Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comparative Accuracy of Lower Limb Osteometry from CT Scans Versus Direct Bone Measurements

Tooba Siddiqui
Cranfield University, Defence Academy of the United Kingdom Shrivenham
Elisa Baldini
East London Forensic Centre
Keiran Kelly
East London Forensic Centre
Claire Carter
East London Forensic Centre
Peter Zioupos
University of Hull
Mark Viner
Cranfield University, Defence Academy of the United Kingdom Shrivenham
Nicholas Márquez-Grant
n.marquezgrant@arch.oxon.net
University of Oxford

Article

Keywords: forensic anthropology, osteoarcheology, UK, biological profile, stature, long bone length, human identification

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11250517/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Radiographic imaging is increasingly used for osteometric data collection in clinical and forensic settings, although measurement accuracy may be affected by imaging protocols and scanning parameters. This study evaluated the agreement between CT-derived lower-limb skeletal measurements and direct dry-bone measurements. Maximum (MLF) and bicondylar (BLF) femoral lengths and maximum (LMT) and malleolar-trochlear (MTL) tibial lengths were recorded from 68 femora and 49 tibiae from anatomical collections using an osteometric board, and through CT images of the same bones. Agreement was assessed using paired comparisons, absolute-agreement intraclass correlation coefficients (ICCs), Bland–Altman analysis, technical error of measurement (TEM), and relative TEM (rTEM). There were statistically significant differences between CT-derived measurements and direct measurements for MLF, BLF, LMT, and MTL; however, mean biases were small, ranging from − 0.46 to 0.76 mm. No significant proportional bias was detected. Intra-observer rTEM ranged from 0.10% to 0.41%, with ICCs of 0.9936–0.9993, while inter-observer rTEM ranged from 0.11% to 0.23%, with ICCs of 0.9975–0.9992. CT-derived femoral and tibial measurements demonstrated excellent agreement with direct dry-bone osteometry. Although small statistically significant method differences were present, their magnitude was unlikely to be practically meaningful. In summary, measuring from CT images provided precise, repeatable, and reproducible dry skeletal measurements and represents a reliable non-invasive alternative for tibia and femora at autopsy, when direct skeletal measurement is unavailable or impractical.

Introduction

Radiographic images such as computed tomography (CT) scans are a well-studied tool for clinical analysis of patients and is also used routinely for post-mortem analysis [1–4]. Its usage has also been applied to forensic anthropology casework and research [5–9]. CT-based osteometry offers a highly practical, non-destructive method for measuring lower limb bones [5–9]. These are utilised in osteometric research fields such as osteoarchaeology, forensic anthropology and orthopaedics as well as in patient care [1–9]. Specifically, within forensic anthropology and osteoarchaeology, femur and tibia are two of the most reliable bones for stature estimation, as well as aiding sex estimation, which are critical criterions of a person's biological profile, necessary for identification of human remains [10–15]. Traditional methods require these measurements to be taken directly from dry bones using osteometric boards; however, advent of CT imaging allows for bones to be measured virtually, allowing for greater accessibility; for instance, living and recently deceased individuals as well as remains that cannot undergo autopsy or physically handled, can also be assessed [2, 6–8].

The standards of osteometry have long been designed to be performed directly on the dry bones preferably on an osteometric board [2, 10], without the need for software if they were CT scanned. Traditional osteometry is also more simplistic and does not require the handler to have knowledge of using specialist software for reading and visualising the radiographic images [2, 8].

Meanwhile, CT scanners have high bone-soft tissue contrast allowing clear imaging of bones in 'wet' environments without needing to remove the bones from the body [2, 15], thus avoiding any invasive autopsy or dissection. Furthermore, CT scan technology enables 3D reconstruction of the bones, which is useful for building models of the patient for anatomical and biomechanical research, as well as for detailed trauma analysis [15]. This technology also allows for cadaver preservation or completeness as invasive extraction of the bones during autopsy is not required [2, 15]. Furthermore, radiographic images can be stored and studied retrospectively without the need to exhume the individual [15]. Overall CT imaging is crucial for when direct measurement is not possible or to minimise invasive autopsy, for example in fresh cadavers, burnt bodies or mummified remains.

However, accuracy of measurements derived from radiographic images depends on slice thickness, voxel size, reconstruction kernel, thresholding, segmentation method, and whether the bone was scanned with or without soft tissue [16–18]. While CT-derived measurements have been shown to be highly precise, a small average difference between CT and direct measurement does not automatically prove interchangeability [2, 15–18]. Further validation for the specific bone, landmark, and measurement type is necessary before making any judgement on this [2, 15–18].

This study aims to evaluate whether CT-derived lower limb skeletal measurements can be treated as interchangeable with dry-bone measurements, specifically for the femur and tibia which is often used in forensic anthropology for stature estimation, orthopaedics and anatomical research.

Materials and Methods

Study Design

A total of 234 measurements were taken using a historical anatomical sample of 68 femora and 49 tibiae. The sample was sourced from Cranfield University's Osteological Collection, and the scanning procedures took place at East London Forensic Centre, where scans were taken by the Lead Radiographer (KK). The bones analysed in this research belonged to individuals who had been deceased for over 100 years, so they are classified as archaeological; therefore, the analysis did not require consent under the Declaration of Helsinki or the Human Tissue Act. Furthermore, their identities are unknown, and antemortem data is unavailable; they are therefore anonymous. Biological sex, age at death, provenance, and population affinity were, therefore, also unknown. The exclusion criteria removed any bones that were incomplete or damaged at the epiphyses, altered by heat, or with any pathology, including congenital malformations, rickets, osteoarthritis, or trauma, that would affect stature estimation.

The scans were taken with a GE Revolution™ Ascend 64-Slice with a slice thickness of 0.6mm. The Anterior-Posterior (AP) scout images were obtained at 120kV and 20 mA (AP scout plane 180 degrees). The scans were then imported onto Horos (v4.01), a free and open-source software (FOSS) program and visualised in 2D view.

Variables and Measurement Protocol

Long-bone measurements typically utilised for stature estimation were selected. These were the v. The term MTL is taken from Jantz et al.'s [19] definition of Trotter's measurement of the tibia, and therefore is meant to replicate Trotter and Gleser's measurements. Direct measurements were taken by positioning the bones on a calibrated osteometric board according to the standard measurement protocol for long bones for stature estimation (see Appendix 1 for full definitions of all measurements taken).

