archaeometry

Archaeometry ••, •• (2019) ••–•• doi: 10.1111/arcm.12480

ADDRESSING DATA COMPARABILITY IN THE CREATION OF COMBINED DATA SETS OF BIOAPATITE CARBON AND OXYGEN ISOTOPIC COMPOSITIONS*

L. A. CHESSON†
PAE, Defense POW/MIA Accounting Agency (DPAA) Laboratory, Joint Base Pearl Harbor-Hickam, HI, USA

M. W. KENYHERCZ
DPAA Laboratory, Joint Base Pearl Harbor-Hickam, HI, USA

L. A. REGAN
Office of Net Assessment, Department of Defense, Washington, DC, USA

and G. E. BERG
DPAA Laboratory, Joint Base Pearl Harbor-Hickam, HI, USA

Before amalgamating published isotope data, comparability should be demonstrated. This paper compares carbon and oxygen isotopic compositions of 30 enamel samples measured by two laboratories. The aims were to see what, if any, isotopic variation was observed, to determine the causes as needed and to correct if possible. Bioapatite was acidified at 90°C in 2006 and at 26°C in 2017, while δ values were corrected via one-point normalization in 2006 and by two-point normalization in 2017. One case (of the 30) produced different δ values between the analysis dates, suggesting contamination. Repeated carbon isotope ratio measurements were not meaningfully different. Repeated oxygen isotope ratio measurements were significantly different, even following correction for acid-carbonate fractionation at different temperatures and the renormalization of 2017 δ values using one point; however, differences were not meaningful for interpretations. Results were used to calculate real interpretative differences (RIDs) for comparing enamel bioapatite as 0.6‰ for $\delta^{13}C$ values and as 1.6‰ for $\delta^{18}O$ values.

KEYWORDS: ANTHROPOLOGY, ARCHAEOLOGY, FORENSIC SCIENCE, ISOTOPIC ANALYSIS, ENAMEL APATITE, SKELETAL REMAINS, AMALGAMATED DATA SETS

INTRODUCTION

Aim and objectives

The aim of this study was to define good practice when amalgamating and using published isotope data of human skeletal remains from the archaeological, anthropological and forensic literature. Of particular focus is a data set of dental remains from US Americans and East Asians published by Regan (2006), as it is relevant to geographical origin questions that the Defense POW/MIA Accounting Agency (DPAA) wants to answer in order to best serve the American

*Received 26 June 2018; accepted 3 April 2019
†Corresponding author: email lesley.chesson@pae.com

public. If the carbon and oxygen isotope data from Regan could be validated, then they could be used to apply testable hypotheses of geographical origin to dental remains; this would be similar to an approach developed previously based on isotopic compositions of bone bioapatite and collagen (Bartelink *et al*. 2014a). The good practice described here to assess the comparability of one specific set of published isotope data with contemporary samples analyses is applicable to any study in which multiple sources of isotope analysis results are amalgamated.

Background to human identification using isotopic analysis

Determining the origin of unidentified human remains is problematic when skeletal elements are few, fragmentary and/or degraded, making the determination of biological profiles difficult or impossible, and DNA analysis can be prohibitively expensive and time consuming. This challenge is particularly acute in the DPAA's efforts to identify US service members missing as a result of past conflicts where there may be high levels of fragmentation and degradation as a result of the circumstances of loss and the passage of time (Holland *et al*. 2008; Emanovsky and Belcher 2012). However, if the geographical region of origin for a set of remains can be established, it facilitates identification in two ways: first, it eliminates non-US service members from the identification pipeline; and second, it can potentially narrow the list of possible individuals to consider for identification based on homes of record of missing personnel.

One tool that is potentially useful in determining geographical origins of skeletal material is stable isotope analysis. Developed primarily in geochemical fields (Fogel *et al*. 1997), applications of the technique have had an important impact on anthropological and archaeological research, beginning with pioneering dietary studies in the late 1970s (e.g., DeNiro and Epstein 1976; Vogel and van der Merwe 1977; Gaffney *et al*. 1978; Lyon and Baxter 1978; van der Merwe and Vogel 1978). Today there are hundreds of additional publications on the isotope analysis of skeletal remains from across the globe (for details, see literature reviews, e.g., Bartelink *et al*. 2014b; Kimmerle and Kamenov 2015; and Chesson *et al*. 2018).

In order to make isotopic comparisons among humans of different geographical origin, isotope data sets of various populations are needed. Large population data sets are relatively scarce in the literature. Considering the publications made to date, a single research group typically generates all data published together as a set, using sample preparation and isotope analysis methods that may or may not be widely adopted by other laboratories. Published isotope data sets can be further divided by type of remains (e.g., bone, teeth) and material tested (e.g., bioapatite carbonate versus phosphate). To characterize a population, or multiple populations, isotopically, the output of many laboratories may be amalgamated.

Publication of isotope data from different human populations provides an opportunity to create amalgamated data sets (e.g., Pollard *et al*. 2011; Lightfoot and O'Connell 2016; Someda *et al*. 2016; Kamenov and Curtis 2017). Unfortunately, adequate assessment of comparability is not always undertaken before isotope data from the anthropological, archaeological and/or forensic literature are directly compiled. A recent study by Pestle *et al*. (2014) on pieces of the same archaeological bone provided to multiple laboratories found that both sample preparation method and the isotope analysis technique significantly impacted the observed isotopic variation of bone collagen and bioapatite. The authors noted, 'Analytical results from different laboratories might not be directly comparable' (15), especially in the case of bioapatite oxygen isotopic compositions.

