Received: 12 February 2020 | Revised: 10 April 2020 | Accepted: 11 May 2020

DOI: 10.1002/oa.2885

RESEARCH ARTICLE

Observer error in bone disease description: A cautionary note

Lucie Biehler-Gomez[1] | Lara Indra[1] | Federica Martino[2] | Carlo Pietro Campobasso[2] | Cristina Cattaneo[1]

[1]LABANOF, Laboratorio Di Antropologia E Odontologia Forense, Sezione Di Medicina Legale, Dipartimento Di Scienze Biomediche per La Salute, Università Degli Studi Di Milano, Milan, Italy

[2]Dipartimento di Medicina Sperimentale, Università degli Studi della Campania "Luigi Vanvitelli", Naples, Italy

Correspondence
Lucie Biehler-Gomez, LABANOF, Laboratorio di Antropologia e Odontologia Forense, Sezione di Medicina Legale, Dipartimento di Scienze Biomediche per la Salute, Università degli Studi di Milano, Via Mangiagalli 37, Milan 20133, Italy.
Email: lucie.biehler@unimi.it

Abstract

The aim of this paper is to examine the accuracy of pathological description on human bones. Ten participants (five forensic pathologists and five anthropologists) were asked to describe 30 bone lesions through observation of the real specimens and photographic images, including character of the lesion, the aspect of the margins and the presence of periosteal new bone, according to recognized and accepted pathological terminology on dry bone. Results were analysed using statistical analysis and interobserver and intraobserver agreements were tested. The anthropologists showed slightly more consistent and accurate results compared with the forensic pathologists, and overall results were better when assessed on the real specimens. Lesion descriptions showed important contradictions and inaccuracies, particularly in the evaluation of the character of the lesion and periosteal new bone, with dramatic potential consequences for the diagnosis of bone disease. This study shows the considerable pitfalls in the assessment of basic pathological bone manifestations and demonstrates the importance of continuing efforts in the standardization of pathological terminology on dry bone.

KEYWORDS

biological profile, bone disease, bone lesions, differential diagnosis, observer error, palaeopathology, terminology

1 | INTRODUCTION

In the anthropology practice, one of the main objectives is the construction of an informative and reliable biological profile. The biological profile includes the estimations of sex, age, ancestry and stature, the recording of anatomical variants, and the analysis of traumatic injuries and pathological bone lesions. The macroscopic diagnosis of pathological conditions is based on the interpretation of the morphological appearance of bone lesions, their position on the bone and their distribution on the skeleton compared with the clinical literature and previously published cases. The first step in the study of bone lesions consists in the recording of pathological bone abnormalities; indeed, as Roberts and Connell (2004, p. 9) mention, "the only way to attempt any form of classification or diagnosis of disease in skeletal remains is with clear and objective description." Thus, a correct description of bone lesions using a comparable terminology understood by all is essential for an accurate diagnosis on dry bone, for the evaluation of proposed diagnoses by other researchers and for the archiving of the documentation of bone lesions, especially in cases where the skeletal material will no longer be available (e.g., skeletons may be reburied).

From the 1980's, increased concern was raised relating to the use of inappropriate terms in describing normal and abnormal bone structures. The *Nomenclature in Paleopathology* (Manchester, Ogden, & Storm, 2016), published in the Paleopathology Association Newsletter, represents one of the most important works aiming to standardize the terminology that may be used to describe bone structures to improve linguistic exactitude. Continuing efforts were realized to

Institution from which the paper emanated: LABANOF, Laboratorio Di Antropologia E Odontologia Forense, Sezione Di Medicina Legale, Dipartimento Di Scienze Biomediche per La Salute, Università Degli Studi Di Milano, Milan, Italy

increase scientific rigour, consensus and accuracy in the documentation and diagnosis of bone disease, in particular with guidelines for the recording of pathological bone abnormalities (Brickley & McKinley, 2004; Buikstra & Ubelaker, 1994; Mitchell & Brickley, 2017) suggesting methodologies for an accurate description of bone lesions and the adaptation of the "Istanbul Protocol Manual on the Effective Investigation and Documentation of Torture and other Cruel, Inhuman or Degrading Treatment or Punishment" (United Nations, 2004), used in forensic sciences to indicate the degree of certainty of the diagnosis of trauma suffered during torture, to its application in the diagnosis of bone disease (Appleby, Thomas, & Buikstra, 2015). In addition, excellent articles have endeavored to standardize the terminology used to describe bone lesions (Buikstra, Cook, & Bolhofner, 2017; Grauer, 2008; Klaus, 2017; Lovell, 2000; Mays, 2018; Miller, Ragsdale, & Ortner, 1996; Ortner, 2003, 2012; Ragsdale, Campbell, & Kirkpatrick, 2018) and alerted on the importance of rigorous descriptions of skeletal lesions as the specific terms selected "describe the interlinked factors of the morphology and underlying pathological process (es) of abnormal new bone formation, abnormal bone loss, and abnormal bone size/shape" (Klaus, 2017, p.98) and thus constitute descriptive tools orienting the differential diagnosis. Nonetheless, there is still considerable variation in the description of bone lesions that may increase interobserver error and ambiguity and ultimately lead to misdiagnoses and misinterpretations among scientists. Consequently, continuing training in the recording of bone lesions is needed.

