QCT) FOR WHOLE-BODY EXPERIMENTAL TEST PMHS (N=14)

PMHS	NRF	L2-L4 Total SDS	L2-L4 Tb SDS	Fem-N Total SDS	Fem-N Tb SDS	Fem-N Inf Ct SDS	Fem-N Sup Ct SDS
PMHS21	0	-1.63	-1.32	-3.51	-2.19	-2.71	-3.09
PMHS13	1	0.03	-0.81	-0.13	-0.98	-1.73	-0.03
PMHS03	3	0.15	-0.52	0.79	0.00	0.79	-0.38
PMHS01	4	1.25	0.45	1.09	0.66	-0.69	-0.28
PMHS02	5	0.91	0.09	-0.55	-0.81	-0.53	-0.40
PMHS06	10	-0.42	-0.43	-0.23	-0.89	0.25	-0.49
PMHS05	14	0.82	0.45	0.43	0.14	0.62	1.18
PMHS10	14	-0.38	0.08	1.07	0.60	0.96	-0.05
PMHS12	16	0.57	0.07	0.18	-0.71	-0.78	-0.42
PMHS11	17	1.07	0.38	-0.99	-0.45	0.00	-0.48
PMHS04	19	-1.22	0.10	-0.19	1.65	0.67	0.12
PMHS22	21	0.10	-0.12	1.18	1.09	1.22	-0.55
PMHS14	34	-0.54	-1.31	-0.32	-1.01	-0.27	-0.38
PMHS09	42	0.75	0.12	-0.82	-0.72	-1.25	0.07

Heatmap legend with reference values at ±0.44 SD:

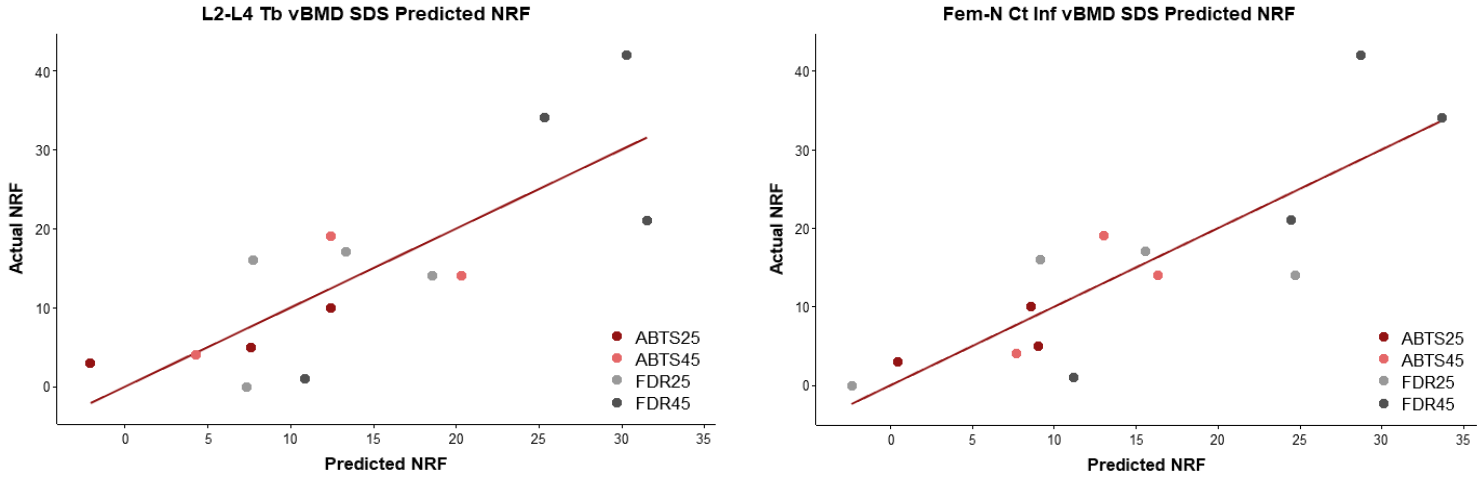

Fig. 5: Predicted NRF from L2-L4 trabecular vBMD SDS (left; p=0.04) and inferior femoral neck cortical vBMD SDS (right; p=0.01) calculated using multiple regression equations (Table AIV) while controlling for test condition (legend) which includes seatbelt type [ABTS (all-belt-to-seat) or FDR (fixed D-ring)] and recline angle (25 or 45 degrees).

