TECHNICAL NOTE

AMERICAN JOURNAL OF
BIOLOGICAL ANTHROPOLOGY
The Official Journal of the American Association of Biological Anthropologists

WILEY

Automated age-at-death estimation by cementochronology: Essential application or additional complication?

Benoit Bertrand[1,2,3] 🔘 | **Martine Vercauteren**[2] | **Eugenia Cunha**[4,5] | **Anne Bécart**[1] | **Didier Gosset**[1] | **Valery Hédouin**[1]

[1]Univ. Lille, CHU Lille, ULR 7367 – UTML&A - Unité de Taphonomie Médico-Légale et d'Anatomie, Lille, France

[2]Université Libre de Bruxelles, Anthropology and Human Genetics, Faculty of Sciences, Brussels, Belgium

[3]Muséum national d'Histoire naturelle, Département Homme et Environnement, UMR 7194 – HNHP, Institut de Paléontologie Humaine, Paris, France

[4]Laboratory of Forensic Anthropology, Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Coimbra, Portugal

[5]National Institute of Legal Medicine and Forensic Sciences, Largo da Sé Nova, Coimbra, Portugal

Correspondence
Benoit Bertrand, UTML&A, Univ. Lille, Institut Médico-Légal, rue A. Verhaeghe, Lille, 59000, France.
Email: benoit.bertrand@univ-lille.fr

Funding information
Conseil Régional Hauts-de-France; Thermo Fisher Scientific

Abstract

It has been repeatedly acknowledged that age-at-death estimation based on dental cementum represents a partial and time-consuming method that hinders adoption of this histological approach. User-friendly micrograph analysis represents a growing request of cementochronology. This article evaluates the feasibility of using a module to accurately quantify cementum deposits and compares the module's performance to that of a human expert. On a dental collection ($n = 200$) of known-age individuals, precision and accuracy of estimates performed by a developed program (101 count/tooth; $n = 20,200$ counts) were compared to counts performed manually (5 counts/tooth; $n = 975$ counts). Reliability of the software and agreement between the two approaches were assessed by intraclass correlation coefficient and Bland Altman analysis. The automated module produced reliable and reproducible counts with a higher global precision than the human expert. Although the software is slightly more precise, it shows higher sensitivity to taphonomic damages and does not avoid the trajectory effect described for age-at-death estimation in adults. Likewise, for human counts, global accuracy is acceptable, but underestimations increase with age. The quantification of the agreement between the two approaches shows a minor bias, and 94% of individuals fall within the intervals of agreement. Automation gives an impression of objectivity even though the region of interest, profile position and parameters are defined manually. The automated system may represent a time-saving module that can allow an increase in sample size, which is particularly stimulating for population-based studies.

KEYWORDS
age-at-death, automation, cementochronology, dental cementum

1 | INTRODUCTION

Age at death is a fundamental feature of the biological profile but also one of the most challenging parameters for physical and forensic anthropologists to estimate. The search for reliable osseous or dental indicators has been ongoing for decades. Popularization of histology and interactions with zoologists has opened the door to a marker that is increasingly considered as an unequivocal age predictor: the dental

—WILEY—

cementum. Cementum refers to the avascular and not innervated mineralized connective tissues covering the root surface. Various types of cementum have been described according to their specific structures and adaptive properties (Schroeder, 1986; Yamamoto et al., 2016) and each shows a layered structure separated by incremental lines under light microscopy. However Acellular Extrinsic Fiber Cementum (AEFC) receives the highest attention due to its constant growth rate (Sequeira et al., 1992). Age estimation by cementochronology is achieved by adding the average counts of cementum annuli performed on cross sections to the age of the first acellular cementum deposit of the specific tooth. Even if there are still ongoing debates regarding the protocol (Bertrand, 2013; Colard et al., 2015; Gualdirusso et al., 2022; Petrovic et al., 2021), the impact of periodontal disease (Broucker et al., 2016; Kagerer & Grupe, 2001) and the imprecision ranges and mean error (Bertrand et al., 2019; Dias et al., 2010; Meinl et al., 2008; Wittwer-Backofen et al., 2004), most of study demonstrate that cementum deposits are highly correlated with age-at-death in known age samples.

Since the first publications related to cementochronology (Charles et al., 1986; Stott et al., 1982), authors have warned about the time and cost required for the acquisition of cementochronology data. These questions have become recurring concerns and unquestionably contribute to the nonadoption of histological investigations of cementum in favor of user-friendly and inexpensive techniques (Cunha et al., 2009). The standard procedure for cementum analysis involves three major steps: (i) histological processing of dental tissues, (ii) exploration of the cross-section and investigation of a region of interest (ROI) under light microscopy, and (iii) counting of acellular cementum deposits. The first step cannot be time-compressed. Indeed, even if standardizations of the protocol improve productivity, only a reduction of the number of dental cross-sections would be time- and cost-saving. However, we previously demonstrated that five slides per tooth in the middle third of the root allows access to the variability of this dental tissue and ensures the location of an appropriate ROI (Bertrand et al., 2019). The exploration of dental preparation has benefited from the progress of digital camera systems over the last two decades. This development is far from anecdotal since it led to the abandonment of counts from rudimentary projections (Condon et al., 1986), even if counting through the eyepiece had persisted for a long time (Gocha & Schutkowski, 2012; Wittwer-Backofen & Buba, 2002). Since the 1980s, the third step, cementum annuli counting, has been performed manually, making the studies tedious and subjective. Therefore, this work's effort focuses on experimenting with a semiautomated counting module. The willingness to move away from human counts has already led to efforts in designing automated tools based on different algorithms (Czermak et al., 2006; Klauenberg & Lagona, 2007). However, 15 years after these publications, none of these tools have been integrated into cementochronological studies. Experiencing inconsistencies in their results, Czermak et al. (2006) concluded that the reproducibility of such an approach would only be accessible if the preparations were derived from a standardized procedure. The standardization of our previously published protocol prompted us to experiment with a computerized procedure.

Since periodic structures similar to alternating opaque and translucent zones in dental cementum are widely used for age determination in sclerochronology for the interpretation of incremental structures such as scales, bones, fin rays, and otoliths (Fablet & Le Josse, 2005; Fisher & Hunter, 2018), we have chosen to take on board some of the developments successfully adopted in marine biology. The software platform selected for this work (Visilog–Thermo Fisher Scientific Inc., formerly Visualization Sciences Group, formerly Noesis) supported specific developments for age estimation in biology by the French Research Institute for Exploitation of the Sea (IFREMER) (Troadec & Benzinou, 2002). To design this software to the requirements and challenges of dental histological micrographs, a specific cementochronology module has been developed in cooperation with engineers from Noesis. This article aims to evaluate the feasibility of adopting a semiautomated module to discriminate and count cementum alternating deposits and aims to compare the software's performance to that of a human expert.

2 | MATERIALS AND METHODS

2.1 | Dental reference sample

The dental material used in this study comprises permanent healthy canines processed in the validation study performed by the authors of this research article (Bertrand et al., 2019). For each individual, chronological age was known, as well as sex and postmortem interval. Teeth originating from five different sources are summarized in Table 1. The first source represents the anatomy laboratory of Lille University (France) and consists of 48 canines of individuals of both sexes curated at the Forensic Taphonomy Unit of the University of Lille (France). The second represents the "Colecção de l'Esqueletos Identificados du Século XXI," curated at the Laboratory of Forensic Anthropology of the University of Coimbra (Portugal) (Ferreira et al., 2014, 2020), and consists of 99 canines. The third represents the Schoten collection curated at the Royal Belgian Institute of Natural Sciences (Belgium) (Orban et al., 2011; Orban & Vandoorne, 2006) and consists of 26 canines. The fourth source represents the Châtelet collection curated at the Royal Belgian Institute of Natural Sciences (Belgium) (Orban et al., 2011; Orban & Vandoorne, 2006) and consists of six canine of individuals of both sexes. The last represents identified soldiers from WWI (Haut-Rhin, France) and consists of 21 canines of males curated at the Forensic Taphonomy Unit of the University of Lille (France). Age and sex distributions of the reference sample are reported in Figure 1. The distribution by five-year age classes shows that even if most age classes are represented, individuals under 45 are underrepresented and consist mainly of males.

