ARTICLE / *ARTICLE*

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

Adéquation, répétabilité et reproductibilité de données ostéométriques obtenues sur os sec, sur des surfaces virtuelles d'os sec et sur des surfaces osseuses d'individus vivants

L. Corron · F. Marchal · S. Condemi · K. Chaumoître · P. Adalian

Abstract Applying dry bone osteometrics to virtual bone surfaces obtained via medical imaging raises the question of consistency between the variables. Variables obtained from virtual bone surfaces also need to be sufficiently repeatable and reproducible to be valid for anthropological studies. This is also true for the landmarks defining these variables and for their acquisition. The consistency between variables taken directly from dry bones and from the virtual surfaces of dry bones was tested on 40 clavicles. 30 virtual surfaces of iliae, fifth lumbar vertebrae, and clavicles reconstructed from computed tomography scans of living individuals were used to test the repeatability and reproducibility of 16 landmarks and 19 variables. Statistical tests, graphical and quantitative error evaluations, and intraclass correlation coefficients were applied. The differences between all variables taken on dry and virtual clavicles were less than ±1 mm. Bland–Altman plots showed more than 95% reliability between variables obtained on dry bone and their virtually reconstructed surfaces, confirming their consistency and thus validating their use in osteometric studies independently of the medium of study. Although not all landmarks were repeatable and reproducible, most variables were. To assess intra- or inter-observer errors, graphical representations or coefficients are more precise and accurate than statistical tests. These two evaluation methods should be given priority to test the repeatability and reproducibility of osteometric variables.

Keywords Osteometrics · Landmarks · Virtual bone surfaces · Consistency · Repeatability · Reproducibility

Résumé L'application directe de mesures définies sur os sec à du matériel obtenu par imagerie médicale pose la question de l'adéquation des variables. Les variables obtenues sur surface virtuelle, les *landmarks* les définissant et leur mode d'acquisition doivent être répétables et reproductibles pour être valides et utilisables dans les études anthropologiques. L'adéquation entre variables prises directement sur os sec et sur surface virtuelle a été testée sur 40 clavicules. Les surfaces de 30 iliums, cinquièmes vertèbres lombaires et clavicules reconstruites à partir de CT scans d'individus vivants ont servi à tester la répétabilité et la reproductibilité de 16 *landmarks* et 19 variables. L'évaluation est fondée sur des tests statistiques, des outils graphiques, des paramètres quantitatifs et des coefficients de corrélation intraclasse. Les différences entre les variables sur os sec et sur leurs homologues virtuels étaient inférieures à 1 mm. Les graphiques de Bland-Altman montrent une adéquation supérieure à 95 % entre les variables. Elles sont donc valides pour des études métriques de la clavicule, quel que soit le support d'étude. La majorité des variables sont répétables et reproductibles, à l'inverse des *landmarks*. Les paramètres quantitatifs ou les représentations graphiques sont plus précis et justes que les tests statistiques pour attester de la significativité d'erreurs. Ces outils semblent à privilégier pour tester la répétabilité et la reproductibilité de variables ostéométriques.

Mots clés Ostéométrie · *Landmarks* · Surfaces osseuses virtuelles · Fiabilité · Répétabilité · Reproductibilité

L. Corron (✉) · F. Marchal · S. Condemi · K. Chaumoître · P. Adalian
UMR 7268 ADES–Aix-Marseille Université–EFS–CNRS,
Faculté de médecine de Marseille secteur Nord,
51 boulevard Pierre-Dramard,
F-13344 Marseille cedex 15, France
e-mail : louise.corron@etu.univ-amu.fr

K. Chaumoître
Service de radiologie et imagerie médicale, CHU Nord,
Assistance Publique des Hôpitaux de Marseille,
chemin des Bourrely,
F-13915 Marseille cedex 20, France

Lavoisier

Introduction

Anthropological studies have benefited greatly from the development of medical imaging techniques such as computed tomodensitometry, magnetic resonance imaging, or microtomodensitometry, which provide extensive, precise, and reliable osteometric data. This type of data can be measured from two-dimensional and/or three-dimensional virtual representations of bones obtained from living individuals using medical imaging techniques or, more frequently in physical anthropology and palaeoanthropology in particular, from dry bones examined directly using microscribes, surface scanners, or the aforementioned medical tools. Two main advantages of using digitised osteological data are the preservation of dry bone specimens and access internal bone structures that would otherwise require the destruction of the material [1].

With two-dimensional virtual bone images, or three-dimensional virtual bone surfaces or volumes now in common use, extensive literature has been produced on the verification of osteological data and data acquisition protocols [2–5]. While direct data acquisition from dry bones requires a sliding calliper, a cephalic compass, or an osteometric board, the procedure is a little more complex when working on virtual representations or reconstructions of bones as several requirements have to be met. These include the development of standard acquisition protocols and segmentation protocols to extract and reconstruct bone surfaces or volumes, as well as verifying consistency between homologous virtual and "dry bone" landmarks to obtain osteometric data [6].

Variables defined from dry bones can be geometric (e.g., maximum length) or anatomical (e.g., anteroposterior diameter at mid-length of a long bone). Such variables can be obtained from virtual bone surfaces using a set of tools that are provided in all virtual reconstruction software. Virtual bone variables can be obtained semiautomatically (e.g., maximum long bone length) or by calculating the distance between two landmarks placed manually or semiautomatically on virtual bone surfaces (e.g., anteroposterior diameter at mid-length of a long bone).

Implying that variables defined by dry bone characteristics are consistently applicable to virtual bone surfaces means that a measurement defined by geometric construction and/or landmarks placed on a virtual bone surface should be equivalent to a dry bone measurement taken with a sliding calliper or any other measuring tool. Bookstein [6] defined several types of landmarks, depending on their anatomical or geometric nature, which can be used to obtain osteometric data. Placing these landmarks implies different levels of difficulty, and different ranges of error for both their placement and the measurement they are defining, depending on their type: a landmark placed on an anatomical structure (Type I) should

logically be more reliable than a landmark obtained by geometric construction (Type II or III).

Testing the consistency, reliability, and accuracy of variables is considered a necessary prerequisite to any work involving biometric data [7]. Despite this consensus, such tests are not always systematic [8]. Moreover, the initial consistency between anatomical or geometric variables defined and measured on dry bones and the same variables measured on medical images of bones (computed tomography [CT] scans, X-rays, etc.) or reconstructed bone surfaces is rarely verified. Several studies have verified the consistency between variables measured directly on skeletal material (dry or fresh bones), with two-dimensional images of bones [9,10], or three-dimensional virtual representations of skeletal elements [3,11–18]. However, the dry and virtual bones were either placed in the same plane (flat surface) for comparison [16,18], or no mention was made of the position of the bones, so the influence of the initial position and orientation of the bones on measurement error was not evaluated. This can be problematical when working on medical imaging data collected from living individuals (often hospital patients) because they are not always placed in the same position during examination. It, therefore, seemed necessary to find a solution to avoid errors due to differences in bone position.

Other studies have tested the influence of the acquisition medium and/or of the software used to obtain variables on the quality of the data obtained [19,20], the reliability of variables or landmark positioning [4,17], and intra- and inter-observer errors (repeatability and reproducibility, respectively) [10,18,21]. However, the definitions of dry bone variables are mostly directly applied to "virtual" osteological material without explicitly testing their consistency, and to our knowledge, no study has been found that has simultaneously tested the repeatability and reproducibility of both landmark positioning and metric data directly obtained from landmark coordinates. Moreover, the studies mentioned above did not all use the same statistical parameters or tests to evaluate errors. The threshold of maximum error was set either at 1 or 2 mm without clear justification [18], and without presenting the range of error obtained for each biometric variable concerned.

This study has two objectives:

- To test the consistency between five variables measured on dry bones and on the surfaces of the same dry bones reconstructed virtually from CT scans. Consistency was evaluated between five dry bone-referenced clavicular variables obtained from both media as was the error relative to the type of material from which the variables are measured (dry bone vs. virtually reconstructed bone surface). A protocol for bone realignment was also elaborated to take errors due to different orientations of anatomical bone faces into account.

- To conduct an extensive evaluation of the repeatability (intra-observer errors) and reproducibility (inter-observer errors) of landmark positioning and osteometric data taken from the reconstructed bone surfaces of clavicles, fifth lumbar vertebrae, and iliae. This was done by comparing the descriptive data, the results of statistical tests, and the parameters and/or graphical representations of error distributions as presented above.

The differences between variables were evaluated using statistical tests, graphical error assessment, quantified error assessment, and consistency coefficients. Most of these methods of evaluation are routinely recommended [22,23] and used for such purposes, but they are seldom compared to one another. This study will also present the consistency rates between these four statistical tools.

Materials

Measurement consistency between dry bones and virtually reconstructed dry bone surfaces was tested on 40 dry clavicles from various archaeological backgrounds. The clavicle was selected because no protocol for measuring "virtual" equivalents for two of the five dry bone variables used here (minimum and maximum diameters) was found in the literature. Moreover, the extreme morphological variability of the clavicle [24,25] made it a good candidate to address the difficulties related to landmarks, variables, and bone orientation.

The clavicles belonged to 30 adults and 10 juveniles of unknown age and sex. Although age was unknown, the maximum length of the adult and juvenile bones varied widely (from 33.20 to 167.50 mm), and the bones had reached different states of maturity (unfused, partially fused, and completely fused epiphyses). Twenty-four adult clavicles were from the "*Cimetière des Trois Maisons*" archaeological site (18th–19th century) in Nancy, NE France [26]. The other six adult clavicles and the 10 juvenile ones belonged to individuals from unknown archaeological backgrounds that are used as reference material for osteology and anatomy courses at UMR 7268 ADES (Anthropologie bioculturelle, Droit, Éthique et Santé Research Unit). All the bones are housed in the collections of UMR 7268 ADES. CT scans of the 40 bones were performed with a 64-row multidetector CT scanner (Somatom Sensation 64, Siemens®, Erlangen, Germany). The scanning parameters were 120 kV, 315 mA, thickness: 0.6 mm. The dry clavicles were all placed parallel to the plane of acquisition, with the sternal ends facing the same direction. However, due to the different positions of stability on the board for each bone, their anterior, posterior, inferior, and superior faces were not always oriented in the same way [16]. Anatomical variability of the clavicle is well known [24]

and concerns the extremities and bone reliefs, such as the conoid tubercle, which can be more or less prominent and participate or not in the clavicle's position of stability. Therefore, the position of stability of the bones could not be systematically considered as their anatomical position. This particularity observed during acquisition enabled us to test whether or not this variability in the orientation of the anatomical faces of the bone could influence the variables.

The sample collected to test the repeatability and reproducibility of variables is made up of 84 juvenile patients (42 males and 42 females of 0–19 years of age) at the hospital services of Marseilles, France (AP-HM), who underwent a CT examination for various medical reasons. Individuals suffering from pathologies that could affect or directly concern skeletal growth and/or development were excluded. The individuals retained had often undergone medical examinations for acute diseases (such as appendicitis), trauma (accidents…), or for forensic purposes (virtual autopsies). The CT scans were collected from the PACS (Picture Archiving and Communication System, McKesson Medical Imaging Group, Richmond, BC, Canada) of the hospital services of Marseilles, France (AP-HM) and anonymized. The CT scans were performed with a 64-row multidetector CT scanner. The scanning parameters were 120 kV, 50–150 mA, thickness 0.6 mm. Most scans were obtained after the patients were injected intravenously with a contrast medium.

The surfaces of the clavicle, the fifth lumbar vertebra and/or the ilium of these individuals were reconstructed for use in this study. These bones represent the three main bone types (long bone, short bone, and flat bone, respectively) found in the human skeleton. This sample and the 19 osteometric variables presented in our study were used for another study on juvenile age estimation: all the variables were used to construct a new method using regression models for age prediction [27]. This is why the consistency, repeatability, and reproducibility of these variables had to be assessed beforehand.

Methods

Dry bone osteometrics

All five clavicular variables used in this study were defined from dry bones by Martin [28], Martin, and Saller [31] and Olivier [32]. They have been applied and/or redefined in several osteometric studies of the adult and juvenile clavicle [29–36]. None of the five clavicular variables tested here are defined by specific and reliable anatomical markers (corresponding to type I landmarks, [6]) on dry bones. Three of them are anatomical (Fig. 1), two are geometric.