Comparability of data is particularly relevant in forensic applications of isotope analysis, as results of data comparisons may be used as evidence in courts of law, or for the identification of an unknown person. In the United States, the admissibility of scientific evidence, such as

isotope analysis data, has been governed since 1923 by legal precedent in the case of *Frye v. United States*, which determined that evidence could be used in the court when it had been generally accepted by the scientific community (Bell 2009; Cerling *et al*. 2016). Further, the Federal Rules of Evidence 702 convey definitions and modes of admission of scientific evidence (Komar and Buikstra 2008). The 1993 case of *Daubert v. Merrell Dow Pharmaceuticals* imparted more responsibility on judges to be the gatekeepers for the admission of scientific evidence in the courtroom. Under the Daubert standard, the judge decides on four major criteria including: if the theory or technique can be or has been tested; if it has been subjected to peer review and publication; if the known or potential rate of error is used; and the degree of the method's acceptance within the relevant scientific community (Komar and Buikstra 2008; Bell 2009).

Selected examples of isotope data compilations

Examples of the aforementioned concerns on isotope data comparability as related to amalgamations of human skeletal remains can be annotated. The selected studies discussed here may not be germane only to this study. They highlight additionally the fact that isotope data compilations are found in a variety of research fields, from archaeology to forensics, and all points in between.

Pollard *et al*. (2011) published a calibration data set to convert dental enamel phosphate $\delta^{18}O$ values to drinking water $\delta^{18}O$ values and to investigate human mobility, despite the fact that 'there must be some doubts regarding the validity of such a combined data set given the differences in the preparation and measurement protocols' (501). Recently, Lightfoot and O'Connell (2016) published a method for identifying human migrants in the archaeological record. They compiled biomineral oxygen isotope data that included tooth enamel carbonate and phosphate plus bone carbonate and phosphate. Examination by site required the combination of results from multiple research groups, without explicit regard to the methods used to prepare and isotopically analyse samples (e.g., the Kaminaljuyu site; Lightfoot and O'Connell 2016, 21). The compiled data were used to classify migrants via the identification of 'outlier' δ values; however, interlaboratory differences may have created the appearance of outliers that were instead artefacts caused by differences in sample preparation and analysis.

As noted previously, Regan (2006) published a large data set of the carbon and oxygen isotopic compositions of dental remains collected from US Americans and East Asians. That data set was used in the investigation of the natal origin of a tooth recovered from a Vietnam War-era plane crash (Holland *et al*. 2012). The tooth collected by Holland *et al*. (2012) was analysed in the same laboratory as samples in the original data set (Regan 2006), which would seemingly provide a straightforward data comparison. What was not clear, however, was that the sample preparation and isotope analysis techniques at the laboratory changed in the intervening years, potentially adding uncertainty to the comparison. While the overall result is likely not in question, the actual δ values of the tooth submitted for isotope analysis by the Holland group might not be directly comparable with the Regan data set.

Someda *et al*. (2016) used the Regan (2006) data to discriminate statistically between US American and Japanese individuals using carbon and oxygen isotopic compositions of tooth enamel. While excellent discrimination between the populations was published (100%), it is difficult to determine the precision of the results. First, different sample preparation methods and isotope analysis techniques were used between the studies and are likely to result in real differences in measured δ values (Pestle *et al*. 2014). Second, at least 25 of the plotted cases are within ±1‰ of the discrimination line (Someda *et al*. 2016, fig. 2), and any slight isotopic variations due to laboratory procedures could create real differences in discrimination ability. Third, as the

authors rightly noted, the used US population is much younger than the Japanese population, and differences in diet over time could impact discriminatory results.

This study

A total of 30 dental samples from the original work of Regan (2006) were reanalysed using a different technique to measure the carbon and oxygen isotopic compositions of bioapatite carbonate. Isotope ratios measured in 2017 were compared with the published isotope analysis results from 2006. Isotopic variation between analysis dates was investigated and correction factors were applied, as needed and feasible. As a consequence, it was possible to estimate the amount of error that may be present within an amalgamated data set and recommend informed parameters (real interpretative differences—RIDs) by which to interpret measured $\delta^{13}C$ and $\delta^{18}O$ values of dental remains.

MATERIALS AND METHODS

Bioapatite preparation

Bioapatite was prepared from modern human teeth. The teeth were collected from donors ($n = 228$) at the US Air Force Academy (AFA) in Colorado Springs, Colorado, USA, as described in detail by Regan (2006). Third molars (M3) were selected. The teeth were first soaked in 3% H_2O_2 for two days, rinsed with tap water, physically scrubbed to remove surface contaminants and ultrasonically cleaned in deionized double-distilled water (DDH_2O) for 30 min. Adherent contaminants, such as alveolar bone remnants, were removed from clean, dry teeth using a dental drill and #8 carbide dental drill bur. Finally, approximately 100–200 mg of pristine enamel powder were drilled from each tooth. Enamel powder was transferred to labelled 1.5-ml microcentrifuge tubes.

In 2005/06, enamel powders were treated with 30% H_2O_2 for 24 h to remove organic contaminants. Approximately 1 ml H_2O_2 was added to each microcentrifuge tube. The powder and hydrogen peroxide were mixed, and the tubes were then uncapped. After 24 h, the tubes were capped and centrifuged; the H_2O_2 was discarded. Samples were rinsed twice with DDH_2O. Samples were next treated with 0.1 N CH_3COOH for 30 min to remove secondary carbonates. Approximately 1 ml acetic acid was added to each microcentrifuge tube. The powder and acid were mixed, and the tubes were then uncapped. After 30 min, the tubes were capped and centrifuged; the CH_3COOH was discarded. Samples were rinsed twice with DDH_2O and placed in an oven to dry at 52°C for 4 days.

In 2017, 30 samples were selected from the original study for reanalysis. Selected samples covered the maximum range in carbon and oxygen isotopic compositions measured previously. The samples were again chemically cleaned. Samples were first treated in microcentrifuge tubes with 3% H_2O_2 for 15 min. Oxidized samples were rinsed three times with deionized water before being treated with 0.1 M CH_3COOH for 15 min. Acid-treated samples were rinsed three times with deionized water, then dried in an oven at 60°C overnight.