TABLE V
MULTIPLE REGRESSIONS PREDICTING NRF CONTROLLING FOR TEST CONDITION (N=14)

Site	Bone Type	Method	F-value	R^2 adjusted	P-value
L2-L4	-	DXA	2.21	27.2%	0.15
Fem-N	-	DXA	1.77	19.2%	0.22
L2-L4	Total	QCT	2.79	35.5%	0.09
	Tb	QCT	4.03	48.3%	**0.04**
Fem-N	Total	QCT	1.72	18.1%	0.23
	Tb	QCT	3.74	45.7%	**0.047**
	Ct Inf	QCT	5.86	59.9%	**0.01**
	Ct Sup	QCT	2.61	33.2%	0.11

IV. DISCUSSION

Overall, results from this study suggest DXA and QCT BMD measurements do not represent bone quality similarly in the hip and spine region for male PMHS. The SDS used in this study was an exercise in categorizing PMHS with respect to their normalized BMD values to investigate error across technologies and at anatomical sites frequently used in assessing bone quality. World Health Organization categories (normal, osteopenic, or osteoporotic) calculated with respect to young reference populations were not utilized here as the goal was to directly compare the agreement of DXA and QCT measures on the same individuals. The total sample (n=83 males) used for the reference population to calculate SDS represented wide age and body size ranges and likely encompassed similar levels of variation in BMD that would be found in a larger population. Although the distributions for L2-L4 SDS and Fem-N SDS for each bone type were statistically similar between DXA and QCT, further investigations into their level of agreement demonstrated weak relationships (Fig. 1-2) and mismatched assessment of PMHS in this sample (Fig. 3-4).

Previously reported correlations between aBMD and HU from opportunistic CT ([24] for a review) were supported in this study but only for the lumbar spine; however, none of the femoral neck comparisons demonstrated significant relationships between DXA and QCT in this sample of male PMHS (Fig. 2). It is likely that the site-specific calibration of HU to vBMD coupled with the use of a threshold to isolate only bone voxels utilized in this study produced a more accurate representation of BMD than opportunistic CT which often employs retrospective phantom-less calculations using a generalized calibration equation for the scanner. Furthermore, despite previous evidence of correlations between aBMD and HU, discrepancies between BMD categorization of bone quality across technologies have been previously demonstrated in both sexes. In a study of postmenopausal females, [25] found significant correlations between DXA and QCT BMD values but a discordance rate of 33% in the lumbar spine. It was suggested the high discordance or mismatched assessment may have been due to QCT having higher sensitivity in detecting low BMD values while the accuracy of DXA declined in these individuals [25]. Interestingly, despite common assumptions concerning females having lower BMD values than males, a higher discordance rate between DXA and QCT was found in elderly Chinese males at 59.1% [26]. Although T-score categorization of PMHS was not employed in the current study, Bland-Altman plots demonstrate the poor level of agreement for both lumbar spine and femoral neck (Fig. 3-4). Only 21/83 (25.3%) and 27/83 (32.5%) of male PMHS fell within the ±0.44 SD threshold for L2-L4 total and trabecular, respectively, whereas the femoral neck ranged from 22.9-30.1% (Table III). This suggests that the error in assessment between technologies is beyond an acceptable level due to expected bias in the systems and reflects mismatched assessment of up to 77.1% of PMHS in this sample.

In previous work, aBMD had been shown to consistently overpredict vBMD in the spine of postmenopausal females [27] as well as elderly males [26] and is attributed to DXA's 2-dimensional limitations and inability to avoid osteophytes, vascular calcifications, and the effects of obesity and spinal surgery. In contrast, our study found no consistent over-or under-predictions of BMD between technologies with no observable pattern in the paired differences in SDS at any site or bone type (Fig. 3-4, Table IV). It also does not appear that the lower end of the BMD distribution is any more discordant than those individuals with higher BMD as the variation in paired differences across the sample is equally distributed across all values of averages (x-axis of Bland-Altman plots; Fig. 3-4). To demonstrate this concept in the subsample (Table IV), there is no clear or consistent positive or

negative trend in paired SDS differences within an individual or within an anatomical site (i.e., the L2-L4 trabecular SDS paired differences are not consistently positive across individuals). The detection rate of bone quality as measured by BMD and categorized by WHO is reported to be higher in QCT (58.2%) compared to DXA (30.1%) leading some to advocate for QCT to be the "gold-standard" of bone quality assessment [28]. However, with increasing evidence of individuals with similar BMD values but different fracture risk, it may be that any measure of BMD is not the strongest or most sensitive predictor of skeletal fragility. Future work should evaluate the relative contribution of BMD to fracture resistance across various anatomical locations, including the ribs, through experimental testing using realistic and dynamic loading rates.