Lavoisier

- Maximum length (Ln): the maximum distance between two points of the sternal and acromial ends of the bone [31,32]. This does not include unfused epiphyses but does include partially and completely fused ones. This variable is measured using a sliding calliper or with an osteometric board;
- Antero-posterior diameter at half-maximum length (AP_diam): the diameter between the anterior and posterior faces of the bone at half the total length of the bone [31,32]. The calliper must be oriented so that the handle is parallel to the anteroposterior axis;
- Supero-inferior diameter at half-maximum length (SI_diam): the diameter between the superior and inferior faces of the bone at half the total length of the bone [31,32]. The calliper must be oriented so that the handle is parallel to the supero-inferior axis;
- Maximum diameter at half-maximum length (Max_diam): the maximum diameter at half of total bone length [31,32]. This is found by turning the calliper around the bone at half-length perpendicularly to the axis of the total bone length until the highest value is obtained;
- Minimum diameter at half-maximum length (Min_diam): the minimum diameter at half the total bone length [31,32]. This is found by turning the calliper around the bone at half-length perpendicularly to the axis of the total bone length until the smallest value is obtained.

Maximum length (Ln), the anteroposterior (AP_diam) and the supero-inferior (SI_diam) diameters had already been tested for reliability and accuracy on the dry bone and on the virtually reconstructed bone surface of the two clavicles from a single individual [18].

The variables were measured using a digital sliding calliper (0.01 mm precision) or an osteometric board (1 mm accuracy) if the clavicles were longer than 150 mm, as this was the maximum distance that could be measured by the
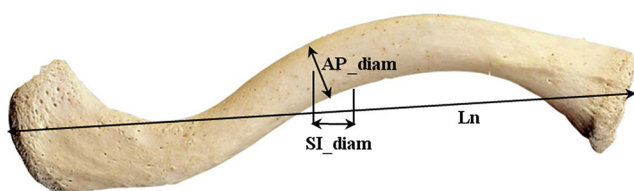
Ln: Maximum length
AP_diam: Antero-Posterior diameter
SI_Supero-inferior diameter

Fig. 1 Variables defined from dry bones by Martin and Saller [28], Martin and Saller [31] and Olivier [32], and measured on dry clavicles and reconstructed bone surfaces of the dry clavicles / *Variables définies sur os sec par Martin et Saller* [28], *Martin et Saller* [31] *et Olivier* [32], *prises sur des clavicules sèches et sur les surfaces osseuses des clavicules sèches reconstruites à partir de données scanner*

calliper. The bones were measured in the reference anatomical plane, with the anterior, posterior, superior, and inferior faces clearly identified.

Scanned bone osteometrics

Segmentation and virtual bone surface reconstructions

The stack of CT scans obtained for each dry clavicle was loaded into the ImageJ® v1.48 Software (National Institute of Health, USA). The grey-scale level corresponding to the separation between bone tissue and its immediate environment (air) was calculated independently using the half-maximum height (HMH) method [37]. The HMH values were then used for semiautomatic segmentation of all bone surfaces and the transformation of two-dimensional objects into three-dimensional bone surfaces. This was done with the Image Segmentation module of the Avizo® v7.0.0 software (Visualizing Sciences Group, SAS).

The bone surfaces of 30 clavicles (left or right), 30 fifth lumbar vertebrae, and 30 iliae (left or right) were also reconstructed from the CT scan slices obtained from living individuals using the HMH method [37] to obtain the grey-scale level between bone and soft tissue on the Image J® and Avizo® softwares.

Realignment of the "virtual" clavicular surfaces

One difficulty encountered when using data taken from CT examinations of living individuals is the variation of the anatomical position and orientation of the bones according to the position of the patient. For example, the patient's arms are mostly lifted over the head, but sometimes they are laid flat, or extended laterally. Arm and shoulder position can vary from patient to patient but also between the left and right sides of the same patient, resulting in different orientations of the bones, and in this case, of the clavicles. Indeed, because of these variations in the patient's position and its important anatomical variability, it can be difficult to find the anatomical orientation of the clavicle. This is not so true for the lumbar vertebrae and the iliae, whose anatomical planes are easier to find. We, therefore, had to verify that the position of the patients did not influence the values of the two anatomical variables taken from that particular bone, to avoid error due to orientation.

A previous study by Brough et al. (2013) showed that variables measured on dry clavicles and on the same clavicles virtually reconstructed from Multi Slice Computed Tomographic slices were comparable [9]. However, the dry and virtual bones were placed in the same plane (flat surface) for comparison, so the influence of the initial position and orientation of the bone was not evaluated.

The idea is to place the reconstructed clavicular surfaces in a common plane defined by three landmarks, preferably of type I, to ensure their reliability and consistency [6]. The extreme morphological variability of the sternal extremity of the clavicle [24] raises questions about placing a landmark there. Similarly, it was often difficult to identify a consistent landmark on the lateral end of the clavicles. Finally, no third point could be identified and qualified as reliable and consistent. To our knowledge, no consensual, reliable, and reproducible plane has been identified for the orientation of virtually reconstructed clavicles [38].

For these reasons, and to avoid errors due to differences in orientation, it was decided to develop a protocol for realigning the clavicles in a common geometric plane directly in Avizo®, without relying on common landmarks or acquisition parameters. Therefore, a new geometric and common orientation of each bone was constructed for the scanned dry bone clavicular surfaces (Fig. 2). With this protocol, the clavicular surfaces are realigned following the same axes: the axis parallel to the maximum length of the bone (x) and the ones perpendicular to it (y and z). Once the bones were realigned on the (x) axis, the anatomical
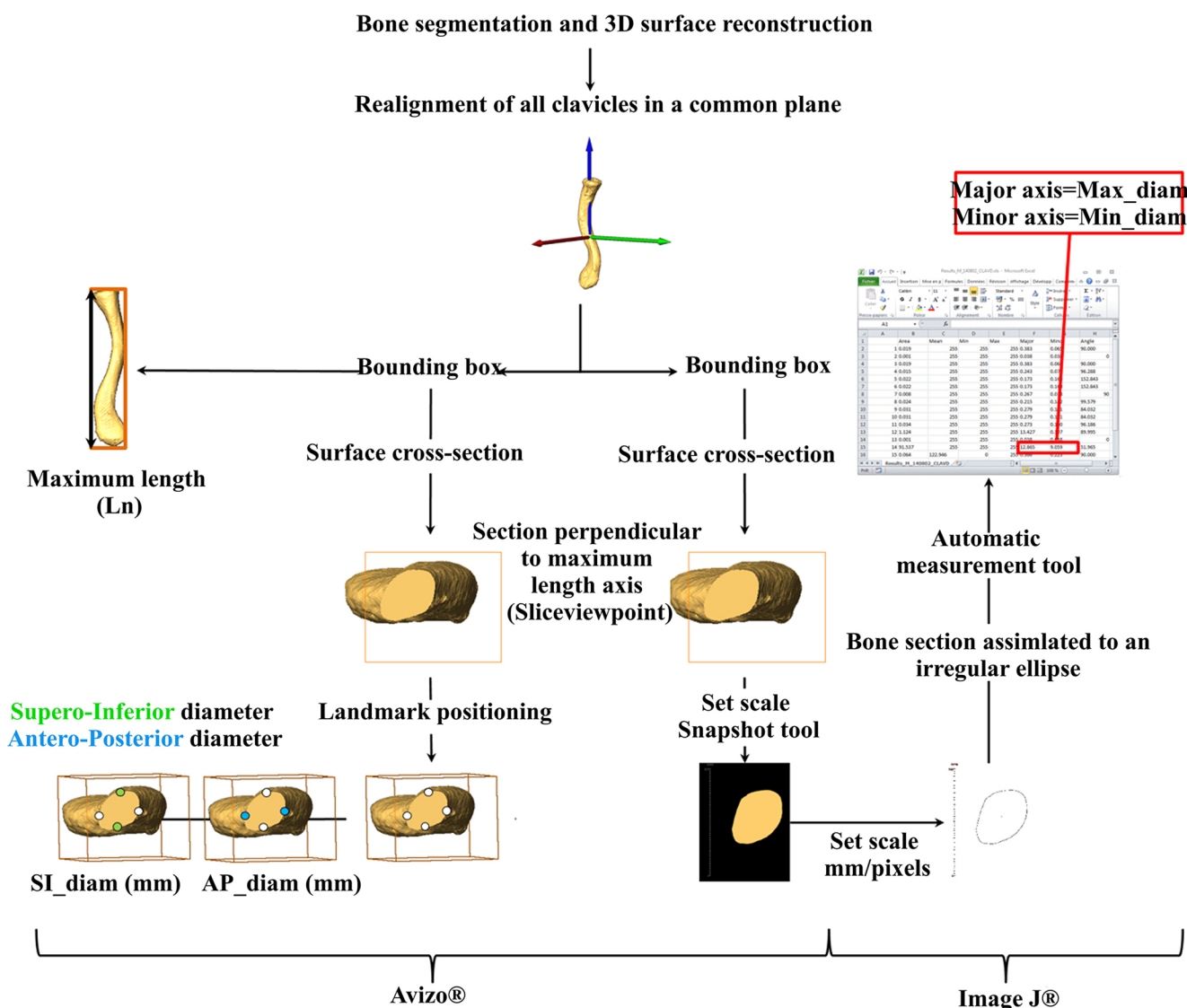
Fig. 2 Protocol for clavicular realignment and osteometric data acquisition, showing the steps for anatomical variables (acquired using landmark coordinates or direct measurement tools in Avizo®) and geometric variables (acquired using measurement tools provided by Avizo® and Image J® Software) / *Étapes du protocole de réalignement des clavicules et d'acquisition des données ostéométriques anatomiques (obtenues par les coordonnées de landmarks ou par mesure directe sur Avizo®) et géométriques (obtenues avec les outils de mesures des logiciels Avizo® et Image J®)*

orientation of the bone was obtained using the acromial surface: we considered the flat surface of the acromial plane as the superior surface of the clavicle, as it is for dry bones [38]. With the clavicular surfaces reoriented along the same axis, the protocol for measuring the bones could be applied.

We then proceeded to realign all the reconstructed clavicular surfaces of the living individuals on a common plane following the same protocol presented above. Once the position and the orientation of the bone were established, the protocol for acquiring the five osteometric variables was applied because the plane allowed landmark positioning to obtain the clavicular variables using the virtual measurement tools provided by the Avizo® Software.

Landmark positioning and landmark-defined variables

The definitions of the variables are found in anatomical, anthropological, and anthropometric manuals and in the articles cited above. The osteometric variables measured on the clavicles and iliae and two of the lumbar variables (Anterior Vertebral Height [AVH] and Upper Vertical Width [UVW]) are well-known anthropological and/or osteological variables defined by several authors for adult bones [28,31,32,39,40] and used in biological anthropology or clinical studies [39,40]. The corresponding variables measured on juvenile

and/or foetal bones have also been defined and/or used in previous works by different authors [34,35,41].

Three sets of landmarks with three coordinates (x, y, z) each were positioned on the reconstructed surfaces of the iliae (four landmarks), fifth lumbar vertebrae (eight landmarks), and clavicles (four landmarks) using the "Landmarks" tool provided by the Avizo® software.

The variables correspond to the Euclidian distance between two landmarks (e.g., A and B) positioned on the reconstructed bone surfaces. They are calculated using the formula for vector length with the landmark coordinates, two by two: $\|(AB)\| = \sqrt{((xA − xB)^2 + (yA − yB)^2 + (zA − zB)^2)}$.

Four iliac landmarks (Fig. 3 and Table 1) were placed directly on the Antero-Superior Iliac Spine (ASIS), the Postero-Superior Iliac Spine (PSIS), the Iliac Crest Summit (ICS), and the most prominent internal point of the acetabular surface of the ilium, the Internal Acetabular Point (IAP): ASIS and PSIS are type I landmarks, ICS and IAP are type II landmarks [6].

