An analysis of interobserver variability in the recording of maxillary sinusitis in human osteoarchaeological remains

Davies-Barrett, A.M.; Casna, M.; Boyd, D.A.; Inskip, S.A.

Received: 26 September 2023 | Revised: 21 February 2024 | Accepted: 22 February 2024

DOI: 10.1002/oa.3293

SHORT REPORT

An analysis of interobserver variability in the recording of maxillary sinusitis in human osteoarchaeological remains

Anna M. Davies-Barrett[1] | Maia Casna[2] | Derek A. Boyd[3] | Sarah A. Inskip[1]

[1]School of Archaeology and Ancient History, University of Leicester, Leicester, UK

[2]Faculty of Archaeology, Leiden University, Leiden, Netherlands

[3]Department of Social, Cultural, and Justice Studies, The University of Tennessee at Chattanooga, Chattanooga, Tennessee, USA

Correspondence
Anna M. Davies-Barrett, School of Archaeology and Ancient History, University of Leicester, University Road, Leicester LE1 7RH, UK.
Email: a.daviesbarrett@leicester.ac.uk

Funding information
Elements of this research were variously funded by a UKRI-FLF research grant (grant no.: MR/T022302/1), awarded to SI; a Dutch Research Council (NWO) grant (grant no.: PGW.21.008), awarded to MC; and by the Thomas Fellowship and McClure Scholarship from the University of Tennessee, Knoxville, awarded to DB.

Abstract

The accurate comparison of prevalence rates of disease between different groups/populations is vital if we seek to contextualize our understanding of the impact of risk factors on health in the past. The majority of bioarchaeological studies of maxillary sinusitis employ the methods and "diagnostic criteria" outlined by Boocock and colleagues in 1995. However, until now, few attempts have been made to assess the inter-rater reliability of these methods. This study presents the results of the analysis of interobserver variability in the recording of bone changes related to maxillary sinusitis among three observers within three human osteoarchaeological populations. The results of Cohen's kappa coefficient tests indicated variability in agreement between different observers. The agreement on the presence/absence of maxillary sinusitis in different osteoarchaeological populations ranged from $\kappa = 0.433$ ("moderate" agreement) to $\kappa = 0.629$ ("substantial" agreement). The agreement on the type of bone change present within affected sinuses was often poor, with almost no to negative agreement for pitting (Observers 1 and 2) and remodeled spicules (Observers 1 and 3). Methodological problems that can impact consistency of results between researchers include variability in sinus preservation, duration of observer experience, the use of different endoscopic equipment, lack of clarity in the original method descriptions, and a deficit in clinical corroborations for bone changes. The results of this study highlight the need to improve standards for the recording of bone changes related to sinusitis to allow for meaningful comparisons of past maxillary sinusitis frequency rates. Further investigations of interobserver variability, incorporating a greater range of variables, are also required.

KEYWORDS
inter-rater reliability, paleopathology, respiratory disease, upper respiratory tract

1 | INTRODUCTION

The development of maxillary sinusitis—inflammation within the maxillary sinuses—can be predisposed by a range of environmental conditions, including poor air quality, as well as genetic, oral, and infectious diseases, among other risk factors (Brook, 2009; Min & Tan, 2015). In human osteoarchaeological remains, bone changes within the maxillary sinuses (Figure 1) have been suggested as evidence of chronic sinusitis (Boocock et al., 1995; Lewis et al., 1995). Calculating past frequency rates of sinusitis can be useful in understanding differential

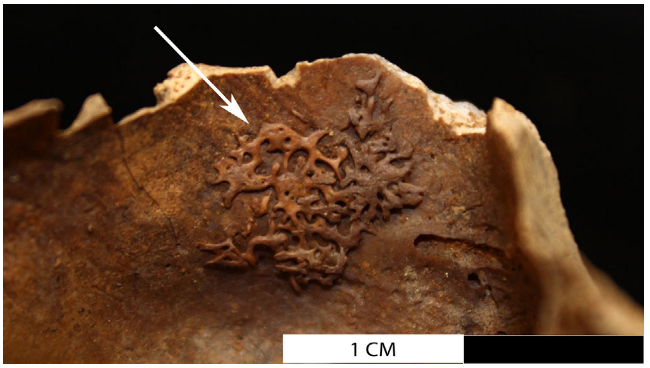

FIGURE 1 Example of bone changes within the maxillary sinus thought to indicate inflammation related to maxillary sinusitis. A patch of spicules is present on the lateral wall of the sinus (St James's Gardens Burial Ground, Sk155211, left maxillary sinus). [Colour figure can be viewed at wileyonlinelibrary.com]

exposure of past populations to poor air quality and other risk factors. Thus, the recording of bone changes within the maxillary sinuses as evidence of sinusitis has become a popular avenue of research in recent years (e.g., Boyd, 2020; Casna et al., 2021; Davies-Barrett, Owens, & Eeckhout, 2021; Davies-Barrett, Roberts, & Antoine, 2021; Riccomi et al., 2021; Zubova et al., 2022). It is common for population prevalence studies to compare results with those produced by other researchers or to incorporate data collected in previous studies (e.g., Davies-Barrett, Roberts, & Antoine, 2021; Riccomi et al., 2021; Roberts, 2007; Zubova et al., 2022). The accurate comparison of prevalence rates of sinusitis within different populations is vital if we seek to further contextualize our understanding of the impact of different risk factors on respiratory health in the past. However, while clinical studies have demonstrated the applicability of interobserver testing in the diagnosis of maxillary sinusitis (e.g., Malina-Altzinger et al., 2015; Timmenga et al., 2002), as of yet, few attempts have been made to investigate the inter-comparability of bioarchaeological results,

