



**HAL**  
open science

## **Assessing biases in the analysis of trabecular bone microarchitecture of non-adult individuals**

Antony Colombo, Bruno Dutailly, Romain Leguay, Olivier Dutour, H el ene  
Coqueugniot

### ► **To cite this version:**

Antony Colombo, Bruno Dutailly, Romain Leguay, Olivier Dutour, H el ene Coqueugniot. Assessing biases in the analysis of trabecular bone microarchitecture of non-adult individuals. *Journal of Archaeological Science: Reports*, 2021, 39, pp.103138. <10.1016/j.jasrep.2021.103138>. <hal-03368597>

**HAL Id: hal-03368597**

**<https://hal.science/hal-03368597v1>**

Submitted on 6 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.



HAL Authorization

**Title: Assessing biases in the analysis of trabecular bone microarchitecture of non-adult individuals**

**Authors' names:** COLOMBO Antony<sup>a,b\*</sup>, DUTAILLY Bruno<sup>c,d</sup>, LEGUAY Romain<sup>e</sup>, DUTOIR Olivier<sup>b,c,f</sup> and COQUEUGNIOT H el ene<sup>b,c,g</sup>

**Authors' affiliations:**

- a- Universit  di Bologna, dipartimento di Scienze Biologiche, Geologiche e Ambientali – BiGeA, Laboratorio di Antropologia fisica, via Francesco Selmi, 3, 40126 Bologna, Italy.
- b- Ecole Pratique des Hautes Etudes-PSL University, Chair of biological anthropology Paul Broca, 4-14 rue Ferrus, F-75014 Paris, France.
- c- UMR 5199 PACEA, University of Bordeaux, CNRS, MCC ; LabEx Sciences arch ologiques de Bordeaux, n  ANR-10-LABX-52, b t B8, all e Geoffroy Saint Hilaire, CS50023, F-33615 Pessac, France.
- d- Bordeaux Montaigne University, UMS 3657 ARCHEOVISION, F-33607 Pessac, France.
- e- Nurea, 37 cours Georges Cl menceau, F-33000 Bordeaux, France.
- f- The University of Western Ontario, Department of Anthropology, N6A-3K7, London, Canada.
- g- Max Planck Institute for Evolutionary Anthropology, Department of Human Evolution, Deutscher Platz 6, D-04103 Leipzig, Germany.

**Keywords:** cancellous bone, volume of interest, 3D, micro-computed tomography

**\* Corresponding author:** Dr Antony Colombo; *address:* Universit  di Bologna, dipartimento di Scienze Biologiche, Geologiche e Ambientali – BiGeA, Laboratorio di Antropologia fisica, via Francesco Selmi, 3, 40126 Bologna, Italy. *Telephone number:* +33.6.08.35.79.44; *e-mail address:* colomboantony@gmail.com; antony.colombo@unibo.it

**Abstract:** Micro-computed tomography ( $\mu$ CT) is a non-destructive x-ray technique that can provide high-resolution images of skeletal material, making it an attractive option for analysing archaeological human remains. However,  $\mu$ CT data acquisition, image segmentation and 3D reconstruction may bias the quantification of trabecular bone microarchitecture (TBMA) parameters. Some biases have been evaluated in previous studies, but never in non-adult individuals at different stages of development. Five non-pathological humeral metaphyses from known-age individuals of various developmental stages were  $\mu$ CT-scanned at a high resolution based on bone size. First, repeatability and reproducibility of the volume of interest selection protocol (VOI, trabecular bone region analysed) have been evaluated and found to produce non-significant variations of the TBMA parameters. Then, each bone was assessed for biasing effects related to image resolution and segmentation, as well as the size and position of the VOI. According to the bias tested, individuals of different developmental stages and parameters are differentially affected. By progressing through the stages of development, the TBMA parameters seem less sensitive to the variation introduced by the bias tests. Perinatal stage and mature adolescent stage distinctly differ from others depending on the tests and parameters measured. Bone volume fraction, trabecular thickness and trabecular number are the most sensitive and correlated parameters to the biases tested. We therefore propose specific recommendations to

assure homologous VOI selection and reliable and comparable results for non-adult TBMA analyses.

## 1. Introduction

The use of 3D imaging methods (*e.g.* CT-scan, photogrammetry, lasergrammetry) in bioarchaeology has greatly increased in recent years. These techniques, especially X-ray-based ones, are non-destructive, preserving the integrity of archaeological material and, in particular, human remains. Imaging allows access to the inner structures of bony elements providing the basis for new bioarchaeological data. Micro-computed tomography ( $\mu$ CT) provides images of bone structures with a level of detail comparable to that of histology for histomorphometry (Fajardo et al., 2002). The 3D microstructures of bone accessed in this case are the trabecular bone microarchitecture (TBMA). TBMA refers to the organisation of the network of the trabeculae, which are the constituent elements of the trabecular bone. Various parameters can be measured to describe it (Bouxsein et al., 2010). The analyses based on such data can inform on factors that influence skeletal modelling and remodelling processes that are crucial for understanding the lifeways of past populations. Nevertheless, TBMA analysis from  $\mu$ CT data is only possible if the potentially bias caused by (1) image acquisition, (2) segmentation and (3) 3D reconstruction on quantified TBMA parameters are identified and controlled.

Reliability of the results is dependent on  $\mu$ CT-scanner performance and appropriate settings (voltage, intensity, captors' size, *etc.*). Due to variation in device performance, it is recommended that all  $\mu$ CT data acquisition for a given study is performed using the same  $\mu$ CT-scanner (Bouxsein et al., 2010; Verdelis et al., 2011). If that is not possible, studies based on data produced by different  $\mu$ CT-scanners should ensure that they are similarly calibrated with equivalent settings. Settings are also dependent on the physical characteristics (*e.g.* thickness, density) of the sample. Moreover, the environment in which the sample is scanned also affects image quality and a better contrast is obtained when the environment is air (Bouxsein et al., 2010). Acquired images have their own "signal to noise" ratio that must be as high as possible and can be improved with higher exposure time (Nägele et al., 2004) and/or a higher frame averaging (Bouxsein et al., 2010) during acquisition.

The resolution of acquisition has a major influence on TBMA quantification (Kim et al., 2004; Kothari et al., 1998; Müller et al., 1996; Peyrin et al., 1998; Sode et al., 2008). Image resolution needs to be as high as possible to obtain the closest results to biological reality. Measurements of TBMA parameters are dependent on both acquisition and reconstruction resolutions (Kothari et al., 1998). For example, if the resolution is not high enough, it can produce up to a 100% over-estimation in comparison to true biological values (Kim et al., 2004). Nevertheless, the size of the anatomical region of interest to be  $\mu$ CT-scanned depends on the resolution of acquisition and *vice versa*. The higher the resolution, the more detailed trabeculae (higher number of voxels for one trabecula), the smaller the acquired anatomical region. It is therefore necessary in the case of human bone analysis to define the anatomical area prior to  $\mu$ CT-acquisition to achieve the highest resolution possible. Some authors argued that a resolution lower than 75  $\mu$ m creates too much measurement error in human TBMA studies to be acceptable (Müller et al., 1996; Salome-Pateyron et al., 1997). Thus, to minimize error, it is recommended to choose a resolution that allows at least three to four voxels in the thickness of trabeculae (Bouxsein et al., 2010; Guldborg et al., 1998).

Segmentation is a form of image processing that automatically or manually separates the raw image into different subsets (Cocquerez and Philipp-Foliguet, 1995). In TBMA studies, segmentation separates the trabecular bone from the background and transforms the  $\mu$ CT image into a binary image composed of only black and white voxels. The segmentation method used affects both the reliability and precision of this separation (Cendre et al., 2000; Elmoutaouakkil et al., 2002; Scherf and Tilgner, 2009; Way et al., 2008). A “simple threshold” segmentation method is enough for  $\mu$ CT images because image noise is very low and the contrast between bone and background is very high (Elmoutaouakkil et al., 2002). However, if the threshold value remains the same from a sample to another, structural measurements will be under- or overestimated (Coleman and Colbert, 2007; Goo et al., 2005; Hara et al., 2002; Way et al., 2008). Therefore, segmentation requires great caution and thresholding that is adapted to each bone sample. A partial volume effect (a blurred transition between two subsets of an image) also indirectly influences the segmentation process (Elmoutaouakkil et al., 2002; Way et al., 2008). It is linked to the image resolution (Pyka et al., 2014) and the characteristics of the CT device.

$\mu$ CT scans can produce very large data sets, requiring high-performance computers for analyses. For this reason, most of the TBMA studies focus on a restricted area of bone tissue: the volume of interest (VOI). A lack of precision in VOI selection can skew the results (Nägele et al., 2004). Indeed, microarchitectural parameters are not affected in the same way for a given VOI if misplaced by only a few millimetres (Kivell et al., 2011; Nägele et al., 2004). Several authors have shown that changes in both VOI location and size have important and variable effects on most of the commonly studied microanatomical parameters, which also vary according to the bone topography (Fajardo and Müller, 2001; Fajardo et al., 2007; Kivell et al., 2011; Lazenby et al., 2011; Maga et al., 2006). The trabecular structure can be site-specific, therefore VOI location must be carefully controlled (Lazenby et al., 2008). VOIs of identical dimensions to study bones with different sizes produce an over-sampling effect, which could amplify the magnitude of differences observed between individuals. Thus, scaling criteria have to be considered (Lazenby et al., 2011). The VOI shape can also affect the results, and a spherical VOI is preferable to a cubic VOI (Ketcham and Ryan, 2004). This is because a sphere is defined by both its centre and radius, and its orientation is then independent of the acquisition axis; spherical VOIs avoid bias due to the variation of bone sample position and orientation within the  $\mu$ CT-scanner.

While there are some published recommendations for TBMA studies of bone development, particularly on how the VOI location (Burrows et al., 2010) and VOI size (Bouxsein et al., 2010; Fajardo and Müller, 2001) should be defined according to the bone length, most of these studies were conducted on bone samples from adult individuals and/or species other than *Homo sapiens*. Bone size, shape, and the extent of mineralisation change during development (Bouxsein et al., 2010). Therefore, a standardised protocol for  $\mu$ CT acquisition is even more important in ontogenetic studies. The potential sources of bias when characterizing TBMA - position and size of VOI, resolution and segmentation of images - are not sufficiently considered for studies of non-adult samples and can be a problem because of the size and shape changes with the development.

This methodological paper aims to evaluate bias effects related to image resolution and segmentation, as well as the size and position of the VOI at different stages of development of non-adult individuals.

The questions asked are: (1) Do known biases affect the measured parameters at each stage of development? (2) If yes, do the biases affect the measured parameters in the same way at each stage of development?

In light of the results, specific recommendations for TBMA studies of non-adults are proposed in order to create reproducible and repeatable VOI selection protocols.

## **2. Material and methods**

### ***2.1. Material***

The humerus proximal metaphysis has been selected for three main reasons: (1) the humerus experiences minimal shape changes throughout human ontogeny and (2) is less biomechanically constrained by body mass and locomotion than bones of the lower limbs (Ruff, 2003) and so likely to exhibit TBMA developmental changes, and (3) its proximal cartilage is responsible for 80% of adult bone length (Pritchett, 1991) and then a better candidate than the distal metaphysis to study bias effect in the context of human bone growth.

To quantify the major sources of bias when studying TBMA of non-adult's bones, we selected a sample of five non-adult proximal humeral metaphyses from collection of individuals of known age-at-death and sex at the Institute of Normal Anatomy, University of Strasbourg (Coqueugniot and Hublin, 2012; Rampont, 1994). A mixed-sex sample of individuals at different stages of development were selected: approximately at birth, during childhood, and before, during, and after adolescence (Table 1). Before adolescence there are no known differences between males and females in TBMA parameters (Modlesky et al., 2011), and after adolescence, sex differences are possible depending on the bone location analysed (Liu et al., 2010). Here, the sex of individuals is not considered as the goal is to identify individual variation of the biases at different stage of development and not to compare individuals between them. Neither the individuals nor the bony elements selected showed any pathological or taphonomic changes.

The sample size might be considered small, but the purpose here is to test the effect of calibrated variations on the main sources of bias at different developmental stages and to quantify the intra-individual variability of the results of the tests. This approach is thus not sample-size dependent.

### ***2.2. Methods***

#### ***2.2.1. $\mu$ CT-scan acquisitions***

Based on the recommendations that trabeculae should be at least three to four voxels thick (Bouxsein et al., 2010; Guldborg et al., 1998) and mean trabecular thickness of the humeral metaphysis ranges from 75 to 112  $\mu$ m (Barvencik et al., 2010),  $\mu$ CT acquisitions must be done with a resolution of 25  $\mu$ m or higher. Thus,  $\mu$ CT acquisitions were performed with the best resolution possible, depending on the size of the bone (detailed in Table 1) with a BIR ACTIS 225/300 micro-CT scanner (Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; acquisition protocol: voltage of 130kV, current of 60 mA, 2500 projections, frame averaging of 3, 200 ms time exposure, rotation of 360°, 0.25 mm brass filter).

### 2.2.2. VOI selection

The protocol below was specifically designed for a study focusing on ontogenetic modification of humerus trabecular structure for which small VOIs were scaled and defined according to humeral length (Colombo, 2014).

The location of the VOI must be homologous inside the proximal humeral metaphyses in order to be compared among individuals. The metaphyseal border (MB) was selected as the bone structure of reference. The MB is the junction line between cortical bone and the growth plate at the bone surface, and it is the only macroscopic structure observable throughout the development of the humeral proximal extremity. Even if the MB disappears after epiphyseal fusion a remnant epiphyseal line is still observable. Starting from this line, for the oldest individuals, it is possible to eliminate what was the epiphysis and to extract from the  $\mu$ CT images a “virtual metaphysis” (Figure 1a).

The TIVMI® software program<sup>1</sup> was used to extract the VOIs. TIVMI® is based on the HMH (Half-Maximum Height) algorithm (Spoor et al., 1993), extended to 3D (Dutailly et al., 2009). It provides volumetric reconstructions which can be more precise than with other algorithms (Guyomarc'h et al., 2012). All the landmarks useful to select VOIs were then created and positioned from the MB. A median plane ( $p$ ) was created from a series of equidistant landmarks positioned on the MB. The first constructed point ( $a2$ ) is the barycentre of the re-distributed projection of MB on  $p$ . The second constructed point ( $a1$ ) is the most superior point of the metaphysis, the furthest point from  $p$ . The third constructed point ( $a3$ ) is the barycentre of the metaphyseal section placed at 10% of humeral length in the distal direction, below  $p$ . These three points produced two axes:  $a2a1$  and  $a2a3$ . These axes permitted the positioning of the centre of the four spherical VOIs. The midpoint of  $a2a1$  is the centre of VOI1. The centre of VOI2, VOI3 and VOI4 were positioned at the quarter, the mid-point and three-quarters length of  $a2a3$  (Figure 1b). The VOIs have a diameter equal to 3% of humeral length. Even if the value has been arbitrarily chosen, it seemed to us a good compromise between the data processed and the computer power. Only VOIs were segmented (for further details, see Colombo, 2014).

