Journal of Archaeological Science 136 (2021) 105300

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: http://www.elsevier.com/locate/jas

Wear of Teeth in Sheep (WoTiS) - A tool for determining the rate of mandibular tooth wear in sheep

Ursula R. Mutze [a,*], Ulrich Mutze [b], Gillian G. Jones [c], Joris Peters [a,d]

[a] Ludwig-Maximilians-Universität München, Tiermedizinische Fakultät, Institut für Paläoanatomie, Domestikationsforschung und Geschichte der Tiermedizin, Kaulbachstraße 37, D-80539, München, Germany

[b] Am Bahndamm 22/1, D-73342, Bad Ditzenbach, Germany

[c] 3 Church Farm Garth Shadwell, Leeds, LS17 8HD, UK

[d] SNSB, Staatssammlung für Anthropologie und Paläoanatomie München, Karolinenplatz 2a, D-80333, München, Germany

ARTICLE INFO

Keywords:
Sheep
Tooth wear rate
Dental ageing
Mandibles
Egypt
Animal necropolis

ABSTRACT

Demographic profiling of a population of sheep buried in toto in the Ptolemaic-Early Roman animal necropolis at Syene/Aswan (Upper Egypt) revealed significantly higher age estimates based on tooth eruption and wear than those based on epiphyseal fusion. Since located in an arid landscape with occasionally heavy dust loads, one plausible assumption would be that at Syene, dental abrasion may have been exceptionally intense. Until now, however, no approach exists to determine the intrinsic rate of tooth wear in archaeological sheep specimens and its effects on demographic profiles. Using occlusal wear patterns observed in mandibular dental rows of large modern sheep populations of known age, we developed a novel method to determine the rate of tooth wear of sheep and estimate the corresponding age at death of the animals. To facilitate such time-consuming analysis, we then developed an IT-tool termed WoTiS (Wear of Teeth in Sheep) that has been coded in R. The paper outlines the principles underlying our approach and guides the reader through our method by way of examples. Discussion focuses on the application of our approach to sheep populations from Celtic Manching (southern Germany) and Late Dynastic-Mamluk Syene (Upper Egypt). Our results underscore the necessity of evaluating the rate of tooth wear in sheep in order to reach valid statements regarding the species' mode of exploitation in ancient cultures.

1. Introduction

In zooarchaeological studies, demographic profiling is essential for understanding human exploitation of livestock species in ancient cultures. Correspondingly, methods for recording and analysing tooth wear have been developed for key farm animals (e.g. Payne 1973, 1987; Grant 1975, 1982), and for estimating age at death there are two large studies, for sheep (Jones 2006) and goats (Deniz and Payne 1979, 1982). While the intensity of tooth wear is known to vary significantly due to different environmental conditions (e.g. Healy and Ludwig 1965; Grant 1978; Twiss 2008; Damuth and Janis 2011; Jones and Sadler 2012a, b), the timing of cheek tooth eruption was found to vary only slightly. Consequently, tooth eruption provides comparatively reliable age estimates for young sheep and goats, while ageing becomes problematic once all permanent teeth have come into wear. The two most frequently applied methods for dental ageing of caprines proposed by Payne (1973) and Grant (1975, 1982) are based on occlusal wear patterns and use different approaches to address this problem: Grant (1982) avoids proposing absolute ages while Payne (1973) suggests absolute ages for the consecutive wear stages based on Silver (1969) and his own observations. His approach was later refined by using a modern control sample of Turkish Angora Goats (Deniz and Payne 1982). For sheep, a similar study was undertaken by Jones (2006). It broadly confirmed the ages proposed by Payne (1973) but refined the consecutive stages, thereby allowing for more detailed analysis. Greenfield and Arnold (2008) also found a good correlation for a modern control sample of sheep and goats from Manitoba, Canada. Available studies thus indicate a broad applicability of Payne's ageing system and the refined system proposed by Jones (2006). Zeder (2006) examined the correlation between age estimates based on dental wear and epiphyseal fusion in wild sheep and goat specimens from Iran and Iraq. She found a good correlation between both, with dental ages in sheep tending to be a little lower than

* Corresponding author.
E-mail address: ursula.mutze@palaeo.vetmed.uni-muenchen.de (U.R. Mutze).

https://doi.org/10.1016/j.jas.2020.105300
Received 13 May 2020; Received in revised form 8 November 2020; Accepted 19 November 2020
Available online 15 March 2021
0305-4403/

epiphyseal ages, especially in females.

We recorded wear stages (after Payne 1987) and epipihyseal fusion data in a sizeable population of sheep buried in toto in the Ptolemaic to Early Roman animal necropolis at Syene (modern Aswan) in Upper Egypt (Mutze et al. in press). In use from the early second century BCE until the late first century CE, excavations not only revealed the presence of numerous sheep (n = 246), but also of dogs (n = 56), calves (n = 13), cats (n = 11) and a single goat. All animals had been deposited in shallow pits without prior mummification. In a number of cases, however, the animals' heads or even their entire bodies had been covered with sherds of large ceramic vessels (Hepa et al. 2018; Mutze et al. in press). Massive dung accumulations indicate that the animals were not only buried but also kept inside the necropolis (Hepa et al. 2018). Since we are dealing with complete individuals, the Syene assemblage offered the exceptional opportunity to compare demographic profiles based on dental (Jones 2006) and epiphyseal ages (Habermehl 1975). In doing so, however, we noted that individual ages based on mandibular tooth wear turned out consistently higher than those based on epiphyseal fusion. Our findings, therefore, clearly deviate from those made by Zeder (2006).

Although epiphyseal fusion can be highly variable as well, there are several arguments that tooth wear is responsible for the observed discrepancy, as we noted extreme tooth wear in some of the dog specimens as well. By far the best evidence, however, is offered by visual comparison of sheep mandibles from the Celtic oppidum of Manching (Bavaria, Germany) and Syene (Upper Egypt) (Fig. 1). While in both specimens, M_3 exhibits wear stage 9G (following Payne 1987), in the Celtic mandible wear in M_1 corresponds to stage 9A and its homologue from Syene fits stage 12A. In other words, demographic profiling based solely on M_1 would estimate the age-at-death somewhere between 1.5 and 4 years for the Manching sheep and between 5 and 7 years (following Jones 2006) for its relative buried at Syene. Even more, if recovered isolated during excavation, most analysts would not take into account the possibility that both M_1 and M_3 pertain to one and the same individual. Consequently, in case isolated teeth were used for demographic profiling, one would obtain a higher proportion of old individuals and a higher average age for the population investigated.

That said, the discrepancy between dental and postcranial findings observed in the Ptolemaic-Roman temple sheep clearly shows the necessity of verifying the rate of wear in cheek teeth prior to discussion and interpretation of age profiles. In our case, two main questions seem particularly relevant. Firstly, intense tooth abrasion and associated high frequency of pathological conditions obviously restricted life expectancy of temple sheep at Syene. So what was the age usually reached by these animals provided they did not die for other reasons? Secondly, does the unexpectedly high tooth wear rate at Syene represent a singular case related to the husbandry conditions in the temple area or are we dealing with a more general phenomenon typical of the region of the First Nile Cataract or even the whole Nile valley?

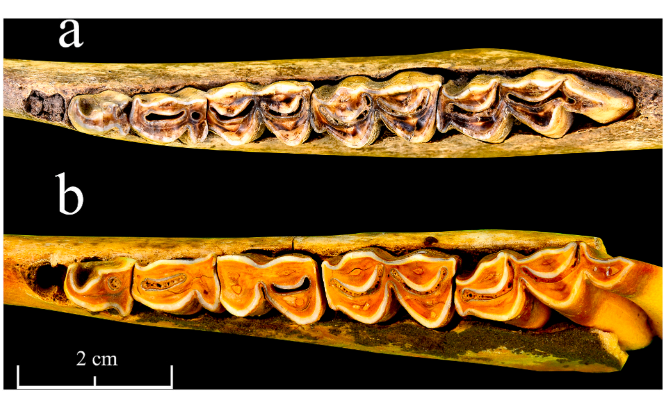

Fig. 1. Sheep mandibles from the Celtic oppidum of Manching, southern Germany (a) and the Ptolemaic-Roman animal cemetery at Syene/Aswan, Upper Egypt (b). Both mandibles exhibit wear stage 9G in the last molar (M_3), whereas in the first molar (M_1) wear stages clearly differ, i.e. 9A (a) vs. 12A (b).

To provide answers to these questions, an approach is needed that evaluates objectively the rate of tooth wear in mandibles, modern as well as ancient. Based on the concept already presented in Mutze et al. (in press), the present paper explains a method which allows the estimation of the rate of tooth wear in an assemblage of sheep mandibles by comparing individual wear states of the different teeth within a single jaw. By means of this method, assemblages of sheep mandibles can be aged and compared with others regarding tooth wear rates in the absence of age determinable postcranial elements. Teeth evaluated are the mandibular deciduous and permanent 4th premolar (dp_4 and P_4) as well as the three molars (M_1, M_2, and M_3). The method presented below is available in the form of the program WoTiS (Wear of Teeth in Sheep), which has been coded in the programming language R. Prior to entering into the details of the R script, however, we will explain our methodological approach assuming the user is carrying it out manually, simply because full understanding of the procedure is essential for interpreting the results obtained.

2. Materials

Our study considers two modern reference materials. The first consists of 1417 observations of 732 live sheep from UK herds examined by one of the authors (Jones 2006). It includes sheep pertaining to several breeds with known ages of up to seven years. This data set is detailed in the Supplements S5.

A second modern reference material comprises 294 mandibles of 149 Karakul sheep housed at the Julius-Kühn-Collection at Halle an der Saale (Germany) and has been recorded by the first author following Payne (1987; cf. Mutze et al. in press). These fat-tailed animals were kept under steady conditions (Frölich 1928) at Halle's 'Haustiergarten'. Since their birth and death dates have been recorded, we know that some individuals even reached 14 years of age.

Taken together, these modern sheep assemblages will be referred to as **standard population** in the following. Merging the data proved essential since the first population suffered from a lack of older individuals and thus of sheep exhibiting advanced wear of late erupting cheek teeth. Combining the data of both populations ensures a uniform age distribution, thereby avoiding a bias of estimated ages by age mimicry (Millard 2006). For the exact procedure of how the two populations were merged, the reader is referred to Supplement S1.