Although bone changes in the sinuses of archaeological individuals were noted as early as the 1960s (Wells, 1964), a pivotal moment in the development of the identification of sinusitis occurred with the publication of "diagnostic criteria" by Boocock et al. (1995). This publication outlined four major types of bone change within the sinus: spicules, remodeled spicules, white pitted bone, and pitting. Since its publication, the majority of sinusitis studies have employed these criteria, with slight modifications. However, issues with the original methods have been raised, which revolve around the lack of clarity in the terminology and original descriptions, and a deficit in clinical research that can be conclusively linked to the bone changes observed in archaeological human remains (Davies-Barrett, Owens, & Eeckhout, 2021; Davies-Barrett, Roberts, & Antoine, 2021; Sundman & Kjellström, 2013). Further, high-quality photos corresponding to detailed descriptions of the pathological changes considered to represent maxillary sinusitis are often lacking in bioarchaeological studies, making accurate inter-study comparisons difficult.

Employing the original criteria presented by Boocock et al. (1995), three observers with variable levels of experience analyzed maxillary sinuses from three osteological collections. We present here the results of the analysis of interobserver variability in the recording of bone changes related to sinusitis within these three populations, and the results of the analysis of intraobserver error in one population, to assess the reliability of this method.

2 | MATERIALS AND METHODS

A total of 200 individuals were included in the study: 64 individuals (119 Sinuses) from Keyserkerk Cemetery, Middenbeemster, Netherlands, and 97 individuals (169 sinuses) from St James's Gardens Burial Ground, London, UK, were analyzed by Observers 1 and 2; 39 individuals (73 sinuses) from St Peter's Church, Barton-upon-Humber, UK, were analyzed by Observers 1 and 3 (Table 1). Each individual was assessed for evidence of bone changes within the maxillary sinuses. All sinuses recorded as over 25% complete by one or both of the observers were included within the sample. Individuals with two complete sinuses that were unobservable macroscopically and had no endoscopic access point in at least one sinus were not included in the sample.

Maxillary sinusitis was recorded as present if one or more of the diagnostic criteria presented by Boocock et al. (1995) were observed. If bone changes were observed within a maxillary sinus, that sinus was categorized as "present" for sinusitis by the observer. If one or both sinuses were observed to have bone changes, that individual was categorized as "present" for sinusitis by the observer. The type(s) of bone change observed were also noted according to the diagnostic criteria. Multiple types of bone change could be recorded as present within a single sinus/individual. As the aim of the study was to investigate the inter-reliability of results using individual researchers' interpretations of the approach outlined by Boocock et al. (1995), the methods used were not discussed between observers in detail prior to undertaking the study. All observations were conducted by each observer independently at different points in time.

Endoscopy of the sinus was applied to individuals from the St James's Gardens Burial Ground and Middenbeemster populations in instances where completeness of the cranium prohibited macroscopic observation of the internal sinus surfaces. The same medical-grade endoscopic equipment, a Karl Storz digital tip flexible endoscope and Karl Storz Tele-pak X for image visualization on the St James's population and a Pentax endoscope (model: FNL-10RBS) on the Middenbeemster population, were used by both observers. The sinuses were accessed via natural openings or the taphonomically damaged medial wall of the sinus, located within the nasal cavity. Too few individuals from Barton-upon-Humber were analyzed using endoscopic equipment to statistically investigate the effect of its use further. Other equipment used by observers was limited to a hand lens (×10 magnification) and lighting aids.

To measure how interobserver tests may be affected by changes in researcher experience, an intraobserver test was also undertaken

TABLE 1 Number of individuals observed from each skeletal population, approximate time period in which observations were undertaken, and approximate level of experience at point of observation.

	No. of individuals observed	Approximate time period when observed			Approximate experience at point of observation[a]		
		Observer 1	Observer 2	Observer 3	Observer 1	Observer 2	Observer 3
St James's gardens burial ground, UK	97	Spring–Summer 2022	Summer 2022	-	8 years	3 years	-
Middenbeemster, Netherlands	64	Macroscopic: Spring 2022; Endoscopic: Winter 2023	Summer–Fall 2020	-	8–9 years	1 year	-
Barton-upon-Humber, UK	39	Summer 2021, Summer 2022	-	Summer 2021	7–8 years	-	2 years

[a]Experience denotes the approximate total length of time spent specializing in the osteoarchaeological study of respiratory diseases, including use of the Boocock et al. (1995) method and use of endoscopy.

by both Observers 1 and 2 on 23 individuals with at least one macroscopically observable maxillary sinus from Middenbeemster. Time elapsed between observations was 1.5 years for Observer 1 and 3 years for Observer 2. A Cohen's kappa coefficient test to measure the level of agreement between observers was applied to the data using SPSS v.28. The 95% confidence interval (CI) was calculated using the standard error value. The kappa statistic (κ) is used as a measure of inter- or intra-rater reliability and its value can range from −1 to +1. Typically, a value of 0.00–0.20 indicates "no to slight" agreement, 0.21–0.40 as "fair" agreement, 0.41–0.60 as "moderate" agreement, 0.61–0.80 as "substantial" agreement, and >0.81 as "almost perfect" agreement (McHugh, 2012).