2.2 | Histological preparation and counting phases

To ensure that the performances of the semiautomated and manual approaches can be compared, each histological preparation was

316 | WILEY AMERICAN JOURNAL OF BIOLOGICAL ANTHROPOLOGY
The Official Journal of the American Association of Biological Anthropologists

BERTRAND ET AL.

TABLE 1 Curation and age, sex distributions of the dental reference sample

Collection	Dental sample		Age range	Sex
Laboratory of Anatomy	$n = 48$		54–90	52.1% M [$n = 25$] - 47.9% F [$n = 23$]
(Lille University, France)	60.4% UC [$n = 29$]	39.6% LC [$n = 19$]	$x = 74.5$; $\sigma = 9.3$	
Colecção de l'Esqueletos Identificados du Século XXI	$n = 99$		25–97	63.6% M [$n = 63$] - 36.4% F [$n = 36$]
(Coimbra University, Portugal)	24.2% UC [$n = 24$]	75.8% LC [$n = 75$]	$x = 69.8$; $\sigma = 16.3$	
Schoten	$n = 26$		19–82	57.7% M [$n = 15$] - 42.3% F [$n = 11$]
(Royal Belgian Institute of Natural Sciences, Belgium)	11.5% UC [$n = 3$]	88.5% LC [$n = 23$]	$x = 53$; $\sigma = 21.3$ years	
Châtelet	$n = 6$		44–83	50% M [$n = 3$] - 50% F [$n = 3$]
(Royal Belgian Institute of Natural Sciences, Belgium)	16.7% UC [$n = 1$]	83.3% LC [$n = 5$]	$x = 61.2$; $\sigma = 15$ years	
Carspach	$n = 21$		19–38	100% M [$n = 21$]
(Archéologie-Alsace, France)	38.1% UC [$n = 8$]	61.9% LC [$n = 13$]	$x = 30.6$; $\sigma = 6.1$ years	
Total	$n = 200$		19–97	63.5% M [$n = 127$] −36.5% F [$n = 73$]
	32.5% UC [$n = 65$]	67.5% LC [$n = 135$]	$x = 64.4$; $\sigma = 19.8$ years	

Abbreviations: F, female; LC, lower canine; M, male; UC, upper canine.
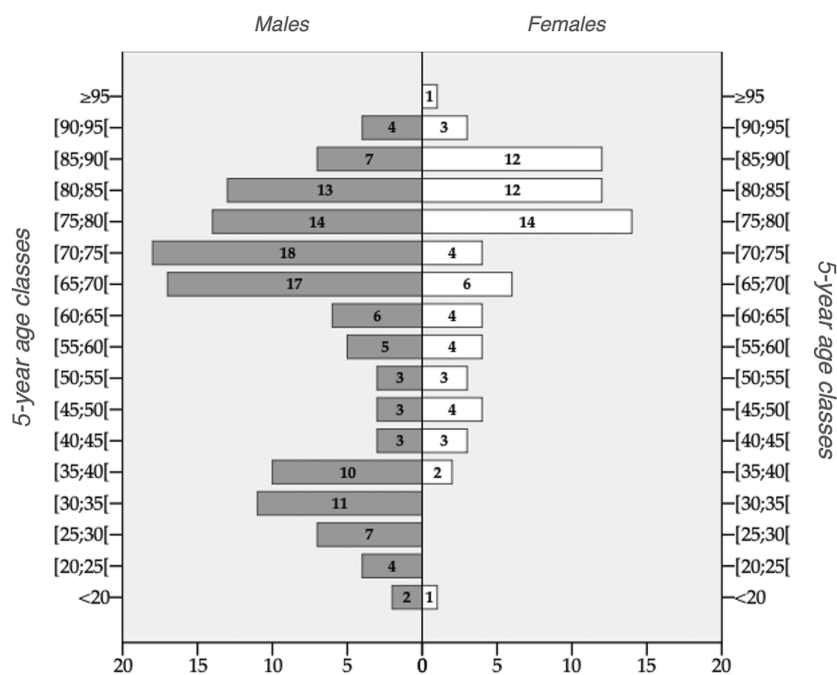
FIGURE 1 Bar chart showing the distribution by age and sex of the dental reference sample grouped by five-years intervals

processed according the ISO-9001 certified protocol (Bertrand, 2013; Colard et al., 2015). Nondecalcified 100-μm cross-sections derived from 200 teeth of known age (five cross-sections per tooth in the middle third of the root) were observed under a light microscope using a Leica® DMEP, and regions of interest showing acellular cementum layering structure were captured as 16-bit JPEG micrographs (1280 × 1024) at 400× magnification with a Leica® DFC280 digital camera and using Leica Application Suite software. All cross-sections were manually assigned an index graded on a five-point rating scale (0: Unreadable; 1: Considerable unreadability; 2: Reasonable readability; 3: Good readability; 4: Clear and unambiguous). This index

assesses the distinctness of acellular cementum structure and not preparation artifacts or taphonomic modifications.

For the manual protocol, hand-operated counts performed by an operator on five cross-sections obtained from each tooth were tracked using the mouse click counter functionality in Adobe® Photoshop CS5. This procedure traditionally adopted by cementochronology users has been repeated on each tooth. A total of 975 counts were manually performed.

For software investigation, we selected one micrograph per individual according to the readability and favoring images graded with the best index. This methodological choice was motivated by

BERTRAND ET AL.

AMERICAN JOURNAL OF
BIOLOGICAL ANTHROPOLOGY
The Official Journal of the American Association of Biological Anthropologists

WILEY

317

the perspective to lighten the sample preparation by reducing the number of cross sections if the software allows it. The selected micrographs were imported in Visilog, which includes a database allowing the storage of processed and to-be-processed images. In the software, each micrograph was identifiable through thumbnails, and metadata associated with the individual can be entered through a form on a windowpane. This solution allows for each file to be filled, for example, name of the image file, name of the site/collection, identification of the individual, sex, age, type of tooth, description, keywords, date of the image processing, and counts if the photograph has been processed (number of cementum rings, mean and standard deviation). These records also allow examination of the library by keyword and allow exportation of all fields or a selection of fields from the search result or the entire data in a Microsoft Excel-compatible file. For each image file, picture quality was automatically assessed. This operator-independent criterion was evaluated in two steps. The first evaluation determined whether the presence of alternating rings was detectable by the software, and the second evaluated the image contrast. The processing steps for quality evaluation are as follows: (1) extraction of the user selection and separation of the three color channels to work only on the green channel; (2) correction of the background by a median filter; (3) calculation of the volume of a gradient oriented in a direction varying by 5° from 0° to 180°; and (4) attribution of a quality grade (quality was assigned to D if the maximum volume value was low; in other cases, the contrast of the image with the highest gradient value was calculated to refine the quality assessment between A and C). This quality estimation allowed the automatic selection of threshold parameters for cementum rings recognition. On the ROI of each micrograph, a set of segments was drawn (Figure 2). The lines traced manually are represented by a central segment surrounded by two external segments and are a simplified representation of a set of parallel invisible segments. It was up to the human operator to choose the origin and end of the central segment. The number of segments and the distance between them (step) were set manually. Additional tools such as noncontiguous segments were available. Along each segment, a command was used to integrate the 10 orthogonal neighboring pixels for each pixel of the current line. The minimum and maximum peak intensity values were determined along the profiles, and visual markers on the main segments allowed the operator to weigh the relevance of automatic detection (Figure 2). The number of annuli detected corresponding to the number of minimums identified along the intensity profile were saved as a Microsoft® Excel file. The image plot and counts were saved for future use. A command allowed the image to be reset to start plotting and processing again. Although the threshold parameters were automatically selected according to the image file quality, these parameters could be set manually to adjust detection through visual markers on the main segments. For each analysis, steps between segments followed the same configuration. Consequently, for each micrograph, 101 counts were performed (1 central segment and 50 lateral and symmetrical segments). A total of 20,200 counts were performed by the software.