As an example of a statistically relevant prehistoric population the Iron Age sheep assemblage from the Celtic oppidum of Manching (Bavaria, Germany) was chosen. Studied since the mid-1950s and published by Boessneck et al. in 1971, the site's substantial faunal assemblage is now housed at the 'Staatssammlung für Anthropologie und Paläoanatomie', Munich. In the frame of this study we recorded 1024 sheep mandibles following Payne (1987). This data set is detailed in the Supplements S6. Since it is essential that our sample comprises exclusively sheep, we verified identification and excluded goats based on the criteria published by Payne (1985) and Halstead and Collins (2002). Mandibles lacking clear taxonomic features were omitted as well. This procedure inevitably biased the age distribution of our data set, since taxonomic classification of mandibles of very old caprines is problematic. Correspondingly, the data set in S6 should not be considered entirely representative of the sheep population kept at Celtic Manching. In the present study, the Manching data served to test the function *WoTiSrandomSampling* (see Fig. 8), necessitating sufficiently large populations, as is the case for Celtic Manching.

Turning to Egypt, two populations have been investigated. The first concerns a large number of burials of sacred sheep (n = 246) from the Ptolemaic-Early Roman animal necropolis at ancient Syene (modern Aswan) situated on the first Nile Cataract. This assemblage produced 212 sufficiently preserved mandibles from 115 individuals (40 males, 45 females and 30 of unknown sex) which we recorded following Payne (1987). The data set is detailed in the Supplements S4. Regarding the sheep herded in the wider environs of Syene/Aswan, we relied on an archaeofaunal study published recently (Sigl 2017). From this assemblage a total of 23 sheep mandibles could be re-examined (following Payne 1987). These specimens date to different time periods, mainly to the Late Dynastic and Ptolemaic periods (722–30 BCE) as well as Mamluk times (1252–1517 CE) (see Supplements S7). This assemblage has been compared with the mandibles from the animal necropolis applying the function *WoTiScompare* (see Fig. 9).

3. Theoretical considerations: principles underlying WoTiS

3.1. Premises and population terms

As stated in the introduction, we recorded occlusal wear states of individual teeth per mandible. Teeth recorded are the mandibular deciduous and permanent 4th premolar (dp_4 and P_4) and all molars (M_1, M_2, M_3).

A first premise underlying our approach is that eruption of cheek teeth is broadly fixed, while tooth wear can be extremely variable, making certain combinations of wear stages possible and others not (for detailed presentation see Mutze et al. in press). Regarding tooth eruption, many factors have been considered influential. For instance, the teeth of rams have been stated to erupt earlier than those of ewes (Simonds 1855; Behr 1928), but this seems to concern primarily the incisors and could not be confirmed for the cheek teeth (Simonds 1855). Similarly, Worley et al. (2016) could not find any significant differences in cheek tooth eruption between ewes, rams and castrates nor between sheep of low and high nutrition. Major discussion also centred around the impact of intensive breeding on the timing of tooth eruption. Indeed, early authors mentioned significantly later times for tooth eruption in both sheep and cattle (Youatt 1834, quoted by Brown 1927; Silver 1969). This is in contrast with modern breeds, wherein teeth were found to erupt much earlier, leading to the conclusion that intensive breeding causes an overall precocity, including tooth eruption. Deniz and Payne (1979) presented a good overview of this discussion and stated that between breeds, slight differences in the timing of tooth eruption exist that do not correlate at all with early or late maturation. They concluded that ‘estimates of age for archaeological specimens based on dental eruption are probably considerably more reliable than has sometimes been supposed.’ This assessment was confirmed by Jones (2006), who found molar eruption to be similar in different breeds, with variation among breeds not much greater than within individual breeds. In her study, eruption in Soay sheep was found mostly similar to that in other breeds, except for a small number of individuals exhibiting slightly delayed eruption (Jones 2006, Figs. 11–13). In sum, cheek teeth eruption can be considered variable within quite narrow limits, which is in marked contrast to the much greater variability of tooth wear.

A second premise concerns the inner structure of sheep teeth, more precisely the relative depth of the infundibula within the crown. Our model assumes that in each type of tooth (i.e. P_4, M_1 - M_3) the infundibula do not vary more than other anatomic features do. Moreover, their extent does not depend significantly either on breed, sex, husbandry conditions or natural variation. Since determined by the inner structure of the teeth, the ratios between the duration of the individual wear stages of a specific type of tooth can be assumed to be identical for all sheep (cf. Mutze et al. in press), a constant quality of fodder presumed.

Summing up, our standard population thus holds two types of information required by our methodological approach: the age at which a specific type of tooth comes into wear (beginning of stage ‘J’), which fits closely the timing of tooth eruption, and secondly, the aforementioned ratios between the durations of the different wear stages for each type of tooth respectively.

In the following, distinct terms for different types of sheep populations will be used that should be defined first. In the context of this study, a sheep **population** is a fictitious or an existing assemblage of sheep (mandibles) related in itself by a (fairly) homogeneous tooth wear rate. This does not necessarily mean that we are dealing with a population in a strict biological sense.

In addition, our approach distinguishes between real populations and ideal populations. The term **real population** refers to present and past populations of sheep, wherein the wear status of the mandibular cheek teeth can be documented. In case of such populations, the rate of tooth wear can be determined, which is the main purpose of the program WoTiS. Tooth wear data recorded from real populations and fed into WoTiS are termed **find populations**. Arguably, determining the tooth wear rate of a population does only make sense if wear has been fairly homogeneous in (most of) its members. This requires that husbandry conditions were broadly constant. Regarding modern populations, written sources can inform about the living and feeding conditions. With archaeological populations, we also assume this to be the case with mandibles recovered from the same spatio-temporal context. If mandibles from different places and/or cultural epochs are pooled to increase sample size, this has to be taken into account when interpreting the results. Although WoTiS has been developed to evaluate archaeological assemblages, obviously any real population including modern sheep can serve as a find population.

In contrast to a real population, an **ideal population** is characterised by the absence of any inter-individual variation regarding the anatomic structure of the teeth and their rates of wear. However, since such a situation never occurs in reality, ideal populations are fictitious, they simply serve as a model for real populations in our approach. The aforementioned **standard population** represents the main model of tooth wear in a sheep population. By definition it exhibits an intensity of wear equal to 1. It is introduced to WoTiS in the form of a ‘phase boundaries file’. It offers for each type of tooth the lower and upper age limits of the consecutive tooth phases (eruption and wear), more precisely the ages marking the beginning and the end of each stage in the standard population. To define these phase boundaries, we used two real sheep populations with known ages, namely a living one studied by Jones (2006) and a historic one composed of Karakul skeletons housed in the Julius-Kühn-Collection (Halle/Saale, Germany). Supplementary file S1 presents the phase boundaries using the code systems proposed by Payne (1987) and Grant (1975), together with a detailed description of how we proceeded to combine these two data sets. In a single sheep individual, as a matter of principle, the upper limit of one stage coincides with the lower limit of the following stage. Since composed of individuals behaving identically, this principle should hold for any ideal population and thus for our standard population as well. When explaining the method below, we will stick to this principle, because it facilitates graphical representation. However, the method also works in the case of overlapping stages, so we tolerate that our phase boundaries file, since derived from real sheep populations, contains some overlaps. These concern the later wear stages in which the upper limit could not be reliably determined due to the restricted life span of the sheep. For a detailed description of how the phase boundaries have been obtained, the reader is referred to Supplements S1.

To conclude, we distinguish between **real populations** of real sheep and **ideal populations**, a model of real populations that ignores any intra-population variability in dental wear and underlying anatomical structures. The **standard population** is meant to be an ideal population, although in our case derived from two real populations. The results of our methodological approach are the more reliable the better the standard population approaches an ideal population. A **find population** is a real population (normally but not necessarily an archaeological one) analysed by our method. Although it is a real population, our methodological approach treats it like an ideal population, necessarily ignoring unpredictable intra-population variability.

3.2. Wear intensity x

We define x as the wear intensity (of one tooth, one mandible or a whole population) relative to the standard population. So if $x = 2$, teeth wear twice as fast as in the standard population and hence reach specific wear stages in half the time since eruption. This also means that $x_{standard} = 1$ by definition. The following sections will describe how a single tooth's age can be written as a function of the wear intensity x, how to use these functions from multiple teeth in a mandible to produce an overall estimate, and how to use the results of multiple mandibles to estimate the wear intensity x of an assemblage. Supplement S11 provides the reader with examples and more explanations for better understanding of the equations.

3.3. Wear intensity x of one tooth

The wear intensity x_{find} of a single find tooth can be expressed by the following equation.

$$x_{\mathrm{find}} = \frac{t_{\mathrm{standard}} - e}{t_{\mathrm{find}} - e} \quad (1)$$

(t_{find} = age at death of the sheep to which the find tooth belonged, $t_{standard}$ = age, at which the wear condition of the find tooth is reached in the standard population, e = beginning of stage 'J' ('just coming into wear' Ewbank et al. 1964) in the tooth type, e.g. M_1, of the find tooth).

Solving equation (1) for t_{find} leads to:

$$t_{\mathrm{find}} = e + \frac{t_{\mathrm{standard}} - e}{x_{\mathrm{find}}} \quad (2)$$

As an essential transformation, equation (2) is now multiplied by x_{find}, thus producing equation (3):

$$t_{\mathrm{find}} \cdot x_{\mathrm{find}} = e \cdot x_{\mathrm{find}} + (t_{\mathrm{standard}} - e) \quad (3)$$

Since this equation turns out to be a linear function of the form $y = m{\cdot}x + b$, it can then be depicted in a Cartesian plane as a straight line with e being the slope and ($t_{standard} - e$) being the y-intercept, whereby $x = x_{find}$ and $y = t_{find}{\cdot}x_{find}$. Since wear stages always have a more or less pronounced temporal extent, exact ages (t_{find}, $t_{standard}$) can only be assigned to borders between two stages. Straight lines produced by equation (3) hence depict borders between wear stages.