3 | RESULTS

Results indicate an overall "moderate" to "substantial" level of agreement in the observation of maxillary sinusitis between observers (Table 2, Figure 2a). Observations of presence/absence in individuals in the Middenbeemster group presented with a lower level of interobserver agreement ($\kappa = 0.433$) than observations in the St James's ($\kappa = 0.592$) and Barton-upon-Humber groups ($\kappa = 0.629$). Observations using endoscopic and macroscopic methods had a similar level of "moderate" agreement within the St James's group but differed in the Middenbeemster population (endoscopic: $\kappa = 0.488$; macroscopic: $\kappa = 0.329$) (Figure 2b).

Agreement in the type of bone change observed within affected sinuses was highly variable, ranging from "substantial" to "negative." Notably, "almost no" to "negative" agreement was observed for the bone change type "pitting" between Observers 1 and 2 within both the St James's ($\kappa = 0.079$) and Middenbeemster ($\kappa = -0.167$) populations. A similar outcome was observed for "remodeled spicules" between Observers 1 and 3 ($\kappa = 0.079$) (Figure 2c). Inter-rater variability in assessment of individual sinuses can be found in Table S1.

In the intra-rater test, Observer 1 presented with a "substantial" agreement in both observations of maxillary sinusitis by individual and by sinus, whereas Observer 2 presented with "fair" ($\kappa = 0.392$) agreement by individual and "moderate" ($\kappa = 0.418$) agreement by sinus (Table 3; Figure 2d). It should be noted that large 95% confidence intervals were present across the majority of statistical tests conducted in this study due to small sample sizes and data variability, meaning kappa value ranges often spanned several different agreement "classes."

4 | DISCUSSION AND CONCLUSIONS

Within palaeopathology, it is well recognized that inter- and intraobserver error tests are a vital part of the observation and diagnosis of pathological lesions; yet these tests are not often undertaken, particularly when such studies use ordinal or nominal data (Mays, 2022; Roberts, 2018). However, the results of the current study highlight why such tests are necessary. The analysis of interobserver variability of evidence for maxillary sinusitis generally presented with a low "moderate" to low "substantial" level of agreement according to kappa coefficients. Overall raw percentage agreement ranged from 71.9% at Middenbeemster to 82.1% at Barton-upon-Humber (Table 2). While this does indicate that the majority of observations were corroborated between observers, approximately 20%–30% of results were unreliable. Additionally, the kappa statistic takes into account the possibility that some agreement may have occurred by random chance. This corrects the percentage of agreement, so that a kappa value of between 0.4 and 0.6 represents a 40%–60% chance-adjusted agreement (McHugh, 2012). Further, large 95% confidence intervals were present due to limited sample sizes and variability in data. Therefore, while the kappa value may have fallen into the "moderate" or "substantial" agreement category, lower confidence intervals may have ranged into the "fair" or "no to slight" agreement categories. These results have serious implications for the comparability of

TABLE 2 An analysis of inter-rater agreement for the presence or absence of maxillary sinusitis according to population, method of observation, and type of bone change.

						Observer 1		Measure of Agreement			
				N		Present *n* (%)	Absent *n* (%)	Raw percentage	Kappa value (κ)	Standard error	95% confidence interval (CI)
Observer 2	St James's gardens burial ground	All individuals		97	Present	40 (41.2)	5 (5.2)	79.4	0.592	0.080	0.435–0.749
					Absent	15 (15.5)	37 (38.1)				
		Observation method	Endoscopic[a]	45	Present	16 (35.6)	2 (4.4)	75.6	0.522	0.120	0.287–0.757
					Absent	9 (20.0)	18 (40.0)				
			Macroscopic[b]	62	Present	26 (41.9)	5 (8.1)	77.4	0.548	0.105	0.342–0.754
					Absent	9 (14.5)	22 (35.5)				
		Type of bone change	Spicules	40	Present	23 (57.5)	8 (20.0)	77.5	0.494	0.135	0.229–0.759
					Absent	1 (2.5)	8 (20.0)				
			Remodeled spicules	40	Present	8 (20.0)	4 (10.0)	70.0	0.348	0.150	0.054–0.642
					Absent	8 (20.0)	20 (50.0)				
			White pitted bone	40	Present	7 (17.5)	2 (5.0)	65.0	0.280	0.132	0.021–0.539
					Absent	12 (30.0)	19 (47.5)				
			Pitting	40	Present	2 (5.0)	12 (30.0)	65.0	0.079	0.128	−0.172–0.330
					Absent	2 (5.0)	24 (60.0)				
	Middenbeemster	All individuals		64	Present	20 (31.3)	9 (14.1)	71.9	0.433	0.113	.212–.654
					Absent	9 (14.1)	26 (40.6)				
		Observation method	Endoscopic[a]	40	Present	12 (30.0)	5 (12.5)	75.0	0.488	0.140	0.214–0.762
					Absent	5 (12.5)	18 (45.0)				
			Macroscopic[b]	21	Present	6 (28.6)	3 (14.3)	66.7	0.329	0.206	−0.075–0.733
					Absent	4 (19.0)	8 (28.1)				
		Type of bone change	Spicules	20	Present	10 (50.0)	5 (25.0)	70.0	0.368	0.194	−0.012–0.748
					Absent	1 (5.0)	4 (20.0)				
			Remodeled spicules	20	Present	5 (25.0)	2 (10.0)	75.0	0.468	0.203	0.070–0.866
					Absent	3 (15.0)	10 (50.0)				
			White pitted bone	20	Present	5 (25.0)	0 (0.0)	85.0	0.667	0.167	0.340–0.994
					Absent	3 (15.0)	12 (60.0)				
			Pitting	20	Present	0 (0.0)	5 (25.0)	65.0	−0.167	0.091	−0.345–0.011
					Absent	2 (10.0)	13 (65.0)				