Isotopic analysis

Stable isotope analysis results of a sample are presented in standard 'delta' notation (as δ values) relative to the virtual material Vienna Peedee belemnite (VPDB) for both carbon and oxygen, following the International Union of Pure and Applied Chemistry (IUPAC) guidelines (Coplen 2011):

$$\delta^{13}C_{VPDB} = \frac{\left(^{13}C/^{12}C\right)_{sample}}{\left(^{13}C/^{12}C\right)_{VPDB}} - 1$$

$$\delta^{18}O_{VPDB} = \frac{\left(^{18}O/^{16}O\right)_{sample}}{\left(^{18}O/^{16}O\right)_{VPDB}} - 1$$

Since δ values are numerically very small, for convenience they are expressed in parts per thousand (‰).

In 2005/06, $\delta^{13}C$ and $\delta^{18}O$ values were measured in 13 analytical sequences ('runs') over a 6-month period using a Micromass PRISM Series II isotope ratio mass spectrometer, operated in dual-inlet mode and with an attached Isocarb common acid bath preparation device, at the Department of Geological Sciences, University of Florida (UF)(Gainesville, FL, USA). Samples were loaded (1.0–1.2 mg) into stainless steel 'boats'. Boats were dropped individually into a common acid bath of H_2PO_4 held at 90°C. The reaction time was 10 min; the resultant CO_2 was collected in a cold finger cooled with liquid nitrogen (LN_2) before further concentration in a second cold finger cooled with LN_2. A typical run included 36 samples and eight replicates of an international reference material (NBS 19).

Samples analysed in 2005/06 at the UF were traceable to the VPDB δ scale because they were analysed alongside the international carbonate reference material NBS 19. Measured $\delta^{13}C$ and $\delta^{18}O$ values of enamel bioapatite samples were normalized via an offset correction using NBS 19 with recommended $\delta^{13}C$ and $\delta^{18}O$ values of 1.95 and − 2.2‰, respectively. The standard deviation (SD) of all replicate analyses of NBS 19 over the 6-month period ($n = 83$) was 0.15‰ for $\delta^{13}C$ and 0.20‰ for $\delta^{18}O$ values.

Six samples were analysed in sextuplicate at the UF. Note that two of the samples analysed in replicate were teeth from the DPAA Laboratory; see Regan (2006) for a description of DPAA samples (CIL identifiers) analysed in the original study. The SDs of sextuplicate $\delta^{13}C$ measurements of CIL-010, CIL-045, AFA-038, AFA-095, AFA-151 and AFA-194 were 0.14, 0.08, 0.16, 0.10, 0.05 and 0.04‰, respectively. The SDs of sextuplicate $\delta^{18}O$ measurements of CIL-010, CIL-045, AFA-038, AFA-095, AFA-151 and AFA-194 were 0.22, 0.23, 0.29, 0.28, 0.20 and 0.09‰, respectively. One replicate analysis of AFA-151 was characterized by a low CO_2 gas yield; the δ values of that analysis were not used in the calculation of SDs.

In 2017, $\delta^{13}C$ and $\delta^{18}O$ values of 30 select AFA samples were measured in a single run using a Finnigan™ MAT 253 isotope ratio mass spectrometer operated in continuous flow mode and attached to a Thermo Scientific GasBench II, at IsoForensics, Inc. (IF) (Salt Lake City, UT, USA). Samples were loaded (1.1–1.2 mg) into 12-ml Exetainer® vials (Labco Ltd, Lampeter, UK), which were flushed with He. Five drops of 105% H_3PO_4, equivalent to 54 ± 3 mg at 20°C with a density of about 1.92 g ml^{-1}, were added to flushed vials. Carbonate was acidified for 24 h at 26°C before the resultant CO_2 was collected via an autosampler for measurement. The single run included all 30 samples, two analysed in duplicate (AFA-099 and AFA-156), and four replicates each of three international reference materials.

Samples analysed in 2017 at IF were traceable to the VPDB δ scale because they were analysed alongside the international reference materials NBS 18, NBS 19 and LSVEC lithium carbonate. Measured $\delta^{13}C$ values of enamel bioapatite samples were normalized via a stretch and shift correction (Dunn and Carter 2018) using NBS 19 and LSVEC, with recommended

$\delta^{13}C$ values of 1.95 and −46.6‰, respectively. Additional reference materials were analysed alongside samples for quality control (QC) and to monitor long-term instrument stability, including USGS44 ($n = 2$) and two in-house materials, a calcium carbonate (CC-1, $n = 4$) and a powdered marble (CC-6, $n = 4$). NBS 18 and the QC materials were used to calculate standard combined uncertainty (u_c), which includes uncertainty from calibration of the normalization reference materials, measurement of those normalization reference materials, normalization itself and the long-term SD of the QC material of interest (Dunn *et al*. 2015).

The maximum u_c of NBS 18 was 0.04‰; its average $\delta^{13}C$ value within the run was −5.04‰ as compared with its recommended $\delta^{13}C$ value of −5.01‰. The maximum u_c of USGS44 was 0.04‰ and its average $\delta^{13}C$ value within the run was −42.17‰ as compared with its current published $\delta^{13}C$ value of −42.15‰. The maximum u_c was 0.04‰ for CC-1 and 0.05‰ for CC-6. The average $\delta^{13}C$ value within the run for CC-1 was −14.02‰ as compared with its in-house, long-term value of −14.07‰. The average $\delta^{13}C$ value within the run for CC-6 was 3.71‰ as compared with its in-house, long-term value of 3.68‰. The SD of duplicate measurements of AFA-099 was 0.06‰; for AFA-156 it was 0.03‰.

Measured $\delta^{18}O$ values of enamel bioapatite samples were normalized to the VPDB δ scale via a stretch and shift correction using NBS 18 and NBS 19, with recommended $\delta^{18}O$ values of −23.01 and −2.2‰, respectively. Materials used for QC purposes included LSVEC, CC-1 and CC-6. The maximum u_c of LSVEC was 0.30‰; its average $\delta^{18}O$ value within the run was −26.6‰ as compared with its recommended $\delta^{18}O$ value of −26.7‰. The maximum u_c was 0.29‰ for CC-1 and 0.15‰ for CC-6. The average $\delta^{18}O$ value within the run for CC-1 was −25.27‰ as compared with its in-house, long-term value of −25.42‰. The average $\delta^{18}O$ value within the run for CC-6 was −3.36‰ as compared with its in-house, long-term value of −3.52‰. The SD of duplicate measurements of AFA-099 was 0.09‰; for AFA-156 it was 0.05‰.