It is important to note that the WHO has acknowledged the sensitivity of BMD classifications of bone quality to anatomical site and measurement technique [1]; however it is common practice to assign bone quality or strength assumptions globally despite large amounts of intra-skeletal variation [5]. For DXA evaluations, it is cautioned that 4 of 10 individuals will demonstrate discordance between the spine and hip [27]. To investigate the implications of the BMD discordance or mismatch in BMD between DXA and QCT identified within this large sample of male PMHS and the saliency for using these measures to predict PMHS thoracic injury severity in experimental testing, a subsample was used. Males in the subsample were selected to be approximately mid-size with initial DXA T-score screenings in the normal or osteopenic range (Table AI) [20]. Interestingly, despite the narrow age range (53-71 years), similarities in body size, and initial DXA screening bone quality assessments, thoracic injuries vary across this sample with NRF ranging from 0 to 42 (Table AI). For detailed summaries on location of rib fractures per PMHS, see [20]. Normalized aBMD (SDS) for L2-L4 and the femoral neck were unable to predict NRF caused by whole-body experimental testing (Table V). This may be a direct result of the non-site-specific nature of the assessment sites used in this study or reflect the low sensitivity of DXA technology. The inability of lumbar spine DXA aBMD to meaningfully predict rib structural properties was also found in [8] for isolated rib dynamic bending tests. Lumbar, femoral neck, and ultradistal radius aBMD standard deviation decreases were not associated with increased rib fractures in males in a sample of real world moderate to severe trauma scenarios [29]. Similarly, median NRF did not differ between normal and low aBMD groups presenting to trauma departments following blunt thoracic trauma [30]. Furthermore, aBMD T-score values for the subsample PMHS did not coincide with expectations of fracture susceptibility (Table AI). For example, PMHS11 and PMHS12 both had L2-L4 and Fem-N aBMD T-scores in the "normal" or "healthy" range yet sustained 17 and 16 NRF, respectively. PMHS09, who sustained the highest NRF (42), was also classified in the "normal" range based on L2-L4 T-score (-0.4) but demonstrated a major discordance, a discrepancy of two WHO categories between T-score values across sites[3], with an osteoporotic femoral neck T-score (-3.0) (Table AI).

In this subsample, multiple regressions including vBMD SDS to predict NRF were generally more successful than aBMD SDS but this was dependent on bone type. Similar investigations with raw aBMD and vBMD produced identical results as SDS data are simply transformed from these values. The strongest predictor of NRF in this sample was Fem-Neck Ct Inf vBMD explaining nearly 60% of rib fractures when controlling for test conditions (Table V, Fig. 5). L2-L4 trabecular vBMD was also able to significantly predict NRF in this subsample (Table V, Fig. 5). Weaver et al. [31] similarly identified the importance of lumbar trabecular vBMD in predicting rib and sternum fractures particularly below a threshold of 145 mg/cc. The relationship between cortical bone of the femoral neck and thoracic injury in the current subsample is unique and contradictory to previous work in males. A one SD decrease in lumbar trabecular spine vBMD resulted in a 3.7-fold increase in rib and sternum fractures in a study of elderly males yet only trabecular femoral neck vBMD (not cortical) resulted in a hazard ratio of 1.26 [32] suggesting assessments of the hip had a weaker relationship with thoracic injury compared to the spine. The inferior cortex of the femoral neck demonstrated the highest percentage of mismatched assessments between technologies (85.7%) for this subsample of males which may suggest that in addition to DXA and QCT capturing different aspects of bone quality with respect to BMD, the femoral neck inferior cortex may be providing the more accurate representations of fracture risk due to its relationship with NRF. It is unknown and should be investigated if similar relationships between vBMD, aBMD, and fracture risk would be found in females in similar experimental tests or for PMHS in differing injurious loading scenarios. Additionally, using non-site specific values of BMD to predict rib fractures is inherently limited due to significant variations in aBMD and vBMD demonstrated within PMHS across the skeleton [5] and even within individual rib elements [33]. Thus, the development and standardization of rib specific assessments of vBMD from whole-body screening CTs may allow for more meaningful prediction of thoracic injuries from experimental testing.

The implications of error in BMD assessment across technologies coupled with the varying evidence for relationships between BMD and fracture risk necessitates further investigations across anatomical sites and in additional PMHS experimental testing scenarios. The case study presented here provides some additional foundational insight into potential issues of the use of BMD in injury biomechanics initially presented by [5]. For example, PMHS21, a 63-year-old male, consistently had the largest paired differences in SDS between DXA and QCT but sustained 0 NRF (Table IV); whereas PMHS09 who sustained the largest number of NRF (42) demonstrated no mismatch between aBMD SDS and vBMD SDS in the L2-L4 Tb and Fem-N Sup Ct measurements. Potentially due to the small subsample (n=14), there was no identifiable trend for individuals falling outside of the maximum acceptable differences threshold with the exception of PMHS21 (Table IV) across all sites and bone types. However, the ±0.44 SD threshold was derived from literature and represents the maximum acceptable difference in detection between DXA systems, which was not investigated in this study. Thus, the choice of this threshold may be too conservative for expected differences in BMD assessment between DXA and QCT. Future work should quantify the meaningful thresholds for cross-technology comparisons of bone quality. Additionally, although one advantage to QCT analysis of BMD is the ability to isolate bone type and remove non-bone voxels, this process may have amplified the discordance in SDS compared to the 2-dimensional DXA aBMD analysis further supporting the need for comparable assessments of BMD. Lastly, BMD is a relatively accessible metric with available technologies (DXA and QCT) for whole-body or *ex vivo* component analysis; however, it is only one component of bone quality. A more advantageous avenue for assessing bone in PMHS for selection criteria, scaling injury risk curves, or predicting injury outcomes such as NRF may be found in measures of gross geometry of both ribs and thorax, rib cortical and trabecular geometry, microarchitecture, and material properties.