In this article, we asked 10 participants (five forensic pathologists and five anthropologists) to fill in a questionnaire for the description of 30 bone lesions according to terminology used in palaeopathological reference works (Aufderheide & Rodríguez-Martín, 1998; Buikstra et al., 2017; Klaus, 2017; Manchester et al., 2016; Mays, 2018; Ortner, 2003, 2011; Ragsdale et al., 2018; Santos & Roberts, 2006). In addition, interobserver and intraobserver errors were tested. It is important to note that the diagnosis of the conditions responsible for the lesions was not evaluated. The aim of this research was to observe the similarities and discrepancies in the description of bone lesions.

2 | MATERIALS AND METHODS

The study consisted in the description of 30 bone lesions from 27 bones; in three instances, two lesions were located on the same bone but in different positions. Twenty-one bones were selected from skeletons of the CAL Milano Cemetery Skeletal Collection with antemortem diagnoses of disease (Cattaneo et al., 2018), and six were archaeological bones from the CAL archaeological collection composed of about 5,000 skeletons from over 50 archaeological excavation sites in Italy. The lesions were undamaged by taphonomic alterations and clearly signalled on the bones, as the aim of this study was not to assess the detection of bone lesions but the variation in the description and classification of bone lesions according to recognized and accepted pathological terminology on dry bone. These bone alterations included lesions due to metastatic carcinoma, multiple myeloma, rheumatoid arthritis and diabetes (documented from associated records), as well as tuberculosis, congenital syphilis, osteomyelitis, rickets, antemortem trauma and periosteal new bone on long bones (from archaeological remains). It is important to note that the diagnoses of the conditions were never asked, only the description of the lesions based on predefined answers.

Ten participants were asked to fill in a questionnaire assessing the types of abnormalities of 30 bone lesions: five forensic pathologists with practical experience in anthropology constituted the "forensic pathologists" group and five anthropologists with experience in forensic anthropology and/or bioarchaeology (varying between 3 and 10 years) formed the "anthropologists" (the experience of the participants is described in Table 1). Anthropologists are all graduates in biology or natural sciences, who, for their MSc and/or PhD, have specialized in forensic anthropology and bioarchaeology, forensic genetics and anthropology or forensic sciences and anthropology (Table 1). The years of experience in Table 1 refer to how long they have been working (through research grants or professional case work) in the field of physical anthropology after their last degree. Forensic pathologists are graduates in medicine who have a specialization in forensic medicine and who have taken up anthropology during their internship in forensic medicine. As with the anthropologists, the years of experience in Table 1 refer to years of practice on dry bone through research grants or casework in forensic anthropology after the end of their specialization.

The participants were not specifically trained before the study on the description of bone lesions so that the results could be an authentic representation of real-life scenarios. Each participant was asked to describe the 30 bone lesions twice, first based on photographic images on the computer (referred to as "computer assessment") and after a few days through the visual observation of the real bone specimens ("visual assessment") to see if any difference was noted.

The description of bone lesions was based on a checklist with multiple predefined answers and divided in various sections, including character of the lesion, dimensions, degree of bone resorption (for osteolytic lesions), margins of the lesion (for osteolytic lesions), periosteal new bone and articular involvement (the original checklist is presented in Supporting Information). For the purpose of this research, the results of three categories will be considered here: character of the lesion, margins and periosteal new bone. The questionnaire of the categories selected for study and submitted to the participants with their predefined answers is presented in Table 2, and an example is provided in Figure 1. The description of bone abnormalities was based on standard and published palaeopathological terminology (Aufderheide & Rodríguez-Martín, 1998; Biehler-Gomez, Giordano, & Cattaneo, 2019; Buikstra et al., 2017; Klaus, 2017; Mays, 2018; Ortner, 2003, 2011; Ragsdale et al., 2018; Santos & Roberts, 2006). These categories were selected because they are particularly informative on the disease process and are therefore of great interpretative value. The correct recording of the character of the lesion (solely bone forming, solely bone destruction or a combination of both components), the presence or absence of bone remodelling on the margins

TABLE 1 Description of the academic background and postacademic experience of the participants