TABLE 2 (Continued)

					Observer 1		Measure of Agreement			
			N		Present *n* (%)	Absent *n* (%)	Raw percentage	Kappa value (*k*)	Standard error	95% confidence interval (CI)
Observer 3	Barton-upon-Humber	All individuals	39	Present	21 (53.8)	0 (0)	82.1	0.629	0.118	0.398–0.860
				Absent	7 (17.9)	11 (28.2)				
		Type of bone change: Spicules	21	Present	8 (38.1)	1 (4.8)	76.2	0.533	0.175	0.190–0.876
				Absent	4 (19.0)	8 (38.1)				
		Remodeled spicules	21	Present	6 (28.6)	7 (33.3)	52.4	0.079	0.201	−0.315–0.473
				Absent	3 (14.3)	5 (23.8)				
		White pitted bone	21	Present	2 (9.5)	0 (0)	66.7	0.246	0.153	−0.054–0.546
				Absent	7 (33.3)	12 (57.1)				
		Pitting	21	Present	3 (14.3)	3 (14.3)	81.0	0.481	0.217	0.056–0.906
				Absent	1 (4.8)	14 (66.7)				

[a]In which both observers employed endoscopy to observe one or both maxillary sinuses.
[b]In which both observers did not employ endoscopy to observe either maxillary sinus.

bioarchaeological studies of sinusitis and indicate that standards are needed to improve consistency in the application of the Boocock et al. (1995) method. The accurate observation of bone changes within the maxillary sinuses is limited by several methodological and practical problems that can impact on consistency of results between researchers. This includes issues surrounding preservation of sinuses and use of endoscopic equipment, level of experience of the observer, lack of clarity in the original method descriptions, and a deficit in clinical corroborations for bone changes.

Firstly, sinus preservation impacts the ability to accurately observe bone changes. Highly incomplete or fragmented sinuses may be difficult to observe for evidence of sinusitis and damage may also strip the sinus of, or obscure, delicate bone changes. Results for macroscopic observation of the Middenbeemster population presented with only a "fair" level of agreement. This sample group was selected for individuals with sinuses that were fully observable without the use of an endoscope, and it was noted that open sinuses from Middenbeemster were in a particularly fragmented and incomplete state. This may have inhibited accurate observation and led to interobserver error. Sundman and Kjellström (2013) investigated the effect of preservation bias on the recording of maxillary sinusitis by grouping sinuses into categories of approximate completeness. They found a significant positive correlation between sinus preservation and observations of sinusitis. Further, they found that more extensive bone changes were observed in more complete sinuses, making it more likely that observers will note bone changes in better preserved sinuses.

Conversely, if the sinuses are highly complete, endoscopy is required for their visualization. It might be expected that agreement of results using an endoscope would be lower due to the inherent difficulties of accessing and fully visualizing the sinuses using such equipment. However, the results of the current study indicate that the use of the same endoscopic equipment by both researchers results in "moderate" agreement. Therefore, the use of endoscopy (if the same medical-grade equipment is employed) does not appear to negatively affect agreement to a greater degree than macroscopic observations. This result was reflected in the results from the St James's sample, which presented with similar "moderate" agreement using both endoscopic and macroscopic methods. However, different types of endoscopic equipment vary greatly, particularly in terms of picture quality and functionality of the scope, often relying on the budget of the researcher. The employment of different endoscopes and the level of experience/training in their use by different researchers could introduce agreement error.

Both intra- and interobserver results from Middenbeemster indicate that duration of experience may have an impact on the accurate observation of bone changes related to sinusitis. Observer 2 assessed individuals from Middenbeemster 2 years prior to their later observations in this study (see Tables 1 and 3) at an early stage of their research career, while Observer 1 had 8–9 years of continuous experience in the analysis of sinusitis at the point of observation. This may be reflected in the less accurate inter-rater results from Middenbeemster, when compared with other sites.