These landmarks define two iliac variables (Fig. 4) expressed in mm [31]:

- Ilium Length (**IL**) is the distance between ASIS and PSIS
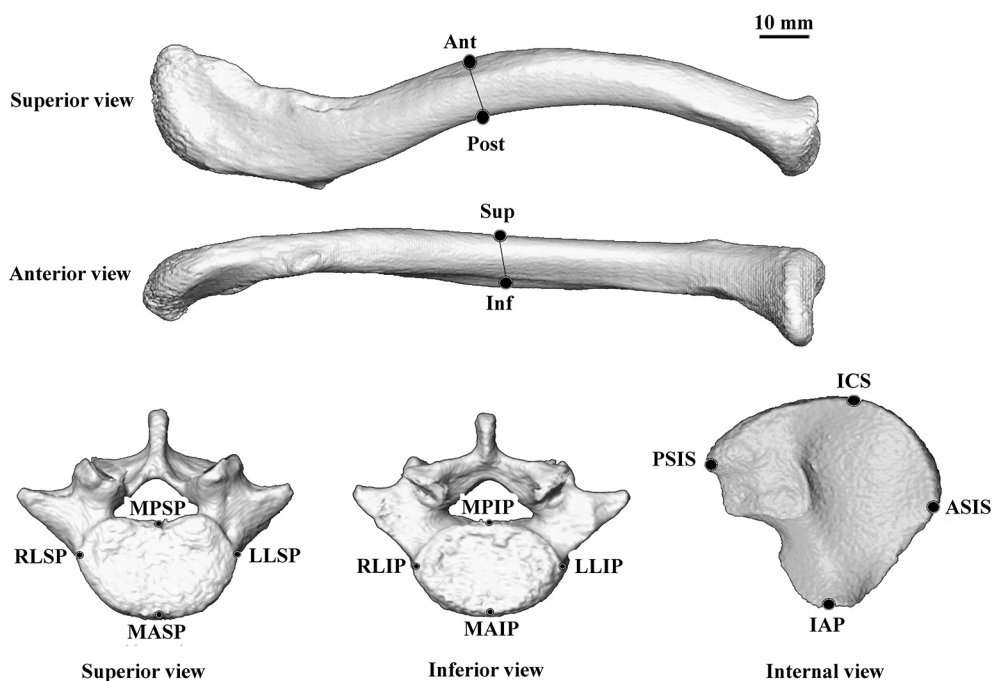- Ilium Width (**IW**) is the maximum distance between ICS and IAP

Fig. 3 Landmarks placed on the reconstructed bone surfaces of the ilium (four landmarks), the fifth lumbar vertebra (eight landmarks) and the clavicle (four landmarks) / *Landmarks placés sur les reconstructions osseuses de l'ilium (quatre landmarks), de la cinquième vertèbre lombaire (huit landmarks) et de la clavicule (quatre landmarks)*

Table 1 Definitions of the landmarks placed on the reconstructed bone surfaces of the iliae, fifth lumbar vertebrae and clavicles and their type according to Bookstein's classification [6] / *Définitions des landmarks placés sur les surfaces osseuses virtuellement reconstruites des iliums, cinquièmes vertèbres lombaires et clavicules et type de landmarks selon la classification de Bookstein* [6]

Bone	Landmarks	Definition	Type [6]
Ilium	ASIS	Antero-Superior Iliac Spine	I
	PSIS	Postero-Superior Iliac Spine	I
	ICS	Iliac Crest Summit: the most cranial point of the iliac crest	II
	IAP	Internal Acetabular Point: the most anterior part of the ilium, and the most superior part of the Y-cartilage that will later ossify and fuse with the ischium and the pubis	II
Fifth lumbar vertebra	MPSP	Median Postero-Superior Point: the intersection between the posterior border of the upper vertebral surface and the mid-sagittal plane of the vertebral body	II
	MASP	Median Antero-Superior Point: the intersection between the anterior border of the upper vertebral surface and the mid-sagittal plane of the vertebral body	II
	RLSP	Right Lateral Superior Point: the most lateral point on the right side of the upper articular surface of the vertebral body	II
	LLSP	Left Lateral Superior Point: the most lateral point on the left side of the upper articular surface of the vertebral body	II
	MPIP	Median Postero-Inferior Point: the intersection between the posterior border of the lower vertebral surface and the mid-sagittal plane of the vertebral body	II
	MAIP	Median Antero-Inferior Point: the intersection between the anterior border of the lower vertebral surface and the mid-sagittal plane of the vertebral body	II
	RLIP	Right Lateral Inferior Point: the most lateral point on the right side of the lower articular surface of the vertebral body	II
	LLIP	Left Lateral Inferior Point: the most lateral point on the left side of the lower articular surface of the vertebral body;	II
Clavicle	Sup	The most superior point of the clavicle at half-length	III
	Inf	The most inferior point of the clavicle at half-length	III
	Ant	The most anterior point of the clavicle at half-length	III
	Post	The most posterior point of the clavicle at half-length	III

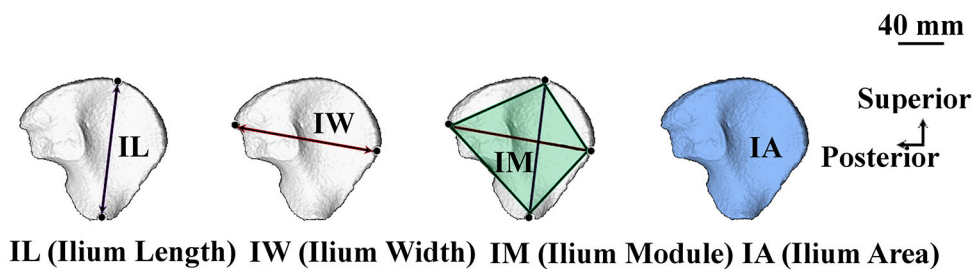**IL (Ilium Length) IW (Ilium Width) IM (Ilium Module) IA (Ilium Area)**

Fig. 4 Osteometric variables obtained from the reconstructed surfaces of the ilium (four variables) / *Variables ostéométriques obtenues sur les reconstructions osseuses de l'ilium (quatre variables)*

Eight vertebral type II landmarks [6] were placed directly on the bone surfaces (Fig. 3 and Table 1): the Median Postero-Superior Point (MPSP), the Median Antero-Superior Point (MASP), the Right Lateral Superior Point (RLSP), the Left Lateral Superior Point (LLSP), the Median Postero-Inferior Point (MPIP), the Median Antero-Inferior Point (MAIP), the Right Lateral Inferior Point (RLIP), and the Left Lateral Inferior Point (LLIP).

These landmarks define eight lumbar variables (Fig. 5) all expressed in mm [40,44]:

- Upper Vertebral Width (**UVW**): the distance between RLSP and LLSP
- Upper Vertebral Length (**UVL**): the distance between MASP and MPSP
- Lower Vertebral Width (**LVW**): the distance between RLIP and LLIP

- Lower Vertebral Length (**LVL**): the distance between MAIP and MPIP
- Anterior Vertebral Height (**AVH**): the distance between MASP and MAIP
- Posterior Vertebral Height (**PVH**): the distance between MPSP and MPIP
- Right Vertebral Height (**RVH**): the distance between RLSP and RLIP
- Left Vertebral Height (**LVH**): the distance between LLSP and LLIP

The first six of these eight osteometric variables measured on the fifth lumbar vertebra were used in a geometric medical and anthropometric study of CT scan slices of adult lumbar vertebrae by Zhou and collaborators in 2000 [39] and in a morphometric study of dissected juvenile lumbar vertebrae by Mavrych and collaborators in 2014 [42]. These variables were adapted from previous anthropometric and morphometric studies of the lumbar spine [40,43–45].

Four type III clavicular landmarks [6] were placed on the superior (Sup), inferior (Inf), anterior (Ant), and posterior (Post) sides of the transversal section at half-length of the realigned clavicles, determined using the Avizo® surface cross-section tool (Fig. 3 and Table 1). These landmarks define two variables (Fig. 6) expressed in mm [31]:

- Antero-posterior diameter at half-maximum length (**AP_diam**): the distance between the Ant and Postlandmarks
- Supero-Inferior diameter at half-maximum length (**SI_diam**): the distance between the Sup and Inf landmarks

Geometric and composite variables

Geometric and composite osteometric variables were either obtained by geometric construction (geometric variables) or calculated from previous variables (composite variables).

Two bidimensional iliac variables (one composite and one geometric) were obtained (Fig. 4). They are expressed in mm^2:

- Ilium Module (**IM**) is the product of IL and IW. IM corresponds to the surface of the quadrilateral whose diagonals are IL and IW. It is a rough geometric estimate of the internal surface of the ilium [27];
- Ilium Area (**IA**) is the exact measurement of the internal bone surface of the ilium as seen perpendicularly when the bone is projected along the iliac plane [46,47]. The iliac plane is an anatomical plane defined by the three following points: ASIS, PSIS, and IAP. To obtain IA, the observer must position the ilium along the iliac plane and face the internal bone surface perpendicularly.

Two composite bidimensional lumbar variables were calculated (Fig. 5) [27]. They are expressed in mm^2:

- The Upper Vertebral Module (**UVM**): UVM is the product of UVL and UVW. UVM is a quadrilateral whose diagonals are UVL and UVW; this variable was constructed to give a geometric approximation of the upper articular surface of the vertebral body;
- The Lower Vertebral Module (**LVM**): LVM is the product of LVL and LVW. LVM is a quadrilateral whose diagonals
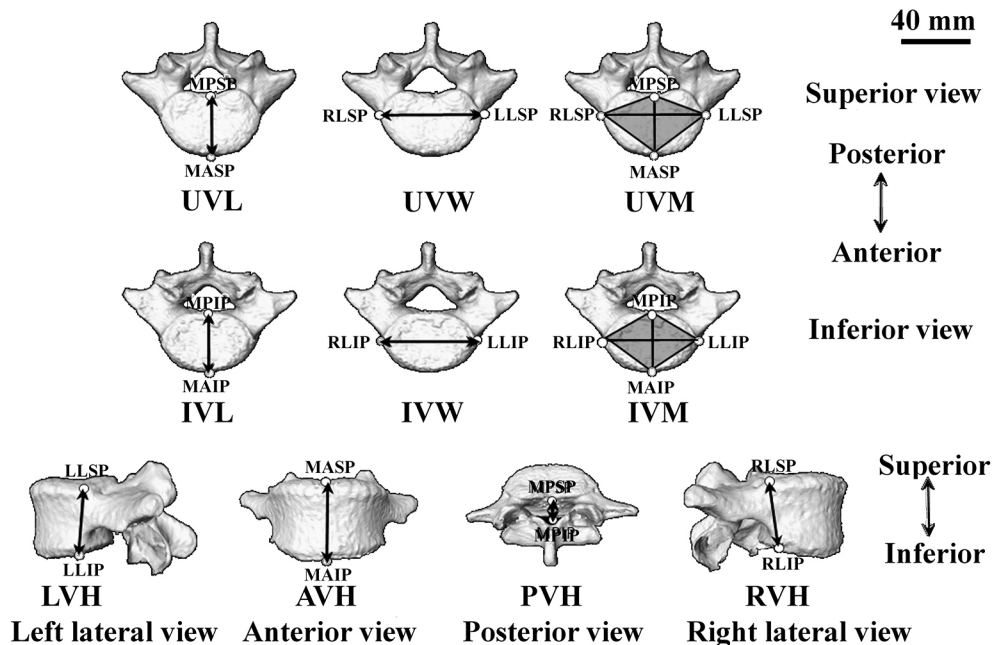