V. CONCLUSIONS

Normalized aBMD and vBMD values (SDS) in a large sample of male PMHS did not characterize bone quality in similar ways with evidence of mismatched assessments which could have meaningful effects on the assumptions made concerning an individual's susceptibility to fracture. Disagreement in assessment between DXA and QCT BMD measures suggests implementing normalized scores (T-scores or Z-scores) in PMHS selection criteria or predicting injury outcomes should be done with caution. Although vBMD SDS predicted rib fractures in this study, the relationship was specific to anatomical site, bone type and only applies to males in the particular experimental scenario represented here as a proof-of-concept exercise. This, coupled with the lack of ability of frequently used aBMD sites (spine and hip) to predict NRF suggests the need to develop rib-specific methods of investigating PMHS bone quality to better improve thoracic injury mitigation.

VI. ACKNOWLEDGEMENT

The authors would like to thank the anatomical donors and the faculty, staff, and students of the Injury Biomechanics Research Center including John H. Bolte IV, Vikram Pradhan, Angela Harden, and Aditi Patel.

VII. REFERENCES

[1] World Health Organization. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: report of a WHO study group [meeting held in Rome from 22 to 25 June 1992]. 1994

[2] El Maghraoui, Mouinga Abayi DA, Rkain H, and Mounach A. Discordance in Diagnosis of Osteoporosis Using Spine and Hip Bone Densitometry. *Journal of Clinical Densitometry*, 2007. 10(2): p. 153-156

[3] Woodson G. Dual X-Ray Absorptiometry T-Score Concordance and Discordance Between Hip and Spine Measurement Sites. *Journal of Clinical Densitometry*, 2000. 3(4): p. 319-324

[4] Faulkner KG, von Stetten E, and Miller P. Discordance in Patient Classification Using T-Scores. *Journal of Clinical Densitometry*, 1999. 2(3): p. 343-350

[5] Haverfield ZA, Hunter RH, Loftis KL, Agnew AM. Skeletal site and method-dependent variability in bone mineral density in injury biomechanics research. *Proceedings of International Research Council on Biomechanics of Injury*, 2022. Porto, Portugal

[6] Stitzel J, Kilgo P, et al. Age thresholds for increased mortality of predominant crash induced thoracic injuries. *Annals of the Advancement of Automotive Medicine*, 2010. 54: p. 41-50

[7] Lee E, Craig M, and Scarboro M. Real-world rib fracture patterns in frontal crashes in different restraint conditions. *Traffic Injury Prevention*, 2015. 16: p. S115-S123