2.3 | Performance indicators and statistical analysis

The reproducibility of the software, in other words its aptitude to produce equal results under identical conditions, was assessed. A randomly selected subsample of 50 micrographs (corresponding to 50 individuals) was imported into the software and blindly reassessed 6 months after the initial automated evaluation. This subsample corresponds to 25% of the dental reference sample and represents according to Buikstra and Ubelaker (1994) a reasonable amount for assessing intraobserver error. The software counts variability between T0 and $T + 6$ months was assessed using the intraclass correlation coefficient (Elie & Colombet, 2011; Shrout & Fleiss, 1979) and the Bland–Altman graphical method (Bland & Altman, 1986; Giavarina, 2015). During this interanalysis evaluation, the ROI and profile positions were selected manually.

Regarding the precision of the counts, we adopted absolute and relative indicators. For software-assisted counts as well as for human counts, the dispersion of annuli counts was evaluated with the standard deviation (σ) and with the help of the coefficient of variation. The standard deviation assesses the annuli counts dispersion in relation to the mean in considering the maximum numbers of counting phases (up to five for the human, 101 for the software). The coefficient of variation defined as the ratio of this standard deviation to the average counts. This relative indicator expressed as a percentage is of special interest since an error of five counts is more consequential when a cross-section displays 20 annuli as opposed to 60 annuli.

Regarding the accuracy, and after having achieved age estimation by adding the mean annuli counts to the age of the 3/4 root completion (AlQahtani et al., 2010; Bosshardt & Schroeder, 1996), we assessed absolute accuracy (Δage) by computing the difference between the estimated age and the chronological age. A value equal zero meant the estimate is perfectly accurate, while a negative or positive value respectively suggested an underestimation or overestimation of age-at-death. We also calculated the relative accuracy (Δ%age) by studying the relative error defined as the ratio of the difference between the estimated age and the chronological age relative to the chronological age. This relative approach is also important since a 10-year accuracy does not have the same weight depending on the subject's age.

In addition to the graphical representations of these indicators, we described the general distribution of each indicator in computing skewness and kurtosis values. Kolmogorov–Smirnov tests were used to explore the distribution of precision and accuracy indicators within the whole sample. For both precision and accuracy indicators, we computed Pearson correlations to examine relationships with known chronological ages.

Performances of the software were compared to that of the human previously established by means of identical methods (Bertrand et al., 2019). Global precision of the counts and accuracy of the estimates were compared between the two approaches as well as the general distribution of each indicator. Finally, the level of agreement between automated and manual approaches was examined using the Bland–Altman graphical method.
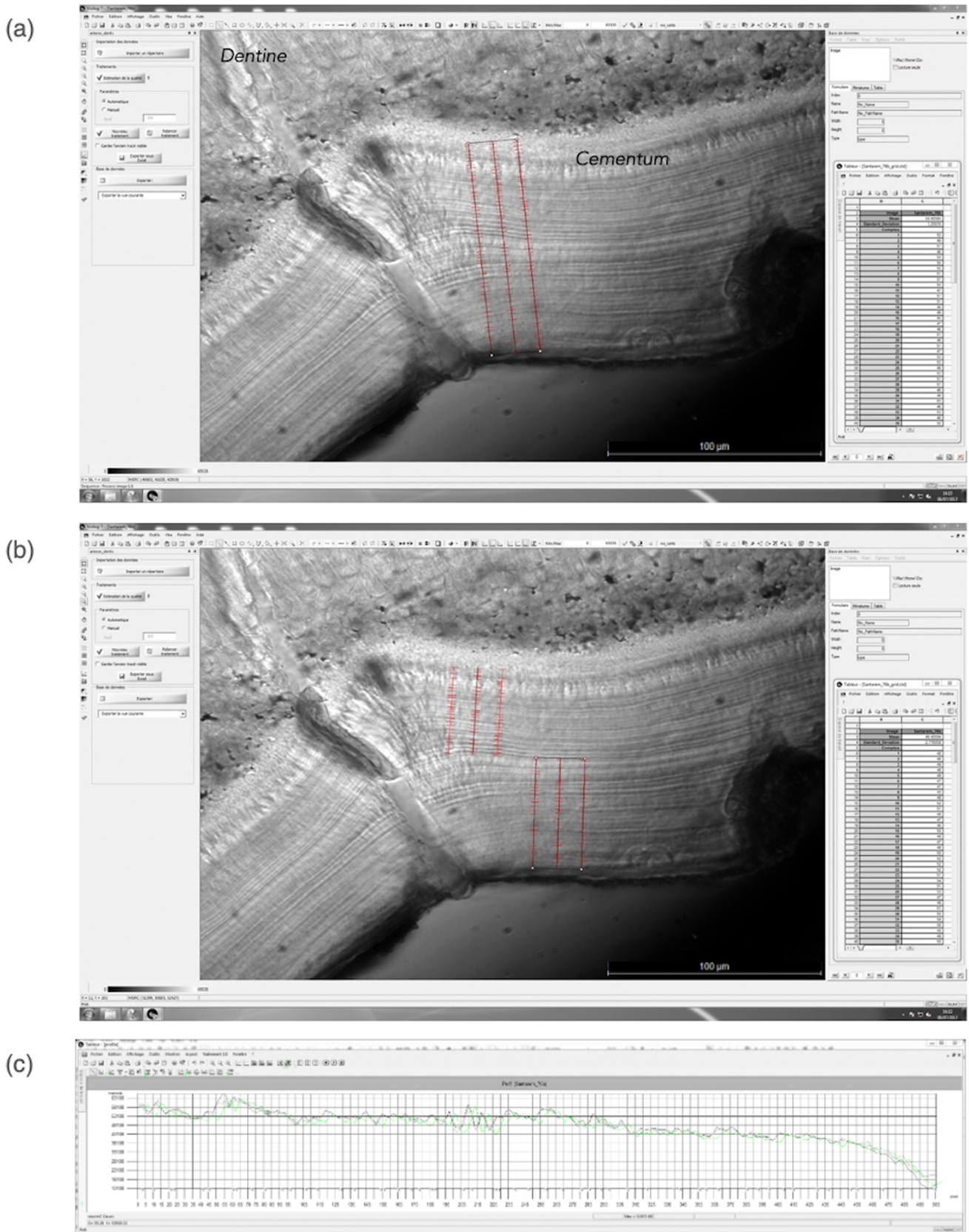
FIGURE 2 Micrograph imported in the image analysis software for automated detection of cementum deposits. The group of parallel segments is drawn by the operator and symbolizes the central segment surrounded by two external segments. Detection can be performed along a single group of segments (a) or from concatenated profiles (b). Intensity profile achieved from one profile along cementum width (c). The scale is 100 μm

Significance was set at $p < 0.05$, and statistical analysis was performed using IBM SPSS Statistics Pack V24.0 (SPSS Inc., Chicago, IL, USA).

3 | RESULTS

3.1 | Reproducibility

The intraclass correlation coefficient was considered to be excellent (ICC = 0.928; 95% CI [0.876–0.958], $n = 50/200$). The Bland–Altman plot examining the extent of disagreement according to the number of annuli shows that the differences between measures are within the $\bar{x} \pm 2\sigma$ limits of agreement (Figure 3) and are congregated around the mean difference estimated at 2.49 annuli. The impression of an accretion around the mean is accentuated by the adaptation of the y-axis scale range to two outliers. This problem was not encountered for the human count, and the reasons may be varied. We investigated whether these two outliers corresponded to any particular age-class or any quality scored by the software and the operator, but we could not find any rational reason. One individual is 41 years and showed an average difference of 25.05 annuli in a positive sense between T0 and $T + 6$ months while the second is 75 years and showed an average difference of 38.96 annuli in a negative sense between T0 and $T + 6$.

We must bear in mind that during semiautomated investigations profiles locations were positioned manually. Without mentioning any malfunction or inadequate settings, a heterogeneous tissue or an artifact at the place where the profiles were repositioned may be the cause of dissimilar interpretations.