3.4. Wear intensity x of one mandible

Each border between two wear stages in each tooth type can be depicted in the described way, whereby the slope (e) as well as the y-intercept ($t_{standard} - e$) depend on the standard population only and not on the find population. This procedure is used to estimate the wear intensity x of single mandibles. We illustrate this by an example, a mandible exhibiting the stage combination 7 (P_4), 9 (M_1), 9 (M_2), 7 (M_3). If we insert in equation (3) the values for e (beginning of stage 'J') and $t_{standard}$ as given by the phase boundaries file, the lower and upper limits of these stages can be defined as follows:

P_4 stage 7, lower limit: $t_{find}{\cdot}x_{find} = 25{\cdot}x_{find} + 10.5$
P_4 stage 7, upper limit: $t_{find}{\cdot}x_{find} = 25{\cdot}x_{find} + 17.7$
M_1 stage 9, lower limit: $t_{find}{\cdot}x_{find} = 2.5{\cdot}x_{find} + 18.7$
M_1 stage 9, upper limit: $t_{find}{\cdot}x_{find} = 2.5{\cdot}x_{find} + 51.5$
M_2 stage 9, lower limit: $t_{find}{\cdot}x_{find} = 10{\cdot}x_{find} + 30.8$
M_2 stage 9, upper limit: $t_{find}{\cdot}x_{find} = 10{\cdot}x_{find} + 75.4$
M_3 stage 7, lower limit: $t_{find}{\cdot}x_{find} = 22{\cdot}x_{find} + 17.8$
M_3 stage 7, upper limit: $t_{find}{\cdot}x_{find} = 22{\cdot}x_{find} + 23.1$

The resulting straight lines are depicted in Fig. 2. Since their slope (e) equals the beginning of stage 'J' in the respective tooth, the lines representing the stage limits of one tooth always run parallel.

Fig. 3 shows all these stage limits in one Cartesian plane. As can be seen, the wear stages of the four teeth intersect in a rather small area filled grey. This area is the graphical representation of the mandible under consideration. In the following, we will refer to such an area of intersect as **matching area**.

Unfortunately, there are stage combinations observed in existing mandibles that do not produce matching areas since the stages of the single teeth do not intersect. We will refer to such a non-intersecting stage combination as **misfit**. Possible reasons for the occurrence of misfits are:

1. The stage limits given by the standard population are an approximation. In other words, the ratios laying at the origin of the equations may not be identical in all sheep.
2. Slight variations in the timing of tooth eruption may lead to a shift of wear stages different from the standard, particularly if this concerns single teeth, which might occur for example as a result of a temporary disease or malnutrition during tooth formation.
3. The proposed method requires a constant x-value throughout the entire lifetime of an individual. In mobile grazing systems, however, this may not always be the case considering, e.g., seasonal change of food supply. Moreover, asymmetric chewing due to transient painful masticatory pathologies causing stronger wear at the healthy side must also be considered.

In sum, the occurrence of misfits is a normal phenomenon in any real population and it would therefore be inappropriate to treat them as outliers. Instead, they should be analysed along with the dental stage combinations producing matching areas.

Inspired by the work of Sen (2012), we have come up with the

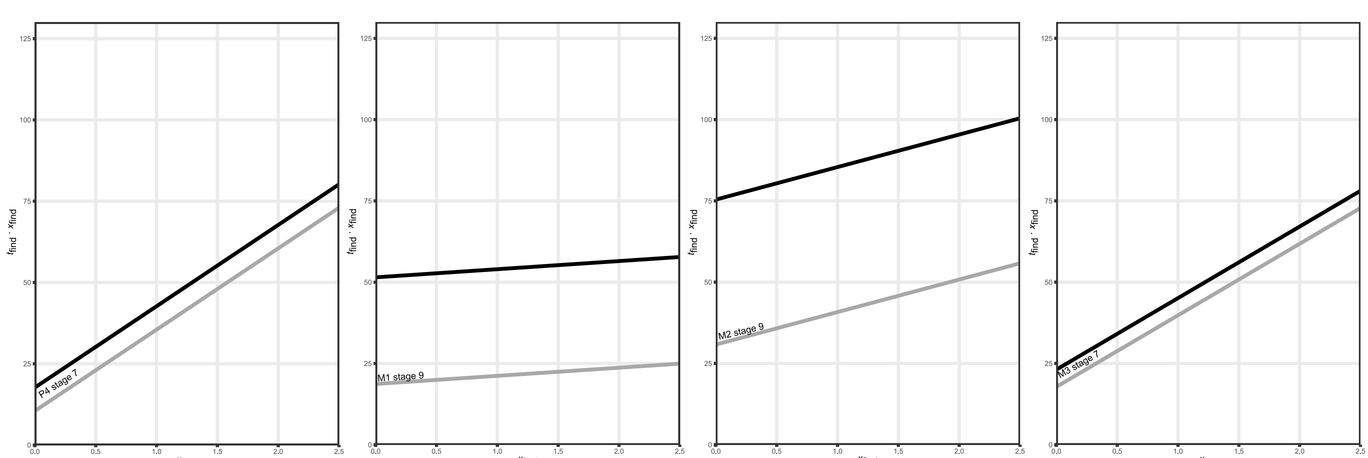

Fig. 2. Lower (grey) and upper (black) limits of the stages 7 (P_4), 9 (M_1), 9 (M_2), 7 (M_3).

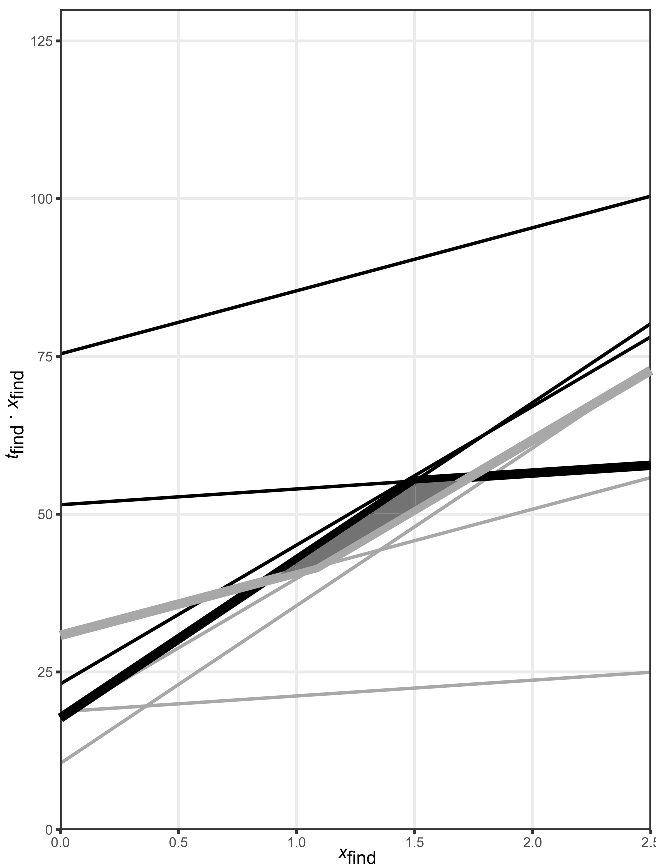

Fig. 3. Lower (grey) and upper (black) limits of the stages 7 (P_4), 9 (M_1), 9 (M_2), 7 (M_3). Matching area filled grey. Collective lower and upper limits bold.

approach of collective lower and upper limits. This allows us to analyse misfits and non-misfits by the same procedure, which is particularly elegant for conversion into a computer program. In Fig. 3 all upper limits are black and all lower limits grey. For each *x*-value, the collective lower limit (grey, bold) corresponds to the highest of the four lower limits and the collective upper limit (black, bold) is defined by the lowest of the four upper limits.

Illustrating a misfit, Fig. 4 shows the collective lower (grey, bold) and upper (black, bold) limits of a mandible with the stage combination 3 (P_4), 9 (M_1), 9 (M_2), 2 (M_3), observed in the assemblage of Celtic Manching. In this case, the collective lower and upper limits do not intersect at all, implying that for every *x*-value, the collective lower limit surpasses the collective upper limit. In such cases, we will use the *x*-value of smallest distance between the two limits in the *y*-direction as best guess (Fig. 4, m).

To conclude, each mandible either produces a matching area with an extent in *x*- and *y*-direction (= regular case) or a vertical line with an exact *x*-value and an extent only in *y*-direction (= misfit).

3.5. Wear intensity x of a population/an assemblage

In regular cases, the matching area exhibits an extent along the *x*-axis encompassing all *x*-values possible for this mandible. In our model, which we call an ideal population, a homogeneous tooth wear rate and thus a single *x*-value is assumed to hold for all mandibles pertaining to a single population. This population specific *x*-value should ideally show up in every matching area. However, due to the fact that the tooth wear rate in a real population is subject to fluctuations, as mentioned above, and due to the occurrence of misfits, this will hardly be the case in practice. Therefore, an *x*-value in best possible accordance with all

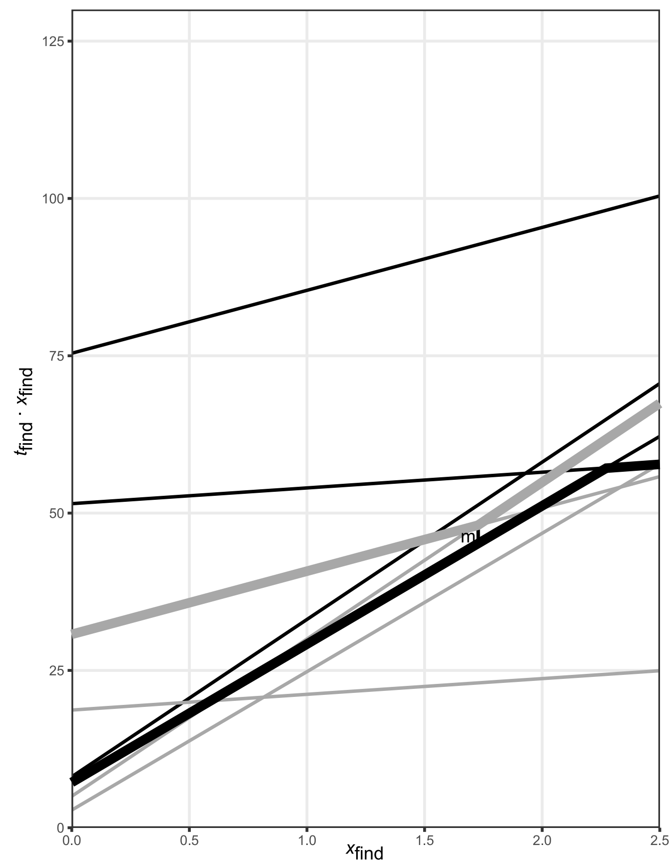

Fig. 4. Lower (grey) and upper (black) limits of the stages 3 (P_4), 9 (M_1), 9 (M_2), 2 (M_3). The misfit value (m) is the smallest vertical distance between the collective lower and upper limits (bold).

