Received: 15 September 2020 | Revised: 29 June 2021 | Accepted: 7 July 2021
DOI: 10.1002/ajpa.24374

American Journal of PHYSICAL ANTHROPOLOGY
WILEY

BRIEF COMMUNICATION

Testing the reliability of CT scan-based dental wear magnitude scoring

Ricardo Miguel Godinho | **Célia Gonçalves**

Interdisciplinary Center for Archaeology and Evolution of Human Behaviour (ICArHEB), Faculdade das Ciências Humanas e Sociais, Universidade do Algarve, Campus Gambelas, Faro, Portugal

Correspondence
Ricardo Miguel Godinho, Interdisciplinary Center for Archaeology and Evolution of Human Behaviour (ICArHEB), University of Algarve, Faculdade das Ciências Humanas e Sociais, Universidade do Algarve, Campus Gambelas, 8005-139, Faro, Portugal.
Email: ricardomiguelgodinho@gmail.com

Funding information
Archaeological Institute of America, Grant/Award Number: The Archaeology of Portugal Fellowship (2016); European Regional Development Fund, Grant/Award Number: ALG-01-0145-FEDER-29680; Fundação para a Ciência e a Tecnologia, Grant/Award Number: DL 57/2016/CP1361/CT0029

Abstract

Objectives: Digital models are now frequently used in biological anthropology (bioanthropology) research. Despite several studies validating this type of research, none has examined if the assessment of dental wear magnitude based on Computerized Tomography (CT) scans is reliable. Thus, this study aims to fill this gap and assess if dental wear magnitude scoring based on CT scans provides results consistent with scoring based on direct observation of the physical specimens.

Materials and Methods: Dental wear magnitude from 412 teeth of 35 mandibles originating from the Portuguese Muge and Sado Mesolithic shell-middens was scored. The mandibles were also CT scanned and visualized using 3D Slicer. CT scan-based scoring of dental wear magnitude was then undertaken. Two scoring rounds were undertaken for each observation method (totaling four scoring rounds) and an intra-observer error test was performed. The averaged results of the two observation methods were compared via boxplots with paired cases.

Results: Intra-observer error was negligible and non-significant. Scoring results are comparable between the two observation methods. Notwithstanding, some differences were found, in which CT scan assessment generally overestimates dental wear when compared to direct observation.

Discussion: Our results generally validate the use of CT scans in studies of dental wear magnitude. Notwithstanding several caveats relating to CT scanning and visualization limitations should be considered to avoid over or under-estimation of dental wear.

KEYWORDS
biological anthropology, CT scans, digitization, teeth, virtual anthropology

1 | INTRODUCTION

Digital models are now routinely used in bioanthropological and palaeoanthropological research to examine, for example, skeletal morphology (Freidline et al., 2012; Freidline et al., 2013; Katz et al., 2017; von Cramon-Taubadel, 2011), dental morphology (Fornai et al., 2015; Fornai et al., 2016; Gómez-Robles et al., 2015; Martinón-Torres et al., 2013; Sarig et al., 2019), skeletal mechanical function (Godinho et al., 2017; Godinho, Fitton, et al., 2018; Godinho & O'Higgins, 2018; Godinho, Spikins, & O'Higgins, 2018), respiratory function (Bastir et al., 2015; Bastir et al., 2020; García-Martínez et al., 2018; Wroe et al., 2018), heat-induced dental morphological changes (Godinho

TABLE 1 Inventory of the teeth used in this study (specimens listed by site and tooth type)

Tooth	Arapouco	Cabeço da Amoreira	Cabeço da Arruda	Cabeço de Pez	Cova da Onça	Moita do Sebastião	Vale de Romeiras	Total
Incisor 1	9	2	8	6	2	18	1	46
Incisor 2	9	2	11	3	2	21	2	50
Canine	8	2	13	3	2	21	4	53
Premolar 1	8	2	16	4	1	25	3	59
Premolar 2	9	2	15	3	2	22	3	56
Molar 1	8	2	14	5	2	27	3	61
Molar 2	6	2	16	2	2	23	3	54
Molar 3	6	2	5	2	2	12	4	33
Total	63	16	98	28	15	169	23	412

et al., 2019; Sandholzer et al., 2013; Sandholzer et al., 2014) and to reconstruct incomplete specimens (Godinho et al., 2020; Godinho & O'Higgins, 2017; Gunz et al., 2009; Gunz et al., 2020; O'Higgins et al., 2019).