Participants	Group	Degree	Specialization	Years of experience
Participant no. 1	Anthropologist	MS	Forensic Anthropology and Bioarchaeology	10
Participant no. 2	Anthropologist	PhD	Forensic Anthropology and Bioarchaeology	8
Participant no. 3	Anthropologist	PhD	Forensic Genetics and Anthropology	5
Participant no. 4	Anthropologist	MS	Forensic Sciences and Anthropology	3
Participant no. 5	Anthropologist	MS	Forensic Anthropology and Bioarchaeology	3
Participant no. 6	Forensic pathologist	MD	Forensic Pathology and Anthropology	3
Participant no. 7	Forensic pathologist	MD	Forensic Pathology and Anthropology	3
Participant no. 8	Forensic pathologist	MD	Forensic Pathology and Anthropology	5
Participant no. 9	Forensic pathologist	MD	Forensic Pathology and Anthropology	3
Participant no. 10	Forensic pathologist	MD	Forensic Pathology and Anthropology	10

TABLE 2 Categories and possible answers submitted during the test and analysed in the present study

Character of lesion	Only osteoblastic
	Mostly osteoblastic, but osteolytic component also present
	Mostly osteolytic, but osteoblastic component also present
	Only osteolytic
Margins[a]	Rounded and remodelled
	No remodelling
Periosteal new bone	Absent
	Woven bone deposition (young periosteal new bone)
	Remodelled lamellar bone (remodelled periosteal new bone)
	Spiculated new bone formation

[a]Only applicable to osteolytic lesions.

of an osteolytic lesion and the presence and type of periosteal new bone can not only indicate an attempt at bone healing but also informs on the rapidity of the process. As Ortner (2012) explains, "in general, the slower the process, the more clearly defined the margins of a destructive lesion will be" on radiographs due to bone remodelling. In osteolytic lesions, well-defined remodelled margins evidence a slow and chronic process, whereas poorly defined margins (without bone remodelling) suggest a rapid and aggressive process (Lovell, 2000; Ortner, 2012). Similarly, in osteoblastic lesions, woven bone (poorly organized porous fibre bone) is the result of a very rapid bone formation, which remodels overtime into smooth and dense compact bone, thus indicating a chronic and relatively slow process (Lovell, 2000; Ortner, 2012). Finally, projecting spicules of bone reveal a very rapid and aggressive process of bone formation (Lovell, 2000; Ragsdale et al., 2018).

Although only one answer was expected in the "character of lesion" and "margins" categories, several answers could be checked in the "periosteal new bone" category. Based on paleopathological literature (Aufderheide & Rodríguez-Martín, 1998; Ortner, 2003, 2011; Ragsdale et al., 2018; Santos & Roberts, 2006), one of the authors (LBG) elaborated the "correct" results, which were then used as reference to calculate the accuracy rates of the participants.

Results were analysed using statistical analysis. Interobserver agreement was realized with Fleiss' Kappa in Excel (Microsoft® Excel® 2016), intraobserver agreement was performed with Cohen Kappa in SPSS (IBM SPSS Statistics for Windows, version 21) and intraobserver agreements for periosteal new bone and accuracy rate calculations were done with Excel (Microsoft® Excel® 2016).

3 | RESULTS

3.1 | Interobserver agreement

Based on Landis and Koch (1977), the results of the Fleiss' Kappa interobserver agreements (Table 3) indicate that among the anthropologists, the interobserver agreement is moderate (scoring 0.41–0.60) to substantial (scoring 0.61–0.80) in all categories except for one: the "computer assessment" of the "character of the lesion," which can be classified as a "fair" agreement (scoring 0.21–0.40). In addition, the agreement values are higher in all three categories of the "visual assessment," whereas the computer-based assessment shows less agreement. Among the forensic pathologists, all values lie within the fair agreement section. Overall, the Fleiss' Kappa values are higher in the "visual assessment" than in the "computer assessment" and in the group of anthropologists than among forensic pathologists (Figure 2). Figure 2 also shows that the category "character of lesion" shows the poorest agreement among participants.

3.2 | Intraobserver agreement

Similarly, the Cohen Kappa intraobserver agreement results (Table 4) show a moderate to substantial intraobserver agreement between "visual" and "computer" assessment for both groups with variation