A Kolmogorov–Smirnov test performed on the distribution of differences between the automated counts at T0 and $T + 6$ months indicates that the distribution does not follow a normal distribution ($p > 0.001$). This result mirrors the bias caused by the two outliers representing 4%

of the measures ($n = 2/50$). By comparison, this mean difference between T0 and $T + 6$ months for an human operator has been estimated at 1.63 annuli. The dispersion series of points observed on the right of the Bland–Altman plot for human counts is much less obvious for automated counts but is still noticeable. This subtly reveals that the software agreements weaken with the number of annuli to be counted.

3.2 | Precision

Regarding the absolute precision, a Kolmogorov–Smirnov test was performed and established that the distribution of standard deviations follows a normal distribution ($p = 0.200$). The negative kurtosis coefficient of the standard deviation distribution (Kurtosis = −0.450) confirmed the observed slight flattening of the distribution, and the low skewness coefficient (Skewness = 0.093) reveals the central position of the distribution peak (Figure 4a). The standard deviations of the software range from 1.34 to 6.42 annuli (95%CI: 3.45–3.72), and the mean is 3.58 annuli, whereas this mean reached 4.85 annuli for a human expert (Bertrand et al., 2019).

To assess the relative precision, we conducted a normality test. This test expresses that relative precision indicates a nonnormal distribution ($p = 0.000$). The positive kurtosis (kurtosis = 30.385) and skewness coefficients (skewness = 4.216) indicate that recurring precision values are approximately 8% (Figure 4b).

As observed for the manual counts, we are far from reaching a symmetrical distribution. This pattern reflects the tendency of a distribution oriented toward low values of the coefficients of variation, thus toward high precision. Automated relative precision ranges from 4.74 to 31.97% (95%CI: 7.6–8.3), but the average coefficient of variation is 7.93%. This average for manual procedure reached 10.13%.
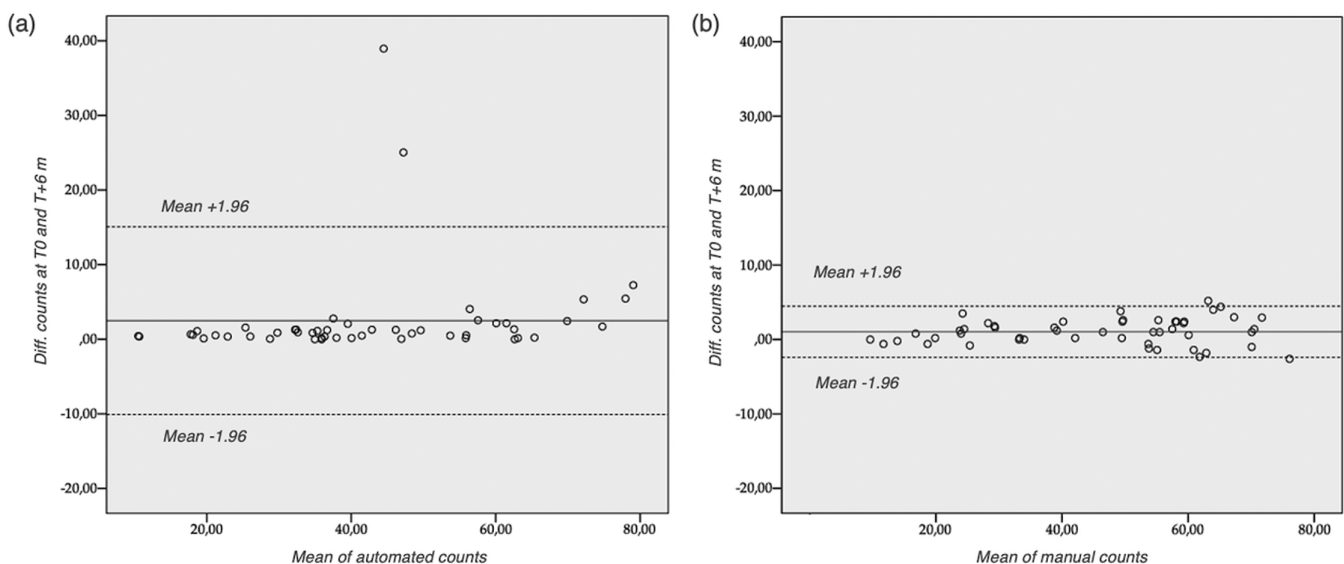

FIGURE 3 Bland–Altman plot showing the concordance between annuli automated counts with an interval of six months (a) compared to an identical approach for manual counts (b) (Bertrand, 2019). The solid line represents the mean difference, and the dotted lines represent the 95% limits of agreement
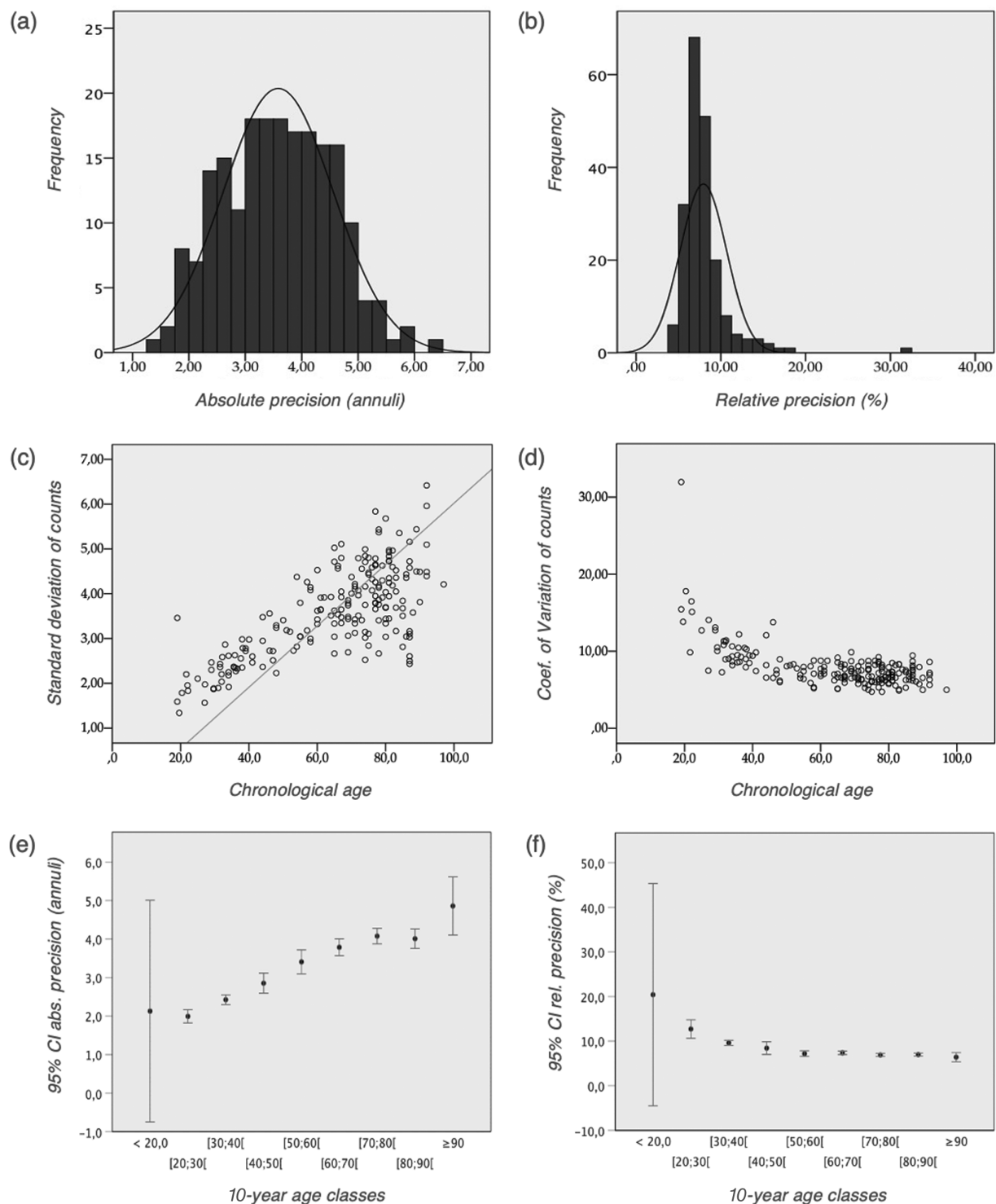
320 | WILEY—**AMERICAN JOURNAL OF BIOLOGICAL ANTHROPOLOGY**
The Official Journal of the American Association of Biological Anthropologists

BERTRAND ET AL.