Statistical analysis

All statistical analyses were performed using Prism for Mac OS X (Version 6.0f). The $\delta^{13}C$ and $\delta^{18}O$ values were tested for normality using the Shapiro–Wilk normality test. Neither the $\delta^{13}C$ nor the $\delta^{18}O$ values measured in either 2005/06 or 2017 were significantly different from a normal distribution. Thus, paired *t*-tests were used to test repeated measures of samples between analysis dates. Correlation between δ values was tested using a Pearson product-moment correlation coefficient.

RESULTS AND DISCUSSION

Results for the isotopic analysis of tooth enamel bioapatite carbonate are presented in Table 1. Henceforth, the 2005/06 analyses are referred to as 2006, for the original publication date.

One sample out of 30 produced highly different δ values between 2006 and 2017: AFA-005 ($\Delta = -1.48$‰ and −2.89‰ for $\delta^{13}C$ and $\delta^{18}O$ values, respectively; Table 1). When examined as $\delta^{13}C$ and $\delta^{18}O$ values—and not as Δ—neither the 2006 nor the 2017 analysis of AFA-005 was an outlier in its respective data set, falling within the range of other measured $\delta^{13}C$ and $\delta^{18}O$ values. Thus, there would be no way to identify easily AFA-005 as an outlier *except* via the comparison of 2006 and 2017 data. One possible explanation for the large differences in δ values in terms of both carbon and oxygen is the contamination of the sample during storage

Table 1 *Results of the isotopic analysis of tooth enamel bioapatite carbonate in 2006 and again in 2017*

	2006 analysis		*2017 analysis*		*Difference (Δ*)*	
Identifier	$\delta^{13}C_{VPDB}$	$\delta^{18}O_{VPDB}$	$\delta^{13}C_{VPDB}$	$\delta^{18}O_{VPDB}$	$\delta^{13}C_{VPDB}$	$\delta^{18}O_{VPDB}$
AFA-005	−11.10	−9.22	−9.62	−6.33	**−1.48**	**−2.89**
AFA-022	−10.10	−3.14	−10.08	−2.93	**−0.02**	**−0.21**
AFA-029	−9.73	−5.75	−9.67	−4.96	**−0.06**	**−0.79**
AFA-031	−10.87	−9.52	−10.76	−9.19	**−0.11**	**−0.33**
AFA-032	−12.52	−12.40	−12.64	−12.19	**0.12**	**−0.21**
AFA-057	−9.61	−8.48	−9.61	−7.77	**0.00**	**−0.71**
AFA-063	−11.23	−12.57	−11.08	−11.12	**−0.15**	**−1.45**
AFA-065	−10.17	−6.79	−10.10	−5.69	**−0.07**	**−1.10**
AFA-086	−10.43	−5.52	−10.62	−4.57	**0.19**	**−0.95**
AFA-099	−8.72	−9.88	−8.62	−9.18	**−0.10**	**−0.70**
AFA-117	−11.06	−8.51	−11.12	−7.58	**0.06**	**−0.93**
AFA-118	−7.77	−6.02	−7.77	−5.24	**0.00**	**−0.78**
AFA-121	−10.59	−11.20	−10.76	−10.91	**0.17**	**−0.29**
AFA-124	−8.33	−6.40	−8.26	−5.74	**−0.07**	**−0.66**
AFA-136	−11.76	−11.46	−11.77	−10.50	**0.01**	**−0.96**
AFA-148	−12.44	−7.34	−12.49	−6.30	**0.05**	**−1.04**
AFA-156	−10.88	−4.11	−11.00	−3.05	**0.12**	**−1.06**
AFA-160	−9.77	−4.13	−9.83	−3.22	**0.06**	**−0.91**
AFA-161	−9.56	−8.41	−9.63	−7.49	**0.07**	**−0.92**
AFA-166	−10.07	−10.18	−10.05	−9.14	**−0.02**	**−1.04**
AFA-168	−11.09	−7.49	−11.19	−6.60	**0.10**	**−0.89**
AFA-172	−9.89	−4.76	−9.91	−3.78	**0.02**	**−0.98**
AFA-184	−12.69	−6.37	−12.81	−5.56	**0.12**	**−0.81**
AFA-205	−7.98	−4.82	−7.95	−3.69	**−0.03**	**−1.13**
AFA-212	−10.57	−10.71	−10.79	−10.28	**0.22**	**−0.43**
AFA-220	−12.88	−4.61	−13.28	−3.68	**0.40**	**−0.93**
AFA-225	−8.03	−7.14	−7.94	−6.55	**−0.09**	**−0.59**
AFA-226	−10.10	−11.60	−10.18	−11.08	**0.08**	**−0.52**
AFA-254	−9.56	−5.75	−9.71	−5.20	**0.15**	**−0.55**
AFA-274	−9.88	−3.75	−10.05	−2.85	**0.17**	**−0.90**
Average (without outlier)					0.05	−0.79
SD (without outlier)					0.12	0.30

Notes: All δ values and differences (Δ) are presented in ‰.

*Differences (Δ) are calculated as the 2006 δ value minus the 2017 δ value. Underlined Δ are outliers within the data set.

or when prepared and/or weighed for analysis in 2017, which impacted the measured δ values of that sample, but no other sample analysed at the same time. It should be noted that AFA-022 was originally analysed in the same run as AFA-005, but no unusually large isotopic differences were observed for AFA-022 between the two analysis dates. The highly different δ values observed for a single case suggests that the inherent error rate between the two analyses is 1/30 or approximately 3%. Therefore, when using these data in future, it should be expected that 3% of the cases could have highly different δ values, but those values will likely fall within the range of the observed isotopic variation for the overall data set.