### 2.2.3. Images segmentation and TBMA parameters

Image segmentation and measurement of the parameters were also performed with the TIVMI® software program. Due to the precise 3D reconstructions done, no measurement differences are observed in relation to threshold variation, a characteristic that underlies reproducibility of this method (Guyomarc'h, 2011).

Five microarchitectural parameters have been measured: the trabecular length (Tb.Le), the trabecular number (Tb.N), the connectivity density (Conn.D), the trabecular thickness (Tb.Th), and the trabecular bone volume fraction (BV/TV). The measurements of these five parameters (for detailed definitions, see Table 2), the only ones available in TIVMI®, were performed using the “skeletonization” algorithm of Palágyi et al. (2001): the *sequential 3D curve-thinning algorithm*, implemented in the software to develop microarchitectural quantification (Colombo, 2014). The “skeletonization” is a process of thinning of a connected tubular network (Blum, 1967). The trabecular network is thinned in each direction of the space and is limited to a central line of one voxel thick: the 3D skeleton (Figure 1b). It keeps the same topological and geometrical

---

<sup>1</sup> TIVMI: Treatment and Increased Vision for Medical Imaging, created and developed by B.D. (<http://projets.pacea.u-bordeaux.fr/TIVMI/>)

characteristics as the original network. Suppression of a constitutive voxel from the 3D skeleton leads to topological changes and rupture of the trabecular network. It exists several skeletonization methods with their own limitations, it is a fruitful topic that won't be discussed here, but for a more complete information on skeletonization, the reader may refer to the review of Saha et al., 2016.

#### 2.2.4. *Bias tests*

To evaluate the biases related to the selection protocol design, ten repetitions of the protocol were performed by two experienced observers (AC, HC) for each 3D reconstructed metaphysis. Error was assessed for the four extracted VOIs by analysing the variability in landmark locations and their TBMA measurements in order to judge the potential effect of landmark variations on TBMA parameters, which is useful to know for the ontogenic study based on these VOIs (Colombo, 2014).

Then, we evaluated the influence of biases linked to VOI size and position, image resolution, and segmentation on TBMA parameters on the VOI2 only. Indeed, beyond the variation related to bone biology (which is not the topic here), the differences that we could observe by analysing VOI1, VOI3, and VOI4 would be related to their positioning which is analysed much more precisely using the protocol below to test the bias linked to the VOI position.

To test the bias related to image resolution, it is possible to artificially degrade an image acquired at a high resolution (resampling method) or to scan the same sample at different resolutions (rescanning method). Previous studies show that these two methods produce different results for the same resolution: results from the resampling method are more similar to those obtained from the highest resolution  $\mu$ CT data compared to results from the rescanning method (Kim et al., 2004; Knowles et al., 2019; Sode et al., 2008). We chose the resampling method for this reason and because we didn't have the opportunity to rescan bone samples. Therefore, the original  $\mu$ CT-scans were virtually degraded from 10 to 100 $\mu$ m. Both the VOI centre and the original axes were unchanged. Segmentation of each degraded  $\mu$ CT image was done in the same way as it was for the original images. The bias linked to image segmentation was tested by changing the threshold from the original threshold value (defined in figures as 0) down to -5,000 and up to +5,000 by increments of 500 grey levels. The bias linked to VOI size was tested using the exact centre of VOI2 and creating several spheres with increasing diameters ranging from 0.5 to 7% using increments of 0.5% of humeral length. Beyond 7% of humeral length, the VOI sphere contains not only trabecular bone but also cortical bone and external background (air). The bias due to VOI position was tested by moving the centre of VOI2 along  $a2a3$  by increments of 0.5% step of humeral length. With each increment,  $\mu$ CT images were segmented using the same threshold.

#### 2.2.5. *Statistics*

VOI selection depends on the placement of  $a1$  and constructions of  $a2$  and  $a3$ . Therefore, these points should always be located in the same place throughout the protocol repetitions. Analysis of inertia of scatter plots composed of the points obtained from the ten repetitions for both intra- and inter-observer testing was done to evaluate both the repeatability and reproducibility of the protocol. Inertia corresponds to the mean of the squared distances between each point and the centroid (the centre of gravity) of the scatter plot. Inertia analysis was complemented by analysis of the coefficients of variation (Saporta, 2011) of the length of  $a2a1$

and  $a2a3$  and the angle between them ( $a123$ ). The coefficient of variation corresponds to the standard deviation mean ratio. Analyses of inertia and coefficients of variation provide insights into the influence of the protocol on the results. Both intra- and inter-observer error tests for the coefficient of variation analysis of each microarchitectural parameter were also calculated. For each parameter (TBMA and those depending on point positioning) and for each protocol iteration, a comparison between two iterations was performed using a Student t-test for paired samples and the normality of distribution was assessed with a Shapiro-Wilk W test.

For each test of bias, a Pearson correlation coefficient assessing the linear relationship was calculated. When there was a non-linear relationship between parameters, a Spearman correlation coefficient was calculated instead. A coefficient close to 0 indicates no link between the parameters; close to 1, a relationship between parameters varying in the same direction; close to -1, a relationship between parameters varying in the opposite direction. The correlation coefficient is considered significantly different from 0 if  $p < 0.05$ . To understand how biases influenced the results, we calculated a ratio between the reference value, which corresponds to the result obtained by the strict application of the defined protocol (*cf.* paragraph 2.2.2.), and calibrated measurement values. If the ratio ranges from 0.95 to 1.05, we considered that the difference between the two values is negligible, not thus exceeding the 5% limit.

### 3. Results

#### 3.1. Intra- and inter-observer error

##### 3.1.1. Landmark placement

For the intra-observer error test,  $a1$  is far from its centroid from 0.072 to 0.52 mm. Lower distances are noted for  $a2$  and  $a3$ .  $a2$  is far from its centroid from 0.037 to 0.196 mm.  $a3$  is far from its own from 0.037 to 0.124 mm (Table 3). Coefficients of variation for angle  $a123$  and segments  $a2a1$  and  $a2a3$  are lower than 0.0117 (Table 3), meaning there is a difference between measurements lower than 1.17%. There are no statistically significant differences for almost all comparisons when iterations are compared for each variable two by two. The mean error between two measurements is equal to  $0.72^\circ$  for  $a123$ , 0.14 mm for  $a2a1$ , and 0.022 mm for  $a2a3$  (Table 3). These results support the low values observed for the coefficients of variation and indicate that from one iteration to another, the protocol was applied similarly.

Regarding the inter-observer test,  $a1$  is far from its centroid from 0.029 to 0.415 mm. Lower distances are highlighted for  $a2$  and  $a3$ .  $a2$  is far from its centroid from 0.132 to 0.272 mm.  $a3$  is far from it from 0.064 to 0.124 mm (Table 3). No significant differences are observed between the two observers. Coefficients of variation for angle and segments are lower than 0.014, meaning a difference between measurements lower than 1.4% (Table 3). Even so, most of the coefficients of variation are higher than those for the intra-observer error test, and almost no statistically significant differences were found for any of the tests performed. The mean error between two measurements is equal to  $0.81^\circ$  for  $a123$ , 0.26 mm for  $a2a1$ , and 0.024 mm for  $a2a3$  (Table 3). Thus, the two observers similarly applied the VOI selection protocol.

##### 3.1.2. Influences on TBMA parameters

The 10 iterations for intra-observer tests performed on each humeral metaphysis provided low coefficients of variation for TBMA parameters, whatever the VOI, mainly showing measurement variation of less than 5%. BV/TV coefficients of variation range from 0.004 to 0.025 (Table 4), indicating measurement variation between 0.4 to 2.5%. For Tb.Le, they range from 0.004 to 0.021 (Table 5). For Tb.Th, they range from 0.002 to 0.010 (Table 6), except for VOI2 of individual e.373, for which the coefficient equals 0.15. This high value is aberrant when compared with the results obtained in any series of measurements. For Tb.N, they range from 0.006 and 0.066 (Table 7); the highest value is obtained on VOI2 for individual e.373. Those of Conn.D range between 0.002 and 0.011 (Table 8).

Regarding repeatability, very few significant differences were found for all parameters. There is no systematic over- or under-estimation between two series of measurements (Tables 4-8). According to the VOI, the mean error between two iterations ranges from 0.07 to 0.17% for BV/TV, from 0.003 to 0.004 mm for Tb.Le, from 0.0005 to 0.0027 mm for Tb.Th, from 1.0 to 20.1 mm<sup>-3</sup> for Tb.N, and from 0.021 to 0.028 for Conn.D (Tables 4-8).

The 10 iterations for inter-observer tests performed by HC on each humeral metaphysis produced low coefficients of variation for TBMA parameters, with all VOIs showing measurement error of less than 5%. Most of the coefficients of the second observer (HC) are slightly higher than the first observer ones (AC). BV/TV coefficients of variation range from 0.004 to 0.054 (Table 4), indicating measurement variation of between 0.4 to 5.4%. For Tb.Le, they range from 0.002 to 0.026 (Table 5), being 0.2 to 2.6% of measurement variation. For Tb.Th, they range from 0.002 to 0.024 (Table 6), being 0.2 to 2.4% of measurement variation. Tb.N ones range from 0.004 to 0.040 (Table 7), being 0.4 to 4% of measurement variation. Those of Conn.D range from 0.001 to 0.012 (Table 8), indicating measurement variation of between 0.1 to 1.2%.

Regarding reproducibility, very few significant differences were found for all parameters. There is no systematic over or under-estimation between two series of measurements (Tables 4-8). According to the VOI, the mean error between two iterations ranges from 0.13 to 0.20% for BV/TV, from 0.003 to 0.005 mm for Tb.Le, from 0.0006 to 0.0016 mm for Tb.Th, from 1.5 to 29.1 mm<sup>-3</sup> for Tb.N, and from 0.021 to 0.029 for Conn.D (Tables 4-8).

## 3.2. Tests of biases

### 3.2.1. Segmentation

BV/TV shows highly significant negative correlations with the segmentation thresholds for all individuals (Figure 2a, Table 9). The proportion of trabecular bone decreases with the increases in grey level values (fewer voxels corresponding to bone are included). From observation (Figure 2a), BV/TV remains similar to the result for the reference value with segmentation varying from 1000 grey levels over or under the reference value.

Tb.Le shows a significant linear correlation with segmentation thresholds only for e.354 (slight and negative, Figure 2b, Table 9). It differs more than 5% from the result for the reference value with an addition of more than 3500 grey levels. For other individuals, Tb.Le does not differ more than 5% whatever the variation in grey levels.

Tb.Th shows highly significant negative linear correlations with segmentation thresholds for all tested individuals (Figure 2c, Table 9). The more the grey level value increases, the thinner the trabeculae are. From observation (Figure 2c), Tb.Th remains less than 5% different from the

result for the reference value with segmentation varying from at least 1500 grey levels over or under the reference value.

Conn.D shows a significant negative linear correlation with segmentation threshold values only for e.352 and e.373 (Figure 2d, Table 9). The increase in grey levels leads to a decrease in the number of trabeculae linked to the same node. However, Conn.D is still similar to the result for the reference value whatever the segmentation threshold used.

Tb.N shows a significant correlation with the segmentation threshold values (Figure 2e, Table 9) for all individuals except e.354. The correlations are positive for e.98 and e.355. Also, e.98 shows an extremely high coefficient of correlation of 0.915 (Table 9) and a much stronger slope than the other individuals (Figure 2e). As grey level values increase, Tb.N also increases. It is the contrary for e.352 and e.373. Regardless of the individual, the Tb.N values differ less than 5% from the result for the reference value, with a segmentation varying from 2500 grey levels over or under the reference value.

### 3.2.2. Resolution

BV/TV shows highly significant positive correlations with the  $\mu$ CT-scan resolution (coefficients are all over 0.879) for all tested individuals (Figure 3a, Table 10). The volumetric ratio increases with decreasing resolution quality. e.98 presents a much stronger slope than the other individuals (Figure 3a).

Tb.Le shows highly significant positive correlations with the resolution (coefficients are all over 0.976) for all tested individuals (Figure 3b, Table 10). The mean length of trabeculae increases with decreasing resolution quality. Measured values differ by more than 5% from the initial degradation value.

Tb.Th shows highly significant positive correlations with the  $\mu$ CT-scan resolution (coefficients are all over 0.974) for all tested individuals (Figure 3c, Table 10). The mean thickness of trabeculae increases with decreasing resolution quality. e.98 presents a much stronger slope than the other individuals (Figure 3c)

Conn.D shows highly significant positive correlations with the  $\mu$ CT-scan resolution for all tested individuals except for e.98 (Figure 3d, Table 10). The mean number of trabeculae linked to the same node increases with decreasing resolution quality.

Tb.N shows highly significant negative correlations with the  $\mu$ CT-scan resolution (coefficients are all under -0.998) for all tested individuals (Figure 3e, Table 10). An exponential decrease in the number of trabeculae is observed with decreasing resolution quality. Measured values differ by more than 5% from the initial degradation value.

### 3.2.3. VOI size<sup>2</sup>

BV/TV shows significant linear correlations with VOI size for all individuals (Figure 4a, Table 11). Measured values increase relatively constantly with VOI diameter.

Tb.Le shows a significant linear correlation with VOI size only for the oldest individuals, e.354 and e.373 (Figure 4b, Table 11). For these individuals, increasing VOI diameter is associated with a decreasing mean trabecular length.

---

<sup>2</sup> Due to its extremely small size, the VOI with a diameter of 0.5% of the humeral length produces random results. Thus, we describe results by excluding those obtained for this VOI.

Tb.Th shows a significant linear correlation with VOI size for all individuals (Figure 4c, Table 11). The correlation is negative for e.98. As VOI diameter increases, the trabeculae are less thick. For the four other individuals, the correlation is positive; thickness increases with VOI diameter.

Conn.D shows a significant linear correlation with VOI size for e.98, e.354 and e.373 (Figure 4d, Table 11). For these three individuals, the bigger the VOI diameter is, the more numerous trabeculae tied to the same node are.

Tb.N shows a positive significant linear correlation with VOI size, for all individuals (Figure 4e, Table 11). The larger the VOI diameter, the greater the number of trabeculae. Independent of the individuals, Tb.N reached 5% of variation rather rapidly.

### 3.2.4. VOI position

BV/TV shows significant strong negative correlations with VOI position (Figure 5a, Table 12) except for e.373 for which the correlation is positive. This means the volume ratio decreases as the VOI gets further to the metaphyseal surface for all individuals, except for e.373. From the reference position, with a shift of 0.5% of humeral length, up or down on  $a2a3$ , BV/TV stays similar to the reference value.

Tb.Le shows significant linear correlations with VOI position (Figure 5b, Table 12) only for e.98 (positive), e.354 (positive) and e.373 (negative). For e.373, the mean length of the trabeculae decreases as the VOI gets further to the metaphyseal surface, on the contrary to the other individuals. However, a shift in the position of -0.5 up to 2% of the humeral length on  $a2a3$ , results in Tb.Le values are similar to the reference value.