Such studies are typically based on photogrammetry, surface scans or CT scans that digitize the relevant specimens and so enable the ensuing computer-based research (Weber, 2015; Weber & Bookstein, 2011). Because it is crucial to ensure that the digital representations of the physical specimens are reliable, several studies have validated the use of digital specimens and have also compared the results of different scanning methods (Aung et al., 1995; Ghoddousi et al., 2007; Katz & Friess, 2014; Slizewski et al., 2010). However, no studies have examined if digital based assessment of dental wear (i.e., the progressive loss of dental tissues due to attrition, abrasion, and erosion) magnitude is reliable yet. This is regrettable because dental wear is critical to examine differences in masticatory mechanics, diet and extra-oral food pre-processing (Bernal et al., 2007; Chattah & Smith, 2006; Clement & Hillson, 2012; Deter, 2009; Kaifu, 1999; Lubell et al., 1994; Molnar, 1971; Molnar et al., 1972; Smith, 1984; Watson, 2008; Watson & Haas, 2017), and also because there is an increasing availability of digitized human skeletal remains (databases such as NESPOS, MorphoSource, MorphoMuseum, The New Mexico Decedent Image Database) that may prompt further dental wear studies. Moreover, the use of online available digitized specimens circumvents the need for handling and direct observation of specimens, and so for traveling (which is currently crucial, as demonstrated by the heavy impact of the Covid-19 pandemic on national and international mobility), provided that digital based dental wear magnitude assessment is reliable. Thus, this study examines if the results of dental wear magnitude scoring based on CT scans and on the direct observation of the physical specimens are consistent or not.

2 | MATERIALS AND METHODS

This study is based on 412 lower teeth originating from seven archaeological sites from the Portuguese Mesolithic Muge and Sado shell-middens (Table 1). Dental wear was first scored for each existing tooth of 35 mandibles according to the scale of Smith (1984; in summary, this scale quantifies the cumulative loss of dental tissues along a scale of eight stages, in which the first stage equals little or no wear and the last stage equals extreme wear with severe or complete loss of the dental crown) via direct observation. The specimens were also CT scanned (Toshiba, 120 kV, voxel size 0.348*0.348*0.3 mm., revolution time 0.75 s, spiral pitch factor 0.94) at the Faculty of Veterinary Medicine of the University of Lisbon. The scans were visualized using the CT Bone filter of the volume rendering module of 3D Slicer (Fedorov et al., 2012), in which the visualization threshold was set manually because the version of the software used did not allow selection of specific threshold values (i.e., there was a slider that did not specify gray level or Hounsfield Unit values). Specifically, the threshold was set when bone started to “wear away” (i.e., started to be excluded from visualization) and so approaches the density of enamel and dentin. Dental wear magnitude was then scored based on CT scan rendering, again according to the scale of Smith (1984).

To assess intra-observer error, dental wear magnitude was scored twice by the same observer (RMG) for each observation method (totaling four rounds of scoring; two based on direct observations and two based on CT scans) with an interval between 9 and 12 months. The results were compared using correlation plots, including the significance of the correlation, of the corrplot R package (Wei et al., 2017). Each correlation plot compared the results of all scoring rounds per tooth. Moreover, descriptive statistics (mean, minimum, maximum and standard deviation) of the differences between direct and CT scan-based scores are also provided, along with the magnitude and direction of individual scoring differences per tooth type.

The results of the two observation methods were then compared for each tooth type via boxplots with paired cases and statistical testing used the Wilcoxon non-parametric test (analyses performed in the ggplot 2 R package; Wickham et al., 2021). Because no meaningful scoring differences were found within observation methods (i.e., scoring 1 and 2 using direct observation, and scoring 1 and 2 using CT scan-based observation), the boxplot comparison between observation methods was based on the average scores of each of the observation methods.