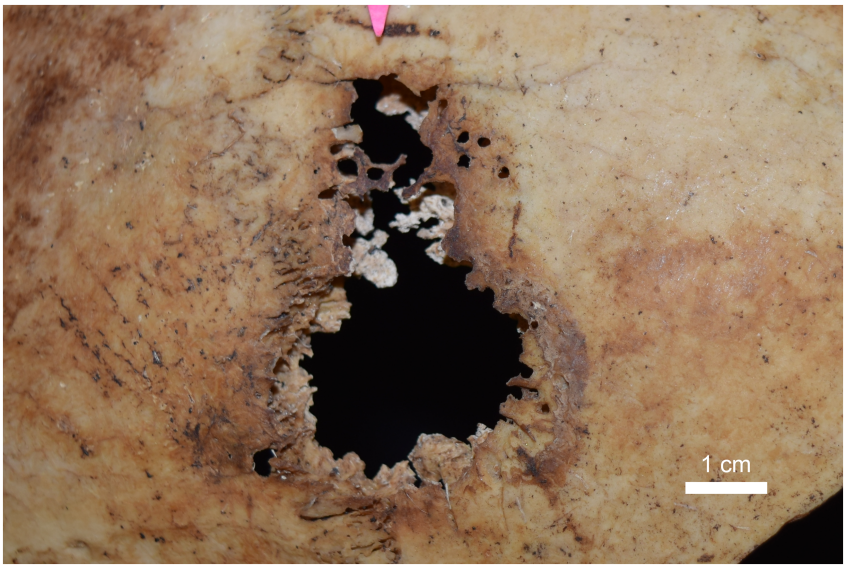

FIGURE 1 Example of bone lesion submitted to the participants. Lateral view of the right side of the frontal bone; the coronal suture is visible on the left of the lesion. The figure shows a large osteolytic lesion with an extensive resorption of the diploe, thus indicating that the lesion originated in the trabecular structure of the bone and perforated the bone cortex through extension of the lesion. In addition, the margins are denticulated, and no bone remodelling/osteoblastic activity (indicating healing) is visible, suggesting a rapid and aggressive process. In fact, this lesion was observed in the cranium of a documented individual from the CAL Milano Cemetery Skeletal Collection (Cattaneo et al., 2018) with a clinical diagnosis of metastatic liver cancer. The expected answers included, for the character of the lesion, "only osteolytic"; for the margins, "no remodelling"; for periosteal new bone, "absent" [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Fleiss' Kappa values for the groups of anthropologists and forensic pathologists, according to each category of description of lesions and divided per type of assessment

Category of lesion description	Anthropologists, Visual assessment	Anthropologists, computer assessment	Forensic pathologists, visual assessment	Forensic pathologists, computer assessment
Character of lesion	0.5088	0.3781	0.2881	0.2594
Margins	0.6252	0.5856	0.3979	0.2782
Periosteal new bone	0.6843	0.6827	0.3200	0.2500
Overall	0.6061	0.5488	0.3353	0.2625
	Substantial agreement	Moderate agreement	Fair agreement	Fair agreement

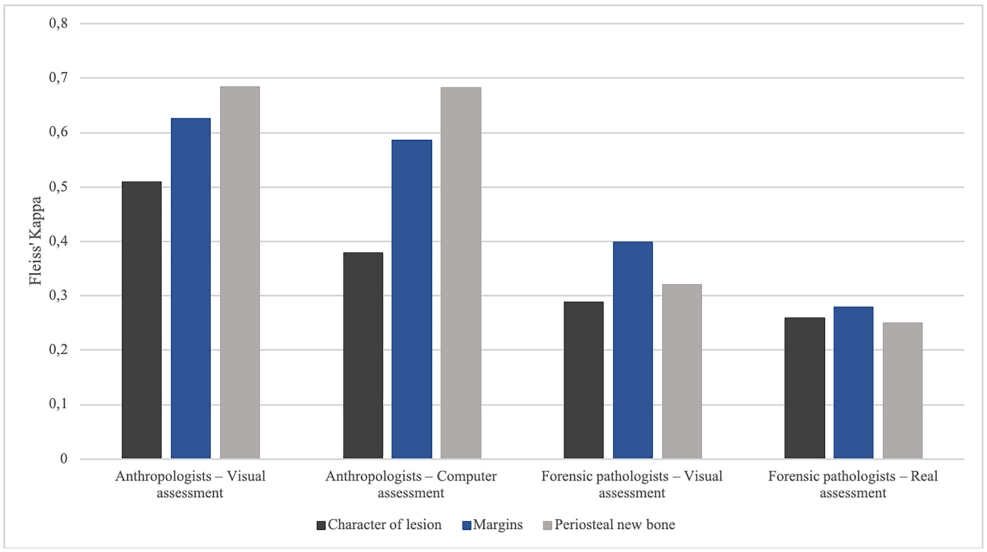

FIGURE 2 Inter-observer variance shown by Fleiss' Kappa between the groups of anthropologists and forensic pathologists according to the type of assessment of the bone lesions [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 4 Cohen Kappa values for all participants per category of description of lesions. All Cohen Kappa values were significant with $\alpha = 0.05$