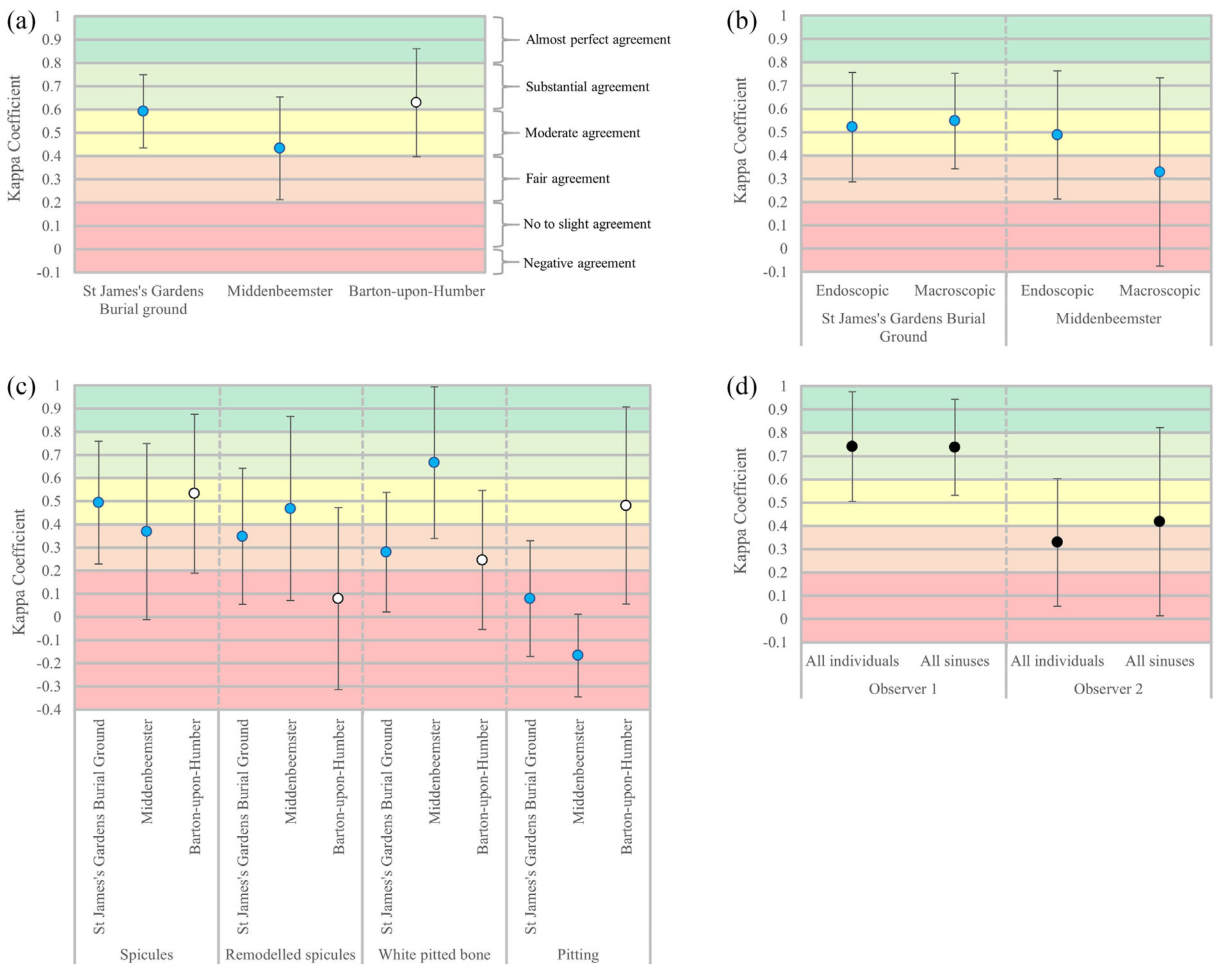

FIGURE 2 Kappa coefficient values (blue, white, and black markers) and positive and negative 95% confidence intervals (black bars) by (a) Interobserver - site, (b) Interobserver - method of observation, (c) Interobserver - type of bone change, (d) Intraobserver - Middenbeemster. Blue marker denotes tests between Observers 1 and 2; white marker denotes tests between Observers 1 and 3; black marker denotes intraobserver tests. [Colour figure can be viewed at wileyonlinelibrary.com]

Further, intraobserver results demonstrate that while the more experienced observer with a smaller duration in time between observations (Observer 1) showed high "substantial" agreement, the less experienced observer with a greater duration of time between observations (Observer 2) demonstrated only a "fair" level of agreement. This likely reflects the fact that, as a researcher develops, they become more confident in the use of methods and in the recognition of certain types of bone changes (whether or not the method is accurately applied). A researcher may also make conscious or subconscious adjustments to the application of this method over extended periods of time. These should be important considerations for any researcher new to using the methods of Boocock et al. (1995) to identify maxillary sinusitis.

Other inconsistencies between researchers may be due to the fact that no clear guidance is given by Boocock et al. (1995) on the extent of bone changes required to determine sinusitis as present. Given that there can be a natural level of minor porosity and irregular surface texture within the maxillary sinuses (Boocock et al., 1995; Lewis et al., 1995; Sundman & Kjellström, 2013), it is ultimately dependent upon the individual researcher to decide whether or not bone changes are extensive enough to represent chronic inflammation. The severity scoring system implemented by Sundman and Kjellström (2013) may help to discriminate between minor sinus surfaces irregularities and more extensive changes caused by chronic inflammation, but adoption of this approach by other researchers has been slow to gain traction due to the clinical implications of the term "severity."