[8] Agnew AM, Murach MM, et al. Sources of Variability in Structural Bending Response of Pediatric and Adult Human Ribs in Dynamic Frontal Impacts. *Stapp car crash journal*, 2018. 62: p. 119-192
[9] Hunter RL, Kang Y-S, Briley KC, and Agnew AM. Feasibility of volumetric bone mineral density (vBMD) to predict rib structural properties. *American Journal of Physical Anthropology*, 2019. 168: p. 110-110
[10] Kang Y-S, Agnew AM, Hong C-B, Icke K, and Bolte IV J. Elderly PMHS Thoracic Responses and Injuries in Frontal Impacts. *Proceedings of International Research Council on the Biomechanics of Injury* 2017. Antwerp, Belgium
[11] Jin X, Begeman P, et al. Comparison of Small Female and Mid-sized Male PMHS Response with an Inflatable Seatbelt System during Frontal Impacts. *Proceedings of International Research Council on the Biomechanics of Injury*, 2019. Florence, Italy
[12] Shaw G, Lessley D, et al. Small female rib cage fracture in frontal sled tests. *Traffic Injury Prevention*, 2017. 18(1): p. 77-82
[13] Shaw G, Lessley D.J, et al. Side Impact PMHS Thoracic Response With Large-Volume Air Bag. *Traffic Injury Prevention*, 2014. 15(1): p. 40-47
[14] Bolte IV J, Fibbi C, et al. Analysis of injury mechanism and thoracic response of elderly, small female PMHS in near-side impact scenarios. *Traffic Injury Prevention*, 2023
[15] Sornay-Rendu E, Boutroy S, Duboeuf F, and Chapurlat RD. Bone Microarchitecture Assessed by HR-pQCT as Predictor of Fracture Risk in Postmenopausal Women: The OFELY Study. *Journal of Bone and Mineral Research*, 2017. 32(6): p. 1243-1251
[16] Boutroy S, Khosla S, et al. Microarchitecture and Peripheral BMD are Impaired in Postmenopausal White Women With Fracture Independently of Total Hip T-Score: An International Multicenter Study. *Journal of Bone and Mineral Research*, 2016. 31(6): p. 1158-1166
[17] Sornay-Rendu E, Munoz F, Garnero P, Duboeuf F, and Delmas PD. Identification of osteopenic women at high risk of fracture: The OFELY study. *Journal of Bone and Mineral Research*, 2005. 20(10): p. 1813-1819
[18] Schuit SCE, van der Klift M, et al. Fracture incidence and association with bone mineral density in elderly men and women: the Rotterdam Study. *Bone*, 2004. 34(1): p. 195-202
[19] WHO. WHO scientific group on the assessment of osteoporosis at primary health care level. *World Healht Organization.* 2004. Brussels, Belgium
[20] Kang Y-S, Stammen J, et al. Thoracic responses and injuries to male post-mortem human subjects (PMHS) in rear-facing seat configurations in high-speed frontal impacts. *Traffic Injury Prevention*, 2023
[21] Bland JM and Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *International Journal of Nursing Studies*, 2010. 47(8): p. 931-936
[22] McMahon K, Kalnins S, Freund J, and Pocock N. Discordance in Lumbar Spine T-Scores and Nonstandardization of Standard Deviations. *Journal of Clinical Densitometry*, 2003. 6(1): p. 1-6
[23] Mueller D, Kutscherenko A, et al. Phantom-less QCT BMD system as screening tool for osteoporosis without additional radiation. *European Journal of Radiology*, 2011. 79: p. 375-381
[24] Gausden EB, Nwachukwu BU, Schreiber JJ, Lorich DG, and Lane JM. Opportunistic Use of CT Imaging for Osteoporosis Screening and Bone Density Assessment: A Qualitative Systematic Review. *JBJS*, 2017. 99(18): 1580-1590
[25] Lin W, He C, et al. Discordance in lumbar bone mineral density measurements by quantitative computed tomography and dual-energy X-ray absorptiometry in postmenopausal women: a prospective comparative study. *The Spine Journal*, 2023. 23(2): p. 295-304
[26] Xu X-m, Li N, et al. Discordance in diagnosis of osteoporosis by quantitative computed tomography and dual-energy X-ray absorptiometry in Chinese elderly men. *Journal of Orthopaedic Translation*, 2019. 18: p. 59-64
[27] Mounach A, Mouinga Abayi, DA, et al. Discordance Between Hip and Spine Bone Mineral Density Measurement Using DXA: Prevalence and Risk Factors. *Seminars in Arthritis and Rheumatism*, 2009. 38(6): p. 467-471
[28] Kulkarni AG, Thonangi Y, et al. Should Q-CT Be the Gold Standard for Detecting Spinal Osteoporosis? *Spine*, 2022. 47(6): E258-E264
[29] Wuermser L, Achenbach SJ, Amin S, Khosla S, and Melton LJ. What accounts for rib fractures in older adults? *Journal of Osteoporosis*. 2011. p. 1-6

[30] Prins JTH, Van Lieshout EMM, Reijnders MRL, Verhofstad MHJ, and Wijffels MME. Rib fractures after blunt thoracic trauma in patients with normal versus diminished bone mineral density: a retrospective cohort study. *Osteoporosis International*, 2020. 31(2): p. 225-231

[31] Weaver A, Beavers K, et al. Lumbar bone mineral density phantomless computed tomography measurements and correlation with age and fracture incidence. *Traffic Injury Prevention*, 2015. 16: p. S153-S160

[32] Chalhoub D, Orwoll ES, et al. Areal and volumetric bone mineral density and risk of multiple types of fracture in older men. *Bone*, 2016. 92: p. 100-106

[33] Haverfield ZA, Agnew AM, and Hunter RL. Differential Cortical Volumetric Bone Mineral Density within the Human Rib. *Journal of Clinical Densitometry*, 2023. 26(2). p. 1-8

VIII. APPENDIX

TABLE AI

DEMOGRAPHICS, ABMD T-SCORE, AND NUMBER OF RIB FRACTURES (NRF) FOR WHOLE-BODY EXPERIMENTAL PMHS (N=14)[20]