	Participants of the group of anthropologists						Participants of the group of forensic pathologists[a]				
Category of lesion description	1	2	3	4	5	Avg.	6	7	8	9	Avg.
Character of lesion	0.632	0.850	0.799	0.818	0.441	0.544	0.949	1.000	0.341	0.351	0.660
Margins	0.639	1.000	0.847	0.947	0.551	0.743	0.947	0.947	0.273	0.533	0.675
Periosteal new bone[b]	0.735	0.860	0.925	0.639	0.753	0.782	0.988	1.000	0.622	0.658	0.817
Total						0.690					0.717

[a]Participant no. 10 did not perform the "visual assessment" of lesions and so the intraobserver agreement could not be calculated.
[b]For the Cohen Kappa calculation of multiple answers in the "periosteal new bone" category, partial agreement was not considered, the answers had to be in the exact same combination to be seen as correct.

between participants from fair to perfect agreements (Landis & Koch, 1977). The average values indicate a slightly greater agreement in the group of forensic pathologists and for the "periosteal new bone" category.

3.3 | Accuracy rates

As seen in Table 5, the highest accuracies of both groups were obtained in the "margins" category and the lowest in the "periosteal new bone" category for the forensic pathologists and "character of the lesion" category for the anthropologists, regardless of the type of assessment used. Overall, "visual" assessment of bone lesions was more often correct than when computer-based and evaluations made by the anthropologists were slightly more accurate than in the group of forensic pathologists. The accuracy rates of the positive identification of the different types of periosteal new bone (Table 6) show poor results, close to 65% (e.g., a periosteal new bone was described when it was absent or vice versa or the wrong type of periosteal new bone was noted as present). In addition, about one in two woven or lamellar new bone depositions (the most commonly found types of periosteal new bone) were wrongly identified (51.1% positive identification).

3.4 | Influence of the experience of the participants

Table 7 shows the correlation between the experience of the participants and their accuracy rates. Surprisingly, participants with less experience showed an overall accuracy in the description of bone lesions slightly higher than participants with 5 years of experience and even 8–10 years of experience. Indeed, participants with 3 years of experience showed a mean accuracy rate of 72.76% on visual assessment and 69% on computer assessment, whereas the accuracy rate of participants with 5 years of experience was 62.29% on real specimens and 55.78% on photographs and that of the participants with 8–10 years of experience was 69.72% on visual assessment.

3.5 | Consistency in the assessment of the lesions

Inconsistencies between the assessment of character of the lesion and their descriptions were noted. For instance, seven participants (four anthropologists and three forensic pathologists) noted some lesions as either osteolytic or osteoblastic (by opposition to manifesting both components) and then listed the presence of periosteal new bone/remodelled margins or bone resorption, respectively, which contradicts the previous assessment of a lesion solely osteolytic or osteoblastic.

4 | DISCUSSION

The interobserver agreement analysis shows that anthropologists tended to agree more in their description of lesions than the forensic pathologists in our study (Table 3). In addition, they were slightly more correct in their description of bone lesions than the participants with a medico-legal background (Table 5). These results were expected as the anthropologists are more frequently confronted to skeletal

TABLE 5 Accuracy rates for the computer-based and "visual" assessments of the characteristics of bone lesions between the groups of anthropologists and forensic pathologists

	Anthropologists		Forensic pathologists	
Category of lesion description	Computer assessment (%)	Visual assessment (%)	Computer assessment (%)	Visual assessment (%)
Character of lesion	64.33	65.83	66.67	66.67
Margins	80.67	80.00	70.67	80.67
Periosteal new bone	64.53	67.83	57.67	62.50
Total	69.84	71.22	65.00	69.95

TABLE 6 Accuracy rates for the positive identifications in the subcategories of periosteal new bone between the groups of anthropologists and forensic pathologists

	Anthropologists		Forensic pathologists	
Subcategories of periosteal new bone	Computer assessment (%)	Visual assessment (%)	Computer assessment (%)	Visual assessment (%)
Absent	77.50	76.25	60.00	70.31
Woven bone	46.67	55.56	40.00	44.44
Lamellar bone	57.78	62.22	57.78	44.44
Spiculated	80	80	70	100
Avg.	65.49	68.51	56.95	64.80

remains than forensic pathologists and so they have higher probabilities to have been confronted to bone lesions, thus explaining a better accuracy in their assessment.

Participants generally tended to describe lesions similarly whether they performed the analysis based on photographs or by visual examination of the actual osseous elements (Table 4). Their answers were more similar on "visual" assessment and were slightly more accurate than when assessed on photographs (Tables 3 and 5). Again, these results were expected because details of bone lesions may be more difficult to recognize on photographs due to the constraint of definition and the lack of perspective. In addition, they demonstrate that visual observation should be preferred to photographs when recording bone lesions.