FIGURE 4 Descriptive statistics of the absolute precision and the relative precision of the counts for the automated approach in cementochronology. a-b: Histograms representing the distribution of the global precision; c-d: Scatter plots displaying the relationship between precision and chronological age; e-f: Error-bar charts indicating 95% confidence intervals of precision by age classes.

3.3 | Accuracy

Regarding the absolute accuracy, a Kolmogorov–Smirnov test showed that the distribution of the absolute error (in years) does not follow a normal distribution ($p > 0.001$). The kurtosis coefficient of the

distribution (kurtosis = 4.158) reflects a peak, and the skewness coefficient indicates that this peak is shifted toward "high" values (skewness = −1.923). As seen with the human counts, the skewness of the distribution indicates a tendency for underestimation and a tendency for the distribution to reach zero value (Figure 5a). The absolute
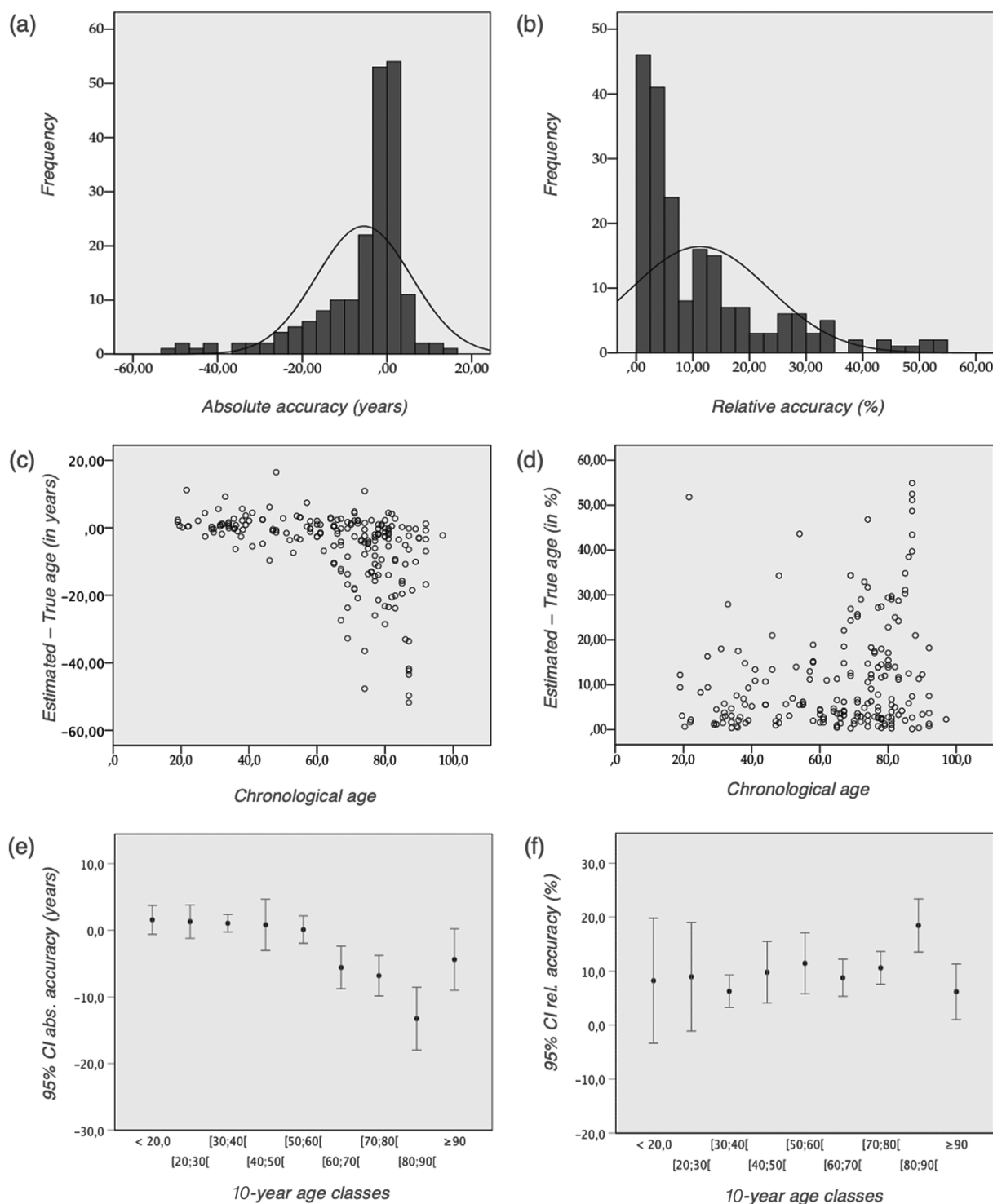
FIGURE 5 Descriptive statistics of the absolute accuracy and the relative accuracy of ages estimates for the automated approach in cementochronology. a-b: Histograms representing the distribution of the global accuracy; c-d: Scatter plots displaying the relationship between accuracy and chronological age; e-f: Error-bar charts indicating 95% confidence intervals of accuracy by age classes.

error ranges from −51.75 years to +16.48 years (95%CI: −7.01 to −3.87) and the average is −5.44 years (−4.53 years for the human expert).

For relative errors (in %), the Kolmogorov–Smirnov test indicates a non-normal distribution (*p* > 0.001). The positive kurtosis coefficient

(kurtosis = 2.230) and the high skewness coefficient (Skewness = 12.152) indicate that the distribution peak shifts to low values of approximately 2% (Figure 5b). The relative error ranges from 0.20% to 54.90% (95%CI: 9.54–12.93), and the average is 11.23% (9.19% for human counting).

322 | WILEY–**AMERICAN JOURNAL OF BIOLOGICAL ANTHROPOLOGY**
The Official Journal of the American Association of Biological Anthropologists

BERTRAND ET AL.

3.4 | Effect of chronological age on the performance of automated estimates

The absolute precision (Figure 5c) shows that the plots draw a linear relationship, suggesting that the older the individual is, the greater the standard deviations of the automated annuli count. This relationship has been demonstrated for human counts. Pearson correlation confirms the strength and statistical significance of this association ($r = 0.721$; $p > 0.001$; $n = 200$). Interestingly, this association is stronger than for human counts ($r = 0.541$; $p > 0.001$; $n = 199$). We investigated the relation between age and imprecision of counts in exploring means and 95%-CI for age classes by means of error bar chart. This graphical representation particularly shows that absolute imprecision of counts increases with age (Figure 5e).

Pearson correlation also confirms the association between chronological age and relative precision ($r = -0.619$; $p > 0.001$; $n = 199$). However, a negative value indicates that the younger the individual is, the lower the precision (higher values of %). This is noticeable on the scatter plot (Figure 5d), where relative precision can reach values above 10% for young individuals. Error bar chart (Figure 5f) shows consistency between age classes but confirms higher imprecision for individuals in the <20 and [20;30] classes and the large dispersion for individual >20.

Regarding the absolute accuracy, the relationship between automated count accuracy and known age is represented by the scatterplot diagram (Figure 4c). Likewise, to human counts, this graphic representation clearly shows that for younger individuals (lower part of the scatterplot), points are stacked along the vertical axis passing through the origin of the x-axis. This diagram also clearly displays that, for older individuals, points spread to the left toward the negative values, indicating underestimation of the estimation. The Pearson correlation test between chronological age and absolute accuracy shows a negative linear relationship ($r = -0.430$; $p > 0.001$; $n = 200$) and confirms, like for the human counts, that the accuracy of the estimates is related to the age of the individual. Graphical representation of 95% CI (Figure 4e) shows a clear decrease of the absolute accuracy for individuals over 60. Higher accuracy for subjects older than 90 is most likely due to underrepresentation of these individuals in the sample.