A particularly low level of agreement was observed for the type of bone change recorded as present, with almost no agreement for pitting between Observers 1 and 2 within the St James's population ($\kappa = 0.079$) and remodeled spicules between Observers 1 and 3. There was also a negative agreement between Observers 1 and 2 for pitting within the Middenbeemster group, indicating that these observers may have been consistently allocating the same morphological bone changes into two different bone type categories. It is evident that variability in the recording of bone change type can differ greatly between sets of observers, as noted by Biehler-Gomez et al. (2020) in a study of inter-rater accuracy between lesion descriptions by

TABLE 3 An analysis of intra-rater agreement for the presence or absence of maxillary sinusitis in the Middenbeemster population (macroscopic observation methods only).

						Measure of agreement			
		N		Present n (%)	Absent n (%)	Raw percentage	Kappa value (κ)	Standard error	95% confidence interval (CI)
Observer 1 First observation: Spring 2022–7 years of experience[a] Second observation: Winter 2023–8 years of experience[a]	All individuals	23	Present	10 (43.5)	1 (4.3)	87.0	0.740	0.140	0.466–1.014
			Absent	2 (8.7)	10 (43.5)				
	All sinuses	38	Present	16 (42.1)	3 (7.9)	86.8	0.737	0.110	0.521–0.953
			Absent	2 (5.3)	17 (44.7)				
Observer 2 First observation: Summer–Fall 2020–1 year of experience[a] Second observation: Winter 2023–4 years of experience[a]	All individuals	23	Present	8 (34.8)	3 (13.0)	69.6	0.392	0.191	0.018–0.766
			Absent	4 (17.4)	8 (34.8)				
	All sinuses	42	Present	12 (28.6)	5 (11.9)	71.4	0.418	0.141	0.142–0.694
			Absent	7 (17.7)	18 (42.9)				

[a]Experience denotes the approximate total length of time spent specializing in the osteoarchaeological study of respiratory diseases, including use of the Boocock et al. (1995) method.

anthropologists and forensic pathologists. This study found that disparities in descriptions occurred independently of the level of experience of the observers and may have been impacted by differences in the extent of training in producing standardized pathological descriptions. The results from the current study highlight an inherent problem with individual interpretations of the diagnostic criteria presented by Boocock et al. (1995). For example, the categories of "pitting" and "white pitted bone" can cause confusion due to similarity in the terminology of "pitting"/"pitted." Additionally, the use of the term "white" does not take into account the range of taphonomic processes that can lead to discoloration of various types within the sinuses, as well as differences in color visualization using an endoscope. For this reason, some researchers have begun to refer to "white pitted bone" as "porous new bone" instead (Davies-Barrett, Owens, & Eeckhout, 2021; Davies-Barrett, Roberts, & Antoine, 2021). Further, distinguishing between spicules and remodeled spicules can be difficult when the extent of remodeling is only subtle (Lee et al., 2024).

Scepticism about "pitting," as originally described by Boocock et al. (1995), and its relationship to maxillary sinusitis has been informally raised among researchers. The roots of the maxillary molars are located directly beneath the sinus, separated by a paper-thin layer of bone which can often be naturally perforated by the apex of the tooth roots (Brook, 2006). Antemortem movement or loss of the dentition is likely to result in remodeling of the alveolar bone and the sinus floor. Pitting or remodeling of the sinus floor may be linked to oral diseases (Hillson, 1996, pp. 284–286; Waldron, 2021, pp. 313–314) rather than chronic sinus inflammation. While this relationship has yet to be fully explored, doubts about the clinical interpretation of different types of bone change may subconsciously affect the choice of the observer when it comes to bone change type category allocation.

It is apparent that the original criteria have left room for departures in the ways that researchers typify bone changes. Although not addressed in the current study, many researchers have also presented adjustments to the original diagnostic criteria. Boocock et al. (1995, p. 486) originally discussed the presence of "thickened and porous" walls with "lobules of white bone" but did not explicitly include this within their criteria. Merrett and Pfeiffer (2000) included additional categories of "plaque," "lobules; and "cysts," while other researchers have included a category of "other," which includes thickening of the walls and lobules (Casna et al., 2021; Davies-Barrett, Owens, & Eeckhout, 2021; Sundman & Kjellström, 2013). Given the paucity in clinical data related to each type of bone change, it may be pertinent to ask whether it is necessary to distinguish between different types of bone change or to ask which bone changes can be conclusively linked to maxillary sinusitis. Although some researchers have noted that different types of bone changes may represent different stages of chronic inflammation (Boocock et al., 1995; Collins, 2019; Lee et al., 2024), it is evident that different pathophysiological processes are occurring to produce different morphological bone change types. Whether or not only some of these processes can be linked to sinusitis requires further clinical confirmation. Unfortunately, this topic is clinically underexplored due to the lack of diagnostic relevance that bone changes within the sinuses have to clinicians (Orlandi et al., 2016). If we were, however, to find that certain bone change types may be linked to specific disease processes or were unrelated to sinus inflammation, accurate recording of the type of bone changes would be required to retrospectively investigate the impact that this finding has on our calculation of prevalence rates.