Participants tended to disagree on their assessment of the character of the lesions, without substantially changing their answers between modalities of assessment (Tables 3 and 4). It is interesting to see that this category, which is the initial interpretation of the lesion and may significantly impact the diagnosis of the disease responsible for the lesion, showed so much variation between participants and such a low accuracy (varying between 64% and 67%; Table 5). Another interesting finding was that in a total of 19 instances, the participants evaluated lesions to be solely osteolytic or osteoblastic but also recorded the presence of bone production or bone resorption, respectively, in other categories. These variations, low accuracies and contradictory assessments testify to a need for continuous training in the description and recording of bone lesions.

The margin category was limited to a binary response: either rounded and remodelled or not remodelled (Table 2); it is therefore understandable that the results of this category were more similar between participants and modalities (Tables 3 and 4) and the most accurate results (Table 5). Nonetheless, they also show that the descriptions of the pathological changes at the margin are the easiest to interpret.

Contrary to the margin category, the interpretation of periosteal new bone posed more difficulties. Although the participants showed some variation in their results between themselves (in particular in the group of forensic pathologists, which may be due to the fact that they are less frequently confronted to pathologic lesions on dry bones than the group of anthropologists, resulting in more uncertainty) and changed very little their results between modalities, this category is also the one with the poorest accuracy (ranging between 58% and 68% and with a 57% to 69% rate of positive identification). In fact, absent periosteal new bone (i.e., the absence of a new bone deposition on the cortical surface) was correctly identified in only 71% of cases. Overall, woven and lamellar new bone (the most commonly encountered types of periosteal new bone) were correctly identified about half of the time (51.1%; Table 6). In two lesions in particular, we could understand that the participants mistook spongiosclerosis (observed in the trabecular bone exposed by an osteolytic perforation of the bone cortex) for the presence of a periosteal new bone. These results demonstrate a poor understanding of the terminology associated to the different types of periosteal new bone, which is particularly alarming considering that these may considerably orient the understanding of the disease process and the diagnosis of the causative agent responsible for the lesion. Ortner (2012, p. 252) said that "careful attention to the characteristics of the abnormal features created by the bone cells provides critical information basic to the description and diagnosis of all skeletal disorders." However, and as highlighted in this study, there is still substantial variation, disagreement and inaccuracies in the description of bone lesions. Consequently, the understanding of the characteristics of bone lesions and the standardization of their description remain a crucial topic to address for the correct recording of bone lesions, the documentation of comparable data and the accurate diagnosis of bone disease.

The study also considered the possible influence of the experience, specialization and training of the participants on the results obtained. Indeed, one expects specialized observers with a long experience to fare much better in the description of bone lesions than young practitioners confronted for the first time with this situation. The study thus involved 10 participants with various experience on dry bone (which can be divided in three groups: 3, 5 and 8–10 years), trained in western Europe (Italy, England, France and Switzerland) including five forensic pathologists (specialized in forensic pathology and forensic anthropology) and five anthropologists (with specialization in forensic anthropology as well as other fields including bioarchaeology, forensic genetics and forensic sciences; Table 1). As a result of the study, the anthropologists showed substantial agreement in their description of bone lesions on real specimens (Fleiss' Kappa value: 0.6061) despite the variation in their years of experience and fields of specialization, suggesting that these factors did not influence their analysis. The forensic pathologists only showed a fair agreement in their results (Fleiss' Kappa values: 0.3353 and 0.2625) in spite of

TABLE 7 Relationship between the accuracy rate of the participants and their experience on dry bone

Experience	3 years						5 years			8–10 years			
Participant	Participant no. 4	Participant no. 5	Participant no. 6	Participant no. 7	Participant no. 9	Total	Participant no. 3	Participant no. 8	Total	Participant no. 2	Participant no. 1	Participant no. 10[a]	Total
Visual assessment (in %)													
Character of lesion	66.67	56.67	96.67	63.33	66.67	70.00	76.67	36.67	56.67	73.33	60.00	-	66.67
Margins	80.00	83.33	100	76.67	86.67	85.33	83.33	56.67	70	80.00	76.67	-	78.34
Periosteal new bone	65.83	77.22	66.67	54.58	50.42	62.94	63.33	57.08	60.20	66.67	61.67	-	64.17
Total	70.83	72.41	87.78	64.86	67.92	72.76	74.44	50.14	62.29	73.33	66.11	-	69.72
Computer assessment (in %)													
Character of lesion	65.00	50.00	93.33	63.33	70.00	68.33	83.33	33.33	58.33	73.33	50.00	73.33	65.55
Margins	76.67	60.00	96.67	76.67	70.00	76.00	93.33	30.00	61.67	80.00	93.33	80.00	84.44
Periosteal new bone	63.75	63.33	87.92	54.58	43.75	62.67	63.88	30.83	47.36	65.00	55.00	63.75	61.25
Total	68.47	57.78	92.64	64.86	61.25	69.00	80.18	31.39	55.78	72.78	66.11	72.36	70.42