Fig. 5 Osteometric variables obtained from the reconstructed surfaces of the fifth lumbar vertebra (10 variables) / *Variables ostéométriques obtenues sur les reconstructions osseuses de la cinquième vertèbre lombaire (dix variables)*
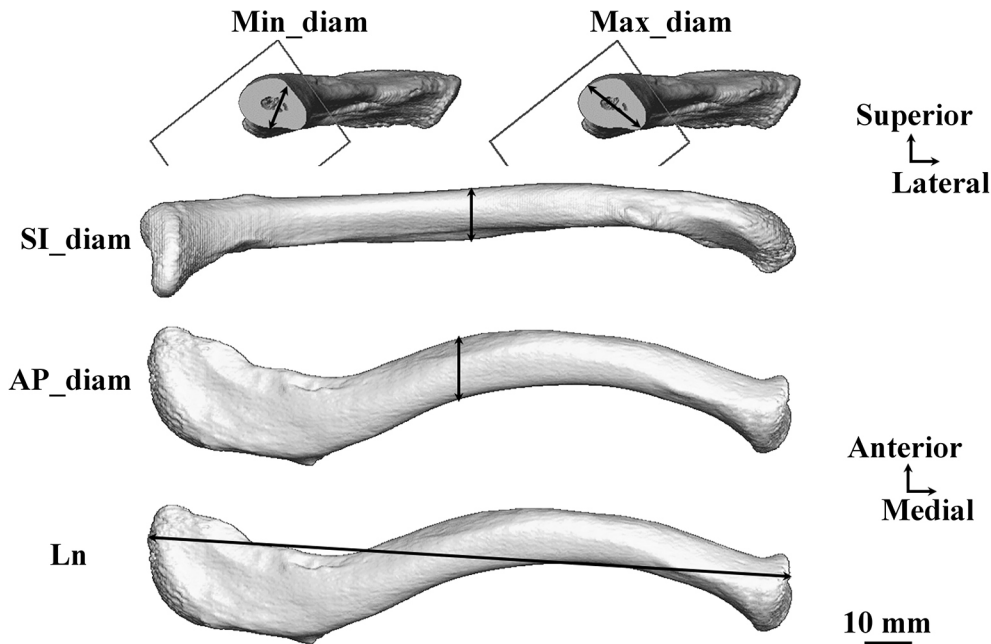
Fig. 6 Five osteometric variables obtained from the virtually reconstructed clavicular surfaces and the dry clavicles. From top to bottom: minimum diameter (Min_diam) and maximum diameter (Max_diam), anteroposterior diameter (AP_diam), supero-inferior diameter (SI_diam), and maximum clavicular length (Ln) / *Cinq variables ostéométriques obtenues sur les surfaces osseuses virtuelles des clavicules et les clavicules sèches. De haut en bas : diamètre minimum (Min_diam), diamètre maximum (Max_diam), diamètre antéropostérieur (AP_diam), diamètre supéro-inférieur (SI_diam) et longueur maximale de la clavicule (Ln)*

are equal to LVL and LVW; this variable was constructed to give a geometric approximation of the lower articular surface of the vertebral body.

Three clavicular variables were measured (Fig. 6). They are expressed in mm [31,32]:

- Maximum length (**Ln**): maximum distance between the medial/sternal extremity and the lateral/acromial extremity. This measurement includes partially and completely fused epiphyses but not unfused clavicular epiphyses. It was taken using the Avizo® "Bounding box" and "Linear Measurement tools";
- Maximum diameter at half-maximum length (**Max_diam**): a geometric variable taken by finding the greatest distance between two points at mid-length of the total length of the clavicle, perpendicular to the principal axis of the bone shaft and passing through the centre of the ellipse formed by the mid-shaft section of the bone;
- Minimum diameter at half-maximum length (**Min_diam**): a geometric variable taken by finding the smallest distance between two points at mid-length of the total length of the clavicle, perpendicular to the principal axis of the bone and passing through the centre of the ellipse formed by the mid-shaft section of the bone.

Landmark positioning and acquisition of variables were done twice by the same observer and once by an independent observer. Both were trained on Avizo®.

Evaluating consistency, repeatability and reproducibility

Normality of the distribution of the distances between homologous landmarks and the differences between variables was assessed using QQ plots (Fig. 7).

Equality between dry bone and scanned bone variables was assessed by comparing dry bone measurements to their scanned counterpart, two by two. Consistency was then tested with a paired Student t-test if the differences followed a normal distribution, or with a signed-rank Wilcoxon test if they did not ($p > 0.05$ for both tests) (Table 2). The 95% confidence intervals (CIs) of the mean differences between variables were also calculated. Bland–Altman plots [47] were made to verify the general distribution of the differences between homologous variables obtained on dry and scanned bone surfaces. These plots are a way of visualising errors by plotting the differences between the paired variables ($d_i = x_i − x_j$) against the mean values of each pair of variables:

$$\bar{x} = (x_i + x_j)/2$$

They allow investigation of the existence of any systematic difference between the variables (i.e., bias) and identification of possible outliers. The limits of the acceptable error rate are represented by 95% CIs, equivalent to a 5% error rate. If more
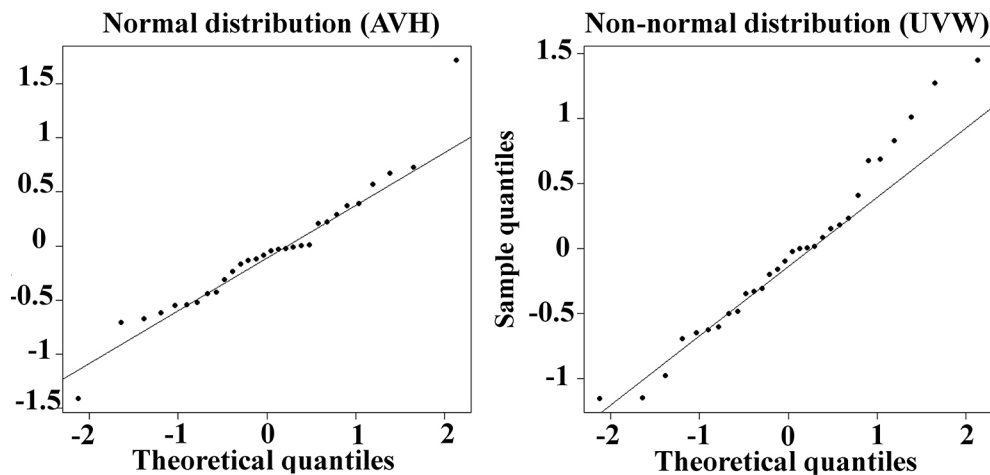
Fig. 7 QQ plot of the differences between two series of measurements for Anterior Vertebral Height (AVH) (left) and Upper Vertical Width (UVW) (right) showing a normal (left) and nonnormal (right) distribution / *QQ-plot de la distribution des différences entre les deux séries de mesures de (gauche) et (droite) montrant une distribution normale (gauche) et non normale (droite)*

Table 2 Different tests and parameters used to evaluate the distribution, consistency, repeatability (intra-observer error) and reproducibility (inter-observer error) of the landmarks and/or osteometric variables / *Tests et paramètres statistiques utilisés pour évaluer la distribution, l'adéquation, la répétabilité (erreur intra-observateur) et la reproductibilité (erreur interobservateur) des landmarks et/ ou des variables ostéométriques*

Aim of the test/parameter	Test/Parameter	Hypotheses/Goals
Type of distribution	QQ-plot	Compare the distribution of the quantiles of variable differences to the normal (theoretical) distribution of the quantiles
Equality of variables measured on dry bones and reconstructed bone surfaces.	Paired *t*-test Paired Wilcoxon test	H0: the mean difference between two variables is zero H1: the mean difference between two variables is different from zero
Repeatability and reproducibility of landmarks and variables		H0: the mean difference between ranks is zero H1: the mean difference between ranks is different from zero
Repeatability and reproducibility of the variables	• Intraclass correlation (ICC)	Assess consistency or agreement of paired quantitative variables made by the same observer of by different observers High coefficients imply high observer agreement
	• Bland-Altman plots	Evaluate the consistency between two different instruments or two variables techniques. If more than 90% of the differences are within the 95% Confidence Interval of the values, the two instruments or techniques are consistent
	• Technical error of measurement (TEM) and percentage of TEM (% TEM)	TEM = Measurement of intra- and inter-observer variables (mm) %TEM = relative TEM. If %TEM≤5%, variables are repeatable and reproducible

than 90% of the points are included within the limits of the CI, errors are accepted as sufficiently low for validation.

Repeatability and reproducibility of landmark positioning and variables were tested on the reconstructed bone surfaces with paired Student *t*-tests if the landmark distances or variables followed a normal distribution, or with signed-rank Wilcoxon tests if they did not ($p > 0.05$ for both tests). Because the tests were used to make multiple comparisons of variables (19 altogether), the Bonferroni adjustment of the *p*-value ($\alpha = 0.05$) was applied. After this adjustment, the threshold for significant intra- or inter-observer errors was therefore 0.05/19 = 0.0026 [49].

Table 3 Results of the statistical tests for consistency of variables measured on dry clavicles and the virtually reconstructed bone surfaces of the same clavicles. 95% CI = 95% confidence interval. *p*-values in italics of the paired *t*-test (upper part of the table) and Wilcoxon tests (lower part of the table) are significant / *Résultats des tests statistiques de l'égalité des variables prises sur clavicules sèches et sur les reconstructions virtuelles des surfaces osseuses de ces mêmes clavicules sèches. IC 95 % = intervalle de confiance à 95 %. Les p-values en italiques des tests de Student sur séries appariées (partie supérieure du tableau) et des tests de Wilcoxon (partie inférieure du tableau) sont significatives*

Measurement	*t*	Df	CI 95% (mm)	Mean difference (mm)	*p*-value
Ln	2.881	39	[0.073; 0.415]	0.244	*0.0064*
Max_diam	1.280	39	[−0.054; 0.240]	0.093	0.208
AP_diam	0.91	39	[−0.101; 0.264]	0.082	0.370
Measurement	*V*		**CI 95% (mm)**	**Pseudo-median (mm)**	***p*-value**
Min_diam	649		[0.073; 0.261]	0.169	*0.0013*
SI_diam	213		[−0.658; −0.065]	−0.326	*0.007*

Ln: Maximum clavicular length; Max_diam: Maximum diameter at half-maximum length; AP_diam: Anteroposterior diameter at half-maximum length; Min_diam: Minimum diameter at half-maximum length; SI_diam: Supero-inferior diameter at half-maximum length.
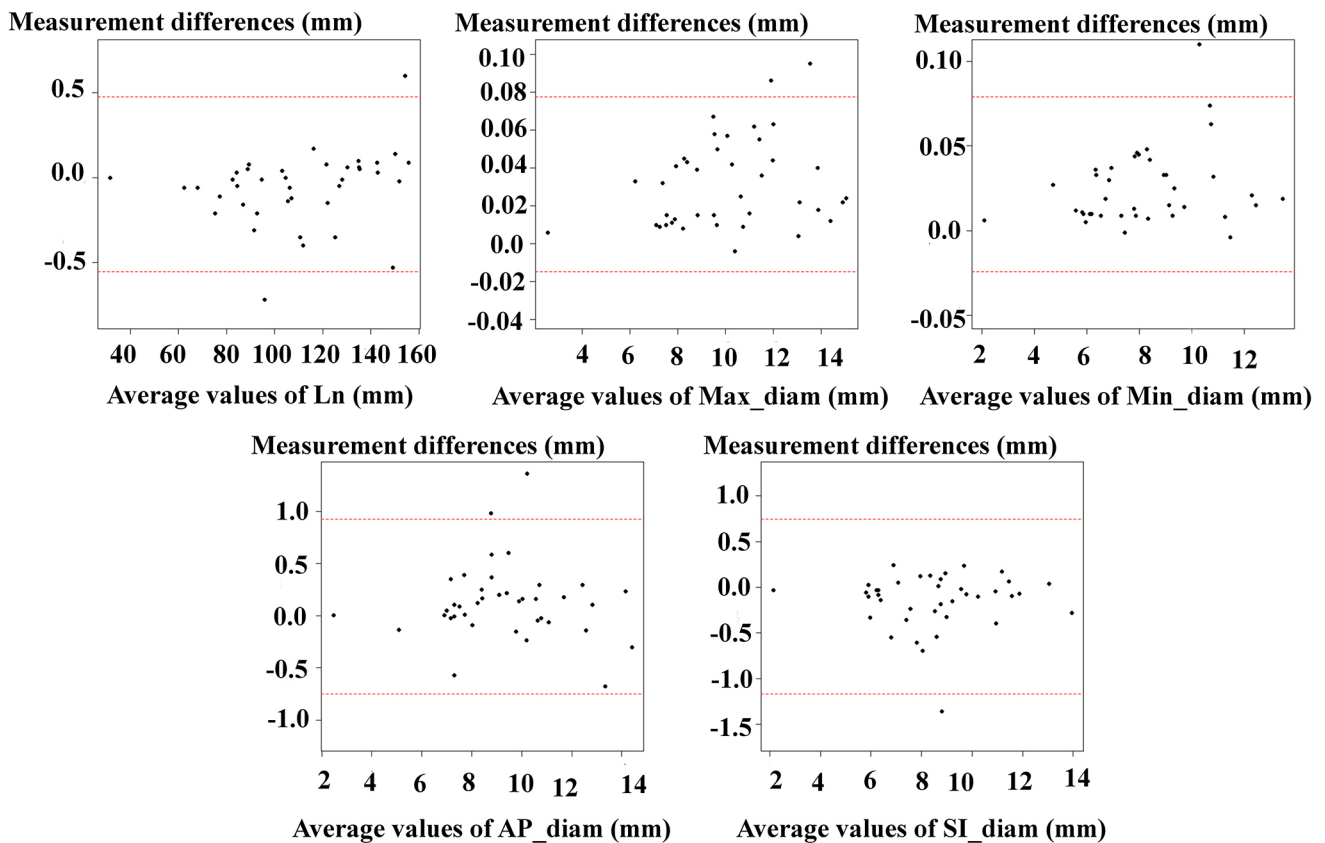
Fig. 8 Bland–Altman plots of the five variables measured on dry clavicles and on the virtually reconstructed surfaces of these dry bones. Red dotted lines correspond to the limits of the 95% confidence interval for the distribution of the errors, in mm / *Graphiques de Bland-Altman des cinq variables obtenues pour les clavicules sèches et pour les surfaces osseuses virtuelles de ces mêmes os. Les pointillés rouges correspondent aux bornes de l'intervalle de confiance à 95 % de la distribution des erreurs, en mm*

