Preliminary Findings from a Visual Pair-Matching Study in a Large Commingled Assemblage

Carrie B. LeGarde[a,b*]

ABSTRACT: Pair-matching is an important component of commingled human remains (CHR) analysis, as it can help to limit the amount of DNA testing and contribute to minimum or most likely number of individual calculations. As commingled assemblages become larger, pair-matching becomes more difficult, and it is unknown if accuracy declines. Therefore, a study to determine the accuracy rates of visual pair-matching for multiple observers with a variety of education and experience levels was conducted on a large commingled assemblage. Because the complete results of this study will not be available until all the DNA testing has been completed, this article focuses on the parameters of the study and interobserver variability in pair-matching as well as the current status of the results. In addition, useful morphology of the humerus for pair-matching is identified.

The sample consisted of left and right humeri ($n = 287$ and $n = 293$, respectively) from the commingled remains of the USS *Oklahoma*, which are currently being segregated and identified by the Defense POW/MIA Accounting Agency. Five anthropologists completed this study with human osteological experience ranging from two to thirteen years, CHR experience ranging from no experience to four years, and the following education levels represented: post-BA ($n = 1$), post-MA ($n = 1$), and post-PhD ($n = 3$).

Those with more CHR experience were more likely to agree than those with PhDs. Participants with the most CHR experience had substantial agreement, spent the most time pair-matching, and had high accuracy.

KEYWORDS: forensic anthropology, humerus, pair-matching, commingled human remains

Introduction

Commingled human remains (CHR) pose a particularly difficult challenge for identification, especially when the commingling is on a large scale. DNA analysis is often heavily relied upon to segregate these remains into discrete individuals, but this can be time consuming and costly. Anthropological methods, such as pair-matching, should be employed when possible to minimize the number of elements that require DNA sampling. When commingled assemblages are represented by relatively small numbers of individuals, visual pair-matching is not a cumbersome task. However, as these assemblages become larger this task becomes more difficult, and it is unknown if accuracy declines. Therefore, a study to determine the accuracy rates of visual pair-matching for multiple observers with a variety of education and experience levels was conducted on a large commingled assemblage. Because the complete results of this study will not be

available until all the DNA testing has been completed, this article is primarily a discussion of the parameters of the study and interobserver variability in pair-matching, useful morphology for pair-matching the humerus, and the current status of the results.

Pair-matching is an important step in determining the minimum number of individuals (MNI) or the most likely number of individuals (MLNI) of a commingled assemblage, which can help with planning and resource allocation and may drive analytical approaches (Konigsberg & Adams 2014; Palmiotto et al. 2019). Konigsberg and Adams (2014) conducted a test of the accuracy of visual pair-matching for the humerus, femur, and tibia. A random sample was selected from two populations, one representing 15 individuals and the second representing 30 individuals, with a 60% recovery rate so there would not be an equal number of pairs. Their test found that visual pair-matching could be accurately performed by an experienced osteologist. The errors that occurred were from overlooking true pairs, which occurred only for the humeri in the smaller sample but for both the tibiae and humeri in the larger sample. This was an important study for investigating the accuracy of visual pair-matching, but it included only one participant (an experienced osteologist) and was with small to moderate commingled assemblages. The current study seeks to expand upon this research by utilizing a large commingled assemblage.

[a]Defense POW/MIA Accounting Agency, Offutt AFB, NE, USA
[b]University of South Florida, Anthropology, Tampa, FL, USA
*Correspondence to: Carrie B. LeGarde, Defense POW/MIA Accounting Agency, 106 Peacekeeper Dr., Bldg. 301, Offutt AFB, NE 68113, USA
E-mail: carrieblegarde@gmail.com

The humerus was chosen for this study for two reasons: (1) there is a relatively high degree of bilateral asymmetry exhibited as compared to other long bones, which could make visual pair-matching more difficult (Byrd 2008; LeGarde 2012); and (2) all left and right humeri in the assemblage were sampled for DNA, which allows for the accuracy of pairs to be determined. The focus on an element that exhibits bilateral asymmetry is important, because other sorting methods may depend on statistical methods, such as osteometric sorting (Byrd & LeGarde 2014; Lynch et al. 2018; Thomas et al. 2013), to assist in pair-matching.