The long bones which were later CT scanned with the GE Revolution™ Ascend 64-Slice, were measured using Horos' perpendicular measuring tool on multiplanar reformats. Consistent positioning and windowing were used to optimise cortical boundary visualisation. Figure 1 shows a MLF measurement taken from one CT image using the same scanner and protocols from a full body PMCT, where MLF measurement was labelled as B-C. Measurements were taken in centimetres. Scans were kept in 2D view as converting to 3D was found to produce more inaccuracies according to [8, 20–22], where measurement differences exceeded the <2mm acceptable range as expressed in [8, 20–22]. Therefore, it was not deemed necessary to add markers to replicate soft tissue, as image was not expanded to a 3D model. Furthermore, depicting soft tissue depth required knowledge of antemortem data, which was not known so adding soft tissue replacement would introduce new inaccuracies.

The long bones, which were later CT scanned with the GE Revolution™ Ascend 64-Slice, were measured using Horos' perpendicular measuring tool on multiplanar reformats. Consistent positioning and windowing were used to optimise cortical boundary visualisation. Figure 1a shows a MLF measurement taken from one CT image of a series of more than a hundred scans we produced in a previous study [23] in which we used the same scanner and protocols from a full body PMCT, and where MLF measurement is labelled as B-C. Measurements were taken in centimetres. In the present study, scans and were kept in 2D view as converting to 3D would have evidently produced more inaccuracies according to [8, 20–22]. Figure 1b shows the same measurement taken from dry bone using an osteometric board; this image was taken from Vasilopoulos et al. [24]. Please note that in [24] this measurement is called 'Femoral Maximum Length' (FML) but in this present study we call it 'Maximum Femoral Length' (MFL) to be consistent with the existing data collection procedures for forensic anthropologists [24–26].

Figure 1 – (a) MLF measurement taken from one CT image of a series of more than a hundred scans we produce in a previous study [23] in which the same scanner and protocols were used (labelled B-C). (b) Image taken from Vasilopoulos et al. [24] similar to our physical measurements we performed in the current study.

All measurements were exported to Microsoft Excel (Office 365) prior to statistical analysis. Raw data from the Excel file was loaded into Python (3.10) using pandas. Analyses were conducted in a Jupyter environment (macOS/Unix), and the primary scientific stack included: numpy, pandas, scikit-learn, seaborn, and matplotlib.

Descriptive Statistics

Statistical analyses were conducted to assess agreement between osteometric data from the CT-derived models and direct dry-bone measurements. Analyses were performed separately for each measurement and the direct dry-bone measurement was treated as the reference value. Differences between methods were calculated as: CT-derived measurement minus direct dry-bone measurement.

Descriptive statistics were calculated for each measurement method, including the mean, standard deviation, minimum, and maximum values. The mean difference between CT-derived and direct measurements was calculated for each osteometric variable to assess the presence and direction of measurement bias. The distribution of paired differences was assessed for normality. Where differences were approximately normally distributed, paired samples t-tests were used to test whether the mean difference between methods differed significantly from zero. Where the assumption of normality was not met, Wilcoxon signed-rank tests were used. Statistical significance was set at 0.05 for paired t-tests, Wilcoxon signed-rank tests and proportional-bias tests, where $p < o.o5$ was considered statistically significant, and $p \geq 0.05$ was considered not statistically significant.

Correlation Analysis

Agreement between CT-derived and direct dry-bone measurements was assessed using intraclass correlation coefficients (ICC). ICC values were calculated separately for each measurement variable using an absolute-agreement model, as the aim was to determine whether CT-derived measurements could be used interchangeably with direct osteometric measurements. ICC values below 0.50 were interpreted as poor agreement, values between 0.50 and 0.75 as moderate agreement, values between 0.75 and 0.90 as good agreement, and values above 0.90 as excellent agreement.

Bland-Altman analysis was used to further assess agreement between the two methods. For each measurement, the difference between CT-derived and direct dry-bone measurements was plotted against the mean of the two methods. The mean difference was used to represent measurement bias, and 95% limits of agreement were calculated as the mean difference ± 1.96 standard deviations of the differences. Bland-Altman plots were examined for systematic bias, proportional bias, and outlying observations. Proportional bias was considered where the magnitude or direction of the difference between methods appeared to vary according to bone size.

Experimental Error and Reliability

The scans were taken by an experienced radiographer (KK) and measurements from the scans were taken by the first author (TS), as well as an experienced anatomical pathology technologist with medical training (EB). Measurement precision was addressed using the Horos perpendicular measuring tool, and the bilateral averaging reduced measurement noise. Experimental error indices (technical error of measurement – TEM and relative technical error of measurement – rTEM%) and reliability (ICC patterns)

were within acceptable ranges for similar anthropometric work, with no indication that measurement noise was driving the observed variance.

Equations as follows:

$$TEM = \sqrt{\frac{\Sigma\, d^2}{2N}}$$

$$\%TEM\ = \frac{TEM}{m} \times\ 100$$

Furthermore, the workflow (PMCT in Horos, consistent multiplanar reconstructions - MPRs, and perpendicular tool) yields reliable data for descriptive and group-level comparisons. Therefore, measurement precision was deemed adequate.

In addition, to assess measurement error, approximately 20% of the skeletal sample was randomly selected for repeat measurements. This subsample consisted of 24 bones (14 femora, 10 tibiae). For intra-observer error, the primary observer (TS) repeated all measurements after a time interval of two weeks without reference to the original results. The TEM and rTEM% were calculated for each variable.

Meanwhile, for inter-observer error, an independent second observer (P2) repeated all of the measurements originally taken by the primary observer (P1) on the same subset of the sample. For each measurement and side, the difference between observers was calculated (P2 – P1), as well as the mean of the two measurements for every specimen. Inter-observer TEM was computed from the inter-observer differences as the square root of the summed squared differences divided by twice the number of paired observations, with %TEM expressed relative to the mean of the paired measurements. Agreement between observers was further assessed using Bland–Altman plots, showing the mean difference (bias) and 95% limits of agreement for each variable. All calculations and graphics were produced in Python.

Results

Descriptive Statistics

Descriptive statistics (Table 1) showed close correspondence between CT-derived and direct dry-bone measurements across all osteometric variables. Mean CT-derived measurements differed from direct dry-bone measurements by less than 1 mm for all variables. Mean dry-bone and CT-derived values were 43.07 and 43.12 cm for MLF, 42.74 and 42.81 cm for BLF, 36.13 and 36.09 cm for LMT, and 35.39 and 35.35 cm for MTL, respectively.