Tests were performed on the distances between two homologous landmarks placed by the same observer twice or by the two observers, and on the differences between homologous variables obtained by the same observer twice or by the two observers (Table 2). The 95% CIs for the distances between homologous landmarks and for differences between variables indicate whether or not the variables were systematically biased (positively or negatively).

Three statistical parameters were also calculated to assess the repeatability and reproducibility of the variables (Table 2):

- The intraclass correlation coefficient (ICC) was used to assess the consistency (repeatability) and agreement (reproducibility) of paired quantitative variables taken by the same observer or by different observers, respectively [50]. ICC gives a composite of intra-observer and inter-observer variability by producing consistency and agreement coefficients.
- Bland–Altman plots [48], as presented above, were built to verify whether or not observer errors are accepted as sufficiently low for validation (if 90% of observer errors are within the range of the 95% CI).

- The technical error of measurement (TEM) is a statistical parameter used to measure the imprecision of variables [2,51]. It is the square root of measurement error variance and is obtained by calculating the difference (d_i) between N repeated measurements of the same variable taken by the same observer or by different observers, to calculate the value of the intra- and inter-observer errors, respectively.

$$TEM = \sqrt{\left(\sum di^2\right)\Big/2N}$$

The size of TEM has been found to be positively correlated with the size of the measurement, meaning that large values of variables are associated with high TEM and vice versa [52]. For this reason, and to be able to compare imprecision between different variables and different samples, Norton and Olds [53] proposed converting absolute TEM into relative TEM, or %TEM.

$$\%TEM = \left(TEM\Big/mean\left(\bar{x}\right)\right)\times100$$

%TEM values below 5% are considered to represent acceptable measurement repeatability and reproducibility.

Table 4 Results of the *t*-tests (upper part of the table) and Wilcoxon tests (lower part of the table) performed to evaluate landmark repeatability on all three bones / *Résultats des tests de Student (partie supérieure du tableau) et de Wilcoxon (partie inférieure du tableau) d'évaluation de la répétabilité des landmarks sur les trois os*

Bone	Intra-observer landmark distance	*t*	Df	95% CI (mm)	Mean difference (mm)	*p*-value
Ilium	ICS–ICS	9.25	29	[2.094; 3.283]	2.688	<0.001
Fifth lumbar vertebra	MPSP–MPSP	8.841	29	[0.644; 1.032]	0.838	<0.001
	MASP–MASP	10.512	29	[0.491; 0.729]	0.610	<0.001
	RLSP–RLSP	9.507	29	[0.632; 0.978]	0.805	<0.001
	LLIP–LLIP	10.803	29	[0.654; 0.959]	0.807	<0.001
Bone	**Intra-observer landmark distance**	**V**		**95% CI (mm)**	**Pseudo-median (mm)**	***p*-value**
Ilium	ASIS–ASIS	465		[0.427; 0.697]	0.547	<0.001
	PSIS–PSIS	465		[0.465; 0.902]	0.623	<0.001
	IAP–IAP	465		[0.815; 1.405]	1.08	<0.001
Fifth lumbar vertebra	LLSP–LLSP	465		[0.630; 0.999]	0.799	<0.001
	MPIP–MPIP	465		[0.419; 0.743]	0.550	<0.001
	MAIP–MAIP	465		[0.479; 0.715]	0.585	<0.001
	RLIP–RLIP	465		[0.529; 0.915]	0.694	<0.001
Clavicle	Sup–Sup	820		[0.407; 0.916]	0.693	<0.001
	Inf–Inf	820		[0.379; 0.769]	0.540	<0.001
	Ant–Ant	820		[0.412; 0.815]	0.566	<0.001
	Post–Post	820		[0.428; 0.822]	0.642	<0.001

ASIS: Antero-Superior Iliac Spine; PSIS: the Postero-Superior Iliac Spine; ICS: the Iliac Crest Summit; IAP: Internal Acetabular Point; MPSP: the Median Postero-Superior Point; MASP: the Median Antero-Superior Point; RLSP: the Right Lateral Superior Point; LLSP: the Left Lateral Superior Point; MPIP: the Median Postero-Inferior Point; MAIP: the Median Antero-Inferior Point; RLIP: the Right Lateral Inferior Point; LLIP: the Left Lateral Inferior Point.

The mean, median, minimum, maximum, and standard deviation of the distances between homologous landmarks placed by both observers and of each variable measured by both observers were calculated (Supplementary Tables S1–S4).

All statistical analyses were performed using the R Software® version 3.1.0 (2014-04-10).

Results

Consistency of variables obtained from dry clavicles and virtual clavicular surfaces

QQ-plots of the differences between the variables taken on dry clavicles and the reconstructed scanned surfaces of these clavicles showed that the differences for Ln, Max_diam, and AP_diam followed a normal distribution, but the differences between Min_diam and SI_diam did not.

Tests for Ln, Min_diam and SI_diam measured on dry and scanned dry bones showed significant differences (Table 3). Therefore, we should reject the hypothesis that the variables Ln, Min_diam and SI_diam are equal, because

the 95% CIs do not include zero. However, for all the 95% CIs, means and medians of the differences between variables are included in the interval [−1 mm; +1 mm] and the mean or pseudo-median differences are lower than 0.5 mm. We can therefore assess that the variables between dry and scanned bone present an error of less than +/− 1 mm in 95% of the cases, which is low enough to consider them as consistent.

Less than 10% of measurement differences are outside the 95% range of the corresponding Bland–Altman plots (Fig. 8). Min_diam and AP_diam have three, SI_diam has two, Ln has one, and Max_diam has none. We can conclude that in all five cases, there is a statistically nonsignificant difference between variables measured on dry bones and the same variables measured on the reconstructed surfaces of the same dry bones.

Repeatability and reproducibility of landmark positioning

The mean distances between homologous landmarks placed by the same observer twice and by two independent

Table 5 Results of the *t*-tests (upper part of the table) and Wilcoxon tests (lower part of the table) performed to evaluate landmark reproducibility on all three bones / *Résultats des tests de Student (partie supérieure du tableau) et de Wilcoxon (partie inférieure du tableau) d'évaluation de la reproductibilité des landmarks sur les trois os*

Bone	Inter-observer landmark distance	*t*	Df	95% CI (mm)	Mean difference (mm)	*p*-value
Ilium	IAP–IAP	11.043	29	[1.417; 2.061]	1.739	<0.001
Fifth lumbar	RLSP–RLSP	9.345	29	[1.125; 1.755]	1.440	<0.001
vertebra	LLSP–LLSP	9.104	29	[1.0956; 1.730]	1.413	<0.001
	MAIP–MAIP	10.422	29	[0.485; 0.723]	0.604	<0.001

Bone	Inter-observer landmark distance	*V*		95% CI (mm)	Pseudomedian (mm)	*p*-value
Ilium	ASIS–ASIS	465		[0.358; 0.661]	0.523	<0.001
	PSIS–PSIS	465		[0.534; 0.844]	0.664	<0.001
	ICS–ICS	465		[1.495; 2.969]	2.082	<0.001
Fifth lumbar	MPSP–MPSP	465		[0.588; 0.92]	0.756	<0.001
vertebra	MASP–MASP	465		[0.552; 1.011]	0.729	<0.001
	MPIP–MPIP	465		[0.699; 1.078]	0.882	<0.001
	RLIP–RLIP	465		[0.654; 1.189]	0.896	<0.001
	LLIP–LLIP	465		[0.625; 0.971]	0.806	<0.001
Clavicle	Sup–Sup	820		[0.480; 0.819]	0.646	<0.001
	Inf–Inf	820		[0.698; 1.436]	1.002	<0.001
	Ant–Ant	820		[0.437; 0.867]	0.594	<0.001
	Post–Post	820		[0.527; 0.940]	0.684	<0.001

ASIS: Antero-Superior Iliac Spine; PSIS: the Postero-Superior Iliac Spine; ICS: the Iliac Crest Summit; IAP: Internal Acetabular Point; MPSP: the Median Postero-Superior Point; MASP: the Median Antero-Superior Point; RLSP: the Right Lateral Superior Point; LLSP: the Left Lateral Superior Point; MPIP: the Median Postero-Inferior Point; MAIP: the Median Antero-Inferior Point; RLIP: the Right Lateral Inferior Point; LLIP: the Left Lateral Inferior Point.

observers on the virtual bone surfaces of the iliae, lumbar vertebrae, and clavicles differ significantly from zero for all landmarks ($p < 0.001$) (Table 4). The 95% CIs of the distances are always smaller than 1 mm, except for the two iliac landmarks ICS and IAP (Table 4). This means that landmarks are predominantly placed within 1 mm from each other by the same observer, except for the type II iliac landmarks ICS and IAP.

These two landmarks are also placed more than 1 mm from each other by two independent observers, along with the type III lumbar landmarks RLSP, LLSP, and the type III clavicular landmark Inf (Table 4). Iliac landmarks IAP, SCI, and clavicular landmark Inf had 95% CIs wider than 2 mm. The other landmark distances had 90% of their distributions within a 95% CI with a 2 mm range, independently of the observer (Tables 4 and 5).

By these observations, we can conclude that two type I iliac landmarks, ASIS and PSIS, are sufficiently repeatable and reproducible. The two type II iliac landmarks, ICS and IAP, are neither sufficiently repeatable nor reproducible.

Seven of the eight lumbar vertebral landmarks are sufficiently repeatable (Table 4). MPSP is neither sufficiently repeatable nor reproducible. Six lumbar landmarks are sufficiently reproducible (Table 4). The two superior lateral lumbar landmarks (RLSP and LLSP) are not.

Three of the four clavicular landmarks (Ant, Post, and Sup) are sufficiently repeatable and reproducible. The Inf landmark is not sufficiently reproducible (Table 5).