Osteometric sorting is utilized in cases of commingling to sort remains based on size. For pair-matching, this method is based on the principle that left and right elements from a single individual will be similar in size, and a reference population is utilized to provide the baseline for normal variation (Byrd 2008; Byrd & LeGarde 2014). This method was developed to be a simple statistical tool with calculations that can easily be done by hand for one-to-one comparisons, and formulae are provided for use with one measurement (Thomas et al. 2013) or multiple measurements (Byrd 2008; Byrd & LeGarde 2014; Lynch et al. 2018). Osteometric sorting can also be used to create a "short list" of possible antimeres to a given element by doing multiple comparisons at once and excluding those that are most different in size (Lynch et al. 2018). This can save time in large commingled assemblages, since it can significantly reduce the number of visual comparisons. However, methods utilizing osteometrics may exclude true pairs exhibiting asymmetry, since these methods are based on the assumption of bilateral symmetry (i.e., left−right=0). Lynch et al. (2018) showed that using the mean left-right difference of the reference sample, rather than zero (proposed by Byrd 2008), was better and lowered the number of false negatives (i.e., excluding a true pair). This is more accurate, since it reflects the fluctuating asymmetry exhibited by the reference population rather than relying on the assumption of bilateral symmetry. Lynch et al. (2018) further suggested that using absolute values of the left-right difference (i.e., left−right=|D|), along with a half-normal data transformation, improved the method further, and this is currently the default in automated processes.

Although osteometric sorting can be a useful tool, the risk of rejecting true pairs because of asymmetry cannot be ignored, as it could have a negative impact on the resolution of a commingled project. It should be investigated whether these asymmetrical pairs would be found with a visual examination. Therefore, the frequency with which true pairs are visually matched but are rejected utilizing osteometric sorting (i.e., asymmetrical pairs) will be noted. Even when size differs significantly there may be particular features and morphology that are useful in determining a pair match. This has not been explored, particularly on a large scale with multiple observers; therefore, this study also captures details about the morphology used for pair-matching the humerus.

Materials and Methods

The sample for this study consists of the commingled skeletal remains of individuals recovered from the USS *Oklahoma,* which are currently being analyzed at the Defense POW/MIA Accounting Agency (DPAA). Based on historical documentation, these remains likely represent 394 individuals (Brown 2019). These individuals are male, between 17 and 52 years of age at death, and have documented ancestries of White, Black, and Asian American. For a full background on this assemblage, see Brown (2019).

After attempts to identify these individuals in the late 1940s were unsuccessful, the remains were buried in separate "bundles" within numerous caskets (Brown 2019; Harris 2010). These bundles represented skeletal elements that were thought to represent a single individual during original identification efforts. Therefore, antimeres from a single bundle, designated by the same "X-" number, are considered a "historical pair." However, these bundles have been shown to be highly commingled (Brown 2019; Brown et al. 2017), so historical pairs are not expected to be a "true pair" (i.e., correct pair). There are 207 humeri historical pairs in the USS *Oklahoma* assemblage. There are likely 14 additional historical pairs, but due to the poor preservation of three caskets the bundles became commingled and historical pairs cannot be determined. There were 117 humeri in bundles without an antimere (i.e., unpaired) and an additional 21 humeri from one bundle that consisted of only humeri. This "humeri bundle" is considered to be humeri that could not be associated to an individual at the time of original analysis and were therefore bundled together.

All humeri in this assemblage, except small unassociated fragments, were sampled for DNA and were included in this study (*N*=580; 287 left and 293 right). This included damaged and incomplete humeri, although the majority are greater than 75% complete and in excellent condition. Each humerus was labeled in ink when it was processed in the laboratory after exhumation in 2015. This label reflects the X-number, which represents the bundle the remains originated from (e.g., X-100A). A tag with the DNA sample number and a unique number, represented by the bundle number and a designator number (e.g., X-100A/201; see Fig. 1), was also attached to each humerus. The label on the bone and the attached tag were visible during the study. The participants were instructed to ignore both as much as possible. All humeri were set out on 10 laboratory tables for analysis, with left and right elements separated to save time for each participant.