Table 1
– Summary of descriptive statistics for 68 femora and tibiae

Bone	Measurement	Dry mean ± SD (in cm)	Dry range (in cm)	CT mean ± SD (in cm)	CT range (in cm)
Femur	MLF	43.07 ± 2.74	37.5–50.4	43.12 ± 2.74	37.57–50.45
Femur	BLF	42.74 ± 2.72	37.4–50.0	42.81 ± 2.72	37.51–50.05
Tibia	LMT	36.13 ± 2.24	30.7–41.0	36.09 ± 2.23	30.65–40.96
Tibia	MTL	35.39 ± 2.19	30.0–39.9	35.35 ± 2.21	29.83–39.87

Correlation Analysis

Paired comparisons (Table 2) identified statistically significant differences between CT-derived and direct dry-bone measurements for MLF (mean difference 0.44 mm, $p < 0.001$), BLF (0.76 mm, $p < 0.001$), LMT (-0.35 mm, $p = 0.014$), and MTL (-0.46 mm, $p < 0.001$). Despite statistical significance, all mean differences were less than 1 mm. Absolute-agreement ICCs ranged from 0.9989 to 0.9993, indicating excellent agreement. Bland–Altman analysis showed small directional biases, with CT values slightly larger for femoral variables and slightly smaller for tibial variables. The 95% limits of agreement were −1.35 to 2.22 mm for MLF, -1.09 to 2.62 mm for BLF, -2.26 to 1.55 mm for LMT, and − 1.82 to 0.89 mm for MTL. No significant proportional bias was detected.

Table 2
– Paired comparison and absolute-agreement ICCs (Mean difference was calculated as CT-derived measurement minus direct dry-bone measurement.)

Measurement	Mean difference – cm (95% CI)	Mean difference (mm)	Normality p	Test	Test statistic	p-value	ICC (95% CI)
MLF	0.044 (0.021, 0.066)	0.44	0.31	Paired t-test	3.936	< 0.001	0.9993 (0.9989, 0.9996)
BLF	0.076 (0.053, 0.099)	0.76	< 0.001	Wilcoxon signed-rank	300.5	< 0.001	0.999 (0.9985, 0.9993)
LMT	-0.035 (-0.063, -0.007)	-0.35	0.039	Wilcoxon signed-rank	395.5	0.014	0.9989 (0.9981, 0.9994)
MTL	-0.046 (-0.066, -0.026)	-0.46	0.109	Paired t-test	-4.671	< 0.001	0.9993 (0.9988, 0.9995)

Bland–Altman statistics (Fig. 3 and Table 3) were calculated using signed CT minus dry-bone differences. Mean absolute differences ranged from 0.69 mm for MTL to 1.04 mm for BLF, and maximum absolute differences did not exceed 2.00 mm. Between 92.65% and 100% of paired measurements were within 2 mm.

The results of the Bland-Altman Agreement analysis between measurements derived from CT scans in comparison with measurements taken from dry bones are presented below in Table 3.

Table 3
– Bland-Altman Agreement summary

Measurement	Bias (mm)	95% limits of agreement (mm)	Proportional bias p	Mean absolute difference (mm)	Maximum absolute difference (mm)
MLF	0.44	-1.35 to 2.22	0.874	0.83	2.0
BLF	0.76	-1.09 to 2.62	0.928	1.04	2.0
LMT	-0.35	-2.26 to 1.55	0.407	0.87	2.0
MTL	-0.46	-1.82 to 0.89	0.053	0.69	1.9

Furthermore, the absolute difference between the measurements can be seen in Fig. 4. Here, positive and negative discrepancies are treated equivalently, and reference lines are included at 1 mm and 2 mm. CT measurements for the femoral measurements were minimally larger than osteometric board measurements, whilst tibial measurements were slightly smaller in the CT images compared to osteometric board measurements.

Experimental Error and Reliability

The relative TEM analysis across dry-bone and CT measurements for intra- and inter-observer error is presented in Fig. 5. All rTEM values were below 0.5%.

Intra-observer analysis (Table 4) demonstrated excellent repeatability for dry-bone and CT-derived measurements. TEM values ranged from 0.43 to 1.48 mm, rTEM values ranged from 0.10% to 0.41%, and ICCs ranged from 0.9936 to 0.9993. Only dry-bone BLF showed a nominally significant repeat difference (bias − 0.50 mm, p = 0.040); its magnitude was small, and reliability remained excellent.

Table 4
– Intra-observer measurement error and reliability

Measurement	Modality	TEM, mm	rTEM, %	ICC (95% CI)	Bias, mm (95% CI)	95% LoA, mm	p-value
MLF	Dry	1.0	0.242	0.9956 (0.9904, 0.9975)	-0.29 (-1.12, 0.54)	-3.1 to 2.53	0.324
	CT	0.461	0.111	0.9991 (0.9974, 0.9997)	0.22 (-0.15, 0.59)	-1.03 to 1.47	0.215
BLF	Dry	0.779	0.19	0.9977 (0.9945, 0.9991)	-0.5 (-1.09, 0.09)	-2.5 to 1.5	0.04
	CT	0.429	0.104	0.9993 (0.9984, 0.9997)	-0.15 (-0.5, 0.2)	-1.35 to 1.05	0.374
LMT	Dry	1.483	0.408	0.9936 (0.9783, 0.9994)	-0.2 (-1.77, 1.37)	-4.51 to 4.11	0.875
	CT	0.579	0.159	0.9988 (0.9959, 0.9996)	0.08 (-0.53, 0.69)	-1.6 to 1.76	0.775
MTL	Dry	0.5	0.141	0.999 (0.9975, 0.9995)	-0.1 (-0.63, 0.43)	-1.55 to 1.35	1.0
	CT	0.717	0.202	0.9981 (0.9936, 0.9992)	-0.04 (-0.8, 0.72)	-2.13 to 2.05	0.908

Inter-observer analysis (Table 5) also demonstrated excellent reproducibility for both modalities. TEM values ranged from 0.46 to 0.82 mm, rTEM values ranged from 0.11% to 0.23%, and ICCs ranged from 0.9975 to 0.9992. No significant inter-observer differences were detected. Biases ranged from − 0.35 to 0.29 mm, indicating no meaningful systematic disagreement between observers.