matching areas and vertical lines of the population must be found or, in other words, an average *x*-value characterising that population termed **population*x***. Due to the fact that the obtained matching areas can be extremely different (some are even infinite to the right on the graphs), describing a mean is not easy. Since the entire method underlying WoTiS was first executed by hand, calculating the mean developed into a multi-stage process, whereby initially a preliminary mean was obtained, which was then adjusted several times leading to the final population*x*. During the conversion into an R script, however, this procedure was not

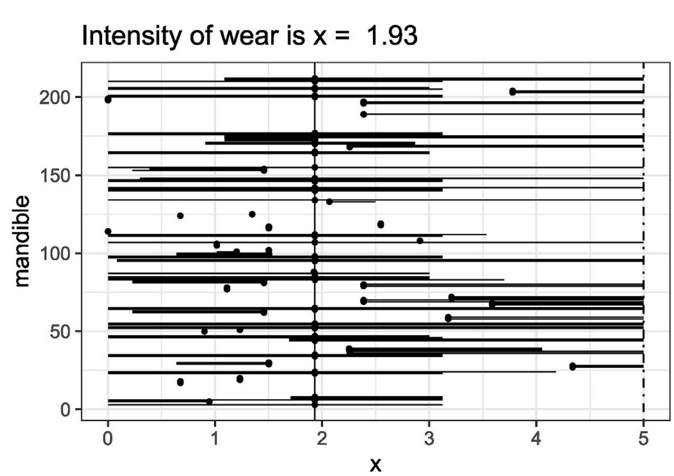

Fig. 5. *x*-values (points) of all mandibles of the population 'Syene necropolis' and resulting population*x* (vertical line).

translated verbatim but reformulated as a non-linear minimization problem with a natural and cheaply computable cost function. Detailing at this point the manual process is therefore superfluous.

Fig. 5 shows the result of this procedure for the population 'Syene necropolis'. The diagram shows the matching areas of all mandibles in this population, but only their extents in x-direction, because the extent in y-direction does not affect the determination of the population$x$. For misfits, only points show up. The y-axis shows the mandibles numbered as in the data set. The vertical line represents the population$x$ determined by WoTiS. To every regular mandible with a matching area, an individual x-value is assigned (points), that is either identical with the population$x$ (in this case the latter is included in the matching area) or as close to it as possible (in this case the population$x$ is not included in the matching area).

One possible usage for the obtained value of population$x$ is the comparison with other populations in order to detect possible differences in husbandry conditions. Here we remind the reader that x is the intensity of tooth wear relative to our standard population. To compare x-values of two populations it is therefore necessary to calculate both of them based on the same standard population. However, we decided not to implement our standard population in the R script but to pass it to WoTiS as a file ('phase boundaries file') allowing the user to work with his/her own standard population. Needless to say, the resulting x-values will not be comparable with those calculated by means of the standard population proposed in this study. To circumvent this problem, another parameter characterizing tooth wear at the population level was defined that is independent of our standard population. This parameter reflects the average duration of stage 9 in M_1 in the find population in months. Termed LSM1 (i.e. long stage M1), it is calculated by WoTiS by the following equation.

$$LSM1 = \frac{\text{duration stage 9 M1 standard}}{\text{population}x}$$

Where this value is high, it implies slow wear and a stage 9 lasting longer. Conversely, a low value points to significant abrasion and a shorter duration of this wear interval. The obvious question with this approach is why stage 9 in M_1 was chosen? Arguably, long-lasting stages have the advantage that minimal differences in the standard populations caused by so-called data noise will have a less marked effect on the calculated duration compared to that in shorter-lasting stages. Moreover, the choice of M_1 is also advantageous because in most standard populations, few M_2 and M_3 are worn down to the level beyond the long-lasting wear stages, i.e. stage 9 in M_2 and stage 11 in M_3, rendering it impossible to determine in a reliable way the upper limits of these stages. Finally, the LSM1-value is a reverse measure of the tooth wear rate – the higher LSM1 the slower the rate of tooth wear– and an innate characteristic of the find population. In sum, since LSM1-values of distinct populations can be compared directly without further manipulation, it is a useful parameter to detect differences in nutrition and thus in husbandry-conditions between find populations when studied by different researchers using different standard populations.

3.6. *Age determination based on the wear intensity x*

Once determined, the parameter population$x$ can be used to estimate the age at death more accurately than with other methods currently available to zooarchaeologists. Considering Figs. 3 and 4, from $y = t_{\text{find}} \cdot x_{\text{find}}$ follows $t_{\text{find}} = \frac{y}{x_{\text{find}}}$ with t_{find} being the age of the animal. Since matching areas extend both in the x- and y-direction, exact ages cannot be calculated, but only an age range, simply because every combination of wear stages will last for some time before abrasion results in the next one. In theory, the exact age of a sheep is to be found between $t_{\text{findmin}} = \frac{y_{\min}}{x_{\text{find}}}$ and $t_{\text{findmax}} = \frac{y_{\max}}{x_{\text{find}}}$. But which x should be used for estimating the age of an individual mandible? Arguably, a mandible's assumed individual x-value is most likely to be found in its matching area, as close to the population$x$ as possible (see Fig. 5). This explains why at first, we intended to use this individual x-value to calculate the individual age of a mandible. But this proved highly problematic in the case of outliers. The latter are not necessarily misfits, as they can generate matching areas, but in the charts these are located very distantly from the population$x$. Correspondingly, age estimates will produce values that make no sense at all when dividing $y_{\min}$ and $y_{\max}$ by the individual x-value. In these cases, using the population$x$ proved much more useful. Provided the individual x-value does not differ significantly from that calculated for the population, the age estimates based on the individual x-value and the population$x$ will not differ significantly either.

For these reasons we decided to always use the population$x$ for estimating individual ages, so that all mandibles are treated identically. A big advantage of doing so is, that criteria for the automated detection of outliers are no longer needed. Of course, for very small find populations or even single mandibles or isolated teeth, the population$x$ may not be reliable. In these cases, the user can (and should) set an x-value or an LSM1-value he considers appropriate for ageing these specimens.

Unfortunately, with outliers not only the x-values but also the y-values proved unreliable, producing erroneous results. Consequently, we decided to use the y-values of the collective upper and lower limits at the population$x$ as $y_{\min}$ and $y_{\max}$. Again, this approach was not only carried out with outliers but also with the entire population, producing more reliable results as reflected by the lower age errors (see below). In case of misfits however, a vertical line is generated, whose x-value will only serve for the determination of the population$x$. For age estimation, as in all mandibles, the y-values of the collective lower and upper limits at the population$x$ are divided by the population$x$ to obtain $t_{\min}$ and $t_{\max}$.

Summing up, of each mandible either the individual x-value of the matching area or the vertical line generated in the case of misfits is used to determine the population$x$. In turn, this population$x$ is then used to calculate the age-at-death of each individual mandible. Taken together, the age-at-death values allow demographic profiling of the entire population to be investigated. Having said that, in the case where population$x$ is not reliable due to small population size, the program WoTiS allows the user to set an x-value or alternatively an LSM1-value manually in order to produce reliable age estimates.

4. Results

The following sections show results of different analyses performed applying WoTiS and also contain some information about settings. For more information about the actual usage of WoTiS, the reader is referred to the User Guide (Supplements S8). For every diagram created by WoTiS and presented here, an input file is provided that allows reproduction of these diagrams (Supplements S9).

4.1. *Analysing a find population with WoTiSmainTask: The sheep buried at the animal necropolis at Syene*

There are three functions embedded in WoTiS: ***WoTiSmainTask***, ***WoTiSrandomSampling***, and ***WoTiScompare***. The latter two have been developed for special test cases. We will return to them later. As the name already suggests, *WoTiSmainTask* does what WoTiS has been developed for: to determine the rate of wear and estimate ages in a find population.

The data set for the necropolis contains one row for every mandible, because the wear stages occasionally differed between the left and right mandible. Individuals with only one mandible preserved are therefore underrepresented. Another possibility would have been to take the average of corresponding wear stages of the left and right side, but we decided to avoid any manipulation of the data.

For an archaeological find population, which of course does not contain data about known ages, WoTiS creates two diagrams, the *intensityOfWear-diagram* and the *ageByWearHistogram*.

(1) The creation of the *intensityOfWear-diagram* (Fig. 5) is controlled by the parameter *createPlotX* (TRUE or FALSE). The diagram shows the matching areas of all mandibles, but only their extents in x-direction, because the extent in y-direction does not affect the determination of the population*x*. For misfits, only points are depicted (see above). The y-axis shows the mandibles, numbered according to their position in the data set.

There are gaps within the diagram, i.e. mandibles for which neither matching area nor a point is depicted, because we decided to exclude mandibles above stage 9 in M_2 or 11 in M_3 (criteria adjustable) from the determination of the population*x* (setting *p2o* for 'purge too old' = TRUE). This is due to the fact that x-values are more homogenous in younger sheep. In older individuals, even small individual variations carry more weight, simply because they take effect over a long time. This fact leads to greater variation of x-values which can affect the population*x* in small populations or in populations with many older individuals. Exclusion of late wear stages is therefore an effective method for minimizing the error rate. In practice, this measure does not have significant disadvantages because settlement refuse in many cases contains a higher proportion of younger individuals. Anyway, WoTiS offers the possibility to disable this setting and use all individuals (*p2o* = FALSE). This might be useful in case of small data sets with only few young individuals. But in these cases, results have to be interpreted very carefully.