Table 6 Results of the t-tests (upper part of the table) and Wilcoxon tests (lower part of the table) performed to evaluate variable repeatability on all three bones. Values in italics indicate significant intra-observer differences after Bonferroni adjustment ($p < 0.0026$) / *Résultats des tests de Student (partie supérieure du tableau) et de Wilcoxon (partie inférieure du tableau) d'évaluation de la répétabilité des variables sur les trois os. Les valeurs en italiques indiquent des différences intraobservateur significatives après l'ajustement de Bonferroni ($p < 0,0026$)*

Bone	Intra-observer error test results						
	Variable	t	Df	95% CI (mm)	Mean difference (mm)	*p*-value	After adjusted *p*-value
Fifth lumbar vertebra	UVL	−0.870	29	[−0.316 ; 0.127]	−0.094	0.391	
	UVW	0.524	29	[−0.192 ; 0.324]	0.066	0.604	
	UVM[a]	−0.45	29	[−13.645 ; 8.724]	−2.461	0.651	
	LVW	−2.252	29	[−0.335 ; -0.016]	−0.175	0.032	
	AVH	−0.587	29	[−0.274 ; 0.152]	−0.061	0.562	
	RVH	1.462	29	[−0.066 ; 0.396]	0.165	0.016	

Bone	Variable	V		95% CI (mm)	Pseudo-median (mm)	*p*-value	
Ilium	IL	331		[0.003; 0.263]	0.146	0.043	
	IW	119		[−0.512; −0.049]	−0.271	0.019	
	IM[a]	143		[−62.538; 1.599]	−31.193	0.067	
	IA[a]	196		[−34.903; 13.518]	−9.193	0.465	
Fifth lumbar vertebra	LVL	154		[−0.238; 0.023]	−0.125	0.109	
	LVM[a]	126		[−10.870; −1.052]	−6.210	0.028	
	PVH	344		[0.056; 0.459]	0.280	0.021	
	LVH	218		[−0.359 ; 0.251]	−0.077	0.777	
Clavicle	Ln	289		[−0.110; 0.020]	−0.035	0.240	
	Max_diam	818		[0.023; 0.038]	0.030	*<0.001*	
	Min_diam	817		[0.017; 0.030]	0.023	*<0.001*	
	AP_diam	564		[0.005; 0.179]	0.093	0.038	
	SI_diam	187		[−0.216; −0.041]	−0.116	*0.002*	

UVL: Upper Vertebral Length; UVW: Upper Vertebral Width; UVM: Upper Vertebral Module; LVW: Lower Vertebral Width; AVH: Anterior Vertebral Height; RVH: Right Vertebral Height; IL: Ilium Length; IW: Ilium Width; IM: Ilium Module; IA: Ilium Area; LVL: Lower Vertebral Length; LVM: Lower Vertebral Module; PVH: Posterior Vertebral Height; LVH: Left Vertebral Height; Ln: Maximum clavicle length; Max_diam: Maximum diameter at half-maximum length; Min_diam: Minimum diameter at half-maximum length; AP_diam: Anteroposterior diameter at half-maximum length; SI_diam: Supero-inferior diameter at half-maximum length.

Repeatability and reproducibility of osteometric variables

The 95% CIs of the errors for lumbar variables LVW, RVH, and PVH, iliac variables IL and IW, and all four clavicular diameters are within less than ±1 mm (Table 6). However, according to the two statistical tests (*t*-test and Wilcoxon test) used for assessing variable repeatability, intra-observer errors differ significantly from zero for 10 variables: the iliac variables IL and IW, the lumbar variables LVW, LVM, PVH, and RVH and the clavicular variables Max_diam, Min_diam, AP_diam, and SI_diam ($p < 0.05$) (Tables 6 and 9). After applying the Bonferroni adjustment, intra-observer errors are relatively important for only two variables: Max_diam and Min_diam. This seems contradictory with the 95% CIs for the distribution of the errors for these two variables, which are very small (Table 6).

The 95% CIs of the inter-observer errors have a range of less than ±0.5 mm for 10 variables: the two iliac variables IL, IW, the four lumbar variables UVL, PVH, LVL, LVW, and the four clavicular diameter variables Max_diam, Min_diam, AP_diam, and SI_diam. The Bland–Altman plots show that 95% CIs for intra- or inter-observer errors generally cover negative and positive values, meaning that none of the observers systematically over- or underevaluates the variables (Fig. 9). However, the *p*-values associated with the *t*-tests or Wilcoxon tests done to assess the reproducibility of variables indicated significant interobserver errors for 14 variables (Table 7): the iliac variables IL, IW, and IM, the lumbar variables UVL, UVM, PVH, RVH, LVL, LVW, and LVM and the four clavicular diameter variables AP_diam, SI-diam, Max_diam, and Min_diam. When the Bonferroni adjustment is applied to the *p*-value, we can conclude that only seven variables present interobserver errors: the iliac variables IW and IM, the lumbar variables LVL, LVM, and RVH and the clavicular variables Max_diam and Min_diam (Table 7).

The ICC coefficients for consistency (repeatability) and agreement (reproducibility) for all variables range from 0.972 for the lowest to 1 for the highest, meaning that all variables are sufficiently repeatable and reproducible (Table 8), with 95% reliability.

Intra- and inter-observer Technical Errors of Measurement (TEM) were lower than 0.8 mm for all unidimensional variables (Table 9). The intra-observer %TEM values for the lumbar variable LVH and the clavicular variable SI_diam were close to 5%; the inter-observer %TEM values for the lumbar variables PVH and RVH were higher than 5%. This
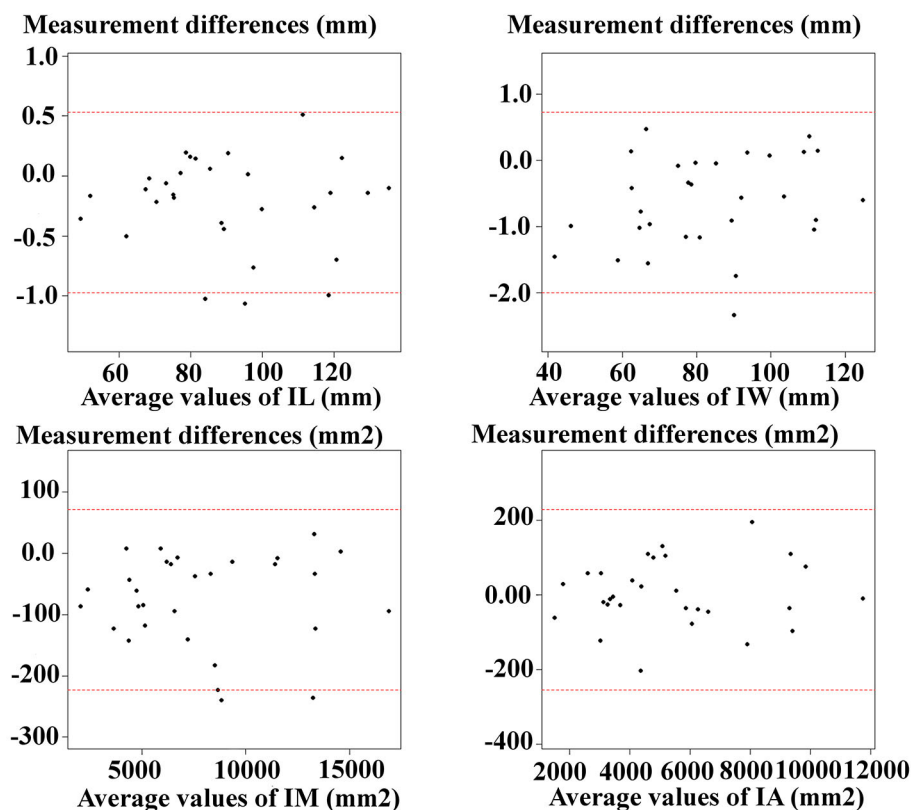

Fig. 9 Bland-Altman plot of the inter-observer differences between iliac variables / *Grahiques de Bland-Altman des erreurs interobservateur pour les variables de l'ilium*

*L*avoisier

Table 7 Results of the *t*-tests (upper part of the table) and Wilcoxon tests (lower part of the table) performed to evaluate the reproducibility of variables on all three bones. Values in italics indicate significant inter-observer differences after Bonferroni adjustment (p < 0.0026) / *Résultats des tests de Student (partie supérieure du tableau) et de Wilcoxon (partie inférieure du tableau) d'évaluation de la reproductibilité des variables sur les trois os. Les valeurs en italique indiquent des différences interobservateur significatives après l'ajustement de Bonferroni (p < 0,0026)*

Bone	Inter-observer error test results						After
	Variable	*t*	Df	CI 95% (mm)	Mean difference (mm)	*p*-value	adjusted *p*-values
Ilium	IW	−5.034	29	[−0.898; −0.379]	−0.638	*<0.001*	
Fifth lumbar	UVL	−2.078	29	[−0.438; −0.003]	−0.221	0.047	
vertebra	UVW	−0.355	29	[−0.290; 0.204]	−0.043	0.726	
	UVM[a]	−2.189	29	[−20.484; −0.694]	−10.589	0.037	
	PVH	2.396	29	[0.065; 0.819]	0.442	0.023	
	AVH	1.737	29	[−0.035; 0.426]	0.196	0.093	
	RVH	6.184	29	[0.514 ; 1.021]	0.767	*<0.001*	
	LVH	1.014	29	[−0.142; 0.422]	0.140	0.319	
Clavicle	Ln	−1.098	39	[−0.105; 0.031]	−0.037	0.279	
Bone	**Variable**	***V***		**CI 95% (mm)**	**Pseudo-median (mm)**	**p-value**	
Ilium	IL	100		[−0.342; −0.057]	−0.173	0.006	
	IM[a]	20		[−104.029; −41.998]	−68.450	*<0.001*	
	IA[a]	224		[−36.777; 34.443]	−3.482	0.871	
Fifth lumbar	LVL	73		[−0.634; −0.213]	−0.426	*<0.001*	
vertebra	LVW	123		[−0.361; −0.028]	−0.185	0.023	
	LVM[a]	58		[−30.325; −12.122]	−21.093	*<0.001*	
Clavicle	Max_diam	720.5		[0.0125; 0.029]	0.020	*<0.001*	
	Min_diam	676.5		[0.0075; 0.022]	0.015	*<0.001*	
	AP_diam	214		[−0.248; −0.032]	−0.123	0.008	
	SI_diam	194		[−0.342; −0.047]	−0.166	0.003	

IW: Ilium Width; UVL: Upper Vertebral Length; UVW: Upper Vertebral Width; UVM: Upper Vertebral Module; PVH: Posterior Vertebral Height; AVH: Anterior Vertebral Height; RVH: Right Vertebral Height; LVH: Left Vertebral Height; Ln: Maximum clavicular length; IL: Ilium Length; IM: Ilium Module; IA: Ilium Area; LVL: Lower Vertebral Length; LVW: Lower Vertebral Width; LVM: Lower Vertebral Module; Max_diam: Maximum diameter at half-meximum length; Min_diam: Minimum diameter at half-maximum length; AP_diam: Anteroposterior diameter at half-meximum length; SI_diam: Supero-inferior diameter at half-maximum length.

means that these two lumbar variables are not measurable by two different observers with an error of less than 5%.

Additional results for landmark and variable repeatability and reproducibility are given in the supplementary tables (Tables S1–S4).

Discussion

While the acquisition of maximum clavicular length (Ln), Max_diam, and Min_diam from dry bones often requires several attempts with the sliding calliper before finding the right points, they can be reliably, semiautomatically, and directly taken using virtual measurement tools provided by image reconstruction and treatment software, provided the bone surfaces are reconstructed correctly and the bones are reoriented in the same plane.