Table 5
Inter-observer measurement error and reliability

Measurement	Modality	TEM, mm	rTEM, %	ICC (95% CI)	Bias, mm (95% CI)	95% LoA, mm	p-value
MLF	Dry	0.655	0.158	0.9981 (0.9961, 0.9987)	-0.29 (-0.81, 0.24)	-2.08 to 1.51	0.09
	CT	0.603	0.145	0.9985 (0.9961, 0.9993)	-0.19 (-0.69, 0.3)	-1.88 to 1.5	0.417
BLF	Dry	0.655	0.159	0.9984 (0.9962, 0.9992)	0.29 (-0.24, 0.81)	-1.51 to 2.08	0.263
	CT	0.464	0.113	0.9992 (0.9975, 0.9998)	0.01 (-0.38, 0.41)	-1.32 to 1.35	0.939
LMT	Dry	0.632	0.176	0.9983 (0.994, 0.9996)	-0.2 (-0.86, 0.46)	-2.0 to 1.6	0.509
	CT	0.7	0.195	0.9978 (0.9935, 0.9991)	-0.14 (-0.88, 0.6)	-2.16 to 1.88	0.678
MTL	Dry	0.707	0.199	0.9981 (0.9938, 0.9991)	0.2 (-0.54, 0.94)	-1.82 to 2.22	0.359
	CT	0.818	0.23	0.9975 (0.9906, 0.9991)	-0.35 (-1.18, 0.48)	-2.63 to 1.93	0.366

Discussion

This research assessed the agreement between osteometric measurements obtained from CT scans and direct dry-bone measurements. This attempted to answer the question on how reliable long bone measurements are when taken from CT images as opposed to direct measurements taken from dry bones with a osteometric board. Overall, the findings demonstrate excellent agreement between the two methods across all four osteometric variables examined. Although statistically significant differences were identified between CT-derived and dry-bone measurements, their magnitude was minimal, with mean differences remaining below 1 mm for each variable.

The descriptive statistics showed close equivalence between CT-derived and direct dry-bone measurements. Mean values for each CT-derived measurement were very similar to the corresponding dry-bone reference values. CT-derived femoral measurements were slightly larger, whereas CT-derived

tibial measurements were slightly smaller. These directional differences were minimal, ranging from −0.46 mm for MTL to 0.76 mm for BLF, suggesting that systematic bias introduced by the CT workflow was small.

Although paired statistical testing identified significant differences for all four measurements, these findings should be interpreted cautiously. With precise measurements and moderate sample sizes, very small differences can reach statistical significance even when their practical effect is negligible [27–28]. In this study, all mean differences were less than 1 mm. In a previous study [29] assessing the usage of computer tomography within the field of forensic anthropology, the acceptable error between two measurements was set to 2mm, while another study [5] reported deviations up to 5mm, which strongly supports the allowance of a mean difference of 1 mm as negligible. Therefore, while CT-derived and dry-bone measurements were not numerically identical, the magnitude of disagreement was limited and CT-derived measurements closely approximated direct osteometric measurements.

The ICC results provide strong support for agreement between CT-derived and direct dry-bone measurements. Absolute-agreement ICC values ranged from 0.9989 to 0.9993, representing excellent agreement. Because an absolute-agreement model was used, these results indicate not only that the methods ranked specimens similarly, but also that their values were nearly identical in scale. This is important because high correlation alone would not establish interchangeability if one method consistently overestimated or underestimated the other [30–31]. The observed ICCs indicate that CT-derived osteometry reproduced direct dry-bone measurements with a high degree of reliability [30–31].

The Bland–Altman analysis further supported this conclusion. Mean biases ranged from − 0.46 to 0.76 mm, and the 95% limits of agreement remained narrow relative to overall bone length. No significant proportional bias was detected for MLF, BLF, LMT, or MTL, although MTL approached the significance threshold ($p = 0.053$). This suggests that CT-derived measurement error did not systematically increase or decrease with bone size and that performance was consistent across the observed measurement range [30–32].

The practical error analysis also demonstrated strong agreement. Mean absolute differences ranged from 0.69 to 1.04 mm, and maximum absolute differences did not exceed 2 mm. Between 92.65% and 100% of paired measurements were within 2 mm. This is particularly relevant for applied osteometric work, where the practical question is not simply whether two methods are statistically identical, but whether discrepancies are sufficiently small for research or casework [33–34]. The present findings therefore indicate that CT-derived lower-limb measurements are sufficiently accurate for applications in which small millimetric deviations are unlikely to alter interpretation.

The experimental error results further strengthen confidence in the measurement protocol. Intra-observer rTEM values ranged from 0.10% to 0.41%, with ICCs between 0.9936 and 0.9993. Inter-observer rTEM values ranged from 0.11% to 0.23%, with ICCs between 0.9975 and 0.9992. No significant inter-observer differences were observed for either dry-bone or CT-derived measurements. These findings

indicate that the measurement definitions were repeatable and reproducible, and that observer-related measurement noise was unlikely to account for the CT–dry differences.

The reliability findings are also important in relation to the imaging workflow. The use of PMCT data, consistent multiplanar reconstructions, and the Horos perpendicular measuring tool produced precise and reproducible measurements. Together, these features likely contributed to the high agreement between CT-derived and dry-bone measurements. The results therefore support the use of CT scans for extracting reliable osteometric data, particularly when direct access to skeletal remains is limited or when preservation, ethical, or logistical considerations make direct measurement impractical.

These findings have broader implications for osteological and forensic practice. CT-derived models offer a non-destructive alternative to direct skeletal measurement and allow digital preservation, re-measurement, and data sharing [1–9, 33]. If CT-derived measurements closely reproduce dry-bone values, as shown here, they can be used to expand osteometric datasets and support analyses where physical remains are unavailable or cannot be handled extensively [1–9, 33]. This is especially relevant for forensic anthropology, bioarchaeology, and museum-based research, where skeletal collections may be fragile, incomplete, restricted, or subject to conservation concerns [1–9].

However, some limitations of this study were also identified. First, although agreement was excellent, the study assessed a limited set of measurements from femora and tibiae. The results should therefore not automatically be generalised to all skeletal elements or more complex anatomical measurements. Linear measurements taken between well-defined landmarks are likely to be more reproducible than measurements involving less distinct landmarks or curved surfaces.

Another limitation is that the analysis focused on agreement between CT-derived and dry-bone measurements, but did not explicitly test how scanner parameters, reconstruction settings, segmentation decisions, or image resolution might influence measurement accuracy. These factors may affect the precision of CT-derived osteometry, particularly for smaller elements or more subtle anatomical landmarks. The present results therefore support the reliability of the specific workflow used here, but additional validation would be useful before applying the method across different scanners, reconstruction protocols, or software environments. In the context of the current study, dry tibiae and femora were scanned, while in real forensic and archaeological cases bones are often found as a part of the whole body with soft tissue and other artefacts, bodies are also found in varying anatomical positions; none of this was studied in the current study and can be the focus of future studies.

Despite these limitations, the findings provide strong evidence that CT-derived osteometric measurements closely approximate direct dry-bone measurements. The combination of sub-millimetric mean biases, excellent ICCs, narrow Bland–Altman limits of agreement, low absolute error, and strong intra- and inter-observer reliability indicates that the CT workflow produced valid and reproducible measurements. Although statistically significant method differences were observed for all variables, these differences were minor and unlikely to compromise biological or anthropological interpretation.