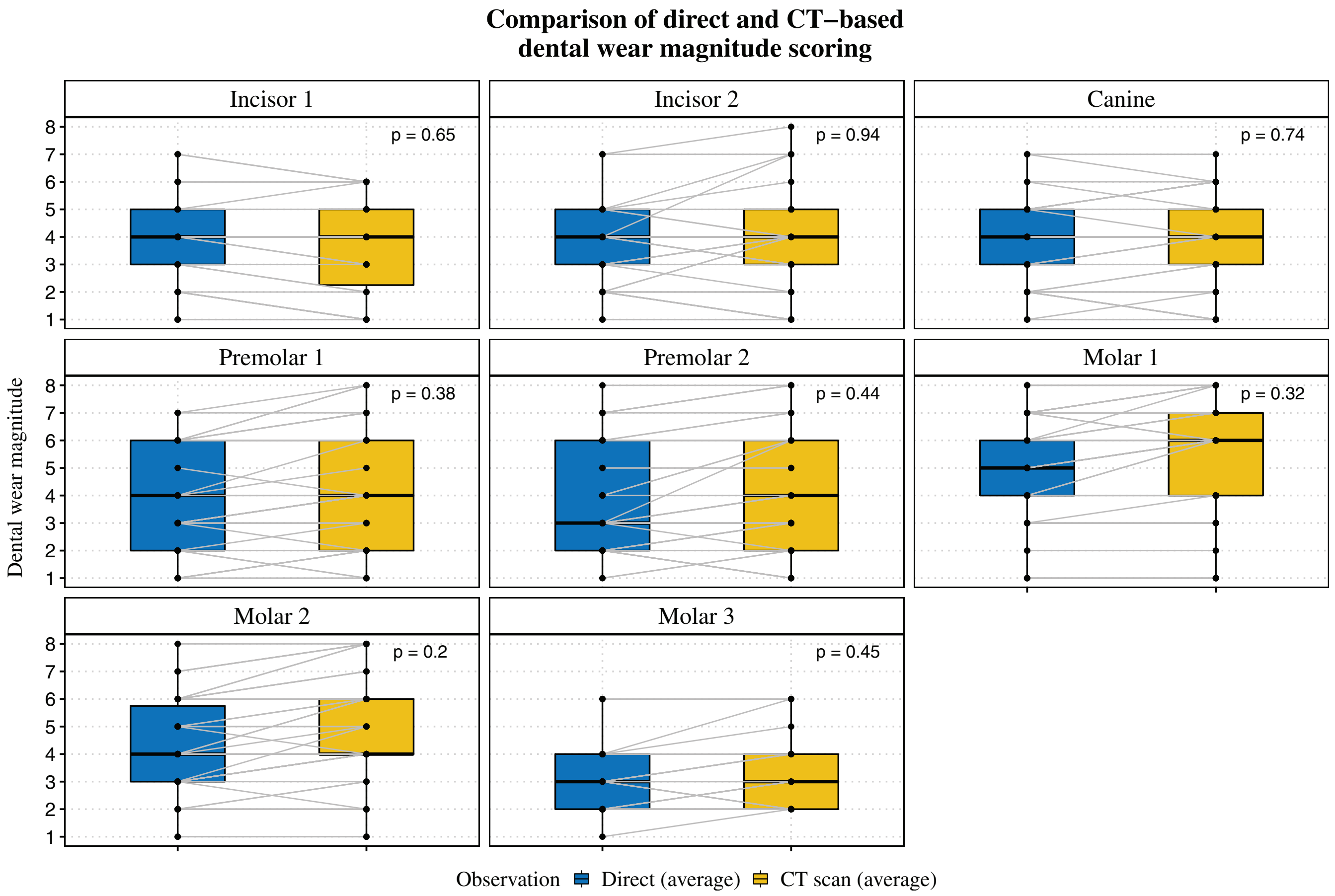

FIGURE 1 Comparison of dental wear scoring based on CT scan or direct observation of teeth. Results show some differences in scoring (which are not statistically significant at the level of $p \leq 0.05$, using the Wilcoxon non-parametric test) based on the two observation methods (which are typically equal or one stage different). Those differences are visible in the gray lines pairing the observations of the direct assessment (blue boxplots) to the corresponding observation of the CT scan-based assessment (yellow boxplots). When scores are equal the line is horizontal, connecting to the same scoring stage (e.g., lines connecting Stage 2 to Stage 2 in Molar 1). When scores are higher or lower, the lines ascend or descend pairing the corresponding observations with different scores. Note that each dot represents all the cases in which dental wear was allocated to that stage (stages with no cases are not represented by a corresponding dot; e.g., Stage 5 of CT scan based assessment of Molar 1)

3 | RESULTS

Analysis of the intra-observer error shows very strongly and statistically correlated scores with negligible differences between direct observations (R^2 values between 0.97 and 1; Figure S1). CT-scan based scores are also very strongly and statistically correlated, although slightly less so than scores based on direct observations (R^2 values between 0.90 and 0.96). Scores are also highly and statistically correlated between the several direct and CT scan-based observations (R^2 values between 0.85 and 0.96), despite less so than within observation methods.

The boxplots with paired cases in Figure 1 show some limited differences in the average scores of the two observation methods. Scoring was typically equal (321/412 = 78%) or one stage higher (34/412 = 8%) or lower (30/412 = 7%) in CT scan-based observations (Table S1). The largest scoring differences were cases in which CT scan-based scoring was two to three stages higher, but these were rare cases (two stages higher: 26/412 = 6%; three stages higher: 1/412 = 0.24%). This overall higher CT scan-based scoring is reflected in some tooth types with somewhat larger medians, and/or larger lower and upper quartiles in the boxplots (e.g., Premolar 2, Molar 1 and Molar 2). Despite this, differences were statistically not significant (at the level of $p \leq 0.05$).

4 | DISCUSSION

Our results show negligible intra-observer error and that dental wear scoring based on direct and CT scan-based observation is very highly correlated with no statistically meaningful differences. Thus, CT scan-based dental wear magnitude scoring is generally reliable. Notwithstanding, some limited scoring differences were found between the two methods. While a restricted proportion of those scoring differences may relate to intra-observer scoring error, the majority of such differences reflect CT scanning and visualization limitations which should be emphasized as potential caveats when examining dental wear magnitude based on CT scans.