The diagram illustrating the relationship between the relative accuracy and chronological age estimates (Figure 4d) shows a densification of the scatterplot toward low values. As with the human counts, older individuals are characterized by high values of percentage errors. The Pearson correlation coefficient indicates a significant linear relationship between these two variables ($r = 0.221$; $p > 0.001$; $n = 200$) and confirms that the older the individual is, the larger the percentage error of the estimate. Error bar chart of relative accuracy (Figure 4f) reveals a level of consistency between age classes. The higher inaccuracy for subjects in the [80;90] class may be due to the overrepresentation of individual aged between 80 and 90 (Figure 1).

3.5 | Validation of the semiautomated method

The purpose of this section is to determine whether the algorithms implemented by the software can differentiate and count acellular cementum annuli with similar performances to the human expert. The whole dental sample ($n = 200$) was considered for this assessment, and the 101 annuli counts analyzed in Visilog on an ROI selected by the operator for each tooth ($n = 20,200$) were compared to the five annuli counts ($n = 975$) performed by the operator on the same teeth.

In terms of absolute precision, the counting phase is characterized by a mean precision of 4.85 annuli for an operator (Bertrand et al., 2019) and 3.58 for the software (Table 2). Similarly, the relative precision is also better for the software (7.93%) than for the analyst (10.13%). The automated system reveals a higher performance, but this observation must be qualified since the operator conducted his counts on five cross-sections for each tooth (involving variability of cementum tissue along the root high), whereas the software was implemented on a single micrograph acquired from a selected slide per tooth.

The mean inaccuracy is greater for the automated approach (absolute accuracy: −5.44 years and relative accuracy: 11.24% on average) than for the manual approach (absolute accuracy: −4.53 years and relative accuracy: 9.19%). This discrepancy highlights the constancy of the reading of the software regardless of the age

TABLE 2 Comparison of performance indicators between a human expert and semiautomated estimates based on cementum annuli counts

|			N	Mean	Min.	Max.	IC-95% Lower bound	Upper bound
Human	Precision	Absolute (σ)	199	4.85	0.00	12.8	4.45	5.25
		Relative (CV)	199	10.13	0.00	28.76	9.4	10.86
	Accuracy	Absolute (Δage)	200	−4.53	−38.5	15.3	−5.58	−3.48
		Relative (Δ%age)	200	9.19	0.13	44.25	8.09	10.29
Software	Precision	Absolute (σ)	200	3.58	1.34	6.42	3.45	3.72
		Relative (CV)	200	7.93	4.74	31.97	7.55	8.31
	Accuracy	Absolute (Δage)	200	−5.44	−51.75	16.48	−7.01	−3.87
		Relative (Δ%age)	200	11.24	0.2	54.9	9.54	12.93
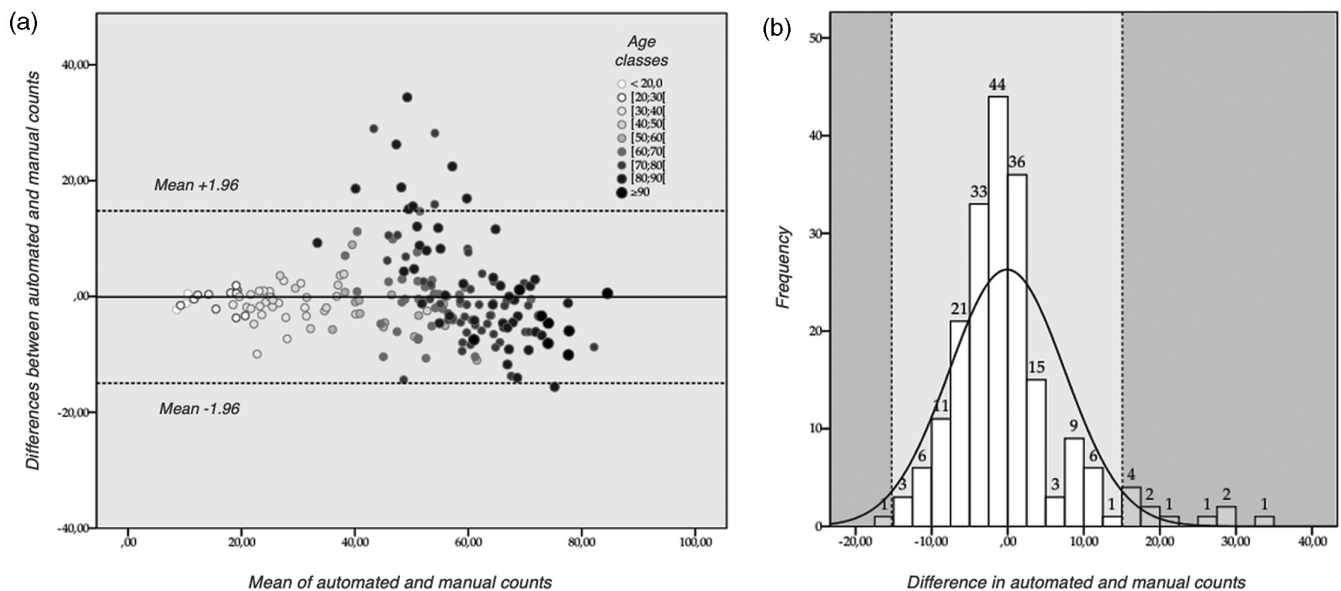
FIGURE 6   Bland–Altman plot (a) and accompanying histogram (b) showing the distribution of mean differences between automated and manual counts

classes or regions of interest and the adaptive capacity of the operator.

Regarding the level of agreement, the 95% agreement interval of the difference between the manual and semiautomated methods was represented on a Bland–Altman plot (Bland & Altman, 1986) (Figure 6a,b). Most of the differences between the two methods lie between the 2σ limits. Ninety-four percent of the individuals ($n = 188/200$) fall within this interval. Some points outside this agreement interval reveal a strong disagreement. These strong divergences are exclusively above the upper limit. A positive value of this difference indicates higher values of the manual counts. This can be interpreted as a higher sensitivity of the operator or software threshold parameters set for a lower sensitivity. It is interesting to note that these major disagreements exclusively concern older individuals in the [70;80] and [80;90] classes.

Note also that both techniques can count the number of annuli similarly on these older individuals. The Bland–Altman representation does not state whether the agreement is appropriate to adopt the semiautomated method but only defines the interval of agreement and quantifies the bias and range of agreement in which 95% of the differences are included. A Kolmogorov–Smirnov test performed on the distribution of differences between the two techniques demonstrates that the distribution of these differences does not follow a normal distribution ($p > 0.001$). Hence, even low, the bias can be statistically considered significant.

4 | DISCUSSION AND CONCLUSION

Based on the known-age dental sample ($n = 200$), 20,200 counts performed by the software (101 counts/tooth) were compared to

975 human counts (5 counts/tooth), and performances were confronted using precision and accuracy and interobservers indicators.

To counts performed by the observer, the automatization of annuli counts demonstrated no considerable variation in repeated measurements with an interval of 6 months. This was a prerequisite for comparing the two methods. The intraclass correlation coefficients and the graphical approach demonstrated suitable agreement between automated and human approaches for estimating age at death. Nevertheless, the automated approach generated outliers that, although they represent only 4% of the readings, have a substantial impact on users' confidence. We have not found a rational explanation to justify these discrepancies, and the origin of inconsistencies may be related to tissue variability, inappropriate settings, or software instability.