In conclusion, CT-derived measurements of the femur and tibia showed excellent agreement with direct dry-bone measurements under the imaging parameters used in this study. The observed differences were small, measurement reliability was high, and no meaningful evidence suggested that measurement error was driving the results. These findings support the use of CT-derived osteometry as a reliable alternative to direct dry-bone measurement for descriptive and group-level analyses, particularly when direct skeletal access is limited or when non-destructive digital methods are preferred.

Abbreviations

Abbreviation	Meaning
CT	Computed Tomography
BLF	Bicondylar Length of Femur
ICC	Intraclass Correlation Coefficient
ICCs	Intraclass Correlation Coefficients
MLF	Maximum Length of Femur
MPRs	Multiplanar Reformations
MTL	Tibia Malleolar-Trochlear Length
LMT	Maximum Length of Tibia
P1	Primary Observer
P2	Second Observer
PMCT	Postmortem Computed Tomography
SD	Standard Deviation
TEM	Technical Error of Measurement
rTEM	Relative Technical Error of Measurement

Declarations

CRediT authorship contribution statement

Acknowledgments

The authors would like to express their sincere gratitude to the staff at East London Forensic Centre for providing the CT scanning technology, and to Professor Andrew Shortland at Cranfield University for providing supervision to the first author in their overarching research.

Tooba Siddiqui: Conceptualisation, Methodology, Formal analysis, Investigation, Writing – original draft, visualisation. **Peter Zioupos**: Writing – review & editing, Supervision. **Keiran Kelly:** Data curation. **Elisa Baldini:** Validation. **Nicholas Marquez-Grant**: Conceptualisation, Writing – review & editing, Supervision.

Data Availability Statement

Data supporting this study are included within the article and/or supporting materials.

Declaration of Competing Interest

The authors declare no competing interests.

Declaration of Funding

None.

Ethical Statement

Ethical approval was gained from Cranfield University (CURES/26951/2026) on 10th January 2026 prior to the start of the research. The bones analysed in this research belonged to individuals who had been deceased for over 100 years, so they are classified as archaeological; therefore, the analysis did not require consent under the Declaration of Helsinki or the Human Tissue Act.

References

1. Guggenberger, R., Pfirrmann, C. W. A., Koch, P. P. & Buck, F. M. Assessment of lower limb length and alignment by biplanar linear radiography: comparison with supine CT and upright full-length radiography. *Am. J. Roentgenol.* **202** (2), W161–W167. 10.2214/ajr.13.10782 (2014).
2. Carew, R. M., Viner, M. D., Conlogue, G., Márquez-Grant, N. & Beckett, S. Accuracy of computed radiography in osteometry: A comparison of digital imaging techniques and the effect of magnification. *J. Forensic Radiol. Imaging*. **19**, 100348. 10.1016/j.jofri.2019.100348 (2019).
3. Stephen, J. M., Calder, J. D., Williams, A. & El Daou, H. Comparative accuracy of lower limb bone geometry determined using MRI, CT, and direct bone 3D models. *J. Orthop. Res.* **39** (9), 1870–1876. 10.1002/jor.24923 (2020).
4. Banik, S. et al. Comparative morphometry of the sacrum and its clinical implications: a retrospective study of osteometry in dry bones and CT scan images in patients presenting with lumbosacral pathologies. *Cureus*; (2022). 10.7759/cureus.22306
5. Colman, K. L. et al. Virtual forensic anthropology: The accuracy of osteometric analysis of 3D bone models derived from clinical computed tomography (CT) scans. *Forensic Sci. Int.* **304**, 109963.

10.1016/j.forsciint.2019.109963 (2019).

6. Kira, K. et al. Stature estimation by semi-automatic measurements of 3D CT images of the femur. *Int. J. Legal Med.* **137** (2), 359–377. 10.1007/s00414-022-02921-y (2022).
7. Gaya-Sancho, B. et al. Comparison of measurements made on dry bone and digital measurements in Anatomage for the sacral bone in a Spanish population. *Sci. Rep.* **13** (1), 20578. 10.1038/s41598-023-48013-8 (2023).
8. Stull, K. E., Tise, M. L., Ali, Z. & Fowler, D. R. Accuracy and reliability of measurements obtained from computed tomography 3D volume rendered images. *Forensic Sci. Int.* **238**, 133–140. 10.1016/j.forsciint.2014.03.005 (2014).
9. Martin-Champetier, A. et al. Optimizing CT scan parameters for dry archaeological bones: a qualitative study as a first step towards standardizing CT acquisition protocols. *Int. J. Osteoarchaeology*. **35**, e3390. 10.1002/oa.3390 (2025).
10. Spradley, M. K. Metric methods for the biological profile in forensic anthropology: sex, ancestry, and stature. *Acad. Forensic Pathol.* **6** (3), 391–399. 10.23907/2016.040 (2016).
11. Howley, D., Howley, P. & Oxenham, M. F. Estimation of sex and stature using anthropometry of the upper extremity in an Australian population. *Forensic Sci. Int.* **287**, 220e1–220e10. 10.1016/j.forsciint.2018.03.017 (2018).
12. Saco-Ledo, G., Porta, J., Duyar, I. & Mateos, A. Stature estimation based on tibial length in different stature groups of Spanish males. *Forensic Sci. Int.* **304**, 109973. 10.1016/j.forsciint.2019.109973 (2019).
13. Scientific Working Group for Forensic Anthropology. Standard for stature estimation in forensic anthropology. *AAFS Stand. Board*, 1. (2019). https://www.aafs.org/sites/default/files/media/documents/045_Std_e1_0.pdf
14. Siddiqui, T. & Zioupos, P. Márquez-Grant, N. Stature estimation in forensic anthropology: addressing the current status, challenges and future. *Forensic Sci.* **6** (1), 23. 10.3390/forensicsci6010023 (2026).
15. Demehri, S. et al. Musculoskeletal CT imaging: state-of-the-art advancements and future directions. *Radiology* **308** (2). https://doi.org/10.1148/radiol.230344 (2023).
16. Engelkes, K. Accuracy of bone segmentation and surface generation strategies analyzed by using synthetic CT volumes. *J. Anat.* **238** (6), 1456–1471. 10.1111/joa.13383 (2020).
17. Ahmed, M. et al. Exploring CT pixel and voxel size effect on anatomic modeling in mandibular reconstruction. *3D Print. Med.* **10** (1). 10.1186/s41205-024-00223-0 (2024).
18. Ogden, K. M. et al. Factors affecting dimensional accuracy of 3-D printed anatomical structures derived from CT data. *J. Digit. Imaging*. **28** (6), 654–663. 10.1007/s10278-015-9803-7 (2015).
19. Jantz, R. L., Jantz, M. & Ousley, S. D. L., Measuring the Tibia: Trotter's error revisited. *Journal of Forensic Sciences*. *65*(6), 2094–; (2097). 10.1111/1556-4029.14550 (2020).