Five participants (including the author) completed this study. Participants for this study were recruited at DPAA,

FIG. 1—*Exemplar humerus with X-number in ink and attached label providing the X-number and a unique number associated with this bone. The DNA sample number would typically be on the reverse side of this label.*

which included anthropologists on staff and interns. Each participant completed a questionnaire prior to beginning the study to obtain information regarding their highest level of attained education, human osteology experience, and CHR experience. Education ranged from completed bachelor's degree to completed doctoral degree (Table 1). All participants had at least 2 years of human osteology experience, but experience with commingled human remains ranged from zero to 4 years.

Participants were not a given particular methodology to follow, because part of this study was to investigate the

TABLE 1—*Background Information for Each Participant in the Study.*

Participant	Education	Osteology Experience	CHR* Experience
P.1	BA	4 years	None
P.2	PhD	2 years	<1 year
P.3	PhD	7 years	<1 year
P.4	PhD	13 years	3 years
P.5	MA	10 years	4 years

*CHR = commingled human remains.

different approaches people may take to visually pair-match on a large scale. Participants were only instructed to document their process, including the time spent pair-matching, and to determine pair matches for every humerus. This included four possible responses for every humerus: Match,[1] Probable Match, Possible Match, and No Match. The three match categories were used to allow for participants to match humeri with differing levels of confidence: Match = confident, Probable Match = fairly confident, and Possible Match = less confident. Although this leaves a level of ambiguity, as one participant may define "fairly confident" differently than another participant, it still provides another level of comparison.

The minimum number of possible outcomes for each participant is 293. This reflects pair-matching every possible left humerus and having six unmatched right humeri. The maximum number of possible outcomes for each participant is 580, which reflects no matches. If each participant paired every possible humerus, the total number of outcomes from all participants could range from 293 (all participants agreeing) to 1,465 (all participants disagreeing). There are over 40,000 possible combinations for matching every humerus, thus complete agreement is highly unlikely. The outcomes are compared to determine the level of agreement between participants. For every humerus, participants can agree or disagree in two ways: agree = same match or no match; disagree = different match or one says no match. For an analysis of overall agreement and disagreement between participants, Fleiss' kappa (K) was calculated using Microsoft Excel, since this statistic can accommodate multiple independent variables and participants and nominal data (Fleiss 1971; Hayes and Krippendorff 2007). Agreement was determined for left and right humeri separately, since K cannot accommodate the "no match" category for both left and right in one analysis. They are expected to be nearly the same, since they overlap on matches and only differ in the determination of no matches. The strength of the agreement for K is generally interpreted as the following: ≤0 = poor, 0.01−0.20 = slight, 0.21−0.40 = fair, 0.41−0.60 = moderate,

1. This category was designated as "Match" to participants in the study, but for clarity in reporting and discussing the categories and their results, it is considered "Confident Match" for the rest of the article.

0.61 – 0.80 = substantial, and 0.81 – 1 = almost perfect (Landis & Koch 1977; Sim & Wright 2005).

Osteometric sorting was done for each match that all participants agreed upon to compare the outcomes of a statistical test and visual assessment. The half-normalized transformation and absolute value *D* methodology proposed by Lynch et al. (2018) was used with an alpha level of 0.05, and all standard measurements available for each humerus. This is the standard methodology utilized by DPAA for creating short lists for pair-matching.