Specifically, incautious selection of visualization thresholds may result in under- or over-estimations of dental wear. This is because CT scanning divides the scanned volume into small elements (voxels), which represent discrete regions of the scanned object (Hofer, 2010). Each voxel is allocated a gray-value that reflects the density of the discrete anatomical region it represents because the X-rays emitted by the CT scanner are attenuated according to the density of that same region (Hofer, 2010). However, if the region represented by the voxel includes materials with distinct densities (e.g., a voxel that includes both enamel and air or enamel and dentin) the density of those materials is averaged, and so the resulting gray value is an artifact of this Partial Volume Averaging Effect (Hofer, 2010). Because visualization of the CT scanned specimens is based on the selection of specific thresholds that define which voxels will be rendered, thresholds should be carefully set to avoid over- or under-selection of voxels that will subsequently impact ensuing visual analyses (Weber & Bookstein, 2011). Moreover, scanning spatial resolution (i.e., voxel size) should be as high as possible, provided it does not impact contrast resolution meaningfully (Hofer, 2010; Weber & Bookstein, 2011). This is because smaller voxels will represent smaller areas, and so will minimize the impact of the Partial Volume Averaging Effect (Weber & Bookstein, 2011).

In studies of dental wear magnitude, if excessively low thresholds are selected, then dental wear may not be apparent and so wear will likely be underestimated (Figure 2(a)). Conversely, if excessively high thresholds are selected, regions of enamel may be excluded and so dental wear may be overestimated (Figure 2(c)). This issue may be further complicated in regions in which enamel has been worn down and only a thin layer of enamel persists (e.g., the occlusal surface of molars). Because of Partial Volume Averaging (see above), such regions may be allocated lower gray values and so may be excluded from visualization in thresholding. Notwithstanding, researchers experienced in working with CT scans will likely not set excessively high or low thresholds and so this should not be a meaningful issue. Regardless, one way to circumvent this issue is to visually compare the visualization results of a given threshold with the physical specimen (either directly or using detailed photographs of the specimen when direct observation of the specimens is not possible), adjust the threshold accordingly and use it as reference. Alternatively, specific thresholds may be calculated and selected using more objective approaches such as the Half Maximum Height Value (Spoor et al., 1993) or the Threshold Mean Value (Weber et al., 1998), provided the visualization software allows inputting specific value based thresholds. Despite some user-specific subjectivity (e.g., selection of the specific location of where the transects are located to quantify gray level or Hounsfield values and so estimate the Half Maximum Height or Threshold Mean Values) these approaches are generally objective and so should minimize an extra subjective step (i.e., the selection of the visualization thresholds) in the assessment of dental wear magnitude. In the case of this study, slight differences in the selection of the thresholds between scoring rounds probably induced rendering differences that were sufficiently meaningful to induce scoring differences between rounds, thus resulting in larger intra-observer error (i.e., lower correlation coefficients) than that emerging from direct observation of specimens. Even though these differences were not statistically meaningful in our study, the use of such more objective approaches to select thresholds could have mitigated this effect and decreased scoring differences between rounds. Moreover, while the scoring differences were non-significant in our study, the selection of excessively high or low thresholds may induce significant over- or under- estimation of dental wear magnitude and so statistically bias the results of the study and any ensuing comparisons.

Our study aimed to assess if medical CT scans provide reliable results in the assessment of dental wear magnitude. This is because these scans