The vertical line represents the population*x*, which is 1.93. This value is in best possible accordance with all matching areas and has been calculated by a linear optimization procedure embedded in R (see above). Additionally, for every mandible a point is depicted which indicates the individual x-value of this mandible. This value is always within the matching area but as close as possible to the population*x*. In this case, the population*x* shows that the teeth of the sheep from the animal necropolis at Syene wore nearly twice as fast ($x_{NecropolisSyene} = 1.93 \approx 2$) as the teeth in our standard population.

(2) The creation of the *ageByWearHistogram* (Fig. 6) is controlled by the parameter *createAgeByWearHistogram* (TRUE or FALSE). The histogram is an established tool for depicting age estimates, since it facilitates the interpretation of age distributions. As mentioned above, WoTiS does not provide exact ages but age ranges, which complicates the creation of a histogram. WoTiS creates the *ageByWearHistogram* in the following way: Since the y-axis depicts the number of mandibles, every mandible occupies a sum bin height of 1. If the calculated age range of the mandible is small and does not exceed the width of one bin, this bin gets the full height of 1. But if the age range of a mandible spans more than one bin, the height of 1 will be split proportionally upon the involved bins. This approach has first been used by Ratcliffe and McCullagh (1998) in crime analysis, where the exact time of the crime committed was unknown but a time range was known (e.g. the time range between people leaving their house and coming back, noticing that a burglary has been committed). Ratcliffe and McCullagh (1998) call such crimes 'aoristic crimes', using a Greek-derived word meaning 'indeterminate' (Ashby and Bowers, 2013). The correct term for the number of mandibles shown in the y-axis of the *ageByWearHistogram* is therefore 'aoristic sum' (David Orton, personal communication, 2019). In WoTiS, there are three settings for histograms: the bin width (*binWidth*), the maximum age to be expected in the population (*ageMaxConceivable*), and the maximum age to be depicted in the graphic (*ageMaxGraphics*). Because the geometry of matching areas sometimes results in large age ranges with upper limits significantly higher than physiologically possible, a maximum age (*ageMaxConceivable*) must be set, which corresponds to the maximum life expectancy in the population under study, which we defined in our study as twelve years. For histogram interpretation, however, particularly the very large age ranges must be carefully examined, because they usually introduce a bias in the last bins.

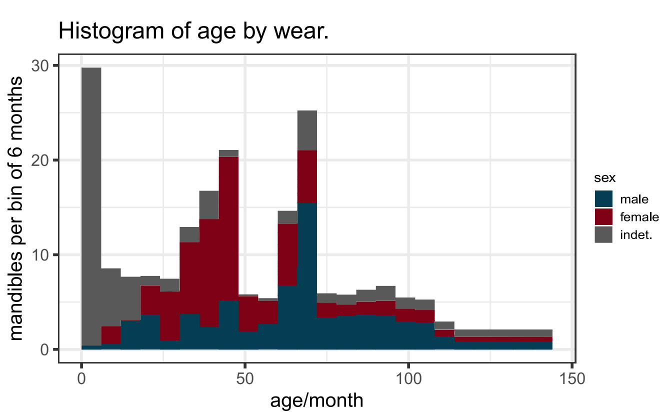

Fig. 6. Age by wear histogram of the sheep from the animal necropolis at Syene.

Fig. 6 shows the age distribution calculated by WoTiS for the sheep from the animal necropolis at Syene. The bin width is six months and the maximum age was set to twelve years (= 144 months). For a better interpretation of the age distribution, the bins have been coloured according to sexes. The latter is not a function of WoTiS though, because this tool was designed to evaluate isolated mandibles, which cannot be sexed. But for interpreting the age distribution of the buried sheep, consideration of the sexes is possible since we are dealing with complete skeletons including skulls.

As can be seen from Fig. 6, many sheep died in their first six months. This is hardly surprising, because even today, lamb mortality amounts to 12–20% worldwide (Bostedt and Dedié 1996). Between six and 30 months mortality is moderate in both sexes. However, between 30 and 48 months it is quite high in females (76%) whilst moderate in males (28%). We explain this discrepancy by conditions associated with reproduction. During gestation, delivery, and lactation, ewes are more prone to disease and ultimately death compared to rams or non-breeding females. Against the background that such early deaths are often associated with first lambing, in the fifth year of life mortality drops to a moderate degree. Between five and six years, however, a third peak occurs. It concerns animals of both sexes, but with a bias towards males. According to our observations, many of these older sheep suffered from severe dental problems, conceivably due to extreme tooth wear, eventually causing death. In some of these animals we noted that the vertebral epiphyses had recently fused which happens between four and five years (Habermehl 1975). The next set of bins comprising animals aged six to nine and a half years are of moderate height and correspond to ca. 33% of the population. The bins between 9.5 and 12 years are exactly of the same height. This suggests that we are dealing with large age ranges (see above) that must be considered artefacts. From the foregoing we conclude that there are three peaks in the age distribution of sheep from the animal necropolis at Syene, corresponding respectively to lamb mortality (0–6 months), breeding ewes (2.5–4 years) and sheep exhibiting severe dental problems (5–6 years).

In addition to the diagrams, WoTiS creates two files: *results.rds* (for opening the rds-file see User Guide) and *logFile.txt*.

The *results.rds* file is a table with one row for each mandible, which contains the following parameters calculated by WoTiS:

xL and *xU* (unit 1): Lowest (L) and highest (U for upper) x-value within the matching area of the mandible.

ageL and *ageU* (unit months): Lowest (L) and highest (U for upper) age in months calculated for this mandible.

misfit (unit months): In case of misfits, this value gives the height of the smallest vertical distance between the collective lower and upper limit, i.e. the height of the vertical line created in misfits (see Fig. 4). Since it is a section of the *y*-axis, it has the unit months. But note that $y = t \cdot x$, so it is not a direct age difference. As explained earlier, if tooth eruption and wear are completely regular (regardless of the rate of tooth wear), misfits will not occur, hence the misfit value will be 0. The bigger the irregularity of a mandible, the bigger the distance between upper and lower limit, i.e. the bigger the misfit-value.

xDev (unit 1): The *x*-deviation is the smallest distance between the population*x* and the *x*-range of the mandible. In the case where the population*x* is included in the mandibles *x*-range, *xDev* is 0. Positive values indicate deviations to the right and negative values to the left.

The *logFile.txt* contains all settings of the current run and the following calculated population parameters. The values for the necropolis are given in Table 1.

populationx (unit 1) see above

LSM1-value (unit months) see above

Mean misfit and *Median misfit*: The arithmetic mean and the median of all misfit values. These values are measures for the 'irregularity' of eruption and wear in the population (see *misfit*). Please note that the eruption times and stage ratios given by the standard population define what is considered 'regular' in this regard. So if many misfits do occur in a find population, i.e. if the mean and median misfits are high, it is possible that the standard population is not appropriate for evaluating this find population. One could then try to accommodate the standard population (by trial and error) to get lower *misfit* values. If there is a system behind the misfits, this will work and will reveal special characteristics of the find population, e.g. late eruption times. If it is not possible to lower the *misfit* values, a possible explanation could be high frequencies of old individuals or dental pathologies (or both).

Standard deviation of misfit

Mean xDev, Median xDev and *Standard deviation of xDev*: Arithmetic mean, median and standard deviation of the *x*-deviation. Since the *x*-deviation can be positive and negative and the population*x* is meant to be in best accordance with all *x*-values of the population, the mean and median *x*-deviation should not be far from 0.

Mean unsigned xDev, Median unsigned xDev and *Standard deviation of unsigned xDev*: The unsigned *x*-deviation is the absolute value of the *x*-deviation. It is always positive and the smaller the mean and the median, the more homogenous the tooth wear rate in the population.

Table 1
Parameters of the different populations.

population		Syene necropolis	Syene settlement	Jones (2006)	Manching
populationx		1.93	1.36	1.08	1.26
*LSM*1 in months		17.0	24.1	30.4	26.0
misfit in months	mean	0.865	0.151	0.446	1.03
	median	0	0	0	0
	Std Dev	3.28	0.501	1.49	2.09
x Deviation	mean	−0.18	0.0332	−0.000472	−0.0211
	median	0	0	0	0
	Std Dev	1.22	0.346	0.258	0.458
unsigned *x* Deviation	mean	0.525	0.222	0.115	0.249
	median	0	0.0441	0	0.0828
	Std Dev	1.11	0.263	0.231	0.385
age Error in months	mean			−0.537	
	median			0	
	Std Dev			4.46	
unsigned age Error in months	mean			2.01	
	median			0.42	
	Std Dev			4.02	
relative age Error in %	mean			−2.02	
	median			0	
	Std Dev			12.9	
unsigned relative age Error in %	mean			7.74	
	median			3.05	
	Std Dev			10.5	

4.2. Analysing a known-age population with WoTiSmainTask: the Jones population

To decide whether or not WoTiS produces reliable results, we tested it with several known-age populations. As an example, we present here the sheep population studied by Jones (2006) consisting of 1417 observations of 732 live sheep of known age. Besides the *ageByWearHistogram* and the *intensityOfWear-diagram*, for known-age populations WoTiS is able to create some more diagrams:

The *TrueAgeHistogram* gives the real age distribution in the population and its creation is controlled by the parameter *createTrueAgeHistogram* (TRUE or FALSE).

If TRUE, the parameter *createPlotComparison* enables two plots:

The *combinedAgeHistogram* shows both the real (green) and the calculated (grey) age distribution. Fig. 7a shows the *combinedAgeHistogram* for the Jones population. The bin width was set to 3 months, which is quite narrow. Therefore, both age distributions differ substantially in some bins but as a whole, they resemble quite closely. The maximum age was set to 100 months and the last bins of the calculated distribution are all of the same height, indicating that they are artefacts, caused by large age ranges (see above). The real age distribution shows that this assessment is correct.

The *ageDifferencesHistogram* (Fig. 7b) uses the same bin width as the age histograms and depicts the unsigned age Errors (see below). Fig. 7b shows that most of the ages could be estimated with an error less than 3 months.

Besides the parameters calculated for find populations, for known-age populations, the *results.rds*-file gives some more parameters concerning the quality of age determination:

The *ageError* (unit months) is the minimum difference between the calculated age range and the true age. It is a negative value if the true age is bigger than the calculated age range, i.e., if the age is underestimated. For overestimated ages it is a positive value. If the true age is within the calculated age range, the age error is 0.