Carbon isotope analysis results

Omitting AFA-005 from consideration, differences between $\delta^{13}C$ values measured in 2006 and 2017 ranged from 0.40‰ to −0.15‰, with a mean Δ (± SD) of 0.05‰ (± 0.12‰) (Table 1). The correlation between 2006 and 2017 $\delta^{13}C$ values was $r = 0.9973$ ($p < 0.0001$; Fig. 1). A paired *t*-test of $\delta^{13}C$ values showed a significant difference between the two analysis dates at $\alpha = 0.05$, but not $\alpha = 0.01$ ($t = 2.158$, $p = 0.0397$). The significant but small mean difference between the analysis dates may be an effect of normalizing to the VPDB δ scale using two points in 2017 versus one point in 2006 (Meier-Augenstein and Schimmelmann 2019). While it is impossible to add reference materials to the 2006 runs, it is possible to normalize the 2017 measurement results using only NBS 19. A paired *t*-test of $\delta^{13}C$ values showed no significant difference between the two analysis dates when 2017 δ values were renormalized via an offset correction using NBS 19 ($t = 0.235$, $p = 0.8156$); data are shown in Figure 2.

Oxygen isotope analysis results

Omitting AFA-005 from consideration, differences ranged from −0.21‰ to −1.45‰ for $\delta^{18}O$ values, with a mean Δ (± SD) of −0.79‰ (± 0.30‰) (Table 1). The correlation between 2006 and 2017 $\delta^{18}O$ values was $r = 0.9948$ ($p < 0.0001$; Fig. 3). A paired *t*-test of $\delta^{18}O$ values showed a significant difference between the two analysis dates ($t = 14.17$, $p < 0.0001$).

As demonstrated by Pestle *et al*. (2014), both preparation and analysis can significantly impact oxygen isotopic variability in bioapatite samples. For example, the use of sodium hypochlorite (NaOCl, or 'bleach') versus H_2O_2 in sample treatment affects measured $\delta^{18}O$ values of bioapatite. In this study, enamel powders were treated with hydrogen peroxide before analysis in both 2006 and 2017; in other words, preparation was similar between the analysis dates. Thus, variation in analysis technique is the more likely cause of the significant difference found between $\delta^{18}O$ values measured in 2006 and 2017.

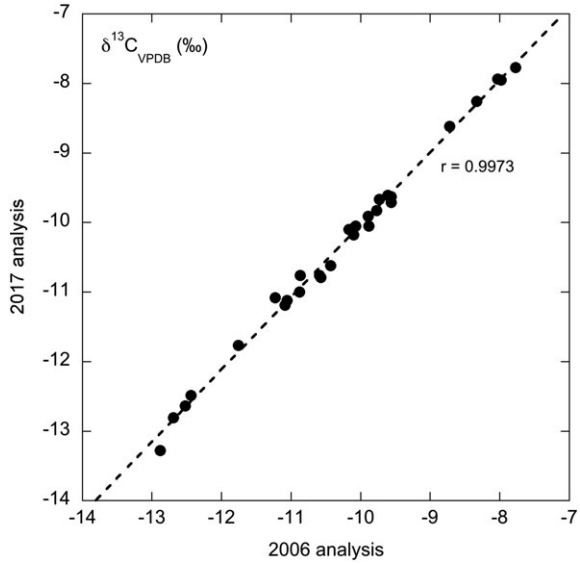

Figure 1 *Correlation of $\delta^{13}C$ values of 29 tooth enamel bioapatite samples ($r = 0.997$), analysed at two laboratories ≥ 10 years apart.*

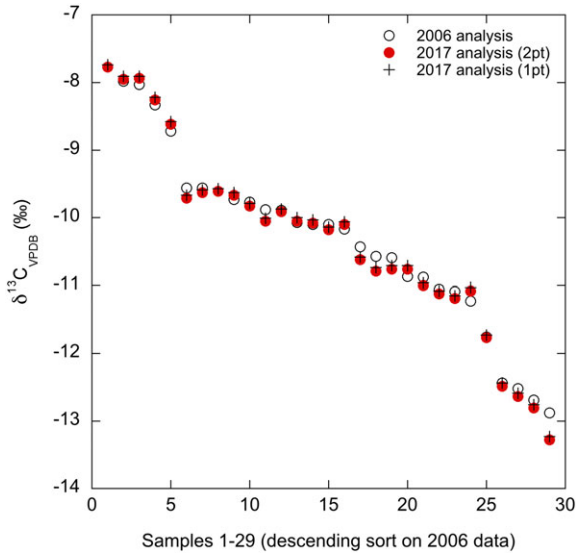

Figure 2 *Comparison of $\delta^{13}C$ values of 29 tooth enamel bioapatite samples analysed in 2006 (open circles) and again in 2017. The 2017 data are normalized to the VPDB δ scale via a stretch and shift correction ('2 pt', closed circles) as well as an offset correction ('1 pt', pluses). Samples are sorted by 2006 δ values and assigned identifiers from 1 to 29. [Colour figure can be viewed at wileyonlinelibrary.com]*

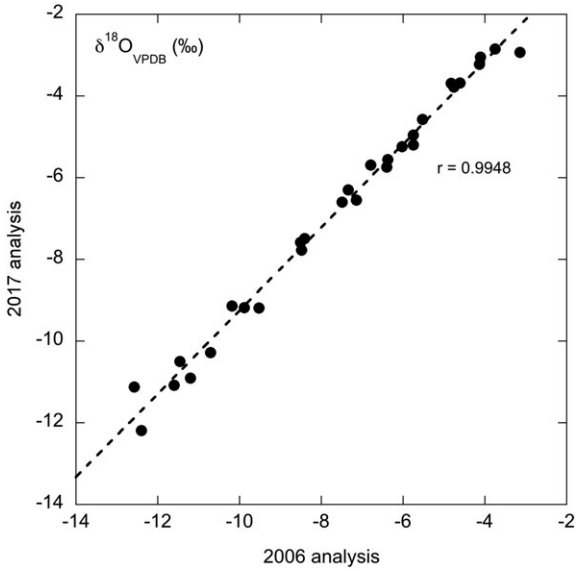

Figure 3 *Correlation of $\delta^{18}O$ values of 29 tooth enamel bioapatite samples (r = 0.995), analysed at two laboratories ≥ 10 years apart.*