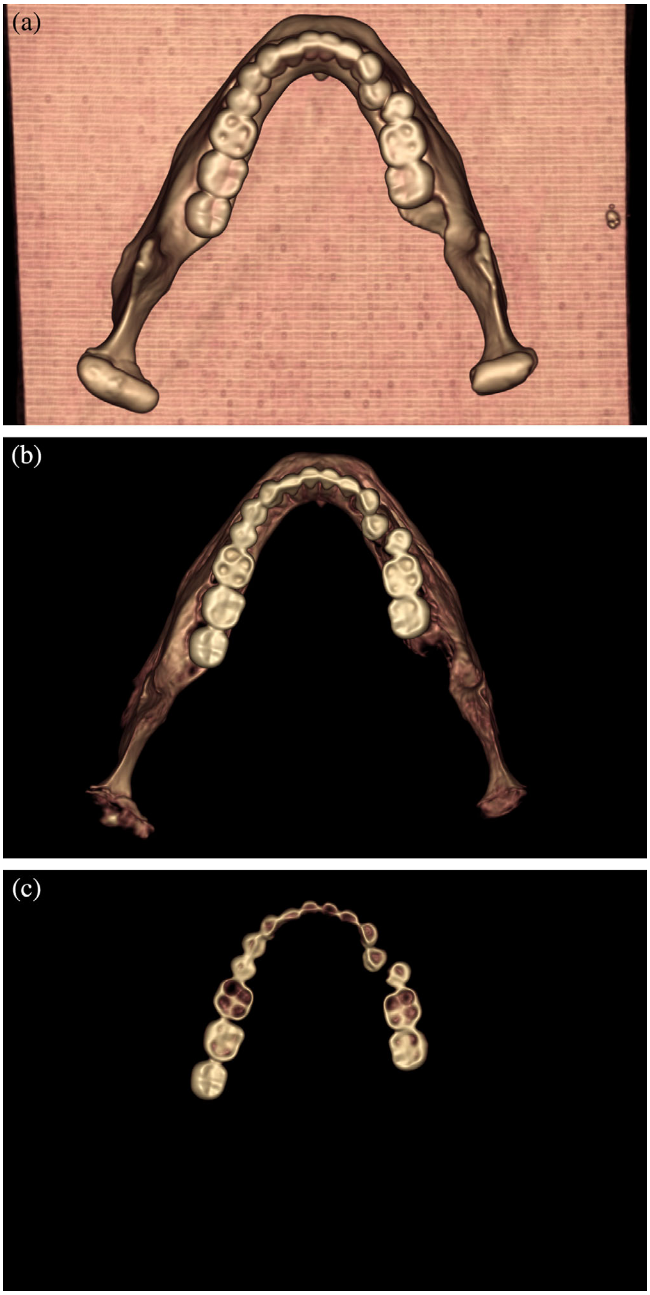

FIGURE 2 Example of the impact of selection of different thresholds on volume rendering and so on dental wear magnitude. (a) Very low threshold with apparently lower dental wear magnitude. (b) Medium threshold in which several voxels representing bone/teeth are excluded and in which dental wear is higher than in (a). (c) Very high threshold in which regions of thinner enamel (along with all bone) are excluded and so suggest higher dental wear than in (a) and (b)

are some of the 3D digitizations most commonly available to researchers. Notwithstanding, other resources may be used for indirect observation of specimens. Photographs may be used to examine dental wear and are cost effective. However, dental wear scoring often requires rotation of specimens to assess, for example, if the dental enamel rim has been worn out completely. This is not possible using 2D representations of specimens and so limits the use of photographs (although this limitation may be mitigated by taking many photos from different relevant perspectives). 3D surface scans of specimens may also be used. Although surface scanners may provide very high resolutions and some capture texture (i.e., color), they do not distinguish materials based on density like CT scans do. Thus, if the texture quality of the surface scans is low, it may be difficult to distinguish enamel from dentin and so score dental wear magnitude reliably. Magnetic resonance imaging (MRI) may also be used to acquire 3D digital volumes of specimens. However, MRI scanning is most efficient at imaging soft tissues and so are not most appropriate for bones/teeth (Weber & Bookstein, 2011). Micro-CT scanning may be used to digitize specimens at a micro-metric scale that should minimize the Partial Volume Averaging Effect greatly, and so provide very reliable dental wear magnitude assessments. However, digitizations of previously micro-CT scanned specimens are very rarely available. Moreover, micro-CT scanning on demand is typically very expensive and often uses tomographs with small scanning chambers that do not fit large skeletal elements (e.g., a mandible does not fit a desktop micro-CT scanner). Scanners with larger scanning chambers allow digitization of larger specimens (e.g., crania), but these are still very difficult to access and scanning even more expensive. Last, micro-CT scans require high-end computational resources that are not easily accessible (e.g., a micro-CT scan of a single tooth can easily achieve 2 GB, which requires 64 GB of ram memory to render).

In summary, medical CT scans of specimens are at this point the most commonly available and feasible 3D volume digitizations available to researchers and so we tested how reliable they are to score dental wear magnitude. Our results validate the use of CT scans, despite the above-mentioned limitations and provided the above-mentioned caveats are considered. Further studies may be conducted to assess the reliability of other alternative methods for indirect assessment of dental wear magnitude.

ACKNOWLEDGMENTS

Ricardo Miguel Godinho is funded by the European Regional Development Fund (FEDER) via the Programa Operacional CRESC Algarve 2020, of Portugal2020 (project ALG-01-0145-FEDER-29680). Célia Gonçalves is funded by the Portuguese Foundation for Science and Technology (FCT; contract reference DL 57/2016/CP1361/CT0029). This research was also partially funded by the Archaeological Institute of America (The Archaeology of Portugal Fellowship). Thanks are also due to Dr Miguel Ramalho† and José António Moita for granting access to the skeletal remains of Muge; Dr. António Carvalho, Drª Luísa Guerreiro and Dr. Paulo Alves for access and help in the scanning of the specimens housed at the Museu Nacional de Arqueologia; Prof. Sandra Jesus and Dr. Óscar Gamboa for CT scanning at the Faculty of Veterinary Medicine of the University of Lisbon.

CONFLICT OF INTEREST

The authors declare no conflict of interests.