Mitochondrial DNA (mtDNA) results are used to determine the accuracy of the visual pair matches, with humeri yielding the same mtDNA sequence a confirmed pair match. However, numerous individuals can have the same mtDNA sequence, so the accuracy of all matches cannot be determined via mtDNA alone. Therefore, a pair match from one of these mtDNA sequences is considered a "potential" match. It is likely that these are true pairs, so they are considered a correct match in determining participant pair-matching accuracy, but are explicitly noted. Although the minority of mtDNA sequences for the USS *Oklahoma* are represented by more than one individual (19%), nearly half of the humeri fall into these sequences (172/399, or 43%). The accuracy of these "potential" pair matches will be determined once all mtDNA testing has been completed and segregation of the USS *Oklahoma* remains has been completed. As of August 15, 2018, mtDNA analysis is 69% (399/580) complete for the humeri in this study.

Results

Methodology was similar for all participants. All participants noted that size was the initial factor for comparing antimeres, followed by general shape and robusticity. All participants stated that the shape of the capitulum and/or trochlea was an important feature for comparison. Age indicators or muscle attachment sites, such as the deltoid tuberosity and medial epicondyle, were other commonly noted features.

Since the humeri were already sorted by side prior to each participant beginning the study, the first step that all participants did was to sort by size. This was primarily done in two ways: continuous sorting from smallest to largest and marking incremental measurements (e.g., 320 mm) with only

the lefts (Participants 1, 2, and 5), or discrete grouping (e.g., 320–330 mm) for both lefts and rights (Participants 3 and 4). For those who created a continuous line of smallest to largest, one right humerus was checked against left humeri from approximately 20 mm below to 20 mm above the right humerus length, even if a potential match was found. For those who created discrete groups, the left and right humeri within the same measurement group (e.g., 320–329 mm) were compared to each other. Participants 1 and 5 did not remove any type of match so that all lefts and rights were compared within a given length range. Participants 2 and 3 removed Confident Matches as they were found, so no further comparisons were made with these humeri. Participant 4 did not note whether any type of match was removed from consideration during the analysis.

After comparing all the right humeri to the left humeri, participants checked through unpaired humeri once more. However, two participants specifically noted that they ran out of time to complete this second check. It is possible that this may lead to a higher number of false negatives (i.e., a left humerus noted as "No Match" but actually having a pair) for these two participants. The amount of time required to complete the pair-matching assessment ranged from 35 to 55 hours, which occurred over approximately three weeks for most participants. This time did not include sorting the humeri by side as this was done before each participant began the study, which saved a considerable amount of time.

The total number of outcomes, which include any match or no match, for each participant ranged from 365 to 425 (Table 2). Participants 1, 2, and 5 have nearly the same number of outcomes (365, 375, and 369, respectively), while Participant 3 has the most (425) and Participant 4 falls in between (395). The higher number of outcomes for Participants 3 and 4 is likely because they did not have any Possible Matches.

Participants 2 through 5 each found approximately 150 Confident Matches, which is nearly twice as many as Participant 1 who noted nearly the same number of Confident Matches and Probable Matches (Table 2). This indicates that participant 1, who had the least amount of CHR experience, was the most conservative. Participants 4 and 5 had the most osteological and commingled experience and spent the most time pair-matching.

When comparing matches, of any category of one participant to another, there is considerable agreement, ranging

TABLE 2—*Results for Each Participant in the Study.*

Participant	Time Spent Matching	Confident Match	Probable Match	Possible Match	No Match	Total Outcomes
P.1	47.0 hrs	88	80	47	150	365
P.2	35.5 hrs	154	30	21	170	375
P.3	35.0 hrs	147	8	0	270	425
P.4	55.0 hrs	156	14	0	225	395
P.5	50.0 hrs	144	32	35	158	369

from 96 (Participants 3 and 4) to 140 (Participants 1 and 5) agreed-upon pairs (Table 3). Participant 1 paired the most humeri, so it is not surprising that this analyst had the highest noted agreed-upon pairs. There are also instances where both participants determined that a particular humerus had a pair, but they disagreed on the antimere. It is interesting to note that Participants 4 and 5, who have the most CHR experience, disagreed on only 14 pairs, which was the lowest number of disagreed-upon pairs. Occurring more often than pair disagreement was one participant pairing a humerus while the other said there was no match. Agreement between most participants was moderate, with Fleiss' kappa ranging from 0.432 to 0.604 (Table 3). Only participants 4 and 5 showed substantial agreement (K = 0.614 and 0.622 for left and right, respectively). Participant 3 had the lowest level of agreement when compared to other participants, which is partially due to having the lowest number of matches (Table 3). There was moderate agreement between participants 2, 3, and 4, who all have a PhD (K = 0.510 and 0.482 for left and right, respectively). This was lower than all between-participant comparisons, except for those with participant 3 (Table 3).