PMHS	Age (yrs)	Height (cm)	Weight (kg)	BMI	NRF	L2-L4 aBMD T-score	Fem-N aBMD T-score
PMHS21	62	172.7	68.5	22.2	0	-0.1	0.1
PMHS13	53	176.3	76.2	25.5	1	1.9	0.1
PMHS03	54	174.0	93.9	29.7	3	1.2	1.7
PMHS01	57	167.0	62.6	21.6	4	1.9	-1.3
PMHS02	64	171.0	62.6	21.6	5	0.3	-1.6
PMHS06	61	176.5	72.6	22.3	10	-1.4	0.3
PMHS05	62	176.0	77.1	24.4	14	-1.0	-1.1
PMHS10	62	177.8	100.7	32.8	14	-0.1	0.5
PMHS12	58	177.8	71.7	22.0	16	1.9	0.8
PMHS11	65	185.4	92.1	26.8	17	1.3	0.4
PMHS04	59	178.0	96.2	30.3	19	-0.1	-0.4
PMHS22	61	176.6	71.7	23.3	21	-1.0	1.2
PMHS14	63	172.3	85.3	28.6	34	-1.6	-2.0
PMHS09	71	187.5	89.4	24.6	42	-0.4	-3.0
Mean±SD	61±5.0	176±4.8	80±12.7	25.4±3.7	14.3±12.3	-	-

TABLE AII

SUBSAMPLE (N=14) WHOLE-BODY PMHS TEST CONDITION, NUMBER OF RIB FRACTURES (NRF), AND SDS VALUES

PMHS	Test Condition	NRF	aBMD SDS		L2-L4 vBMD SDS		Fem-N vBMD SDS			
			L2-L4	Fem-N	Total	Tb	Total	Tb	Inf Ct	Sup Ct
PMHS21	FDR25	0	0.33	0.13	1.30	0.99	3.63	2.31	2.83	3.21
PMHS13	FDR45	1	0.99	0.33	0.95	1.79	0.44	1.30	2.05	0.35
PMHS03	ABTS25	3	0.54	1.6	0.39	1.05	0.82	1.60	0.81	1.98
PMHS01	ABTS45	4	1.01	-0.63	-0.24	0.56	-1.72	-1.29	0.06	-0.36
PMHS02	ABTS25	5	-0.07	-0.80	-0.98	-0.16	-0.25	0.01	-0.27	-0.41
PMHS06	ABTS25	10	-1.19	0.03	-0.77	-0.76	0.27	0.92	-0.22	0.52
PMHS05	ABTS45	14	-0.96	-0.41	-1.78	-1.45	-0.85	-0.56	-1.04	-1.60
PMHS10	FDR25	14	-0.34	0.34	0.04	-0.42	-0.73	-0.27	-0.62	0.38
PMHS12	FDR25	16	1.00	0.60	0.43	0.94	0.41	1.30	1.36	1.01
PMHS11	FDR25	17	0.62	0.54	-0.46	0.24	1.53	0.99	0.54	1.02
PMHS04	ABTS45	19	-0.36	0.06	0.86	-0.46	0.24	-1.60	-0.62	-0.07
PMHS22	FDR45	21	-0.91	1.59	-1.02	-0.79	0.40	0.49	0.36	2.13
PMHS14	FDR45	34	-1.33	-1.08	-0.79	-0.02	-0.77	-0.08	-0.82	-0.71
PMHS09	FDR45	42	-0.51	-1.43	-1.26	-0.64	-0.62	-0.71	-0.19	-1.50

TABLE AIII

DETAILED TABLES FOR MULTIPLE REGRESSIONS PREDICTING NRF CONTROLLING FOR TEST CONDITION (N=14) FOR DXA

Model	Term	Coefficient (β)/Constant*	T-Value	Term p-value	R^2_{Adj}	Model p-value
L2-L4 aBMD	SDS	-6.20	-1.68	0.128	27.2%	0.148
	ABTS25	4.49*	0.74	0.481		
	ABTS45	7.20*	0.84	0.422		
	FDR25	8.72*	1.07	0.314		
	FDR45	17.25*	2.15	0.060		
Fem-N aBMD	SDS	-4.68	-1.28	0.232	19.2%	0.219
	ABTS25	7.29*	1.13	0.287		
	ABTS45	3.47*	0.37	0.717		
	FDR25	6.31*	0.75	0.473		
	FDR45	16.47*	1.92	0.087		