Tb.Th shows significant linear correlations with VOI position, except for e.354 (Figure 5c, Table 12). The correlation is positive for individuals e.98 and e.373; the further the VOI is to the top of the metaphysis, the thicker the trabeculae are. The correlations are negative for e.352 and e.355. From the reference position, with a shift of 0.5% of humeral length up or down on  $a2a3$ , Tb.Th stays close to the reference value. For individuals e.352 and e.355 whatever the shift in the position, Tb.Th remains less than 5% different from the reference value.

Conn.D shows strong slope differences and significant linear correlations with VOI position for e.98 (negative) and e.373 (positive) (Figure 5d, Table 12). For e.373, the mean number of trabeculae linked to the same node increases as the VOI gets further to the metaphyseal surface, on the contrary to e.98. However, whatever the shift in the position analysed, Conn.D results are similar to those of the reference value.

Tb.N shows a significantly strong negative linear correlation with VOI position for all individuals, except e.373 for which it is positive (Figure 5e, Table 12). Also, e.98 shows a stronger slope than the other individuals. This means the number of trabeculae decreases as the VOI gets further to the metaphyseal surface (faster for e.98), in contrast to e.373. However, from the reference position, with a shift of 0.5% of humeral length up or down on  $a2a3$ , Tb.N stays close to the reference value, except in the case of individual e.98.

## 4. Discussion

This study investigated the effects of the four main potential biases when quantifying the trabecular bone microarchitecture on a sample representing five different developmental stages. The  $\mu$ CT image resolution and segmentation depends only on the image itself and can generate

independent bias. The size and the position of the volume of interest are dependent on the image and also on the characteristics of the trabecular bone associated with individual developmental stage, environment of development and its constraints (such as body weight, activity and locomotion, nutrition, *etc.*), and topographical anatomical variation. In addition, each bias could also be differentially impacted by the extent of observer experience, as VOI positioning and image segmentation are steps that can be affected by the observer experience level.

#### **4.1. VOI selection protocol**

With the VOI selection protocol that we developed here for the humerus, measurement differences are very low and not significantly different between two repetitions of the protocol on the same individual sample; coefficients of variation are lower than 5% for both intra- and inter-observer measurement differences. This result demonstrates that the VOI selection protocol is applicable as it is, and attests to the good reproducibility and repeatability of the VOI protocol selection. Moreover, the possible slight variations in VOI positioning have negligible effects on TBMA parameters quantification. Even if a more simplified protocol could reduce the intra- and inter-observer error even further, our semi-automated protocol, based on the easily observable metaphyseal border, drastically limits the potential effect of differences between observers on the results.

To compare TBMA of individuals from different species, which present an important variability in body sizes and shapes, the most possible homologous VOI must be selected (Kivell, 2016; Kivell et al., 2011; Lazenby et al., 2011). Through its development, the body size and shape of humans change, as well as the biomechanics (Ruff, 2003; Tardieu, 1999). As TBMA scales allometrically, at least in the humerus and the femur (Ryan and Shaw, 2013), the same patterns have to be considered to compare TBMA of non-adult human remains from different stages of development. Our results show that homologous regions in the humerus metaphysis can be selected thanks to the metaphyseal border identifiable at any stage of development.

#### **4.2. Biases tests**

Each trabecular bone parameter analysed is differentially sensitive to the tested biases (Table 13). We will discuss them one by one.

BV/TV and Tb.Th are highly negatively correlated with the segmentation variations for all individuals (Table 9). Each individual presents a decreasing bone volumetric ratio with the increased threshold value for segmentation partly explained by thinner and thinner trabeculae. In addition, BV/TV is also related to the quantity of trabeculae in the VOI. The perinatal individual (e.98) has more numerous trabeculae, the preadolescent (e.355) has less numerous trabeculae, and the child (e.352) as the mature adolescent (e.373) show numerous trabeculae. Furthermore, BV/TV and Tb.Th are the most sensitive to variation in segmentation as they show results differing by more than 5%, with variation in the segmentation threshold around the reference value as early as 5,000 grey levels (Table 13).

Tb.Le, Conn.D and to a lesser degree Tb.N are less and/or not significantly correlated to segmentation variations (Table 9). Indeed, these three parameters are only slightly or not at all dependent on variation in segmentation since they do not differ from more than 5% considering the conditions of the test or from more than 7,500 grey levels.

When segmentation varies, a certain number of voxels allocated to the bone can be added or removed (Hara et al., 2002). Therefore, trabecular thickness and the volumetric part of

trabecular bone do greatly vary with segmentation. Tb.N, Conn.D and Tb.Le depend only on the 3D skeleton. Consequently, whatever the over- or under-selection of voxels during segmentation, no or fewer variations are observed for connectivity, number and length of trabeculae as the 3D skeleton is the central axis of the trabecular structure.

Several previous studies tested different methods of segmentation to evaluate their effect on the measurement of TBMA parameters. All of them seem to agree that segmentation methods become an issue when the image resolution is too low. Indeed, with high resolution, from one segmentation method to another, results are similar (e.g. Christiansen, 2016; Waarsing et al., 2004). But even though a manual segmentation method is to be avoided because it is time-consuming and observer experience dependant (Scherf and Tilgner, 2009), it is mandatory to control automated or semi-automated segmentation and to correct them manually if necessary. While some might think that manual correction of segmentation could bias the measurements, our results showed that the segmentation can be done with a certain level of freedom without significantly influencing the measurements.

It is known that the higher the resolution, the more detailed the image, and the more reliable and precise the results will be (Müller et al., 1996; Peyrin et al., 1998), making them comparable among individuals. Indeed,  $\mu$ CT-scans of poor quality provide images highlighting structural variations by merging or losing some trabecular material (Hara et al., 2002; Kim et al., 2004; Kothari et al., 1998). The lowest resolution leads to degradation of the 3D skeleton which becomes less precise and does not reveal the smallest structural details. For example, structural differences between pathological and non-pathological individuals cannot be distinguished if the resolution is too low (Isaksson et al., 2011).

Even if the method chosen (image degradation rather than rescanning) to evaluate the bias linked to the resolution can, in a way, minimize the differences measured, each parameter for each individual is correlated to resolution degradations. The trabecular structure seems to be affected in the same way for all of the individuals with a strong positive correlation between the resolution degradation and BV/TV, Tb.Le, Tb.Th and Conn.D, and a strong negative correlation between the resolution degradation and Tb.N. As the resolution becomes lower, the trabecular bone will have an increasing bone volumetric ratio with longer, thicker, more connected and less numerous trabeculae (Table 10). Moreover, Tb.Le and Tb.N are much more sensitive to resolution degradations than BV/TV, Tb.Th and Conn.D (Table 13) as they are directly dependent on the “3D skeleton”. Indeed, Tb.Le and Tb.N differ by more than 5% from the initial degradation of the resolution, meaning that the maximum resolution that should be used is 10  $\mu$ m for perinatal individuals and 20  $\mu$ m for older individuals. On the contrary, BV/TV is the least sensitive parameter to resolution degradation (results do not differ from more than 5%), then Conn.D and Tb.Th. They accept degradations respectively from initial resolution to 20-50  $\mu$ m, 10-45  $\mu$ m and 10-40  $\mu$ m depending on the individuals (Table 13). With the general observation that the older the individuals are, the more degraded the resolution can be. Individuals of different developmental stages are differentially sensitive to degradation. Depending on the parameters, to obtain similar results,  $\mu$ CT scan resolution can support degradation until 20  $\mu$ m for e.98, 25  $\mu$ m for e.352, 35  $\mu$ m for e.355, 50  $\mu$ m for e.354 and 45  $\mu$ m for e.373.

The previous studies testing the effect of image resolution on TBMA parameters were conducted, to our knowledge, on adult individuals. According to our results, if a multi-parameter analysis is carried out, the acquisition resolution must be lower than 10  $\mu$ m for perinatal

individuals and lower than 20  $\mu\text{m}$  for older non-adult individuals in order to reliably compare individuals from different stages of development.

Also, to go further on the effect of resolution and segmentation interaction, if the chosen resolution is not high enough, partial volume effects (blurred delimitation at the interface between two materials) represents an additional bias that can affect the segmentation (Elmoutaouakkil et al., 2002). According to the filter used during  $\mu\text{CT}$  acquisition, partial volume effects differ considerably, even if the use of filters does not impact the 3D surface reconstructions of large structures such as the cranium (Guyomarc'h, 2011).

By testing the VOI size effect, it is possible to learn about bone microstructure organisation, in our case, from the centre to the periphery of the humeral metaphysis. Thus, no significant correlation between VOI size increase and the TBMA parameters would mean that TBMA is homogeneous in the metaphysis.

TBMA parameters are differently affected from one individual to another. BV/TV, Tb.Th and Tb.N are significantly positively correlated to increasing VOI size (except Tb.Th of e.98 which is negatively correlated). Tb.Le is negatively correlated for the two oldest and Conn.D is positively for the youngest and the two oldest individuals (Table 11).

BV/TV and Tb.N are the most sensitive parameters (Table 13). They differ by more than 5% from the reference values (*i.e.* measures for the VOI with a diameter of 3% of the humeral length), from an amplitude of the VOI size change, around the initial size (3% of the humeral length), equal to 1.5% of the humeral length. For instance, for the individual e.373 (Table 13), BV/TV does not differ by more than 5% from BV/TV measured for VOIs with a size ranging from 2.5 to 4 % of humeral length (*i.e.* an amplitude of the VOI size change equals to 1.5%).

Also, by progressing through the stages of development, from birth to adolescence, the number of sensitive parameters increases. These sensitive parameters differ by more than 5% from the reference values with smaller amplitude of the VOI size change around the initial size. Indeed, for e.98, only Tb.N is very sensitive; e.352 has sensitive Tb.N and BV/TV, e.355 has sensitive Tb.N, BV/TV and Tb.Th; e.354 has sensitive Tb.N, BV/TV and Tb.Le; and e.373 has sensitive Tb.N, BV/TV, Tb.Th and Tb.Le (Table 13).

During development, the trabecular bone structure becomes more specialised and more heterogeneous through remodelling. If the VOI is too big, intra-bone variation cannot be observed, whereas an undersized VOI can produce random results. In agreement with Lazenby et al. (2011), we consider that VOI should be scaled to bone size. A VOI diameter from 3 to 4% of humeral length seems to us appropriate for the analysis of humeral TBMA changes during development. Indeed, a small VOI (diameter lower than 2% of the humeral length) would not contain enough trabeculae to well represent the microarchitecture of the region of interest. A larger VOI could include the transition between trabecular and cortical bone tissues. It would prevent highlighting the heterogeneity of the trabecular organisation in the case of a multiple VOIs analysis.

BV/TV, Tb.Th and Tb.N are significantly negatively correlated to the position of the VOI, except for the oldest, mature individual. Tb.Le and Conn.D are only correlated for the youngest and the oldest individuals. Tb.Le is positively correlated for the youngest individual and negatively correlated for the oldest one. It is the contrary for Conn.D. By moving the VOI along the axis of the metaphysis, the affected parameters are different from an individual to another (Table 12).

BV/TV and Tb.N are the most sensitive parameters (Table 13). They differ by more than 5% from the reference values from a range of VOI shift, around the initial position, equals to 0%

of the humeral length. For instance, for the individual e.373 (Table 13), BV/TV does not differ by more than 5% from BV/TV measured for VOIs shifted from -0.5 to +0.5% of humeral length (*i.e.* a range of the VOI shift equals to 1%). Also, by progressing through the stages of development, from birth to adolescence, the combination of sensitive parameters changes. Indeed, e.98 has sensitive BV/TV, Tb.N, Tb.Le, and Tb.Th; and e.352 and e.355 have sensitive BV/TV and Tb.N; e.354 has sensitive BV/TV, Tb.N, and Tb.Le; and e.373 has sensitive BV/TV, Tb.N, and Tb.Th (Table 13).

We observed that all of the measured parameters vary by less than 5% from the reference value by a shift on  $a2a3$  of 0.5% of humeral length up or down. Therefore, to avoid the VOI position effect and to have homologous and comparable VOIs between individuals, we consider that the VOI centre shifting must not exceed, in our case, 0.28 mm for the youngest individual and 1.63 mm for the oldest (because of the scaled VOI), whatever the direction of the shift. We judge variations inferior to these upper limits as negligible. In the case of studying the humerus, our VOI selection protocol based on the metaphyseal border minimizes the error associated with VOI position.

Variation in both size and position of the VOI influence the results for TBMA characterisation (Chappard et al., 2008; Kivell et al., 2011; Lazenby et al., 2011). These studies, done on adult individuals from different bone locations and species led to several observations and conclusions. The human iliac crest biopsies analysis of Chappard et al. (2008) concluded that TBMA parameters must be selected according to the type of study (*i.e.*, longitudinal or not). In agreement with our results, depending on the stage of development analysed, researchers should select the most appropriate parameters for that stage and adapt the  $\mu$ CT image treatment and VOI selection accordingly. Lazenby et al. (2011) on humans and chimpanzee first metacarpals provided methodological recommendations for cross-species comparisons as the VOI have to be scaled. Developing individuals experience changes in size and shape, and their microarchitecture scales allometrically (Ryan and Shaw, 2013). Comparisons between individuals from different developmental stages require scaled VOI too. By studying non-human primate third metacarpals and capitates, Kivell et al. (2011) showed that the trabecular parameters are not equally sensitive to variation in VOI size and position, concluding that researchers should be cautious when inferring intra- and interspecific differences, or comparing results of different studies. We suggest that this advice should also be applied when studying individuals of different developmental stages, as our results suggest that TBMA parameters are not equally sensitive from one stage of development to another.

The concerns about oversampling or error linked to displacements of VOI and size differences also led to the development of methods for analysing the whole anatomical region of interest in its entirety (Gross et al., 2014; Kivell, 2016) or with a multiple VOIs analysis (Sylvester and Terhune, 2017) by mapping TBMA parameters. We agree that one single VOI is not enough to analyse TBMA of a specific bony element, even if it is a large one (*e.g.* Scherf et al., 2013) since it will not reflect the heterogeneity of the bone and its functional specificity. Indeed, we demonstrated that by changing the VOI size, TBMA is not homogeneous from the centre to the periphery of the metaphysis. Also, we suggest that intra-individual variability can be interpreted by comparing a series of VOIs between themselves. In studies interested in developmental processes, multiplication of VOIs permits detection of structural organisation heterogeneity and growth dynamics (Colombo, 2014; Saers et al., 2020).

### ***4.3. Some observations on TBMA variations related to bone development***

Observations made by testing biases related to the size and position of the VOI can be interpreted regarding developmental processes even if this issue was not the main topic of our article. We consider as important to be mentioned, because trabecular bone analysis during development is quite rare (*e.g.* Colombo et al., 2019; Ryan and Krovitz, 2006; Ryan et al., 2017; Saers et al., 2020) in both biomedical (*e.g.*, mainly studied for senescence and fracture healing) and anthropological literature.