In the isotope analysis of carbonate as CO_2, oxygen is partitioned into more than one product during acid digestion:

$$-CO_3^{2-} + 2H^+ \rightarrow CO_2 + H_2O$$

There is a temperature-dependent isotopic fractionation between oxygen in the measured product gas (CO_2) and the carbonated mineral (Passey *et al*. 2007). That fractionation is different between inorganic minerals, such as the pure carbonate reference material NBS 19, and biominerals, such as bioapatite, and even between 'modern' and archaeological enamel bioapatite (Passey *et al*. 2007; Kusaka and Nakano 2014). At 25°C, the apparent fractionation factor (α_{25}) is assumed to be the same for all carbonated minerals, regardless of whether they are pure carbonates or biominerals (Swart *et al*. 1991; Passey *et al*. 2007; Kusaka and Nakano 2014) and equal to 1.01025 (Friedman and O'Neil 1977). In the 2017 analysis, acidification took place at 26°C and the apparent fractionation can thus be considered the same for all sample and reference material measurements made within that run. In contrast, the acidification of bioapatite samples and the reference material NBS 19 in the 2006 runs took place at 90°C.

The apparent fractionation during acid reaction of carbonate was not the same for samples and reference materials at the elevated temperature used in 2006. To possibly correct for the temperature of acidification, fractionation factors were applied to 'raw' $\delta^{18}O$ values measured in 2006. The apparent fractionation factor for NBS 19 analysed at 90°C came from Passey *et al*. (2007, tab. 1): $\alpha_{90} = 1.00818$. For the apparent fractionation factor of the bioapatite samples, α_{90} was calculated following the recommendation of Passey *et al*. for modern tooth enamel: $\alpha_{90} = 1.00312 + 635/363.15^2 = 1.00794$ (range = 1.00747–1.00840; see Passey *et al*. 2007, eqn (3)). These acid-carbonate fractionation factors were applied before $\delta^{18}O$ values of samples were normalized to the VPDB δ scale via offset correction using NBS 19 (see the additional supporting information). The $\delta^{18}O$ values of the 2006 analysis thus changed from those originally published by Regan (2006); see α data in Figure 4 and the additional supporting information. Following α correction, a paired *t*-test of $\delta^{18}O$ values showed a significant difference between 2006 and 2017 analyses ($t = 8.719$, $p < 0.0001$), although the mean difference between the analysis dates was reduced from $-0.79‰$ to $-0.49‰$.

The significant difference between the two analysis dates was further investigated by renormalizing 2017 $\delta^{18}O$ values to the VPDB scale using an offset correction as opposed to a stretch and shift correction. A paired *t*-test of $\delta^{18}O$ values again showed a significant difference between the two analysis dates when the 2017 δ values were normalized using NBS 19 only; this was true if either the original 2006 $\delta^{18}O$ values were used in the comparison ($t = 10.85$, $p < 0.0001$) or the α-corrected 2006 $\delta^{18}O$ values were used in the comparison ($t = 5.707$, $p < 0.0001$); data are shown in Figure 4. However, the mean difference was smaller when the offset-normalized 2017 $\delta^{18}O$ values were compared with the α-corrected 2006 $\delta^{18}O$ values than when the offset-normalized 2017 $\delta^{18}O$ values were compared with the original 2006 $\delta^{18}O$ values: $-0.36‰$ versus $-0.66‰$, respectively.

Interpreting significant differences

The Regan (2006) data set and the results of this current study were used to describe RIDs for the $\delta^{13}C$ and $\delta^{18}O$ values of dental remains. RIDs are similar to the minimum meaningful differences (MMD) calculated by Pestle *et al*. for comparing bone bioapatite δ values, which were described as 'an empirically derived threshold by which the significance of values obtained in different

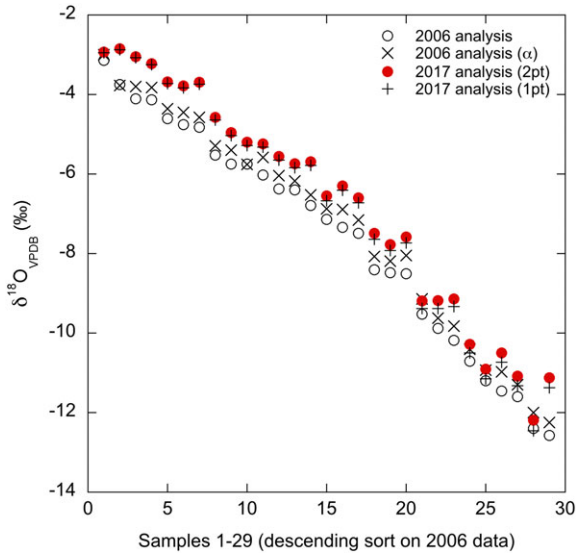

Figure 4 *Comparison of $\delta^{18}O$ values of 29 tooth enamel bioapatite samples analysed in 2006 and again in 2017. The 2006 data are presented as originally published (open circles) and following α-correction for temperature of acidification ('α', crosses). The 2017 data are normalized to the VPDB δ scale via a stretch and shift correction ('2 pt', closed circles) as well as an offset correction ('1 pt', pluses). Samples are sorted by the original 2006 δ values and assigned identifiers from 1 to 29. [Colour figure can be viewed at wileyonlinelibrary.com]*

laboratories might be judged' (Pestle *et al*. 2014, 2). The MMD for bioapatite carbon as published by Pestle *et al*. was 1.2‰, while the MMD for bioapatite oxygen was 3.1‰—but these were based on bone, which is thought to be more susceptible to diagenesis that could impact δ values than tooth enamel (Lee-Thorp 2002); in addition, results used to calculate MMD by Pestle *et al*. (2014) included data from bone bioapatite prepared using different methods and measured using different isotope analysis techniques. This current study focused on enamel bioapatite and only the isotope analysis technique—and not the sample preparation method—differed between 2006 and 2017 δ values.