20. Decker, S. J., Hoegstrom, E. J. & Hilbelink, D. R. Virtual anatomy: three-dimensional computer modeling and measurement of human cranial anatomy. American Academy of Forensic Sciences. Colorado Springs, CO, 312 (2008).
21. Robinson, C. et al. Anthropological measurement of lower limb and foot bones using multi-detector computed tomography. *J. Forensic Sci.* **53** (6), 1289–1295. 10.1111/j.1556-4029.2008.00875.x (2008).
22. Verhoff, M. A. Digital forensic osteology–possibilities in cooperation with the Virtopsy project. *Forensic Science International. 174(2–3)*, 152–156; (2008). 10.1016/j.forsciint.2007.03.017
23. Siddiqui, T. et al. Reassessing stature estimation in modern Britain: A PMCT-based evaluation and alternative to Trotter and Gleser in forensic anthropology casework. *Forensic Sci. Int.* **388**, 113078. 10.1016/j.forsciint.2026.113078 (2026).
24. Vasilopoulos, A. et al. Odontoid process and femur: a novel bond in anatomy. *Cureus* **12** (3), e7372. 10.7759/cureus.7372 (2020).
25. Langley, N. R. et al. Data for validation of osteometric methods in forensic anthropology. *Data Brief.* **19**, 21–28. 10.1016/j.dib.2018.04.148 (2018).
26. Stevens, N. T., Steiner, S. H. & MacKay, R. J. Assessing agreement between two measurement systems: an alternative to the limits of agreement approach. *Stat. Methods Med. Res.* **26** (6), 2487–2504. 10.1177/0962280215601133 (2015).
27. Serdar, C. C., Cihan, M., Yücel, D. & Serdar, M. A. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochemia Med.* **31** (1), 27–53. 10.11613/bm.2021.010502 (2021).
28. Andrade, C. Sample size and its importance in research. *Indian J. Psychol. Med.* **42** (1), 102–103 (2020). 10.4103/IJPSYM.IJPSYM_504_19.
29. Colman, K. L. et al. The geometrical precision of virtual bone models derived from clinical computed tomography data for forensic anthropology. *Int. J. Legal Med.* **131** (4), 1155–1163. 10.1007/s00414-017-1548-z (2017).
30. Ranganathan, P., Pramesh, C. & Aggarwal, R. Common pitfalls in statistical analysis: measures of agreement. *Perspectives Clin. Research* **8**(4), 187 ; 10.4103/picr.picr_123_17
31. Liu, J. et al. Correlation and agreement: overview and clarification of competing concepts and measures. *Shanghai Archives Psychiatry*. **28** (2), 115–120. 10.11919/j.issn.1002-0829.216045 (2016).
32. Coskun, A. Bias in laboratory medicine: the dark side of the moon. *Annals Lab. Med.* **44** (1), 6–20. 10.3343/alm.2024.44.1.6 (2024).
33. Corron, L., Marchal, F., Condemi, S., Chaumoître, K. & Adalian, P. Evaluating the consistency, repeatability, and reproducibility of osteometric data on dry bone surfaces, scanned dry bone surfaces, and scanned bone surfaces obtained from living individuals. *Bulletins et Mémoires de La. Société d'Anthropologie de Paris*. **29** (1–2), 33–53. 10.1007/s13219-016-0172-7 (2016).

34. Sjovøld, T. Stature estimation from the skeleton. *Encyclopedia Forensic Sci.* 276–284. doi.org/10.1006/rwfs.2000.0770 (2000).
35. CRediT authorship contribution statement.