All outcomes were compared and examined for agreement between participants. There were a total of 794 different outcomes: 389 different pair matches and 405 humeri with no match (Table 4). The overall agreement between participants is moderate, with a Fleiss' kappa of 0.587 and 0.596 for left and right, respectively. All participants agreed on 103 outcomes: 36 no matches (19 left and 17 right humeri) and 67 matches (of any level). As of 15 August 2018, 54 (80.6%) of these agreed upon matches have yielded mtDNA results and have been confirmed as true pairs. In addition, 40 (59.7%) of these congruent matches are also historical

TABLE 3—*Number of agreed-upon pairs between participants (*bottom*, white cells) and the agreement (K) between participants for left/right humeri (*upper*, gray cells). The center diagonal numbers in bold italics represent the total number of pairs identified by each participant.*

	P.1	P.2	P.3	P.4	P.5
P.1	*215*	.522/.520	.432/.485	.561/.566	.566/.593
P.2	124	*205*	.457/.431	.560/.540	.604/.593
P.3	101	97	*155*	.505/.521	.502/.514
P.4	126	121	96	*170*	.614/.622
P.5	140	139	108	137	*211*

TABLE 4—*The Number of Participants Agreeing on Matches and No Matches.*

Number of Participants	Matches	No Match	Total	Percent
5 (All agree)	67	36	103	13.0
4	60	61	121	15.2
3	40	78	118	14.9
2	39	101	140	17.6
1 (No agreement)	183	129	312	39.3
Total	389	405	794	100

TABLE 5—*Number of Matches with mtDNA Results by Participant.*

Participant	Confident Matches	Number Incorrect	Number Correct*	% Correct
Historical pairs	132	63	69 (26)	52
P.1	63	1	62 (22)	98
P.2	94	11	83 (35)	88
P.3	87	14	73 (29)	84
P.4	101	7	94 (39)	93
P.5	93	1	92 (35)	99

*Number in parentheses indicates the number of the total that are potential pairs.

pairs, which suggests that these have the most distinct morphology. Osteometric sorting excluded four (6%) of these congruent matches as a "statistical pair," and therefore they would not have shown up on a short list of possible antimeres. This is slightly greater than what would be expected for Type II errors with an alpha level of 0.05.

The total number of humeri pairs based on mtDNA results is 133, of which 56 are potential pairs (i.e., from an mtDNA sequence with MNI > 1) and 133 do not currently have a pair. These numbers, as well as the accuracy rates given below, will continue to change as mtDNA results are obtained, but they are provided here to show the current accuracy of each participant. A Confident Match is considered correct if both humeri have the same mtDNA sequence. The accuracy rate is the percent of correct Confident Matches out of the total number of Confident Matches for each participant (Table 5). Only accuracy results for those categorized as a Confident Match are shown here, because this relates to pair matches that analysts are most confident in and would be comfortable identifying them as belonging to a single individual. Accuracy for Confident Matches ranges from 84% to 99% (73/87; 92/93) for the participants in this study, and the historical pairs are currently 52% (69/132) correct.

Discussion

This study demonstrates that pair-matching in a large commingled skeletal assemblage is feasible and well worth the time required. In general, participants took a full 40-hour week to pair-match the humeri, which spanned over multiple weeks for most participants. However, this is significantly less time than DNA testing, which can take months or years, depending on the number of samples and the resources and workload of the DNA laboratory. Pair-matching is an effective method in determining the MLNI represented by an assemblage (Konigsberg & Adams 2014), and it is recommended to do at the beginning of a project to reduce the number of samples submitted for DNA testing, if applicable, and therefore save time and money (Palmiotto et al. 2019).