TABLE AIV

DETAILED TABLES FOR MULTIPLE REGRESSIONS PREDICTING NRF CONTROLLING FOR TEST CONDITION (N=14) FOR QCT

Model	Term	Coefficient (β)/Constant*	T-Value	Term p-value	R^2_{Adj}	Model p-value
L2-L4 Total vBMD	SDS	-6.73	-2.08	0.067	35.5%	0.093
	ABTS25	2.95*	0.50	0.627		
	ABTS45	6.79*	0.84	0.420		
	FDR25	11.00*	1.39	0.199		
	FDR45	17.98*	2.39	**0.040**		
L2-L4 Tb vBMD	SDS	-8.00	-2.76	**0.022**	48.3%	**0.038**
	ABTS25	6.33*	1.24	0.245		
	ABTS45	2.41*	0.33	0.750		
	FDR25	8.90*	1.30	0.225		
	FDR45	18.84*	2.80	**0.021**		
Fem-N Total vBMD	SDS	-3.62	-1.22	0.252	18.1%	0.230
	ABTS25	7.00*	1.08	0.307		
	ABTS45	2.52*	0.26	0.798		
	FDR25	9.12*	1.02	0.333		
	FDR45	17.01*	1.99	0.078		
Fem-Neck Tb vBMD	SDS	-8.68	-2.62	**0.028**	45.7%	**0.047**
	ABTS25	13.29*	2.25	0.051		
	ABTS45	-10.95*	-1.11	0.297		
	FDR25	7.85*	1.13	0.287		
	FDR45	19.35*	1.86	0.095		
Fem-N Inf Ct vBMD	SDS	-7.84	-3.53	**0.006**	59.9%	**0.013**
	ABTS25	6.84*	1.53	0.161		
	ABTS45	1.31*	0.20	0.845		
	FDR25	12.94*	2.07	0.069		
	FDR45	20.39*	3.43	**0.008**		
Fem-N Sup Ct vBMD	SDS	-4.90	-1.97	0.081	33.2%	0.106
	ABTS25	9.43*	1.56	0.153		
	ABTS45	-0.42*	-0.05	0.964		
	FDR25	9.19*	1.17	0.272		
	FDR45	15.40*	1.97	0.080		

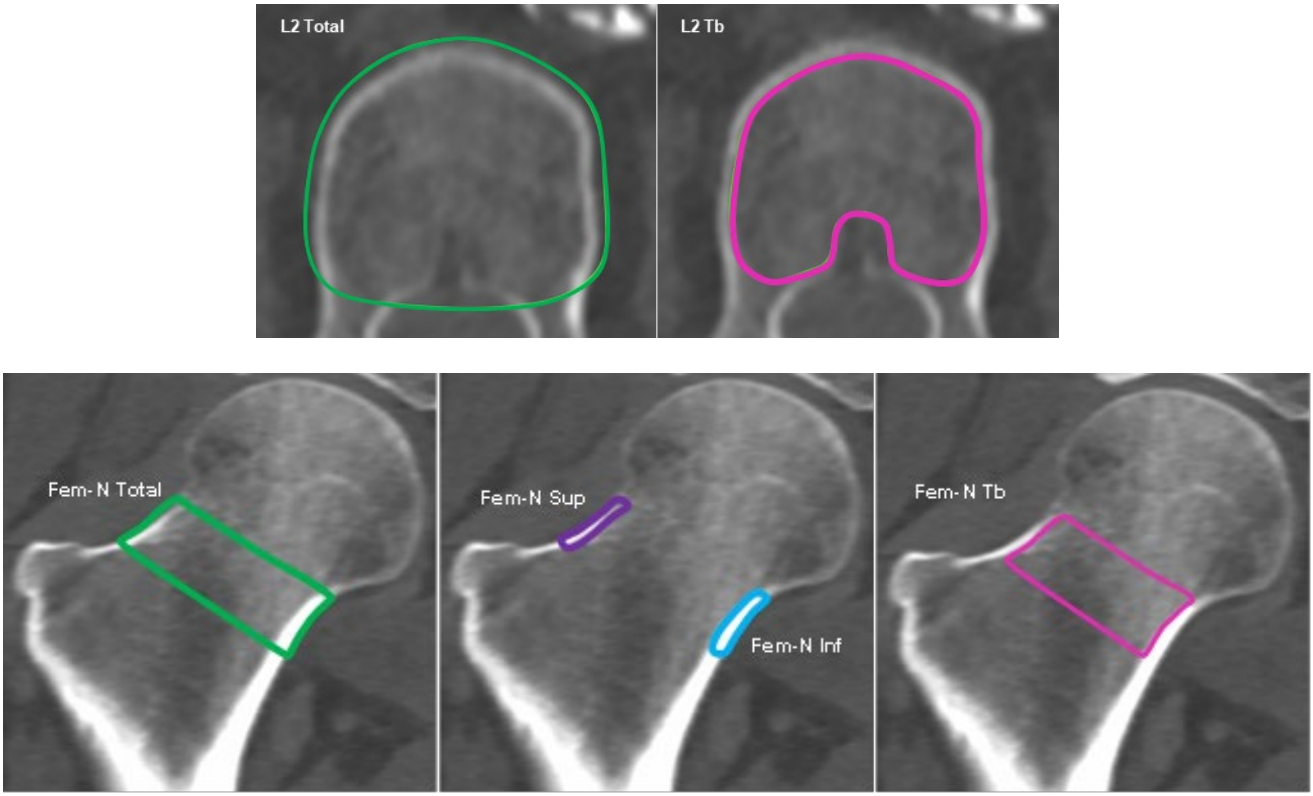

Fig. AI: Representative blunt VOIs (without threshold applied) for each anatomcial location and bone type. The lumbar spine VOIs are collected for each vertebra (L2, L3, and L4) and averaged for an L2-L4 total (trabecular and cortical) and L2-L4 Tb (trabecular). The femoral neck VOIs are Fem-N Total (trabecular and cortical), Fem-N Sup (cortical), Fem-N Inf (cortical), and Fem-N Tb (trabecular).