Figures

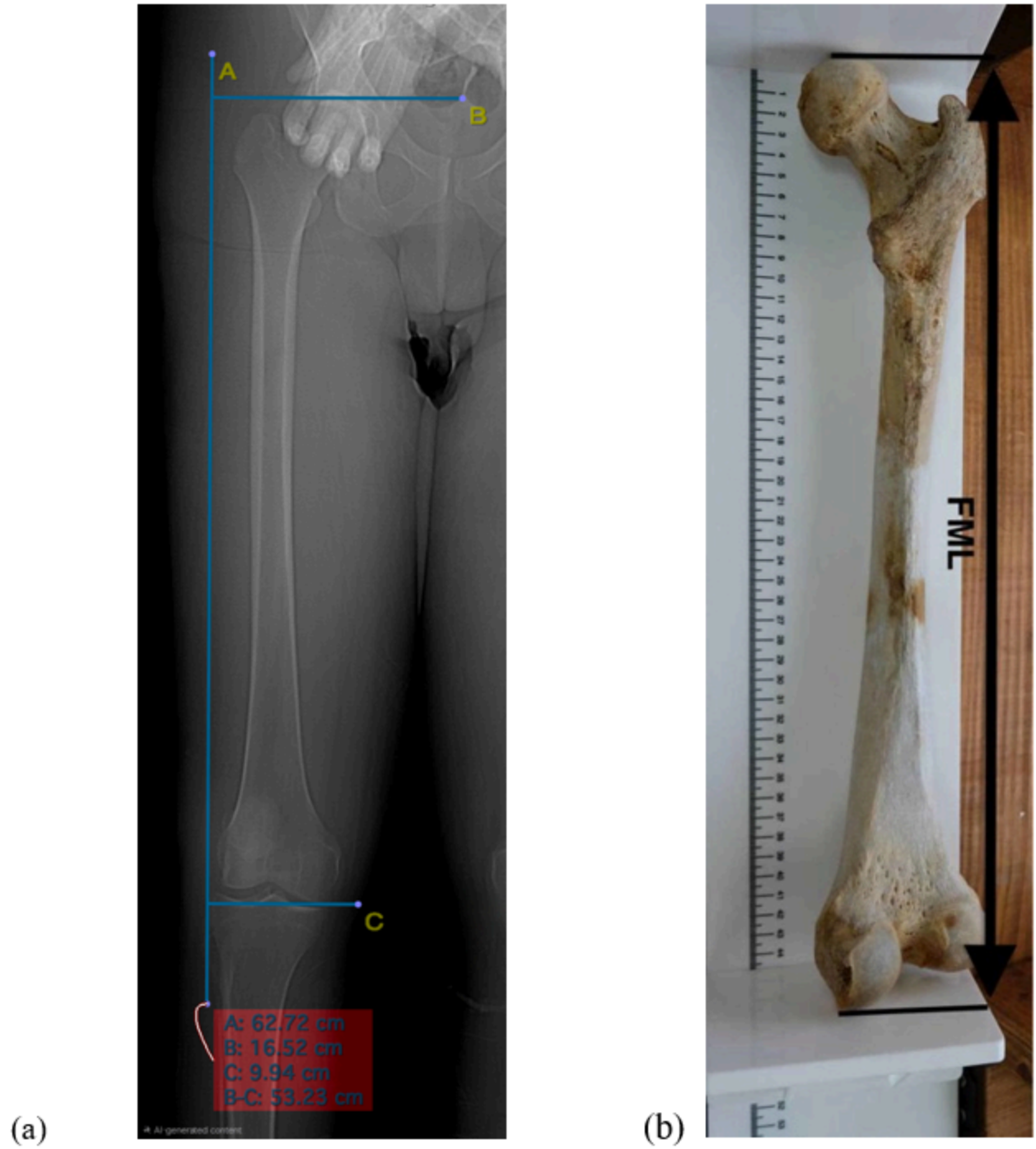


Figure 1

(a) MLF measurement taken from one CT image of a series of more than a hundred scans we produce in a previous study [23] in which the same scanner and protocols were used (labelled B-C). (b) Image taken from Vasilopoulos et al. [24] similar to our physical measurements we performed in the current study.

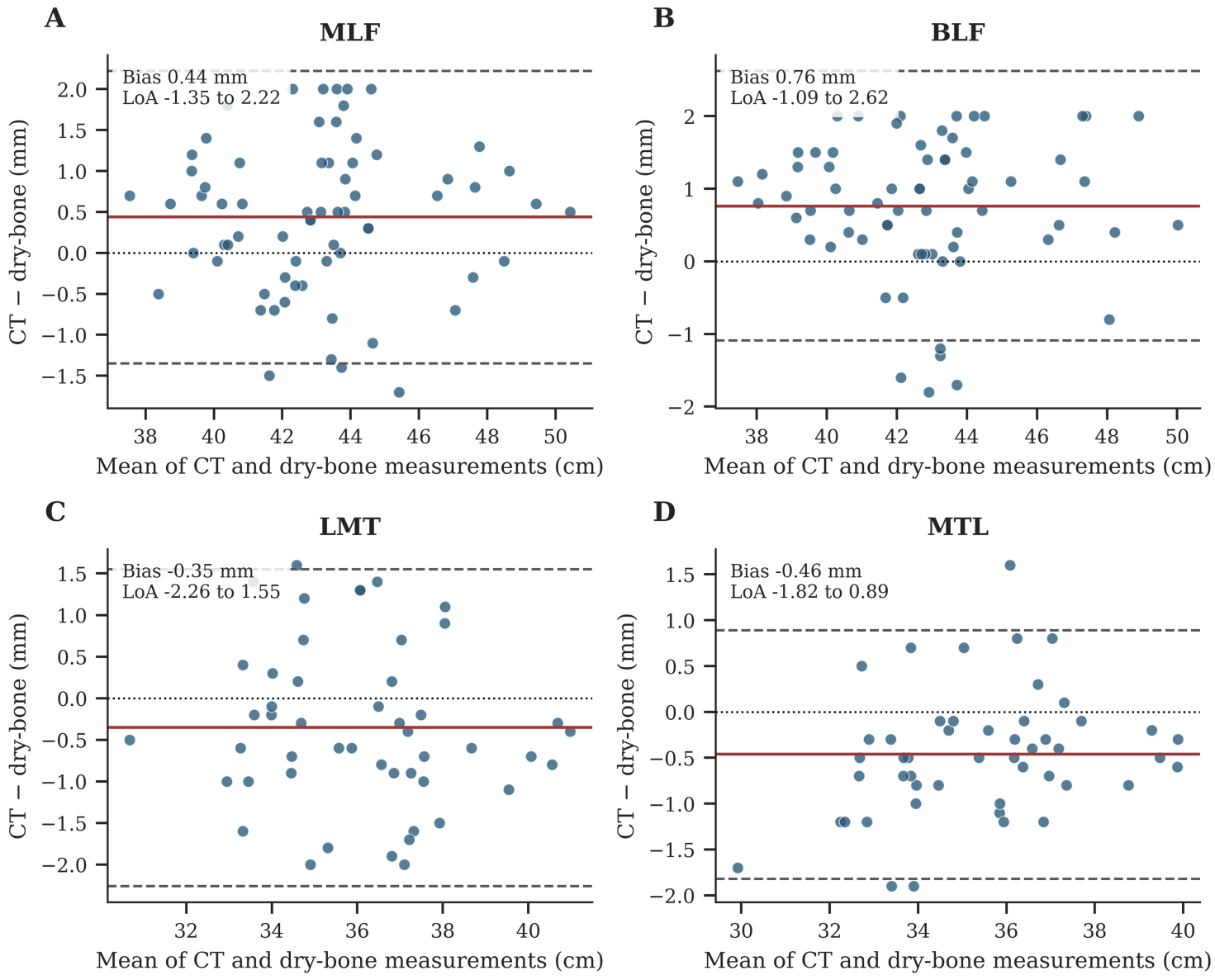


Figure 2

Figure 3 - Bland–Altman plots comparing CT-derived and direct dry-bone measurements.