The methodology of all participants shows the importance of size for pair-matching. The first step for all participants was

to sort by maximum length of the humerus, with some participants using robusticity as a secondary sorting parameter within length groups. Since statistical methods rely on size, this study illustrates the usefulness of osteometric sorting for pair-matching as a first step in creating smaller groupings for visual comparisons. To account for possible asymmetry and ensure the true pair is considered, it may be necessary to be highly conservative and utilize a 99% confidence interval.

This study shows an overall moderate level of agreement between participants. The two participants with the most CHR experience showed substantial agreement ($K = 0.614$ [left] and 0.622 [right]), while those with a PhD showed moderate agreement ($K = 0.510$ [left] and 0.482 [right]). This suggests that those with more CHR experience are more likely to agree on pair-matching than those with a more advanced degree.

Although there appears to be a fairly large number of humeri where there was no agreement and only one participant chose a particular outcome (39.3%; see Table 4), there was still some level of agreement for most humeri. The majority of humeri had only two or fewer outcomes (426, or 73.4%), which shows that it was more common to have three or more participants agreeing on an outcome and then one or two participants choosing a different outcome. This was often the majority agreeing on a pair match, while the other participant(s) concluded no match, or vice versa. It was less common for it to be differing pair matches. In addition, of the 183 pairs with no agreement, 124 (67.8%) were Probable Matches or Possible Matches, indicating there was some level of uncertainty for the majority of these. There were only two humeri that had five different outcomes (i.e., each participant paired it to a different humerus), and only another 33 (5.7%) with four different outcomes. It is important to consider the large number of possible matches for each humerus when looking at agreement between participants. The agreement between participants was considered moderate, but the magnitude of kappa can be influenced by many categories because there is more potential for disagreement (Sim & Wright 2005). In addition, the cutoffs for determining the magnitude of kappa (i.e., moderate or substantial) are arbitrary and must be considered in the context of the variables (Landis & Koch 1977; Sim & Wright 2005). There were nearly 300 categories in this analysis, which could be driving kappa lower, so the overall agreement of 0.587 (left) and 0.596 (right) could be considered moderate to substantial.

The osteometric sorting results for the congruent matches showed that true pairs found during visual assessment were excluded in statistical analysis. This shows the value of visual comparisons. Although the use of the absolute value D and half-normalized transformation has been suggested as an improvement to lower the number of false negatives (i.e., a true pair excluded with osteometric sorting), using the absolute value of the left-right differences ignores

directional asymmetry, which shifts the mean and narrows the distribution. When using numerous measurements, this can cause minimal differences to become significant and lead to the rejection of an antimere that is nearly identical. Using the reference sample mean D rather than zero and/or fewer measurements may actually be preferable and will be explored further once all true pairs can be determined. In addition, a slightly higher percentage than expected of false negatives occurred (6%) when utilizing osteometric sorting; however, this was a relatively small sample with only 67 pairs tested and will also be investigated further once all mtDNA testing is complete.

These preliminary results show that visual pair-matching has relatively high accuracy, ranging from 84% to 99%, regardless of education or experience level; however, the accuracy does appear to increase with experience as well as time spent analyzing pairs. Participants 2 and 3 spent the least amount of time pair-matching (35.5 and 35 hours, respectively) and had the most incorrect Confident Matches (Table 5). They also both had less than one year of CHR experience. Although Participant 1 had no CHR experience and has high accuracy (98%), this participant was conservative when determining Confident Matches and took more time pair-matching. Similar to Konigsberg and Adams's (2014) pair-matching test, overlooking a true pair was more likely to occur than a false positive ("incorrect pair"). This is not surprising due to the large number of No Matches noted by all participants, which ranged from 150 to 270 (Table 2), and will be explored fully when mtDNA testing is complete.

A limitation of this study was the small number of participants. Due to time and space requirements, only five participants could fully complete the study. Two of these individuals stated that they could have spent additional time double-checking the unpaired humeri, including participant 5, who spent 50 hours pair-matching. This indicates that these participants are more likely to have false negatives because they may have noted a pair match during the review of unpaired humeri.