The *relAgeError* (relative age error, unit %) is the age error as a percentage of the true age of the animal. This parameter is of more general validity, because age errors naturally tend to be bigger in older individuals. An age error of 6 months is acceptable in an 8 years old animal but not acceptable in a 6 months old individual. Therefore, the relative age error is a better indicator of the quality of age determination. As with the age Error, the relative age error can be negative or positive.

As in find populations, the *logFile.txt* gives the settings and the population parameters. In addition to the parameters given for find populations, for known-age populations the age error parameters are given as well:

Mean ageError, Median ageError and *Standard deviation of ageError*: Arithmetic mean, median and standard deviation of age error. If the tooth wear rate of the population has been correctly determined, the mean and median should be not far from 0 since positive and negative values should be balanced.

Mean unsigned ageError, Median unsigned ageError and *Standard deviation of ageError*: The unsigned age error is the absolute value of the age error, i.e. it is always positive. Therefore, the mean always differs from 0.

Mean relative ageError, Median relative ageError and *Standard deviation of relative ageError*: Arithmetic mean, median and standard deviation of the relative age error.

Mean relative unsigned ageError, Median relative unsigned ageError, Standard deviation of relative unsigned ageError: The relative unsigned age error is the absolute value of the relative age error.

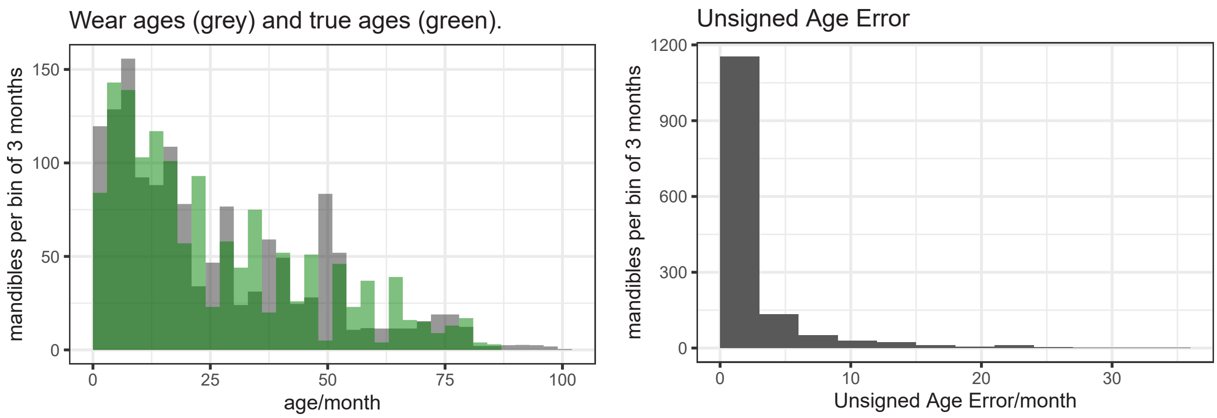

Fig. 7. *CombinedAgeHistogram* and *ageDifferencesHistogram* for the population investigated by Jones (2006).

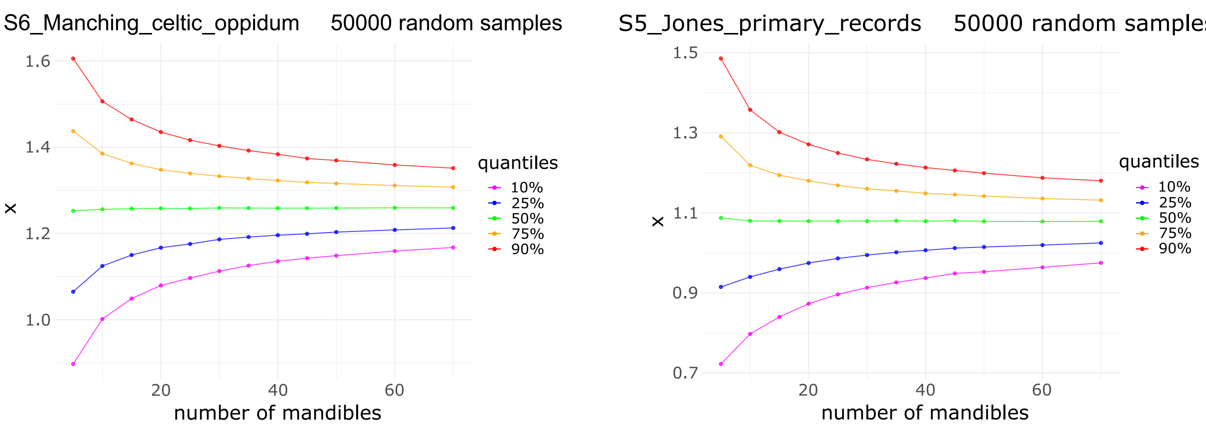

Fig. 8. Scattering of *x*-values based on random sampling of 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, and 70 mandibles (50,000 runs) from Celtic Manching (a) and modern UK (b).

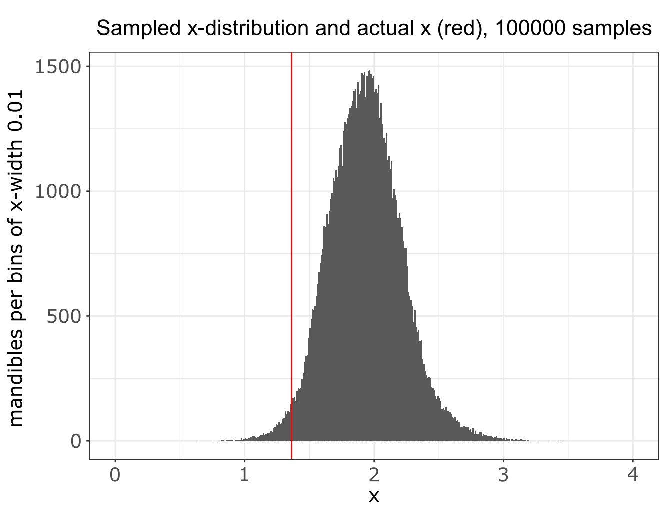

Fig. 9. Results of the comparison between the sheep populations *Syene*$_{\text{necropolis}}$ and *Syene*$_{\text{settlement}}$ by means of *WoTiScompare*.

The population parameters for the Jones population are given in Table 1.

4.3. WoTiSrandomSampling: Scattering of x-values in relation to sample size

Since WoTiS is meant to be used for analysing archaeological specimens, sample size is a central issue. Faunal assemblages from residential areas usually do not contain large numbers of sheep mandibles. Needless to say, the values of population*x* calculated by WoTiS are more reliable the larger the population size. Therefore, the interpretation of results requires a correct assessment of their reliability, i.e., how much scattering of *x*-values should be considered normal for a certain sample size.

To answer this question, another function was embedded into WoTiS: *WoTiSrandomSampling*, which performs numerous runs of *WoTiSmainTask* using random samples of certain sizes out of the find file under consideration. Any run of *WoTiSrandomSampling* is controlled by three additional parameters: *nS*, which is the number of runs per sample size, *nL*, which is a vector giving the different sample sizes and *seed*. Any integer value for *seed* different from 0 makes the random selections reproducible when we rerun the program with all parameters unchanged, including *seed*. Fig. 8 shows the results for two different populations. The first is an assemblage of 1024 sheep mandibles from the Celtic oppidum of Manching (Germany, see above). The second is the population studied by Jones (2006), consisting of 1417 observations of 732 live sheep. For both populations, sample sizes of 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, and 70 mandibles have been analysed. For each sample size, 50,000 runs were performed, each providing an *x*-value. The diagrams give the 10%, 25%, 50%, 75% and 90% quantiles of the calculated *x*-values. Using *WoTiSrandomSampling* only makes sense with big data sets. The data set has to be much greater than the samples drawn from it to avoid analysing the same mandibles many times. With more than 1000 mandibles each, the two data sets exploited here are both large enough and comparable in size. Interestingly, the two scattergrams of *x*-values resemble each other quite closely (Fig. 8), suggesting that this may be a regular pattern in large find populations. Essential for the interpretation of these plots is that they were generated with the setting *p2o* = TRUE (for ‘purge too old’). This means that mandibles exhibiting too advanced wear for reliable *x*-determination (see section 4.1) had been excluded from the data set prior to random sampling. In the graphs, the number of mandibles in the *x*-axis is the effective sample size actually used for calculating *x*. To estimate the reliability of *x* in a find assemblage, its effective sample size has to be compared with Fig. 8. When analysing the find file with *p2o* = TRUE, the effective sample size is given by the *logFile.txt* in the sentence: *Determination of x was done by analysing the wear stages of xxx out of yyy mandibles. WoTiSrandomSampling* also creates a histogram for each sample size, showing the distribution of obtained *x*-values.

4.4. Comparing two populations with WoTiScompare: sheep mandibles from the necropolis and from the settlement at Syene

To answer the question, if the unexpectedly high tooth wear rate of the sheep buried at the animal necropolis at Syene was a local phenomenon due to the husbandry conditions at the site or if it should be assumed for larger stretches of the Nile Valley, it was necessary to compare this assemblage with another one from the same region. Since the faunal remains from the settlement of Syene, studied by Johanna Sigl for her dissertation (Sigl 2017), were available, we decided to reexamine the sheep mandibles (since she used a different code system). These mandibles have been excavated from sites within the modern town of Aswan and date to different times with focus on the Late and Ptolemaic period and the Mamluk time. Even though mandibles of all time periods were pooled, the sample size was very small (23 mandibles, 1 too old, thus 22 mandibles). The result was as follows:

$x_{\text{necropolis}} = 1.93$

$LSM1_{necropolis} = 17.0$

$x_{\text{settlement}} = 1.36$

$LSM1_{settlement} = 24.1$

There seems to be a marked difference between the two data sets, indicating that tooth wear in the necropolis was much more pronounced than in the settlement of Syene. But this result is inconclusive because of the small sample size of settlement mandibles. The presented results of *WoTiSrandomSampling* (Fig. 8) show how much scattering of *x* is to be expected for the different sample sizes. Since the two populations from Syene differ significantly in size (with *p2o* = TRUE:$n_{\text{necropolis}} = 112$, $n_{\text{settlement}} = 22$), it is for the moment difficult to decide whether the difference between the two *x*-values is beyond coincidence.