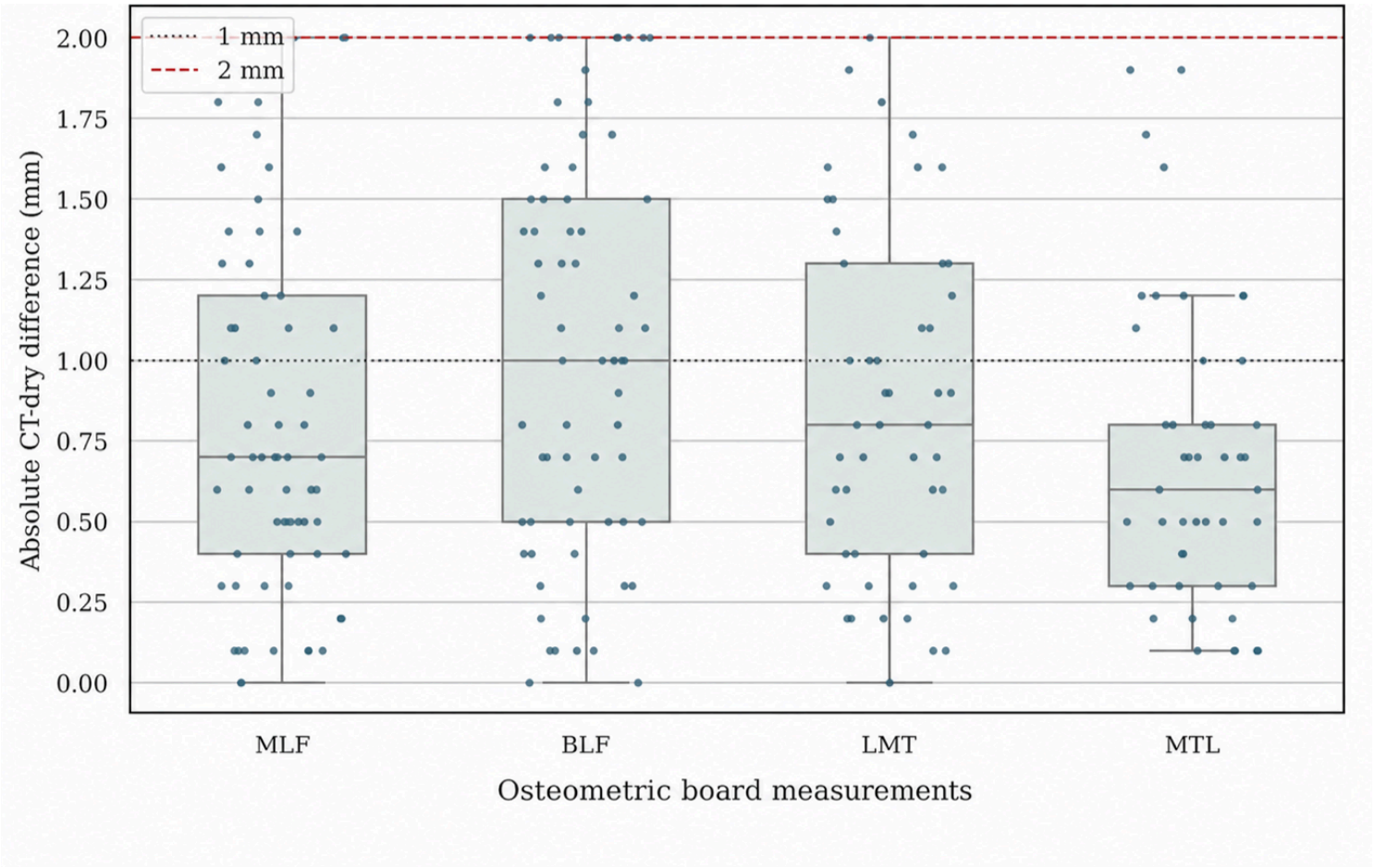


Figure 3

Figure 4 – Distribution of absolute CT–dry-bone measurement differences

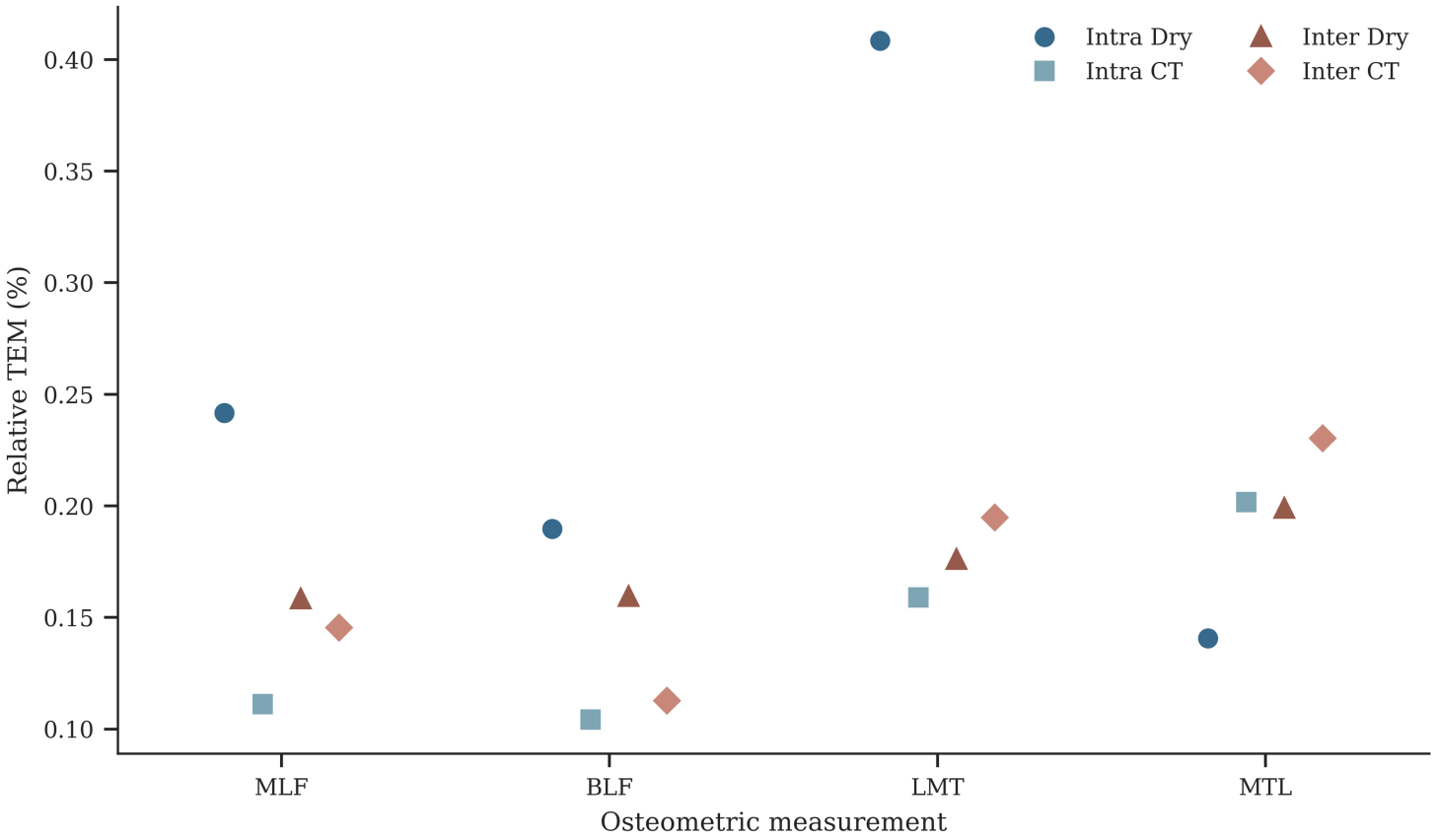


Figure 4

Figure 5 – Relative technical error of measurement for intra- and inter-observer assessments.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Paper3Data.xlsx