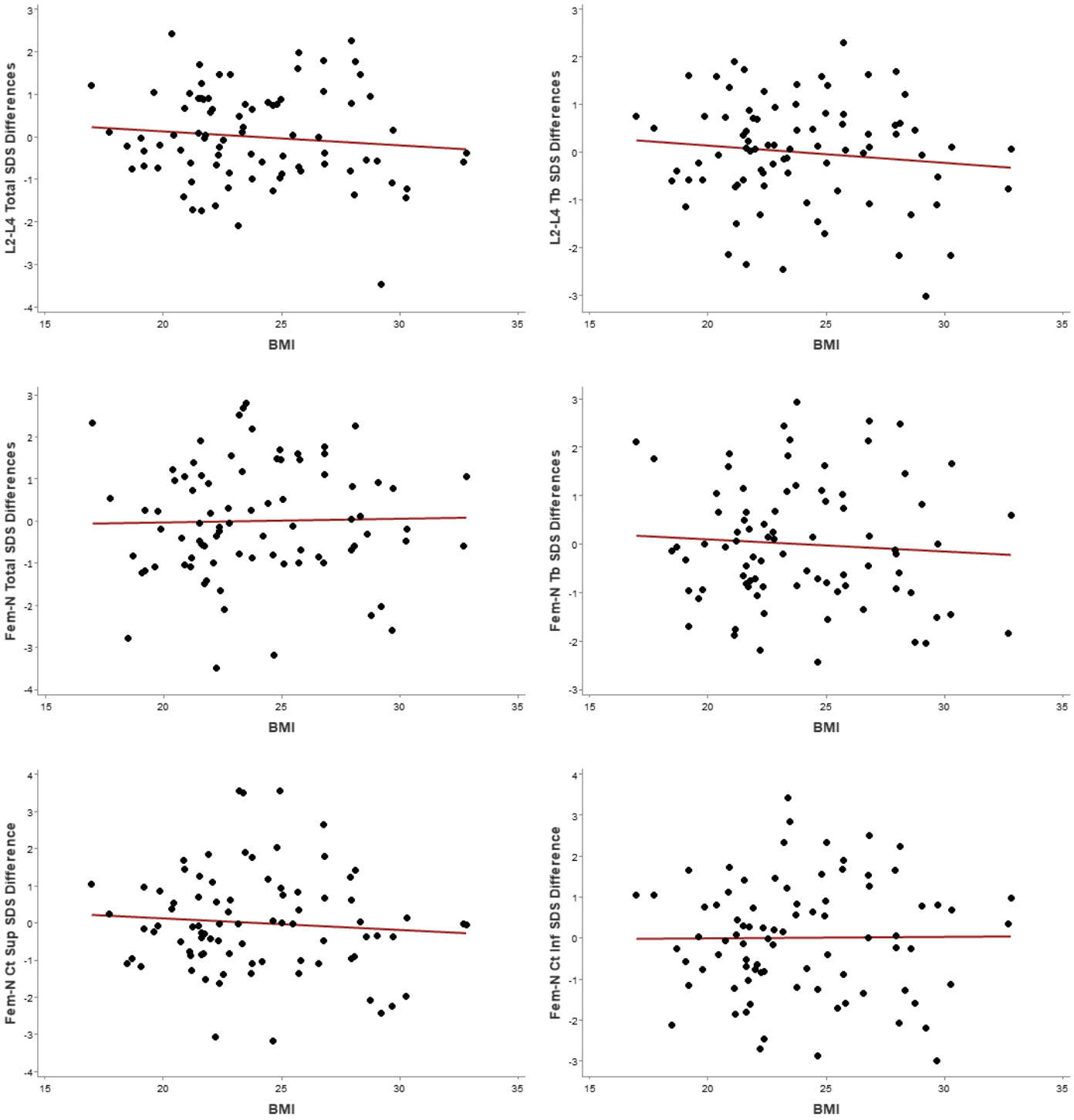

Fig. AII: Linear regressions for SDS differences and BMI at each anatomical site. No significant relationships were found between differences and body size suggesting that differential error due to larger BMI does not contribute to the discordance in assessments demonstrated in this sample. L2-L4 Total SDS p=0.35 (top left); L2-L4 Tb SDS p=0.30 (top right); Fem-N Total SDS p=0.84 (middle left); Fem-N SDS p=0.54 (middle right); Fem-N Sup SDS p=0.47 (bottom left); Fem-N Inf SDS p=0.94 (bottom right).