To calculate MMD, Pestle *et al*. (2014) used four times the SD of both laboratories in any given comparison, stating that this would account for 95% of the laboratory error. However, twice the SD accounts for approximately 96% of the variation within a normally distributed sample and accomplishes the same goal. The RID for tooth bioapatite $\delta^{13}C$ values can thus be defined by halving the published MMD of 1.2‰ to 0.6‰. Omitting the outlier AFA-005, no individual carbon Δ calculated between the two analysis dates was > 0.6‰; indeed, the largest absolute Δ was 0.40‰ (Table 1). For oxygen, halving the MMD of 3.1‰ to 1.6‰ defines that as the RID for tooth bioapatite $\delta^{18}O$ values. Omitting the outlier, AFA-005, no sample had an oxygen Δ between 2006 and 2017 that was > 1.6‰ (Table 1).

CONCLUSIONS

This study compared carbon and oxygen isotopic compositions of 30 bioapatite (tooth enamel) samples measured by two facilities ≥ 10 years apart. A single sample outlier was found upon examination of differences between 2006 and 2017 isotope analyses, which may have been caused by sample contamination. This observation suggests an inherent error rate of approximately 3%

for any future use of this combined data set. The results of this validation study were used to define RIDs for tooth enamel bioapatite carbonate isotope analysis as 0.6‰ for $\delta^{13}C$ values and 1.6‰ for $\delta^{18}O$ values. These RIDs can be used as the maximums for any isotopic variation between repeated sample analyses, and if the variation exceeds these limits, the sample(s) should be discarded. The number of discarded samples will inform the user about the amount of error inherent in an amalgamated isotope data set.

For others interested in validating published isotope data and using combined data sets of the isotopic compositions of human remains from the archaeological, anthropological and/or forensic literature, the following good practice guidelines are recommended:

- Assess comparability between δ values by reanalyzing a statistically significant fraction of samples at another laboratory. Any isotopic differences observed should be thoroughly investigated—and addressed—before amalgamating data. RIDs should be employed, and the error inherent in the combined isotope data set should be reported. Tested data sets that are highly disparate or those that cannot be adequately assessed for comparability should *not* be combined.
- When using isotope data from the published literature, careful attention should be paid to the methods used in sample preparation. This is especially true for human skeletal remains, where the method used in sample preparation can have a significant effect on measured δ values (Pestle *et al*. 2014). Based on the current understanding of isotopic variation introduced during preparation, amalgamation of isotope data from only samples prepared in a similar manner is recommended.
- Careful attention should be paid to the methods used in isotope data handling. Details on data handling should be fully reported with each published isotope data set. Some guidelines on reporting isotope data are available for archaeology (Szpak *et al*. 2017; Roberts *et al*. 2018) and forensics (Dunn *et al*. 2017).

ACKNOWLEDGEMENTS

The USAFA Institutional Review Board (IRB) granted IRB exempt status to donated contemporary teeth from the US Air Force Academy (HQ USAFA IRB FAC2005026H); for details, see Regan (2006). The authors thank Dr Jason Curtis, University of Florida, for assistance with sample analysis in 2005/06. Mr Michael Lott, Mr Thuan Chau and Mr John Howa, IsoForensics, Inc., assisted with sample and data analysis in 2017/18. Ms Amelia Edwards, DPAA Laboratory, helped with logistics during sample handing and transfer. DPAA Laboratory Director, Dr John Byrd, and Science Director, Dr Philip Berran, were particularly supportive of this project, for which the authors are grateful. An anonymous reviewer provided helpful input that improved the manuscript text and data presentations.

REFERENCES

Bartelink, E. J., Berg, G. E., Beasley, M. M., and Chesson, L. A., 2014a, Application of stable isotope forensics for predicting region of origin of human remains from past wars and conflicts, *Annals of Anthropological Practice*, **38**(1), 124–36.

Bartelink, E. J., Berry, R., and Chesson, L. A., 2014b, Stable isotopes and human provenancing, in *Advances in forensic human identification* (eds. X. Mallett, T. Blythe, and R. Berry), 165–92, CRC Press, Boca Raton, FL.

Bell, S., 2009, Forensic chemistry, *Annual Review of Analytical Chemistry*, **2**(1), 297–319.

Cerling, T. E., Barnette, J. E., Bowen, G. J., Chesson, L. A., Ehleringer, J. R., Remien, C. H., Shea, P., Tipple, B. J., and West, J. B., 2016, Forensic stable isotope biogeochemistry, *Annual Review of Earth and Planetary Sciences*, **44**(1), 175–206.

Chesson, L. A., Tipple, B. J., Youmans, L. V., O'Brien, M. A., and Harmon, M. M., 2018, Forensic identification of human skeletal remains using isotopes: A brief history of applications from archaeological dig sites to modern crime scenes, in *New perspectives in forensic human skeletal identification* (eds. K. Latham, E. J. Bartelink, and M. Finnegan), 157–73, Elsevier, London, UK.

Coplen, T. B., 2011, Guidelines and recommended terms for expression of stable-isotope-ratio and gas-ratio measurement results, *Rapid Communications in Mass Spectrometry*, **25**(17), 2538–60.

DeNiro, M. J., and Epstein, S., 1976, You are what you eat (plus a few permil): The carbon isotope cycle in food chains, *Geological Society of American Abstracts with Programs*, **8**, 834–5.

Dunn, P., and Carter, J., 2018, *Good practice guide for isotope ratio mass spectrometry*, Second edn, available online: http://www.forensic-isotopes.org/gpg.html

Dunn, P. J. H., Hai, L., Malinovsky, D., and Goenaga-Infante, H., 2015, Simple spreadsheet templates for the determination of the measurement uncertainty of stable isotope ratio delta values, *Rapid Communications in Mass Spectrometry*, **29**(22), 2184–6.