The extreme morphological variability of the clavicle [25,38] had already raised questions over perspective and bone positioning in the use of two-dimensional medical images of this bone for forensic purposes [16,54]. In this study, clavicular morphology raised questions about orientation and landmark positioning on the reconstructed surfaces to acquire anatomical variables. The two clavicular variables AP_diam and SI_diam are well known to anthropologists, well defined [33,35] and "easy" to acquire on dry bones. Landmarks had been previously defined to acquire these variables on virtual bone surfaces [18], but significant errors had been identified. One explanation was that the landmarks were not type I (anatomical) so could not be placed with sufficient precision [6,55]. This study shows that four

Table 8 Correlation coefficients for the repeatability (consistency) and reproducibility (agreement) of variables. ICC = intraclass correlation coefficient. CI = confidence interval of the ICC. All *p*-values are significant (*p*<0.001) / *Coefficients de corrélation de la répétabilité (consistency) et de la reproductibilité (agreement) des variables. ICC = coefficient de corrélation intraclasse. CI = intervalle de confiance du ICC. Toutes les p-values sont significatives (p < 0,001)*

Bone	Variable	*F*	95% CI	ICC (consistency)	*F*	CI 95%	ICC (agreement)
Ilium	IL	11655	[0.999; 1]	0.9998	10875	[0.999; 1]	0.9998
	IW	4087	[0.999; 1]	0.9996	2026	[0.998; 1]	0.9990
	IM	6491	[0.999; 1]	0.9997	5419	[0.999; 1]	0.9996
	IA	3782	[0.999; 1]	0.9995	2009	[0.998; 1]	0.9990
Fifth lumbar vertebra	UVL	514	[0.992; 0.998]	0.996	490	[0.992; 0.998]	0.996
	UVW	789	[0.995; 0.999]	0.997	817	[0.995; 0.999]	0.998
	UVM	847	[0.995; 0.999]	0.998	996	[0.996; 0.999]	0.998
	LVL	798	[0.995; 0.999]	0.997	226	[0.982; 0.996]	0.991
	LVW	1713	[0.998; 0.999]	0.999	930	[0.996; 0.999]	0.998
	LVM	2570	[0.998; 1]	0.999	473	[0.991; 0.998]	0.996
	PVH	325	[0.987; 0.997]	0.994	104	[0.961; 0.991]	0.981
	AVH	401	[0.990; 0.998]	0.995	316	[0.987; 0.997]	0.994
	RVH	367	[0.989; 0.997]	0.995	137	[0.970; 0.993]	0.986
	LVH	215	[0.981; 0.996]	0.991	231	[0.982; 0.996]	0.991
Clavicle	Ln	48765	[1; 1]	1	73579	[1; 1]	1
	Max_diam	18799	[1; 1]	1	391	[0.990; 0.997]	0.995
	Min_diam	15290	[1; 1]	1	391	[0.990; 0.997]	0.995
	AP_diam	136	[0.973; 0.992]	0.985	167	[0.978; 0.994]	0.988
	SI_diam	76.2	[0.952; 0.986]	0.974	71	[0.948; 0.985]	0.972

IL: Ilium Length; IW: Ilium Width; IM: Ilium Module; IA: Ilium Area; UVL: Upper Vertebral Length; UVW: Upper Vertebral Width; UVM: Upper Vertebral Module; LVL: Lower Vertebral Length; LVW: Lower Vertebral Width; LVM: Lower Vertebral Module; PVH: Posterior Vertebral Height; AVH: Anterior Vertebral Height; RVH: Right Vertebral Height; LVH: Left Vertebral Height; Ln: Maximum clavicular length; Max_diam: Maximum diameter at half-maximum length; Min_diam: Minimum diameter at half-maximum length; AP_diam: Anteroposterior diameter at half-meximum length; SI_diam: Supero-inferior diameter at half-maximum length.

landmarks can be placed on the anterior, posterior, inferior, and superior sides of the bone at mid-length and reliably define the two clavicular variables AP_diam and SI_diam on virtual surfaces.

Our simple protocol for bone realignment ensures that clavicles are placed along the same geometric plane before these two anatomical variables are measured. The anatomical orientation of the bones is obtained by this common geometric plane and based on the assumption that the acromial surface is equivalent to the superior plane. The variability in clavicular morphology could possibly bring the latter anatomical postulate into question, and it should therefore be verified by a more extensive study on the variability of clavicular morphology. The advantage of using geometric variables (Max_diam and Min_diam) is their independence from anatomical orientation, which induces fewer errors due to varying orientations. This could explain why the errors between the clavicular variables AP_diam and SI_diam

measured on dry and reconstructed bone surfaces are slightly higher than the geometric variables. The protocol for measuring diameters is only applicable to fully preserved clavicles, as there is no anatomical marker at mid-shaft. Therefore, a precise and reliable location for the landmarks depends on the total length of the bone [56], which in this study is measured by very low error rates.

Even if not all landmarks are repeatable or reproducible, the variables, whether defined by landmarks, obtained by geometric construction or composite, are both repeatable and reproducible with a reliability higher than 95%, with the exception of the lumbar variables RVH and PVH, which present %TEM values above 5% and significant inter-observer errors (Tables 8 and 10). This exception cannot be explained by the type of landmark (type II) or by anatomical location (upper central and lateral ridges of the lumbar vertebral body) as neither of these two factors is specific to these variables or landmarks. Based on these results,

Table 9 Intra- and inter-observer technical errors of measurement (TEM), relative TEM (%TEM) and associated coefficients of reliability. Values in italics indicate significant intra- or inter-observer differences for the corresponding variables (TEM comparable or higher than 5%) / *Erreurs techniques de mesure absolues (TEM) et relatives (%TEM), coefficients de fiabilité associés. Les valeurs en italique indiquent des différences intra- ou interobservateur non négligeables pour les variables correspondantes (%TEM proches ou supérieures à 5 %)*

Bone	Variable	Intra-observer		Inter-observer		Intra-observer reliability	Inter-observer reliability
		TEM (mm)	% TEM	TEM (mm)	% TEM		
Ilium	IL	0.299	0.332	0.309	0.343	0.999	0.999
	IW	0.466	0.562	0.661	0.795	0.999	0.999
	IM[a]	68.438	0.863	74.739	0.940	0.999	0.999
	IA[a]	62.460	1.122	86.118	1.547	0.999	0.999
Fifth lumbar vertebra	UVL	0.418	1.898	0.433	1.960	0.996	0.996
	UVW	0.469	1.291	0.461	1.268	0.997	0.998
	UVM[a]	21.283	2.479	19.887	2.304	0.998	0.998
	LVL	0.341	1.557	0.647	2.928	0.997	0.991
	LVW	0.322	0.900	0.438	1.225	0.999	0.998
	LVM[a]	12.021	1.429	28.462	3.344	0.999	0.996
	PVH	0.448	3.228	0.769	*5.587*	0.994	0.981
	AVH	0.398	2.434	0.450	2.774	0.995	0.994
	RVH	0.446	3.379	0.720	*5.581*	0.994	0.985
	LVH	0.584	4.505	0.535	4.173	0.991	0.991
Clavicle	Ln	0.178	0.156	0.165	0.145	0.999	0.999
	Max_diam	0.031	0.361	0.222	2.110	0.999	0.993
	Min_diam	0.031	0.297	0.194	2.294	0.999	0.993
	AP_diam	0.322	3.267	0.315	3.153	0.983	0.983
	SI_diam	0.415	*4.758*	0.399	4.553	0.965	0.969

IL: Ilium Length; IW: Ilium Width; IM: Ilium Module; IA: Ilium Area; UVL: Upper Vertebral Length; UVW: Upper Vertebral Width; UVM: Upper Vertebral Module; LVL: Lower Vertebral Length; LVW: Lower Vertebral Width; LVM: Lower Vertebral Module; PVH: Posterior Vertebral Height; AVH: Anterior Vertebral Height; RVH: Right Vertebral Height; LVH: Left Vertebral Height; Ln: Maximum clavicular length; Max_diam: Maximum diameter at half-maximum length; Min_diam: Minimum diameter at half-maximum length; AP_diam: Anteroposterior diameter at half-maximum length; SI_diam: Supero-inferior diameter at half-maximum length; [a]: Indicates values in mm^2.

it seems that the presence of substantial intra- and inter-observer errors in landmark positioning does not seem to induce important intra- or inter-observer errors for the corresponding variables.

The eight lumbar variables defined by type II landmarks and the two clavicular variables defined by type III landmarks did not systematically present higher error rates than the iliac variable IL defined by type I landmarks.

Variables obtained without the direct or indirect (for composite variables) use of landmarks (clavicular variables Ln, Max_diam, Min_diam, and the iliac variable IA) and variables obtained using landmarks did not present substantially different intra- or inter-observer error rates. The use of landmarks or not to obtain osteometric variables and the type of landmark used to obtain measurements do not seem to have

a significant influence on the presence of intra- or inter-observer errors or on the range of those errors, although this could be tested in another study. Possible sources of error could be unclear landmark definitions (especially for type II and III), or the influence of the quality of acquisition and reconstruction modes, which lead to various degrees of smoothing of the reconstructed bone surfaces that can affect anatomical structures. Placing landmarks is more difficult when the anatomical reliefs are smoothed or when segmentation is not optimal [6]. However, this study showed that even with bone surfaces reconstructed from CT scans of living patients, using a simple segmentation method (HMH) to separate bone surfaces from all the surrounding soft tissue, a large variety of osteometric variables could be obtained with sufficient repeatability and reproducibility [18].

Inter-observer errors are more often significant than intra-observer errors (Table 10). For example, the Student and Wilcoxon tests detected seven variables for which inter-observer errors are significant against three with significant intra-observer errors. However, the interpretations of the *t*-tests or Wilcoxon tests often contradict the conclusions obtained using the 95% CIs for error distributions, Bland–Altman plots, TEM or ICC (Table 10), even after applying the Bonferroni adjustment to the *p*-value (Tables 6 and 7). Indeed, the Bonferroni adjustment systematically indicates a lower number of variables for which both intra- and inter-observer errors are significant. This indicates that errors could be overestimated if this correction is not applied. It therefore seems important to use it, to avoid falsely rejecting variables that apparently present significant errors. Moreover, the tests do not give consistent results for the same types of variables. The ranges of intra- and inter-observer errors for unidimensional and for bidimensional variables (IM and IW for the ilium, RVH and LVM for the lumbar vertebrae) are not always consistent with the conclusions obtained with the adjusted or nonadjusted *p*-values of the statistical tests (Table 10). Graphical or quantified evidence of errors can be used to assess the repeatability and reproducibility of any type of data, without having to test the normality of the distribution, as is the case in order to apply statistical tests.

This study provides clear examples of the limited relevance of using statistical tests, the caution needed when applying and interpreting them and their associated *p*-values

Table 10 Summary of the results of the tests and parameters for evaluating repeatability and reproducibility of the variables. "Checks" indicate sufficient repeatability or reproducibility, crosses indicate insufficient repeatability or reproducibility / *Récapitulatif de l'ensemble des tests et paramètres d'évaluation de la répétabilité et la reproductibilité des variables de l'étude. Les « tick » indiquent une répétabilité ou une reproductibilité suffisante, les croix indiquent une répétabilité ou une reproductibilité insuffisantes*

Bone	Variable	Repeatability				Reproducibility			
		Wilcoxon *t*-test	Bland-Altman	ICC	TEM − % TEM	Wilcoxon *t*-test	Bland-Altman	ICC	TEM − % TEM
Ilium	IL	✓	✓	✓	✓	✓	✓	✓	✓
	IW	×	✓	✓	✓	×	✓	✓	✓
	IM	✓	✓	✓	✓	×	✓	✓	✓
	IA	✓	✓	✓	✓	✓	✓	✓	✓
Fifth lumbar vertebra	UVL	✓	✓	✓	✓	✓	✓	✓	✓
	UVW	✓	✓	✓	✓	✓	✓	✓	✓
	UVM	✓	✓	✓	✓	✓	✓	✓	✓
	LVL	✓	✓	✓	✓	×	✓	✓	✓
	LVW	✓	✓	✓	✓	✓	✓	✓	✓
	LVM	✓	✓	✓	✓	×	✓	✓	✓
	PVH	✓	✓	✓	✓	✓	✓	✓	×
	AVH	✓	✓	✓	✓	✓	✓	✓	✓
	RVH	✓	✓	✓	✓	×	✓	✓	×
	LVH	✓	✓	✓	✓	✓	✓	✓	✓
Clavicle	Ln	✓	✓	✓	✓	✓	✓	✓	✓
	Max_diam	×	✓	✓	✓	×	✓	✓	✓
	Min_diam	×	✓	✓	✓	×	✓	✓	✓
	AP_diam	✓	✓	✓	✓	✓	✓	✓	✓
	SI_diam	✓	✓	✓	✓	✓	✓	✓	✓

IL: Ilium Length; IW: Ilium Width; IM: Ilium Module; IA: Ilium Area; UVL: Upper Vertebral Length; UVW: Upper Vertebral Width; UVM: Upper Vertebral Module; LVL: Lower Vertebral Length; LVW: Lower Vertebral Width; LVM: Lower Vertebral Module; PVH: Posterior Vertebral Height; AVH: Anterior Vertebral Height; RVH: Right Vertebral Height; LVH: Left Vertebral Height; Ln: Maximum clavicular length; Max_diam: Maximum diameter at half-maximum length; Min_diam: Minimum diameter at half-maximum length; AP_diam: Anteroposterior diameter at half-maximum length; SI_diam: Supero-inferior diameter at half-maximum length.

to assess observer or measurement agreement [57,58] and the necessity of using corrective factors such as Bonferroni for multiple testing. In comparison, other descriptive statistical parameters such as ICC, TEM, and Bland–Altman plots can be used to evaluate the repeatability and reproducibility of any type of osteometric variable without worrying about such limitations.