4.1 | Limitations of the study and future approaches

This study presents only a preliminary exploration of interobserver variability in the recording of maxillary sinusitis. It was limited by inconsistency in access to collections and equipment, leading to gaps in time in the observations of a single sample. Further, the intraobserver error test included only a reduced sample size conducted by Observers 1 and 2 only, due to access, equipment, and time constraints. These limitations reflect the general nature of osteological analysis, which is reliant on the often-restricted availability of access to equipment and collections. Additionally, large confidence interval ranges were present in kappa value calculations, indicating that future studies with larger sample sizes are needed to further explore this topic. There is still much to investigate, including additional analysis of the effects of sinus completeness, preservation, and taphonomic alterations on reliability of observations, different adaptations to the methods, the use of different types of endoscopic equipment, the inclusion of a greater number of observers, and further investigation into the effects of variable levels of observer experience on intra- and inter-rater reliability.

There will always be an inherent level of independent interpretation among researchers of methods which observe variable bone changes. However, attempts can be made to decrease interobserver variability. To improve consistency of recording, concerted efforts are required by researchers to openly discuss approaches employed and publish with greater clarity the methods they use. It is evident that, while the diagnostic criteria presented by Boocock et al. (1995) have provided an invaluable foundation for the identification of chronic sinus inflammation within the past, these methods require further standardization. Greater consideration of up-to-date standardized palaeopathological terminology (such as those recommended by Appleby et al. (2015) and Manchester et al. (2016)) are required in future method developments. Given the increasing number of studies that incorporate and compare prevalence rates of sinusitis, a revision of the methods, including intra- and interobserver testing of any new developments, is long overdue.

ACKNOWLEDGMENTS

Our heartfelt acknowledgements go to Raguwinder (Bindy) Sahota (University of Leicester) and Karl Stortz; the Forensic Anthropology Center (University of Tennessee at Knoxville); Sarah Schrader and Rachel Schats (Leiden University); the Middenbeemster Historical Society and the community of Middenbeemster; Simon Mays (Historic England); Kevin Booth (English Heritage); Andrea Bradley and Hayley James (HS2); and Michael Henderson and Louise Fowler (MOLA).

CONFLICT OF INTEREST STATEMENT

The authors report there are no conflicts of interests to declare.

DATA AVAILABILITY STATEMENT

Raw data pertaining to this study can be requested by contacting the corresponding author.

ORCID

Anna M. Davies-Barrett https://orcid.org/0000-0002-1555-1468
Maia Casna https://orcid.org/0000-0002-4209-7889
Derek A. Boyd https://orcid.org/0000-0003-1444-0536

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

Biehler-Gomez, L., Indra, L., Martino, F., Campobasso, C. P., & Cattaneo, C. (2020). Observer error in bone disease description: A cautionary note. *International Journal of Osteoarchaeology*, *30*(5), 607–615. https://doi.org/10.1002/oa.2885

Boocock, P. A., Roberts, C. A., & Manchester, K. (1995). Maxillary sinusitis in Medieval Chichester, England. *American Journal of Physical Anthropology*, *98*(4), Article 4–Article 495. https://doi.org/10.1002/ajpa.1330980408

Boyd, D. A. (2020). Respiratory stress at the periphery of industrial-era London: Insight from parishes within and outside the city. In T. K. Betsinger & S. N. DeWitte (Eds.), *The bioarchaeology of urbanization: The biological, demographic, and social consequences of living in cities* (pp. 379–402). Springer International Publishing. https://doi.org/10.1007/978-3-030-53417-2_15

Brook, I. (2006). Sinusitis of odontogenic origin. *Otolaryngology–Head and Neck Surgery*, *135*(3), 349–355. https://doi.org/10.1016/j.otohns.2005.10.059

Brook, I. (2009). Sinusitis. *Periodontology 2000*, *49*(1), 126–139. https://doi.org/10.1111/j.1600-0757.2008.00293.x

Casna, M., Burrell, C. L., Schats, R., Hoogland, M. L. P., & Schrader, S. A. (2021). Urbanization and respiratory stress in the northern Low Countries: A comparative study of chronic maxillary sinusitis in two early modern sites from the Netherlands (AD 1626–1866). *International Journal of Osteoarchaeology*, *31*(5), 891–901. https://doi.org/10.1002/oa.3006

Collins, C. (2019). *The palaeopathology of maxillary sinusitis, otitis media and mastoiditis in Medieval Iceland: Assessing the prevalence and aetiology of chronic upper respiratory disease and the presence of tuberculosis using microscopy, endoscopy and CT* [Phd, University of Reading]. http://centaur.reading.ac.uk/84764/

Davies-Barrett, A. M., Owens, L. S., & Eeckhout, P. A. (2021). Paleopathology of the Ychsma: Evidence of respiratory disease during the Late Intermediate Period (AD 1000-1476) at the Central Coastal site of Pachacamac, Peru. *International Journal of Paleopathology*, *34*, 63–75. https://doi.org/10.1016/j.ijpp.2021.06.001

Davies-Barrett, A. M., Roberts, C. A., & Antoine, D. (2021). Time to be nosy: Evaluating the impact of environmental and sociocultural changes on maxillary sinusitis in the Middle Nile Valley (Neolithic to Medieval periods). *International Journal of Paleopathology*, *34*, 182–196. https://doi.org/10.1016/j.ijpp.2021.07.004

Hillson, S. (1996). *Dental anthropology*. Cambridge University Press. https://doi.org/10.1017/CBO9781139170697