A source of potential bias in this study was the visible label and attached tag (Fig. 1) that indicated the specific bundle the humeri originated from (i.e., provenience). Two humeri with the same bundle number would indicate a "historical pair," and therefore someone in the past may have thought that they belonged to a single individual (Brown 2019; Harris 2010). The participants knew this, so this could have influenced the determination of a pair match, particularly if it was a match the participant was less confident about. However, the participants also knew that a high degree of commingling is present between bundles and that a historical pair does not indicate a true pair. In some cases, seeing two humeri with the same bundle number could make a participant more skeptical and not determine a match.

Conclusions

Space and time permitting, visual pair-matching on a large scale can be done to assist in the segregation of commingled human remains. Sorting humeri by size was an important first step taken by all participants, followed by comparisons of robusticity and morphology. All participants noted the capitulum and/or trochlea as an important morphological feature for pair-matching the humerus. Morphological comparisons were shown to be invaluable in cases of left-right size differences, as evidenced by true pairs identified via visual pair-matching being excluded with osteometric sorting.

Overall, the participants showed a moderate to substantial level of agreement. The participants with the most commingled human remains experience showed substantial agreement. They also spent more time pair-matching, which corresponded with higher pair-matching accuracy. Participants with a PhD had only moderate agreement, and the two with less than one year of CHR experience spent the least amount of time pair-matching and had the lowest accuracy. This suggests that CHR experience is a more important factor than education level for visual pair-matching agreement and accuracy. These results are preliminary, and accuracy rates may change as additional DNA results are received.

Acknowledgments

References

Brown CA. The USS *Oklahoma* Identification Project. *Forensic Anthropology* 2019;2(2):102–112.

Brown CA, LeGarde CB, Damann FE, Byrd JE. Assessing the accuracy of historical associations in a commingled assemblage. In: Proceedings of the 69th Annual Meeting of the American Academy of Forensic Sciences, February 13–18, 2017; New Orleans, LA.

Byrd JE. Models and methods for osteometric sorting. In: Adams BJ, Byrd JE, eds. *Recovery, Analysis, and Identification of Commingled Human Remains*. Totowa, NJ: Humana Press; 2008:199–220.

Byrd JE, LeGarde CB. Osteometric sorting. In: Adams BJ, Byrd JE, eds. *Commingled Human Remains: Methods in Recovery, Analysis, and Identification*. San Diego: Academic Press; 2014:167–191.

Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 1971;76(5):378–382.

Harris H. History of the sinking of USS *Oklahoma* and subsequent attempts to recover and identify her crew. Available at: http://www.public.navy.mil/bupers-npc/support/casualty/Documents/POW%20MIA/USS%20OKLAHOMA%20(BB-37).pdf . Created March 1, 2010. Updated May 4, 2015. Accessed June 1, 2018.

Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures* 2007;1(1):77–89.

Konigsberg LW, Adams BJ. Estimating the number of individuals represented by commingled human remains: A critical evaluation of methods. In: Adams BJ, Byrd JE, eds. *Commingled Human Remains: Methods in Recovery, Analysis, and Identification*. San Diego: Academic Press; 2014:193–220.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159–174.

LeGarde CB. *Asymmetry of the Humerus: The Influence of Handedness on the Deltoid Tuberosity and Possible Implications for Osteometric Sorting* [master thesis]. Missoula: University of Montana; 2012.

Lynch JJ, Byrd JE, LeGarde CB. The power of exclusion using automated osteometric sorting: Pair-matching. *Journal of Forensic Sciences* 2018;63(2):371–380.

Palmiotto A, Brown CA, LeGarde CB. Estimating the number of individuals in a large commingled assemblage. *Forensic Anthropology* 2019;2(2):129–138.

Sim J, Wright CC. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy* 2005;85(3):257–268.

Thomas RM, Ubelaker DH, Byrd JE. Tables for the metric evaluation of pair-matching of human skeletal elements. *Journal of Forensic Sciences* 2013;58(4):952–956.