Rather than relying on the single statistical parameter that is a *p*-value, this study confirms that it appears easier, more relevant, and more valid to conclude on variable repeatability or reproducibility by comparing the range of errors to the range of values for the variables and using graphical representations of the dispersions of errors [23].

Conclusions

This study assesses the consistency between variables defined for dry clavicles and measured on dry clavicles and bone surfaces virtually reconstructed from CT scans of the same dry clavicles. Osteometric, geometric, and composite variables such as the ones used here can be obtained from virtual bone surfaces reconstructed from CT scans with sufficient repeatability and reproducibility. Neither the type of bone (long, flat, short), nor the type of landmark used to obtain the variables seem to be associated with the presence of large intra- or inter-observer errors, although this last point would require further investigation on a more diversified sample of bones. In this study, the repeatability and reproducibility of variables are apparently not significantly affected by errors in landmark positioning. Nevertheless, any biometric study involving both landmarks and variables should systematically check for intra- and inter-observer errors for both landmark positioning and variables.

With the example of verifying repeatability and reproducibility, this study has shown the limitations and pitfalls of interpreting errors solely on the basis of uncorrected *p*-values associated with statistical tests (Student, Wilcoxon), which give inconsistent results and therefore should not be used without corrective factors, such as the Bonferroni adjustment. However, to obtain a complete and accurate overview of the data, the statistical interpretation of any type of error should be primarily based on graphical representations of the data (e.g., Bland–Altman plots) or coefficients to assess reliability (e.g., ICC), and should quantify errors (e.g., by using TEM and %TEM).

References

1. Weber GW, Bookstein FL (2011) Virtual anthropology: a guide to a new interdisciplinary field. Springer, Wien
2. Mueller WH, Martorell R (1988) Reliability and accuracy of measurement. In: Lohman TG, Roche AF, Martorell R (eds) Anthropometric standardisation reference manual. Human Kinetic Books, Champaign, IL, pp 83–6
3. Hildebolt CF, Vannier MW, Knapp RH (1990) Validation study of skull three-dimensional computerized tomography measurements. Am J Phys Anthropol 82:283–94
4. Richtsmeier JT, Paik CH, Elfert PC, et al (1995) Precision, repeatability, and validation of the localization of cranial landmarks using computed tomography scans. Cleft Palate Craniofac J 32:217–27
5. Jamison PL, Zegura SL (1974) A univariate and multivariate examination of measurement error in anthropometry. Am J Phys Anthropol 40:197–204
6. Bookstein FL (1991) Morphometric tools for landmark data: geometry and biology. Cambridge University Press, New York, 435 pp
7. Dirkmaat D, Cabo LL, Ousley SD, et al (2008) New perspectives in forensic anthropology. Yearb Phys Anthropol 51:33–52
8. Harris EF, Smith RN (2009) Accounting for measurement error: a critical but often overlooked process. Arch Oral Biol 54S:S107–S17
9. Brough AL, Bennett J, Morgan B, et al (2013) Anthropological measurement of the Juvenile clavicle using multi-detector computed tomography — affirming reliability. J Forensic Sci 58:946–51
10. Robinson C, Eisma R, Morgan B, et al (2008) Anthropological measurement of lower limb and foot bones using multi-detector computed tomography. J Forensic Sci 53:1289–95
11. Goto R, Nicholas Mascie-Taylor CG (2007) Precision of measurement as a component of human variation. J Phys Anthropol 26:253–56
12. Sicotte M, Ledoux M, Zunzunegui MV, et al (2010) Reliability of anthropometric measures in a longitudinal cohort of patients initiating ART in West Africa. BMC Med Res Method 10:102
13. Citardi MJ, Hermann B, Hollenbeak CS, et al (2001) Comparison of scientific calipers and computer-enabled CT review for the measurement of skull base and craniomaxillofacial dimensions. Skull Base 11:5–12
14. Waitzman AA, Posnick JC, Armstrong DC, et al (1992) Craniofacial skeletal measurements based on computed tomography: Part I. Accuracy and reproducibility. Cleft Palate Craniofac J 29:217–27
15. Lopes PM, Moreira CR, Perrella A, et al (2008) 3-D volume rendering maxillofacial analysis of angular measurements by multislice CT. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 105:224–30
16. Brough AL, Rutty GN, Black S, Morgan B (2012) Post-mortem computed tomography and 3D imaging: anthropological applications for juvenile remains. Forensic Sci Med Pathol 8:270–79
17. Lou L, Lagravere MO, Compton S, et al (2007) Accuracy of measurements and reliability of landmark identification with

computed tomography (CT) techniques in the maxillo-facial area: a systematic review. Oral Surg Oral Med Oral Patho Oral Radiol Endod 104:402–11

18. Stull KE, Tise ML, Ali Z, et al (2014) Accuracy and reliability of measurements obtained from computed tomography 3D volume rendered images. Forensic Sci Int 238:133–40

19. Guyomarc'h P, Santos F, Dutailly B, et al (2012) Three-dimensional computer-assisted craniometrics: a comparison of the uncertainty in measurement induced by surface reconstruction performed by two computer programs. Forensic Sci Int 219:221–27

20. Shirley NR, Ramirez-Montes PA (2015) Age estimation in forensic anthropology: Quantification of observer error in phase versus component-based methods. J Forensic Sci 60:107–11

21. Utermohle CJ, Zegura SL (1981) Intra- and inter-observer error in craniometry: a coutionary tale. Am J Phys Anthropol 57:303–10

22. Ferrante L, Cameriere R (2009) Statistical methods to assess the reliability of measurements in the procedures for forensic age estimation. Int J Legal Med 123:277–83

23. Santos F (2014) Evaluer les erreurs de mesure en anthropométrie. Personal communication, CNRS, UMR 5199 PACEA, 13 pp

24. Langley-Shirley N, Jantz RL (2010) A Bayesian approach to age estimation in modern Americans from the clavicle. J Forensic Sci 55:571–83

25. Cook A, Sripada S, Soames R, et al (2013) Assessing the variability in clavicular anatomy: a cadavaric study. Orthop Proc 95-B:25

26. Dohr M (2012) Nancy, Meurthe-et-Moselle. Ilôt Berger-Levrault. Du village Saint-Dizier au cimetière des Trois Maisons. Rapport Final d'Opération de fouille préventive 2010, INRAP Grand Est Nord (eds) : Service Régional de l'Archéologie de Lorraine, 4 vol.

27. Corron L (2016) Juvenile age estimation in physical anthropology: a critical review of existing methods and the application of two standardised methodological approaches. Université Aix-Marseille, Marseille

28. Martin R, Saller K (1957) Lehrbuch der Anthropologie in systematischer Dartstellung mit besonderer Berücksichtigung der anthropologischer Methoden, Vol. I. Stuttgart, Gustav Fischer Verlag, 661 pp

29. Scheuer L, Black S (2000) Developmental juvenile osteology. Gray Publishing, E.A. Press, San Diego, 587 pp

30. Scheuer L, Black S (2004) The juvenile skeleton. E.A. Press, London, 485 pp

31. Martin R, Saller K (1959) Lehrbuch der Anthropologie in systematischer Darstellung mit besonderer Berücksichtigung der anthropologischen Methoden, Vol. II. Gustav Fischer Verlag, Stuttgart, 690 pp

32. Olivier G (1965) Anatomie anthropologique. Vigot frères (eds), Paris, 488 p.

33. Moore-Jansen PM, Ousley SD, Jantz RL (1994) Data collection procedures for forensic skeletal material. The University of Tennessee Department of Anthropology and Forensic Anthropology Center, Knoxville

34. Black S, Scheuer L (1996) Age changes in the clavicle: from the early neonatal period to skeletal maturity. Int J Osteoarchaeol 6:425–34

35. Buikstra J, Ubelaker D (1994) Standards for data collection from human skeletal remains. In: Buikstra J (ed) Proceedings of a seminar at the Field Museum of Natural History. Arkansas Archaeological Survey, Fayetteville, 272 p.

36. Miles AEW, Bulman JS (1995) Growth curves of immature bones from a Scottish island population of sixteenth to mid-nineteenth century: Shoulder girdle, ilium, pubis and ischium. Int J Osteoarchaeol 5:15–27

37. Spoor CF, Zonneveld FW, Macho GA (1993) Linear measurements of cortical bone and dental enamel by computed tomography: Applications and problems. Am J Phys Anthropol 91:469–84

38. Voisin JL, Balzeau A (2004) Structures internes claviculaires chez *Pan, Gorilla* et *Homo*. Méthodes d'analyses et résultats préliminaires. Bull Mem Soc Anthropol 16:5–16

39. Zhou SH, McCarty ID, McGregor AH, et al (2000) Geometrical dimensions of lower lumbar vertebrae—analysis of data from digitized CT images. Eur Spine J 9:242–48

40. Berry JL, Moran JM, Berg WS, et al (1987) A morphometric study of human lumbar and selected thoracic vertebrae. Spine (Phila Pa 1976) 12:362–67

41. Fazekas IG, Kosa F (1978) Forensic fetal osteology. Akademiai Kiado, Budapest, 414 pp

42. Mavrych V, Bolgova O, Ganguly P, et al (2014) Age-related changes of lumbar vertebral body morphometry. Austin J Anat 1:1014–21

43. Panjabi MM, Goel V, Oxland T, et al (1992) Human lumbar vertebrae quantitative three-dimensional anatomy. Spine (Phila Pa 1976) 17:299–306

44. Fang D, Cheung K, Ruan D, et al (1994) Computed tomographic osteometry of the Asian lumbar spine. J Spinal Disord Tech 7:307–16

45. Jones RAC, Thomson LG (1968) The narrow lumbar canal — a clinical and radiological review. J Bone Joint Surg 50:595–605

46. Adalian P (2001) Évaluation mulitparamétrique de la croissance fœtale—Application à la détermination de l'âge et du sexe. Faculté de Med., Univ. Méditerranée, Marseille, 269 pp

47. Marchal F (1997) L'os coxal des Hominidés fossiles (Tome I). Univ. Méditerranée, Marseille, 337 pp

48. Bland M, Altman D (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307–10

49. Shaffer JP (1995) Multiple hypothesis testing. Annu Rev Psychol 46:561–84

50. Fisher RA (1925) Statistical methods for research workers. Oliver & Boyd, Edinburgh, 145 p.

51. Ulijaszek SJ, Kerr DA (1999) Anthropometric measurement error and the assessment of nutritional status. Br J Nutr 82:165–77

52. Ross WD, Kerr DA, Carter JEL, et al (1994) Anthropometric techniques; precision and accuracy. In: Carter JEL, Ackland TR (eds) Kinanthropometry in aquatic sports. A study of world class athletes. Human Kinetics Books, Champaign, IL, pp 158–73

53. Norton K, Olds T (1996) Anthropometrica: A textbook of body measurement for sports and health courses, ed. Australian Sports Commission, University of New South Wales Press, 413 p.

54. Stephan CN, Guyomarc'h P (2014) Quantification of perspective-induced shape changes of clavicles at radiography and 3D scanning to assist human identification. J Forensic Sci 59:447–53

55. Sholts SB, Flores L, Walker PL, Wärmländer SKTS (2011) Comparison of coordinate measurement precision of different landmark types on human crania using a 3D laser scanner and a 3D digitiser: Implications for applications of digital morphometrics. Int J Osteoarchaeol 21:535–43

56. Sladek V, Berner M, Galeta P, et al (2010) Technical note: the effect of midshaft location on the error ranges of femoral and tibial cross-sectional parameters. Am J Phys Anthropol 141:325–32

57. Lin LI-K (1989) A concordance correlation coefficient to evaluate reproducibility. Biometrics 45:255–68

58. Colquhoun D (2014) An investigation of the false discovery rate and the misinterpretation of p-values. R Soc Open Sci 1:2-16