AUTHOR CONTRIBUTIONS

Ricardo Miguel Godinho: Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; writing - original draft. **Célia Gonçalves:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; writing - original draft.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Ricardo Miguel Godinho https://orcid.org/0000-0003-0107-9577

REFERENCES

Aung, S., Ngim, R., & Lee, S. (1995). Evaluation of the laser scanner as a surface measuring tool and its accuracy compared with direct facial anthropometric measurements. *British Journal of Plastic Surgery*, *48*(8), 551–558.

Bastir, M., García-Martínez, D., Estalrrich, A., García-Tabernero, A., Huguet, R., Ríos, L., Barash, A., Recheis, W., de la Rasilla, M., & Rosas, A. (2015). The relevance of the first ribs of the El Sidrón site (Asturias, Spain) for the understanding of the Neandertal thorax. *Journal of Human Evolution*, *80*, 64–73. https://doi.org/10.1016/j.jhevol.2014.10.008

Bastir, M., García-Martínez, D., Torres-Tamayo, N., Palancar, C. A., Beyer, B., Barash, A., Villa, C., Sanchis-Gimeno, J. A., Riesco-López, A., Nalla, S., Torres-Sánchez, I., García-Río, F., Been, E., Gómez-Olivencia, A., Haeusler, M., Williams, S. A., & Spoor, F. (2020). Rib cage anatomy in *Homo erectus* suggests a recent evolutionary origin of modern human body shape. *Nature Ecology & Evolution*, *4*, 1178–1187. https://doi.org/10.1038/s41559-020-1240-4

Bernal, V., Novellino, P., Gonzalez, P. N., & Perez, S. I. (2007). Role of wild plant foods among late Holocene hunter-gatherers from central and North Patagonia (South America): An approach from dental evidence. *American Journal of Physical Anthropology*, *133*(4), 1047–1059. https://doi.org/10.1002/ajpa.20638

Chattah, N. L.-T., & Smith, P. (2006). Variation in occlusal dental wear of two chalcolithic populations in the southern Levant. *American Journal of Physical Anthropology*, *130*(4), 471–479. https://doi.org/10.1002/ajpa.20388

Clement, A. F., & Hillson, S. W. (2012). Intrapopulation variation in macro tooth wear patterns—A case study from Igloolik, Canada. *American Journal of Physical Anthropology*, *149*(4), 517–524. https://doi.org/10.1002/ajpa.22153

Deter, C. A. (2009). Gradients of occlusal wear in hunter-gatherers and agriculturalists. *American Journal of Physical Anthropology*, *138*(3), 247–254. https://doi.org/10.1002/ajpa.20922

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J. V., Pieper, S., & Kikinis, R. (2012). 3D slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, *30*(9), 1323–1341. https://doi.org/10.1016/j.mri.2012.05.001

Fornai, C., Benazzi, S., Gopher, A., Barkai, R., Sarig, R., Bookstein, F. L., Hershkovitz, I., & Weber, G. W. (2016). The Qesem cave hominin material (part 2): A morphometric analysis of dm2-QC2 deciduous lower second molar. *Quaternary International*, *398*, 175–189. https://doi.org/10.1016/j.quaint.2015.11.102

Fornai, C., Bookstein, F. L., & Weber, G. W. (2015). Variability of *Australopithecus* second maxillary molars from Sterkfontein member 4. *Journal of Human Evolution*, *85*, 181–192. https://doi.org/10.1016/j.jhevol.2015.05.013

Freidline, S. E., Gunz, P., Harvati, K., & Hublin, J. J. (2012). Middle Pleistocene human facial morphology in an evolutionary and developmental context. *Journal of Human Evolution*, *63*(5), 723–740. https://doi.org/10.1016/j.jhevol.2012.08.002

Freidline, S. E., Gunz, P., Harvati, K., & Hublin, J.-J. (2013). Evaluating developmental shape changes in *Homo antecessor* subadult facial morphology. *Journal of Human Evolution*, *65*(4), 404–423. https://doi.org/10.1016/j.jhevol.2013.07.012

García-Martínez, D., Torres-Tamayo, N., Torres-Sánchez, I., García-Río, F., Rosas, A., & Bastir, M. (2018). Ribcage measurements indicate greater lung capacity in Neanderthals and lower Pleistocene hominins compared to modern humans. *Communications Biology*, *1*(1), 117. https://doi.org/10.1038/s42003-018-0125-4

Ghoddousi, H., Edler, R., Haers, P., Wertheim, D., & Greenhill, D. (2007). Comparison of three methods of facial measurement. *International Journal of Oral and Maxillofacial Surgery*, *36*(3), 250–258. https://doi.org/10.1016/j.ijom.2006.10.001

Godinho, R. M., Fitton, L. C., Toro-Ibacache, V., Stringer, C. B., Lacruz, R. S., Bromage, T. G., & O'Higgins, P. (2018). The biting performance of *Homo sapiens* and *Homo heidelbergensis*. *Journal of Human Evolution*, *118*, 56–71. https://doi.org/10.1016/j.jhevol.2018.02.010

Godinho, R. M., & O'Higgins, P. (2017). Virtual reconstruction of cranial remains: The *H. heidelbergensis*, Kabwe 1 fossil. In D. Errickson & T. Thompson (Eds.), *Human remains - another dimension: The application of 3D imaging in funerary context* (pp. 135–147). Elsevier.