First, with the four biases tested, the particularity of perinatal trabecular bone microarchitecture is evidenced. The correlations often show stronger slopes associated with strong coefficients and reactions to bias tests opposite to other individuals. A microarchitecture with thinner and more numerous trabeculae can explain it. These specific characteristics change drastically during the first months of the development (Ryan et al., 2017) and are not specific to modern humans, as also highlighted for Neandertals (Chevalier et al., 2021).

By increasing the VOI size, information about the heterogeneity of the bone is highlighted according to the developmental stage. At all stages, the bone volume fraction increases from the centre to periphery and can be partly explained by an increasing number of trabeculae. Other parameters can change but are different from one developmental stage to another. In addition to the increasing number of trabeculae, from the centre to the periphery of the metaphysis, the perinatal individual has thinner and more connected trabeculae, the child and the preadolescent show thicker trabeculae, the adolescent and mature adolescent show shorter, thicker and more connected trabeculae.

With VOI shifting along the metaphysis axis, interpretations can be made regarding remodelling and the production of bone tissue through developmental stages. Indeed, the trabecular bone furthest from the physis is the first to be produced and consequently the oldest and the most remodelled. Thus, the analysis of TBMA through different location gives an idea of the role of the remodelling process through the different stages of development. Therefore, in getting further from the metaphyseal surface, the four youngest individuals show a decreasing bone volume ratio explained partly by a decreasing number of trabeculae. These four individuals still have productive physis. In addition to the decreasing number of trabeculae, the perinatal individual has longer, thicker and less connected trabeculae, the child and the preadolescent have thicker trabeculae and the adolescent has longer trabeculae. On the contrary, the mature adolescent shows an increasing bone volume fraction explained by smaller, thicker, more connected and numerous trabeculae. It seems that trabecular bone differs from a developmental stage to another in both heterogeneity and remodelling processes which increases with age, supporting the findings of Ryan et al. (2017).

## **5. Conclusions**

With only one individual per stage of development and only one anatomical region studied, it is not possible to assess how the TBMA parameters vary within individuals, among individuals from a same stage of development, and across the stages of development according to the bias tested. But, the design of our study allowed us to address the question: does TBMA characterisation differ according to biases at different stages of development?

To sum up, although we focused on the humerus only, we can provide general recommendations for studying non-adult trabecular bone, particularly concerning the VOI selection protocol and test biases.

The few available studies evaluating biases of  $\mu$ CT-scans analyses were performed on adult individuals (Lazenby et al., 2011; Maga et al., 2006), elderly ones (Isaksson et al., 2011; Kim et al., 2004; Sode et al., 2008), non-human primates (Fajardo and Müller, 2001; Kivell et al., 2011; Lazenby et al., 2011; Maga et al., 2006) or other species (Bouxsein et al., 2010; Sode et al., 2008). To our knowledge, no recommendations were presented for studies specifically dealing with human development, and yet, image analyses of bone development processes require a specific and rigorous protocol to obtain the most accurate results. Growth processes lead to bone size increase, shape changes, and modifications due to bone mineralization (Scheuer and Black, 2000). Also, changes in biomechanical constraints directly influence the developmental characteristics of bone tissue (Tardieu, 1999; Tardieu et al., 2013). It is thus necessary to adjust the acquisition parameters of the  $\mu$ CT scanner (*e.g.* X-ray tube power) to bone mineralization and thickness of the sample, as well as to match up the thickness of the acquisition resolution to the size of the bone and the area of interest. We must be cautious especially in a developmental context because of the diverse array of morphological changes that take place.

In addition, we demonstrated that, depending on the source of bias tested, different TBMA parameters are the most influenced ones. Moreover, considering each type of bias one by one, individuals of different developmental stages and parameters are differentially affected. Indeed, by progressing through the stages of development, the TBMA parameters seem less sensitive to the variation introduced by the bias tests. Perinatal stage, represented by individual e.98 and mature adolescent stage represented by individual e.373 often differ from others depending on the tests and parameters measured. Finally, bone volume fraction, trabecular thickness and trabecular number are the most sensitive and correlated parameters.

When studying the developmental processes of archaeological non-adult trabecular bone, in order to produce more reliable measurements, we suggest basing the VOI selection protocol on a landmark that can be systematically identified at any stage of development, as is the metaphyseal border for the humerus. This protocol must be repeatable and reproducible, based on a homologous region from an individual to another. Then, it should be tested that the VOI is not moving more than 5% of the bone length to obtain negligible variation and comparable VOIs from an individual to another. If there is no specific selection of TBMA parameters when researchers are designing their study, the resolution of the acquisition must be higher than 20  $\mu$ m for non-adult individuals and 10  $\mu$ m for perinatal individuals. Even if the legitimacy of VOI-based studies can be debated, in order to not obscure possible topological intra-osseous TBMA variation, we suggest that multiple VOIs with a diameter ranging from 3 to 4% of bone length should be chosen, as always with the greatest precautions.

## **Acknowledgements**

This research was a part of a doctoral project funded by the Aquitaine Regional Council (France), as part of the LaScArBx ANR-10-LABX-52 research program. Authors are indebted to Jérôme Bouzillard and Prof. Pascal Desbarats (UMR 5800 LaBRI, University of Bordeaux, France) for computer science assistance; to Heiko Temming and Prof. Jean-Jacques Hublin (Max Planck Institute, Leipzig) for  $\mu$ CT acquisitions; and to Frédéric Santos (UMR 5199 PACEA) for his precious help with statistics; to Prof. Jean-Luc Kahn (Normal Anatomy Institute, University of

Strasbourg) for his kind assistance and for granting access to the Strasbourg osteological collection. The authors are grateful to Dr Kate McGrath for English language editing and to both reviewers for their comments which helped to improve the manuscript.

## References

- Barvencik, F., Gebauer, M., Beil, F.T., Vettorazzi, E., Mumme, M., Rupprecht, M., Pogoda, P., Wegscheider, K., Rueger, J.M., Puschel, K., Amling, M., 2010. Age- and sex-related changes of humeral head microarchitecture: Histomorphometric analysis of 60 human specimens. *J. Orthop. Res.* 28, 18-26.
- Blum, H., 1967. A transformation for extracting new descriptors of shape, in: Wathen-Dunn, W. (Ed.), *Models for the perception of speech and visual form*, MIT Press, Cambridge, pp. 362-380.
- Bouxsein, M.L., Boyd, S.K., Christiansen, B.A., Guldborg, R.E., Jepsen, K.J., Müller, R., 2010. Guidelines for assessment of bone microstructure in rodents using micro-computed tomography. *J. Bone Miner. Res.* 25, 1468-1486.
- Burrows, M., Liu, D., McKay, H., 2010. High-resolution peripheral QCT imaging of bone microstructure in adolescents. *Osteoporosis Int.* 21, 515-520.
- Cendre, E., Kaftandjian, V., Peix, G., Jourlin, M., Mitton, D., Babot, D., 2000. An investigation of segmentation methods and texture analysis applied to tomographic images of human vertebral cancellous bone. *J. Microsc.* 197, 305-316.
- Chevalier, T., Colard, T., Colombo, A., Golovanova, L., Doronichev, V., Hublin, J.-J., 2021. Early ontogeny of humeral trabecular bone in Neandertals and recent modern humans. *J. Hum. Evol.* 154, 102968.
- Chappard, C., Marchadier, A., Benhamou, C.-L., 2008. Side-to-side and within-side variability of 3D bone microarchitecture by conventional micro-computed tomography of paired iliac crest biopsies. *Bone* 43, 203-208.
- Christiansen, B.A., 2016. Effect of micro-computed tomography voxel size and segmentation method on trabecular bone microstructure measures in mice. *Bone Rep.* 5, 136-140.
- Cocquerez, J.-P., Philipp-Foliguet, S., 1995. *Analyse d'images : filtrage et segmentation*, Masson, Paris.
- Coleman, M.N., Colbert, M.W., 2007. Technical note: CT thresholding protocols for taking measurements on three-dimensional models. *Am. J. Phys. Anthropol.* 133, 723-725.
- Colombo, A., 2014. *La micro-architecture trabéculaire de l'os en croissance : variabilité tridimensionnelle normale et pathologique analysée par microtomodensitométrie [Thèse de doctorat]*. UMR 5199 PACEA, University of Bordeaux Talence, p. 286.
- Colombo, A., Stephens, N.B., Tsegai, Z.J., Bettuzzi, M., Morigi, M.P., Belcastro, M.G., Hublin, J.J., 2019. Trabecular analysis of the distal radial metaphysis during the acquisition of crawling and bipedal walking: a preliminary study. *Bull. Mém. Soc. Anthropol. Paris* 31, 43-51.
- Coqueugniot, H., Hublin, J.J., 2012. Age-related changes of digital endocranial volume during human ontogeny: Results from an osteological reference collection. *Am. J. Phys. Anthropol.* 147, 312-318.
- Dutailly, B., Coqueugniot, H., Desbarats, P., Gueorguieva, S., Synave, R., 2009. 3D surface reconstructing using HMH algorithm. *Proceedings - International Conference on Image Processing, ICIP art. no. 5413911, 2505-2508.*

Elmoutaouakkil, A., Peyrin, F., Elkafi, J., Laval-Jeantet, A.M., 2002. Segmentation of cancellous bone from high-resolution computed tomography images: influence on trabecular bone measurements. *Medical Imaging, IEEE Transactions on* 21, 354-362.

Fajardo, R.J., Müller, R., 2001. Three-dimensional analysis of non-human primate trabecular architecture using micro-computed tomography. *Am. J. Phys. Anthropol.* 115, 327-336.

Fajardo, R.J., Müller, R., Ketcham, R.A., Colbert, M., 2007. Nonhuman anthropoid primate femoral neck trabecular architecture and its relationship to locomotor mode. *Anat. Rec.* 290, 422-436.

Fajardo, R.J., Ryan, T.M., Kappelman, J., 2002. Assessing the accuracy of high-resolution X-ray computed tomography of primate trabecular bone by comparisons with histological sections. *Am. J. Phys. Anthropol.* 118, 1-10.

Goo, J.M., Tongdee, T., Tongdee, R., Yeo, K., Hildebolt, C.F., Bae, K.T., 2005. Volumetric measurement of synthetic lung nodules with multi-detector row CT: Effect of various image reconstruction parameters and segmentation thresholds on measurement accuracy. *Radiology* 235, 850-856.

Gross, T., Kivell, T.L., Skinner, M.M., Nguyen, N.H., Pahr, D.H., 2014. A CT-image-based framework for the holistic analysis of cortical and trabecular bone morphology. *Palaeontol. Electron.* 17.

Guldberg, R.E., Hollister, S.J., Charras, G.T., 1998. The accuracy of digital image-based finite element models. *J. Biomech. Eng.* 120, 289-295.

Guyomarc'h, P., 2011. Reconstitution faciale par imagerie 3D : variabilité morphométrique et mise en oeuvre informatique, École doctorale Sciences de l'environnement, Université Bordeaux 1, Talence, p. 284.

Guyomarc'h, P., Santos, F., Dutailly, B., Desbarats, P., Bou, C., Coqueugniot, H., 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, 221-227.

Hara, T., Tanck, E., Homminga, J., Huiskes, R., 2002. The influence of microcomputed tomography threshold variations on the assessment of structural and mechanical trabecular bone properties. *Bone* 31, 107-109.

Isaksson, H., Töyräs, J., Hakulinen, M., Aula, A.S., Tamminen, I., Julkunen, P., Kröger, H., Jurvelin, J.S., 2011. Structural parameters of normal and osteoporotic human trabecular bone are affected differently by microCT image resolution. *Osteoporosis Int.* 22, 167-177.

Ketcham, R.A., Ryan, T.M., 2004. Quantification and visualization of anisotropy in trabecular bone. *J. Microsc.* 213, 158-171.

Kim, D.G., Christopherson, G.T., Dong, X.N., Fyhrie, D.P., Yeni, Y.N., 2004. The effect of microcomputed tomography scanning and reconstruction voxel size on the accuracy of stereological measurements in human cancellous bone. *Bone* 35, 1375-1382.

Kivell, T.L., 2016. A review of trabecular bone functional adaptation: What have we learned from trabecular analyses in extant hominoids and what can we apply to fossils? *J. Anat.* 228, 569-594.

Kivell, T.L., Skinner, M.M., Lazenby, R.A., Hublin, J.J., 2011. Methodological considerations for analyzing trabecular architecture: an example from the primate hand. *J. Anat.* 218, 209-225.

Knowles, N.K., Ip, K., Ferreira, L.M., 2019. The Effect of Material Heterogeneity, Element Type, and Down-Sampling on Trabecular Stiffness in Micro Finite Element Models. *Ann. Biomed. Eng.* 47, 615-623.

Kothari, M., Keaveny, T.M., Lin, J.C., Newitt, D.C., Genant, H.K., Majumdar, S., 1998. Impact of spatial resolution on the prediction of trabecular architecture parameters. *Bone* 22, 437-443.

Lazenby, R.A., Angus, S., Cooper, D.M.L., Hallgrímsson, B., 2008. A three-dimensional microcomputed tomographic study of site-specific variation in trabecular microarchitecture in the human second metacarpal. *J. Anat.* 213, 698-705.

Lazenby, R.A., Skinner, M.M., Kivell, T.L., Hublin, J.J., 2011. Scaling VOI size in 3D  $\mu$ CT studies of trabecular bone: a test of the over-sampling hypothesis. *Am. J. Phys. Anthropol.* 144, 196-203.

Liu, D., Burrows, M., Egeli, D., McKay, H., 2010. Site specificity of bone architecture between the distal radius and distal tibia in children and adolescents: An HR-pQCT study. *Calcified Tissue Int.* 87, 314-323.

Maga, M., Kappelman, J., Ryan, T.M., Ketcham, R.A., 2006. Preliminary observations on the calcaneal trabecular microarchitecture of extant large-bodied Hominoids. *Am. J. Phys. Anthropol.* 129, 410-417.

Modlesky, C.M., Bajaj, D., Kirby, J.T., Mulrooney, B.M., Rowe, D.A., Miller, F., 2011. Sex differences in trabecular bone microarchitecture are not detected in pre and early pubertal children using magnetic resonance imaging. *Bone* 49, 1067-1072.

Müller, R., Koller, B., Hildebrand, T., Laib, A., Gianolini, S., Rüeegsegger, P., 1996. Resolution dependency of microstructural properties of cancellous bone based on three-dimensional  $\mu$ -tomography. *Technol. Health Care* 4, 113-119.

Nägele, E., Kuhn, V., Vogt, H., Link, T.M., Müller, R., Lochmüller, E.-M., Eckstein, F., 2004. Technical considerations for microstructural analysis of human trabecular bone specimens excised from various skeletal sites. *Calcified Tissue Int.* 75, 15-22.