Since zooarchaeologists are often dealing with very small samples, we developed the function *WoTiScompare* that addresses the problem of comparing two populations of different size. *WoTiScompare* has been designed to test the following null hypothesis:

With regard to populationx, the smaller assemblage ($Syene_{\text{settlement}}$) could be a subset of the larger assemblage ($Syene_{\text{necropolis}}$).

To test this hypothesis, *WoTiScompare* determines the values of population*x* of many randomly selected subsets of the larger assemblage, whereby the size of each subset equals the size of the smaller assemblage. The population*x* values obtained are shown in a histogram (Fig. 9), wherein the population*x* of the smaller assemblage shows up as a red vertical line. *WoTiScompare* furthermore compares the *x*-value of the smaller assemblage with all the *x*-values obtained from the subsets of the larger assemblage and calculates the ratio of *x*-values below or above the *x*-value expressed by the red line. In our case, the $Syene_{\text{settlement}}$ population consists of 23 mandibles, while the $Syene_{\text{necropolis}}$ population contains 212 mandibles. With the setting *p2o* = TRUE, 22 mandibles can be retained from the settlement and 112 from the necropolis whose teeth allow for reliable *x*-determination. If *nS* = 100,000, *WoTiScompare* will randomly select 100,000 subsets of 22 mandibles out of the 112 mandibles from $Syene_{\text{necropolis}}$. The parameter *seed* can be used again to render the random selections reproducible. Fig. 9 shows the distribution of *x*-values obtained for the 100,000 samples (*seed* ← 1, *xBinwidth* ← 0.01). Logically, the peak of the histogram corresponds to the population*x* of $Syene_{\text{necropolis}}$, i.e. 1.93. The population*x* of $Syene_{\text{settlement}}$ (1.36) is indicated by the vertical red line. Out of the 100,000 samples of 22 mandibles, only 2005 yielded *x*-values equal or below 1.36, implying that the probability of the 22 sheep mandibles from $Syene_{\text{settlement}}$ representing a subset of the $Syene_{\text{necropolis}}$ sheep population is only 2.005%. The *logfile* produced by *WoTiScompare* gives this value as $p = 0.02005$. Since $p < 0.05$, the null hypothesis should be rejected (Field et al. 2012). It follows that the difference between the two populations is significant. This is strongly indicative of the existence of distinct husbandry conditions of $Syene_{\text{necropolis}}$ sheep, resulting in a significantly higher tooth wear rate.

5. Discussion and conclusion

5.1. WoTiSmainTask and the relevance of differing wear intensities

WoTiSmainTask determines the wear intensity *x* of a find population and provides age estimation based on *x*. In the case of known age populations it also compares estimated and known ages.

The comparison of estimated and known ages of the sheep population studied by Jones (2006) (Fig. 7) shows a convincing agreement. However, it must be admitted that comparison with this population holds two dangers:

(1) It is part of the standard population. Millard (2006) has outlined that the age distribution of the standard population can bias age estimation, a phenomenon called ‘age mimicry’.
(2) The population studied by Jones (2006) is very large and *x* can therefore be determined much more accurately than should be expected in most archaeological find populations.

Regarding age mimicry, we have made efforts to balance the age distribution of the standard population by involving a second population. Additionally, the standard population plays a different role in our approach than outlined by Millard (2006). Nevertheless, the quality of age determination by *WoTiSmainTask* should be tested with more known age populations, ideally exhibiting wear intensities different from 1.

Another major question focuses on the statistical significance and practical relevance of the calculated differences in the wear intensity *x*. To test whether the differences between our four populations are statistically significant, Welch’s *t*-test has been performed. It is a two-sample test used to test the hypothesis that two populations represented by two samples have equal means, where sample size and variances may be different. This is essential since in our case, sample size is extremely different and variances should be assumed to differ as well, since they depend on the uniformity of feeding and overall husbandry practices within the population. Since find populations are not necessarily populations in a biological sence and ancient husbandry conditions are usually unknown, variances of find populations will hardly be equal. The parameters used by Welch’s *t*-test are the mean (= population*x*), the sample size (= sample size after *p2o*), and the standard deviation (= standard deviation of *xDev*). The respective p-values are given by Table 2.

Statistically speaking, the difference between the Celtic population of Manching and the sheep from Syene settlement turned out not to be significant ($p = 0.096 > 0.05$). This can be explained by their similarity of population*x* (1.26 vs. 1.36) and the very small sample size of the Syene population (= 22). However, all other combinations of populations yielded significant differences. But what is the practical relevance of these differences?

When using an inappropriate *x*-value for age determination, differences between calculated and real ages depend on the tooth under consideration, with differences between early and late erupting teeth

Table 2
p-values calculated via Welch's *t*-test.

	Syene necropolis	Syene settlement	Jones (2006)	Manching
Syene necropolis		2.876 E−05	1.392 E−11	3.088 E−08
Syene settlement	2.876 E−05		5.061 E−04	9.564 E−02
Jones (2006)	1.392 E−11	5.061 E−04		0
Manching	3.088 E−08	9.564 E−02	0	

increasing with rising population*x*. Consequently, in cases when population*x* is high, demographic profiles based essentially on isolated M_1 and M_2 will produce age-at-death histograms that are skewed more or less heavily to the right. Conclusions drawn from such profiles regarding sheep management as well as the animals' life expectancy are misleading and thus unsuitable to address the species' mode of exploitation in ancient times. Fig. 10 illustrates this by comparing age distributions based on M_1/M_2 (grey) and P_4/M_3 (hatched) for the sheep buried at the animal necropolis of Syene. The ages have been calculated on the basis of x_{Jones} (=1.08) and the discrepancy between the two distributions could be interpreted as showing that this *x*-value is inadequate for this population. However, it turns out that the discrepancy is not significantly reduced by using x = 1.93 (Table 1). Discussion of this discrepancy has to take into account that Fig. 10 refers to a selection of specimens (131 out of 212).

The foregoing shows the importance of considering wear intensity when reconstructing age profiles of ancient populations. *WoTiSmainTask* offers the possibility to do this in a transparent way. Of course, the accuracy of the results achieved with WoTiS stands and falls with the quality of the standard population. Therefore we invite the users to question the phase boundaries provided by our standard population and to try out alternatives. With corresponding standard populations WoTiS could also be applied to other hypsodont taxa, e.g., *Capra* or *Bos*.

5.2. *WoTiSrandomSampling and the impact of sample size*

As always in archaeology and archaeozoology, sample size is a key issue. Or better said, since the sample size cannot be selected freely, the question of crucial importance is the extent to which the sample size affects the reliability of the results.

The function *WoTiSrandomSampling* has been developed to address this question. It performs numerous runs of *WoTiSmainTask* using random subsamples of certain sizes out of a find file and examines the distribution of *x*-values obtained for these subsamples. To obtain smooth curves, an extremely large number of subsamples must be examined,

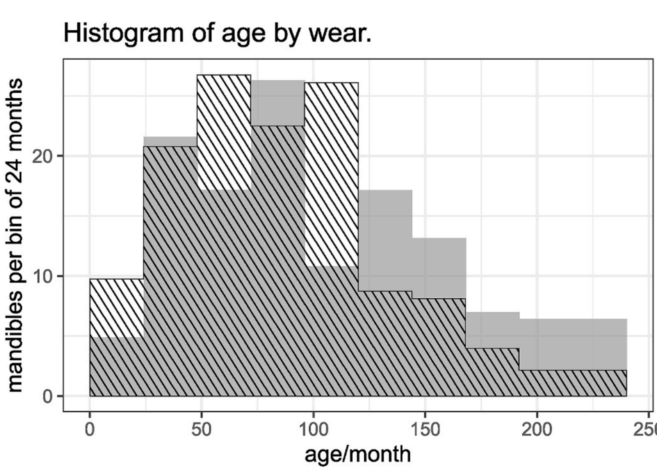

Fig. 10. Age distributions of the population $Syene_{necropolis}$ calculated from M_1/M_2 (grey) and P_4/M_3 (hatched), based on x_{Jones} (=1.08). Only specimens with information for all four teeth.

which takes a lot of time. The graphics shown in Fig. 8 have been calculated by WoTiS in about 30 days each, with the speed naturally depending on the available hardware. Accordingly, this function is not intended to be performed routinely by every user.

However, once created, the graphics of the Jones- and Manching populations (Fig. 8) show a very similar pattern of the relationship between sample size and the scattering of *x*-values. In both, differences of obtained *x*-values are extreme if only five mandibles are considered, but the curve quickly flattens out. It should be noted, however, that even with very large samples, a certain amount of variation in *x*-values can still be expected.

In this respect, the question of the perfect sample size cannot be answered. The larger the sample, the more reliable the *x*-value. The transitions are smooth. However, it can be seen that the curve becomes quite flat from a sample size of 25 mandibles compared to its previous course, which indicates that for this sample size the *x*-value can already be narrowed down quite well. The interpretation of the results obtained with WoTiS should always be done carefully and with consideration of the possible variation of *x*-values.

For the question of the statistical significance of a difference between *x*-values of two populations, sample size is also an important parameter. It can be taken into account via Welch's *t*-test (see section 5.1). For comparing two populations of very different size, the function *WoTiScompare* has been developed (see below).

5.3. *WoTiScompare and implications regarding husbandry conditions of sheep at the animal necropolis of Syene*

WoTiScompare compares two populations of different size by determining the values of population*x* of many randomly selected subsets of the larger assemblage, whereby the size of each subset equals the size of the smaller assemblage. If the number of runs is large enough to get a smooth curve, this offers a good assessment of the probability of the difference in wear intensity being coincidental.

The comparison between the sheep mandibles from the necropolis and from the settlement at Syene (Fig. 9) reveals a very low probability of coincidence (2.005%), confirming statistical significance of the difference. We consider this a useful tool in addition to established statistical tests (like Welch's *t*-test, see above), since it accounts for sample size in a straightforward way.

In the present case, the result confirms the assumption that the tooth wear rate observed in the necropolis sheep was extreme not only in comparison with Central European reference populations, but also with sheep mandibles from contexts that are comparable in space and time, which suggests that the former were exposed to special husbandry conditions.