Godinho, R. M., & O'Higgins, P. (2018). The biomechanical significance of the frontal sinus in Kabwe 1 (*Homo heidelbergensis*). *Journal of Human Evolution*, *114*, 141–153. https://doi.org/10.1016/j.jhevol.2017.10.007

Godinho, R. M., O'Higgins, P., & Gonçalves, C. (2020). Assessing the reliability of virtual reconstruction of mandibles. *American Journal of Physical Anthropology*, *172*(4), 723–734. https://doi.org/10.1002/ajpa.24095

Godinho, R. M., Oliveira-Santos, I., Manuel Francisco, C.,. P., Maurício, A., Valera, A., & Gonçalves, D. (2019). Is enamel the only reliable hard tissue for sex metric estimation of burned skeletal remains in biological anthropology? *Journal of Archaeological Science: Reports*, *26*, 101876. https://doi.org/10.1016/j.jasrep.2019.101876

Godinho, R. M., Spikins, P., & O'Higgins, P. (2018). Supraorbital morphology and social dynamics in human evolution. *Nature Ecology & Evolution*, *2*(6), 956–961. https://doi.org/10.1038/s41559-018-0528-0

Godinho, R. M., Toro-Ibacache, V., Fitton, L. C., & O'Higgins, P. (2017). Finite element analysis of the cranium: Validity, sensitivity and future directions. *Comptes Rendus Palevol*, *16*(5), 600–612. https://doi.org/10.1016/j.crpv.2016.11.002

Gómez-Robles, A., Bermúdez de Castro, J. M., Martinón-Torres, M., Prado-Simón, L., & Arsuaga, J. L. (2015). A geometric morphometric analysis of hominin lower molars: Evolutionary implications and overview of postcanine dental variation. *Journal of Human Evolution*, *82*, 34–50. https://doi.org/10.1016/j.jhevol.2015.02.013

Gunz, P., Kozakowski, S., Neubauer, S., Le Cabec, A., Kullmer, O., Benazzi, S., Hublin, J.-J., & Begun, D. R. (2020). Skull reconstruction of the late Miocene ape *Rudapithecus hungaricus* from Rudabánya, Hungary. *Journal of Human Evolution*, *138*, 102687. https://doi.org/10.1016/j.jhevol.2019.102687

Gunz, P., Mitteroecker, P., Neubauer, S., Weber, G. W., & Bookstein, F. L. (2009). Principles for the virtual reconstruction of hominin crania. *Journal of Human Evolution*, *57*(1), 48–62. https://doi.org/10.1016/j.jhevol.2009.04.004

Hofer, M. (2010). *CT teaching manual. A systematic approach to CT reading* (2nd ed.). Thieme.

Kaifu, Y. (1999). Changes in the pattern of tooth wear from prehistoric to recent periods in Japan. *American Journal of Physical Anthropology*, *109*(4), 485–499.

Katz, D., & Friess, M. (2014). Technical note: 3D from standard digital photography of human crania—A preliminary assessment. *American Journal of Physical Anthropology*, *154*(1), 152–158. https://doi.org/10.1002/ajpa.22468

Katz, D. C., Grote, M. N., & Weaver, T. D. (2017). Changes in human skull morphology across the agricultural transition are consistent with softer diets in preindustrial farming groups. *Proceedings of the National Academy of Sciences*, *114*, 9050–9055. https://doi.org/10.1073/pnas.1702586114

Lubell, D., Jackes, M., Schwarcz, H., Knyf, M., & Meiklejohn, C. (1994). The Mesolithic-Neolithic transition in Portugal: Isotopic and dental evidence of diet. *Journal of Archaeological Science*, *21*(2), 201–216.