Palágyi, K., Balogh, E., Kuba, A., Halmai, C., Erdohelyi, B., Sorantin, E., Hausegger, K., 2001. A Sequential 3D Thinning Algorithm and Its Medical Applications, *Information Processing in Medical Imaging*, Springer Berlin Heidelberg, pp. 409-415.

Peyrin, F., Salome, M., Cloetens, P., Laval-Jeantet, A.M., Ritman, E., Ruegsegger, P., 1998. Micro-CT examinations of trabecular bone samples at different resolutions: 14, 7 and 2 micron level. *Technol. Health Care* 6, 391-401.

Pritchett, J.W., 1991. Growth plate activity in the upper extremity. *Clin. Orthop. Relat. Res.*, 235-242.

Pyka, G., Kerckhofs, G., Schrooten, J., Wevers, M., 2014. The effect of spatial micro-CT image resolution and surface complexity on the morphological 3D analysis of open porous structures. *Mater. Charact.* 87, 104-115.

Rampont, M., 1994. Les squelettes, os et dents de fœtus, nouveaux-nés et enfants du musée anatomique de Strasbourg. Aspects historiques et catalogue., Université Strasbourg 1, Strasbourg, p. 170.

Ruff, C., 2003. Ontogenetic adaptation to bipedalism: Age changes in femoral to humeral length and strength proportions in humans, with a comparison to baboons. *J. Hum. Evol.* 45, 317-349.

Ryan, T.M., Krovitz, G.E., 2006. Trabecular bone ontogeny in the human proximal femur. *J. Hum. Evol.* 51, 591-602.

Ryan, T.M., Raichlen, D.A., Gosman, J.H., 2017. Structural and Mechanical Changes in Trabecular Bone during Early Development in the Human Femur and Humerus, in: Percival, C.J., Richtsmeier, J.T. (Eds.), *Building Bones: Bone Formation and Development in Anthropology*, Cambridge University Press, Cambridge, pp. 281-302.

Ryan, T.M., Shaw, C.N., 2013. Trabecular bone microstructure scales allometrically in the primate humerus and femur. *Proc. R. Soc. B Biol. Sci.* 280.

Saers, J.P.P., Ryan, T.M., Stock, J.T., 2020. Baby steps towards linking calcaneal trabecular bone ontogeny and the development of bipedal human gait. *J. Anat.* 236, 474-492.

Saha, P.K., Borgefors, G., Sanniti di Baja, G., 2016. A survey on skeletonization algorithms and their applications. *Pattern Recognition Letters* 76, 3-12.

Salome-Pateyron, M., Peyrin, F., Borras, G., Cloetens, P., Laval-Jeantet, A.M., 1997. Quantification de l'architecture osseuse par microtomographie 3D utilisant le rayonnement synchrotron Seizième Colloque GRETSI, Grenoble, pp. 55-58.

Saporta, G., 2011. Probabilités, analyse des données et statistique, Editions TECHNIP, Paris.

Scherf, H., Harvati, K., Hublin, J.J., 2013. A comparison of proximal humeral cancellous bone of great apes and humans. *J. Hum. Evol.* 65, 29-38.

Scherf, H., Tilgner, R., 2009. A new high-resolution computed tomography (CT) segmentation method for trabecular bone architectural analysis. *Am. J. Phys. Anthropol.* 140, 39-51.

Scheuer, L., Black, S., 2000. Developmental juvenile osteology, Academic Press, London.

Sode, M., Burghardt, A.J., Nissenson, R.A., Majumdar, S., 2008. Resolution Dependence of the Non-metric Trabecular Structure Indices. *Bone* 42, 728-736.

Spoor, C.F., Zonneveld, F.W., Macho, G.A., 1993. Linear measurements of cortical bone and dental enamel by computed tomography : applications and problems. *Am. J. Phys. Anthropol.* 91, 469-484.

Sylvester, A.D., Terhune, C.E., 2017. Trabecular mapping: Leveraging geometric morphometrics for analyses of trabecular structure. *Am. J. Phys. Anthropol.* 163, 553-569.

Tardieu, C., 1999. Ontogeny and phylogeny of femoro-tibial characters in humans and hominid fossils: Functional influence and genetic determinism. *Am. J. Phys. Anthropol.* 110, 365-377.

Tardieu, C., Bonneau, N., Hecquet, J., Boulay, C., Marty, C., Legaye, J., Duval-Beaupère, G., 2013. How is sagittal balance acquired during bipedal gait acquisition? Comparison of neonatal and adult pelvis in three dimensions. Evolutionary implications. *J. Hum. Evol.* 65, 209-222.

Verdelis, K., Lukashova, L., Atti, E., Mayer-Kuckuk, P., Peterson, M.G.E., Tetradis, S., Boskey, A.L., van der Meulen, M.C.H., 2011. MicroCT morphometry analysis of mouse cancellous bone: Intra- and inter-system reproducibility. *Bone* 49, 580-587.

Waarsing, J.H., Day, J.S., Weinans, H., 2004. An improved segmentation method for in vivo  $\mu$ CT imaging. *J. Bone Miner. Res.* 19, 1640-1650.

Way, T.W., Chan, H.P., Goodsitt, M.M., Sahiner, B., Hadjiiski, L.M., Zhou, C., Chughtai, A., 2008. Effect of CT scanning parameters on volumetric measurements of pulmonary nodules by 3D active contour segmentation: A phantom study. *Phys. Med. Biol.* 53, 1295-1312.

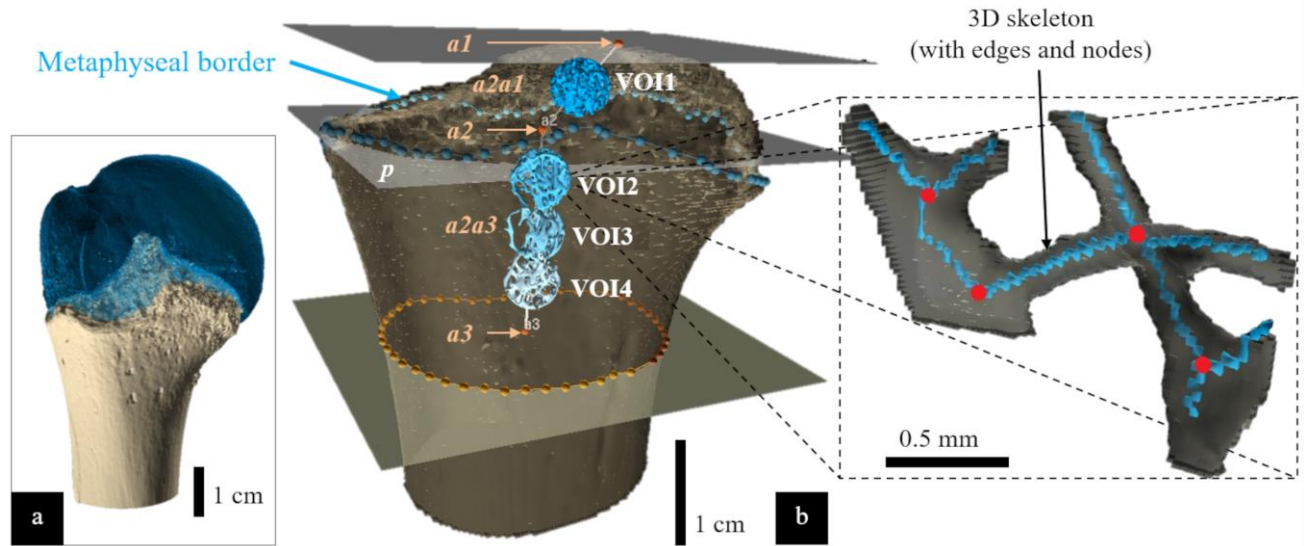


Figure 1 - Protocol for VOI selection: (a) virtual separation of fusing epiphysis (blue part) from the metaphysis, (b) metaphyseal border allowing the creation of the medium plane ( $p$ ) from which  $a1$ ,  $a2$  and  $a3$  are created; VOI1 centre is placed at the middle of the axis  $a2a1$ ; VOI2, VOI3, and VOI4 are located at the quarter, middle and third quarter of the axis  $a2a3$ ; zoom shows skeletonization result as the “3D skeleton” composed with edges and nodes.

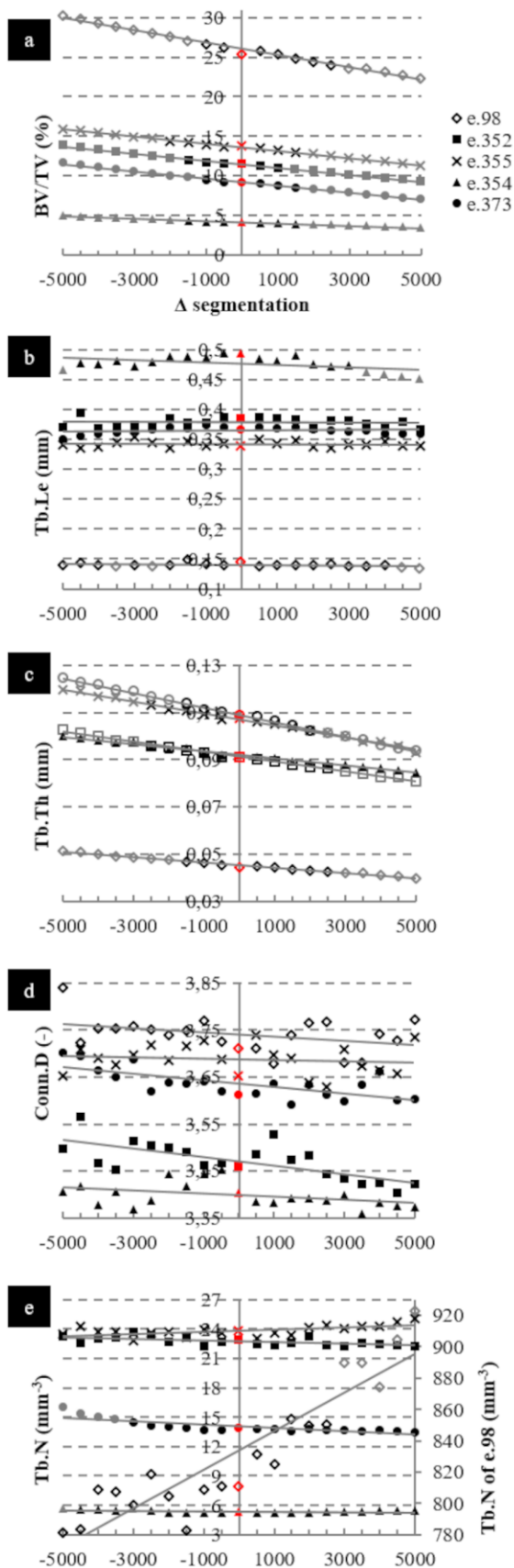


Figure 2 - Effect of image segmentation on: (a) trabecular bone volume fraction (BV/TV), (b) trabecular length (Tb.Le), (c) trabecular thickness (Tb.Th), (d) Connectivity (Conn.D), (e) trabecular number (Tb.N). Red dots represent reference values resulting from the application of protocol for VOI selection, black dots represent values differing from less than 5% of the reference value, grey dots differ from more than 5%.

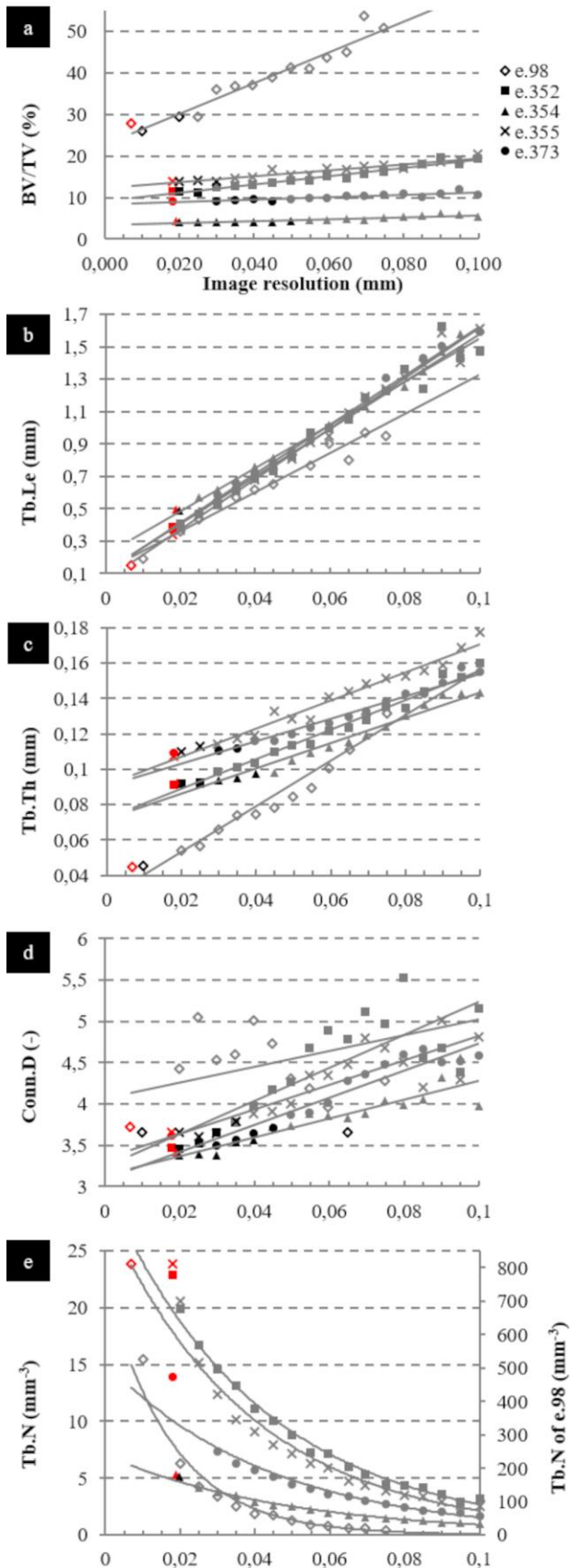


Figure 3 - Effect of image resolution on: (a) trabecular bone volume fraction (BV/TV), (b) trabecular length (Tb.Le), (c) trabecular thickness (Tb.Th), (d) Connectivity (Conn.D), (e) trabecular number (Tb.N). Red dots represent reference values resulting from the application of protocol for VOI selection, black dots represent values differing from less than 5% of the reference value, grey dots differ from more than 5%.