The mean precision for automated counts has been estimated to be 3.6 annuli and ranges from 1.3 to 6.4. This mean precision reaches 4.9 annuli for the operator, but the interval is twice as broad and reaches 12.8 annuli. Regarding only the absolute precision, the software appears to be more effective. Regarding relative precision, the software values range from 4.7% to 32% and are similar to that of the operator, which reaches 29%. Globally, the advantage is also given to the software, with a relative precision of 7.9% against 10.1% for manual technique. The higher precision achieved by the software must be qualified. Due to methodological choice whose purpose was to move forward on the simplification of the protocol if a software is adopted, human counts performed on different cross-sections could explain greater variation. Thus, the operator's imprecision could integrate greater biological variability along the dental root height. Future development could fruitfully explore this issue further by processing automated counts on different cross sections of the same tooth. Adapting this procedure would also overcome some taphonomic

obstacles that render some preparations hardly usable for software applications.

Interestingly, the accuracy of the automated approach is established to be −5.4 years and yielded less accurate age estimates than the human (−4.5 years). In both methods, the absolute accuracy covers wide ranges. The relative accuracy is also comparable for both approaches (11.2% for the software and 9.2% for the manual) but again, the software led to somewhat less accurate estimates than human.

Based on our observations, the software performs more precise but less accurate counts compared with the human. This result which might seem paradoxical demonstrates that although computer-assisted estimations are more consistent in the annuli counts, the dental cementum patterns, often heterogeneous in distribution and optically anisotropic, can make automated image analysis vulnerable to irregularities and therefore wind up having somewhat greater age estimation error.

Although more precise, the software does not avoid the trajectory effect. For both approaches, the variance in cementum annuli counts increases with increasing age. The relative imprecision of the software shows variation in subjects under the age of 30. We suspect that specific taphonomic conditions, such as infiltrations recurrently seen on cross sections derived from WWI soldiers, might be responsible for this variation in disturbing annuli detection (Figure 7). We believe that the human eye can adapt and perceives more efficiently the cementum alternating structure despite the staining due to iron oxide infiltration.

The comparison between the two techniques shows that 94% of the estimates are within the agreement interval. In other words, 6% of the counts represent a disagreement, and if we base our reasoning on a significance level set at 5%, this proportion is mathematically significant. Interestingly, major disagreements concern individuals over 60 years old. It is then difficult not to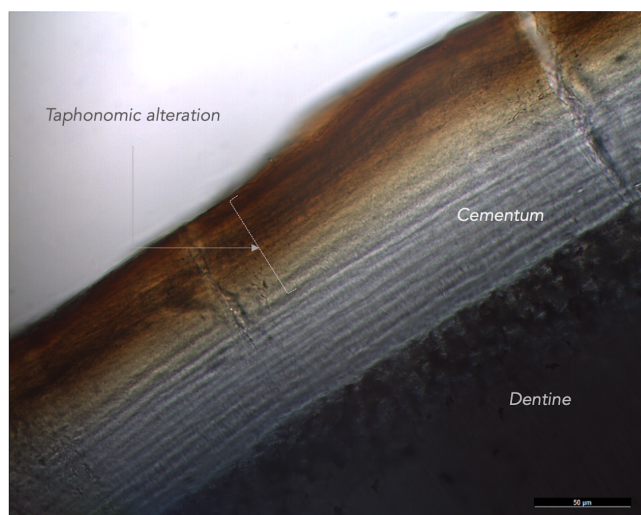 link this observation to the age-related decreasing distinctness of cementum annuli. It should also be noted that cementochronology is a two-tier performance method and that a deterioration of performance for individuals over 60 already has been demonstrated. However, these discrepancies do not undermine the use of the software on individuals over the age of 60 since majority of them are within the agreement interval.

The subjectivity of annuli interpretation and counts, which may appear to some practitioners more of an art than science, is often advanced as a justification for designing an automated system. The computerization of data acquisition gives the impression of greater objectivity, but we must keep in mind that the system is only semiautomated and that the region of interest and profile locations are selected manually. Moreover, the operator always conducts visual control of the detected annuli along the generated segments and checks, accepts or adjusts the detection parameters. Subjectivity is always perceived as a bias but can also be seen as an asset. Visual perception is indeed a complex process for which a retinal image, even if incomplete, is not an obstacle for a human expert. The eye can identify a cementum annulus, ignore artifacts, search for discontinuous cementum lines and complete discontinuities that the software is unable to interpret, and adapt. Advances in analytical tools such as machine learning already applied in sclerochronology (Hoekendijk et al., 2021; Moen et al., 2018) and improvement of open-source plugins for analyzing micrographs may represent a precious perspective for future applications.

The analytical tool tested here does not really limits human intervention but assign new tasks to the operator. The use of a software shifts subjectivity from the annuli identification and counting action to the ROI and parameters selection. Furthermore, a visual inspection is regularly required to check the adequacy and effectiveness of selected settings. In this respect, the adoption of an automated system is not entirely objective and may not represent a time-saving solution. Nonetheless, automated image analysis allows substantial increases in the numbers of counts and investigated samples. Hence, even though we cannot recommend the adoption of image analysis for case studies where human counts are faster and already beset with difficulties, we believe that it is certainly for population-based studies that such an approach can reveal its potential.

FIGURE 7 Transverse section of a canine extracted from one of the WWI soldiers showing taphonomic alteration (iron oxide infiltration). The scale is 50 μm

AUTHOR CONTRIBUTIONS
Benoit Bertrand: Conceptualization (lead); formal analysis (lead); funding acquisition (lead); investigation (lead); methodology (lead); writing – original draft (lead). **Martine Vercauteren:** Methodology (supporting); project administration (supporting). **Eugenia Cunha:** Conceptualization (supporting); investigation (supporting); project administration (lead); resources (equal). **Anne Bécart:** Conceptualization (supporting); resources (equal). **Didier Gosset:** Project administration (lead). **Valery Hédouin:** Funding acquisition (lead); project administration (lead); resources (equal); supervision (lead).

ACKNOWLEDGMENTS

BERTRAND ET AL.

AMERICAN JOURNAL OF
BIOLOGICAL ANTHROPOLOGY
The Official Journal of the American Association of Biological Anthropologists

WILEY | 325

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this article.

DATA AVAILABILITY STATEMENT

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

ORCID

Benoit Bertrand 🔟 https://orcid.org/0000-0001-7640-6929

REFERENCES

AlQahtani, S. J., Hector, M. P., & Liversidge, H. M. (2010). Brief communication: The London atlas of human tooth development and eruption. *American Journal of Physical Anthropology*, 142(3), 481–490. https://doi.org/10.1002/ajpa.21258

Bertrand, B. (2013). Procedure ISO-9001—Cementochronologie, référence: MOP-Dap-04, Indice a.

Bertrand, B., Cunha, E., Bécart, A., Gosset, D., & Hédouin, V. (2019). Age at death estimation by cementochronology: Too precise to be true or too precise to be accurate? *American Journal of Physical Anthropology*, 169(3), 464–481. https://doi.org/10.1002/ajpa.23849

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 327, 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Bosshardt, D. D., & Schroeder, H. E. (1996). Cementogenesis reviewed: A comparison between human premolars and rodent molars. *Anatomical Record*, 245(2), 267–292. https://doi.org/10.1002/(SICI)1097-0185(199606)245:2<267::AID-AR12>3.0.CO;2-N

de Broucker, A., Colard, T., Penel, G., Blondiaux, J. J., & Naji, S. (2016). The impact of periodontal disease on cementochronology age estimation. *International Journal of Paleopathology*, 15, 128–133. https://doi.org/10.1016/j.ijpp.2015.09.004

Buikstra, J. E., & Ubelaker, D. H. (1994). Standards for data collection from human skeletal remains. In *Proceedings of a seminar at the Field Museum of natural history* (Vol. 68). Arkansas Archeological Survey.