Martinón-Torres, M., Spěváčková, P., Gracia-Téllez, A., Martínez, I., Bruner, E., Arsuaga, J. L., & Bermúdez de Castro, J. M. (2013). Morphometric analysis of molars in a middle Pleistocene population shows a mosaic of 'modern' and Neanderthal features. *Journal of Anatomy*, *223*(4), 353–363. https://doi.org/10.1111/joa.12090

Molnar, S. (1971). Human tooth wear, tooth function and cultural variability. *American Journal of Physical Anthropology*, *34*(2), 175–189. https://doi.org/10.1002/ajpa.1330340204

Molnar, S., Barrett, M. J., Brian, L., Brace, C. L., Brose, D. S., Dewey, J. R., Frisch, J. E., Ganguly, P., Gejvall, N.-G., Greene, D. L., Kennedy, K. A. R., Poirier, F. E., Pourchet, M. J., Rhine, S., II, Turner, C. G., II, Valen, L. V., Koenigswald, G. H. R. V., Wilkinson, R. G., Wolpoff, M. H., & Wright, G. A. (1972). Tooth Wear and culture: A survey of tooth functions among some prehistoric populations [and comments and reply]. *Current Anthropology*, *13*(5), 511–526. https://doi.org/10.1086/201284

O'Higgins, P., Fitton, L. C., & Godinho, R. M. (2019). Geometric morphometrics and finite elements analysis: Assessing the functional implications of differences in craniofacial form in the hominin fossil record. *Journal of Archaeological Science*, *101*, 159–168. https://doi.org/10.1016/j.jas.2017.09.011

Sandholzer, M., Baron, K., Heimel, P., & Metscher, B. (2014). Volume analysis of heat-induced cracks in human molars: A preliminary study. *Journal of Forensic Dental Sciences*, *6*(2), 139–144. https://doi.org/10.4103/0975-1475.132545

Sandholzer, M. A., Walmsley, A. D., Lumley, P. J., & Landini, G. (2013). Radiologic evaluation of heat-induced shrinkage and shape preservation of human teeth using micro-CT. *Journal of Forensic Radiology and Imaging*, *1*(3), 107–111. https://doi.org/10.1016/j.jofri.2013.05.003

Sarig, R., Fornai, C., Pokhojaev, A., May, H., Hans, M., Latimer, B., Barzilai, O., Quam, R., & Weber, G. W. (2019). The dental remains from the early upper Paleolithic of Manot cave, Israel. *Journal of Human Evolution*, 102648. https://doi.org/10.1016/j.jhevol.2019.102648

Slizewski, A., Friess, M., & Semal, P. (2010). Surface scanning of anthropological specimens: Nominal-actual comparison with low cost laser scanner and high end fringe light projection surface scanning systems. *Quartär*, *57*, 179–187.

Smith, B. H. (1984). Patterns of molar wear in hunter–gatherers and agriculturalists. *American Journal of Physical Anthropology*, *63*(1), 39–56. https://doi.org/10.1002/ajpa.1330630107

Spoor, C. F., Zonneveld, F. W., & Macho, G. A. (1993). Linear measurements of cortical bone and dental enamel by computed-tomography - applications and problems. *American Journal of Physical Anthropology*, *91*(4), 469–484.

von Cramon-Taubadel, N. (2011). Global human mandibular variation reflects differences in agricultural and hunter-gatherer subsistence strategies. *Proceedings of the National Academy of Sciences*, *108*(49), 19546–19551. https://doi.org/10.1073/pnas.1113050108

Watson, J. T. (2008). Changes in food processing and occlusal dental wear during the early agricultural period in Northwest Mexico. *American Journal of Physical Anthropology*, *135*(1), 92–99. https://doi.org/10.1002/ajpa.20712

Watson, J. T., & Haas, R. (2017). Dental evidence for wild tuber processing among Titicaca Basin foragers 7000 ybp. *American Journal of Physical Anthropology*, *164*(1), 117–130. https://doi.org/10.1002/ajpa.23261

Weber, G. W. (2015). Virtual anthropology. *American Journal of Physical Anthropology*, *156*, 22–42. https://doi.org/10.1002/ajpa.22658

Weber, G. W., & Bookstein, F. L. (2011). *Virtual anthropology - a guide for a new interdisciplinary field*. Springer-Verlag.

Weber, G. W., Recheis, W., Scholze, T., & Seidler, H. (1998). Virtual anthropology (VA): Methodological aspects of linear and volume measurements - first results. *Collegium Antropologicum*, *22*(2), 575–584.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., & Zemla, J. (2017). corrplot: Visualization of a Correlation Matrix. R package version 0.89. URL: https://cran.r-project.org/web/packages/corrplot/index.html

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., & Dunnington, D. (2021). ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. R package version 3.3.4. URL: https://cran.r-project.org/web/packages/ggplot2/index.html

Wroe, S., Parr, W. C. H., Ledogar, J. A., Bourke, J., Evans, S. P., Fiorenza, L., Benazzi, S., Hublin, J.-J., Stringer, C., Kullmer, O., Curry, M., Rae, T. C., & Yokley, T. R. (2018). Computer simulations show that Neanderthal facial morphology represents adaptation to cold and high energy demands, but not heavy biting. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1876).

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Godinho, R. M., & Gonçalves, C. (2021). Testing the reliability of CT scan-based dental wear magnitude scoring. *American Journal of Physical Anthropology*, 1–7. https://doi.org/10.1002/ajpa.24374