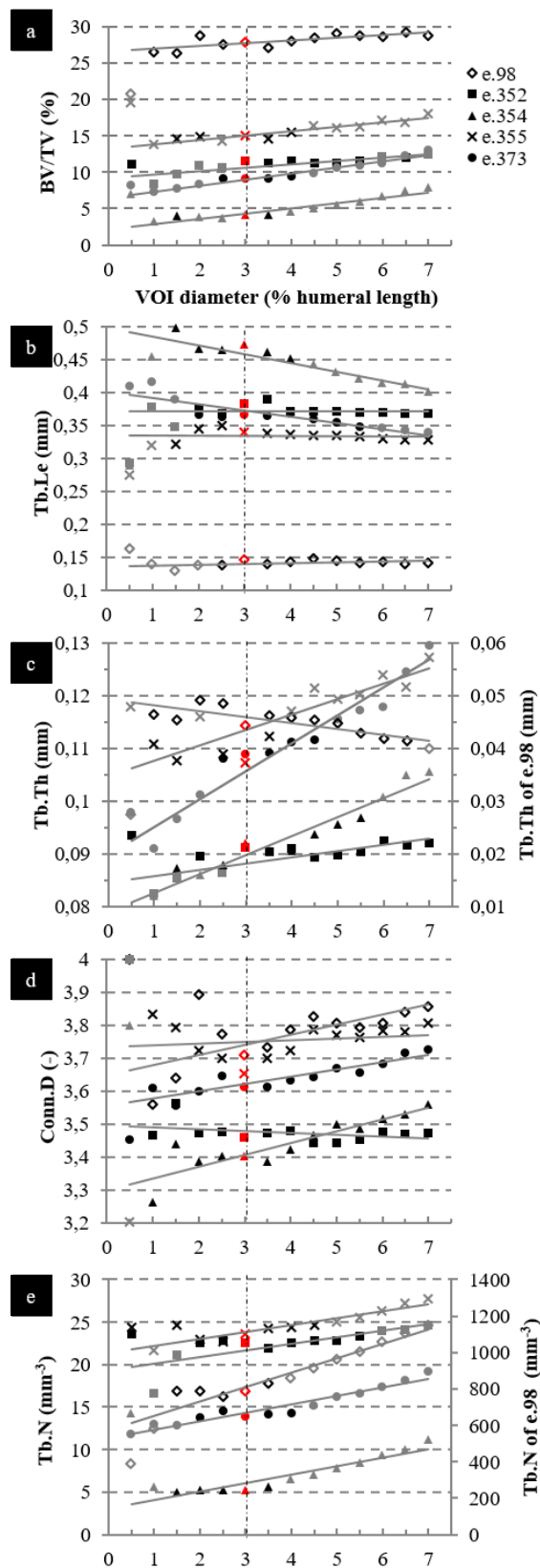


Figure 4 - Effect of VOI size on: (a) trabecular bone volume fraction (BV/TV), (b) trabecular length (Tb.Le), (c) trabecular thickness (Tb.Th), (d) Connectivity (Conn.D), (e) trabecular number (Tb.N). Red dots represent reference values resulting from the application of protocol for VOI selection, black dots represent values differing from less than 5% of the reference value, grey dots differ from more than 5%.

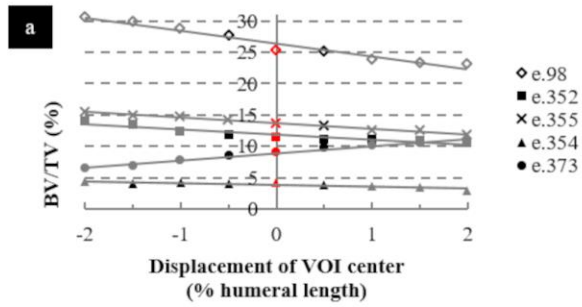
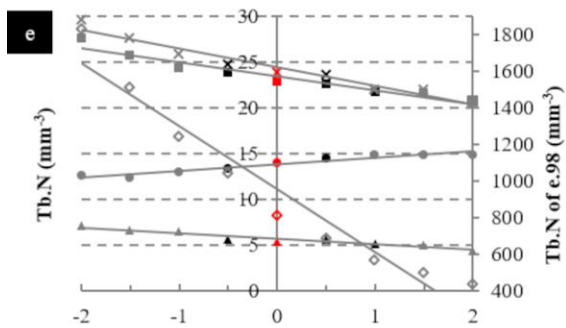
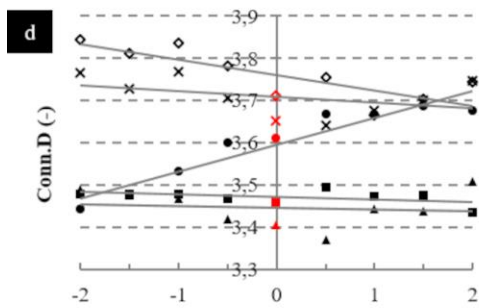
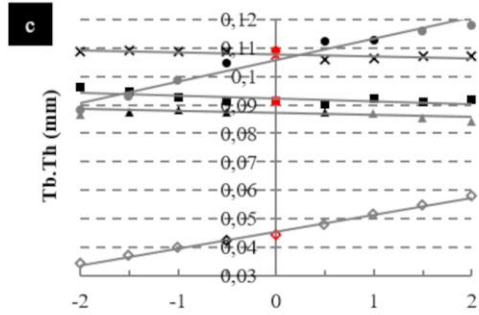
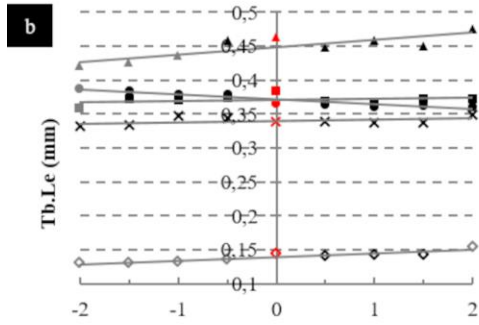


Figure 5 - Effect of VOI position on: (a) trabecular bone volume fraction (BV/TV), (b) trabecular length (Tb.Le), (c) trabecular thickness (Tb.Th), (d) Connectivity (Conn.D), (e) trabecular number (Tb.N). Red dots represent reference values resulting from the application of protocol for VOI selection, black dots represent values differing from less than 5% of the reference value, grey dots differ from more than 5%.



**Table 1 – Description of the sample from the skeletal collections of identified individuals at the Institute of Normal Anatomy, University of Strasbourg (F: female, M: male, I: indeterminate, L: left, R: right).**

Individual (collection number)	Age (years)	Sex	Humerus	Maximal length (cm)	Resolution (mm)
e.98 (1887-88/68-embr.98)	0.00	F	L	5.52	0.007
e.352 (1880/8-embr.352)	5.00	I	L	15.6	0.018
e.355 (1897-98/76-embr.355)	10.00	M	R	18.9	0.018
e.354 (1892-93/284-embr.354)	14.00	F	L	24.7	0.019
e.373 (1897-98/112-embr.373)	18.75	M	L	32.6	0.018

**Table 2 – Descriptions and definitions of the measured parameters.**

Parameter	Abbreviation	Units	Biological definition	Graphic definition (with the “3D skeleton”)	Measurement method
Trabecular length	<b>Tb.Le</b>	mm	Average trabecular length inside the VOI.	Mean length of the edges.	Each edge voxel is linked to others by a face, an edge or a vertex. These relationships between voxels determine the value of the total length of the edge.
Trabecular number	<b>Tb.N</b>	mm <sup>-3</sup>	Number of trabeculae per unit of volume	Number of edges constituting the skeleton.	Simple count of the number of edges.
Connectivity density	<b>Conn.D</b>	-	Average number of trabeculae linked to the same junction. The higher the number, the more the structure is connected.	Mean number of edges per node.	Calculate from the histogram of edge distribution by type of node.
Trabecular thickness	<b>Tb.Th</b>	mm or µm	Average thickness of the trabeculae inside the VOI.	Double average distance between 3D skeleton and border of 3D reconstruction.	For each trabecula, 6 probes are sent from the 3D-skeleton voxels in the 6 directions of the space (-x, +x, -y, +y, -z et +z). Probes with the greatest distances are deleted, because they are not transverse to the trabecula axis. Other probes (e.g. -x, +x et -y, +y) are added up, giving 2 diameters. Diameter mean generates trabecular thickness for each voxel. The mean of trabecular thickness from each voxel gives Tb.Th.
Trabecular bone volume fraction	<b>BV/TV</b>	%	Quantity of trabecular bone inside the VOI.	-	Ratio between the number of white voxels and the total number of voxels inside the VOI.

**Table 3 – Results of the intra-observer and interobserver error estimations on landmarks positioning: repeatability and reproducibility are respectively represented by statistical significance of measurement differences between two iterations of the first observer (Obs 1 vs. Obs 1) and between iterations of the two observers (Obs 1 vs. Obs 2) (n: number of iterations or number of differences measured, m: mean,  $\sigma$ : standard deviation, cv: coefficient of variation,  $p < 0.05$ : number of statistically significant tests).**

	Individual	n	Inertia (mm)			Angle a123 (°)			a1a2 (mm)			a2a3 (mm)		
			a1	a2	a3	m	$\sigma$	cv	m	$\sigma$	cv	m	$\sigma$	cv
Obs 1	e.98	10	0.520	0.037	0.037	149.17	0.84	0.0056	2.37	0.01	0.0050	5.56	0.00	0.0001
	e.352	10	0.039	0.151	0.092	146.62	0.76	0.0052	10.90	0.09	0.0085	15.87	0.04	0.0027
	e.355	10	0.122	0.137	0.073	152.55	0.99	0.0065	11.51	0.13	0.0117	19.01	0.01	0.0006
	e.354	10	0.072	0.093	0.071	153.51	0.46	0.0030	16.31	0.06	0.0036	24.82	0.02	0.0007
	e.373	10	0.362	0.196	0.124	151.40	0.48	0.0032	19.15	0.12	0.0063	32.66	0.02	0.0005
Obs 2	e.98	10	0.415	0.272	0.064	150.00	1.19	0.0080	2.38	0.03	0.0123	5.57	0.01	0.0026
	e.352	10	0.260	0.183	0.124	146.87	1.54	0.0105	10.87	0.15	0.0140	15.87	0.03	0.0019
	e.355	10	0.050	0.136	0.092	153.03	0.72	0.0047	11.53	0.09	0.0074	19.00	0.02	0.0010
	e.354	10	0.029	0.132	0.115	153.49	0.40	0.0026	16.17	0.11	0.0065	24.81	0.02	0.0007
	e.373	10	0.393	0.228	0.109	151.37	1.01	0.0067	19.15	0.18	0.0096	32.66	0.02	0.0006
						m	$\sigma$	$p < 0.05$	m	$\sigma$	$p < 0.05$	m	$\sigma$	$p < 0.05$
Repeatability	<b>Obs 1 vs. Obs 1</b>	225	-	-	-	0.72	0.67	3/45	0.14	0.14	0/45	0.022	0.024	0/45
Reproducibility	<b>Obs 1 vs. Obs 2</b>	500	-	-	-	0.81	0.71	3/100	0.26	0.63	4/100	0.024	0.022	3/100

**Table 4 - Results of the intra-observer and interobserver error estimations on BV/TV (trabecular bone volume fraction) for each Volume of Interest (VOI) : repeatability and reproducibility are respectively represented by statistical significance of measurement differences between two iterations of the first observer (Obs 1 vs. Obs 1) and between iterations of the two observers (Obs 1 vs. Obs 2) (n: number of iteration or number of differences measured, m: mean,  $\sigma$ : standard deviation, cv: coefficient of variation,  $p < 0.05$ : number of statistically significant tests).**

BV/TV (%)	Individual	n	VOI1			VOI2			VOI3			VOI4		
			<i>m</i>	$\sigma$	<i>cv</i>	<i>m</i>	$\sigma$	<i>cv</i>	<i>m</i>	$\sigma$	<i>cv</i>	<i>m</i>	$\sigma$	<i>cv</i>
Obs1	e.98	10	36.70	0.189	0.005	25.96	0.293	0.011	22.82	0.560	0.025	25.95	0.180	0.007
	e.352	10	16.68	0.088	0.005	11.44	0.133	0.012	10.08	0.084	0.008	7.47	0.033	0.004
	e.355	10	27.20	0.161	0.006	4.12	0.022	0.005	2.65	0.036	0.014	0.65	0.016	0.025
	e.354	10	12.44	0.076	0.006	13.54	0.081	0.006	12.12	0.064	0.005	10.35	0.039	0.004
	e.373	10	9.90	0.200	0.020	9.06	0.093	0.010	10.01	0.100	0.010	7.52	0.047	0.006
Obs2	e.98	10	36.74	0.379	0.010	26.05	0.633	0.024	23.20	0.214	0.009	26.08	0.223	0.009
	e.352	10	16.65	0.128	0.008	11.45	0.117	0.010	10.08	0.104	0.010	7.49	0.034	0.004
	e.355	10	12.64	0.130	0.010	4.07	0.023	0.006	2.71	0.045	0.017	0.69	0.037	0.054
	e.354	10	27.26	0.142	0.005	13.56	0.055	0.004	12.11	0.092	0.008	10.52	0.300	0.028
	e.373	10	9.77	0.133	0.014	9.05	0.092	0.010	10.01	0.092	0.009	7.55	0.080	0.011
			<i>m</i>	$\sigma$	$p < 0.05$	<i>m</i>	$\sigma$	$p < 0.05$	<i>m</i>	$\sigma$	$p < 0.05$	<i>m</i>	$\sigma$	$p < 0.05$
Repeatability	<b>Obs 1 vs. Obs 1</b>	225	0.168	0.134	0/45	0.141	0.167	3/45	0.176	0.322	1/45	0.071	0.100	2/45
Reproducibility	<b>Obs 1 vs. Obs 2</b>	500	0.204	0.169	4/100	0.173	0.266	3/100	0.160	0.280	3/100	0.126	0.164	3/100

**Table 5 - Results of the intra-observer and interobserver error estimations on Tb.Le (trabecular length) for each Volume of Interest (VOI) : repeatability and reproducibility are respectively represented by statistical significance of measurement differences between two iterations of the first observer (Obs 1 vs. Obs 1) and between iterations of the two observers (Obs 1 vs. Obs 2) (n: number of iteration or number of differences measured, m: mean,  $\sigma$ : standard deviation, cv: coefficient of variation,  $p < 0.05$ : number of statistically significant tests).**