Charles, D. K., Condon, K., Cheverud, J. M., & Buikstra, J. E. (1986). Cementum annulation and age determination in *Homo sapiens*. I. Tooth variability and observer error. *American Journal of Physical Anthropology*, 71(3), 311–320. https://doi.org/10.1002/ajpa.1330710306

Colard, T., Bertrand, B., Naji, S., Delannoy, Y., & Bécart, A. (2015). Toward the adoption of cementochronology in forensic context. *International Journal of Legal Medicine*, 132(4), 1117–1124. https://doi.org/10.1007/s00414-015-1172-8

Condon, K., Charles, D. K., Cheverud, J. M., & Buikstra, J. E. (1986). Cementum annulation and age determination in *Homo sapiens*. II. Estimates and accuracy. *American Journal of Physical Anthropology*, 71(3), 321–330. https://doi.org/10.1002/ajpa.1330710307

Cunha, E., Baccino, E., Martrille, L., Ramsthaler, F., Prieto, J., Schuliar, Y., Lynnerup, N., & Cattaneo, C. (2009). The problem of aging human remains and living individuals: A review. *Forensic Science International*, 193(1–3), 1–13. https://doi.org/10.1016/j.forsciint.2009.09.008

Czermak, A., Czermak, A., Ernst, H., & Grupe, G. (2006). A new method for the automated age-at-death evaluation by tooth-cementum annulation (TCA). *Anthropologischer Anzeiger; Bericht über die biologisch-anthropologische Literatur*, 64(1), 25–40. http://www.ncbi.nlm.nih.gov/pubmed/16623086

Dias, P. E. M., Beaini, T. L., & Melani, R. F. H. (2010). Age estimation from dental cementum incremental lines and periodontal disease. *Journal of Forensic Odonto-Stomatology*, 28(1), 13–21. http://www.ncbi.nlm.nih.gov/pubmed/21239858

Elie, C., & Colombet, I. (2011). Méthodes d'estimation de la reproductibilité. *Sang Thrombose Vaisseaux*, 23(3), 138–145. https://doi.org/10.1684/stv.2011.0592

Fablet, R., & Le Josse, N. (2005). Automated fish age estimation from otolith images using statistical learning. *Fisheries Research*, 72(2–3), 279–290. https://doi.org/10.1016/j.fishres.2004.10.008

Ferreira, M. T., Coelho, C., Makhoul, C., Navega, D., Gonçalves, D., Cunha, E., & Curate, F. (2020). New data about the 21st century identified skeletal collection (University of Coimbra, Portugal). *International Journal of Legal Medicine.*, 135, 1087–1094. https://doi.org/10.1007/s00414-020-02399-6

Ferreira, M. T., Vicente, R., Navega, D., Gonçalves, D., Curate, F., & Cunha, E. (2014). A new forensic collection housed at the University of Coimbra, Portugal: The 21st century identified skeletal collection. *Forensic Science International*, 245, 202.e1–202.e5. https://doi.org/10.1016/j.forsciint.2014.09.021

Fisher, M., & Hunter, E. (2018). Digital imaging techniques in otolith data capture, analysis and interpretation. *Marine Ecology Progress Series*, 598(July), 213–231. https://doi.org/10.3354/meps12531

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica*, 25(2), 141–151. https://doi.org/10.11613/BM.2013.003

Gocha, T. P., & Schutkowski, H. (2012). Tooth cementum annulation for estimation of age-at-death in thermally altered remains. *Journal of Forensic Sciences*, 58(Suppl. 1), 1–5. https://doi.org/10.1111/1556-4029.12023

Gualdi-russo, E., Saguto, I., Frisoni, P., Neri, M., Mongillo, J., & Rinaldo, N. (2022). Age estimation using tooth cementum annulations: Bias and sources of inaccuracy. *Age estimation using tooth cementum annulations: bias and sources of inaccuracy.*, 27(5), 141.

Hoekendijk, J. P. A., Kellenberger, B., Aarts, G., Brasseur, S., Poiesz, S. S. H., & Tuia, D. (2021). Counting using deep learning regression gives value to ecological surveys. *Scientific Reports*, 11(1), 1–12. https://doi.org/10.1038/s41598-021-02387-9

Kagerer, P., & Grupe, G. (2001). On the validity of individual age-at-death diagnosis by incremental line counts in human dental cementum. Technical considerations. *Anthropologischer Anzeiger; Bericht Über Die Biologisch-Anthropologische Literatur*, 59(4), 331–342. http://www.ncbi.nlm.nih.gov/pubmed/11838046

Klauenberg, K., & Lagona, F. (2007). Hidden Markov random field models for TCA image analysis. *Computational Statistics & Data Analysis*, 52(2), 855–868. https://doi.org/10.1016/j.csda.2007.06.022

Meinl, A., Huber, C. D., Tangl, S., Gruber, G. M., Teschler-Nicola, M., & Watzek, G. (2008). Comparison of the validity of three dental methods for the estimation of age at death. *Forensic Science International*, 178(2–3), 96–105. https://doi.org/10.1016/j.forsciint.2008.02.008

Moen, E., Handegard, N. O., Allken, V., Albert, O. T., Harbitz, A., & Malde, K. (2018). Automatic interpretation of otoliths using deep learning. *PLoS One*, 13(12), e0204713. https://doi.org/10.1371/journal.pone.0204713

Orban, R., Eldridge, J., & Polet, C. (2011). Potentialités et historique de la collection de squelettes identifiés de Schoten (Belgique, 1837-1931). *Anthropologica et Praehistorica*, 122(1), 19–62.

Orban, R., & Vandoorne, K. (2006). Les squelettes humains de Koksijde (Coxyde) et Schoten: deux collections remarquables conservés à l'Institut royal des Sciences naturelles de Belgique. In Y. Ardagna, B. Bizot,

326 WILEY— AMERICAN JOURNAL OF BIOLOGICAL ANTHROPOLOGY
The Official Journal of the American Association of Biological Anthropologists

BERTRAND ET AL.

G. Boetsch, & X. Delestre (Eds.), *Les collections ostéologiques humaines: gestion, valorisation et perspectives, Table ronde (25–26 avril 2003)* (pp. 79–84). Association Provence Archéologie.

Petrovic, B., Pantelinac, J., Capo, I., Miljkovic, D., Popovic, M., Penezic, K., & Stefanovic, S. (2021). Using histological staining techniques to improve visualization and interpretability of tooth cementum annulation analysis. *International Journal of Morphology*, *39*(1), 216–221. https://doi.org/10.4067/S0717-95022021000100216

Schroeder, H. E. (1986). *The periodontium*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-71261-6

Sequeira, P., Bosshardt, D. D., & Schroeder, H. E. (1992). Growth of acellular extrinsec fiber cementum (AEFC) and density of inserting fibers in human premolars of adolescents. *Journal of Periodontal Research*, *27*(2), 134–142. http://www.ncbi.nlm.nih.gov/pubmed/1532202

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Stott, G. G., Sis, R. F., & Levy, B. M. (1982). Cemental annulation as an age criterion in forensic dentistry. *Journal of Dental Research*, *61*(6), 814–817. https://doi.org/10.1177/00220345820610063401

Troadec, H., & Benzinou, A. (2002). Estimation de l'âge assistée par ordinateur. In J. Panfili, H. de Pontual, H. Troadec, & P. J. Wright (Eds.), *Manuel de sclérochronologie des poissons* IRD Éditio (pp. 199–242). OpenEdition Books. https://doi.org/10.4000/books.irdeditions.20901

Wittwer-Backofen, U., & Buba, H. (2002). Age estimation by tooth cementum annulation: Perspectives of a new validation study. In R. D. Hoppa & J. W. Vaupel (Eds.), *Paleodemography: Age distributions from skeletal samples* (Vol. 44, pp. 107–128). Cambridge University Press. https://doi.org/10.1016/S0047-2484(03)00046-0

Wittwer-Backofen, U., Gampe, J., & Vaupel, J. W. (2004). Tooth cementum annulation for age estimation: Results from a large known-age validation study. *American Journal of Physical Anthropology*, *123*(2), 119–129. https://doi.org/10.1002/ajpa.10303

Yamamoto, T., Hasegawa, T., Yamamoto, T., Hongo, H., & Amizuka, N. (2016). Histology of human cementum: Its structure, function, and development. *Japanese Dental Science Review*, *52*(3), 63–74. https://doi.org/10.1016/j.jdsr.2016.04.002