[a]Participant no. 10 did not perform the "visual assessment" of lesions and so the accuracy rate could not be calculated.

having the same specialization. In this group, although the specialization cannot be a determining factor, the different level of experience of the participants may explain the variation in the results. Yet Participants 8 and 10, with, respectively, 5 and 10 years of experience, did not evidence a higher accuracy rate than Participant 6, with only 3 years of experience on dry bone (Table 7). In fact, Table 7 demonstrates that the participants with less experience (3 years; Participants 4, 5, 6, 7 and 9; accuracy rates: 72.76% in real specimens and 69% on photographs) had an overall slightly better accuracy in the description of bone lesions than those with a longer experience (5 years; Participants 3 and 8; accuracy rates: 62.29% on real specimens and 55.78% on photographs; and 8–10 years; Participants 1, 2 and 10; accuracy rate: 69.72% on real specimens). Finally, Participants 6 and 8 are forensic pathologists trained in Italy, yet the former shows the highest accuracy rates of the study and the latter the poorest (Tables 1 and 7). Therefore, the variation of results observed in this study cannot be explained by differences in years of experience, fields of specialization or even geographical place of training of the participants. It is probable that specific training in bone pathology would have increased the accuracy rates and reduced the variation observed in our results. Nonetheless, we decided not to train the participants beforehand in order not to skew the results and give an authentic representation of real-life scenarios, to which any of the participants selected for this study could be confronted (as describing bone lesions forms the basis of diagnosis in bone disease). Consequently, a possible explanation for these results is that the description of bone lesions constitutes a specific skillset that is not entirely obtained through the experience of the forensic pathologist or anthropologist or their specialty in forensic anthropology or bioarchaeology, despite their ability, range of knowledge and practice. The results of this study thus testify to the need for more specific and thorough training in the description of bone lesions to all practitioners on dry bone, regardless of their field of specialization or experience.

Previous studies (Bridges, 1993; Waldron & Rogers, 1991) have already shown the existing interobserver disagreement in the scoring and analysis of bone lesions (specifically, due to osteoarthritis) and alerted on the need for a standardized methodology in the analysis of bone lesions. In the present study, however, we also evidenced critical issues in lesion recording, in particular when assessing the character of lesion (bone forming or bone destruction) as well as the presence of periosteal bone forming lesions (regardless of the exact subcategory), namely, inconsistencies, confusions and low accuracies. The results of this exploratory study thus argue in favour of continuing efforts in the training on the description of the characteristics of bone lesions.

5 | CONCLUSION

The accurate description of bone lesions is an essential step for the correct diagnosis of bone disease. In this study, we submitted 30 bone lesions to 10 anthropologist and forensic pathologist participants. As a result, anthropologists showed slightly more consistent and accurate results compared with the forensic pathologists, and overall results were better when assessed by visual examination of the bone lesions. Lesion descriptions, however, showed important inaccuracies and contradictions, particularly in the evaluation of the character of the lesion and periosteal new bone, with dramatic potential consequences for the diagnosis of bone disease. We can only recommend continuing training and communication on the identification of the characteristics of bone lesions for the correct recording of bone lesions, the documentation of comparable data and the accurate diagnosis of bone disease. This study shows the considerable pitfalls in the description and classification of basic pathological bone manifestations and demonstrates the importance of continuing efforts in the standardization of pathological terminology on dry bone.

CONFLICT OF INTEREST

None.

FUNDING

This research did not receive any specific grants from public, commercial and not-for-profit funding agencies.

ORCID

Lucie Biehler-Gomez https://orcid.org/0000-0001-6674-7850
Lara Indra https://orcid.org/0000-0001-6644-2133

REFERENCES

Appleby, J., Thomas, R., & Buikstra, J. (2015). Increasing confidence in paleopathological diagnosis–Application of the Istanbul terminological framework. *International Journal of Paleopathology*, *8*, 19–21. https://doi.org/10.1016/j.ijpp.2014.07.003

Aufderheide, A. C., & Rodríguez-Martín, C. (1998). *The Cambridge encyclopedia of human paleopathology*. Cambridge, United Kingdom: Cambridge University Press.

Biehler-Gomez, L., Giordano, G., & Cattaneo, C. (2019). The overlooked primary: Bladder cancer metastases on dry bone. A study of the 20th century CAL Milano Cemetery Skeletal Collection. *International Journal of Paleopathology*, *24*, 130–140. https://doi.org/10.1016/j.ijpp.2018.10.005

Brickley, M., & McKinley, J. I. (2004). *Guidelines to the standards for recording human remains*. Southampton: British Association for Biological Anthropology and Osteoarchaeology and Institute of Field Archaeologists.