Massive dung accumulations in the courtyards of the animal necropolis at Syene indicate that the animals have not only been buried there, but also lived there in enclosures, surrounded by walls (Hepa et al. 2018; Mutze et al. in press). The high wear intensity suggests that they were not able to search for food, but had to eat what was put at their disoposal. This fodder probably was contaminated with dust and grit during transportation or inside the enclosures by trampling and pulling it over the ground (Mutze et al. in press). In view of their specific situation, the extreme tooth wear rate observed in this sheep population seems therefore hardly representative for sheep herds allowed to pasture freely in the Nile valley and adjacent areas. Because of the small sample size of the settlement population, additional sheep mandibles from this and other Egyptian sites should be examined to consolidate this result.

Regarding the age distribution of the sheep at the necropolis (Fig. 6), we clearly see three peaks corresponding to the groups with the greatest health risks, namely lambs (0–6 months), breeding ewes (2.5–4 years) and older sheep with extremely worn teeth (5–6 years). Together with the fact that no traces of killing were found on any of the skeletons, we can conclude, that we are dealing with perished animals instead of human controlled kill-off.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are indebted to Renate Schafberg, curator of the Julius-Kühn-Collection, Martin-Luther-Universität Halle-Wittenberg for access to the Karakul skeletons and to Johanna Sigl for access and help with the sheep mandibles from the settlement of Syene and for fruitful discussion. David Orton kindly commented on an earlier version of the manuscript. We are grateful to Cornelius von Pilgrim, Wolfgang Müller, Mariola Hepa, and Lea Kreuzburg for the efficient and enjoyable cooperation on the animal necropolis at Syene. Funding was provided by the Deutsche Forschungsgemeinschaft (DFG) in the frame of the project "Die Tempelherde und ihr Bestattungsplatz im ptolemäisch-römischen Syene aus archäozoologischer, archäologischer und religionshistorischer Perspektive" (DFG project: 392758722), which is very much appreciated. Travel to and accommodation in Elephantine, Aswan were kindly funded by the Ludwig Borchardt-Stiftung, Schaffhausen, Switzerland.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.17605/OSF.IO/HZCKS.

Credit author statement

Ursula R. Mutze: Conceptualization, Methodology, Formal analysis, Visualization, Writing – original draft. Ulrich Mutze: Software, Formal analysis, Visualization, Writing – review & editing. Gillian G. Jones: Investigation, Data curation, Writing – review & editing. Joris Peters: Resources, Supervision, Funding acquisition, Writing – review & editing.

References

Ashby, M.P.J., Bowers, K.J., 2013. A comparison of methods for temporal analysis of aoristic crime. Crime Science 2 (1), 381. https://doi.org/10.1186/2193-7680-2-1.

Behr, E., 1928. Die Abnutzung der Zähne bei den Karakul-Schafen mit besonderer Berücksichtigung der Abnutzungsanomalien und der Altersbestimmung nach dem Zahnwechsel. Kühn-Archiv (1992) 18, 70–114.

Boessneck, J., von den Driesch, A., Meyer-Lemppenau, U., Wechsler-von Ohlen, E., 1971. Die Tierknochenfunde aus dem Oppidum von Manching, Volume 6. of Manching. Franz Steiner Verlag, Wiesbaden.

Bostedt, H., Dedié, K., 1996. Schaf- und Ziegenkrankheiten: 92 Tabellen. In: UTB für Wissenschaft Große Reihe Veterinärmedizin. Ulmer, Stuttgart.

Brown, G.T., 1927. Dentition as Indicative of the Age of the Animals of the Farm, 8 edition. John Murray, London.

Damuth, J., Janis, M., 2011. On the relationship between hypsodonty and feeding ecology in ungulate mammals, and its utility in palaeoecology. Biol. Rev. 86, 733–758.

Deniz, E., Payne, S., 1979. Eruption and wear in the mandibular dentition of Turkish Angora goats in relation to ageing sheep/goat mandibles from archaeological sites. In: Kubasiewicz, M. (Ed.), Archaeozoology 1, 153–163. Sczeczin.

Deniz, E., Payne, S., 1982. Eruption and wear in the mandibular dentition as a guide to ageing Turkish Angora Goats. In: Wilson, Bob, Grigson, Caroline, Payne, Sebastian (Eds.), *Ageing And Sexing Animal Bones from Archaeological Sites*, BAR British Series. BAR, Oxford, ISBN 0860541924, pp. 155–204.

Ewbank, J.M., Phillipson, D.W., Whitehouse, R.D., 1964. Sheep in the Iron Age: a method of study. Proc. Prehist. Soc. 30, 423–426.

Field, A., Miles, J., Field, Z., 2012. Discovering Statistics Using R, reprint edition. Sage, Los Angeles, Calif, ISBN 9781446200469.

Frölich, G., 1928. Die Zucht des Karakulschafes am Tierzuchtinstitut der Universität Halle als Versuchs- und Forschungsobjekt. Kühn-Archiv (1992) 18, I–XIII.

Grant, A., 1975. The animal bones. In: Cunliffe, Barry (Ed.), Excavations at Portchester Castle. 1: Roman, Reports of the Research Committee of the Society of Antiquaries of London. Society of Antiquaries, London, ISBN 978-0500770245, pp. 378–408.

Grant, A., 1978. Variation in dental attrition in mammals and its relevance to age estimation. In: Brothwell, D.R., Thomas, Kenneth D., Clutton-Brock, Juliet (Eds.), Research Problems in Zooarchaeology. University College London Institute of Archaeology Publications, Taylor and Francis, London, ISBN 0905853075, pp. 103–106.

Grant, A., 1982. The use of tooth wear as a guide to the age of domestic ungulates. In: Wilson, Bob, Grigson, Caroline, Payne, Sebastian (Eds.), *Ageing And Sexing Animal Bones from Archaeological Sites*, BAR British Series. BAR, Oxford, ISBN 0860541924, pp. 91–108.

Greenfield, H.J., Arnold, E.R., 2008. Absolute age and tooth eruption and wear sequences in sheep and goat: determining age-at-death in zooarchaeology using a modern control sample. JAS 35, 836–849.

Habermehl, K.H., 1975. Die Altersbestimmung bei Haus- und Labortieren ; mit 22 Tabellen. Parey, ISBN 3489683161. Berlin.

Halstead, P., Collins, P., 2002. Sorting the sheep from the goats: morphological distinctions between the mandibles and mandibular teeth of adult Ovis and Capra. JAS. 29, 545–553.

Healy, W.B., Ludwig, T.G., 1965. Wear of sheep's teeth: I. The role of ingested soil. New Zealand Journal of Agricutural Research 8 (4), 737–752. https://doi.org/10.1080/00288233.1965.10423710.

Hepa, M., Müller, W., Mutze, U.R., 2018. Neuentdeckung in Assuan: Ein ungewöhnlicher Tierfriedhof. Antike Welt (2), 25–29.

Jones, Gillian G., 2006. Tooth eruption and wear observed in live sheep from Butser Hill, the Cotswold Farm Park and five farms in the Pentland Hills, UK. In: Ruscillo, D. (Ed.), Recent Advances in Ageing and Sexing Animal Bones. Oxbow Books, Oxford, pp. 155–178.

Jones, G.G., Sadler, P., 2012a. A review of published sources for age at death in cattle. Environ. Archaeol. 17 (1), 1–10.

Jones, G.G., Sadler, P., 2012b. Age at death in cattle: methods, older cattle and known-age reference material. Environ. Archaeol. 17 (1), 11–28.

Millard, A.R., 2006. A bayesian approach to ageing sheep/goats from toothwear. In: Ruscillo, D. (Ed.), Recent Advances in Ageing and Sexing Animal Bones. Oxbow Books, Oxford, pp. 145–154.

Mutze, U.R., Müller, W., Hepa, M., Peters, J., in press. Old dentitions and young post-crania: Sheep burials in the Ptolemaic-Early Roman animal necropolis at Syene/Upper Egypt. In Daujat, J., Hadjikoumis, A., Berthon, R., Chahoud, J., Kassianidou, V., Vigne, J.-D. (Eds.), Archaeozoology of Southwest Asia and Adjacent Areas XIII., Atlanta.

Payne, S., 1973. Kill-off patterns in sheep and goats: the mandibles from Asvan Kale. Anatol. Stud. 23, 281–303.

Payne, S., 1985. Morphological distinctions between the mandibular teeth of young sheep, Ovis, and goats, Capra. JAS. 12, 139–147.

Payne, S., 1987. Reference codes for wear states in the mandibular cheek teeth of sheep and goats. JAS. 14, 609–614.

Ratcliffe, J.H., McCullagh, M.J., 1998. Aoristic crime analysis. Int. J. Geogr. Inf. Sci 12 (7), 751–764.

Sen, S., 2012. Intersection of Half Planes and Duality. https://www.youtube.com/watch?v=0-k4xsvnnXU.

Sigl, J., 2017. *Syene II: Die Tierfunde aus den Grabungen von 2000-2009: Dissertation, Beiträge zur ägyptischen Bauforschung und Altertumskunde* 19. PeWe-Verlag, ISBN 9783935012225.

Silver, I.A., 1969. The ageing of domestic animals. In: Brothwell, Don, Higgs, Eric (Eds.), Science in Archaeology. Thames and Hudson, London, pp. 283–302.

Simonds, J.B., 1855. On the teeth of the ox, sheep, and pig, as indicative of the age of the animal. J. Roy. Agric. Soc. Engl. 15, 276–362.

Twiss, K.C., 2008. An assessment of the archaeological applicability of faunal ageing methods based on dental wear. Int. J. Osteoarchaeol. 18, 329–351.

Worley, F., Baker, P., Popkin, P.R.W., Hammon, A., Payne, S., 2016. The Sheep Project (2): The effects of plane of nutrition, castration and the timing of first breeding in ewes on dental eruption and wear in unimproved Shetland sheep. JAS Reports 6, 862–874.

Youatt, W., 1834. Cattle: Their Breeds, Management, and Diseases. Baldwin & Cradock, London.

Zeder, M.A., 2006. Reconciling rates of long bone fusion and tooth eruption and wear in sheep (Ovis) and goat (Capra). In: Ruscillo, D. (Ed.), Recent Advances in Ageing and Sexing Animal Bones. Oxbow Books, Oxford, ISBN 1842171224, pp. 87–118.