Dunn, P. J. H., Salouros, H., Carter, J. F., and Doyle, S. P., 2017, Forensic application of stable isotope delta values: Proposed minimum requirements for method validation, *Rapid Communications in Mass Spectrometry*, **31**(17), 1476–80.

Emanovsky, P. D., and Belcher, W. R., 2012, The many hats of a recovery leader: Perspectives on planning and executing worldwide forensic investigations and recoveries at the JPAC Central Identification Laboratory, in *A companion to forensic anthropology* (ed. D. C. Dirkmaat), 567–92, John Wiley & Sons, Ltd.

Fogel, M. L., Tuross, N., Johnson, B. J., and Miller, G. H., 1997, Biogeochemical record of ancient humans, *Organic Geochemistry*, **27**(5–6), 275–87.

Friedman, I., and O'Neil, J. R., 1977, Compilation of stable isotope fractionation factors of geochemical interest, U.S. Geological Survey.

Gaffney, J. S., Irsa, A. P., Friedman, L., and Slatkin, D. N., 1978, Natural $^{13}C/^{12}C$ ratio variations in human populations, *Biological Mass Spectrometry*, **5**(8), 495–7.

Holland, T., Byrd, J., and Sava, V., 2008, Joint POW/MIA Accounting Command's Central Identification Laboratory, in *The forensic anthropology laboratory* (eds. M. Warren, H. Walsh-Haney, and L. Freas), 47–63, CRC Press.

Holland, T. D., Berg, G. E., and Regan, L. A., 2012, Identification of a United States airman using stable isotopes, *Proceedings of the American Academy of Forensic Sciences*, **18**, 420–1.

Kamenov, G. D., and Curtis, J. H., 2017, Using carbon, oxygen, strontium, and lead isotopes in modern human teeth for forensic investigations: A critical overview based on data from Bulgaria, *Journal of Forensic Sciences*, **62**(6), 1452–9.

Kimmerle, E. H., and Kamenov, G. D., 2015, Linking identity with landscape: Osteological and Sr-Pb isotopic methods for biogeoreference, in *Biological affinity in forensic identification of human skeletal remains* (eds. G. E. Berg and S. C. Ta'ala), 239–56, CRC press.

Komar, D. A., and Buikstra, J. E., 2008, *Forensic anthropology: Contemporary theory and practice*, Oxford University Press.

Kusaka, S., and Nakano, T., 2014, Carbon and oxygen isotope ratios and their temperature dependence in carbonate and tooth enamel using a GasBench II preparation device: Letter to the editor, *Rapid Communications in Mass Spectrometry*, **28**(5), 563–7.

Lee-Thorp, J. A., 2002, Two decades of progress towards understanding fossilization processes and isotopic signals in calcified tissue minerals, *Archaeometry*, **44**(3), 435–46.

Lightfoot, E., and O'Connell, T. C., 2016, On the use of biomineral oxygen isotope data to identify human migrants in the archaeological record: Intra-sample variation, statistical methods and geographical considerations, *PLoS ONE*, **11**(4), e0153850.

Lyon, T. D. B., and Baxter, M. S., 1978, Stable carbon isotopes in human tissues, *Nature*, **273**(5665), 750–1.

Meier-Augenstein, W., and Schimmelmann, A., 2019, A guide for proper utilisation of stable isotope reference materials, *Isotopes in Environmental and Health Studies*, **55**(2), 113–28.

Passey, B. H., Cerling, T. E., and Levin, N. E., 2007, Temperature dependence of oxygen isotope acid fractionation for modern and fossil tooth enamels, *Rapid Communications in Mass Spectrometry*, **21**(17), 2853–9.

Pestle, W. J., Crowley, B. E., and Weirauch, M. T., 2014, Quantifying inter-laboratory variability in stable isotope analysis of ancient skeletal remains, *PLoS ONE*, **9**(7), e102844.

Pollard, A. M., Pellegrini, M., and Lee-Thorp, J. A., 2011, Technical note: Some observations on the conversion of dental enamel $\delta^{18}O_p$ values to $\delta^{18}O_w$ to determine human mobility, *American Journal of Physical Anthropology*, **145**(3), 499–504.

Regan, L. A., 2006, Isotopic determination of region of origin in modern peoples: Applications for identification of U.S. war-dead from the Vietnam conflict, Ph.D., University of Florida.

Roberts, P., Fernandes, R., Craig, O. E., Larsen, T., Lucquin, A., Swift, J., and Zech, J., 2018, Calling all archaeologists: Guidelines for terminology, methodology, data handling, and reporting when undertaking and reviewing stable isotope applications in archaeology, *Rapid Communications in Mass Spectrometry*, **32**(5), 361–72.

Someda, H., Gakuhari, T., Akai, J., Araki, Y., Kodera, T., Tsumatori, G., Kobayashi, Y., Matsunaga, S., Abe, S., Hashimoto, M., Saito, M., Yoneda, M., and Ishida, H., 2016, Trial application of oxygen and carbon isotope analysis in tooth enamel for identification of past-war victims for discriminating between Japanese and US soldiers, *Forensic Science International*, **261**, 166.e1–5–.

Swart, P. K., Burns, S. J., and Leder, J. J., 1991, Fractionation of the stable isotopes of oxygen and carbon in carbon dioxide during the reaction of calcite with phosphoric acid as a function of temperature and technique, *Chemical Geology: Isotope Geoscience section*, **86**(2), 89–96.

Szpak, P., Metcalfe, J. Z., and Macdonald, R. A., 2017, Best practices for calibrating and reporting stable isotope measurements in archaeology, *Journal of Archaeological Science: Reports*, **13**, 609–16.

van der Merwe, N. J., and Vogel, J. C., 1978, ^{13}C content of human collagen as a measure of prehistoric diet in woodland North America, *Nature*, **276**(5690), 815–16.

Vogel, J. C., and van der Merwe, N. J., 1977, Isotopic evidence for early maize cultivation in New York state, *American Antiquity*, **42**(2), 238–42.

SUPPORTING INFORMATION

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

Data S1. Supporting information