Lee, M. J., Siek, T. J., & Hirst, C. S. (2024). Chronic maxillary sinusitis in palaeopathology: A review of methods. *International Journal of Paleopathology*, *44*, 51–64. https://doi.org/10.1016/j.ijpp.2023.11.005

Lewis, M. E., Roberts, C. A., & Manchester, K. (1995). Comparative study of the prevalence of maxillary sinusitis in later Medieval urban and rural populations in Northern England. *American Journal of Physical Anthropology*, *98*(4), 497–506. https://doi.org/10.1002/ajpa.1330980409

Malina-Altzinger, J., Damerau, G., Grätz, K. W., & Stadlinger, P. B. (2015). Evaluation of the maxillary sinus in panoramic radiography—A comparative study. *International Journal of Implant Dentistry*, *1*(1), 17. https://doi.org/10.1186/s40729-015-0015-1

Manchester, K., Ogden, A., & Storm, R. (2016). Nomenclature in palaeopathology. *Palaeopathology Newsletter*, *175*. https://www.paleopathology.org/resources/Documents/PPA%20Monographs/Nomenclature%20in%20Palaeopathology%20Web%20Document.pdf

Mays, S. (2022). The macroscopic study of human skeletal paleopathology. In *The Routledge handbook of paleopathology*. Routledge. https://doi.org/10.4324/9781003130994-3

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282. https://doi.org/10.11613/BM.2012.031

Merrett, D. C., & Pfeiffer, S. (2000). Maxillary sinusitis as an indicator of respiratory health in past populations. *American Journal of Physical Anthropology*, *111*(3), 3. https://doi.org/10.1002/(SICI)1096-8644(200003)111:3<301::AID-AJPA2>3.0.CO;2-0

Min, J.-Y., & Tan, B. K. (2015). Risk factors for chronic rhinosinusitis. *Current Opinion in Allergy and Clinical Immunology*, *15*(1), 1–13. https://doi.org/10.1097/ACI.0000000000000128

Orlandi, R. R., Kingdom, T. T., Hwang, P. H., Smith, T. L., Alt, J. A., Baroody, F. M., Batra, P. S., Bernal-Sprekelsen, M., Bhattacharyya, N., Chandra, R. K., Chiu, A., Citardi, M. J., Cohen, N. A., DelGaudio, J., Desrosiers, M., Dhong, H.-J., Douglas, R., Ferguson, B., Fokkens, W. J., … Kennedy, D. W. (2016). International consensus statement on allergy and rhinology: Rhinosinusitis. *International Forum of Allergy & Rhinology*, *6*(S1), S22–S209. https://doi.org/10.1002/alr.21695

Riccomi, G., Casaccia, J., Minozzi, S., Felici, C., Campana, S., & Giuffra, V. (2021). Maxillary sinusitis as a respiratory health indicator: A bioarchaeological investigation into medieval central Italy. *International Journal of Paleopathology*, *35*, 40–48. https://doi.org/10.1016/j.ijpp.2021.09.001

Roberts, C. A. (2007). A bioarcheological study of maxillary sinusitis. *American Journal of Physical Anthropology*, *133*(2), 792–807. https://doi.org/10.1002/ajpa.20601

Roberts, C. A. (2018). Guidance on recording palaeopathology (abnormal variation). In P. D. Mitchell & M. B. Brickley (Eds.), *Updated guidelines to the standards for recording human remains* (pp. 44–47). CIfA Chartered Institute for Archaeologists/British Association for Biological Anthropology and Osteoarchaeology. https://www.archaeologists.net/sites/default/files/14_Updated%20Guidelines%20to%20the%20Standards%20for%20Recording%20Human%20Remains%20digital_0.pdf

Sundman, E. A., & Kjellström, A. (2013). Chronic maxillary sinusitis in Medieval Sigtuna, Sweden: A study of sinus health and effects on bone preservation. *International Journal of Osteoarchaeology*, *23*(4), 447–458. https://doi.org/10.1002/oa.1268

Timmenga, N., Stegenga, B., Raghoebar, G., van Hoogstraten, J., van Weissenbruch, R., & Vissink, A. (2002). The value of waters' projection for assessing maxillary sinus inflammatory disease. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, *93*(1), 103–109. https://doi.org/10.1067/moe.2002.120056

Waldron, T. (2021). *Palaeopathology* (2nd ed.). Cambridge University Press.

Wells, C. (1964). Chronic sinusitis with alveolar fistulae of Mediæval date. *The Journal of Laryngology & Otology*, *78*(3), 320–322. https://doi.org/10.1017/S0022215100062137

Zubova, A. V., Moiseyev, V. G., Ananyeva, N. I., Stulov, I. K., & Andreev, E. V. (2022). Chronic maxillary sinusitis recorded in archaeological samples: Geographical distribution and predisposing factors. *Archaeology, Ethnology and Anthropology of Eurasia*, *50*(1), 1. https://doi.org/10.17746/1563-0110.2022.50.1.147-157

SUPPORTING INFORMATION

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

How to cite this article: Davies-Barrett, A. M., Casna, M., Boyd, D. A., & Inskip, S. A. (2024). An analysis of interobserver variability in the recording of maxillary sinusitis in human osteoarchaeological remains. *International Journal of Osteoarchaeology*, e3293. https://doi.org/10.1002/oa.3293