Bridges, P. S. (1993). The effect of variation in methodology on the outcome of osteoarthritic studies. *International Journal of Osteoarchaeology*, *3*, 289–295. https://doi.org/10.1002/oa.1390030407

Buikstra, J. E., Cook, D. C., & Bolhofner, K. L. (2017). Introduction: Scientific rigor in paleopathology. *International Journal of Paleopathology*, *19*, 80–87. https://doi.org/10.1016/j.ijpp.2017.08.005

Buikstra, J. E., & Ubelaker, D. H. (1994). *Standards for data collection from human skeletal remains*. Arkansas: Arkansas Archaeological Survey Report Number 44: Fayetteville.

Cattaneo, C., Mazzarelli, D., Cappella, A., Castoldi, E., Mattia, M., Poppa, P., … Biehler-Gomez, L. (2018). A modern documented Italian identified skeletal collection of 2127 skeletons: the CAL Milano Cemetery Skeletal Collection. *Forensic Science International*, *287*, 219.e1–219.e5. https://doi.org/10.1016/j.forsciint.2018.03.041

Grauer, A. L. (2008). Macroscopic analysis and data collection in palaeopathology. In R. Pinhasi, & S. Mays (Eds.), *Advances in human*

palaeopathology (pp. 57–76). Chichester, England: Wiley Online Library.

Klaus, H. D. (2017). Paleopathological rigor and differential diagnosis: Case studies involving terminology, description, and diagnostic frameworks for scurvy in skeletal remains. *International Journal of Paleopathology*, *19*, 96–110. https://doi.org/10.1016/j.ijpp.2015.10.002

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. https://doi.org/10.2307/2529310

Lovell, N. C. (2000). Paleopathological description and diagnosis. In S. R. Saunders, & A. M. Katzenberg (Eds.), *Biological anthropology of the human skeleton* (pp. 217–248). New York, USA: Wiley-Liss New York.

Manchester, M., Ogden, A., Storm, R. (2016). *Nomenclature in paleopathology*. Paleopathology Association Newsletter.

Mays, S. (2018). How should we diagnose disease in palaeopathology? Some epistemological considerations. *International Journal of Paleopathology*, *20*, 12–19. https://doi.org/10.1016/j.ijpp.2017.10.006

Miller, E., Ragsdale, B. D., & Ortner, D. J. (1996). Accuracy in dry bone diagnosis: A comment on palaeopathological methods. *International Journal of Osteoarchaeology*, *6*, 221–229. https://doi.org/10.1002/(SICI)1099-1212(199606)6:3<221::AID-OA267>3.0.CO;2-2

Mitchell PD, Brickley M. 2017. Updated guidelines to the standards for recording human remains. Chartered Institute for Archaeologists.

Ortner, D. J. (2003). *Identification of pathological conditions in human skeletal remains* (3rd ed.). New York: Smithsonian Institution Press.

Ortner, D. J. (2011). What skeletons tell us. The story of human paleopathology. *Virchows Archiv*, *459*, 247–254. https://doi.org/10.1007/s00428-011-1122-x

Ortner, D. J. (20120. Differential diagnosis and issues in disease classification. In A. L. Grauer (Ed.). *A companion to paleopathology* (pp. 250–267). Chichester, England: Wiley Online Library. https://doi.org/10.1002/9781444345940.ch14

Ragsdale, B. D., Campbell, R. A., & Kirkpatrick, C. L. (2018). Neoplasm or not? General principles of morphologic analysis of dry bone specimens. *International Journal of Paleopathology*, *21*, 27–40. https://doi.org/10.1016/j.ijpp.2017.02.002

Roberts, C. A., & Connell, B. (2004). Guidance on recording palaeopathology. In *Guidelines to the standards for recording human remains* (Vol. 102) (pp. 34–39). Southampton: British Association for Biological Anthropology and Osteoarchaeology and Institute of Field Archaeologists. https://doi.org/10.1063/1.2756072

Santos, A. L., & Roberts, C. A. (2006). Anatomy of a serial killer: Differential diagnosis of tuberculosis based on rib lesions of adult individuals from the Coimbra identified skeletal collection, Portugal. *American Journal of Physical Anthropology*, *130*, 38–49. https://doi.org/10.1002/ajpa.20160

United Nations. (2004). *Istanbul protocol manual on the effective investigation and documentation of torture and other cruel, inhuman or degrading treatment or punishment*. New York/Geneva: Office of The United Nations High Commissioner for Human Rights.

Waldron, T., & Rogers, J. (1991). Inter-observer variation in coding osteoarthritis in human skeletal remains. *International Journal of Osteoarchaeology*, *1*, 49–56. https://doi.org/10.1002/oa.1390010107

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: Biehler-Gomez L, Indra L, Martino F, Campobasso CP, Cattaneo C. Observer error in bone disease description: A cautionary note. *Int J Osteoarchaeol*. 2020;30:607–615. https://doi.org/10.1002/oa.2885