Tb.Le (mm)	Individual	n	VOI1			VOI2			VOI3			VOI4		
			<i>m</i>	$\sigma$	<i>cv</i>	<i>m</i>	$\sigma$	<i>cv</i>	<i>m</i>	$\sigma$	<i>cv</i>	<i>m</i>	$\sigma$	<i>cv</i>
Obs1	e.98	10	0.133	0.003	0.0190	0.140	0.002	0.015	0.139	0.003	0.021	0.138	0.002	0.013
	e.352	10	0.344	0.005	0.0140	0.373	0.007	0.019	0.377	0.004	0.010	0.392	0.006	0.014
	e.354	10	0.378	0.002	0.0046	0.483	0.004	0.008	0.462	0.004	0.009	0.405	0.005	0.012
	e.355	10	0.338	0.002	0.0046	0.343	0.003	0.009	0.344	0.005	0.014	0.364	0.003	0.009
	e.373	10	0.369	0.002	0.0063	0.364	0.004	0.010	0.361	0.002	0.005	0.365	0.002	0.004
Obs2	e.98	10	0.135	0.002	0.012	0.139	0.003	0.020	0.141	0.003	0.018	0.140	0.004	0.026
	e.352	10	0.344	0.003	0.010	0.376	0.005	0.012	0.381	0.006	0.015	0.390	0.007	0.018
	e.354	10	0.378	0.001	0.004	0.483	0.006	0.012	0.460	0.002	0.005	0.408	0.008	0.020
	e.355	10	0.337	0.002	0.007	0.343	0.002	0.007	0.343	0.004	0.011	0.364	0.005	0.014
	e.373	10	0.370	0.002	0.006	0.366	0.002	0.005	0.360	0.001	0.004	0.366	0.002	0.006
			<i>m</i>	$\sigma$	<i>p &lt; 0.05</i>	<i>m</i>	$\sigma$	<i>p &lt; 0.05</i>	<i>m</i>	$\sigma$	<i>p &lt; 0.05</i>	<i>m</i>	$\sigma$	<i>p &lt; 0.05</i>
Repeatability	<b>Obs 1 vs. Obs 1</b>	225	0.003	0.003	2/45	0.004	0.004	3/45	0.004	0.003	2/45	0.004	0.004	2/45
Reproducibility	<b>Obs 1 vs. Obs 2</b>	500	0.003	0.002	1/100	0.004	0.004	6/100	0.004	0.004	4/100	0.005	0.005	3/100

**Table 6 - Results of the intra-observer and interobserver error estimations on Tb.Th (trabecular thickness) for each Volume of Interest (VOI) : repeatability and reproducibility are respectively represented by statistical significance of measurement differences between two iterations of the first observer (Obs 1 vs. Obs 1) and between iterations of the two observers (Obs 1 vs. Obs 2) (n: number of iteration or number of differences measured, m: mean,  $\sigma$ : standard deviation, cv: coefficient of variation,  $p < 0.05$ : number of statistically significant tests).**

Tb.Th (mm)	Individual	n	VOI1			VOI2			VOI3			VOI4		
			<i>m</i>	$\sigma$	<i>cv</i>	<i>m</i>	$\sigma$	<i>cv</i>	<i>m</i>	$\sigma$	<i>cv</i>	<i>m</i>	$\sigma$	<i>cv</i>
Obs1	e.98	10	0.0298	0.0002	0.0055	0.0454	0.0003	0.007	0.0598	0.0003	0.004	0.0706	0.0007	0.010
	e.352	10	0.0898	0.0005	0.0056	0.0916	0.0007	0.008	0.0920	0.0003	0.004	0.0772	0.0004	0.005
	e.354	10	0.1002	0.0004	0.0040	0.0914	0.0002	0.002	0.0808	0.0004	0.005	0.0664	0.0003	0.005
	e.355	10	0.1251	0.0004	0.0029	0.1072	0.0004	0.004	0.1097	0.0004	0.004	0.1040	0.0002	0.002
	e.373	10	0.1026	0.0007	0.0065	0.1136	0.0170	0.150	0.1135	0.0007	0.006	0.1165	0.0005	0.004
Obs2	e.98	10	0.0299	0.0003	0.010	0.0453	0.0005	0.010	0.0600	0.0005	0.008	0.0709	0.0007	0.009
	e.352	10	0.0896	0.0007	0.008	0.0914	0.0004	0.005	0.0916	0.0005	0.005	0.0772	0.0005	0.006
	e.354	10	0.1006	0.0004	0.004	0.0914	0.0002	0.002	0.0816	0.0006	0.007	0.0659	0.0004	0.007
	e.355	10	0.1254	0.0004	0.003	0.1073	0.0003	0.003	0.1100	0.0005	0.004	0.1056	0.0026	0.024
	e.373	10	0.1020	0.0007	0.006	0.1086	0.0008	0.007	0.1136	0.0007	0.006	0.1166	0.0006	0.005
			<i>m</i>	$\sigma$	<i>p</i> < 0.05	<i>m</i>	$\sigma$	<i>p</i> < 0.05	<i>m</i>	$\sigma$	<i>p</i> < 0.05	<i>m</i>	$\sigma$	<i>p</i> < 0.05
Repeatability	<b>Obs 1 vs. Obs 1</b>	225	0.0005	0.0004	5/45	0.0027	0.0104	1/45	0.0005	0.0004	2/45	0.0005	0.0004	4/45
Reproducibility	<b>Obs 1 vs. Obs 2</b>	500	0.0006	0.0005	7/100	0.0016	0.0074	0/100	0.0006	0.0005	6/100	0.0008	0.0012	7/100

**Table 7 - Results of the intra-observer and interobserver error estimations on Tb.N (trabecular number) for each Volume of Interest (VOI) : repeatability and reproducibility are respectively represented by statistical significance of measurement differences between two iterations of the first observer (Obs 1 vs. Obs 1) and between iterations of the two observers (Obs 1 vs. Obs 2) (n: number of iteration or number of differences measured, m: mean,  $\sigma$ : standard deviation, cv: coefficient of variation,  $p < 0.05$ : number of statistically significant tests).**

Tb.N (mm <sup>-3</sup> )	Individual	n	VOI1			VOI2			VOI3			VOI4		
			<i>m</i>	$\sigma$	<i>cv</i>	<i>m</i>	$\sigma$	<i>cv</i>	<i>m</i>	$\sigma$	<i>cv</i>	<i>m</i>	$\sigma$	<i>cv</i>
Obs1	e.98	10	3077.8	83.8	0.0272	817.5	20.1	0.025	431.7	3.6	0.008	363.6	5.1	0.014
	e.352	10	40.4	0.5	0.0113	22.7	0.3	0.013	19.3	0.2	0.013	17.4	0.2	0.014
	e.354	10	20.8	0.2	0.0083	5.3	0.0	0.009	4.2	0.1	0.017	1.9	0.0	0.018
	e.355	10	48.7	0.4	0.0072	23.5	0.1	0.006	20.1	0.2	0.011	16.7	0.2	0.013
	e.373	10	21.6	0.8	0.0351	14.2	0.9	0.066	14.2	0.1	0.006	10.3	0.1	0.010
Obs2	e.98	10	3033.7	71.5	0.024	818.9	32.7	0.040	424.8	7.3	0.017	364.6	7.2	0.020
	e.352	10	40.5	0.5	0.012	22.7	0.3	0.013	19.1	0.2	0.010	17.7	0.3	0.017
	e.354	10	21.2	0.3	0.012	5.3	0.1	0.011	4.2	0.1	0.013	1.9	0.0	0.025
	e.355	10	49.0	0.4	0.008	23.4	0.2	0.011	20.1	0.2	0.008	16.6	0.3	0.019
	e.373	10	21.1	0.5	0.022	13.9	0.1	0.007	14.2	0.1	0.005	10.3	0.1	0.012
			<i>m</i>	$\sigma$	$p < 0.05$	<i>m</i>	$\sigma$	$p < 0.05$	<i>m</i>	$\sigma$	$p < 0.05$	<i>m</i>	$\sigma$	$p < 0.05$
Repeatability	<b>Obs 1 vs. Obs 1</b>	225	20.1	49.2	1/45	4.9	11.8	2/45	1.0	2.1	1/45	1.3	3.0	3/45
Reproducibility	<b>Obs 1 vs. Obs 2</b>	500	19.1	47.3	3/100	5.9	15.1	7/100	1.8	4.3	3/100	1.5	3.4	13/100

**Table 8 - Results of the intra-observer and interobserver error estimations on Conn.D (connectivity) for each Volume of Interest (VOI) : repeatability and reproducibility are respectively represented by statistical significance of measurement differences between two iterations of the first observer (Obs 1 vs. Obs 1) and between iterations of the two observers (Obs 1 vs. Obs 2) (n: number of iteration or number of differences measured, m: mean,  $\sigma$ : standard deviation, cv: coefficient of variation,  $p < 0.05$  : number of statistically significant tests).**

Conn.D (-)	Individual	n	VOI1			VOI2			VOI3			VOI4		
			m	$\sigma$	cv	m	$\sigma$	cv	m	$\sigma$	cv	m	$\sigma$	cv
Obs1	e.98	10	4.127	0.036	0.009	3.744	0.027	0.007	3.745	0.031	0.008	3.779	0.040	0.011
	e.352	10	3.554	0.007	0.002	3.472	0.028	0.008	3.467	0.020	0.006	3.369	0.020	0.006
	e.354	10	3.535	0.016	0.005	3.400	0.012	0.004	3.465	0.019	0.005	3.410	0.021	0.006
	e.355	10	3.689	0.012	0.003	3.679	0.020	0.006	3.768	0.031	0.008	3.630	0.029	0.008
	e.373	10	3.550	0.017	0.005	3.607	0.014	0.004	3.674	0.014	0.004	3.741	0.015	0.004
Obs2	e.98	10	4.129	0.032	0.008	3.733	0.032	0.009	3.747	0.036	0.010	3.792	0.045	0.012
	e.352	10	3.542	0.019	0.005	3.483	0.026	0.008	3.468	0.025	0.007	3.365	0.025	0.007
	e.354	10	3.540	0.021	0.006	3.389	0.015	0.004	3.472	0.024	0.007	3.390	0.026	0.008
	e.355	10	3.687	0.018	0.005	3.695	0.019	0.005	3.749	0.027	0.007	3.621	0.024	0.007
	e.373	10	3.547	0.005	0.001	3.601	0.018	0.005	3.677	0.013	0.003	3.746	0.011	0.003
			m	$\sigma$	$p < 0.05$	m	$\sigma$	$p < 0.05$	m	$\sigma$	$p < 0.05$	m	$\sigma$	$p < 0.05$
Repeatability	<b>Obs 1 vs. Obs 1</b>	225	0.021	0.020	2/45	0.023	0.019	3/45	0.026	0.022	4/45	0.028	0.024	4/45
Reproducibility	<b>Obs 1 vs. Obs 2</b>	500	0.0214	0.018	7/100	0.0243	0.021	7/100	0.0271	0.022	11/100	0.0293	0.025	3/100

**Table 9 - Pearson Correlation Coefficients (PCC) and associated *p*-values highlight the linear link and its significance between bias due to image segmentation and the measured microarchitectural parameters (BV/TV: trabecular bone volumetric fraction, Tb.Le: trabecular length, Tb.Th: trabecular thickness, Conn.D: connectivity density, Tb.N: trabecular number, *p* in bold shows statistically significant correlation).**

Individual	BV/TV		Tb.Le		Tb.Th		Conn.D		Tb.N	
	PCC	<i>p</i>	PCC	<i>p</i>	PCC	<i>p</i>	PCC	<i>p</i>	PCC	<i>p</i>
e.98	-0.996	< <b>0.001</b>	-0.357	0.112	-0.996	< <b>0.001</b>	-0.377	0.092	0.915	< <b>0.001</b>
e.352	-0.998	< <b>0.001</b>	-0.05	0.829	-0.998	< <b>0.001</b>	-0.711	< <b>0.001</b>	-0.639	<b>0.002</b>
e.355	-0.999	< <b>0.001</b>	-0.097	0.677	-0.997	< <b>0.001</b>	-0.131	0.57	0.616	<b>0.003</b>
e.354	-0.998	< <b>0.001</b>	-0.465	<b>0.034</b>	-0.996	< <b>0.001</b>	-0.399	0.073	-0.384	0.086
e.373	-0.998	< <b>0.001</b>	0.142	0.54	-0.999	< <b>0.001</b>	-0.68	<b>0.001</b>	-0.817	< <b>0.001</b>

**Table 100 - Pearson Correlation (PCC) or Spearman Correlation Coefficients (SCC) and associated *p*-values highlight the correlation and its significance between bias due to image resolution and the measured microarchitectural parameters (BV/TV: trabecular bone volumetric fraction, Tb.Le: trabecular length, Tb.Th: trabecular thickness, Conn.D: connectivity density, Tb.N: trabecular number, *p* in bold shows statistically significant correlation).**

Individual	BV/TV		Tb.Le		Tb.Th		Conn.D		Tb.N	
	PCC	<i>p</i>	PCC	<i>p</i>	PCC	<i>p</i>	PCC	<i>p</i>	SCC	<i>p</i>
e.98	0.97	< <b>0.001</b>	0.976	< <b>0.001</b>	0.974	< <b>0.001</b>	0.28	0.331	-1	< <b>0.001</b>
e.352	0.979	< <b>0.001</b>	0.984	< <b>0.001</b>	0.992	< <b>0.001</b>	0.83	< <b>0.001</b>	-0.998	< <b>0.001</b>
e.355	0.964	< <b>0.001</b>	0.993	< <b>0.001</b>	0.987	< <b>0.001</b>	0.867	< <b>0.001</b>	-1	< <b>0.001</b>
e.354	0.902	< <b>0.001</b>	0.995	< <b>0.001</b>	0.982	< <b>0.001</b>	0.921	< <b>0.001</b>	-1	< <b>0.001</b>
e.373	0.879	< <b>0.001</b>	0.996	< <b>0.001</b>	0.982	< <b>0.001</b>	0.948	< <b>0.001</b>	-1	< <b>0.001</b>

**Table 11 - Pearson Correlation Coefficients (PCC) and associated *p*-values highlight the linear link and its significance between bias due to VOI (Volume of Interest) size and the measured microarchitectural parameters (BV/TV: trabecular bone volumetric fraction, Tb.Le: trabecular length, Tb.Th: trabecular thickness, Conn.D: connectivity density, Tb.N: trabecular number, *p* in bold shows statistically significant correlation, PCC in *italics* takes into account each value, PCC in plain font excludes values obtained from a VOI with a diameter of 0.05 % of the humeral length).**

Individual	BV/TV		Tb.Le		Tb.Th		Conn.D		Tb.N	
	PCC	<i>p</i>	PCC	<i>p</i>	PCC	<i>p</i>	PCC	<i>p</i>	PCC	<i>p</i>
e.98	<i>0.695</i>	<b>0.006</b>	<i>-0.091</i>	<i>0.756</i>	<i>0.035</i>	<i>0.906</i>	<i>0.224</i>	<i>0.442</i>	<i>0.949</i>	< <b>0.001</b>
	0.768	<b>0.002</b>	0.471	0.104	-0.832	< <b>0.001</b>	0.653	<b>0.016</b>	0.963	< <b>0.001</b>
e.352	<i>0.775</i>	<b>0.001</b>	<i>0.401</i>	<i>0.156</i>	<i>0.505</i>	<i>0.066</i>	<i>-0.508</i>	<i>0.064</i>	<i>0.612</i>	<b>0.02</b>
	0.862	< <b>0.001</b>	-0.026	0.933	0.797	<b>0.001</b>	-0.391	0.187	0.778	< <b>0.001</b>
e.355	<i>0.650</i>	<b>0.012</b>	<i>0.371</i>	<i>0.191</i>	<i>0.754</i>	<b>0.002</b>	<i>0.484</i>	<i>0.08</i>	<i>0.835</i>	< <b>0.001</b>
	0.956	< <b>0.001</b>	-0.066	0.83	0.873	< <b>0.001</b>	0.211	0.489	0.913	< <b>0.001</b>
e.354	<i>0.341</i>	<i>0.232</i>	<i>-0.069</i>	<i>0.814</i>	<i>0.862</i>	< <b>0.001</b>	<i>0.143</i>	<i>0.627</i>	<i>0.293</i>	< <b>0.001</b>
	0.937	< <b>0.001</b>	-0.901	< <b>0.001</b>	0.969	< <b>0.001</b>	0.888	< <b>0.001</b>	0.938	< <b>0.001</b>
e.373	<i>0.963</i>	< <b>0.001</b>	<i>-0.904</i>	< <b>0.001</b>	<i>0.964</i>	< <b>0.001</b>	<i>0.204</i>	<i>0.484</i>	<i>0.966</i>	< <b>0.001</b>
	0.979	< <b>0.001</b>	-0.878	< <b>0.001</b>	0.968	< <b>0.001</b>	0.905	< <b>0.001</b>	0.958	< <b>0.001</b>

**Table 112 - Pearson Correlation Coefficients (PCC) and associated *p*-values highlight the linear link and its significance between bias due to VOI (Volume of Interest) position and the measured microarchitectural parameters (BV/TV: trabecular bone volumetric fraction, Tb.Le: trabecular length, Tb.Th: trabecular thickness, Conn.D: connectivity density, Tb.N: trabecular number, *p* in bold shows statistically significant correlation).**

Individual	BV/TV		Tb.Le		Tb.Th		Conn.D		Tb.N	
	PCC	<i>p</i>	PCC	<i>p</i>	PCC	<i>p</i>	PCC	<i>p</i>	PCC	<i>p</i>
e.98	-0.980	< <b>0.001</b>	0.935	< <b>0.001</b>	0.998	< <b>0.001</b>	-0.807	<b>0.009</b>	-0.97	< <b>0.001</b>
e.352	-0.960	< <b>0.001</b>	0.324	0.395	-0.709	<b>0.033</b>	-0.458	0.215	-0.969	< <b>0.001</b>
e.355	-0.994	< <b>0.001</b>	0.441	0.234	-0.813	<b>0.008</b>	-0.397	0.29	-0.976	< <b>0.001</b>
e.354	-0.917	<b>0.001</b>	0.838	<b>0.005</b>	-0.426	0.254	-0.143	0.714	-0.969	< <b>0.001</b>
e.373	0.986	< <b>0.001</b>	-0.945	< <b>0.001</b>	0.973	< <b>0.001</b>	0.949	< <b>0.001</b>	0.959	< <b>0.001</b>

**Table 123 – Sensitivity of the trabecular bone microarchitectural (TBMA) parameters to the variation in bias tests. For the biases due to image segmentation, VOI position and size, values in each cell correspond to the range of the variation in bias test around the reference value from which the TBMA parameter measured does not vary by more than 5%. For the bias due to image segmentation, the bold numbers indicate a range under 7,500 grey levels. For the bias due to VOI size, the bold numbers indicate a range under 4% of the humeral length\*. For the bias due to VOI position, the bold numbers indicate a range under 3% of humeral length\*\*. For the bias due to image resolution, values in each cell correspond to the lowest resolution for which the TBMA parameter measured does not differ from more than 5% in comparison to the reference value. The bold numbers indicate the lowest resolution for each individual from which at least one TBMA parameters does not differ by more than 5%; the italic numbers correspond to the highest resolution for each individual from which at least one TBMA parameters differs by more than 5% (Ind.: individual, BV/TV: trabecular bone volumetric fraction, Tb.Le: trabecular length, Tb.Th: trabecular thickness, Conn.D: connectivity density, Tb.N: trabecular number).**

Ind.	BV/TV	Tb.Le	Tb.Th	Conn.D	Tb.N	Most sensitive parameter(s) to bias test
<b>Segmentation (-)</b>						
e.98	<b>3,500</b>	9,000	<b>3,500</b>	10,000	7,500	BV/TV +Tb.Th
e.352	<b>2,500</b>	10,000	<b>5,000</b>	10,000	10,000	BV/TV +Tb.Th
e.355	<b>3,500</b>	10,000	<b>4,500</b>	10,000	10,000	BV/TV +Tb.Th
e.354	<b>3,000</b>	7,500	10,000	10,000	10,000	BV/TV
e.373	<b>2,500</b>	10,000	<b>3,500</b>	10,000	8,000	BV/TV +Tb.Th
<b>Resolution (<math>\mu\text{m}</math>)</b>						
e.98	<b>20</b>	<10	10	10	<10	Tb.Le + Tb.N
e.352	<b>25</b>	<20	<b>25</b>	<b>25</b>	<20	Tb.Le + Tb.N
e.355	30	<20	25	<b>35</b>	<20	Tb.Le + Tb.N
e.354	<b>50</b>	<20	40	40	20	Tb.Le + Tb.N
e.373	<b>45</b>	<20	35	<b>45</b>	<20	Tb.Le + Tb.N
<b>VOI size (% humeral length)</b>						
e.98	6	4.5	5.5	6	<b>2</b>	Tb.N
e.352	<b>3.5</b>	5	5	6	<b>3.5</b>	Tb.N + BV/TV
e.355	<b>2.5</b>	5.5	<b>2.5</b>	6	<b>3</b>	Tb.N + BV/TV +Tb.Th
e.354	<b>2</b>	<b>2.5</b>	4	6	<b>2</b>	Tb.N + BV/TV +Tb.Le
e.373	<b>1.5</b>	<b>3.5</b>	<b>2.5</b>	6	<b>2</b>	Tb.N + BV/TV + Tb.Th + Tb.Le
<b>VOI position (% humeral length)</b>						
e.98	<b>1</b>	<b>1.5</b>	<b>0.5</b>	4	<b>0</b>	BV/TV + Tb.Le +Tb.Th + Tb.N
e.352	<b>1.5</b>	3.5	4	4	<b>1</b>	BV/TV + Tb.N
e.355	<b>0.5</b>	4	4	4	<b>1.5</b>	BV/TV + Tb.N
e.354	<b>2</b>	<b>2.5</b>	<b>2</b>	4	<b>1.5</b>	BV/TV + Tb.Le + Tb.N
e.373	<b>1</b>	3.5	<b>2</b>	4	<b>1</b>	BV/TV + Tb.Th + Tb.N

\*For example, Tb.N for e.373 does not differ by more than 5% to the reference value (*i.e.* VOI diameter equals 3% of humeral length) for VOI diameters ranging from 2% to 4% of the humeral length (see Figure 4e), being a range of 2% of the humeral length.

\*\* For example, BV/TV for e.354 does not differ by more than 5% to the reference value (*i.e.* original VOI position with the VOI selection protocol) with a shifting of the VOI centre ranging from -1.5% to +0.5% of the humeral length (see Figure 5a), being a range of 2% of the humeral length.

## **SUPPLEMENTARY INFORMATION**

**Title: Assessing biases in the analysis of trabecular bone microarchitecture of non-adult individuals**

**Authors' names:** COLOMBO Antony<sup>a,b\*</sup>, DUTAILLY Bruno<sup>c,d</sup>, LEGUAY Romain<sup>e</sup>, DUTOIR Olivier<sup>b,c,f</sup> and COQUEUGNIOT H el ene<sup>b,c,g</sup>

### **Authors' affiliations:**

- a- Universit  di Bologna, dipartimento di Scienze Biologiche, Geologiche e Ambientali – BiGeA, Laboratorio di Antropologia fisica, via Francesco Selmi, 3, 40126 Bologna, Italy.
- b- Ecole Pratique des Hautes Etudes-PSL University, Chair of biological anthropology Paul Broca, 4-14 rue Ferrus, F-75014 Paris, France.
- c- UMR 5199 PACEA, University of Bordeaux, CNRS, MCC ; LabEx Sciences arch ologiques de Bordeaux, n  ANR-10-LABX-52, b t B8, all e Geoffroy Saint Hilaire, CS50023, F-33615 Pessac, France.
- d- Bordeaux Montaigne University, UMS 3657 ARCHEOVISION, F-33607 Pessac, France.
- e- Nurea, 37 cours Georges Cl menceau, F-33000 Bordeaux, France.
- f- The University of Western Ontario, Department of Anthropology, N6A-3K7, London, Canada.
- g- Max Planck Institute for Evolutionary Anthropology, Department of Human Evolution, Deutscher Platz 6, D-04103 Leipzig, Germany.

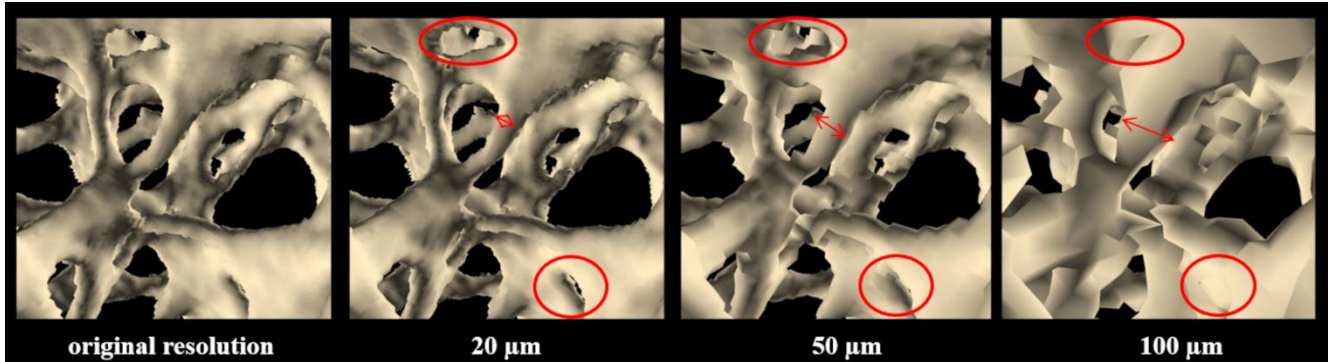
**Keywords:** cancellous bone, volume of interest, 3D, micro-computed tomography

**\* Corresponding author:** Dr Antony Colombo; *address:* Universit  di Bologna, dipartimento di Scienze Biologiche, Geologiche e Ambientali – BiGeA, Laboratorio di Antropologia fisica, via Francesco Selmi, 3, 40126 Bologna, Italy. *Telephone number:* +33.6.08.35.79.44; *e-mail address:* [colomboantony@gmail.com](mailto:colomboantony@gmail.com); [antony.colombo@unibo.it](mailto:antony.colombo@unibo.it)

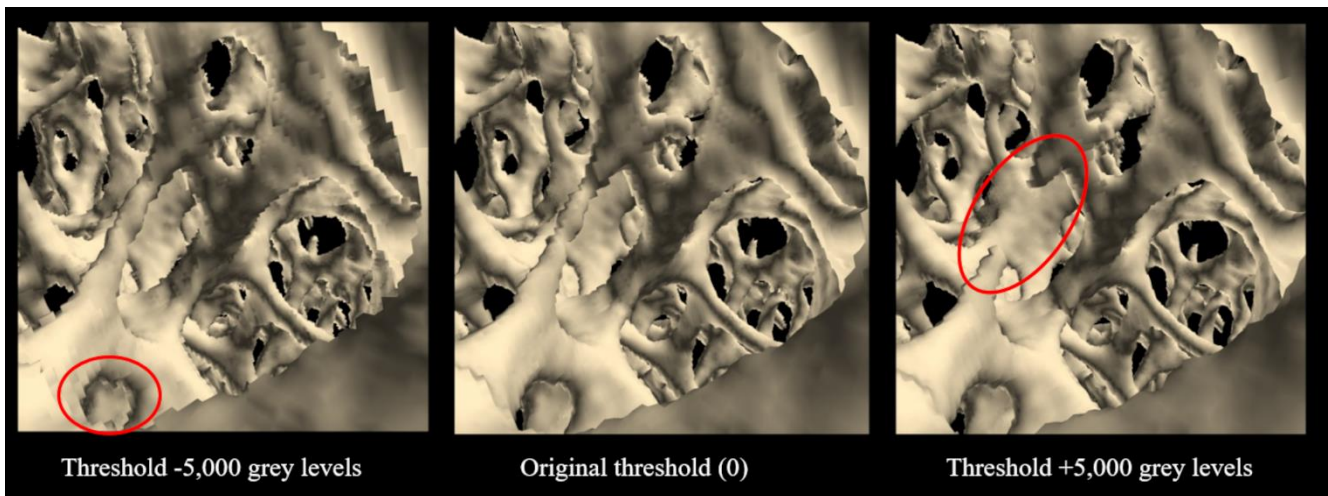
**Abstract:** Micro-computed tomography ( $\mu$ CT) is a non-destructive x-ray technique that can provide high-resolution images of skeletal material, making it an attractive option for analysing archaeological human remains. However,  $\mu$ CT data acquisition, image segmentation and 3D reconstruction may bias the quantification of trabecular bone microarchitecture (TBMA) parameters. Some biases have been evaluated in previous studies, but never in non-adult individuals at different stages of development. Five non-pathological humeral metaphyses from known-age individuals of various developmental stages were  $\mu$ CT-scanned at a high resolution based on bone size. First, repeatability and reproducibility of the volume of interest selection protocol (VOI, trabecular bone region analysed) has been evaluated and found to produce non-significant variations of the TBMA parameters. Then, each bone was assessed for biasing effects related to image resolution and segmentation, as well as the size and position of the VOI. According to the bias tested, individuals of different developmental stages and parameters are differentially affected. By progressing through the stages of development, the TBMA parameters seem less sensitive to the variation introduced by the bias tests. Perinatal stage and mature

adolescent stage distinctly differ from others depending on the tests and parameters measured. Bone volume fraction, trabecular thickness and trabecular number are the most sensitive and correlated parameters to the biases tested. We therefore propose specific recommendations to assure homologous VOI selection and reliable and comparable results for non-adult TBMA analyses.

This supplementary information contains 2 figures to illustrate the effect of resolution and (Figure SI1) and segmentation (Figure SI2).



**Figure SI1:** Effect of resolution degradation from the resolution of acquisition to 20, 50 and 100  $\mu\text{m}$ . Red circles are showing the filling of the originally empty spaces and the red arrows the widening of the trabeculae (individual e.354, resolution of acquisition: 19  $\mu\text{m}$ ).



**Figure SI2:** Effect of segmentation on the trabecular microarchitecture in comparison to the original threshold (at the center). On the left, original segmentation was down to -5,000 grey levels, showing the filling of the original empty space with addition of voxels (red circle). On the right, the original segmentation was up to +5,000 grey levels, showing the disappearance of the central trabecula (red circle) and the modification of the microarchitecture (individual e.373, resolution of acquisition: 18  $\mu\text{m}$ ).